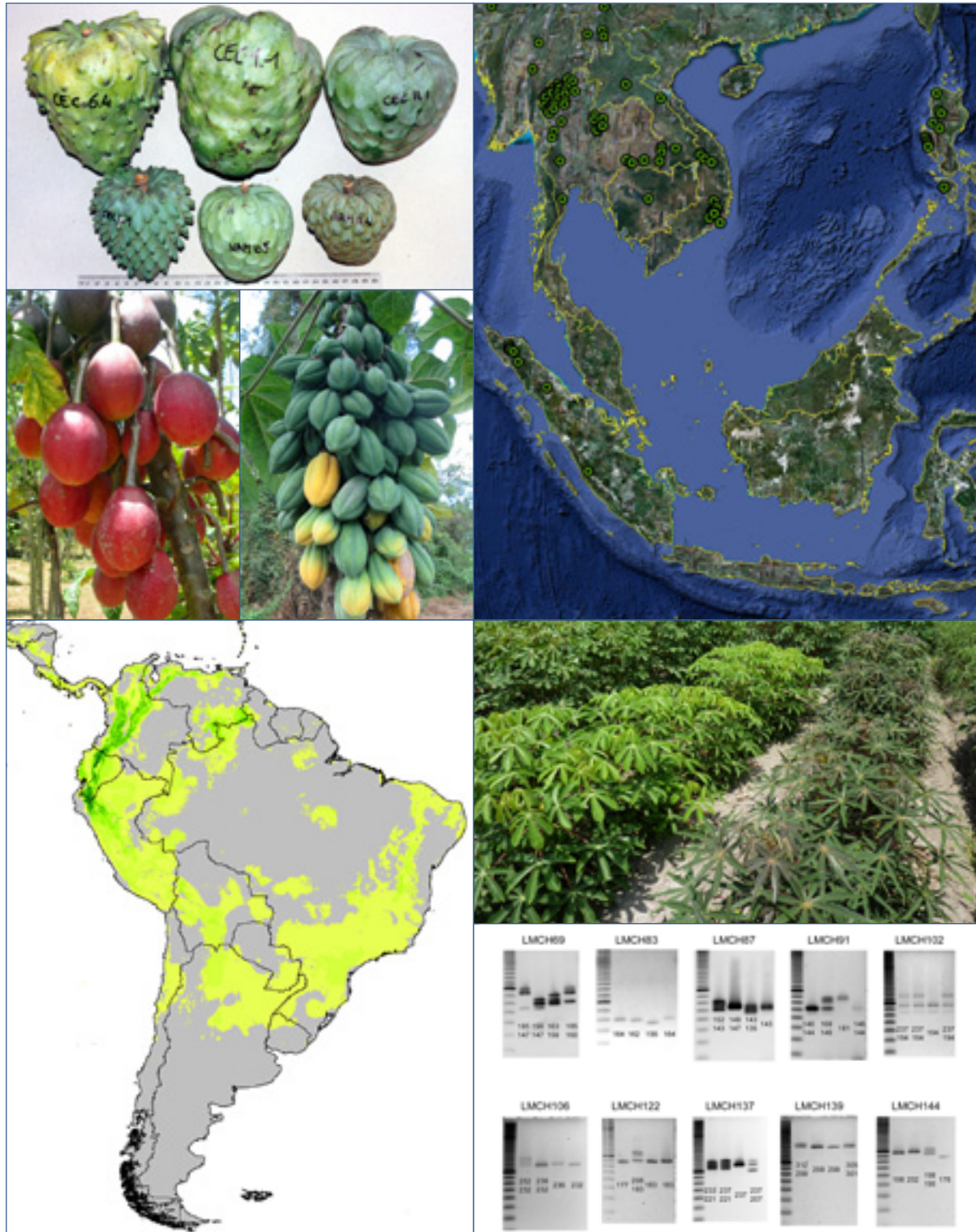


Manuel de formation à l'analyse spatiale de la diversité et de la distribution des plantes

Xavier Scheldeman et Maarten van Zonneveld



Bioversity International fait partie du Groupe Consultatif sur la Recherche Agricole Internationale (CGIAR), qui œuvre à réduire la faim, la pauvreté et la dégradation de l'environnement dans les pays en développement, en générant et en partageant des connaissances, technologies et politiques pertinentes en matière agricole. Cette recherche, focalisée sur le développement, est conduite par un Consortium de 15 centres CGIAR qui travaillent avec des centaines de partenaires dans le monde et qui sont financés par un fonds alimenté par de multiples bailleurs.

Le statut international de Bioversity lui est conféré par un accord d'établissement qui, en 2010, avait été signé par les gouvernements suivants : Algérie, Australie, Belgique, Bénin, Bolivie, Brésil, Burkina Faso, Burundi, Cameroun, Chili, Chine, Congo, Costa Rica, Côte d'Ivoire, Chypre, Cuba, Danemark, Equateur, Egypte, Ethiopie, Ghana, Grèce, Guinée, Hongrie, Ile Maurice, Inde, Indonésie, Iran, Israël, Italie, Jordanie, Kenya, Malaisie, Mali, Mauritanie, Maroc, Norvège, Oman, Ouganda, Pakistan, Panama, Pérou, Pologne, Portugal, République Tchèque, Roumanie, Russie, Sénégal, Slovaquie, Soudan, Suisse, Syrie, Tunisie, Turquie et Ukraine.

Les programmes de recherche de Bioversity reçoivent un appui financier de plus de 150 donateurs, incluant des gouvernements, des fondations privées et des organismes internationaux. Des renseignements complémentaires sur les donateurs et les activités de recherche sont donnés dans les rapports annuels de Bioversity, qui sont disponibles sous forme imprimée en adressant un courriel à biodiversity-publications@cgiar.org ou à partir du site Web www.biodiversityinternational.org.

Les appellations géographiques employées dans cette publication ainsi que la présentation de matériel ne sont en aucun cas le signe d'une opinion, quelle qu'elle soit, exprimée par Bioversity ou le CGIAR quant au statut légal d'un pays, d'un territoire, d'une ville ou une zone ou l'autorité qui les dirige, ou sur la délimitation de ses frontières géographiques ou administratives. De même, les opinions exprimées sont celles des auteurs et ne reflètent pas nécessairement celles de ces organisations.

La mention d'une marque déposée n'est faite qu'à titre d'information et ne constitue pas le cautionnement du produit.

Citation de la version originale en anglais : Scheldeman, Xavier and van Zonneveld, Maarten. 2010. Training Manual on Spatial Analysis of Plant Diversity and Distribution. Bioversity International, Rome, Italy. 179pp.
ISBN 978-92-9043-880-9

Citation de la version française : Scheldeman, Xavier et van Zonneveld, Maarten, 2012. Manuel de formation à l'analyse spatiale de la diversité et de la distribution des plantes. Bioversity International, Rome, Italie. 186 pp.
ISBN 978-92-9043-917-2

Bioversity International
Via dei Tre Denari, 472/a,
00057 Maccarese, Rome, Italie
Tél. : (39) 0661181

Illustrations de couverture : Crédits : Photographies : Xavier Scheldeman, Bioversity International;
Cartes : Maarten van Zonneveld, Bioversity International ; Images d'électrophorèse d'ADN en gel : Iñaki Hormaza, IHSM la Mayora, Consejo Superior de Investigaciones Científicas, Espagne.

© Bioversity International, 2012

Table des matières

REMERCIEMENTS	5
INTRODUCTION	6

SECTION A ÉLÉMENTS DE BASE ET PRÉPARATION DES DONNÉES

1	INSTALLATION DES PROGRAMMES ET DE DONNÉES D'EXEMPLE À ANALYSER	11
1.1	TÉLÉCHARGEMENT ET INSTALLATION DE DIVA-GIS	11
1.1.1	Comment installer DIVA-GIS	12
1.2	INSTALLATION DE MAXENT	16
1.2.1	Comment installer Java	16
1.2.2	Comment installer Maxent	17
1.3	INSTALLATION DE GOOGLE EARTH	18
1.3.1	Comment installer Google Earth	18
1.4	DONNÉES À ANALYSER	20
2	PRÉPARER ET IMPORTER LES DONNÉES VERS DIVA-GIS ET MAXENT	21
2.1	PRÉPARATION ET IMPORTATION DES POINTS DE PRÉSENCE	24
2.1.1	Comment convertir des données du format Degrés, Minutes, Secondes (<i>Degrees, Minutes, Seconds</i> [DMS]) vers le format Degrés Décimaux (<i>Decimal Degrees</i> [DD])	25
2.1.2	Comment importer dans DIVA-GIS des points de présence géoréférencés	28
2.1.3	Comment importer dans Maxent des points de présence géoréférencés	29
2.2	IMPORTATION DE DONNÉES CLIMATIQUES DANS DIVA-GIS ET MAXENT	34
2.2.1	Comment importer des données climatiques dans DIVA-GIS	35
2.2.2	Comment importer des données climatiques dans Maxent	36
2.3	SOURCES DE DONNÉES SPATIALES ET AUTRES DONNÉES PERTINENTES	38
3	ÉLÉMENTS DE BASE DE L'ANALYSE SPATIALE DANS DIVA-GIS	41
3.1	VISUALISATION DANS DIVA-GIS	41
3.1.1	Comment effectuer des visualisations de base à l'aide des fichiers vecteur	43
3.1.2	Comment réaliser des visualisations simples en utilisant des rasters	50
3.1.3	Comment combiner des rasters	57
3.1.4	Comment extraire des valeurs de rasters en se basant sur des points de présence	60
3.1.5	Comment créer des couches climatiques personnalisées	62
3.1.6	Comment importer des données climatiques dans DIVA-GIS	66
3.1.7	Comment créer des fichiers CLM dans DIVA-GIS	69

3.2	EXPORTER DES COUCHES VERS GOOGLE EARTH	74
3.2.1	Comment exporter des données vers Google Earth	74
3.3	EDITION DE CARTES ET FINALISATION D'UN PROJET	77
4	CONTRÔLE QUALITÉ	79
4.1	CONTRÔLE QUALITÉ FONDÉ SUR L'INFORMATION RELATIVE AUX UNITÉS ADMINISTRATIVES	80
4.1.1	Comment vérifier la qualité des données en se fondant sur les données d'unité administrative du passeport	80
4.2	CONTRÔLE QUALITÉ PAR L'IDENTIFICATION DE POINTS ATYPIQUES	87
4.2.1	Comment identifier des valeurs atypiques sur la base des données environnementales	88
 SECTION B		
ANALYSE DES DONNÉES		
5	ANALYSE SPATIALE DE LA DIVERSITÉ EN VUE DE PLANIFIER LA CONSERVATION	97
5.1	RICHESSSE DES ESPÈCES	99
5.1.1	Comment réaliser une analyse spatiale de la richesse spécifique	99
5.2	ANALYSE DE LA DIVERSITÉ INTRA-SPÉCIFIQUE BASÉE SUR DES DONNÉES PHÉNOTYPIQUES	111
5.2.1	Comment réaliser une analyse de diversité spatiale en utilisant des données phénotypiques	112
5.3	ANALYSE DE DIVERSITÉ INTRA-SPÉCIFIQUE FONDÉE SUR LES DONNÉES DE MARQUEURS MOLÉCULAIRES	118
5.3.1	Comment réaliser une analyse spatiale de diversité à base de données de marqueurs moléculaires	119
5.4	IMPLICATIONS POUR LA FORMULATION DE STRATÉGIES DE CONSERVATION	132
5.4.1	Comment identifier les zones prioritaires pour la conservation <i>in situ</i> ou pour la collecte de matériel génétique	133
6	MODÉLISATION ET ANALYSE DE LA DISTRIBUTION D'ESPÈCES	139
6.1	ANALYSE DE LA NICHE EFFECTIVE D'UNE ESPÈCE	140
6.1.1	Comment analyser et comparer les niches effectives de différentes espèces	141
6.2	MODÉLISATION DE LA DISTRIBUTION POTENTIELLE D'UNE ESPÈCE	147
6.2.1	Comment modéliser la distribution naturelle potentielle d'une espèce végétale	148
6.3	MODÉLISATION DE L'IMPACT DU CHANGEMENT CLIMATIQUE SUR LA DISTRIBUTION DES ESPÈCES	159
6.3.1	Comment évaluer l'impact du changement climatique sur la distribution des espèces	159
6.4	IDENTIFICATION DES LACUNES DANS LES COLLECTIONS D'ESPÈCES DE PLANTES SAUVAGES	167
6.4.1	Comment identifier des lacunes possibles dans les collections	167
 ANNEXE - LISTE DES COMMANDES EN ANGLAIS ET LEURS ÉQUIVALENTS EN FRANÇAIS		181

Remerciements

Les auteurs sont redevables à de nombreux collègues et pairs de Bioversity et d'autres organismes pour leur importante contribution à l'élaboration de ce manuel en anglais, en espagnol et en français.

Nous aimerions remercier Nora Castañeda (Bioversity/CIAT) qui, avec sa maîtrise des méthodes de SIG permettant de conduire l'analyse spatiale de données sur la biodiversité, a participé à la phase initiale d'élaboration. Nous aimerions également remercier tout particulièrement Colm Bowe (Centre international des espèces agricoles sous-utilisées, International Centre for Underutilized Crops, Royaume-Uni) dont les commentaires sur les premières versions du manuel ont contribué à établir sa forme actuelle et Jesús Salcedo pour son précieux travail de mise en forme du manuel.

Nous sommes particulièrement reconnaissants aux « pionniers de DIVA-GIS », Robert Hijmans (Université Davis de Californie, Etats-Unis), Luigi Guarino (Global Crop Diversity Trust, Italie) et Andy Jarvis (Centro Internacional de Agricultura Tropical, CIAT, Colombie), pour leurs commentaires constructifs. Leur soutien et leurs encouragements ont fortement motivé les auteurs pour avancer dans la préparation de ce manuel.

De nombreuses autres personnes ont aussi fourni des commentaires sur les diverses versions préliminaires du manuel. Nous aimerions exprimer notre gratitude à nos collègues de Bioversity, Karen Amaya, Margarita Baena, Michele Bozzano, Gea Galuzzi, Prem Mathur, Imke Thormann, Veerle Van Damme et Barbara Vinceti, ainsi qu'à Sixto Iman de l'Instituto Nacional de Investigación Agraria, INIA, Pérou et Diana Lara de la Corporación Nacional de Investigación y Fomento Forestal, CONIF, Colombie.

Nous aimerions aussi remercier Nicole Hoagland qui, avec sa précieuse assistance éditoriale, a considérablement amélioré la lisibilité du manuel.

Les exemples utilisés dans ce manuel ont été testés lors de nombreuses sessions de formation. Nous remercions les organisateurs et bailleurs de fonds de ces sessions organisées entre Mars 2006 et Novembre 2010 en Argentine (Bariloche), au Chili (Pucón), en Colombie (Cali et Cartagène), au Costa Rica (Turrialba), en Ethiopie (Addis Abeba), au Mali (Bamako) et au Mexique (Mérida) mais aussi tout spécialement les participants qui ont suivi ces formations ; leurs commentaires ont en effet grandement contribué à l'amélioration du contenu de ce manuel.

Enfin, la production de ce manuel n'aurait jamais été possible sans le soutien financier de l'Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, l'INIA d'Espagne (via le projet « Renforcement de la collaboration régionale pour la conservation et l'utilisation durable des ressources génétiques forestières en Amérique latine et en Afrique Sub-saharienne ») et de la Coopération autrichienne pour le développement, ADC en anglais, via le projet « Développement des capacités de formation et des ressources humaines pour la gestion de la biodiversité forestière ». Nous aimerions remercier Ricardo Alia et Santiago Martinez de l'INIA pour leurs commentaires positifs et encourageants durant tout le processus d'élaboration du manuel.

La traduction du manuel en français a été effectuée par Nathalie Chenin, traductrice indépendante. La révision et l'édition de la traduction ont été réalisées par Claudine Picq, Bioversity International. La mise en page a été financée par le projet de recherche du CGIAR sur le changement climatique, l'agriculture et la sécurité alimentaire (Climate Change, Agriculture and Food Security, CCAFS).

Introduction

La diversité végétale est vitale pour la survie et le bien-être de l'humanité. Nombre d'espèces végétales domestiquées sont cruciales pour la sécurité alimentaire mondiale, tandis que d'autres sont d'une grande importance dans des domaines tels que la production de bois ou de biocarburant. Outre les espèces cultivées, de nombreuses plantes sauvages jouent encore un rôle important dans la satisfaction des besoins locaux en nourriture, carburant, médicaments et matériaux de construction ; les espèces sauvages apparentées aux plantes cultivées ont aussi une grande importance pour les programmes d'amélioration des plantes. Il existe des centaines d'espèces et de variétés, aujourd'hui sous-utilisées, qui présentent des caractères intéressants pour satisfaire les besoins actuels et futurs. Et la valeur de nombreuses autres espèces végétales reste toujours à découvrir.

La Convention sur la diversité biologique (CDB), établie en 1992, appelle à une stratégie mondiale pour la conservation des plantes (CDB, 2009a). En plus de efforts de la CDB, le Plan d'action mondial (PAM) pour la conservation et l'utilisation durable des ressources phytogénétiques pour l'alimentation et l'agriculture (RFGAA) adopté en 1996 et le Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture (ITPGRFA) (FAO, 2009), entré en vigueur en 2004, ont été formulés pour focaliser les efforts sur le potentiel de la biodiversité agricole et son importance pour la production agricole. Ces cadres internationaux visent à accroître l'information et les actions en faveur de la conservation et de l'utilisation de la diversité végétale. Par exemple, l'Article 7 de la CDB appelle à identifier et faire un suivi de la biodiversité, avec une attention particulière sur les espèces et les variétés qui offrent le meilleur potentiel en matière d'utilisation durable, et qui requièrent des mesures urgentes de conservation (CDB, 2009b). De la même manière, au sein du GPA, l'Action prioritaire 1 appelle à un suivi et un inventaire accrus des RFGAA, tandis que l'Action Prioritaire 4 vise à promouvoir la conservation in situ des plantes sauvages apparentées aux espèces cultivées et des espèces sauvages utiles à la production alimentaire. En outre, l'Action prioritaire 7 recommande la collecte planifiée et ciblée des ressources phytogénétiques pour l'alimentation et l'agriculture. L'importance de ces activités est encore confirmée dans l'Article 5 de l'ITPGRFA. Au delà de ces priorités, chacun de ces cadres internationaux met l'accent sur le besoin de renforcer les capacités locales à conduire les recherches relatives à la diversité et aux ressources génétiques.

L'analyse spatiale peut améliorer considérablement la compréhension et le suivi de la biodiversité étant donné que les résultats de l'analyse spatiale permettent de formuler et mettre en œuvre des stratégies de conservation mieux ciblées et donc plus efficaces. Les résultats des études spatiales peuvent fournir une information cruciale sur la biodiversité présente dans des zones géographiques données et peuvent être utilisés à des fins diverses : par exemple, pour évaluer le statut actuel de conservation de certaines espèces végétales ou identifier les zones prioritaires de conservation. L'information spatiale, combinée aux données disponibles de caractérisation et/ou d'évaluation, s'est aussi avérée utile à la gestion efficace des banques de gènes (par exemple, par la définition de collections minimales, l'identification de lacunes dans les collections, etc.). Les outils des Systèmes d'Information Géographique (SIG) (Guarino *et al.*, 2002) permettent de conduire des analyses complexes combinant différentes sources de données (spatiales) et de générer des cartes lisibles, facilitant ainsi l'assimilation des résultats par les autorités concernées et encourageant le développement et la mise en œuvre des politiques de conservation. Ces dernières années, les avancées technologiques et la disponibilité croissante d'ordinateurs et de récepteurs de système de localisation mondial (GPS en anglais) ont conduit à une application accrue de l'analyse à base de SIG. L'accès et l'usage généralisés de l'Internet ont aussi révolutionné le partage de données sur la biodiver-

sité, la géographie et l'environnement. Le Système mondial d'information sur la biodiversité (GBIF en anglais) est une plateforme qui donne un accès public aux données sur la biodiversité issues des muséums nationaux, des herbiers et des banques de gènes dans le monde. En Octobre 2010, le GBIF contenait environ 39 millions d'observations géoréférencées de plantes (GBIF, 2009).

Ce manuel de formation est à l'intention des scientifiques (professionnels et étudiants) qui travaillent avec des données sur la biodiversité et qui souhaitent acquérir la capacité d'utiliser efficacement les programmes d'analyse spatiale dans les applications SIG. Il a été conçu pour être utilisé en auto-formation, mais il peut aussi être utilisé dans des ateliers de formation. Le manuel décrit comment réaliser des analyses de base en matière de biodiversité et d'écologie qui s'appuient sur les applications SIG. Les résultats de ces analyses offrent une meilleure compréhension de la structure spatiale de la diversité végétale, aidant ainsi à améliorer les efforts de conservation. Le manuel se concentre sur les plantes qui peuvent améliorer le bien-être des hommes (par exemple, les plantes cultivées ou leurs parentes sauvages) et/ou celles qui sont menacées. On y explique comment réaliser des analyses de diversité inter-spécifique et intra-spécifique en utilisant divers types de données : données d'occurrence d'espèces, de caractérisation morphologique (données phénotypiques) ou de marqueurs moléculaires (séquence ADN ou poids moléculaire). Bien que ce manuel se focalise sur la diversité végétale, de nombreuses analyses décrites s'appliquent aussi à l'étude d'autres organismes, comme les animaux et les champignons.

La publication de ce manuel répond au nombre croissant de sollicitations en matière de développement des capacités dans l'analyse spatiale des données sur la biodiversité reçues par Bioversity International. Les auteurs ont mis au point un ensemble d'instructions par étapes, accompagnées d'une série d'analyses, fondées sur des logiciels libres et disponibles gratuitement : DIVA-GIS, un programme de SIG conçu spécialement pour l'analyse spatiale de la biodiversité et Maxent, un programme de modélisation de la distribution d'espèces. Ce manuel ne vise pas à illustrer l'utilisation de chaque fonction ou option de DIVA-GIS ou de Maxent, mais se concentre plutôt sur l'utilisation d'outils SIG pour contribuer à répondre à des questions courantes relatives à l'analyse spatiale des données sur la biodiversité. Bien que tout au long du manuel on souligne l'importance d'un échantillonnage correct, cet ouvrage ne peut pas entrer dans le détail des théories de l'échantillonnage ; il ne traite pas non plus en détail de l'analyse statistique des données sur la biodiversité ; il indique plutôt les ouvrages auxquels le lecteur peut se référer lorsque des méthodes et programmes statistiques sont cités dans le texte.

En suivant les instructions et en menant à bien les analyses présentées dans ce manuel, les utilisateurs seront à même de conduire des analyses de base en matière de diversité spatiale pour répondre à des préoccupations communes à la recherche en biologie de la conservation et ressources phytogénétiques.

Références

- CDB. 2009a. Convention sur la diversité biologique. Stratégie mondiale pour la conservation des plantes [en ligne]. Disponible à : <http://www.cbd.int/gspc/>
- CDB. 2009b. Convention sur la diversité biologique. Article 7. Identification et surveillance [en ligne]. Disponible à : <http://www.cbd.int/convention/articles/?a=cbd-07>
- FAO. 2009. Le Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture [en ligne]. Disponible à : <http://www.planttreaty.org/fr>

GBIF. 2009. The Global Biodiversity Information Facility (GBIF) [en ligne]. Disponible à : <http://www.gbif.org/>

Guarino L, Jarvis A, Hijmans RJ, Maxted N. 2002. Geographic Information Systems (GIS) and the conservation and use of plant genetic resources. In: Engels JMM, Ramanatha Rao V, Brown AHD, Jacson MT, éditeurs. Managing plant genetic diversity. International Plant Genetic Resources Institute (IPGRI) Rome, Italie. pp. 387-404.

Section A

Éléments de base et préparation des données

Chapitre 1

Installation des programmes et de données d'exemple à analyser

Afin de réaliser les analyses présentées dans ce manuel, vous aurez besoin des programmes informatiques suivants : DIVA-GIS, Maxent, Google Earth et Excel. Les trois premiers sont disponibles en ligne gratuitement à :

- <http://www.diva-gis.org>
- <http://www.cs.princeton.edu/~schapire/maxent>
- <http://earth.google.com>

Le programme Excel, en revanche, n'est pas gratuit mais fait partie de la suite Microsoft Office, qui est installée sur la plupart des ordinateurs. Si vous n'avez pas accès à Excel, vous pouvez utiliser Calc, un logiciel libre disponible dans la suite OpenOffice qui peut être téléchargé à <http://www.openoffice.org/fr/>.

Ce manuel se réfère aux versions les plus récentes des programmes indiqués ci-dessus, à la date de publication du manuel. Nous espérons que les exemples et les instructions présentés dans ce texte demeureront valides dans un futur proche mais nous reconnaissons que ceux-ci peuvent changer avec les nouvelles versions de logiciel et que ces programmes peuvent devenir obsolètes. C'est pourquoi il est recommandé que les logiciels et les sites où se trouvent ces programmes soient consultés régulièrement pour toute mise à jour qui s'avérerait nécessaire.

Les paragraphes suivants fournissent des instructions détaillées pour télécharger et installer DIVA-GIS et Maxent sur votre ordinateur.

1.1. Téléchargement et installation de DIVA-GIS

DIVA-GIS est le principal programme utilisé pour mener les analyses spatiales présentées dans ce manuel et visualiser leurs résultats. L'installation de ce programme est plutôt simple. Tout ce qui est requis est la Version 7.3.0 (*diva730.zip*) du logiciel d'installation, disponible à <http://www.diva-gis.org/download>.

Note

DIVA-GIS a été développé pour le système d'exploitation Windows, mais il tourne aussi sur Mac OSX (voir le site Web de DIVA-GIS pour plus d'information).

Le site Web de DIVA-GIS offre aussi des données climatiques (voir la Section 2.3), des couches thématiques, un manuel utilisateur, une documentation de formation complémentaire et des liens vers d'autres publications et ressources Web pertinents.

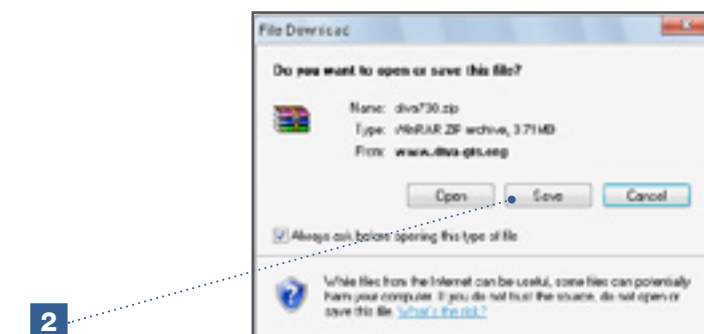
1.1.1. Comment installer DIVA-GIS

Etapes :

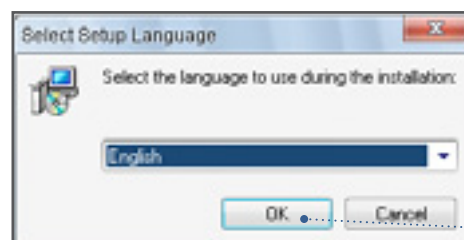
1. Télécharger le logiciel d'installation compressé à partir de l'adresse URL : <http://www.diva-gis.org/download>.



2. Sauvegarder le fichier sur votre disque dur.



3. Lancer le programme d'installation DIVA-GIS, en décompressant d'abord le fichier que vous avez sauvegardé à l'étape 2 ci-dessus. Pour décompresser le fichier, utilisez un logiciel libre comme 7-zip (<http://www.7-zip.org>) ou Izarc (<http://www.izarc.org>). Après avoir décompressé le fichier, vous pourrez voir le fichier *setup.exe*. Cliquez sur *setup.exe*. La fenêtre *Select Setup Language* sera alors affichée ; sélectionnez la langue pour le processus d'installation.



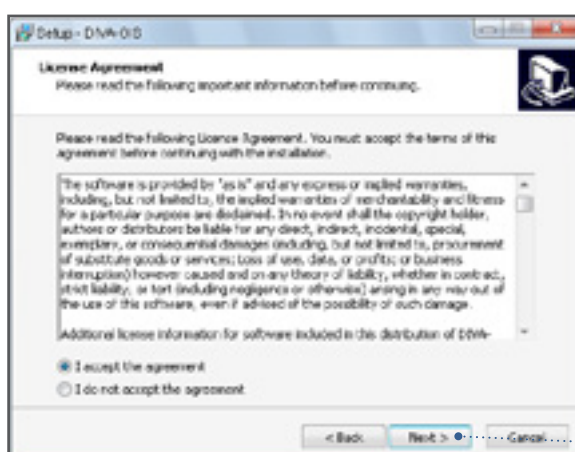
Note

Le programme DIVA-GIS lui-même n'est disponible qu'en anglais

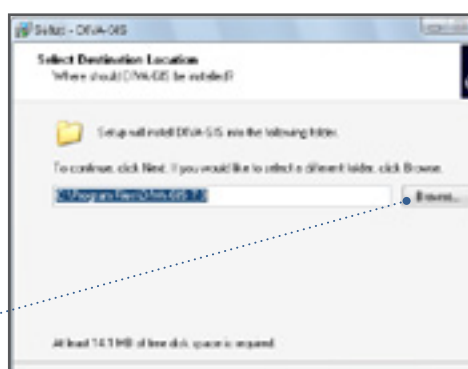
4. Le message de Bienvenue sera affiché. Cliquez sur *Next* pour continuer.



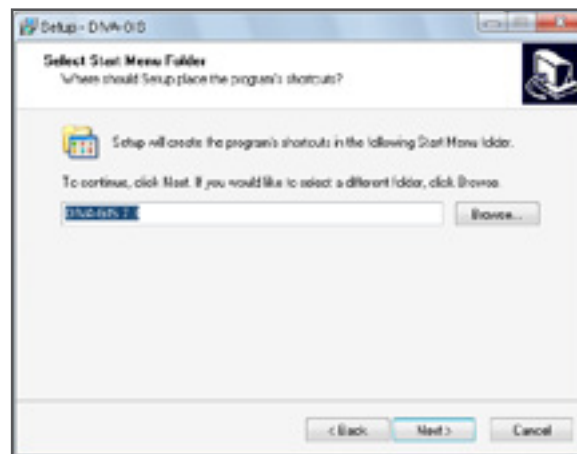
5. Lisez attentivement les termes de l'accord de la licence du logiciel DIVA-GIS. Si vous les acceptez, sélectionnez la case : *I accept the agreement*. Pour continuer, cliquez sur *Next*. Si vous indiquez que vous n'êtes pas d'accord avec les termes de la licence, le processus d'installation de DIVA-GIS sera avorté et le programme ne sera pas installé.



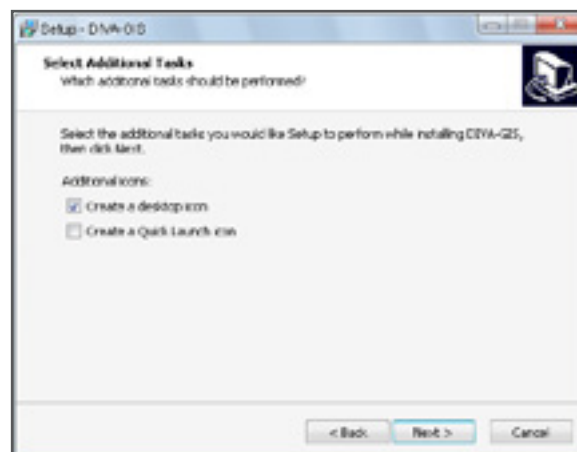
6. Indiquez le chemin d'accès au dossier où vous souhaitez sauvegarder le programme sur votre disque dur. La meilleure option peut être d'utiliser l'emplacement par défaut qui est proposé automatiquement par le programme d'installation. Si vous préférez installer le programme ailleurs, sélectionnez le chemin d'accès en utilisant le bouton *Browse*.



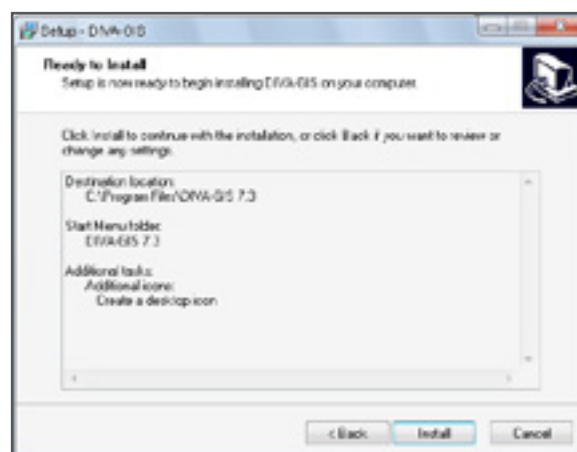
7. Indiquez le répertoire d'où vous aurez directement accès au programme dans le *Startup Menu*. Nous recommandons d'utiliser l'option par défaut.



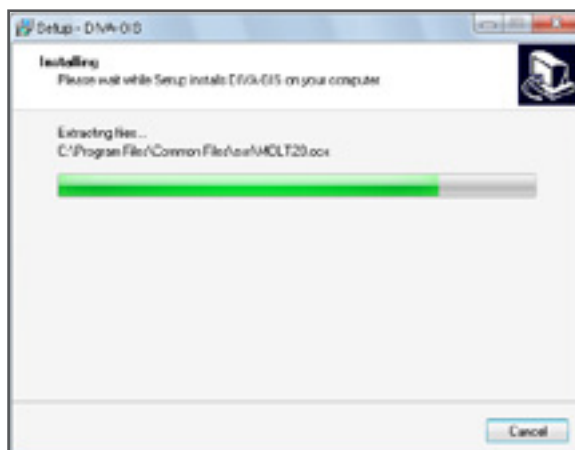
8. Sélectionnez la configuration désirée pour les icônes d'accès rapide à DIVA-GIS.



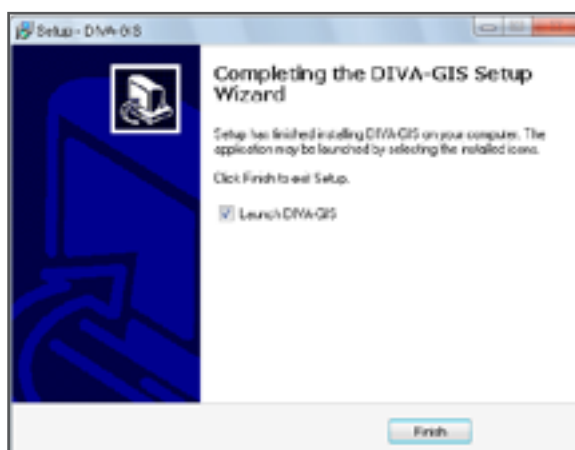
9. Une fenêtre résumant le chemin d'accès au dossier et les options que vous avez sélectionnées s'affichera. Relisez l'information pour vous assurer qu'elle est correcte et continuez en cliquant sur *Install*.



10. L'installation sera lancée et les fichiers requis seront installés sur votre ordinateur.



11. Si toutes les étapes sont franchies correctement, le processus d'installation devrait se dérouler sans problème. Pour terminer, cliquez sur *Finish*.



1.2. Installation de Maxent

Le programme Maxent sera utilisé pour l'analyse de modèles de distribution des espèces décrits dans le Chapitre 6 de ce manuel. Pour télécharger et utiliser Maxent correctement, vous aurez besoin d'avoir Java sur votre ordinateur. La plupart des ordinateurs modernes ont Java installé par défaut, mais le programme est aussi disponible gratuitement à l'adresse : <http://www.java.com/fr/>. Si Maxent ne fonctionne pas sur votre ordinateur, il est vraisemblable que Java n'y a pas été installé.



1.2.1. Comment installer Java

Étapes :

1. Si votre navigateur est Internet Explorer, allez sur : http://www.java.com/fr/download/help/ie_online_install.xml.

Si votre navigateur est Firefox, allez sur : http://www.java.com/fr/download/help/firefox_online_install.xml.
2. Cliquez sur *Install* pour lancer l'installation

2



1.2.2. Comment installer Maxent

Maxent est un programme libre qui peut être téléchargé à l'adresse suivante : <http://www.cs.princeton.edu/~schapire/maxent/>. Le fichier exécutable du programme est téléchargé directement, il n'y a donc pas besoin d'une installation séparée.

Étapes :

1. Les documents qui expliquent les concepts utilisés par Maxent peuvent être trouvés sur le site Web cité ci-dessus.
2. Avant de télécharger le programme, fournissez l'information de contact demandée.
3. Enfin, acceptez l'accord de licence et téléchargez le programme.



4. Les fichiers du programme Maxent peuvent être téléchargés séparément ou ensemble dans un format compressé (*maxent.zip*).



5. Sauvegardez les fichiers dans un emplacement pratique sur votre disque dur. Pour lancer Maxent, double-cliquez sur le fichier JAR (*.jar).

Note

Si vous avez le logiciel de compression de données Winrar installé sur votre ordinateur, il se peut qu'il soit associé aux fichiers JAR,



6. Si vous utilisez fréquemment Maxent, il peut être utile de créer un raccourci vers le fichier JAR (*.jar) sur le bureau de votre ordinateur.

1.3. Installation de Google Earth

Google Earth est devenu récemment un outil important pour visualiser et partager l'information géographique. Le programme permet de visualiser des données géoréférencées sur la distribution d'un taxon combinées avec des images satellitaires de grande qualité. La disponibilité croissante de ces images va sans aucun doute stimuler l'utilisation de Google Earth et il est probable qu'apparaissent des applications similaires sur le marché. Avec Google Earth, les couches peuvent être visualisées à diverses échelles, du niveau mondial (continent, pays, etc.), jusqu'à des niveaux aussi spécifiques qu'un arbre (qui pourrait correspondre à un point d'occurrence dans une base de données), fournissant ainsi de nombreuses possibilités de visualisation.

Cependant, lorsqu'on utilise Google Earth, on risque de faire des interprétations à haute résolution à partir de données d'une précision réduite. Par exemple, vous pouvez savoir que vous avez géoréférencé des occurrences d'espèces avec une précision de 5 km, alors que la résolution des images de Google Earth peut aller jusqu'à 1 x 1 m. Dans cet exemple, essayer de relier le jeu de données sur les espèces à l'image de Google Earth et à sa résolution n'aurait pas de sens.

1.3.1. Comment installer Google Earth

L'installation de Google Earth est très simple et consiste à télécharger le programme d'installation, à l'ouvrir et à suivre les instructions affichées à l'écran. Pour faire tourner Google Earth, vous devez être en ligne (c'est-à-dire connecté à l'Internet).

Etapes :

1. Téléchargez le programme d'installation de Google Earth à l'adresse : <http://earth.google.com> (après avoir accepté les conditions).



2. Ouvrez le fichier d'installation téléchargé.
3. Suivez les instructions affichées à l'écran. L'installation commencera en téléchargeant des données supplémentaires.

1.4. Données à analyser

Pour conduire les analyses décrites dans les sections suivantes de ce manuel, vous devrez installer les programmes requis et **télécharger des données d'exemples pour chaque section** qui sont disponibles sur le site web de Bioversity à l'adresse : http://www.bioversityinternational.org/training/training_materials/GIS_manual. A chaque section correspond un jeu séparé de données. Vous devrez placer ces jeux de données dans un répertoire unique pour faciliter leur accès par DIVA-GIS.

La plupart des analyses présentées dans ce manuel sont fondées sur des données climatiques détaillées (à une résolution de 2,5 minutes), ce qui entraîne le téléchargement de gros fichiers de données qui requièrent une bonne connexion Internet.

- Si vous bénéficiez d'une bonne connexion Internet, nous vous suggérons de télécharger les jeux de données dans un fichier zip. Quand ce fichier sera décompressé, cela créera un nouveau répertoire où toutes les données seront déjà organisées séparément pour être utilisées avec ce manuel.
- Si votre connexion Internet est assez lente, vous pouvez télécharger les données section par section. Quand vous décompresserez les différents fichiers zip, les données seront également organisées en fichiers séparés comme le requiert le manuel.
- Si vous rencontrez des difficultés pour télécharger les données, n'hésitez pas à contacter les auteurs du manuel aux adresses électroniques suivantes : bioversity-colombia@cgiar.org, x.schelde@gmail.com ou m.vzonneveld@gmail.com. Si nécessaire, des données de plus basse résolution (données climatiques avec moins de détail) peuvent être fournies ou envoyées en format DVD.

Les jeux de données fournis avec le manuel sont basés sur les études existantes mais ont été modifiés afin de les adapter aux analyses ou pour protéger la propriété intellectuelle, par exemple en y introduisant des erreurs ou de légères modifications (dans le cas de données non publiées). La Section 2.3 fournit des sources de données spatiales réelles qui peuvent être utilisées pour réaliser des études spécifiques.

Chapitre 2

Préparer et importer les données vers DIVA-GIS et Maxent

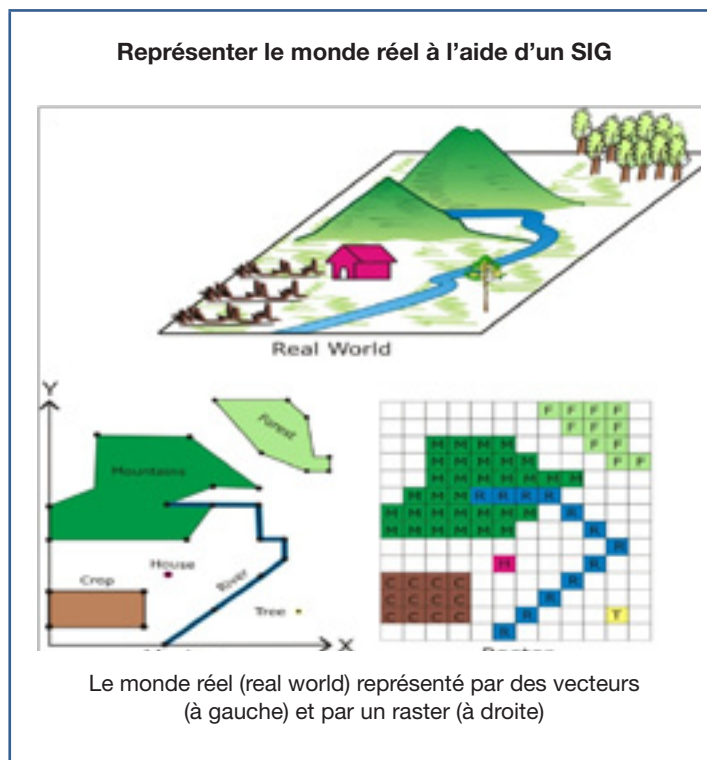
L'analyse spatiale de la biodiversité est basée sur les données d'observation. Les données d'observation sont des aperçus de la présence d'espèces, de caractères ou d'allèles dans le temps et l'espace. Pour analyser des données d'observation, les utilisateurs peuvent travailler sur leurs propres données, sur des données disponibles gratuitement via des plateformes internationales comme le GBIF (<http://www.gbif.org/>) ou une combinaison des deux. Nombre des données sur le portail GBIF sont des observations historiques tirées d'herbiers et de banques de données, qui peuvent ne pas refléter la présence actuelle des taxons, du fait de processus écologiques ou d'interventions humaines récents, telles que la conversion de forêt à l'agriculture et d'autres changements dans l'utilisation des terres. Néanmoins, de telles données restent utiles pour appréhender les processus écologiques et génétiques en jeu dans la distribution spatiale de la diversité des plantes. Les données d'observation doivent être organisées selon le format spécifié par le logiciel de SIG utilisé. Ce chapitre explique le type de données requis pour les analyses spatiales et la modélisation de la distribution des espèces avec DIVA-GIS et Maxent et comment préparer et formater ces données.

Les analyses décrites dans ce manuel sont basées sur deux composantes de données d'observation : les données de passeport et les données associées. Les données de passeport de base consistent en un code d'identification (ID), de l'identification taxonomique d'une plante observée individuellement (ou d'un groupe d'individus) et de la localisation du site de collecte ou d'observation. Les analyses présentées dans ce manuel utilisent principalement les données de passeport, couramment dénommées « points de présence ». Les points de présence peuvent être associés à des données additionnelles qui décrivent le site de collecte ou d'observation (par exemple, la couverture végétale, le type de sol, etc.), la date de la collecte, la source des coordonnées, le nom du ou des collecteur(s) et l'institution qui héberge ou conserve le spécimen¹. L'information sur la date de collecte fournit une dimension temporelle aux analyses et peut être utilisée pour analyser les tendances dans la distribution des espèces (par exemple, pour suivre de possibles cas d'érosion génétique).

A l'opposé des points de présence se situent les données d'absence. Bien que les données d'absence puissent aussi être pertinentes dans les analyses spatiales, par exemple pour suivre les tendances dans la distribution d'une espèce, il est souvent difficile de comprendre les raisons concrètes de l'absence de taxons dans une unité géographique, ce qui complique l'utilisation de ces données dans les analyses écologiques (telles que celles qui sont expliquées dans le Chapitre 6). Un taxon peut avoir disparu du fait de perturbations d'origine humaine ; son absence d'un lieu peut aussi être expliquée par des limitations dans sa dispersion ou par des changements de conditions environnementales. De plus, il est toujours possible que la présence d'un spécimen ait pu tout simplement échapper aux collecteurs. Les points de présence, peuvent être reliés à des facteurs plus crédibles, tels que les variables climatiques. C'est pourquoi les analyses dans ce manuel sont basées uniquement sur les points de présence.

¹ Dans le domaine de la conservation des ressources génétiques des plantes cultivées, de telles données sont considérées comme « données de passeport ». Pour plus d'information, consulter la page Web de Bioversity sur les 'Descripteurs de passeport pour de multiples espèces cultivées' sur le site web de Bioversity à l'adresse suivante : www.bioversityinternational.org/index.php?id=19&user_bioversitypublications_pi1%5bshowUid%5d=2192

Les spécimens sont la plupart du temps associés à des données qui donnent de plus amples détails sur les caractéristiques individuelles de la plante et à des données utilisées pour évaluer la diversité intra-spécifique ou les caractéristiques agronomiques qui peuvent être associés à des informations spatiales (par exemple des données climatiques). Les données associées peuvent donc comprendre des informations sur les caractères morphologiques (comme la taille, la forme, la couleur), la physiologie (comme le nombre de jours avant la germination, ou avant la floraison), l'évaluation (comme le rendement, la tolérance aux stress abiotiques ou biotiques) ou la séquence ADN (comme des données de marqueurs moléculaires). La Section 5.2 présente une analyse de la diversité fondée sur les caractères morphologiques (diversité phénotypique) alors que l'analyse présentée dans la Section 5.3 est fondée sur les données de marqueurs moléculaires.



Les logiciels de SIG peuvent traiter différents types de données spatiales : les données vectorielles et les données raster (en quadrillage). Les données vectorielles sont constituées d'un ensemble de points géoréférencés qui forment des groupes de points, des lignes ou des polygones et représentent des zones géographiques réelles. Dans ce manuel, les points géoréférencés de présence de plantes individuelles (points de présence) ou les couches d'unités administratives des pays (polygones) servent d'exemples de données vectorielles. Le format de dossier des données vectorielles le plus couramment utilisé est *shapefile*.

DIVA-GIS utilise des *shapefiles* pour représenter des données vectorielles. Un *shapefile* consiste en (au moins) trois fichiers avec le même nom de fichier mais avec une extension différente. Les extensions de ces fichiers sont : '.shp', '.shx' et '.dbf'. L'ensemble formé par ces trois fichiers constitue un *shapefile*. Le fichier SHP (*.shp) est le fichier principal dans lequel sont stockées les *spatial features* (objets spatiaux) et qui apparaît dans la boîte de dialogue lorsqu'on ouvre le fichier vecteur en utilisant DIVA-GIS. Les attributs non spatiaux des objets sont stockés dans le fichier dBase IV (*.dbf), tandis que le fichier SHX (*.shx) contient l'information d'indexation du fichier vecteur – elle est utilisée par les programmes pour accéder plus rapidement au fichier .shp et le relier aux enregistrements dans le fichier dBase IV (.dbf).

Les données environnementales de zones géographiques spécifiques peuvent aussi être géoréférencées et sont normalement organisées en raster. Un raster est un ensemble de cellules de taille identique, où chaque cellule contient une valeur d'une variable d'intérêt (comme la température, le type de sol). La taille optimale des cellules dépend de la taille de la zone géographique et de l'objectif de l'étude. La taille de la zone étudiée est désignée « étendue », tandis que le niveau de détail (donné par la taille de la cellule) est appelé « résolution ». L'usage de différentes tailles de cellules se traduit par des rasters aux résolutions différentes (de plus petites cellules engendrent des rasters de plus haute résolution). Tout au long de ce manuel, diverses tailles de cellules de raster sont utilisées, de un (1) degré (approximativement 111 km à l'équateur) jusqu'à 30 secondes (approximativement 1 km à l'équateur). Bien que ce soient des unités géographiques assez grandes, ces tailles sont appropriées pour de nombreux types d'analyses spatiales et de modélisation de la distribution des espèces au niveau national ou régional. Ces types d'analyses sont présentés dans les Chapitres 5 et 6.

Tailles approximatives des unités géographiques (à l'équateur ²), arrondies au km	
Degrés	Taille
1 degré	111 km
10 minutes	18 km
5 minutes	9 km
2,5 minutes	5 km
30 secondes	1 km

Les SIG et les programmes de modélisation de la distribution des espèces utilisent différents types de fichiers lorsqu'ils traitent des données raster. Les données raster utilisées par le programme DIVA-GIS sont sous forme de grille, mais le programme permet aussi d'importer ou d'exporter d'autres types de fichiers raster (voir les Analyses 3.1.5 et 3.1.6).

Dans DIVA-GIS, un raster est constitué de deux fichiers : un fichier GRD (*.grd) et un fichier GRI (*.gri), mais seul le fichier GRD est affiché quand le raster est ouvert dans DIVA-GIS. Le fichier GRD contient l'information générale sur le format du raster, comme la position des coins du raster, le nombre de colonnes et de lignes et la taille de cellule ; le fichier GRI contient les valeurs de chaque cellule du raster et il est donc notablement plus gros.

Quand on importe les données raster dans Maxent, il est recommandé d'utiliser des fichiers raster de type ASCII (qui peuvent être créés dans DIVA-GIS à partir des fichiers GRD). Un fichier ASCII est un fichier unique (fichier *.asc). Le logiciel ESRI ArcGIS® a ses propres types de fichiers, mais il peut aussi importer des fichiers ASCII.

² Vers les pôles, les distances métriques en longitude (est-ouest) deviennent plus courtes, jusqu'à se réduire à zéro aux pôles Nord et Sud (où l'on peut « faire le tour du monde » en quelques mètres seulement). Ce manuel utilisera des tailles de cellules en degrés, minutes et secondes, avec lesquelles il est plus facile de travailler qu'avec des tailles de cellules en distances métriques.

2.1. Préparation et importation des points de présence

Comme indiqué précédemment, les points de présence peuvent être extraits des données d'inventaires de végétation, d'expéditions de collecte de matériel végétal ou de sources publiques comme le GBIF. La Section 2.3 fournit des liens vers plusieurs sources d'information de ce type. Les appareils GPS sont maintenant largement utilisés pour géoréférencer les points de présence. Ceux-ci peuvent être organisés en fichier Excel et ensuite convertis dans les formats adaptés aux analyses spatiales en utilisant des programmes de SIG comme DIVA-GIS, ou des programmes de modélisation de la distribution d'espèces comme Maxent. Cette section décrit les données minimales requises pour chaque point de présence, comment formater ces données sous Excel et comment importer ensuite de telles données dans DIVA-GIS et Maxent.

Pour commencer à préparer une base de données avec des points de présence, il faut prendre en compte les éléments suivants :

a. Information de base

Les points de présence doivent comprendre les données de passeport de base d'une plante individuelle ou d'un groupe d'individus dans une unité géographique donnée. Ils doivent comprendre au moins quatre éléments : un code d'identification (ID), le nom taxonomique de la plante (ou du groupe d'individus) et les coordonnées longitude et latitude. Ces types de points sont couramment utilisés en analyse spatiale de la diversité et de la distribution géographique.

b. Stocker les coordonnées

Quand des coordonnées géographiques sont utilisées dans DIVA-GIS, Maxent et d'autres SIG et programmes de modélisation de la distribution des espèces, il est préférable qu'elles soient reportées en utilisant un système de coordonnées latitude/longitude et présentées en Degrés Décimaux (*Decimal Degrees*, DD) : format DD.DDDD). Lorsque les coordonnées sont disponibles en Degrés, Minutes et Secondes (*Degrees, Minutes, Seconds*, DMS) : format DD°MM'SS") ou en Degrés Minutes (*Degrees, Minutes* DM) : format DD°MM.MM'), l'information devrait être convertie en DD pour être utilisée dans un SIG ou un programme de modélisation de la distribution d'espèces. Il faut alors employer la formule de conversion de données :

$$\text{Degrés décimaux} = [(\text{Degrés (°)} + \text{Minutes (')} / 60 + \text{Secondes (")} / 3600)] * H$$

H = 1 lorsque la coordonnée est dans l'hémisphère Est (E) ou l'hémisphère Nord (N)

H = -1 lorsque la coordonnée est dans l'hémisphère Ouest (O) ou l'hémisphère Sud (S)

Longitude	Degrés, Minutes et Secondes	Degrés Décimaux	Latitude	Degrés, Minutes et Secondes	Degrés Décimaux
Hémisphère Est	60°20'15" E	+ 60,3375	Hémisphère Nord	24°00'45" N	+ 24,0125
Hémisphère Ouest	60°20'15" W	- 60,3375	Hémisphère Sud	24°00'45" S	- 24,0125

Il est recommandé que les points DD aient une précision d'au moins quatre décimales. Sous Excel, la conversion de données DMS ou DM en format DD est possible à l'aide des fonctions texte (les données DMS sont stockées sous forme de texte) *RIGHT/MID/LEFT*.

Note

Faites attention à ne pas donner une fausse impression de précision en mettant trop de décimales lorsque vous convertissez des données du format DMS au format DD. Une coordonnée en format DMS dont seules les valeurs en degrés et minutes sont indiquées (par exemple, 60°05') a une précision véritable de deux décimales, mais elle peut être présentée en format DD comme une coordonnée à quatre décimales (60,0833). Lorsque les coordonnées du point de présence sont au format DMS avec moins de quatre décimales (seulement des degrés et minutes ou seulement des degrés), ajoutez un champ supplémentaire pour spécifier la précision des coordonnées, avant de les convertir de DMS à DD.

En dehors du système de coordonnées latitude/longitude, il existe aussi le système Universal Transverse Mercator (UTM en anglais) qui est couramment utilisé. Bien que DIVA-GIS puisse tourner en utilisant le système UTM, il est préférable d'utiliser les coordonnées latitude/longitude (en format DD) du fait qu'elles ont plus de chances d'être compatibles avec les couches thématiques disponibles (information sur les unités administratives, sur le climat ou sur la couverture végétale). Les coordonnées UTM peuvent être converties en latitude/longitude en degrés décimaux à l'aide des feuilles de calcul d'Excel qui peuvent être téléchargées sur Internet sur des sites comme : <http://www.uwgb.edu/dutchs/usefuldata/utmformulas.htm>.

L'option *Projection* dans le menu *Tools* de DIVA-GIS vous permet aussi de convertir les formats de latitude et de longitude en UTM et vice versa.

PROGRAMMES ET FICHIERS DE DONNEES UTILISEES DANS CETTE SECTION**Programmes :**

- DIVA-GIS
- Excel
- Maxent et Java

Fichiers de données :

- Carpeta 2.1 Importing observation data*
(Importación de datos de presencia)
- *Vcundinamarcensis_DMSdata.xls*
 - *Vasconcellea.xls*

2.1.1. Comment convertir des données du format Degrés, Minutes, Secondes (Degrees, Minutes, Seconds [DMS]) vers le format Degrés Décimaux (Decimal Degrees [DD])

Cet exercice utilise comme exemple le fichier Excel *Vcundinamarcensis_DMSdata.xls* ; le fichier se trouve dans le répertoire de données qui accompagne ce chapitre et qui comprend sept feuilles de calcul dans lesquelles les coordonnées DMS sont converties par étapes au format DD.

Etapes :

1. Ouvrez le fichier Excel [Vcundinamarcensis_DMSdata.xls](#) et sélectionnez la feuille de calcul *Start data*. Notez que la latitude et la longitude sont présentées en format DMS.

ID	Species	Latitude	Longitude	Country	Altitude
1	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
2	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
3	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
4	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
5	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
6	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes
7	Vcundinamarcensis	5°42'N	75°00'W	Colombia	Andes

- Allez à *Columns* dans la feuille de calcul. Créer des colonnes séparées pour les degrés, les minutes et les secondes de la latitude et de la longitude (Colonnes E, F, G et L, M, N). La colonne *Lat* (Colonne D) a été créée pour indiquer si la coordonnée de latitude est dans l'hémisphère Nord, N (1), ou Sud, S (-1). La colonne *Lon* (Colonne K) a été créée pour indiquer si la coordonnée de longitude est dans l'hémisphère Est, E (1), ou Ouest, W (-1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		SPECIES	LATITUDE	Lat	degrees	minutes	seconds	decimal degrees latitude	LONGITUDE	Lon	degrees	minutes	seconds	decimal degrees longitude
1	3432	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
2	3837	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
3	2846	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
4	2848	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
5	3338	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
6	3430	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000

- Ouvrez la feuille de calcul *DD formula*. La formule de conversion de la latitude et de la longitude de DMS vers DD ($=D2*(E2+(F2/60)+(G2/3600))$) a été insérée dans les colonnes de latitude et de longitude nommée *decimal degrees* (Colonnes H et O).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		SPECIES	LATITUDE	Lat	degrees	minutes	seconds	decimal degrees latitude	LONGITUDE	Lon	degrees	minutes	seconds	decimal degrees longitude
1	3432	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
2	3837	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
3	2846	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
4	2848	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
5	3338	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
6	3430	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000

- Ouvrez la feuille de calcul *Formulas*. Des formules particulières (basées sur les fonctions texte d'Excel *RIGHT/MID/LEFT*) sont insérées dans les colonnes nommées *Lat*, *Lon*, *degrees*, *minutes* et *seconds* (Colonnes D et K, colonnes E, F, G et L, M, N) afin de séparer les valeurs des hémisphères, degrés, minutes et secondes tirées des colonnes *Latitude* et *Longitude* (Colonnes C et J) et les placer dans des colonnes distinctes.
- Ouvrez la feuille de calcul *Copied formulas*. Les formules sont copiées dans chaque ligne pour déterminer la latitude et la longitude en DD pour chaque point de présence.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		SPECIES	LATITUDE	Lat	degrees	minutes	seconds	decimal degrees latitude	LONGITUDE	Lon	degrees	minutes	seconds	decimal degrees longitude
1	3432	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
2	3837	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
3	2846	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
4	2848	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
5	3338	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000
6	3430	V. sublineatus	07°40'N	1	7	40	00	7.6667	75°00'W	-1	75	00	00	-75.0000

- Ouvrez la feuille de calcul *Values*. Copiez les coordonnées calculées en DD (Colonnes H et O) et collez les valeurs, à l'aide de l'option *Paste special*, dans les colonnes correspondantes de *latitude* et *longitude* (Colonnes I et P). Ceci est nécessaire du fait que DIVA-GIS ne peut pas importer de nouvelles valeurs DD présentées sous forme de formules.
- Ouvrez la feuille de calcul *Final*. Les coordonnées en DMS sont maintenant converties en DD et la base de données de points de présence est prête à être importée dans DIVA-GIS. Il est important de conserver les coordonnées DMS originales afin de pouvoir retrouver toute erreur qui peut s'être produite au cours du calcul des coordonnées.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		SPECIES	Latitude	Longitude	COUNTRY	ACID	LATITUDE DMS	LONGITUDE DMS	Latitude DD	Longitude DD				
1	3432	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				
2	3837	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				
3	2846	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				
4	2848	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				
5	3338	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				
6	3430	V. sublineatus	7°40'N	75°00'W	Colombia	Antioquia	07°40'N	75°00'W	7.6667	-75.0000				

Géoréférencement de l'information de présence

Il arrive parfois que les points de présence soient dépourvus de coordonnées géographiques et n'aient qu'une description de leur localisation sous forme de données relatives aux unités administratives. Ces données peuvent être catégorisées comme suit :

- Pays
- Niveau administratif 1 (Adm1) : état, département, région, province (d'un pays)
- Niveau administratif 2 (Adm2) : province, canton, municipalité
- Localité : grande ville, ville, parc national, etc.

	A	B	C	D	E	F
1	ID	Taxon	Country	Adm1	Adm2	Locality
2	1	<i>Capsicum chacoense</i> Jacq.	Bolivia	Pando	Nuevo Surata	Cobija
3	2	<i>Capsicum chacoense</i> Jacq.	Bolivia	Pando	Nuevo Surata	Cobija
4	65	<i>Capsicum bursadense</i> L.	Bolivia	Beni	Vista Alegre	Pibaraña
5	72	<i>Capsicum eximium</i> Harms	Bolivia	Tarija	Mendoza	San Lorenzo
6	73	<i>Capsicum eximium</i> Harms	Bolivia	Tarija	Mendoza	San Lorenzo

Exemple de données relatives à une unité administrative de Bolivie avec des accessions de *Capsicum* originaires du pays.

Dans de tels cas, on peut s'aider des bases de données connues sous le nom de « *gazetteers* ». Ce sont des dictionnaires géographiques qui contiennent des listes d'unités administratives (comme des municipalités), avec leurs coordonnées géographiques respectives que l'on peut utiliser pour attribuer de l'information géoréférencée sur des points d'intérêt. Des dictionnaires géographiques contenant les données des unités administratives de la plupart des pays sont disponibles en ligne et peuvent être téléchargés gratuitement à partir du site web de DIVA-GIS : <http://www.diva-gis.org/gdata>. Les fichiers sont généralement en format dBase IV (*.dbf), et peuvent être ouverts et consultés sous Excel (à l'aide de l'option *Edit/Find*).

Mais il n'est pas toujours possible de géoréférencer un site à l'aide d'un dictionnaire géographique, car l'information sur la localité peut être incomplète ou même indisponible. Une autre difficulté intervient lorsque plusieurs endroits portent le même nom (voir le tableau ci-dessous). Dans ce cas, les différences entre les niveaux administratifs supérieurs, comme Adm1, peuvent aider à faire une distinction entre les endroits et à résoudre le conflit.

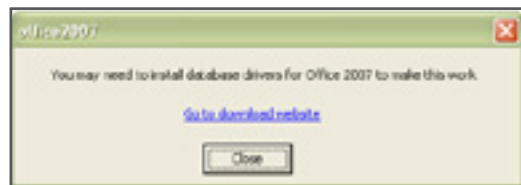
Outre les dictionnaires géographiques, d'autres programmes comme Biogeomancer, disponible à l'adresse : <http://www.biogeomancer.org>, permettent de géoréférencer des points à partir de l'information disponible sur le site.

A	B	C	D	E	F	G	H	I
13104	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13107	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13108	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13109	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13110	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13111	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13112	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00
13113	Jalapa	MEX	Jalisco	Jalisco	Jalisco	19 05 00	101 05 00	19 05 00

Dictionnaire géographique du Mexique qui présente plus d'un endroit portant le nom 'Jalapa'.

2.1.2. Comment importer dans DIVA-GIS des points de présence géoréférencés

Comme il a été indiqué précédemment, les points de présence peuvent être importés depuis Excel dans DIVA-GIS. Si vous essayez d'importer des données directement depuis Excel 1997-2003, un signal d'alerte comme ci-dessous apparaîtra à l'écran.



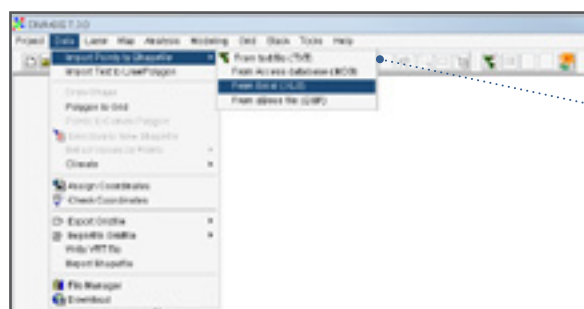
Le signal d'alerte indique que vous devez installer les pilotes de bases de données pour Office 2007. Cliquer sur *Go to download website* pour aller sur le Centre de Téléchargement de Microsoft où vous pourrez obtenir ces pilotes.

Note

Si vous utilisez un autre logiciel que Excel pour gérer votre base de données de passeport, ou si vous rencontrez encore des problèmes pour importer des points de présence depuis Excel dans DIVA-GIS après avoir installé les pilotes, vous pouvez importer les points dans DIVA-GIS à partir de fichiers texte (*.txt), CSV (*.csv) ou dBase IV (*.dbf). L'importation de points à partir de ces types de fichiers vers DIVA-GIS est analogue à l'introduction des points à partir d'Excel (en suivant les étapes expliquées dans cet exemple), si ce n'est que la base de données des points de présence sera sauvegardée dans un autre type de fichier. Pour des explications plus complètes sur l'importation de points de présence à partir de ces types de fichiers, vous pouvez consulter le Manuel d'exploitation de DIVA-GIS, disponible en ligne à l'adresse : <http://www.diva-gis.org/documentation>.

Etapes de l'importation de données depuis Excel dans DIVA-GIS :

1. Excel 2007 : Sauvegardez vos points de présence dans un fichier Excel 1997-2003 (*.xls) et fermez Excel 2007. [Si vous sauvegardez dans un fichier Excel (*.xlsx), les points de présence ne peuvent plus être importés dans DIVA-GIS]. Pour les versions antérieures d'Excel : sauvegardez vos points de présence dans un fichier Excel (*.xls).
2. Ouvrez DIVA-GIS et allez à *Data/Import Points/From Excel (.XLS)*

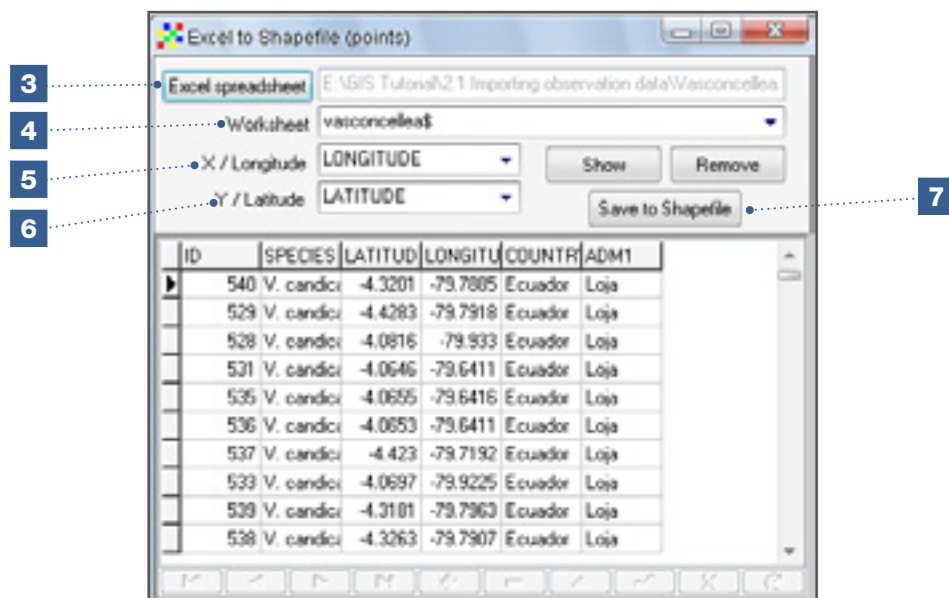


3. Dans *Excel spreadsheet*, sélectionnez le fichier Excel qui contient les points de présence que vous voulez importer (*Vasconcellea.xls*) dans DIVA-GIS.

Note

Fermez toujours le fichier Excel avant d'importer les données dans un fichier vecteur dans DIVA-GIS.

4. Sous *Worksheet*, sélectionnez la feuille de calcul Excel qui contient les données que vous voulez importer dans DIVA-GIS.
5. Sélectionnez la colonne qui contient les coordonnées de longitude.
6. Sélectionnez la colonne qui contient les coordonnées de latitude.
7. Cliquez sur *Save to Shapefile* pour générer un fichier vecteur (*.shp).



2.1.3. Comment importer dans Maxent des points de présence géoréférencés

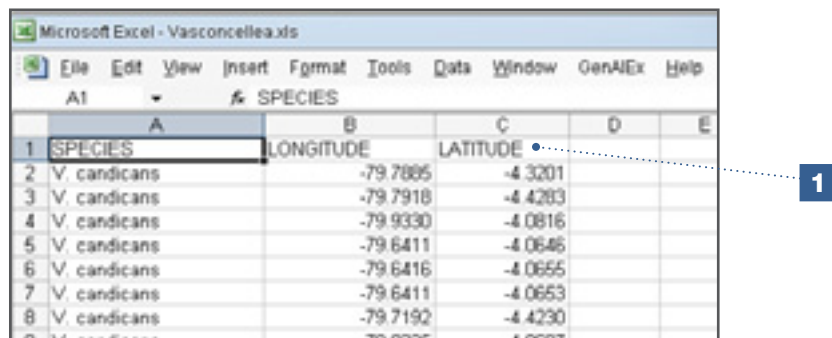
Maxent est un programme de modélisation de la distribution d'espèces utilisé pour prédire la distribution potentielle d'une ou de plusieurs espèces. Comme DIVA-GIS, Maxent travaille avec des données de présence géoréférencées. Cette section illustre comment importer dans Maxent des points de présence géoréférencés. La manière dont fonctionne Maxent est présentée de façon détaillée dans la Section 6.2.

Dans Maxent, les données doivent être importées sous forme d'un fichier CSV (*.csv) qui doit comprendre trois champs fixes (colonnes) correspondant aux catégories suivantes : Espèce, Longitude et Latitude. Les colonnes doivent être listées dans cet ordre (mais des colonnes additionnelles avec des informations supplémentaires sont autorisées). Le fichier à trois colonnes peut être préparé en commençant avec une base de données originellement construite sous Excel et ensuite sauvegardée au format CSV (*.csv). Il y a de légères différences dans la procédure de création des fichiers CSV selon que l'on utilise Excel 2007 ou Excel 1997-2003.

Etapes de la préparation d'un fichier *Comma Separated Values (CSV)* pour Maxent sous Excel 1997-2003

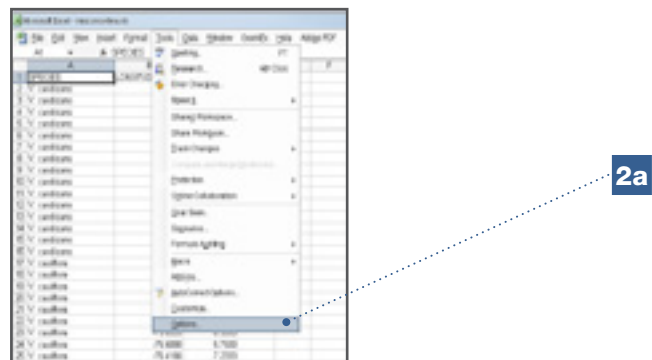
Dans l'exemple suivant, on utilise le fichier Excel *Vasconcellea.xls* qui se trouve dans le répertoire de données qui accompagne cette section.

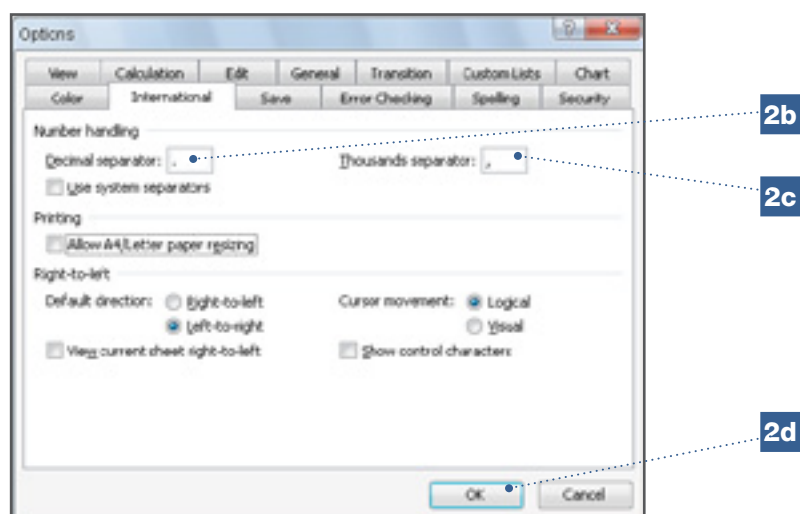
1. Dans une nouvelle feuille Excel, copiez les données des variables suivantes : Espèce, Longitude et Latitude, **dans cet ordre précis**.



	A	B	C	D	E
1	SPECIES	LONGITUDE	LATITUDE		
2	V. candicans	-79.7885	-4.3201		
3	V. candicans	-79.7918	-4.4283		
4	V. candicans	-79.9330	-4.0816		
5	V. candicans	-79.6411	-4.0646		
6	V. candicans	-79.6416	-4.0655		
7	V. candicans	-79.6411	-4.0653		
8	V. candicans	-79.7192	-4.4230		

2. Avant de commencer, assurez-vous que les décimales sont séparées par des points :
 - a. Allez à *Tools/Options*
 - b. Dans l'onglet *International*, sélectionnez *points* [.] pour séparer les décimales
 - c. Dans l'onglet *International*, sélectionnez *commas* [,] pour séparer les milliers
 - d. Cliquez sur *OK*.





3. N'utilisez aucune virgule parmi les caractères de la colonne des noms d'espèces ni dans aucune des colonnes additionnelles, s'il en existe. Ceci évitera les erreurs involontaires qui peuvent apparaître quand on sépare l'information du fichier CSV (*.csv) (par exemple dans l'information sur l'unité administrative).
4. Sauvegardez le fichier en format CSV (*.csv).

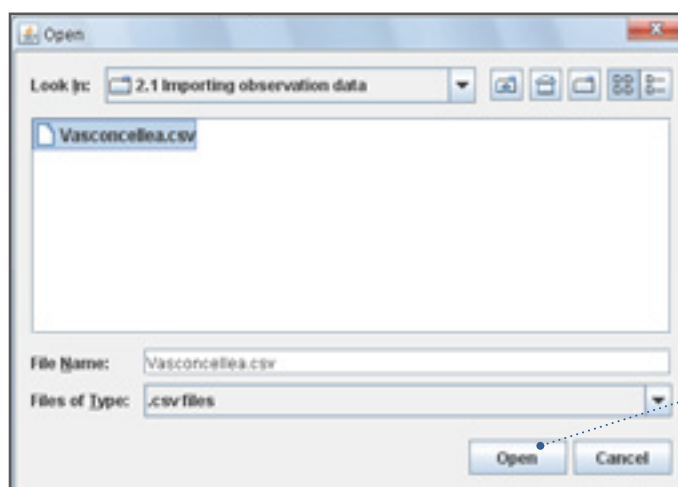
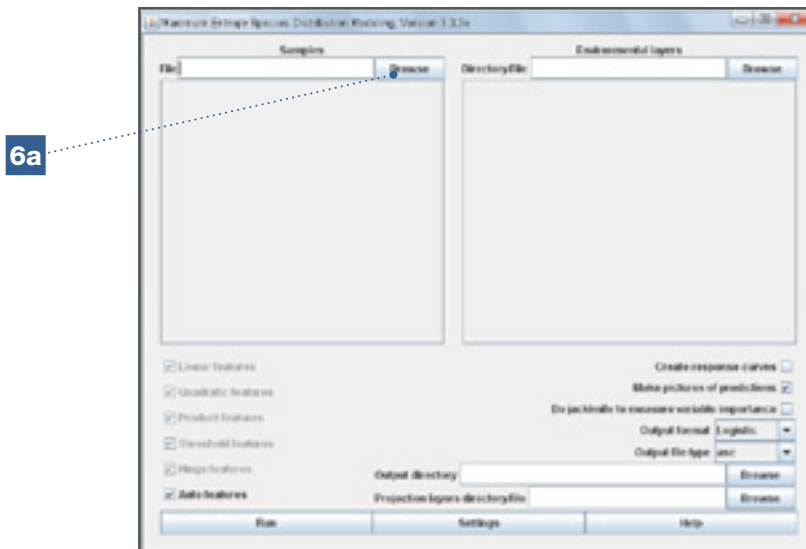


5. Ouvrez le fichier CSV (*.csv) dans Excel pour vérifier que les coordonnées ont toutes les décimales requises.

6. Ouvrez Maxent et, sous l'onglet *Samples*, indiquez le fichier CSV (*.csv) qui contient les points de présence géoréférencés.

Pour sélectionner le fichier :

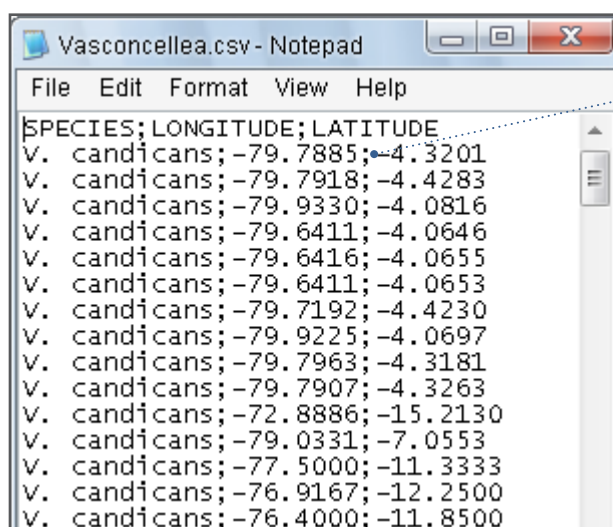
- a. Allez à l'option *Browse*
- b. Ouvrez le fichier CSV (*.csv).



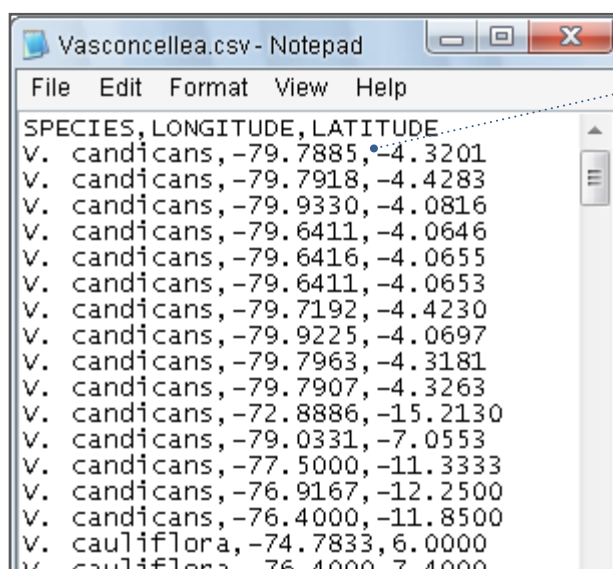
Etapes pour préparer un fichier CSV pour Maxent sous Excel 2007 :

1. Dans une nouvelle feuille de calcul, copiez-les données pour les variables : Espèce, Longitude et Latitude **dans cet ordre précis**.
2. Vérifiez que les décimales dans Excel sont séparées par des points
 - a. Allez à *Office Button/Excel*
 - b. Dans l'onglet *Advanced*, sélectionnez *points* [.] pour séparer les décimales
 - c. Dans l'onglet *Advanced*, sélectionnez *commas* [,] pour séparer les milliers.

3. Sauvegardez le fichier CSV (*.csv) et ouvrez-le ensuite avec *Notepad*.
4. Assurez-vous que les différentes données sont séparées par des virgules. Si ce n'est pas le cas, effectuez les corrections nécessaires. Vous pouvez pour cela utiliser la fonction *Replace* dans le menu *Edit/Replace*.
 - a. Délimitations par points-virgules
 - b. Remplacement par des virgules
 - c. N'oubliez pas de sauvegarder les modifications effectuées dans le fichier
 - d. Enfin, suivez l'Etape 6 de la section précédente (consacrée à Excel 1997-2003) et ouvrez le fichier CSV (*.csv) dans Maxent.



4a



4b

2.2. Importation de données climatiques dans DIVA-GIS et Maxent

L'option *Bioclim/Domain* du menu *Modeling* dans DIVA-GIS vous permet de réaliser diverses analyses basées sur les données climatiques. Ces analyses comprennent, par exemple, l'identification de points aberrants connus comme valeurs atypiques (voir le Chapitre 4), la caractérisation d'une niche écologique, la prédiction de la distribution potentielle d'une espèce avec l'analyse des lacunes que l'on peut en déduire ou l'analyse des impacts du changement climatique (voir le Chapitre 6).

Pour réaliser l'une de ces analyses, vous devez d'abord importer les données climatiques dans DIVA-GIS. De telles données sont disponibles gratuitement depuis des bases de données mondiales. La base de données la plus couramment référencée est WorldClim (Hijmans et al., 2005), qui utilise 19 variables bioclimatiques dérivées (Busby, 1991, voir tableau) en plus des données mensuelles du climat (températures maximale et minimale et précipitations). Comparées aux paramètres couramment utilisés de la température et des précipitations, les 19 variables bioclimatiques sont plus directement reliées aux aspects physiologiques de la croissance des plantes et leurs valeurs ne dépendent pas du moment où un état particulier intervient, c'est-à-dire que peu importe que le mois le plus chaud soit Juillet (hémisphère Nord) ou Janvier (hémisphère Sud). Certaines variables bioclimatiques sont de simples paramètres climatiques basiques types (comme BIO1, la température moyenne annuelle, ou BIO12, le niveau cumulé annuel des précipitations), tandis que d'autres combinent température et précipitations en une seule variable (comme BIO 18, niveau cumulé des précipitations durant le trimestre le plus chaud). D'autres variables encore intègrent des aspects de saisonnalité (comme BIO4 pour la température, BIO15 pour les précipitations), qui peuvent aussi être importants pour déterminer la distribution d'une espèce.

Las 19 Variables bioclimatiques ³

BIO1	=	Température moyenne annuelle
BIO2	=	Ecart diurne moyen (température maximale – température minimale ; moyenne mensuelle)
BIO3	=	Isothermalité (BIO1/BIO7) * 100
BIO4	=	Saisonnalité de la température (Coefficient de variation)
BIO5	=	Température maximale de la période la plus chaude
BIO6	=	Température minimale de la période la plus froide
BIO7	=	Ecart annuel de température (BIO5-BIO6)
BIO8	=	Température moyenne du trimestre le plus humide
BIO9	=	Température moyenne du trimestre le plus sec
BIO10	=	Température moyenne du trimestre le plus chaud
BIO11	=	Température moyenne du trimestre le plus froid
BIO12	=	Précipitations annuelles
BIO13	=	Précipitations de la période la plus humide
BIO14	=	Précipitations de la période la plus sèche
BIO15	=	Saisonnalité des précipitations (Coefficient de variation)
BIO16	=	Précipitations du trimestre le plus humide
BIO17	=	Précipitations du trimestre le plus sec
BIO18	=	Précipitations du trimestre le plus chaud
BIO19	=	Précipitations du trimestre le plus froid

³ La liste originale et des informations complémentaires sont disponibles à : <http://www.worldclim.org/bioclim>

Les données climatiques sont importées dans DIVA-GIS à partir de fichiers CLM (*.clm), disponibles sur le site Web de DIVA-GIS (<http://www.diva-gis.org/climate>) ou bien sont préparées pour une zone d'étude spécifique à partir des données climatiques globales (voir l'Analyse 3.1.6). Les paramètres de base (les températures maximale et minimale et les précipitations), de même que les valeurs des 19 variables bioclimatiques peuvent être extraits à partir des fichiers CLM dans DIVA-GIS, soit sous forme de couches, soit sous forme de données associées à des points d'observation. Des fichiers CLM de résolutions diverses sont disponibles à l'adresse : <http://www.diva-gis.org/climate>.

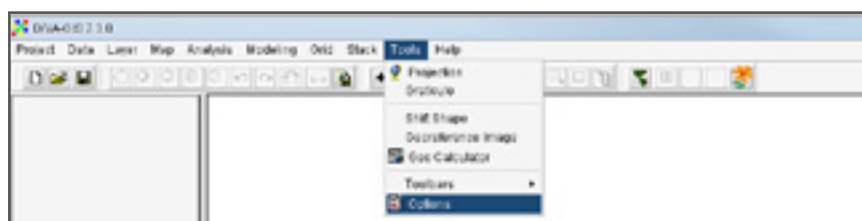
PROGRAMMES ET FICHIERS DE DONNÉES À UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel • Maxent et Java 	Fichiers de données : <p>Répertoire 2.2 Importing climate data (Importer des données climatiques)</p> <ul style="list-style-type: none"> • Répertoire diva_worldclim_2-5min (fichier zip) téléchargé depuis http://www.diva-gis.org/climate • Répertoire wclim_eth_2-5min_ascii (fichiers asc)

2.2.1. Comment importer des données climatiques dans DIVA-GIS

Cette section explique comme importer des données climatiques dans DIVA-GIS à partir de fichiers CLM (*.clm). DIVA-GIS utilise le format CLM pour stocker et lire les données climatiques spatialisées.

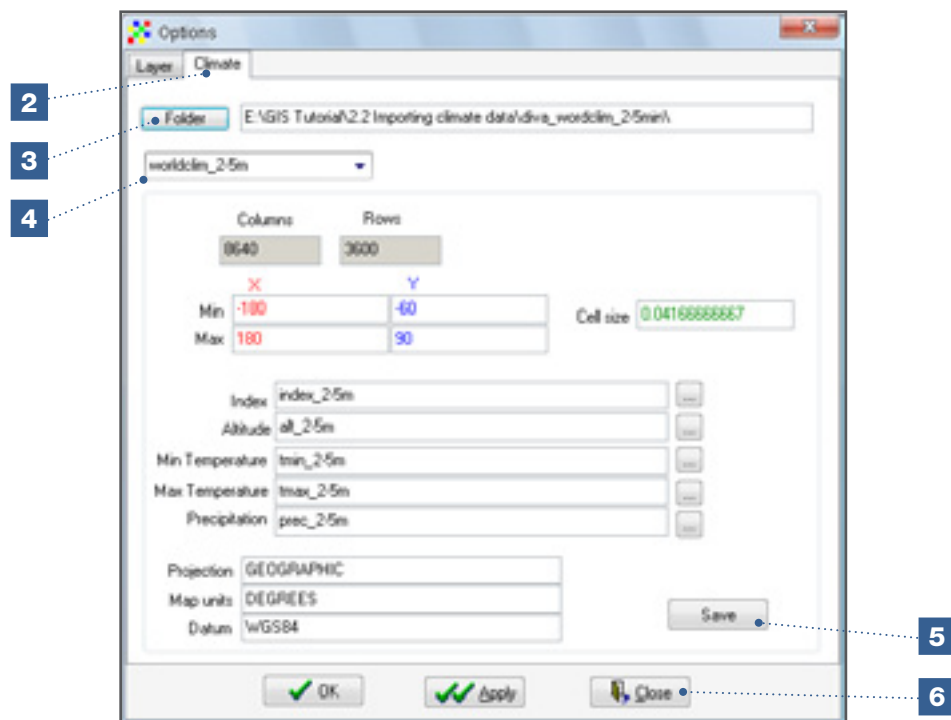
Etapes :

1. Copiez ou extrayez les données climatiques à partir du fichier compressé *diva_worldclim_2-5min.zip* vers un répertoire de votre ordinateur. Dans cet exemple, les données sont sauvegardées en utilisant le chemin *C:\Program Files\DIVA-GIS\environ*.
2. Allez à l'onglet *Tools/Options* dans DIVA-GIS et sélectionnez l'onglet *Climate*.



3. Dans la fenêtre *Folder*, indiquez l'emplacement du répertoire qui contient les données climatiques que vous aimeriez importer. Dans cet exemple, les données se trouvent dans le répertoire : *C:\Program Files\DIVA-GIS\environ*.
4. Sélectionnez *worldclim_2-5m* comme base de données climatiques.
5. Cliquez sur *Apply* et sur *Save* pour en faire la base de données par défaut pour l'analyse des données climatiques dans DIVA-GIS.
6. Cliquez sur *Close* pour fermer la fenêtre *Options*.

7. Retournez à *Tools/Options/Climate* pour visualiser les données climatiques sélectionnées et pour vérifier si les données climatiques ont été effectivement ajoutées.

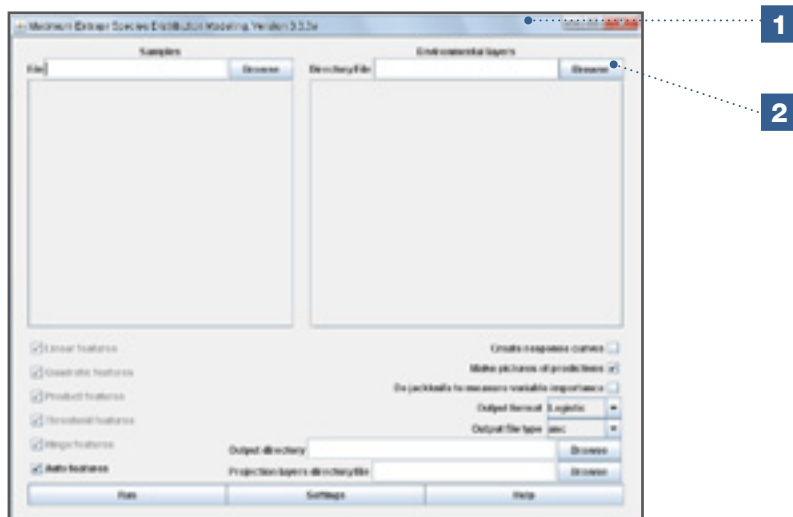


2.2.2. Comment importer des données climatiques dans Maxent

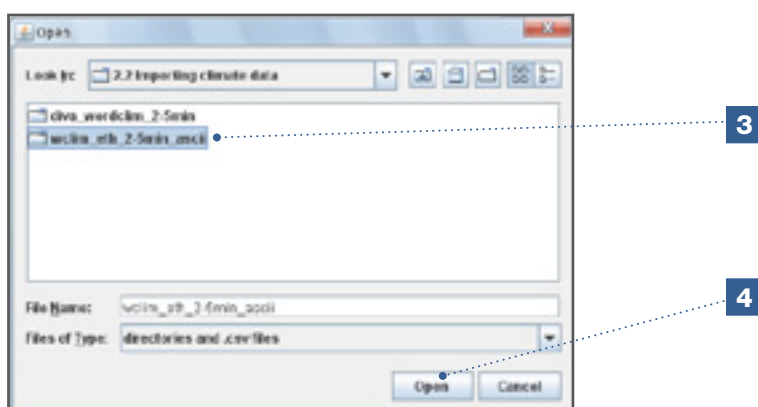
Afin que le processus de modélisation des données dans Maxent se déroule sans interruptions, il est recommandé d'utiliser des données d'un raster environnemental au format ASCII (*.asc). Les fichiers ASCII (*.asc) peuvent être créés dans DIVA-GIS avec l'option *Data/Export_Gridfiles*. L'Analyse 3.1.5 illustre comment préparer des données climatiques ASCII pour une zone d'étude donnée. La Section 6.2 explique comment utiliser Maxent avec des 19 variables bioclimatiques pour la prédiction de la distribution potentielle d'espèces. Pour l'exemple décrit ci-dessous, on utilise les 19 rasters au format ASCII (*.asc) placés dans le répertoire *wclim_eth_2-5min_ascii*.

Etapes :

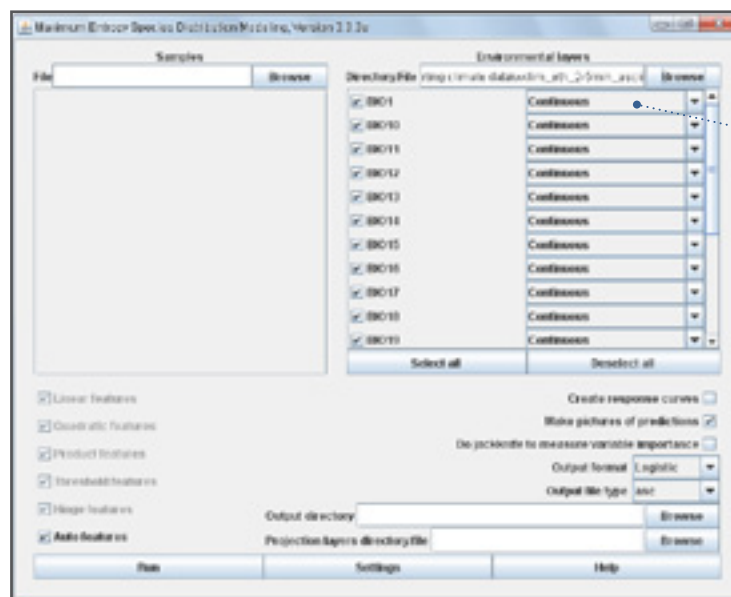
1. Ouvrez Maxent et indiquez dans la fenêtre *Environmental layers* quels sont les fichiers raster à inclure.
2. Utilisez le bouton *Browse* pour trouver le répertoire : *wclim_eth_2-5min_ascii*.



3. Une fois que vous avez trouvé le répertoire, sélectionnez-le en cliquant **une seule fois** (ne pas double-cliquer ni ouvrir le répertoire).
4. Cliquez sur *Open*.



5. Tous les fichiers raster stockés dans le répertoire apparaîtront automatiquement dans l'interface de Maxent.



Les couches environnementales sont maintenant prêtes à être utilisées dans les analyses de modélisation de la distribution des espèces qui est expliquée dans la Section 6.2.

2.3. Sources de données spatiales et autres données pertinentes

Il y a un nombre croissant d'organismes qui donnent accès gratuitement à leurs données spatiales, généralement dans le cadre d'un réseau pour lequel l'Internet facilite le partage et le téléchargement des données. Cette section liste les principales ressources en ligne d'où l'on peut récupérer des données environnementales, géographiques et de passeport.

Points de présence d'espèces

- Le Système mondial d'information sur la biodiversité (GBIF) : <http://www.gbif.org/>
- La liste des fournisseurs de données du GBIF peuvent être trouvés à l'adresse suivante : <http://data.gbif.org/datasets/>

Données de géoréférencement et données de niveau national

- Biogeomancer : <http://www.biogeomancer.org/>
- Geolocate : <http://www.museum.tulane.edu/geolocate/>
- Geonames : <http://www.geonames.org/>
- Google Geocoder : <https://developers.google.com/maps/>
- Données de niveau national : <http://www.diva-gis.org/gdata>

Information taxonomique

- Tropicos, Jardin botanique du Missouri : <http://www.tropicos.org/>
- Taxonomie des plantes de GRIN : <http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl>

Données environnementales

- WorldClim – données climatiques actuelles, futures et passées : <http://www.worldclim.org/>
- Information sur les sols du monde (ISRIC) : <http://www.isric.org/>
- Données d'altitude (STRM 90 mètres) : <http://www.cgiar-csi.org/data/elevation/item/45-srtm-90m-digital-elevation-database-v41>
- Downscaled GCM Data Portal (Portail pour télécharger les projections de données sur le climat mondial) : <http://www.ccafs-climate.org/>

Couverture végétale

- Global Land Cover Facility (Données sur la couverture végétale au niveau mondial) : <http://glcf.umd.edu/index.shtml>
- Global Land Cover 2000 (Données sur la couverture végétale au niveau mondial) : <http://bioval.jrc.ec.europa.eu/products/glc2000/glc2000.php>
- World Database on Protected Areas (WDPA, Base de données mondiale sur les zones protégées) : <http://www.wdpa.org/>
- World Wildlife Data and Tools (Données et outils sur les espèces sauvages au niveau mondial) : <http://www.worldwildlife.org/science/data/item1872.html>

Divers

- FAO Geonetwork (Réseau d'information géographique de la FAO) : <http://www.fao.org/geonetwork/srv/fr/main.home>
- CGIAR Consortium for Spatial Information (Consortium d'information spatiale du CGIAR) : <http://www.cgiar-csi.org/>
- DIVA-GIS : <http://www.diva-gis.org/Data>

Références

Busby JR. 1991. BIOCLIM a bioclimatic analysis and prediction system. In: Margules CR, Austin MP, éditeurs. Nature Conservation: Cost Effective Biological Surveys and Data Analysis. CSIRO, Canberra. pp. 64–68.

Hijmans RJ, Cameron SE, Parra JL., Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Chapitre 3

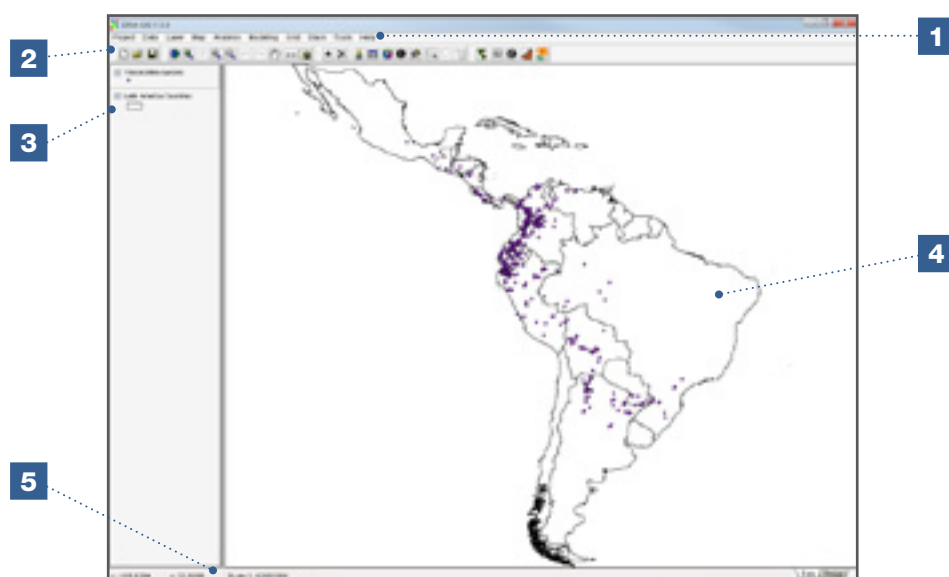
Éléments de base de l'analyse spatiale dans DIVA-GIS

Ce chapitre illustre comment utiliser les outils de base dans le programme DIVA-GIS pour réaliser des analyses spatiales courantes. Si vous souhaitez en savoir plus, vous pouvez consulter le Manuel de DIVA-GIS, disponible en ligne à l'adresse : <http://www.diva-gis.org/documentation>.

3.1. Visualisation dans DIVA-GIS

Avant de lancer les processus de visualisation, il est important de connaître les cinq sections qui sont présentées dans l'écran de travail DIVA-GIS.










1. **Menu bar** : facilite l'accès à toutes les commandes de DIVA-GIS
2. **Toolbar** : fournit un accès rapide à la plupart des tâches couramment utilisées
3. **Legend** : liste toutes les couches de la carte sur l'écran. La carte sélectionnée ressort. Les couches peuvent être visualisées ou cachées en cochant ou décochant les cases devant chaque nom de couche
4. **Map** : visualise de la carte courante
5. **Status bar** : Indique les coordonnées du point où se trouve le curseur, l'échelle de la carte et les position et valeur du raster.



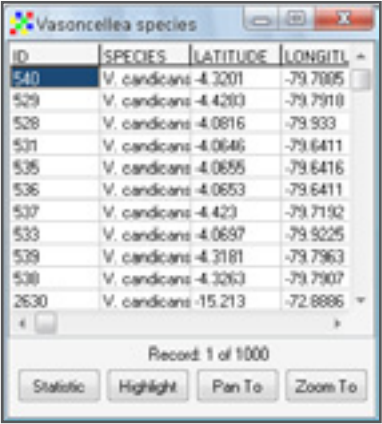
Le tableau suivant montre les principales commandes de navigation qui vous permettent de visualiser et explorer les divers aspects des cartes. Chaque commande est activée en cliquant sur le bouton correspondant.

Note

Selon l'information attribuée à chaque carte, certains boutons peuvent être inactifs.

Boutons	Commandes
	<i>Zoom in</i> (zoom avant sur un point avec le clic gauche de la souris ou zoom avant sur une zone en traçant un rectangle tout en tenant appuyé le bouton gauche de la souris)
	<i>Zoom out</i> (zoom arrière depuis un point avec le clic gauche de la souris)
	<i>Pan</i> (déplace la zone visible de la carte en tenant appuyé le bouton gauche de la souris en la déplaçant)
	<i>Zoom to active layer</i> (zoom arrière jusqu'à montrer l'ensemble de la couche active courante)
	<i>Zoom to full extent</i> (zoom arrière montrant l'étendue maximale de toutes les couches)
	<i>Information</i> (fournit l'information sur l'élément identifié par la souris dans la couche sélectionnée)
	<i>Remove Layer</i> (élimine la couche qui a été sélectionnée)
	<i>Table button</i> (voir les explications complètes ci-dessous)
	<i>Distance button</i> (permet de calculer la distance entre différentes caractéristiques)

Le bouton *Table* permet d'accéder au tableau des attributs d'un fichier vecteur (ces attributs peuvent être explorés avec les boutons *Statistic*, *Highlight*, *Pan to* et *Zoom to*). Le tableau est un fichier en 'lecture seule' et ne peut pas être modifié. Pour modifier les données, vous devez accéder au fichier original utilisé pour générer le fichier vecteur (voir 2.1.2). Il est cependant recommandé de ne pas modifier le fichier original mais de préparer un nouveau fichier de tableau avec les données modifiées. De cette manière, si une modification est invalide, vous pourrez toujours retourner aux données originales.



ID	SPECIES	LATITUDE	LONGITUDE
540	V. candicans	-4.3001	-79.7005
529	V. candicans	-4.4203	-79.7910
528	V. candicans	-4.0816	-79.933
531	V. candicans	-4.0646	-79.6411
535	V. candicans	-4.0655	-79.6416
536	V. candicans	-4.0653	-79.6411
537	V. candicans	-4.423	-79.7192
533	V. candicans	-4.0697	-79.9225
539	V. candicans	-4.3181	-79.7963
530	V. candicans	-4.3263	-79.7907
2630	V. candicans	-15.213	-72.8886

Record: 1 of 1000

Statistic Highlight Pan To Zoom To


- **Statistic** : dans le cas de valeurs numériques, le bouton *Statistic* fournit, dans une fenêtre séparée, un résumé des statistiques de base des variables numériques.

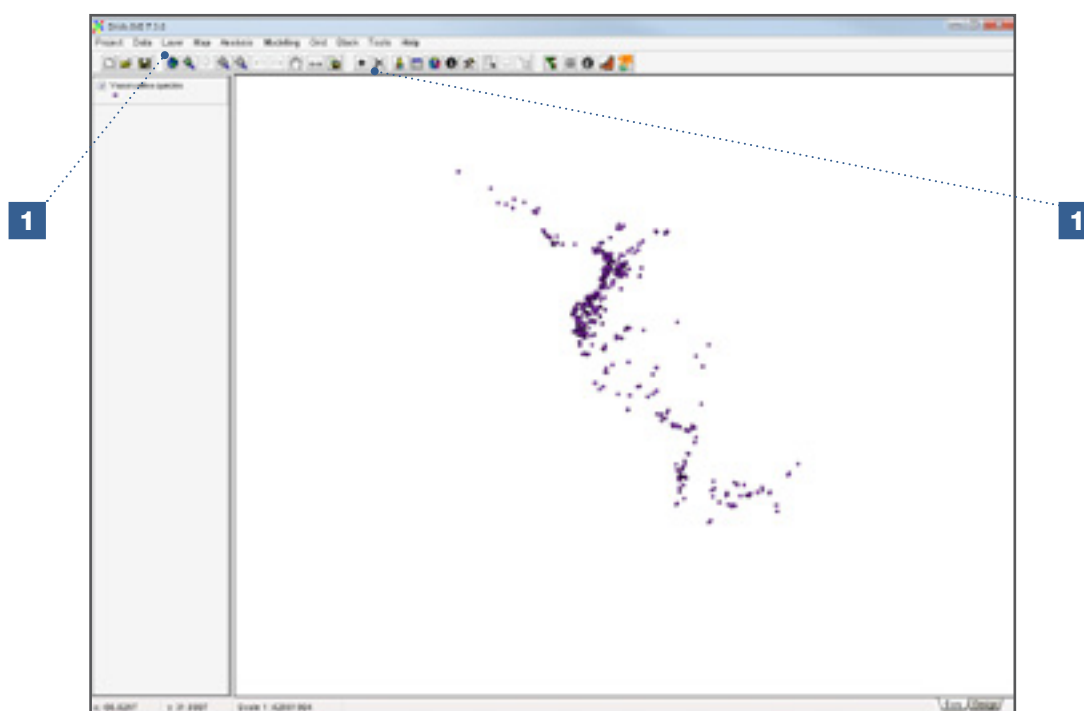
- **Highlight** : souligne les points de présence, les lignes ou les polygones sélectionnés pendant quelques secondes.
- **Pan to** : déplace la carte (avec la même échelle de zoom) vers l'emplacement où se trouve le point/ligne/polygone de présence que vous souhaitez visualiser peut se trouver hors de la portion de la carte initialement visualisée.
- **Zoom to** : cette option déplace aussi le point/ligne/polygone de présence vers le centre de la portion de carte visualisée, mais accroît le zoom.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel 	Fichiers de données : <p>Répertoire 3.1 Basic elements (Eléments de base)</p> <ul style="list-style-type: none"> • Vasconcellea species (shp, shx, dbf) • Latin America countries (shp, shx, dbf) • COL_ADM1 (shp, shx, dbf) • Mean temperature Latin America 10 min (grd, gri) • Precipitation Latin America 10 min (grd, gri) • World_adm0 (shp, shx, dbf) <p>Pour les analyses suivantes, vous devrez avoir importé les données climatiques de WorldClim à 2,5 minutes dans DIVA (voir la Section 2.2)</p>

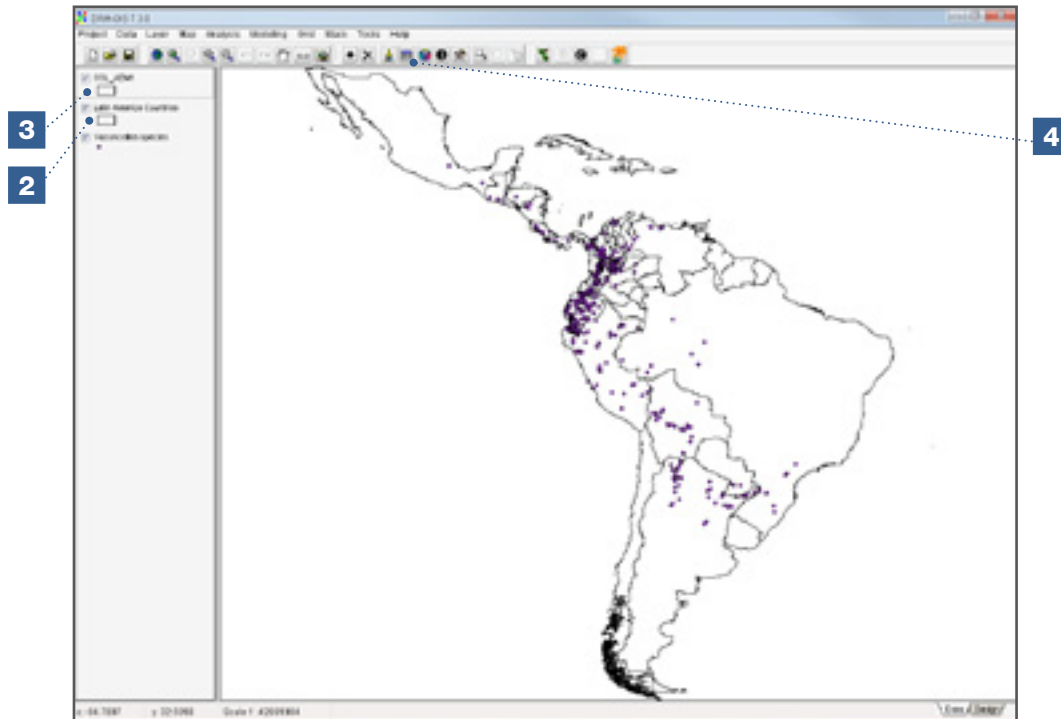
3.1.1. Comment effectuer des visualisations de base à l'aide des fichiers vecteur

Etapes :

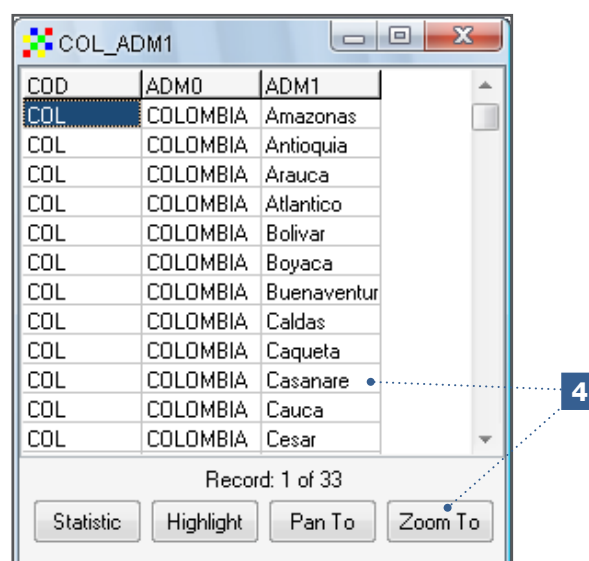
1. Ouvrez la couche de points de *Vasconcellea* (*Vasconcellea species.shp*) en utilisant l'icône  ou l'option du menu : **Layer/Add Layer**. Cette visualisation simple des points de présence ne fournissant que très peu d'information ; il est nécessaire d'ajouter d'autres couches pour compléter ces données.



2. Ajoutez la couche qui comprend les pays d'Amérique latine (fichier : *Latin America Countries.shp*). Cette couche devrait s'afficher immédiatement. Au fur et à mesure que l'on rajoute des couches, leurs noms apparaissent dans la colonne de gauche. Cochez (ou décochez) les cases devant les couches que vous souhaitez visualiser (ou cacher).
3. La Colombie est utilisée ici comme exemple pour fournir des détails supplémentaires durant l'analyse. Ajoutez la couche contenant les départements de Colombie (*COL_ADM1.shp*).

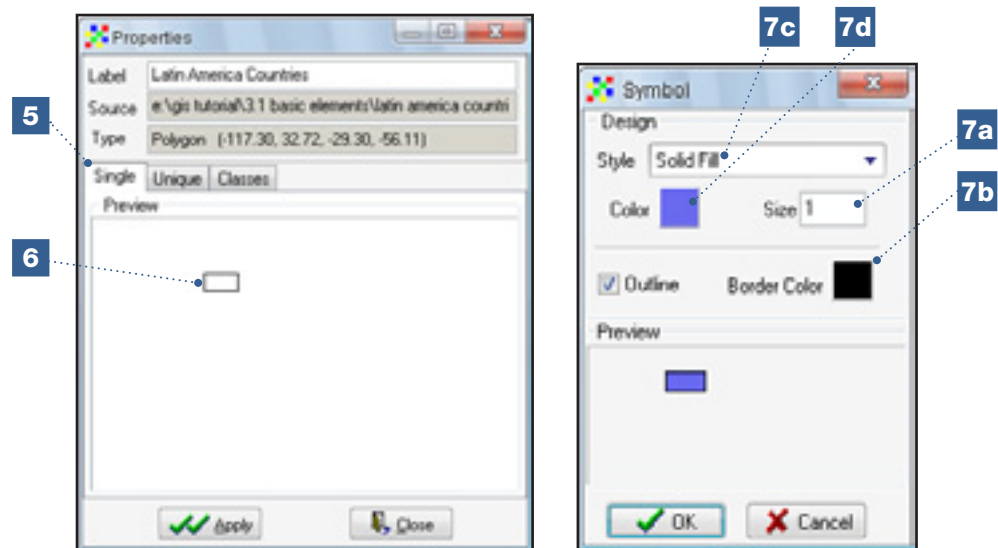


4. En sélectionnant la couche nouvellement ajoutée, on obtiendra l'information sur les divisions administratives de Colombie. Cliquez sur cette couche puis sur l'icône de tableau. L'information du tableau s'ouvrira et permettra de chercher des caractéristiques spécifiques dans le fichier *.shp. Par exemple, on pourra voir où se trouve le Département du Valle del Cauca, en utilisant les boutons *Highlight*, *Pan To* et *Zoom To*.

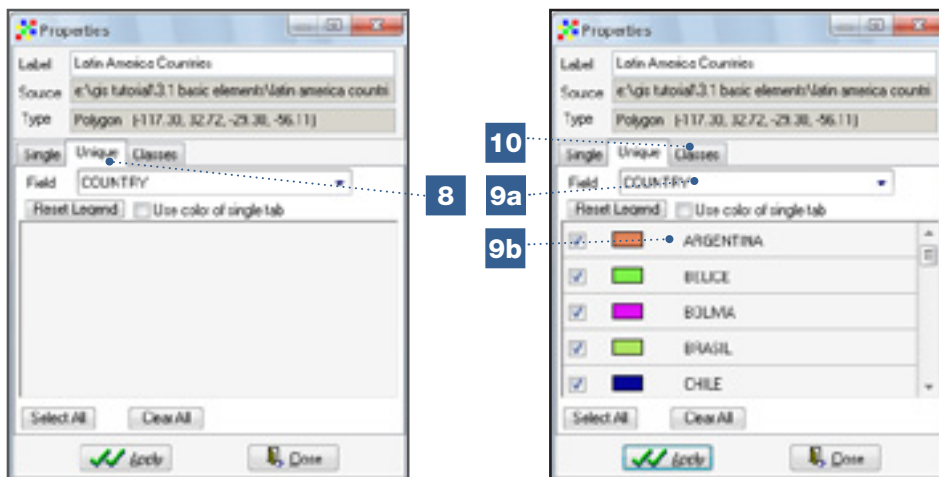


5. Cachez maintenant (décochez) les couches des espèces de *Vasconcellea* et retirez la couche des départements de Colombie. Étendez la carte jusqu'à visualiser tous les pays d'Amérique latine. Pour modifier la légende d'une couche, double-cliquez sur la couche à modifier. Pour cette analyse, cliquez sur la couche : *Latin America Countries*. La fenêtre *Properties* sera immédiatement affichée, avec l'onglet *Single* comme première option. Cette fenêtre vous permettra d'apporter des modifications à la légende.
6. L'onglet *Single* affiche un rectangle transparent. Un double clic sur le rectangle ouvre la fenêtre *Symbol*.
7. Les attributs de l'option *Symbol* vous permettent de changer certaines propriétés des polygones telles que :
 - a. L'épaisseur du contour
 - b. La couleur du contour
 - c. Le style de remplissage du polygone
 - d. La couleur du polygone.

Essayez de changer les valeurs par défaut de ces propriétés et observez ce qui se passe après avoir cliqué sur *OK* dans la fenêtre *Symbol* et *Apply* dans la fenêtre *Properties*. Pour une meilleure visualisation, le style recommandé est : *Solid Fill*.



8. Le mieux est d'utiliser des couleurs différentes pour les valeurs des différents attributs. Dans la fenêtre *Properties*, sélectionnez l'onglet *Unique*.
9. Dans l'onglet *Field*, sélectionnez le champ à utiliser pour créer la nouvelle légende de la couche. Il est important de sélectionner un champ ayant un nombre limité de classes. Si vous utilisez un champ qui contient des valeurs numériques (comme une surface), il se peut que le processus dans DIVA-GIS ralentisse fortement (puisque le programme considérera chaque valeur comme une classe différente). Pour cet exercice, sélectionnez le champ : *Country* [9a], et cliquez sur le bouton *Reset Legend*. Enfin, cliquez sur *Apply*. La légende de la couche sélectionnée sera affichée (les couleurs peuvent aussi être changées dans ce menu en double-cliquant sur le rectangle qui indique la couleur de chaque classe [9b]). Afin d'assigner une couleur différente à chaque classe, il est important d'avoir un style de remplissage sélectionné dans l'onglet *Single* (comme par exemple, *Solid Fill*).



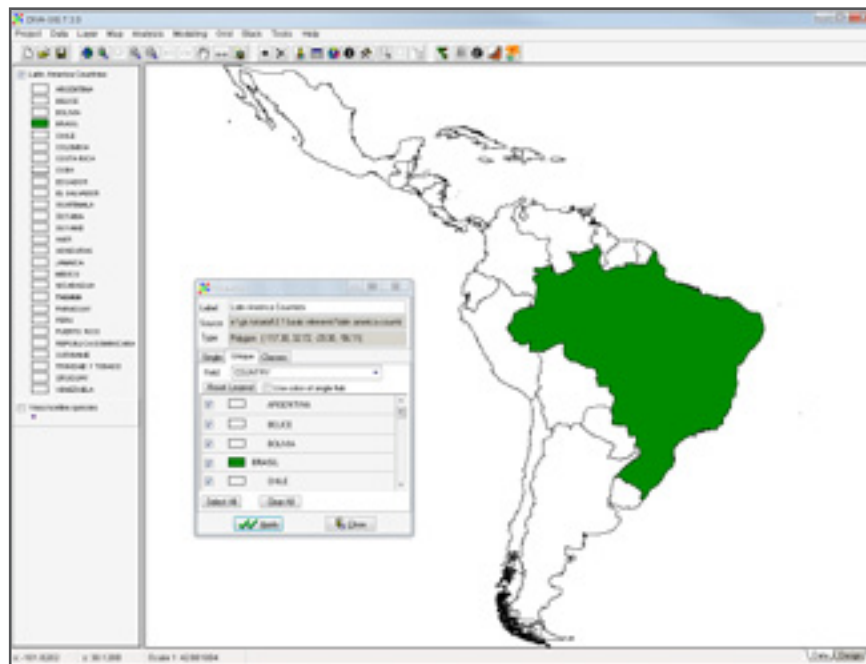
10. Un processus similaire est utilisé pour les valeurs numériques dans l'onglet *Classes* (ce processus n'étant pas utilisé couramment, il n'est pas inclus dans ce manuel.)

Si vous avez suivi les étapes correctement, la carte ci-dessous devrait être affichée (les couleurs peuvent différer) :



Travail personnel : Colorer le Brésil en vert et éliminer la couleur dans les autres pays.

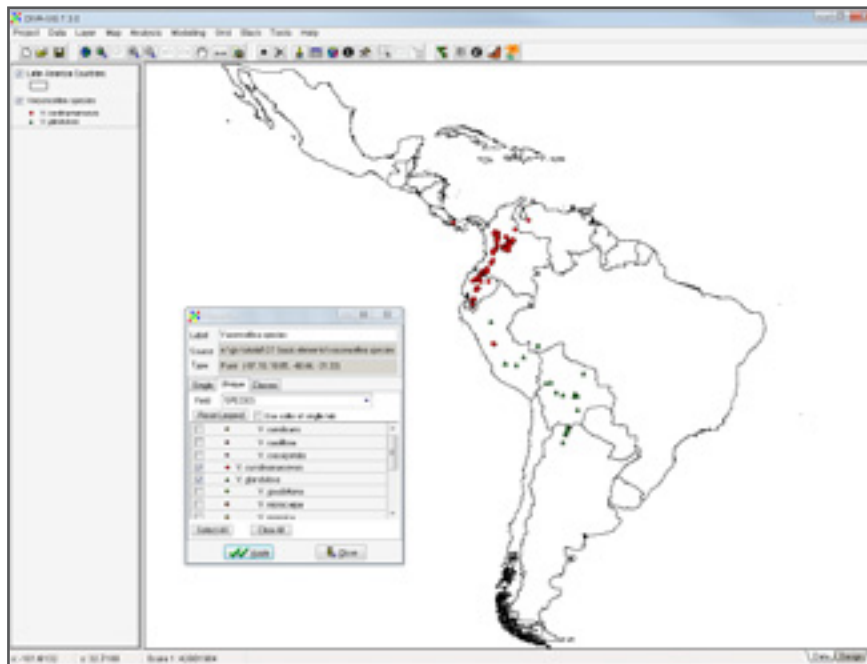
Résultat :




11. Maintenant, retirez la légende des pays (retournez à l'onglet *Single* et éliminez *Solid Fill*) et visualisez à nouveau la couche avec les points de *Vasconcellea*. Cette fois-ci, assignez une couleur différente à chaque espèce suivant la procédure expliquée dans les étapes précédentes.

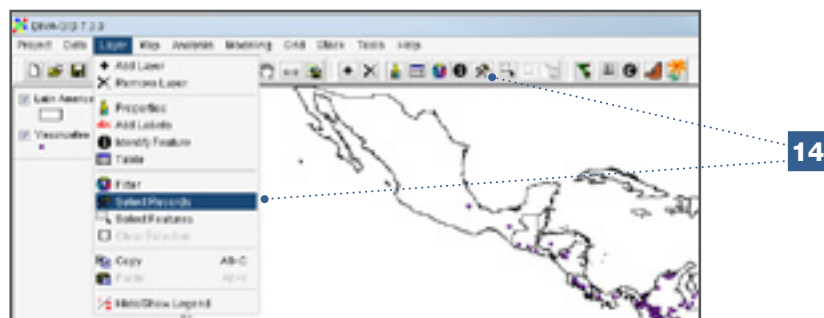


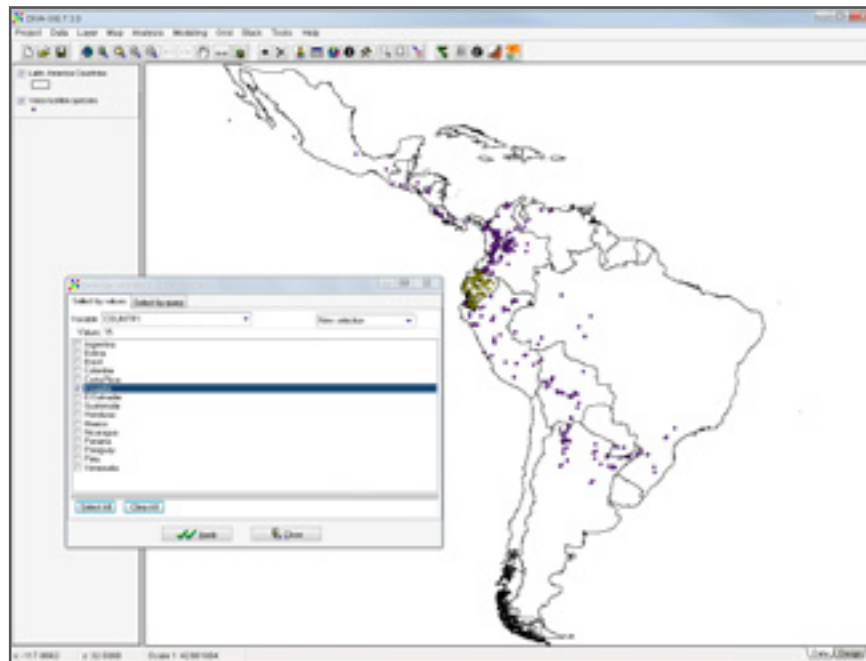
12. A l'aide des options de la fenêtre *Properties*, assignez un cercle rouge aux données de présence de l'espèce *V. cundinamarcensis* et un triangle vert aux données pour l'espèce *V. glandulosa*. Les autres espèces devraient rester invisibles.



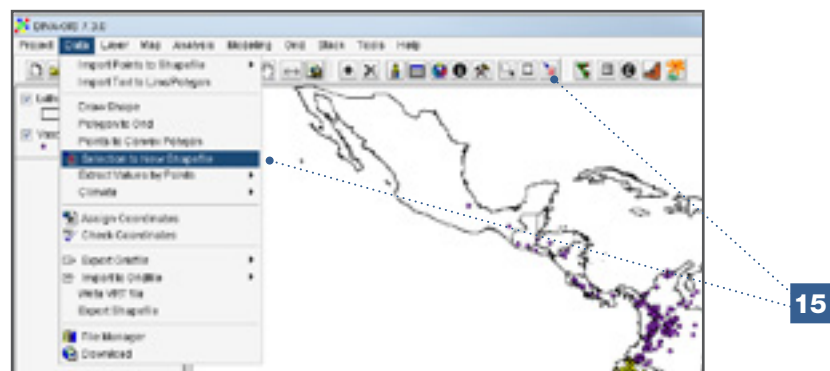
Il est important de réaliser qu'une sélection au niveau légende ne changera pas le contenu des fichiers vecteurs. Pour réaliser une analyse sur un sous-ensemble de données, vous devez d'abord sélectionner le sous-ensemble et le sauvegarder comme un fichier vecteur différent. Retournez à l'onglet *Single* dans la légende et sélectionnez un seul symbole pour tous les points. Toutes les observations sont à nouveau visualisées avec un seul symbole.


13. Maintenant, sélectionnez (n'affichez pas) un groupe donné d'espèces de *Vasconcellea* en utilisant l'une des alternatives fournies par DIVA-GIS dans le menu *Layer* : *Select Records* ou *Select Features* (ces options sont aussi disponibles en accès rapide dans la barre d'outils : boutons  et , respectivement).
14. Cliquez sur *Select Records* pour sélectionner des groupes correspondant à une variable donnée (*Select by values*) ou à une combinaison de variables (*Select by query*). Pour cette analyse, toutes les occurrences de *Vasconcellea* en Equateur ont été sélectionnées. Les points jaunes indiquent le groupe sélectionné.

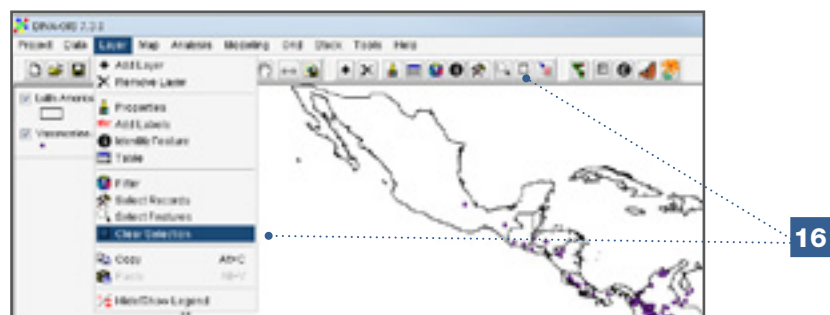




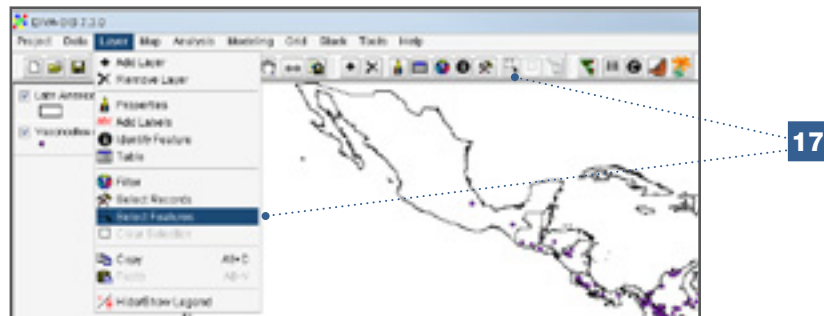
15. L'étape suivante consiste à sauvegarder la sélection en tant que nouveau fichier vecteur. Dans le menu *Data*, cliquez sur *Selection to New Shapefile* pour sauvegarder la sélection (espèces de *Vasconcellea* en Equateur) comme une couche additionnelle qui s'affiche maintenant sur la barre de légende. Cachez cette nouvelle couche.



16. Parfois, il est important de retirer des éléments sélectionnés. Pour ce faire, sélectionnez la couche qui contient les éléments sélectionnés et allez sur *Clear Selection* dans le menu *Layer*, ou bien cliquez sur le bouton correspondant dans la barre d'outils (). Les points sélectionnés (en jaune) ne sont alors plus surlignés.



17. Maintenant, effectuez une nouvelle sélection en désignant les points directement sur la carte à l'aide de l'option *Select Features* dans le menu *Layer*. Pour cet exercice, continuez à travailler avec la couche des espèces de *Vasconcellea* (gardez-la sélectionnée).



Le curseur prendra aussitôt la forme d'une croix (+), vous permettant de sélectionner manuellement un groupe de points (gardez le bouton gauche de la souris appuyé et tirez le curseur sur le groupe de points que vous voulez sélectionner). Les points sélectionnés vont changer de couleur. Une fois encore, la sélection peut être sauvegardée comme une nouvelle couche (Etape 15).

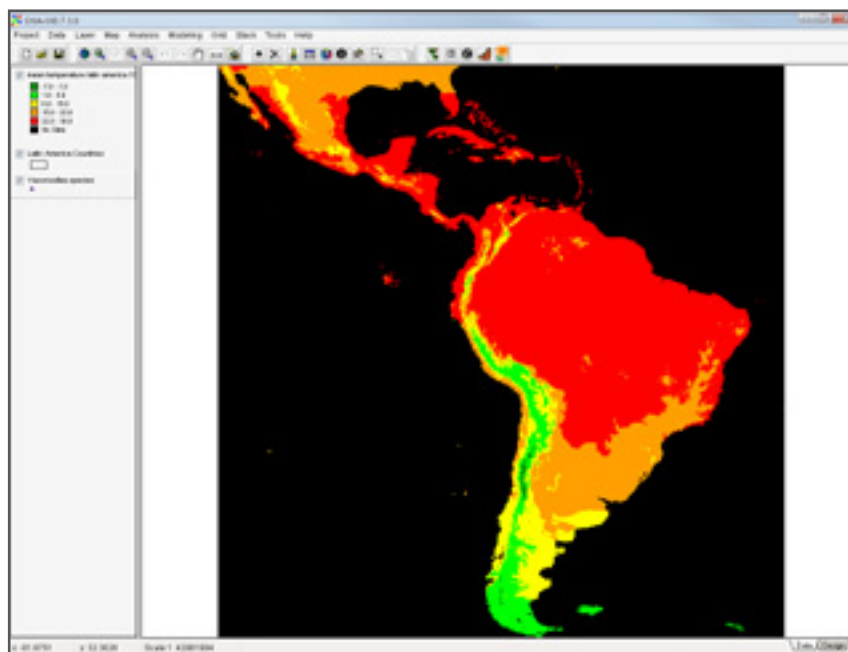


3.1.2. Comment réaliser des visualisations simples en utilisant des rasters

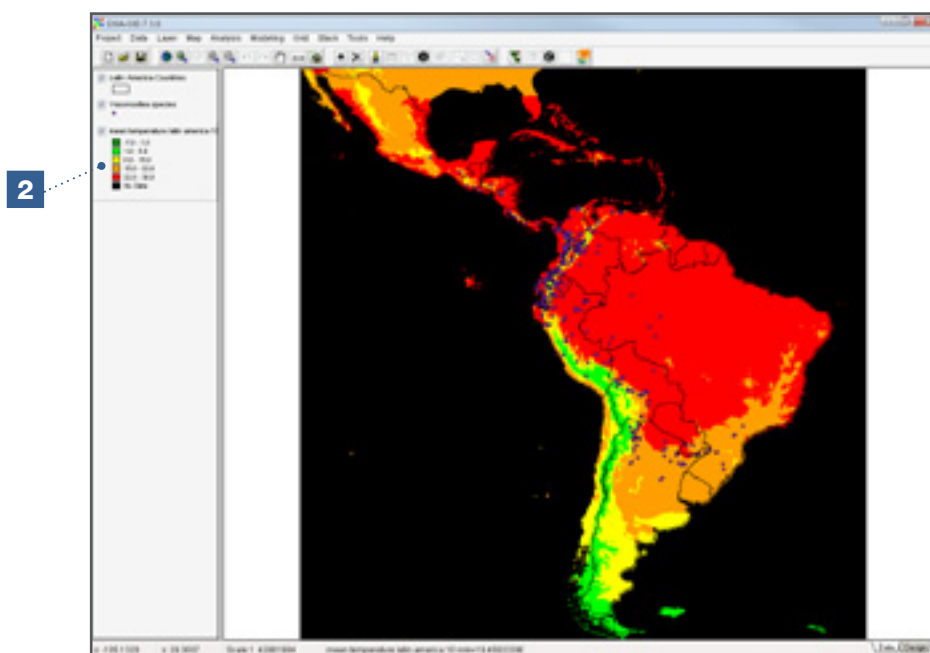
Etapes :



Comme indiqué au Chapitre 2, DIVA-GIS gère aussi les données raster (les fichiers raster ont une extension *.grd dans DIVA-GIS). En fait, la plupart des analyses décrites dans ce manuel se traduiront sous forme de rasters.

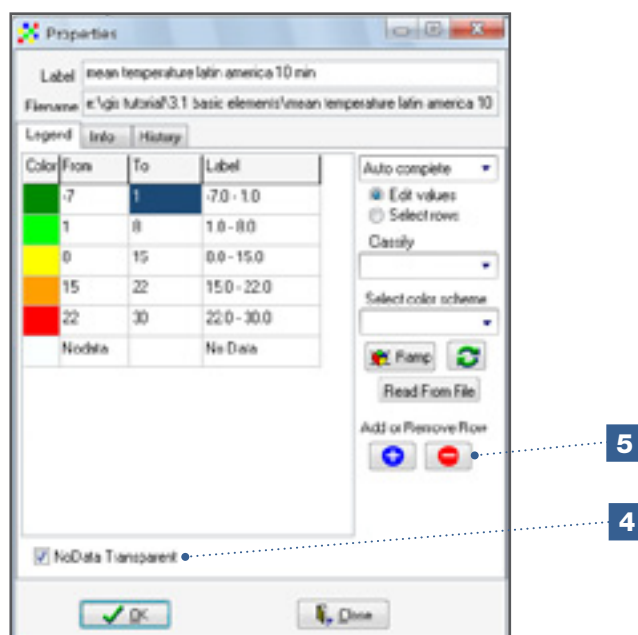
1. Ouvrez les couches des espèces de *Vasconcellea* et des *Latin America Countries*. Ajoutez la couche contenant les valeurs des températures maximales annuelles en Amérique latine (fichier : *Mean temperature Latin America 10 min.grd*). Cette couche correspond à un raster, qui est ajouté au panneau de légende de la même manière que précédemment avec un fichier vecteur.



2. Notez que la nouvelle couche recouvre les précédentes (c'est-à-dire les frontières de pays et les points de présence). L'ordre dans lequel les couches sont listées dans la colonne de légende correspond à celui dans lequel elles sont affichées (superposées). Pour modifier l'ordre, réorganisez les couches en les tirant avec la souris vers la position désirée. Pour continuer, remplacez la couche de température tout en bas afin de permettre aux autres couches (pays et points d'occurrence de l'espèce) d'être affichées en haut.



3. Les différentes classes de température sont exprimées sous forme d'intervalles (cette information apparait dans la légende). Par défaut, DIVA-GIS utilise cinq classes avec des intervalles égaux, ce qui est rarement la meilleure manière de visualiser les résultats. Les classes peuvent être ajustées à l'aide du menu *Propriétés* auquel on accède en double-cliquant sur la couche.
4. Sélectionnez la case *NoData Transparent* en bas de la fenêtre pour éliminer la couleur noire des cellules qui ne contiennent pas de données.
5. Assignez les nouvelles classes (par intervalles de cinq degrés) en commençant par la température minimale de -10°C jusqu'à atteindre la température maximale de +30°C. Cette classification formera huit classes (un intervalle de 40 degrés divisé par pas de 5 degrés). Vous devrez pour cela ajouter trois lignes dans le tableau par défaut. Les boutons plus et moins ( ) vous permettent d'ajouter ou de retirer des classes, selon le besoin.



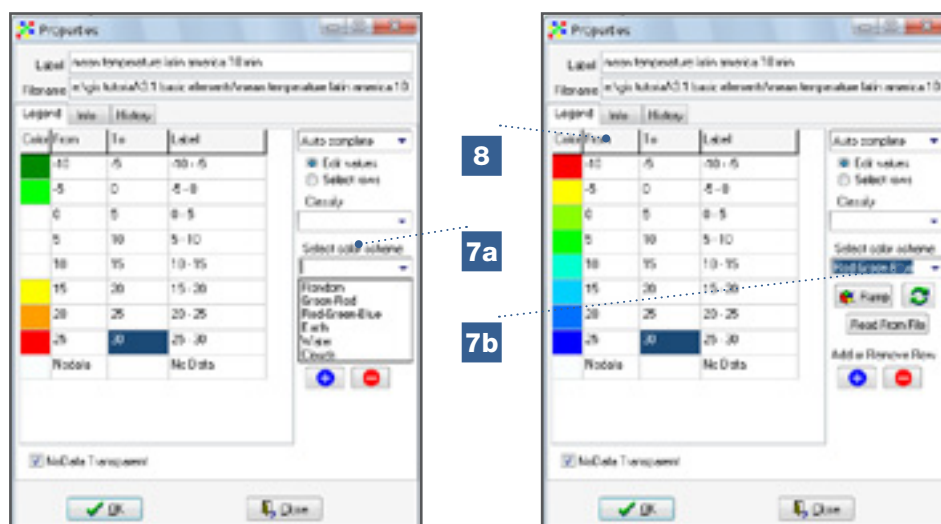
6. Avant de réorganiser les intervalles, ajustez les paramètres suivants :
 - a. Sélectionnez les commandes *Auto complete* et *Edit Values* (affichées dans la colonne de gauche) pour pouvoir saisir les valeurs dans les colonnes *From* et *To*.
 - b. Saisissez les valeurs des nouvelles classes dans les cellules, en commençant par le haut du tableau (*From* : -10 ; *To* : 5).
 - c. Après avoir saisi la première valeur, les cellules (ainsi que leurs étiquettes) dans la colonne *From* changeront automatiquement dans la ligne suivante ; il vous suffit de saisir les valeurs dans la colonne *To*.



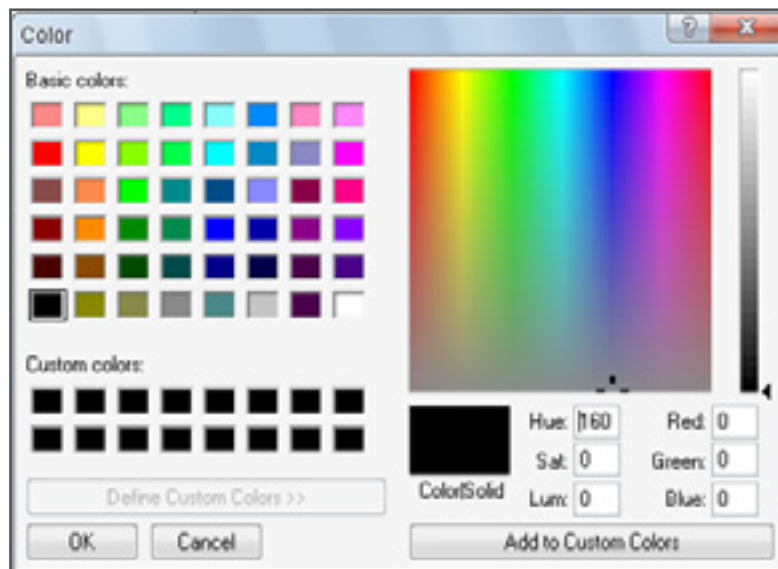
Note

Il est aussi possible de modifier manuellement les valeurs dans les colonnes *From* et *To*. Pour ce faire, sélectionnez l'option *Manual* dans la première case (voir 6a). Ceci vous permettra de modifier les valeurs dans ces colonnes. En choisissant cette option, faites attention car vous risquez d'oublier certaines valeurs, qui ne seront alors pas affichées sur la carte.

7. La couleur de chaque intervalle de température peut être modifiée à l'aide de la commande *Select Color Scheme* en sélectionnant les options disponibles, comme *Red-Green-Blue*. Les zones plus froides seront alors affichées en rouge et les plus chaudes en bleu (7a). Vous pouvez inverser l'ordre des couleurs en cliquant sur le bouton à deux flèches, surligné ci-dessous (7b).



8. Vous pouvez aussi personnaliser les couleurs, si vous préférez. Double-cliquez sur la première couleur et choisissez la nuance désirée (vous pouvez sélectionner soit *Basic colors*, soit *Custom colors*). Répétez ces actions pour les couleurs dans chaque classe.

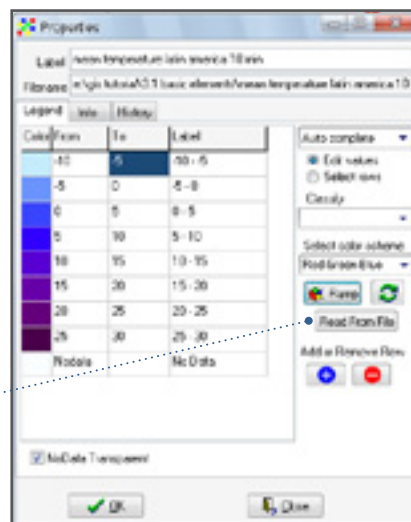


9. Le bouton *Ramp* vous permet de ne sélectionner que les nuances extrêmes ; cet outil suggérera automatiquement l'échelle des nuances intermédiaires. Pour cette analyse, sélectionnez les nuances extrêmes : bleu pâle et rouge bordeaux.
10. Dans certains cas, il faut que les légendes soient cohérentes entre différentes cartes. Pour utiliser l'information de légende d'un raster existant, cliquez sur l'onglet *Read From File* et indiquez l'emplacement du fichier (répertoire).

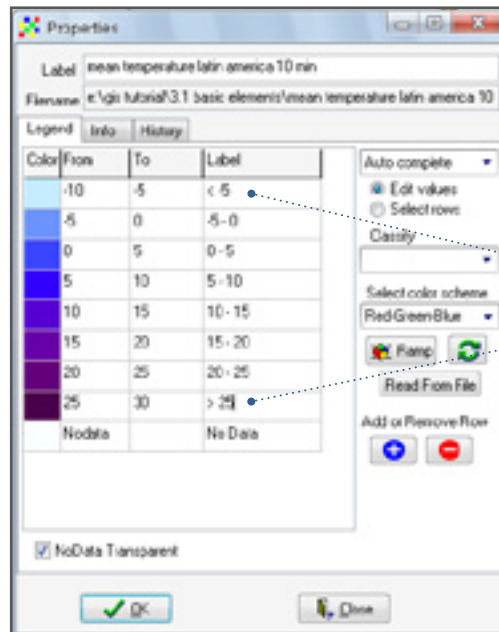


9

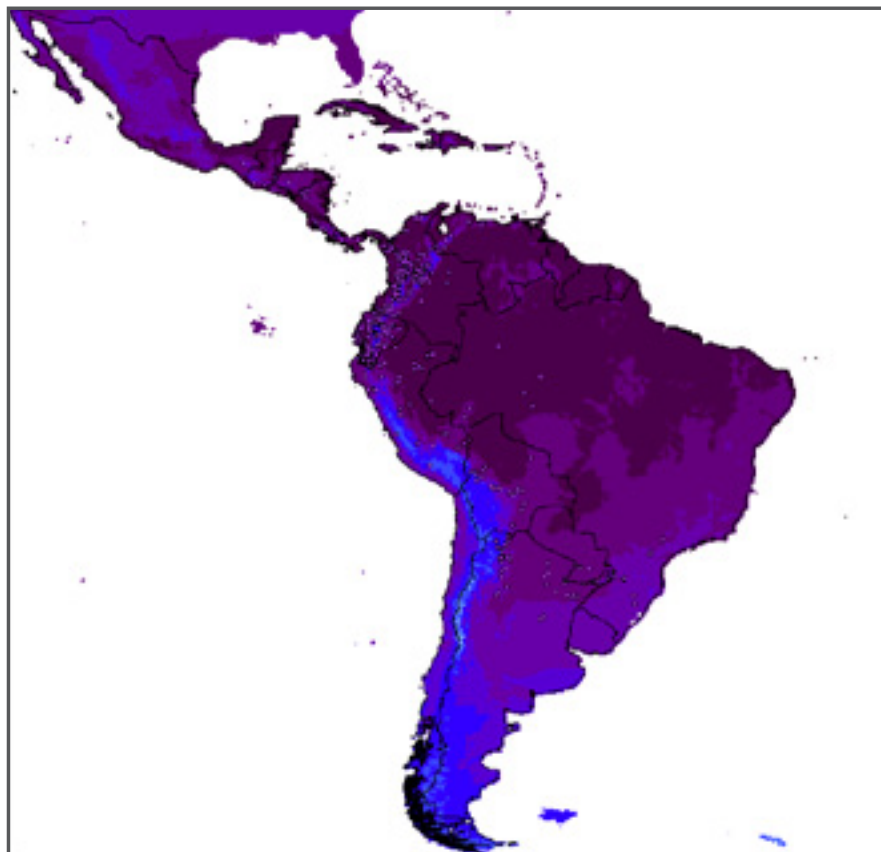
10



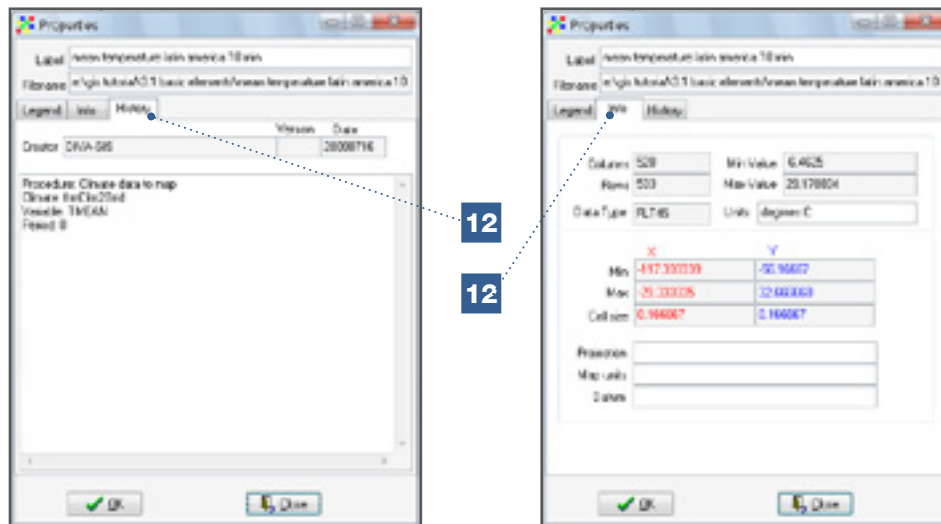
11. L'étiquette (texte) qui décrit chaque classe peut aussi être modifiée manuellement. Pour cet exercice, changez l'étiquette de la première classe : *From* [-10 - -5] *To* [< -5] et pour la dernière classe : *From* [25 - 20] *To* [> 25]. Cliquez sur *OK* pour afficher ces modifications sur la carte.



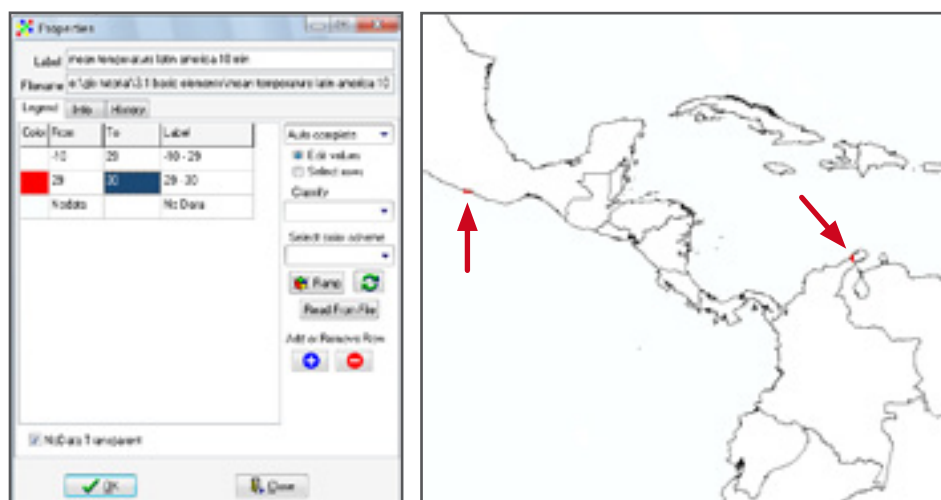
Si vous avez correctement suivi la procédure, la carte affichée devrait ressembler à celle-ci :



12. Le bouton *Properties* comprend aussi des onglets supplémentaires pour d'autres informations. Dans l'onglet *History*, vous trouverez l'information sur la manière dont la couche a été générée (ce qui peut être utile en cas de résultat inattendu ou pour trouver les détails sur la manière dont l'analyse précédente a été réalisée), tandis que dans l'onglet *Info* vous trouverez l'information pertinente sur le raster, y compris sa résolution et ses valeurs maximale et minimale.



13. Dans cette analyse, cliquer sur l'onglet *Info* vous permet de voir que la température maximale est de 29,17°C, mais l'emplacement exact sur la carte qui correspond à cette valeur extrême est difficile à localiser.
14. Essayez de trouver les sites présentant la température la plus haute (29,17°C) en Amérique latine. Pour ce faire, cliquez sur l'onglet *Legend* et créez une classe unique qui contient les données de température supérieure à 29°C. Sélectionnez *red* comme couleur pour cette classe. La carte résultante devrait afficher quelques cellules du Mexique et de la Colombie. Maintenant, vous devez vérifier chaque cellule pour identifier celle qui présente la plus haute température. La barre d'état en bas de l'écran donne la valeur exacte pour chaque cellule. L'endroit le plus chaud en Amérique latine est situé dans le Sud du Mexique.

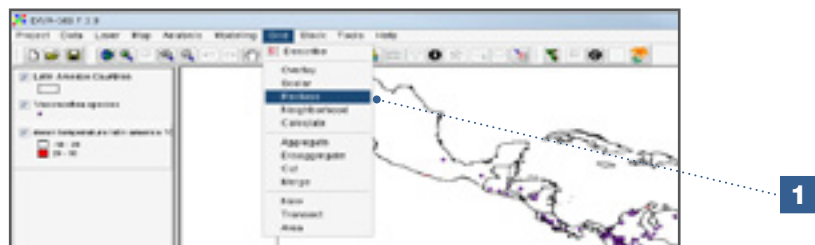


3.1.3. Comment combiner des rasters

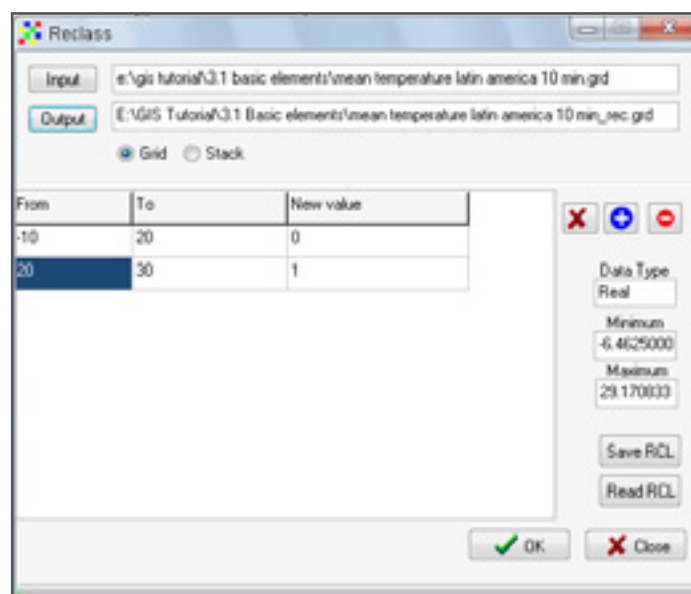
De même que pour les légendes des couches de vecteurs, les modifications dans les légendes des rasters n'affectent que l'affichage, et non l'information originale du fichier. Il est parfois intéressant de combiner des zones sélectionnées dans différents rasters ; par exemple, les zones chaudes d'un raster de température avec les zones sèches d'un raster de précipitations. Les rasters originaux contiennent de l'information à la fois sur les zones sélectionnées et sur celles qui ne le sont pas, ce qui les rend peu utiles pour de telles combinaisons. De ce fait, de nouveaux rasters, dédiés aux zones sélectionnées, doivent être créés pour combiner l'information qui vous intéresse.

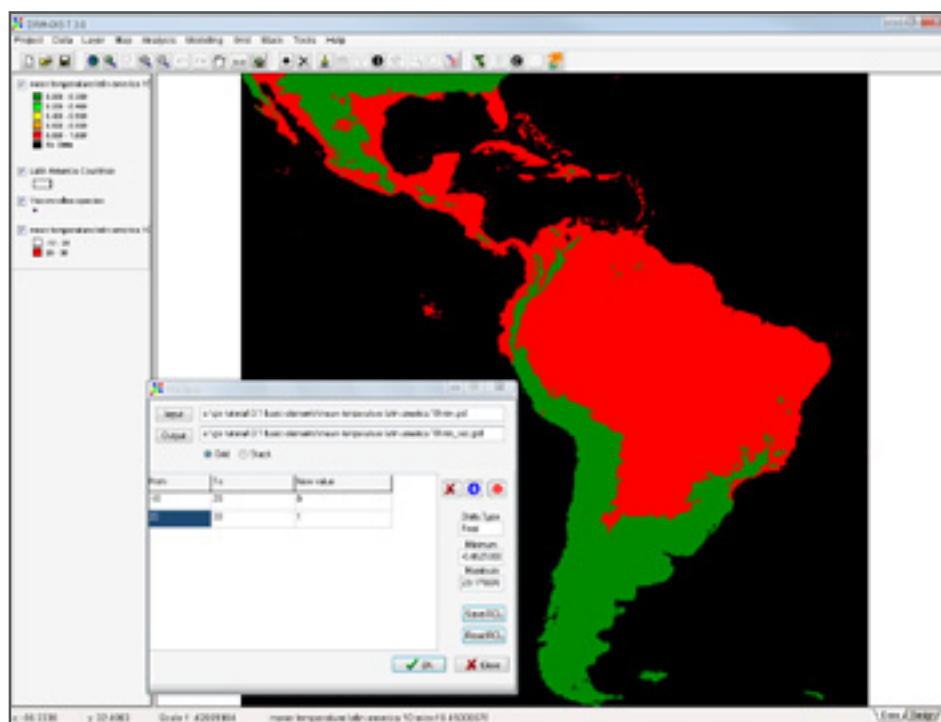
Etapes :

1. Pour cet exercice, sélectionnez toutes les zones d'Amérique latine présentant une température moyenne $> 20^{\circ}\text{C}$ et des précipitations < 1000 mm. Commencez par la température (couche : *Mean temperature Latin America 10 min.grd*), en sélectionnant toutes les zones dont la température moyenne dépasse 20°C . Dans le menu, allez sur *Grid/Reclass*.

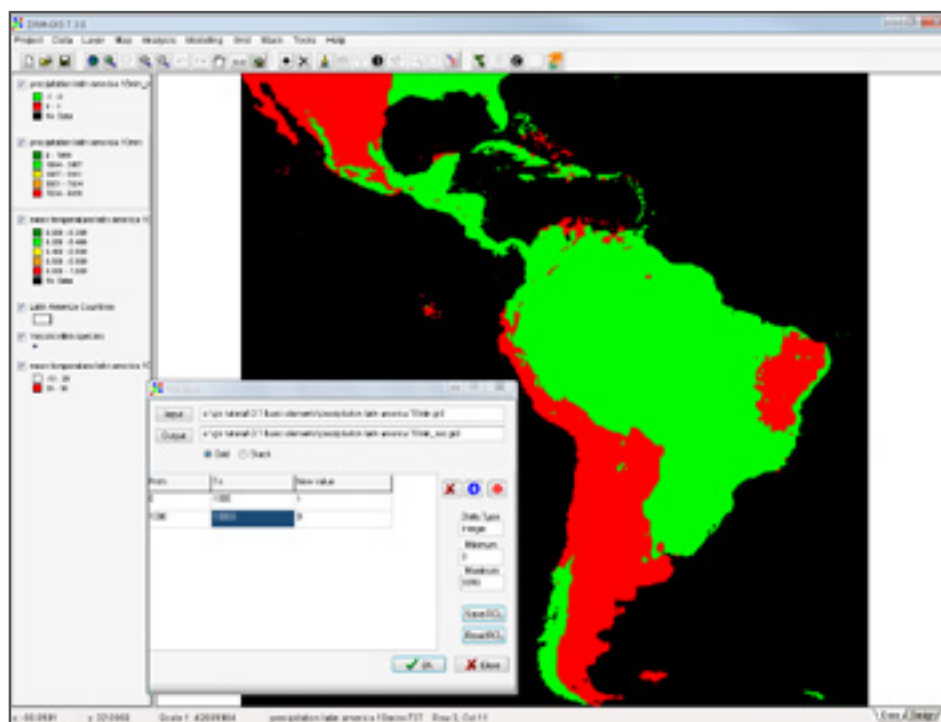


2. Transformez les valeurs en classes (assurez-vous de couvrir l'intervalle entier, de la valeur minimale à la valeur maximale), en assignant une valeur de zéro (0) aux cellules présentant des valeurs inférieures à 20°C et une valeur de un (1) à celles ayant une température supérieure à 20°C . Assignez un nom au nouveau raster généré, à l'aide du bouton *Output*. L'affichage qui en résulte doit seulement comprendre les cellules ayant une température moyenne supérieure ou égale à 20°C .

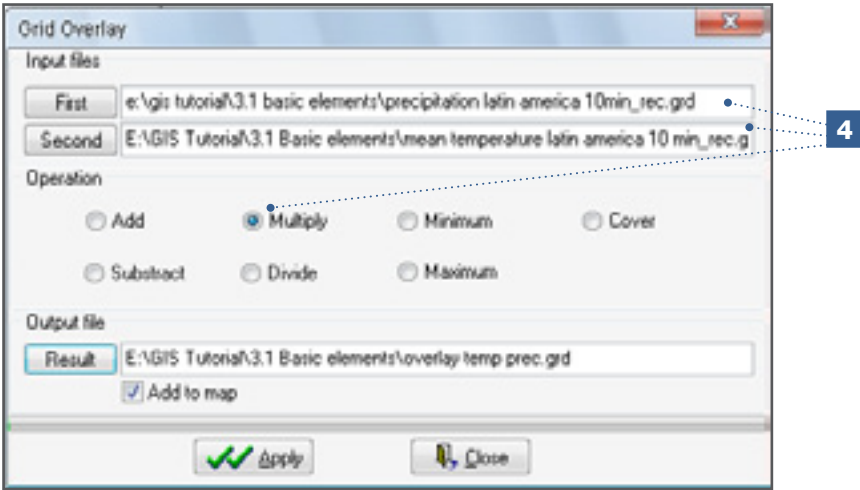




3. Ajoutez le raster des précipitations (*Precipitation Latin America 10min.grd*). En utilisant l'option *Reclass*, vous pouvez sélectionner les cellules où les précipitations sont en dessous de 1000 mm (voir l'Etape 1).



4. DIVA-GIS vous permet de combiner les rasters à l'aide de l'outil *Grid Overlay*. Vous devez maintenant identifier les zones qui présentent une température moyenne supérieure ou égale à 20°C et des précipitations annuelles en dessous de 1000 mm. Dans l'onglet *Input files*, sélectionnez les deux couches générées au cours des étapes précédentes et cliquez ensuite sur *Multiply*. Observez, dans le tableau ci-dessous comment se font les calculs.



	Température moyenne =>20 °C et Précipitations <1000 mm	Température moyenne =>20 °C et Précipitations >1000 mm	Température moyenne <20 °C et Précipitations <1000 mm	Température moyenne <20 °C et Précipitations >1000 mm
Valeur de la température de la cellule du raster	1	1	0	0
Valeur des précipitations de la cellule du raster	1	0	1	0
Combinaison (multiplication)	1	0	0	0

Le résultat est un raster où les cellules, avec une valeur combinée de un (1), indiquent les zones qui satisfont les deux conditions.



Il existe beaucoup d'autres options pour visualiser et manipuler les couches de vecteurs et rasters. Parmi les plus importantes :

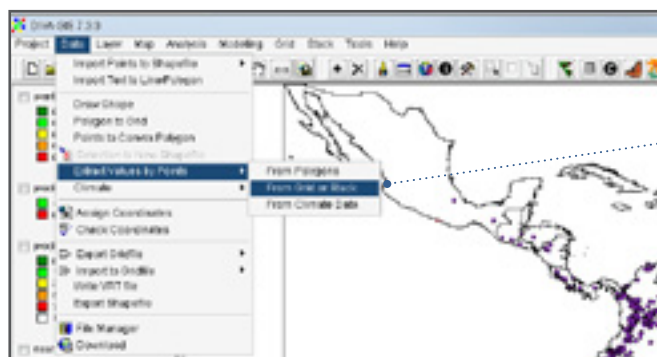
- *Grid/Aggregate* vous permet de combiner des cellules et diminuer la résolution d'un raster (voir la Section 6.4) et *Grid/Disaggregate* de diviser des cellules et accroître la résolution (notez que ceci peut engendrer une fausse impression de précision). Ces processus garantissent que deux rasters ont la même résolution et étendue, ce qui est indispensable pour pouvoir les combiner.
- *Stack/Calculate* vous permet de combiner trois rasters ou plus avec les mêmes caractéristiques (en terme de résolution, origine et taille) à l'aide de manipulations plus complexes que celles effectuées par *Overlay*.
- L'option *Grid/Cut* vous permet d'élaguer une portion d'un raster et de créer un jeu de données qui ne comprend que la zone (connue par son étendue) sur laquelle vous travaillez. Cette option est très utile quand vous utilisez une quantité importante de données. Elle permettra de réduire le temps de calcul lorsque vous lancerez une analyse comme la modélisation de la distribution d'une espèce, ainsi qu'il est expliqué au Chapitre 6. Ceci est particulièrement important lorsque vous utilisez un jeu de données volumineux.
- DIVA-GIS vous permet de modifier la projection des fichiers vecteurs. En sélectionnant *Tools/Projection*, vous pouvez effectuer des conversions entre différentes projections, comme entre latitude/longitude et UTM.

3.1.4. Comment extraire des valeurs de rasters en se basant sur des points de présence

Etapes :

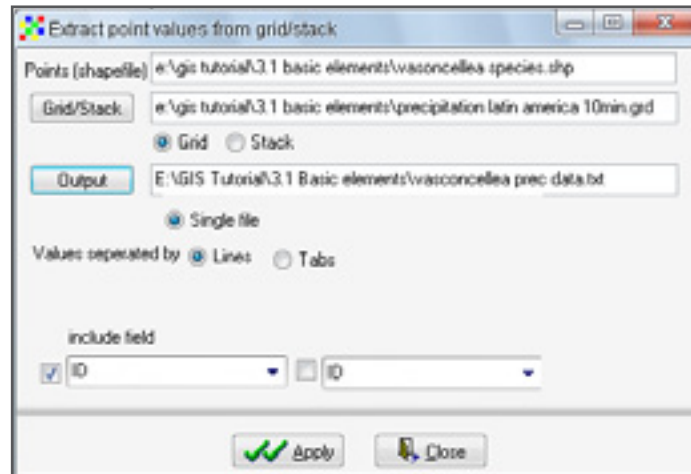
Les étapes précédentes décrivent comment afficher et manipuler des couches de vecteurs et rasters. Il sera souvent utile de combiner des données de vecteurs et rasters relatives à la même zone géographique. Dans l'exercice présenté ci-dessous, la valeur des précipitations est extraite du raster climatique pour chaque point de la base de données de *Vasconcellea* (fichier vecteur) en se basant sur sa localisation géographique. Cachez toutes les couches générées précédemment.

1. Sélectionnez la couche qui contient les points de *Vasconcellea* (*Vasconcellea species.shp*) et allez sur le menu *Data*. Sélectionnez *Extract Value by Points*, puis *From Grid or Stack*. Un *stack* est un groupe de rasters qui partagent des mêmes caractéristiques (en résolution, origine et taille).



1

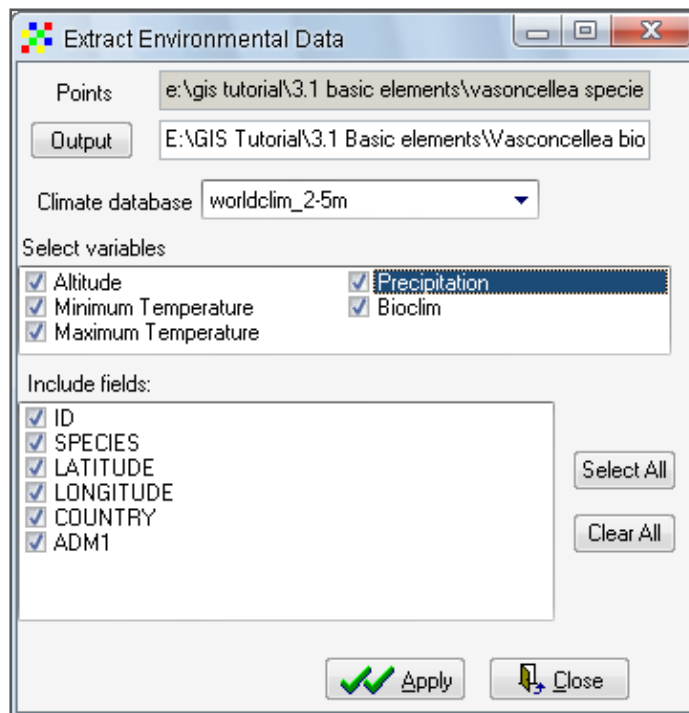
- Sélectionnez l'option *Grid* et marquez le raster dont vous voulez extraire les données (cet exercice utilise le fichier raster contenant les données de précipitations *Precipitation Latin America 10min.grd*). Laissez les valeurs par défaut (*Single File, Lines*) et insérez un nom pour l'*Output File* (qui sera un fichier texte [*.txt]). Vous devez utiliser les champs d'identification du fichier vecteur (*ID*) car ceux-ci sont des champs clés pour combiner le fichier texte généré avec les données de la base de données d'observation (fichier *.xls ou *.dbf).



- Lorsque vous ouvrirez le fichier résultant dans Excel, vous remarquerez que les identifications (*ID*) sont combinées avec les valeurs raster extraites. Vous pouvez maintenant utiliser l'information des deux sources, par exemple pour une analyse de niche écologique. Ceci est expliqué plus en détail dans la Section 6.1.

Microsoft Excel - vasconcellea prec data.txt									
File Edit View Insert Format Tools Data Window Services Help Adobe PDF									
A1 grid									
	A	B	C	D	E	F	G	H	I
1	grid	ID	values						
2	precipitation latin america 10min	540	1139						
3	precipitation latin america 10min	529	1010						
4	precipitation latin america 10min	528	1362						
5	precipitation latin america 10min	531	556						
6	precipitation latin america 10min	535	596						
7	precipitation latin america 10min	536	598						
8	precipitation latin america 10min	537	1019						
9	precipitation latin america 10min	533	1362						
10	precipitation latin america 10min	530	1139						
11	precipitation latin america 10min	538	1139						
12	precipitation latin america 10min	2530	526						
13	precipitation latin america 10min	542	484						
14	precipitation latin america 10min	543	21						
15	precipitation latin america 10min	544	22						
16	precipitation latin america 10min	2525	503						
17	precipitation latin america 10min	2010	2406						
18	precipitation latin america 10min	2913	2912						
19	precipitation latin america 10min	2915	1716						
20	precipitation latin america 10min	2920	3117						

4. Enfin, si vous souhaitez extraire toutes les données climatiques, vous pouvez répéter l'opération, mais au lieu de sélectionner l'option *Extract Values by Points/From Grid or Stack*, vous devrez sélectionner l'option *Extract Values by Points/From Climate data*. Ceci peut se faire si les fichiers CLM (*.clm) contenant les données climatiques sont connectés à DIVA-GIS (voir les étapes décrites dans l'Analyse 2.2.1). Le fichier résultant sera un fichier texte qui comprendra toutes les données climatiques, ainsi que les données originales (si elles ont été sélectionnées).



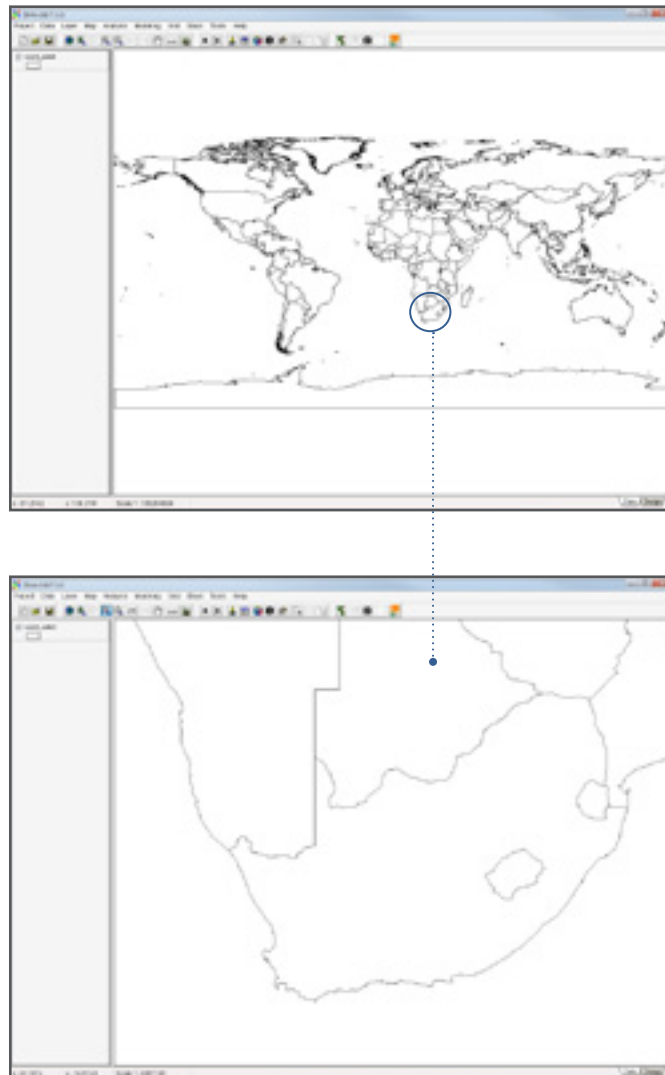
3.1.5. Comment créer des couches climatiques personnalisées

Etapes :

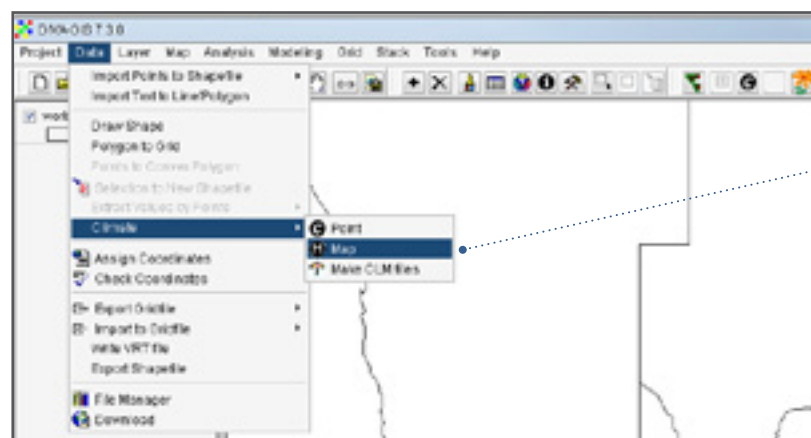
Outre l'extraction de données climatiques spécifiques d'un site à partir de l'information contenue dans les fichiers CLM (*.clm), DIVA-GIS permet aussi de générer des couches climatiques spécifiques à un paramètre (comme celles utilisées dans les Analyses 3.1.2, 3.1.3 et 3.1.4). Ces rasters vous permettent non seulement de mieux comprendre les conditions climatiques de la zone étudiée mais ils peuvent aussi être utilisés dans la modélisation de la distribution d'espèces (pourvu que tous les rasters partagent les mêmes propriétés).

Dans l'exercice présenté ci-dessous, vous apprendrez comment réaliser des couches climatiques en format ASCII (*.asc) pour l'Afrique du Sud en utilisant les fichiers CML (*.clm) à 2,5 minutes (contenant les données climatiques) (qui ont été liés à DIVA-GIS dans le Chapitre 2).

1. Ouvrez *World_adm0* dans DIVA-GIS et zoomez sur l'Afrique du Sud.

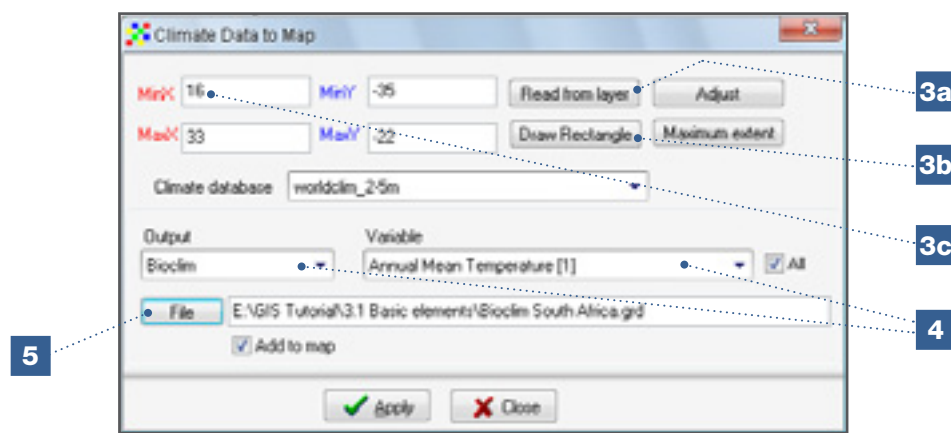


2. Allez sur *Data/Climate/Map*. Ceci ouvre la fenêtre *Climate Data to Map*.



3. Dans cette fenêtre, sélectionnez les propriétés du raster dont vous avez besoin pour la couche climatique. Vous pouvez le faire en :
 - a. Sélectionnant *Read from Layer* pour utiliser les caractéristiques d'un raster sélectionné dans la légende de DIVA-GIS. C'est utile quand vous souhaitez combiner des rasters et vous assurer qu'ils conservent les mêmes propriétés.
 - b. Sélectionnant *Draw rectangle* pour définir une zone en traçant, à l'aide de la souris, un rectangle dans la fenêtre DIVA-GIS.
 - c. Saisissant manuellement, et définissant les valeurs minimale et maximale des données de longitude (X) et de latitude (Y).

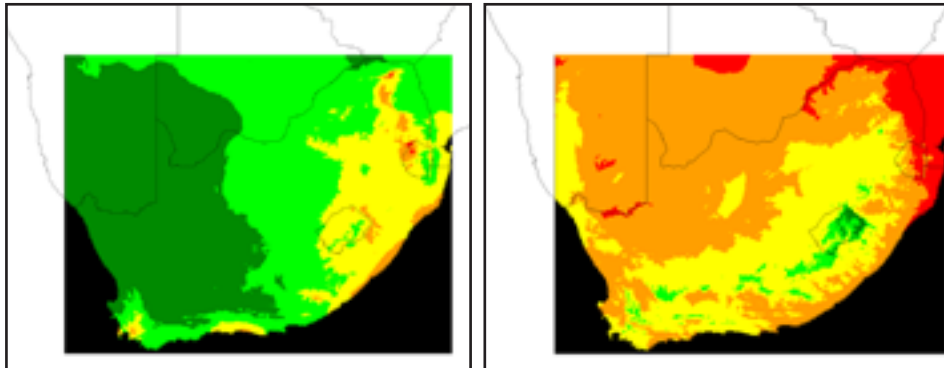
Dans cet exemple, c'est la dernière option qui est choisie : saisissez les valeurs fixes 16 et 33 pour X et -35 et -22 pour Y.



4. En plus de la définition du raster, vous devrez indiquer les couches climatiques à cartographier, que vous trouverez dans les menus déroulants *Output* et *Period/Variable*. On peut choisir les données simples de température et de précipitations (disponibles à intervalles mensuels) ou les variables bioclimatiques qui ont été présentées plus haut (voir le Chapitre 2). Comme la modélisation de la distribution d'espèces dans le Chapitre 6 sera fondée sur ces variables bioclimatiques, c'est celles-ci qui seront utilisées aussi dans cet exercice. Assurez-vous de cocher *All*, ce qui génèrera des couches pour les 19 variables bioclimatiques basées sur un raster identique.
5. Définissez dans l'onglet *File* le répertoire dans lequel les couches climatiques seront générées.
6. Dans le répertoire sélectionné, les 19 fichiers raster Bioclim seront créés comme suit :

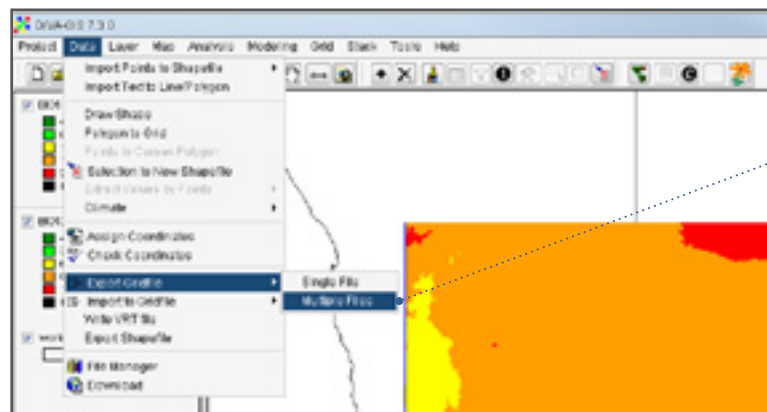
[illegible]

7. Grâce au tableau de la Section 2.2, dérivé des données issues de <http://www.worldclim.org/bioclیم>, nous savons que BIO1 se réfère à la température moyenne annuelle et BIO2 aux précipitations annuelles. Ajoutez ces deux couches à la carte.



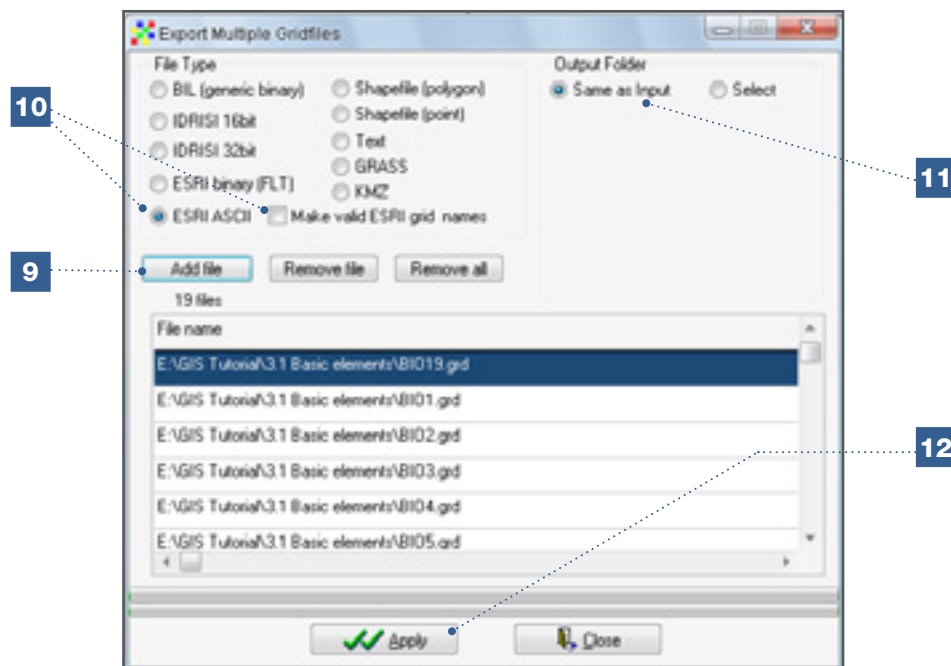
Comme noté dans l'introduction du Chapitre 2, les programmes de SIG et de modélisation de la distribution d'espèces utilisent divers types de fichiers raster. DIVA-GIS utilise le format .grd tandis que Maxent travaille mieux avec des rasters en format ASCII (.asc). DIVA-GIS vous permet d'importer et d'exporter des rasters dans des formats de fichiers compatibles avec différents programmes.

8. Afin d'utiliser les couches climatiques générées pour la modélisation de la distribution d'espèces dans Maxent, vous devez les convertir en fichiers ASCII (*.asc) en utilisant les options *Data/Export Gridfiles/Multiple Files*.



9. A l'aide du bouton *Add File*, sélectionnez les fichiers raster (*.grd) que vous souhaitez convertir en ASCII.
10. Pour exporter des fichiers raster *.grd en ASCII, sélectionnez l'option *ESRI ASCII*. Lorsque vous exportez des fichiers raster *.grd en ASCII, vous pouvez choisir de formater les noms de fichiers en sélectionnant l'option *Make valid ESRI grid names*. Cependant, ne choisissez pas cette option pour cette analyse.
11. Sélectionnez *Output Folder* et sauvegardez les rasters nouvellement formatés dans le même répertoire que les rasters du type du fichier original (option *Same as input*) ou dans un autre répertoire (*Select*).

12. Cliquez sur *Apply* pour lancer l'opération.



13. Les fichiers ASCII générés (*.asc) peuvent maintenant être utilisés dans Maxent pour la modélisation de la distribution d'espèces (voir l'Analyse 2.2.2).

3.1.6. Comment importer des données climatiques dans DIVA-GIS

Les couches climatiques créées dans la section précédente (à partir de *Data/Climate/Map*) sont dérivées des fichiers CLM (*.clm), en utilisant les données de résolution de 2,5 minutes décrites dans le Chapitre 2 pour DIVA-GIS. Le site Web de WorldClim (<http://www.worldclim.org/>) fournit des données climatiques plus fines (jusqu'à 30 secondes ou 1 km à l'équateur). Pour analyser une petite zone utilisant des points de présence de grande précision, ces données climatiques pourraient être les plus appropriées. Malgré tout, des rasters à une résolution de 30 secondes pour le monde entier prennent beaucoup d'espace sur un ordinateur, c'est pourquoi il est utile de télécharger les couches climatiques à 30 secondes de résolution par mailles (en anglais : « tiles ») de 30 x 30 degrés à partir du site Web de WorldClim (<http://www.worldclim.org/tiles.php>). Cette section illustre comment générer des fichiers climatiques raster (*.grd) à partir d'un fichier BIL (*.bil) de 30 secondes de résolution pour une région spécifique de 30 x 30 degrés, disponible depuis WorldClim.

Des couches sur le climat futur, que l'on peut télécharger à partir de WorldClim (<http://www.worldclim.org/futdown.htm>) ou du portail de données à échelle réduite GCM (<http://www.ccafs-climate.org/>) peuvent être importées dans DIVA-GIS de la même manière. Dans la Section 6.3, l'utilisation de données sur le climat futur dans les études d'impact du changement climatique sur les distributions d'espèces et la diversité est décrite avec plus de détail.

Etapes :

1. Allez sur le site Web : <http://www.worldclim.org/tiles.php>. Sélectionnez la maille 33 en cliquant dessus (ces quatre fichiers sont aussi inclus dans le répertoire *Basic Elements*).

Importer des rasters ArcGIS depuis DIVA-GIS

Des fichiers rasters créés dans DIVA-GIS peuvent aussi être importés dans ArcGIS. Pour ce faire, vous devez commencer par convertir les fichiers raster (*.grd) depuis DIVA-GIS en fichiers ASCII (*.asc) en utilisant l'option *ESRI ASCII* ou en les convertissant en fichiers FLOAT (*.flt), à l'aide de l'option *ESRI binary (FLT)*. Ces fichiers peuvent ensuite être importés dans ArcGIS.

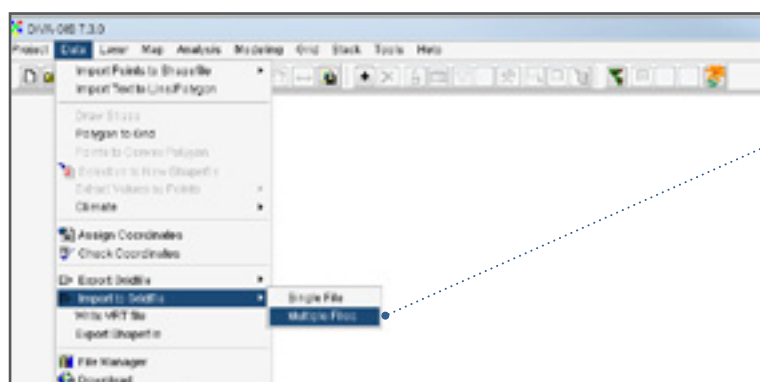
2. Téléchargez les jeux de données zippés pour *Minimum temperature*, *Maximum temperature*, *Precipitation* (il s'agit de données mensuelles) et *Altitude*. Sauvegardez les jeux de données sur votre ordinateur.

Note

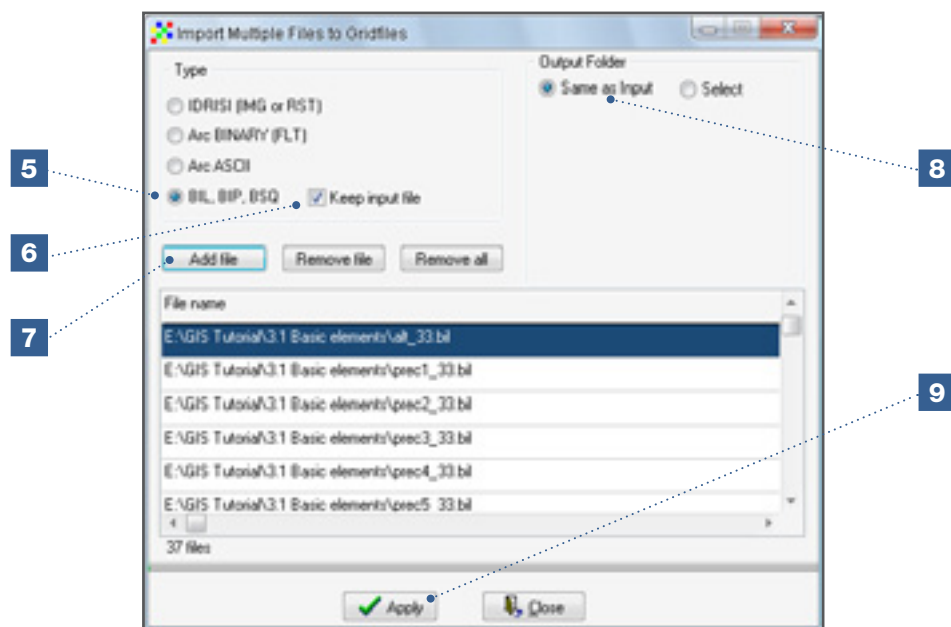
Vous remarquerez que les 19 variables Bioclim peuvent aussi être téléchargées. C'est utile pour les analyses de modélisation de distribution d'espèces avec Maxent (voir les Sections 6.2, 6.3 et 6.4). Mais pour cette analyse, on utilise seulement les couches climatiques mensuelles de *Minimum Temperature*, *Maximum Temperature*, *Precipitation* et le fichier raster *Altitude*. Vous utiliserez aussi ces couches dans l'Analyse 3.1.7 pour créer votre propre fichier CLM (*.clm).



3. Extrayez les couches des fichiers zippés.
4. Dans DIVA-GIS, allez sur *Data/Import to Gridfile/Multiple Files*.



5. Sélectionnez l'option *BIL, BIP et BSQ*.
6. Gardez les options par défaut du fichier d'entrée.
7. Cliquez sur *Add file* et sélectionnez les couches climatiques mensuelles de *Minimum Temperature*, *Maximum Temperature*, *Precipitation* et le fichier raster *Altitude* que vous voulez convertir en fichier *.grd.
8. Sauvegardez les fichiers raster générés (*.grd) dans le même répertoire *Output Folder* que les fichiers sous leur format original (*Same as input*). Vous pouvez aussi choisir de les sauvegarder dans un répertoire différent (*Select*).
9. Cliquez sur *Apply* pour lancer l'opération.



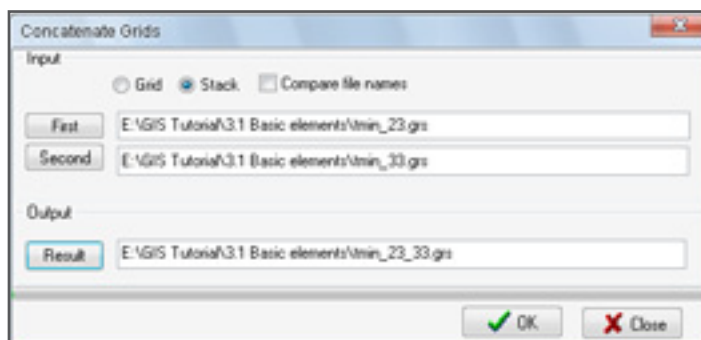
10. Maintenant vous pouvez ouvrir les couches climatiques dans DIVA-GIS à l'aide de l'option *Add layer*.

Note

La même procédure doit être suivie pour importer dans DIVA-GIS des données raster sous les formats suivants : IDRISI, Arc BINARY et Arc ASCII. Par exemple, les données sur le climat futur du portail de données à échelle réduite GCM (<http://www.ccafs-climate.org/>) sont disponibles au format ASCII.

Fusionner des rasters

Lorsque vous utilisez une couche climatique à 30 secondes de résolution sur une zone donnée, il est possible que celle-ci recouvre totalement ou partiellement plusieurs mailles de 30 x 30 degrés adjacents. Après avoir généré les fichiers raster (*.grd), vous pouvez fusionner les couches climatiques des mailles adjacentes dans DIVA-GIS à l'aide de l'option *Grid/Merge*. Commencez par regrouper les jeux de données à l'aide de l'option *Stack/Make Stack*, puis fusionnez-les à l'aide de *Grid/Merge*



Dans l'image ci-dessus, 12 couches de température minimale des zones 23 et 33 sont fusionnées.

3.1.7. Comment créer des fichiers CLM dans DIVA-GIS

Dans la section précédente, on a généré des jeux de données contenant des couches de précipitations et de températures minimales et maximales mensuelles. Quand on effectue une analyse écologique dans DIVA-GIS, il est souvent plus facile de travailler avec des fichiers CLM (*.clm) couvrant les 19 paramètres bioclimatiques qu'avec des couches climatiques individuelles. Par exemple, les données climatiques des fichiers CLM sont utilisées dans l'Analyse 3.1.4 et dans les Sections 4.2 et 6.1. Le site Web de DIVA-GIS <http://www.diva-gis.org/climate> fournit des bases de données (fichiers *.clm) des conditions climatiques actuelles jusqu'à une résolution de 2,5 minutes (5 km à l'équateur). Mais il se peut que vous vouliez préparer des fichiers CLM avec des données climatiques de 30 secondes de résolution, comme celles générées dans l'Analyse 3.1.6 ou comme plus haut avec les données du climat futur.

Cette section montre comment créer des fichiers CLM dans DIVA-GIS. A titre d'exemple, dans cet exercice, les fichiers CLM seront préparés à partir des jeux de données générés dans l'Analyse 3.1.6 (*Minimum temperature*, *Maximum temperature*, *Precipitation* et *Altitude*). Gardez en mémoire que tous les rasters doivent avoir la même couverture et la même résolution, comme c'est le cas pour les jeux de données générés précédemment. La préparation des fichiers CLM dans DIVA-GIS est aussi expliquée par Ramirez et Bueno-Cabrera (2009).

Étapes :

1. Chaque jeu de données (*Minimum temperature*, *Maximum temperature*, *Precipitation*) contient 12 fichiers correspondant à des valeurs mensuelles. Assurez-vous que les noms de ces fichiers sont différenciés par des numéros de 1 à 12 en mettant chaque numéro à la fin du nom de chaque fichier. A l'exception du numéro qui le termine, le nom devrait être le même pour chacun des 12 fichiers. Pour chaque jeu de données, c'est le préfixe qui sera utilisé pour développer le fichier CLM. Si ce n'est pas le cas, les noms de fichier doivent être modifiés en conséquence. Pour cette analyse, les jeux de données de la maille 33 et les noms des fichiers *.grd et *.gri doivent être modifiés manuellement.

Dans le cas du jeu de données *Precipitation* :

- *prec1_33.grd* et *prec1_33.gri* deviennent *prec1.grd* et *prec1.gri*,
- *prec2_33.grd* et *prec2_33.gri* deviennent *prec2.grd* et *prec2.gri*...
- *prec12_33.grd* et *prec12_33.gri* deviennent *prec12.grd* et *prec12.gri*.

Dans le cas du jeu de données *Minimum Temperature* :

- *tmin1_33.grd* et *tmin1_33.gri* deviennent *tmin1.grd* et *tmin1.gri*,
- *tmin2_33.grd* et *tmin2_33.gri* deviennent *tmin2.grd* et *tmin2.gri*...
- *tmin12_33.grd* et *tmin12_33.gri* deviennent *tmin12.grd* et *tmin12.gri*.

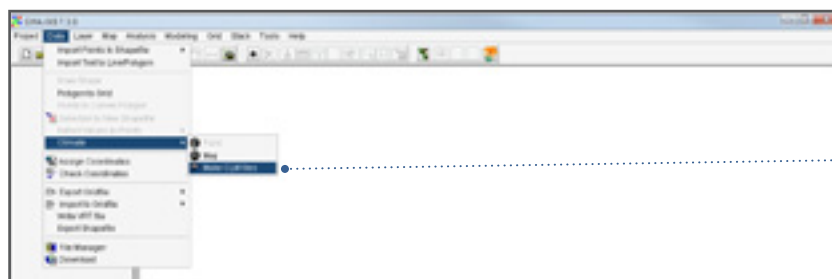
Dans le cas du jeu de données *Maximum Temperature* :

- *tmax1_33.grd* et *tmax1_33.gri* deviennent *tmax1.grd* et *tmax1.gri*,
- *tmax2_33.grd* et *tmax2_33.gri* deviennent *tmax2.grd* et *tmax2.gri*...
- *tmax12_33.grd* et *tmax12_33.gri* deviennent *tmax12.grd* et *tmax12.gri*.

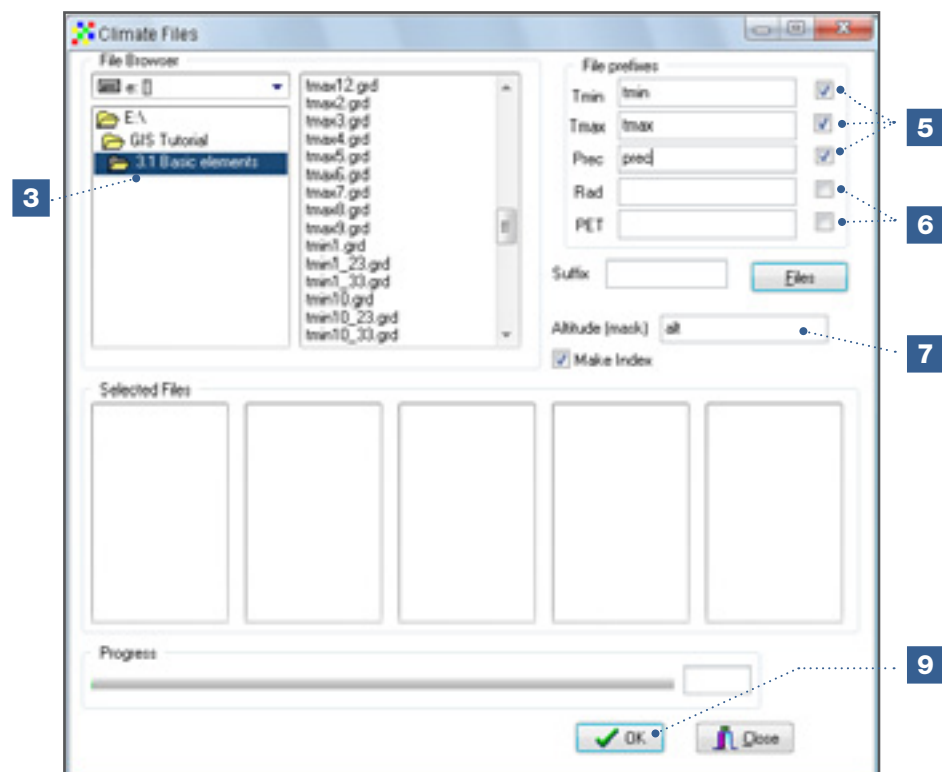
Note

Il est plus facile de renommer les 72 fichiers en utilisant un programme qui permet de renommer un groupe de fichiers (comme *Rename Master*).

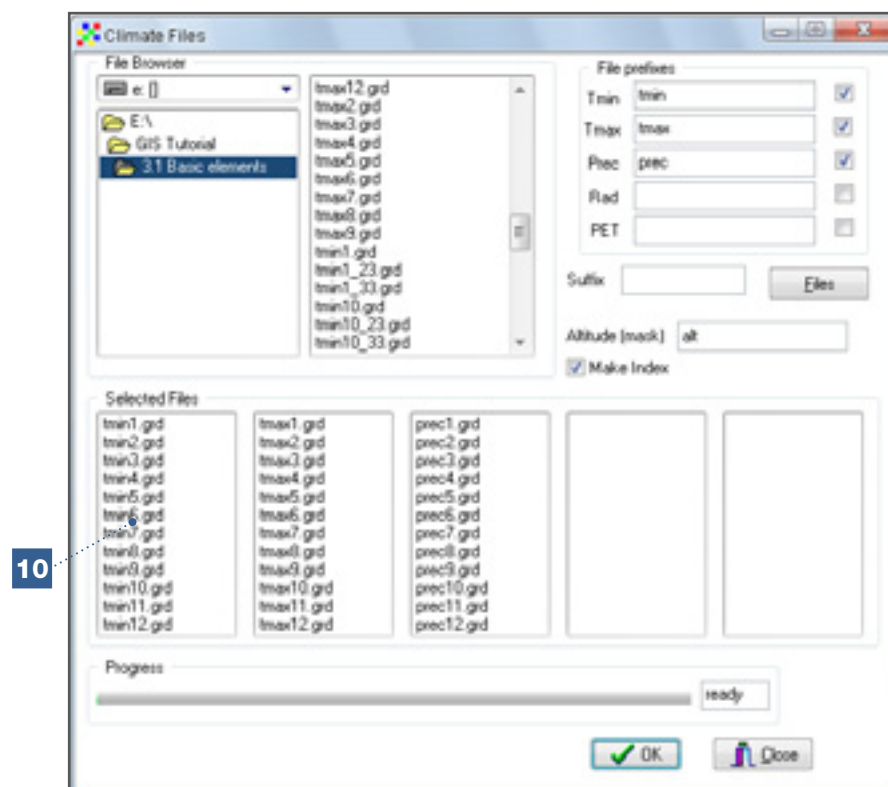
2. Après avoir modifié les noms des fichiers, allez sur *Data/Climate/Make CLM files*.



3. Dans le menu *File Browser*, sélectionnez le répertoire où vous avez sauvegardé les fichiers raster (*.grd) *Minimum temperature*, *Maximum temperature*, *Precipitation* et *Altitude*.
4. Tous les fichiers raster (*.grd) dans le répertoire indiqué sont affichés sur la droite dans le menu *File Browser*.
5. Indiquez les préfixes de fichier raster dans les cases *T min* (température minimale), *T max* (température maximale) et *Prec* (précipitations). Les préfixes comprennent tous les caractères avant le numéro (1-12). Pour cette analyse, le préfixe de *T min* est *tmin* ; le préfixe de *T max* est *tmax* et le préfixe de *Prec* est *prec*. Assurez-vous que les cases sur la droite des paramètres sélectionnés sont cochées.
6. Il y aussi une option qui permet d'inclure des données de radiation (rad) et d'évapotranspiration potentielle (PET). De telles données ne sont pas nécessaires pour les 19 variables bioclimatiques, et ces données ne sont pas actuellement disponibles auprès de WorldClim. Elles ne seront donc pas utilisées dans cette analyse.
7. Indiquez la couche d'altitude (alt).
8. Conservez toutes les autres options par défaut.
9. Cliquez sur OK pour lancer l'opération.



10. Lorsque vous aurez cliqué sur OK, les couches des trois jeux de données seront automatiquement importées dans les trois premières colonnes des fichiers sélectionnés.



Note

Générer des fichiers CLM (*.clm) peut prendre un certain temps si votre ordinateur n'est pas très puissant. Dans ce cas, il vaut mieux utiliser des couches climatiques de moindre résolution (5 ou 10 minutes).

Les fichiers CLM (*.clm) sont sauvegardés comme suit :

- climate.cli
- alt.clm
- index.clm
- tmin.clm
- tmax.clm
- prec.clm

Le fichier (*.cli) contient l'information sur les caractéristiques des fichiers CLM (*.clm). Il est recommandé de modifier le nom du fichier *climate.cli* en un nom plus spécifique afin de pouvoir distinguer les fichiers CLM (*.clm) des autres bases de données climatiques CLM. Dans cette analyse, par exemple, le nom du fichier *climate.cli* peut être changé en *wclim_t33_30sec.cli*. Ne changez pas les noms des fichiers CLM (*.clm) ; sinon ils ne seront pas reconnus dans DIVA-GIS.

Assurez-vous que les propriétés des rasters sont identiques avant de les combiner

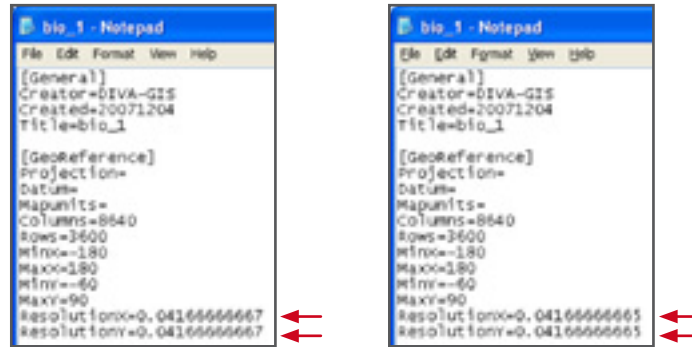
Des erreurs peuvent se produire au cours des calculs dans DIVA-GIS lors de la combinaison de deux rasters ou plus (par exemple en cas de recouvrement) ou lorsque deux *stacks* (groupes) de rasters sont utilisés en entrée de Maxent (pour comparer les zones de distribution potentielle d'une espèce dans différents scénarios environnementaux ou sous des climats différents ; voir la Section 6.3). Ces erreurs sont généralement dues aux différences entre les propriétés des rasters.

Les propriétés des rasters, c'est-à-dire leur résolution (la taille des cellules), et leur étendue [le nombre de lignes et de colonnes, les coins ou les sommets (les valeurs min et max de X et Y) des rasters], doivent être identiques pour que l'on puisse les combiner dans DIVA-GIS, comme nous l'avons expliqué dans l'Analyse 3.1.3, et pour que l'on puisse les utiliser en entrée de Maxent après les avoir convertis au format ASCII (*.asc). Pour ce faire, dans DIVA-GIS, les rasters devraient être créés avec des propriétés identiques, comme indiqué dans l'Analyse 3.1.5.

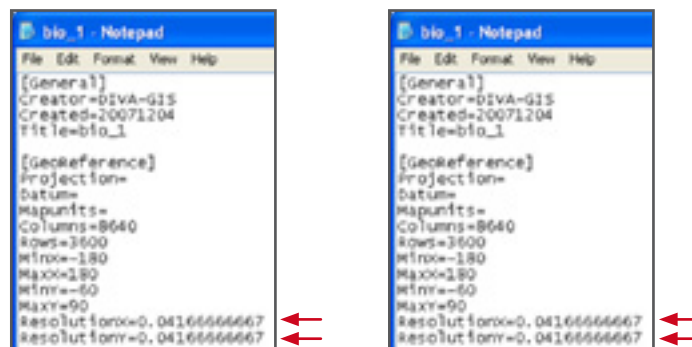
Néanmoins, après avoir utilisé ces options pour créer des rasters aux propriétés identiques, il peut quand même subsister de légères différences dans les décimales de la résolution et/ou dans les coordonnées. Ceci peut arriver si des rasters aux propriétés identiques sont créés à partir de jeux de données différents importés dans DIVA-GIS (comme des jeux de données du climat actuel et de climats futurs, ou des jeux de données sur le sol et sur le climat).

[...suite page suiv.]

C'est pourquoi, au cas où une erreur se produise dans les calculs en combinant des rasters dans DIVA-GIS ou dans la modélisation de la distribution d'espèces dans Maxent, il est recommandé de vérifier les différences en résolution et/ou entre les coordonnées des coins des rasters. Vous pouvez effectuer cette vérification en ouvrant les fichiers raster (*.grd) à l'aide de *Notepad*. Regardez l'exemple ci-dessous et notez l'information dans le raster (*.grd) pour *Mean Annual Temperature* (BIO1) dans les conditions du climat actuel et du climat futur en Asie du Sud-Est.



Bien que le nombre de colonnes et de lignes et les coordonnées des coins (*MinX*, *MaxX*, *MinY*, *MaxY*) des deux rasters soient les mêmes, il y a cependant une différence dans la onzième décimale de la résolution (*Resolution X* et *Resolution Y*). La différence est extrêmement faible et elle n'entraînera pas de variations dans la visualisation, mais elle génèrera une erreur qui empêchera de comparer les aires de distribution potentielle pour les climats actuel et futur lors de la conversion des rasters au format ASCII, adapté à Maxent. Pour résoudre ce problème, les valeurs différentes peuvent être modifiées manuellement dans les fichiers de documentation des rasters (*.grd) afin de les rendre équivalentes.



Lorsque la onzième décimale a été corrigée, les résolutions sont identiques et les rasters peuvent être combinés. Cette solution est bien adaptée pour corriger toutes sortes de différences mineures qui peuvent se trouver dans les coordonnées des coins (*MinX*, *MaxX*, *MinY*, *MaxY*).

3.2. Exporter des couches vers Google

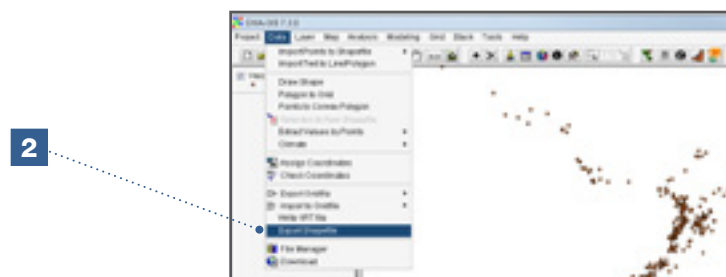
Google Earth utilise des fichiers aux formats particuliers nommés *.kml et *.kmz qui sont une exclusivité de ce programme. De ce fait, les couches en provenance de Google Earth doivent être converties en fichiers sous ces formats pour que les données puissent être visualisées avec Google Earth. La version 7.1 de DIVA-GIS inclut une option d'exportation de couches vecteur et raster vers Google Earth.

PROGRAMMES ET FICHIERS DE DONNÉES À UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Google Earth 	Fichiers de données : <p>Répertoire 3.2 Export to Google Earth (Exporter vers Google Earth)</p> <ul style="list-style-type: none"> • <i>Vasconcellea species</i> (shp, shx, dbf) • <i>Distribution Pinus kesiya</i> (grd, gri)

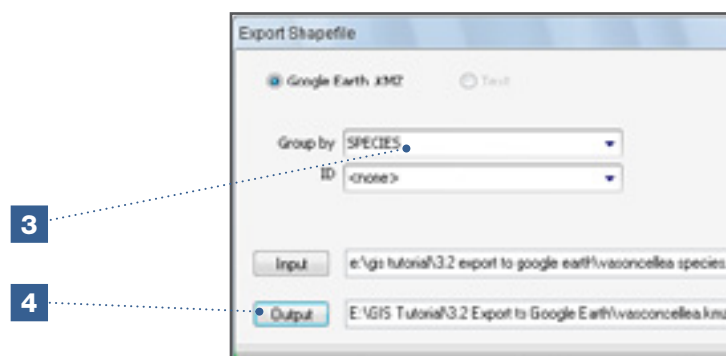
3.2.1. Comment exporter des données vers Google Earth

Etapes :

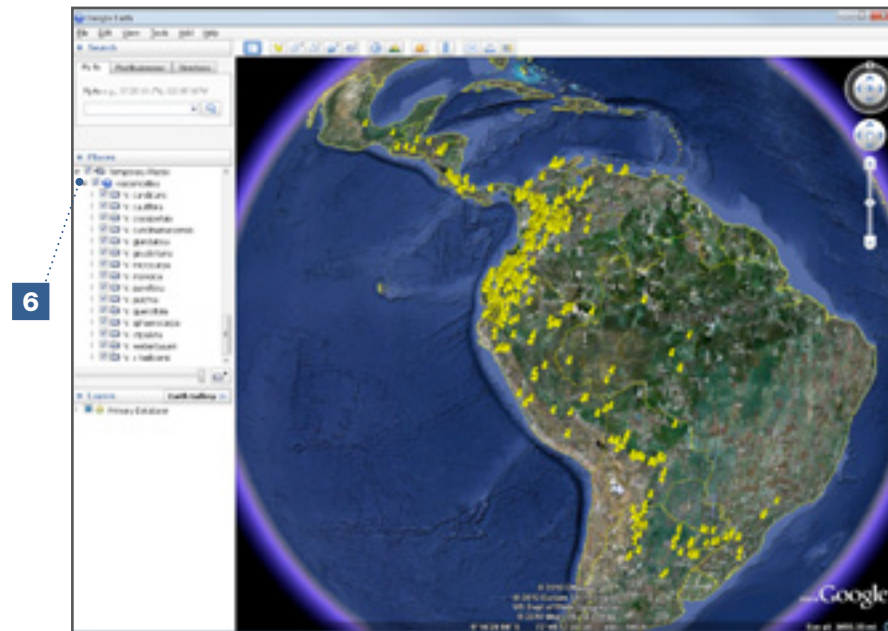
1. Pour exporter un fichier vecteur (*.shp) vers Google Earth, il faut d'abord l'ouvrir avec DIVA-GIS. Pour cette analyse, ouvrez le fichier *Vasconcellea species.shp*.
2. Sélectionnez *Data/Export Shapefile*.



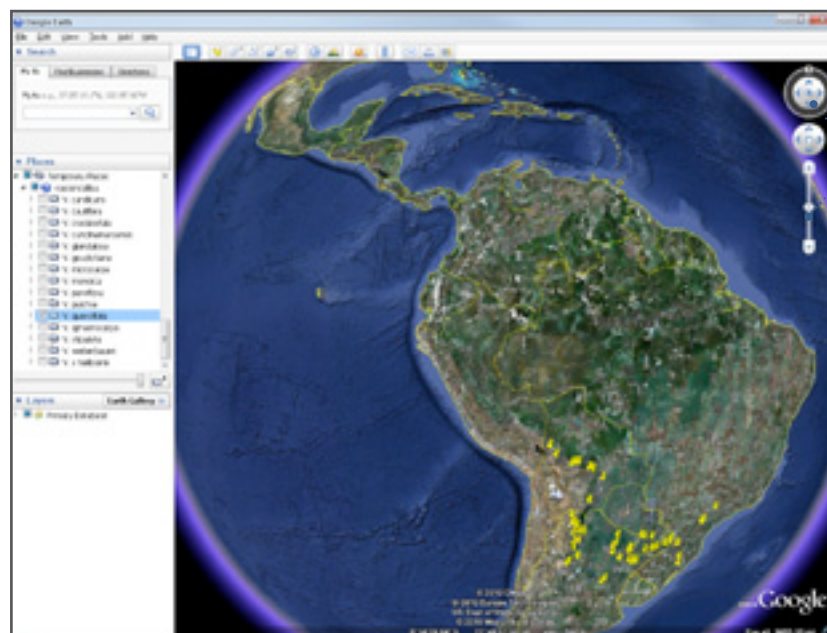
3. Dans la fenêtre *Export Shapefile*, avec l'option *Group by*, sélectionnez *Species*. Ceci permettra de visualiser une espèce particulière dans Google Earth. L'option *ID* vous permet d'ajouter du texte à chaque point. Puisque le fichier des espèces de *Vasconcellea* contient de nombreux points, il est recommandé pour cette analyse de sélectionner l'option : *<none>*.
4. Le bouton *Output* vous permet de nommer le fichier *.kmz nouvellement créé.



5. Allez dans le répertoire où le fichier *.kmz a été sauvegardé et ouvrez-le.
6. Dans la fenêtre *Places*, vous trouverez le répertoire *Temporary Places*, qui contient le fichier nouvellement créé. Cliquez sur le signe triangulaire et remarquez que le fichier *.kmz contient des sous-groupes. Sélectionnez uniquement le sous-groupe de l'espèce *V. quercifolia*.



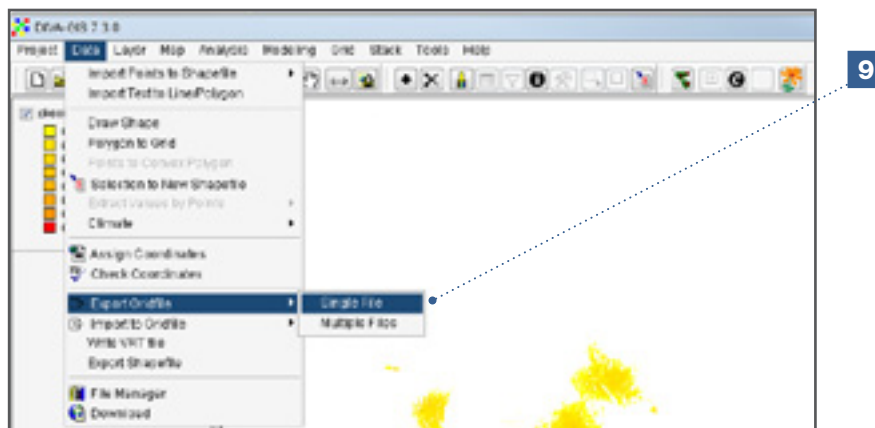
7. A l'aide des outils *Pan* et *Zoom* dans Google Earth, localisez la zone au Nord de l'Argentine où l'on trouve beaucoup de *Vasconcellea quercifolia* (autour de la ville de Salta). Remarquez que certains points se trouvent dans le centre de la ville, ce qui montre le danger d'utiliser une trop haute résolution. Il est vraisemblable que l'on n'a pas pu géoréférencer ces points avec la même précision ou résolution que celle du fond fourni par les images satellitaires de Google Earth.



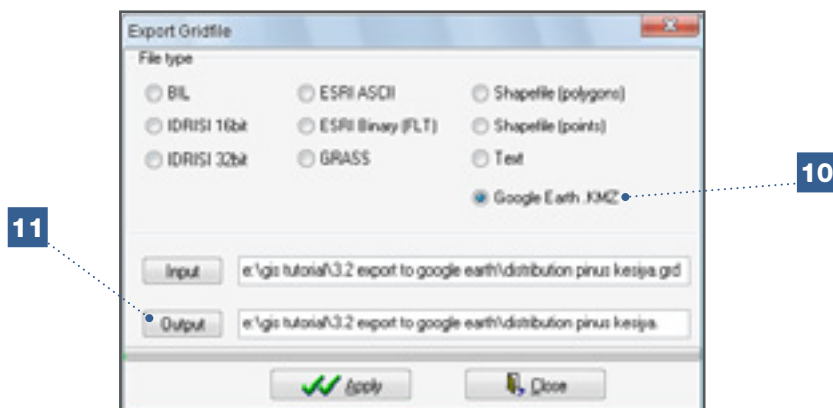


L'activité suivante consiste à visualiser un raster, comme ceux qui résultent d'une analyse de biodiversité (voir le Chapitre 5) ou d'une modélisation de distribution d'espèces (voir le Chapitre 6). Dans cette analyse, la modélisation de la distribution de *Pinus kesiya* (un pin d'Asie), est utilisée à titre d'exemple.

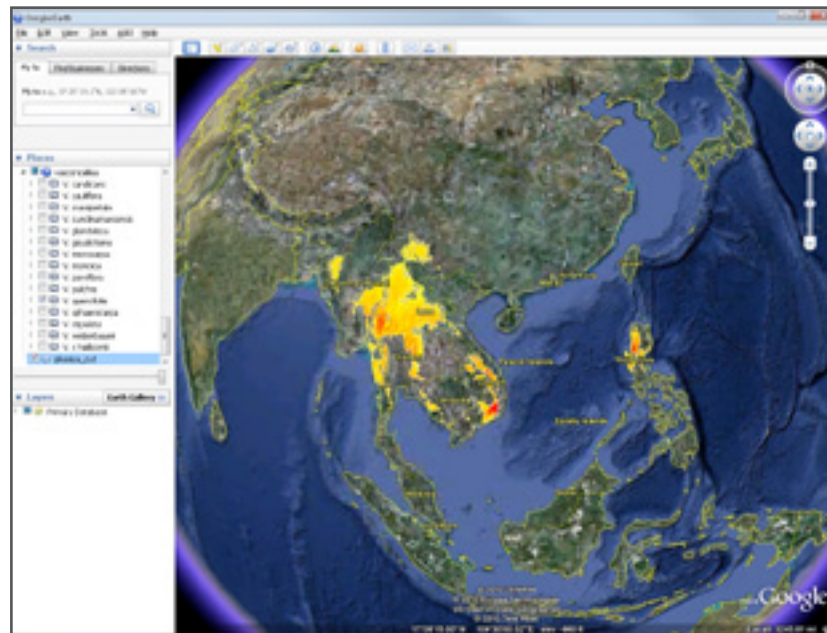
8. Dans DIVA-GIS, ouvrez le fichier : *Distribution Pinus kesiya.grd*.
9. Dans le menu *Data*, allez sur *Data/Export Gridfile/Single File*.



10. La fenêtre *Export Gridfile* s'affichera automatiquement ; sélectionnez *Google Earth.KMZ*.
11. Utilisez le bouton *Output* pour définir l'emplacement et le nom du fichier *.kmz.



12. Enfin, allez dans le répertoire où a été sauvegardé le fichier *.kmz et ouvrez-le.





3.3. Edition de cartes et finalisation d'un projet

Si vous souhaitez inclure des cartes dans des documents ou des rapports, vous pouvez personnaliser la combinaison des couches générées, avec l'option *Design*, accessible dans le coin en bas à droite de l'écran. Cette option vous permet de réaliser des éditions simples de la visualisation courante (comme ajouter l'échelle, une flèche indiquant le Nord ou ajouter du texte).



Après avoir personnalisé la carte, vous pouvez sauvegarder l'image dans un fichier *.png, *.tif ou *.bmp, ou bien la copier directement dans le presse-papier. Le menu *Map*, dans l'onglet *Data*, comprend l'option *Map to Image*, qui vous permet de sauvegarder la carte éditée dans un fichier *.png, *.tif ou *.bmp, mais pas de poursuivre l'édition.

Comment sauvegarder les projets dans DIVA-GIS

Les fichiers vecteur et raster sont sauvegardés automatiquement quand ils sont générés ou modifiés dans DIVA-GIS. Une combinaison particulière de fichiers peut être sauvegardée comme un projet, défini comme une combinaison de différents fichiers raster et vecteur. Si vous souhaitez reprendre votre travail, il vous suffit alors d'ouvrir de nouveau le projet : ceci ouvrira automatiquement toutes les couches qui constituent la carte, afin que vous n'ayez pas besoin de la recréer. Utilisez l'icône  et sélectionnez *Project/Save (as)* pour sauvegarder un projet. Pour ouvrir un projet qui a été sauvegardé antérieurement, utilisez l'option *Project/Open* ou cliquez sur l'icône .

Un projet est sauvegardé sous forme d'un fichier *.div. Ce fichier ne contient que l'emplacement sur le disque dur (répertoire) de chaque couche du projet. C'est pourquoi il ne peut être ouvert que sur le même ordinateur que celui sur lequel il a été créé à l'origine (pourvu qu'il n'y ait pas eu de modification dans l'emplacement des couches). Pour partager ou échanger des projets, allez sur l'option *Project/Export Project*. Avant d'exporter, convertissez le fichier *.div en fichier interchangeable *.dix. Souvenez-vous qu'un fichier *.dix peut être très volumineux, selon son contenu (résolution et nombre de couches), ce qui le rend difficile à échanger. Les fichiers vecteur et raster individuels peuvent aussi être partagés, mais assurez-vous de partager tous les fichiers qui constituent une couche donnée :

- Fichiers vecteur : *.shp, *.shx et *.dbf
- Fichiers raster : *.grd et *.gri (les fichiers image *.bpw et *.bmp sont générés automatiquement lorsque vous visualisez un fichier raster, mais ils ne constituent pas une partie essentielle du fichier).

Références

DIVA-GIS. 2005. User Manual, version 5.2 [en ligne]. Disponible à l'adresse : http://www.diva-gis.org/docs/DIVA-GIS5_manual.pdf.

Ramirez J, Bueno-Cabrera A. 2009. Working with climate data and niche modelling. Creation of bioclimatic variables. Tutoriel [en ligne]. Disponible à l'adresse : http://gisweb.ciat.cgiar.org/GCMPPage/docs/tutorial_bcvars_creation.pdf.

Chapitre 4

Contrôle qualité

Lorsqu'on entreprend une analyse spatiale de données sur la biodiversité, l'un des principaux objectifs est de fournir une information qui permette de conduire des politiques efficaces et de prendre les bonnes décisions en matière de préservation et d'exploitation des ressources naturelles. La formulation de stratégies appropriées de gestion et de conservation nécessite des jeux de données de grande qualité et d'une précision suffisante (Chapman, 2005a, 2005b). L'utilisation d'informations incorrecte ou de mauvaise qualité peut avoir des conséquences graves sur la pertinence et le caractère approprié des recommandations, des décisions et même des investissements qui en résultent. Les données des banques de gènes et des herbiers (disponibles au travers de réseaux de biodiversité comme le GBIF : voir la Section 2.3) sont de plus en plus utilisées dans les études biogéographiques ; mais de telles données proviennent de tierces parties et leur origine est souvent inconnue, ce qui rend le problème de la qualité des données d'autant plus pertinent.

L'objectif spécifique de ce chapitre est de montrer aux utilisateurs comment identifier d'éventuels points de présence erronés en utilisant différents outils, et comment prendre les mesures correctrices afin d'assurer un haut niveau de qualité des données. Une mauvaise qualité des données peut résulter de causes diverses, comme des erreurs dans la description des sites, des coordonnées imprécises, ou même des erreurs ou des changements dans l'identification taxonomique. L'enregistrement des coordonnées sur le terrain est souvent entachée d'erreurs, surtout lorsqu'il faut retranscrire les données ou quand l'information est saisie dans une base de données. L'enregistrement d'un géoréférencement fait à distance ou à partir d'un bureau sur la base de la description du site de collecte, peut aussi être une source de mauvaise qualité des données.

Deux aspects clés de la qualité des données sont l'exactitude et la précision des coordonnées géographiques. L'exactitude des coordonnées détermine la capacité de représenter correctement le site de collecte ou d'observation d'où est issu le point de présence. La précision se réfère au niveau de détail des coordonnées nécessaire pour représenter efficacement le site décrit. On peut évaluer la précision en examinant la méthode utilisée pour déterminer les coordonnées (par exemple, à l'aide d'une carte, ou avec un GPS) ou à partir du nombre de décimales incluses dans les coordonnées, comme discuté auparavant dans la Section 2.1. Un défaut d'exactitude dans les données analysées conduira inévitablement à des erreurs dans les résultats de l'analyse, tandis le manque de précision induira souvent des conclusions d'une utilité réduite, du fait de la basse résolution des données. Les données peuvent être très précises, mais inexactes, et elles peuvent aussi être exactes mais très imprécises.

Le programme DIVA-GIS offre la possibilité d'afficher les problèmes de qualité, ce qui peut vous aider à prendre des décisions concernant la correction ou l'élimination de points de présence de mauvaise qualité. Le logiciel propose des outils utiles pour identifier les erreurs éventuelles dans les coordonnées à partir de l'information sur l'unité administrative disponible dans les données passeport (voir la Section 4.1) ou pour identifier les points de présence douteux en se basant sur des conditions climatiques atypiques, qui peuvent traduire une mauvaise identification taxonomique ou des coordonnées erronées (voir la Section 4.2).

Un autre aspect important de la qualité des données, qu'il est difficile d'évaluer, est le préjugé que l'on peut avoir quand un échantillonnage n'est pas totalement représentatif de la zone étudiée et qu'on peut y remédier avec une bonne stratégie de collecte de données. Néanmoins, de nombreuses analyses spatiales de biodiversité sont effectuées avec des données prov-

enant partiellement ou entièrement de banques de gènes ou d'herbier ; or de telles données ne sont habituellement pas générées dans le but d'effectuer des études biogéographiques, et elles résultent souvent de collectes *ad hoc*, d'échantillonnages non systématiques et d'efforts d'échantillonnage inégaux (Chapman, 2005a). Il est fréquent que les spécimens aient été collectés dans des zones faciles d'accès ou des zones où l'on sait que se trouve une espèce, et cela réduit la représentativité des données (Hijmans *et al.*, 2000). De ce fait, la population échantillonnée peut être ou ne pas être représentative de l'espèce sur les plans environnemental ou géographique, puisque les données fournissent une information sur des structures de répartition qui correspondent seulement aux sites collectés, plutôt qu'à l'ensemble de la zone étudiée (Williams *et al.*, 2002). Il y a plusieurs méthodes pour réduire le préjugé statistique, mais celles-ci ne résolvent que partiellement le problème. Les Chapitres 5 et 6 donnent plus de détail sur ces méthodes (comme le voisinage circulaire, la raréfaction et la modélisation de la distribution d'espèces).

4.1. Contrôle qualité fondé sur l'information relative aux unités administratives

Une manière d'évaluer l'exactitude d'un point de présence consiste à comparer les données sur l'unité administrative contenues dans ses données de passeport avec l'information sur l'unité administrative extraite des couches thématiques en se référant aux coordonnées géographiques du point. Pour cela, les données de passeport doivent inclure des données sur le pays et, de préférence, sur des unités administratives de niveau inférieur. Le menu *Data* de DIVA-GIS fournit l'option *Check Coordinates*, qui permet de vérifier la qualité des coordonnées à partir des informations sur les unités administratives.


PROGRAMMES ET FICHIERS DE DONNÉES À UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel 	Fichiers de données : <p>Répertoire 4.1 <i>Quality control – Administrative units</i> (Contrôle qualité – Unités administratives)</p> <ul style="list-style-type: none"> • Vasconcellea final errors (shp, shx, dbf) • Latin America countries (shp, shx, dbf) • Latin America Adm 01 (shp, shx, dbf)¹

4.1.1. Comment vérifier la qualité des données en se fondant sur les données d'unité administrative du passeport

Dans l'exercice suivant, vous utiliserez une portion du jeu de données qui contient de l'information sur la distribution géographiques des papayes d'altitude collectée au cours d'une étude sur la papaye d'altitude conduite en Amérique latine (Scheldeman *et al.*, 2007). Certaines erreurs ont été introduites dans le jeu de données pour les besoins de cette analyse. La version propre de ce jeu de données sera utilisée dans l'analyse de la diversité interspécifique de la Section 5.1.

¹ Notez que la couche *Latin America Adm 01* contient certaines imprécisions au niveau administratif 1 qui, si elles n'ont pas d'impact sur les analyses dans cet exercice, pourraient empêcher l'usage de ce fichier dans d'autres analyses. Pour obtenir des cartes au niveau Adm 01 à jour, vous pouvez consulter le site Web de DIVA-GIS (<http://www.diva-gis.org/Data>) ou la base de données mondiales sur les zones administratives (GADM en anglais, <http://www.gadm.org/>).

Etapes de la mise en œuvre de la fonction *Check Coordinates* du menu *Data* de DIVA-GIS :

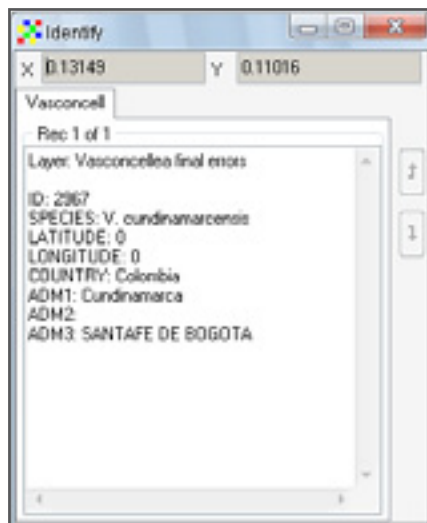
1. La première manière d'identifier des points potentiellement erronés est de visualiser le jeu de données sur une carte. Ouvrez le fichier *Vasconcellea final errors.shp* pour visualiser les points de collecte de *Vasconcellea* en Amérique latine. Pour voir tous les points sur la carte, utilisez le bouton *Zoom to Theme*, qui affiche toute l'étendue des jeux de données ouverts.
2. Maintenant, ajoutez le fichier polygone avec les pays d'Amérique latine (*Latin America countries.shp*). Des erreurs de données évidentes apparaissent immédiatement puisque certains points se trouvent en dehors de la zone étudiée (l'Amérique latine). Pour accéder aux données de passeport de chaque point et déterminer l'identification (*ID*), utilisez le bouton *Information*  dans le menu principal.



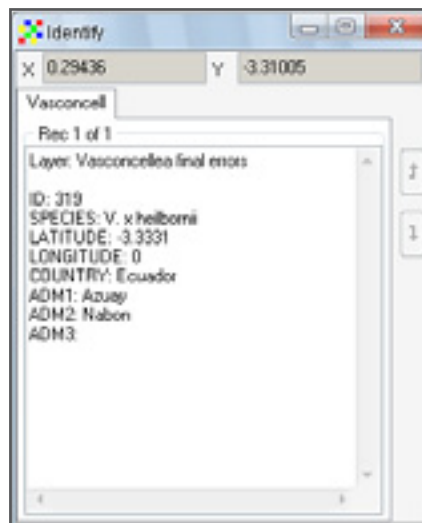
Conformément à l'objectif de cette analyse, les données présentées illustrent trois erreurs typiques, toutes évidentes, du fait que les points se trouvent en dehors de la zone d'étude :

- **Cas A : ID 2967.** Les coordonnées du point valent 0,0 (ce qui place ce point devant les côtes d'Afrique). Cette erreur se produit dans la base de données si une donnée est absente, du fait que certains programmes remplissent automatiquement les cellules vides avec des valeurs à zéro (0).
- **Cas B : ID 319.** La situation est similaire à la précédente, mais dans ce cas-ci, l'information n'est incorrecte que pour une des coordonnées (la longitude). Dans une base de données, il se peut qu'une des coordonnées géographiques d'un point de présence manque. On peut se poser la question : Que se passerait-il si la valeur de latitude était zéro (0) au lieu de celle de la longitude ? Est-ce que l'erreur serait aussi facile à identifier ?
- **Cas C : ID 1669.** Ce point est éloigné de la zone d'étude. Ici, l'erreur vient vraisemblablement de l'omission du signe moins affecté aux points situés dans les latitudes Sud ou les longitudes Ouest, ce qui se traduit par une position incorrecte sur la carte.

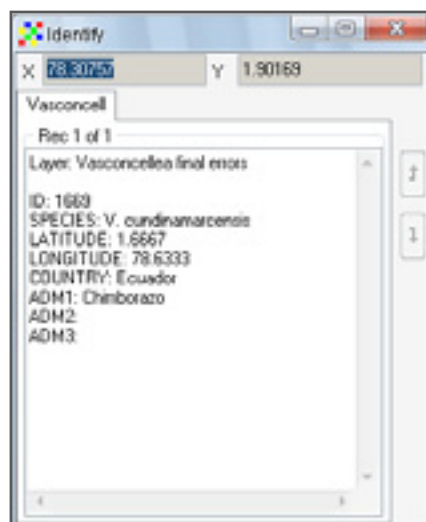
Cas A



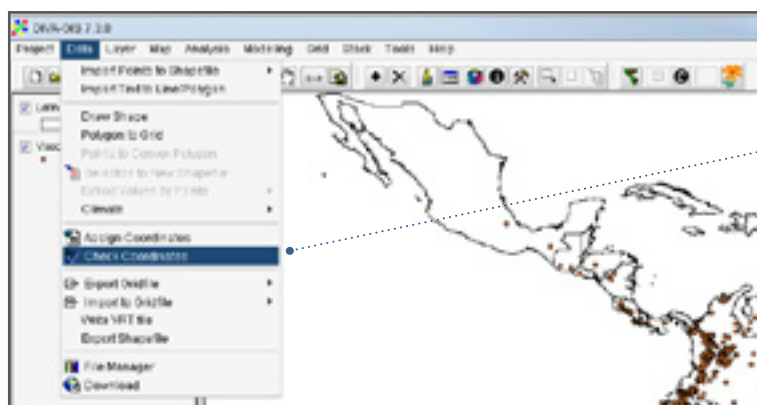
Cas B



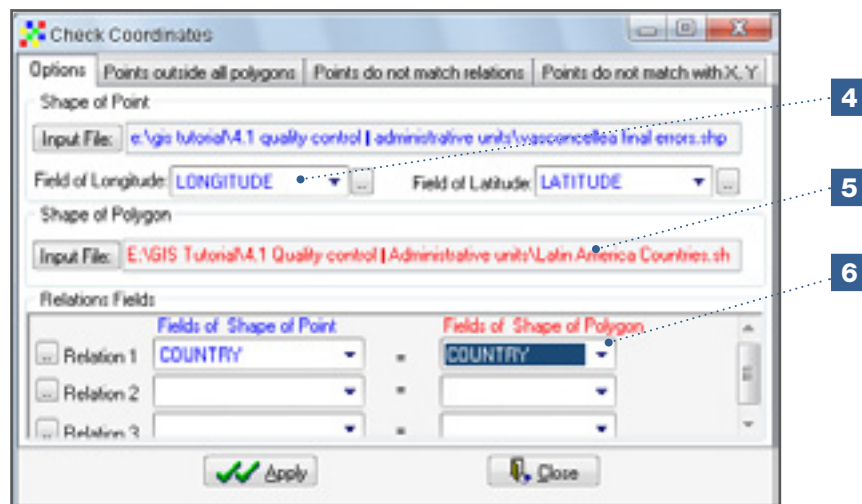
Cas C



3. Si ces erreurs sont aisément détectées par simple visualisation de la carte, la base de données dans son entier doit aussi être vérifiée à la recherche d'erreurs moins évidentes. Pour ce faire, allez sur le menu *Data* et sélectionnez *Check Coordinates*. Si la couche contenant les points a été sélectionnée, celle-ci deviendra l'*Input* ; sinon cliquez sur *Input File* et sélectionnez la couche qui contient les observations.



4. Sélectionnez le fichier contenant les données de longitude et de latitude.
5. Sélectionnez le fichier polygone qui contient les données d'unité administrative (*Latin America countries.shp*) en commençant par les données de niveau supérieur (la couche avec les pays).
6. Indiquez la relation entre les deux fichiers (*Country-Country*) et cliquez sur *Apply*. DIVA-GIS recherchera immédiatement les incohérences éventuelles dans l'information relative aux deux couches.



7. Le premier groupe d'incohérences concerne les points qui se trouvent en dehors de tous les polygones (onglet *Points outside all polygons*). Dans cette analyse, ces points sont situés en dehors du polygone de l'Amérique latine. Cette option est très utile pour identifier les points qui ont des localisations douteuses, comme le point d'une espèce de plante qui serait situé dans l'océan.

No Reg	X	Y	ID	SPECIES	LATITUDE	LONGITUDE	COL
6	0	-3.3331	319	V. x helbonii	-3.3331	0	Ecu
321	-00.8419	-1.5667	796	V. parviflora	-1.5667	-00.8419	Ecu
561	-00.6033	0	1367	V. parviflora	0	-00.6033	Ecu
622	78.6333	1.6667	1669	V. cundinam	1.6667	78.6333	Ecu
830	0	0	2967	V. cundinam	0	0	Colo

Dans la quatrième colonne de l'information affichée, vérifiez l'identification (*ID*) des données erronées. Avec cette information, vous pouvez revenir au fichier de données original [dans Excel ou dans un fichier dBase IV (*.dbf)] et vérifiez les détails afin de déterminer comment corriger l'erreur. Les options *Pan to* et *Zoom to* (voir le Chapitre 3) permettent aussi d'identifier d'autres points qui pourraient contenir des erreurs. Ces points problématiques peuvent aussi être exportés dans un fichier texte en utilisant l'option *Export*.

Note

Il est fortement recommandé de documenter en détail toutes les modifications apportées à la base de données, y compris les rapports d'erreur, et de conserver la base de données originale intacte, en créant de nouvelles versions corrigées pour poursuivre le traitement.

La liste générée par DIVA-GIS à l'étape 7 rapporte cinq points potentiellement erronés, dont trois ont déjà été détectés (ID : 2967, 319 et 1669) sur la carte. Les autres points suspects sont :

- **ID 1367**, qui est situé à 500 km de la côte de l'Equateur. Le niveau de Zoom utilisé dans l'analyse n'a pas permis de détecter ce point à première vue. Ce point a aussi une latitude égale à zéro (0), ce qui explique probablement cette erreur.
- **ID 796**, qui est situé dans l'océan, à environ 1 km de la côte équatorienne. Sachant que les données de passeport de ce point indiquent un site de collection très proche de la côte, l'erreur est plus vraisemblablement due à une légère imprécision du GPS ou de l'opération de géoréférencement manuel du site. Ceci peut être corrigé en assignant une coordonnée située dans la zone d'étude.

Que faire avec les points erronés ?

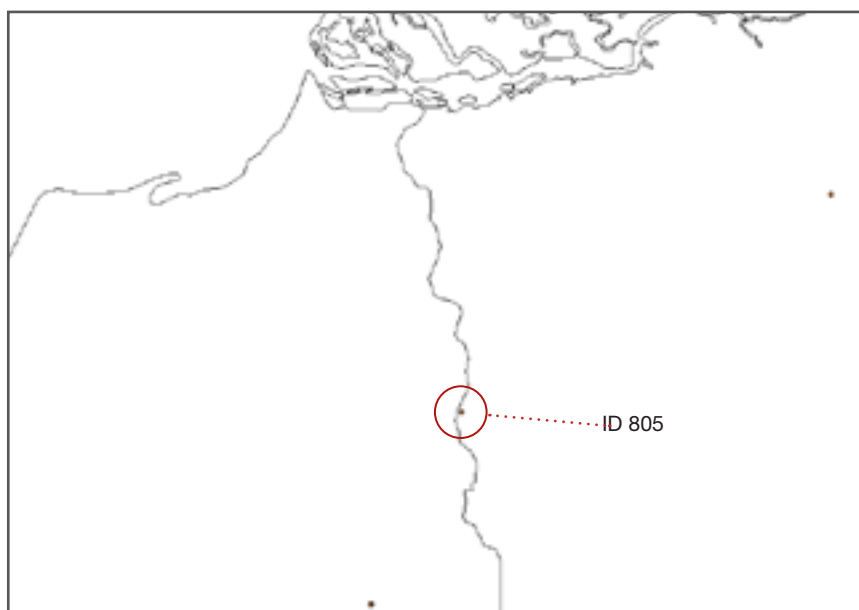
Il y a trois options pour traiter les erreurs : les corriger, les détruire ou les conserver. La décision la plus appropriée dépend de plusieurs facteurs. Deux facteurs importants sont l'échelle de l'étude (mondiale, continentale, nationale ou locale) et le nombre de points disponibles pour le taxon étudié. Dans l'un ou l'autre cas, la première étape consiste à localiser l'information originale et à vérifier l'origine du point (d'où l'importance d'entretenir les carnets de terrain). Cette étape peut vous permettre de corriger l'erreur immédiatement. Si l'information originale n'est pas disponible (ce qui est souvent le cas), la prochaine étape consiste à identifier la cause possible de l'erreur (voir l'exemple ci-dessus pour le point ID 1669, où le seul problème est l'omission du signe moins). Si ces deux étapes ne fonctionnent pas, vous devriez envisager d'éliminer ce point mais avant de le faire, réfléchissez aux conséquences d'une telle élimination sur les résultats (assurez-vous que vous ne biaisez pas les résultats). Vous préfererez peut-être conserver les données erronées si vous travaillez à très grande échelle et si le point problématique est considéré comme important (par exemple, dans une étude visant à connaître le nombre d'espèces par pays, la localisation exacte au sein d'un pays est moins importante).

- Maintenant, regardez les autres erreurs. Dans le menu *Check Coordinates*, allez sur l'onglet *Points do not match relations*. Le résultat de la comparaison des données de passeport avec leur localisation effective sur la carte est affiché, il révèle les points dont l'information sur le pays dans le passeport ne coïncide pas avec la localisation du pays dans le point sur la carte.

No Reg	X	Y	Point COUN	Poly COUN	ID	SPECIES	LAT
326	-80.1915	-3.6033	Peru	ECUADOR	805	V. pinnifrons	-3.60
483	-72.7333	8.3	Brazil	COLOMBIA	1143	V. microcap	8.3
695	-65.7333	-2.6166	Brazil	BRASIL	2724	V. microcap	-2.6

Cette analyse vous permet d'identifier d'autres erreurs, au delà de celles détectées après la première analyse décrite à l'étape 7 :

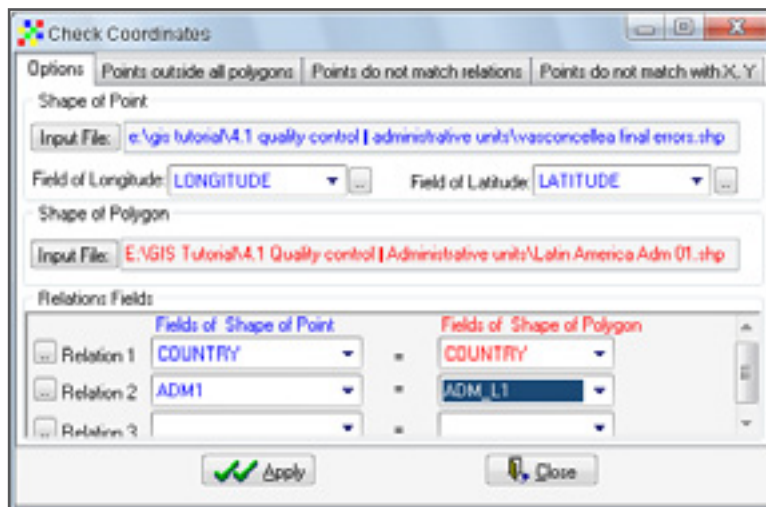
- **ID 805** : Selon ses données de passeport, ce spécimen a été collecté au Pérou, mais selon la carte, le point est situé en Equateur. Comme l'imprécision dans les coordonnées est inférieure à 1 km, cela semble provenir d'une erreur lors du géoréférencement ou de l'enregistrement des coordonnées. Le collecteur peut avoir traversé la frontière sans s'en rendre compte, ce qui pourrait être l'origine de l'incohérence dans les données administratives. C'est un problème courant et la décision est difficile, du fait que l'erreur pourrait être dans la partie descriptive du passeport alors que les coordonnées sont correctes. De nombreuses études autorisent une marge d'erreur autour des frontières des unités administratives, par exemple 5 km.



- **ID 1143** : Le point est situé en Colombie, alors que son information de passeport indique qu'il a été collecté au Brésil. Considérant qu'il y a un large espace entre le site et la frontière brésilienne, il s'agit clairement d'une erreur qui doit être résolue (si la cause du problème peut être identifiée). Si c'est impossible, le point doit lors être éliminé. Dans cet exemple, le problème vient d'un signe négatif (-) manquant dans une des coordonnées.
- **ID 2724** : Le point illustre aussi un problème courant, causé par des erreurs ou des imperfections dans l'orthographe des unités administratives (par exemple « Brazil » au lieu de « Brasil »). Ce point n'est manifestement pas erroné, et la différence d'orthographe peut être facilement corrigée, bien qu'il ne soit pas toujours nécessaire de le faire.

Contrôle des données pour les unités administratives de niveau inférieur

Lorsque des données de passeport contiennent une information sur un niveau administratif inférieur, comme les départements ou les provinces, il est recommandé de lancer une vérification des erreurs à ce niveau également. On procède initialement comme expliqué aux étapes 1 à 5 ci-dessus. Mais à l'étape 6, où les relations sont indiquées, il faut ajouter la couche contenant les données d'unité administrative de niveau inférieur (*Latin America Adm 01.shp*). Cela fait, poursuivez avec les étapes 7 et 8 en ajoutant simplement un niveau de plus dans le champ relation à l'aide des outils *Fields of Shape of Point* et *Fields of Shape of Polygon*.



On voit apparaître de nouveaux points de présence aux coordonnées douteuses en vérifiant ce deuxième niveau de données. Ces nouveaux points s'ajoutent aux points précédemment décrits (ID 805, ID 1143, ID 2724), qui étaient erronés au niveau pays, et aux points (ID 319, ID 796, ID 1367, ID 1669 et ID 2967), situés en dehors de la zone d'étude. Il s'agit de :

- **ID 689** : Ce point est situé sur la frontière entre deux provinces de l'Equateur (Morona Santiago et Chimborazo). Comme dans le cas d'ID 805, le problème peut venir d'une coordonnée imprécise mais l'explication la plus probable est que le collecteur a traversé la frontière provinciale sans le savoir, ce qui l'a conduit à donner une description erronée. Lorsque la distance entre le point et la frontière est faible (comme c'est le cas ici), il n'y a généralement pas besoin de faire de correction.
- **ID 2729** : La latitude de ce point est égale à zéro (0) – une erreur qui a été décrite précédemment (voir le point d'ID 319). Ce point illustre clairement le besoin de contrôler les niveaux administratifs inférieurs, puisque ce type d'erreur ne place pas toujours le point dans l'océan ou dans un autre site improbable.
- **ID 2870** : Le site de collecte a des noms différents au niveau départemental de la Colombie. Lorsque ce point a été enregistré, le nom du département du site était « Cundinamarca », mais plusieurs zones de ce département (y compris la zone de collecte) ont ensuite été rattachées au « Distrito Capital ». Le changement du nom d'une unité administrative est souvent source d'erreurs lorsque l'on utilise des données historiques des herbiers ou des muséums.
- **ID 2906** : Ce point contient une erreur d'orthographe – une erreur courante lorsque l'on utilise des caractères accentués comme le « ñ » qui n'existent pas dans toutes les langues.
- **ID 2943** : Un autre exemple de valeur nulle (0) affectée à la latitude.
- **ID 3023** : Un problème dans le nom du département : « Valle » (forme abrégée courante) pour « Valle del Cauca » (le nom complet).
- **ID 3068** : Ce point présente le même problème que celui décrit pour le point d'ID 2729 ci-dessus.

4.2. Contrôle qualité par l'identification de points atypiques

Des points atypiques, ou aberrants, sont des points de présence situés en dehors des limites de la distribution environnementale normale d'une espèce. Les données atypiques occupent une place ambivalente dans les analyses spatiales de biodiversité. Les méthodes d'identification de points atypiques peuvent aider à détecter des points de présence erronés dans un jeu de données, qui devront être éliminés pour assurer la qualité des données. Un environnement atypique peut signaler des données de présence incorrectes résultant de différents types d'erreurs : des coordonnées géographiques ou une détermination taxonomique erronées, ou l'introduction d'individus dans des endroits qui ne correspondent pas à leur aire naturelle (comme des systèmes de production ou des jardins botaniques).

D'un autre côté, un point atypique peut aussi signaler un individu ou un groupe d'individus qui se sont adaptés à des facteurs environnementaux différents de ceux qui règnent là où vivent le plus naturellement individus et populations de l'espèce en question. L'individu ou le groupe d'individus que l'on trouve en un point atypique a ou ont pu acquérir des caractéristiques uniques et fournir un matériel génétique intéressant pour des programmes d'amélioration qui recherchent une résistance adaptée à des stress environnementaux comme la sécheresse ou des températures extrêmes. Des points atypiques peuvent aussi dénoter une information biaisée dans l'échantillonnage, comme peut en produire le sous-échantillonnage de l'aire géographique ou de l'écosystème où se trouvent ces points atypiques.

Bien que les diverses explications des points atypiques mentionnées ci-dessus (erreur, individus poussant dans un climat extrême, information biaisée dans l'échantillonnage) soient toutes valables dans l'analyse spatiale de la biodiversité, il est souvent difficile de déterminer exactement pourquoi ces points sont atypiques. Il est, de ce fait, malaisé de décider ce que l'on doit en faire, à savoir les éliminer ou les conserver dans le jeu de données. La première étape pour déterminer l'action corrective appropriée consiste à localiser l'information originale et vérifier l'origine des points. Mais si les données ont été obtenues d'une tierce partie (comme le GBIF), cela est souvent impossible et il faut donc bien réfléchir avant de décider d'éliminer ou de conserver des points atypiques.

Parfois, il est évident que les points sont erronés lorsque, par exemple, ils reflètent un climat complètement différent du reste ; ces points seront donc facilement identifiables. Dans certains cas malgré tout, la différence entre les points atypiques (suspects) et le reste des données ne sera pas si évidente et il vous sera plus difficile de déterminer un seuil à partir duquel un point peut être considéré comme atypique. Diverses méthodes statistiques (mono ou multi-variables) sont disponibles pour identifier de tels points (voir Chapman, 2005a). Ci-dessous sont décrites deux méthodes permettant de détecter les points atypiques dans DIVA-GIS :

- *Reverse jackknife* (Chapman, 2005b). Cette méthode est recommandée pour des jeux de données dont les valeurs sont distribuées normalement, comme ceux qui contiennent de nombreuses observations pour chaque taxon.
- *1.5x interquartile range (1.5 IQR)* (DIVA-GIS 2005). Cette méthode est recommandée pour les jeux de données avec un nombre limité d'observations par taxon (par exemple $n < 20$).

Bien que ces méthodes soient pratiques pour détecter des valeurs atypiques, elles ne garantissent pas que toutes les aberrations détectées correspondent effectivement à des erreurs ; certaines valeurs atypiques peuvent correspondre à des points valides. Avant de décider de retirer des points, il faut notamment tenir compte de l'objectif de l'étude. Si l'objectif est de modéliser la distribution d'espèces afin d'identifier des aires de production appropriées, il vaut alors mieux n'utiliser que les enregistrements sûrs et ne pas retenir les points atypiques dans l'analyse. Mais si l'objet de l'étude est d'identifier des écotypes adaptés à des conditions qui sortent de l'ordinaire, les points atypiques sont alors intéressants et il peut être utile de les inclure dans l'étude. Comme toujours, il convient de maintenir un registre de toutes les modifications apportées au fichier ou jeu de données original (Chapman, 2005a).

PROGRAMMES ET FICHIERS DE DONNÉES À UTILISER DANS CETTE SECTION

Programmes :	Fichiers de données :
<ul style="list-style-type: none"> • DIVA-GIS • Excel 	<p>Répertoire 4.2 <i>Quality Control – Atypical points</i> (Contrôle qualité – Points atypiques)</p> <ul style="list-style-type: none"> • Vcundinamarcensis_outliers (shp, shx, dbf) • Vcandicans_outliers (shp, shx, dbf) • Latin America countries (shp, shx, dbf) <p>Pour cette analyse, vous devez avoir importé dans DIVA-GIS les données de climat de WorldClim à 2,5 min (voir la Section 2.2)</p>

4.2.1. Comment identifier des valeurs atypiques sur la base des données environnementales

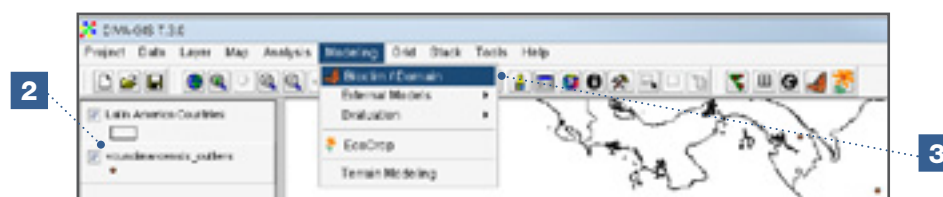
Dans cette analyse vous apprendrez comment identifier des valeurs atypiques avec DIVA-GIS et des données relatives à *V. cundinamarcensis* et *V. candicans* (Scheldeman *et al.*, 2007). Des erreurs ont été volontairement introduites dans la base de données pour les besoins de cet exercice. On peut identifier les erreurs là où le nom d'espèce a été remplacé par le mot « Error » dans le jeu de données. Le jeu de données de *V. cundinamarcensis* est assez grand (144 observations) pour être révisé avec la méthode *Reverse jackknife*, tandis que le jeu de données de *V. candicans* est plus petit (17 points) et devra être révisé avec la méthode *1.5 IQR*.

Etapes :

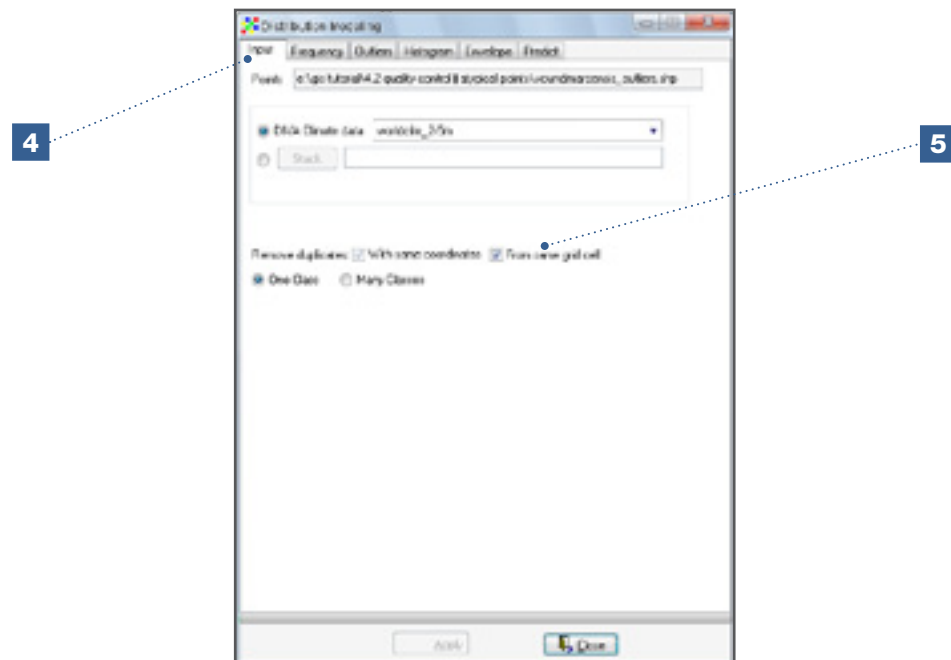
1. Ouvrez les fichiers vecteur (*.shp) *Vcundinamarcensis_outliers.shp* et *Latin America countries* dans DIVA-GIS.



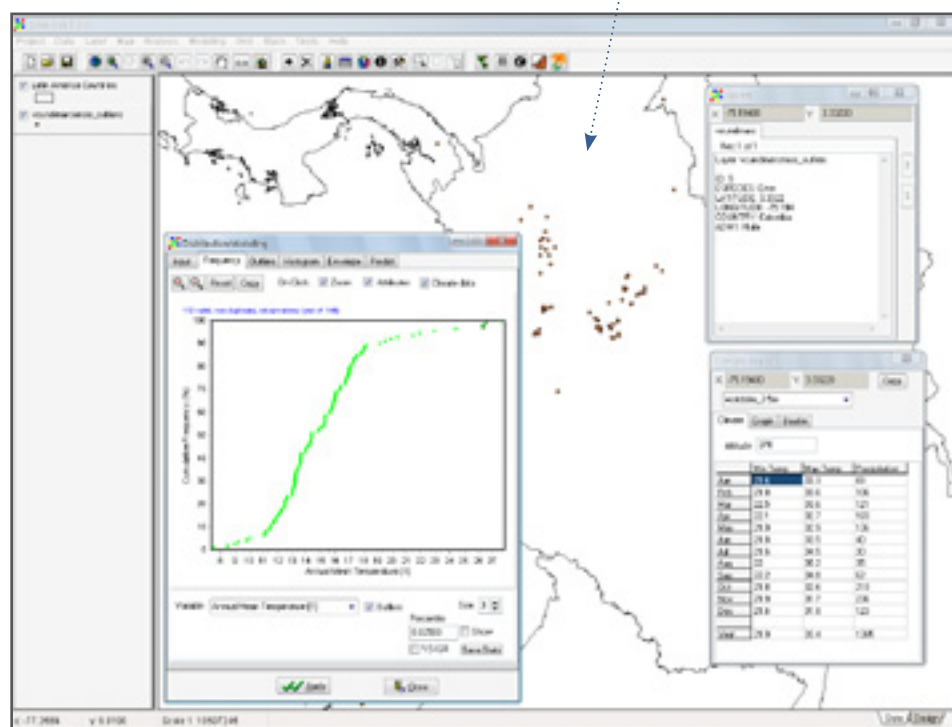
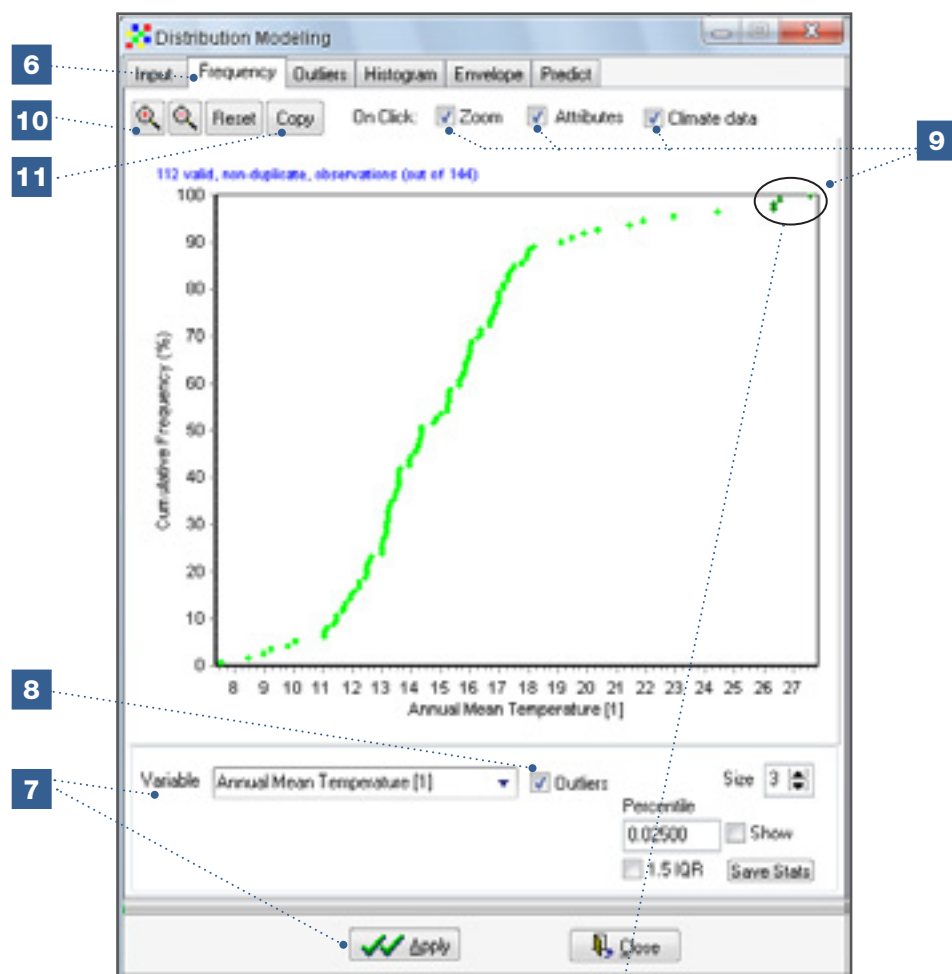
2. Pour détecter les points atypiques avec la méthode *Reverse jackknife*, sélectionnez le fichier des points en cliquant sur le répertoire dans la légende.
3. Dans le menu *Modeling*, allez sur *Bioclim/Domain*.



4. Le fichier *Vcundinamarcensis_outliers.shp* et le fichier de données sur le climat *worldclim_2-5min* devraient apparaître automatiquement dans les fenêtres *Points* et *DIVA Climate data*. Si ce n'est pas le cas, ouvrez-les manuellement.
5. Gardez cochée l'option *From same grid cell* afin d'exclure les doublons dans les cellules des variables climatiques.



6. Allez sur l'onglet *Frequency* pour comprendre la distribution climatique des points de présence de *V. cundinamarcensis*.
7. Sélectionnez la variable climatique qui vous intéresse. Pour cette analyse, la variable sélectionnée est le fichier *Annual Mean Temperature*. Cliquez sur *Apply* pour voir les résultats sur le graphe.
8. A l'aide de l'option *Outliers*, les données relevées comme atypiques par la méthode *Reverse jackknife* seront surlignées en vert foncé.
9. Cochez les cases *Zoom*, *Attributes* et *Climate data* et cliquez sur un point quelconque dans le graphe. Les données de présence et de climat correspondant à ce point seront affichées et le point sera surligné sur la carte pendant quelques secondes.
10. Pour ajuster la taille du graphe, utilisez les boutons *Zoom in* ou *Zoom out*, situés en haut à gauche du menu.
11. L'option *Copy* vous permet de copier et coller le graphe dans tout autre type de document.



Normalement, le classement d'un point comme atypique résulte d'une analyse intégrant une combinaison de variables bioclimatiques. Dans l'onglet *Outliers*, vous pouvez définir les conditions et les variables bioclimatiques importantes pour une espèce donnée. Vous pouvez aussi indiquer la quantité minimale de variables environnementales requises pour considérer qu'un point de présence est atypique (avec les valeurs extrêmes pour chaque variable climatique).

Note

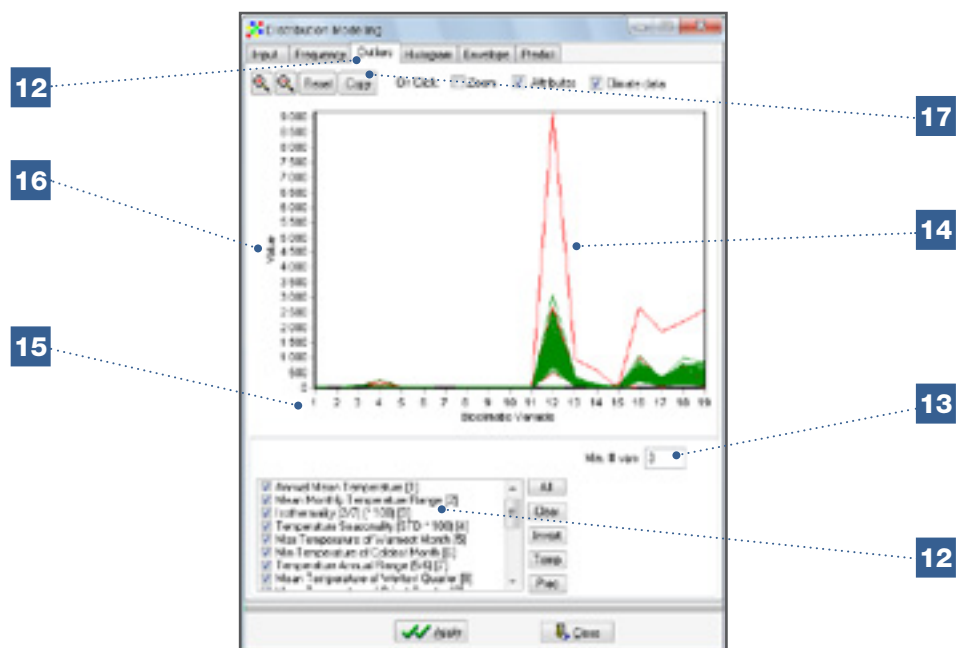
Il arrive qu'il y ait plusieurs points atypiques au même endroit. Vous pouvez le voir dans l'encadré *Attributes* qui vous indiquera s'il y a plus d'un point de présence (par exemple, *Rec 1 of 2*). Dans ce cas, l'encadré *Attributes* vous permettra d'afficher les données de passeport des différents points atypiques situés au même endroit. Pour ce faire, cliquez sur les flèches situées dans la partie droite de la fenêtre où sont présentées les données de passeport pour chaque point.

12. Allez dans l'onglet *Outliers* et sélectionnez les variables climatiques que vous souhaitez intégrer dans l'analyse des points atypiques. Pour cette analyse, toutes les variables ont été sélectionnées. Vous pouvez aussi sélectionner un ensemble de variables climatiques que vous considérez comme des variables clé pour l'espèce étudiée.
13. Sélectionnez le nombre minimum de variables pour lesquelles un point de présence devrait avoir des valeurs inhabituelles pour être considéré comme atypique. Dans cette analyse, trois variables ont été sélectionnées.
14. Les lignes sur le graphique représentent les points de présence. Les lignes rouges représentent des points atypiques selon les conditions spécifiées. Double-cliquez sur l'une des lignes, en vous assurant que les options : (a) *Attributes* et (b) *Climate data* ont été sélectionnées.

L'information suivante devrait s'afficher automatiquement en cliquant sur une ligne :

- a. Données de présence (*Attributes*)
 - b. Données climatiques.
15. L'axe des X du graphe montre les variables bioclimatiques : 1) *Mean Annual Temperature* ; 2) *Mean Monthly Temperature Range*, etc. (voir le tableau des variables bioclimatiques au Chapitre 2).
 16. L'axe des Y du graphe montre les valeurs de chaque variable bioclimatique pour chaque point de présence.

17. La option *Copy* le permet de copier et coller le graphique en n'importe quel type de document.



14a



14b



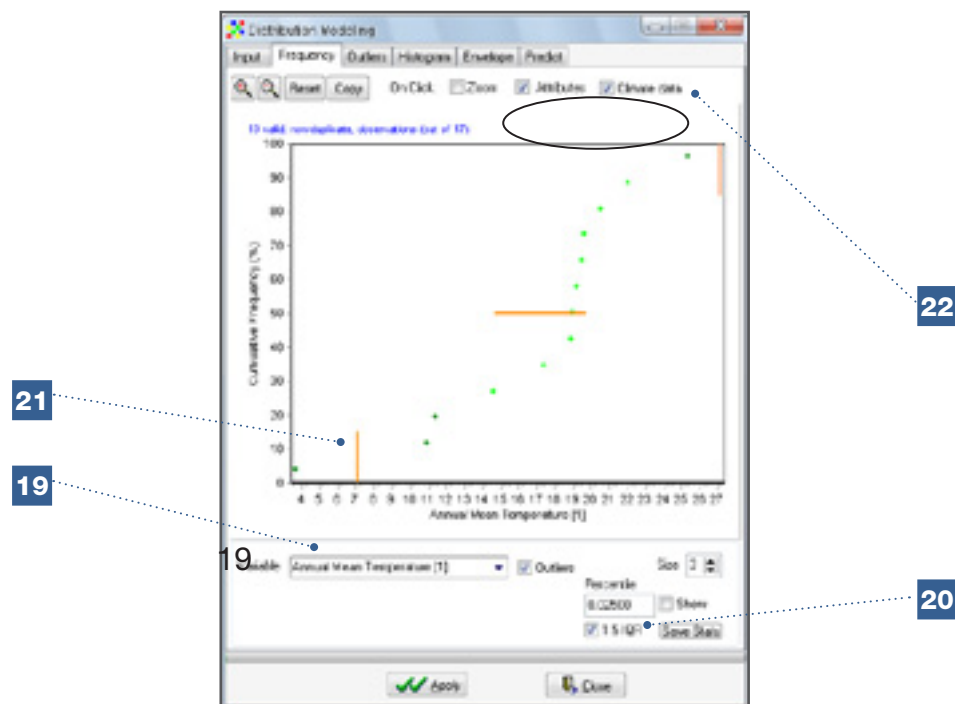
Travail personnel : Trouvez tous les points erronés dans la base de données des points de présence de *Vasconcellea cundinamaricensis* (les points indiqués comme erreurs).

Points de présence de petits jeux de données

Les points de présence dans de petits jeux de données peuvent être testés pour déterminer s'ils sont atypiques à l'aide de la méthode *1.5 IQR* disponible dans l'option *Frequency*.

18. Pour identifier les points atypiques dans le jeu de données de *V. candicans* à l'aide de la méthode *1.5 IQR*, commencez par suivre les étapes 1 à 6 ci-dessus en utilisant le nouveau jeu de données.

19. Définissez les variables climatiques sur la base desquelles vous voulez appliquer la méthode de détection de points atypiques $1.5 IQR$. Dans cette analyse, sélectionnez *Annual Mean Temperature*.
20. Sélectionnez la fenêtre $1.5 IQR$ pour établir les limites de l'intervalle de valeurs.
21. Les points atypiques sont ceux qui tombent en dehors des limites $1.5 IQR$ définies par les deux lignes extérieures.
22. Lorsque les cases *Climate data* et *Attributes* sont cochées, l'information suivante devrait s'afficher automatiquement en cliquant sur un point :
 - a. Données de présence (*Attributes*)
 - b. Données climatiques.



22a



22b

	Min Temp	Max Temp	Precipitation
Jan	-1	11	119
Feb	-0.4	10.5	134
Mar	-0.6	10.3	161
Apr	-2	10.7	87
May	-4.8	10.2	26
Jun	-6.0	11.2	6
Jul	-7	11.3	5
Aug	-7.1	11.1	11
Sep	-5.1	11.3	40
Oct	-3.8	10.3	86
Nov	-3	11.4	78
Dec	-2.1	11.3	87
Year	-3.6	11.1	840

Note

Dans l'onglet *Frequency*, vous pouvez surligner simultanément les points atypiques identifiés avec la méthode *Reverse jackknife (Outliers)* et ceux identifiés avec la méthode 1.5IQR. Notez que la méthode *Reverse jackknife* n'est pas adaptée pour de petits jeux de données, car elle est trop rigoureuse et réduirait significativement le nombre d'observations (y compris en ôtant des données correctes).

Travail personnel : Détectez les points erronés dans la base de données des points de présence de *Vasconcellea candicans* en utilisant différentes variables climatiques pour appliquer la méthode 1.5IQR.

Références

- Chapman AD. 2005a. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Chapman AD. 2005b. Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- DIVA-GIS. 2005. User Manual, version 5.2 [en ligne]. Disponible à l'adresse : http://www.diva-gis.org/docs/DIVA-GIS5_manual.pdf. Lien vérifié en mai 2012.
- Hijmans RJ, Garrett KA, Huamán Z, Zhang DP, Schreuder M, Bonierbale M. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology* 14(6):1755-1765.
- Scheldeman X, Willemen L, Coppens D'eeckenbrugge G, Romeijn-Peeters E, Restrepo MT, Romero Motoche J, Jimenez D, Lobo M, Medina CI, Reyes C, Rodriguez D, Ocampo JA, Van Damme P, Goetghebeur P. 2007. Distribution, diversity and environmental adaptation of highland papaya (*Vasconcellea* spp.) in tropical and subtropical America. *Biodiversity and Conservation* 16(6):1867-1884.
- Williams PH, Margules CR, Hilbert, DW. 2002. Data requirements and data sources for biodiversity priority area and selection. *Journal of Bioscience* 27(4):327-338.

Section B

Analyse des données

Chapitre 5

Analyse spatiale de la diversité en vue de planifier la conservation

On trouve un certain nombre de déclarations sur l'état actuel de la biodiversité, dans des articles et rapports consacrés à la conservation. Par exemple, des commentaires comme « 75 % de la diversité a été perdue » ou « ce site est riche en biodiversité ». Mais pour que de telles affirmations soient utilisées dans les processus d'aide à la décision ou pour établir des politiques, elles doivent être crédibles et fondées sur des méthodologies bien établies capables d'évaluer des niveaux de diversité basés sur des mesures comparables. Bien que ceci puisse paraître simple et évident, de nombreuses questions se posent encore quant à la meilleure manière de mesurer la diversité et de comparer et combiner les résultats d'études différentes. La disponibilité et l'utilisation de diverses options d'analyse sont souvent sources de confusion quant à l'exactitude des résultats issus d'une multitude d'études : les résultats peuvent être difficiles à comparer, ce qui nuit à leur validité. Ce chapitre vise à améliorer la compréhension du lecteur quant aux options disponibles lorsqu'on entreprend l'analyse spatiale de la diversité et les implications associées en termes de conservation. Le chapitre fournit également des lignes directrices pour sélectionner les méthodologies appropriées pour conduire les analyses et interpréter leurs résultats.

Le premier défi, quand on conduit n'importe quel type d'analyse de diversité, est de déterminer le niveau approprié auquel travailler. La diversité végétale peut être étudiée au niveau de la communauté (écosystème), au niveau des espèces ou au niveau génétique. Ce manuel est consacré uniquement à l'étude de la diversité au niveau de l'espèce et au niveau génétique. Au niveau de l'espèce, l'unité de diversité observée est l'espèce, dont on note la présence ou l'absence à un endroit donné. En matière de diversité génétique, l'unité de diversité observée peut être un trait phénotypique (le produit de l'expression d'un ou plusieurs gène[s]) ou une séquence de paires de bases de l'ADN (analyse des marqueurs neutres ou d'ADN fonctionnel connu, fondée sur des séquences ou des poids moléculaires).

La diversité (des espèces, variétés ou allèles) dans des sous-unités distinctes au sein d'une zone d'étude (connue sous le nom de diversité *alpha*) est le sujet principal de l'analyse spatiale de la diversité. Les sous-unités d'une zone d'étude – déjà identifiées – peuvent se référer à des sites, des unités administratives ou des cellules raster de toute taille choisie. Dans ce manuel, on utilise les cellules raster pour représenter des sous-unités de diversité. L'avantage d'utiliser des cellules raster est que celles-ci permettent de comparer la présence/absence d'espèces, de caractéristiques et d'allèles dans trois sous-unités de taille géographique similaire sur l'ensemble de la zone étudiée. Dans certains cas, des données vecteur sont utilisées, par exemple pour comparer le nombre d'espèces entre différents pays.

La mesure la plus directe de la diversité *alpha* consiste à compter le nombre d'unités de diversité observées (par exemple, le nombre d'espèces ou le nombre d'allèles par sous-unités de la zone d'étude). Connue sous le nom de , ce type de mesure est simple et assez facile à interpréter (voir les Sections 5.1 et 5.3). Les analyses dans ce chapitre se concentreront surtout sur ce type de mesure car elle est utilisée couramment pour évaluer la diversité au niveau spécifique. La richesse allélique est aussi considérée comme une mesure clé pour analyser la conservation de la diversité génétique d'une espèce d'intérêt (Frankel *et al.*, 1995 ; Petit *et al.*, 1998).

Un inconvénient de la mesure de richesse, cependant, est qu'elle dépend du nombre d'échantillons considérés dans chaque sous-unité de la zone d'étude. On trouve couramment de plus hauts niveaux de diversité lorsque de nombreux spécimens ont été collectés, tandis que le sous-échantillonnage conduit souvent à estimer de plus bas niveaux de diversité ; de tels résultats ne sont pas toujours exacts (Hijmans *et al.*, 2000). On peut remédier à une telle erreur en utilisant des méthodes analytiques plus complexes comme la raréfaction, qui recalcule la diversité mesurée dans chaque sous-unité de l'aire d'étude sur la base d'un nombre identique normalisé d'échantillons (Petit *et al.*, 1998 ; Leberg, 2002) (voir la Section 5.3). Malgré cela, la richesse d'un site spécifiques reste difficile à évaluer avec cette alternative, particulièrement quand on dispose d'un nombre limité d'observations. C'est souvent le cas lorsqu'on utilise un raster de haute résolution (par exemple des cellules de 1 ou 5 km) dans une analyse de diversité spatiale à grande échelle (par exemple, l'analyse décrite dans les Sections 5.1, 5.3 et 5.4). Dans cet exemple, il est impossible que chaque cellule du raster (souvent au nombre de plusieurs milliers) possède un grand nombre d'observations. Le Chapitre 6 examine comment appliquer la modélisation de la distribution d'espèces pour traiter ce problème d'échantillonnage incomplet ou biaisé.

Un autre inconvénient de mesurer la richesse est que la méthodologie ne prend pas en compte les proportions relatives des nombres d'unités de diversité observées. Par exemple, dans un site où 150 observations sont enregistrées pour un total de trois espèces, il peut y avoir 50 observations pour chaque espèce ou 148 pour une espèce et une seule pour chacune des deux autres. De toute évidence, le premier cas est plus riche en diversité que le second. Plusieurs indices, comme les indices de *Shannon* et de *Simpson*, ont été développés pour évaluer la diversité en tenant compte des proportions respectives de chaque espèce dans la zone d'étude, ce que l'on appelle aussi la mesure de l'*homogénéité*. Dans le cas de l'analyse de diversité à base de marqueurs moléculaires, des indices de fréquence allélique (comme mesure d'homogénéité) sont utilisés afin d'assurer que l'on prend en considération les proportions relatives des différents allèles de chaque locus dans la zone étudiée. Il y a cependant une difficulté, quand on utilise des indices qui prennent en charge les proportions ; ceux-ci ne sont en effet pas appropriés pour des cellules qui contiennent un nombre limité d'observations, ce qui est généralement le cas des analyses basées sur des rasters à haute résolution. Donc, lorsqu'on travaille avec des rasters à haute résolution, la richesse peut rester la méthode la plus appropriée pour mesurer la diversité.

Outre la richesse et l'homogénéité, il existe d'autres mesures, spécifiques de certains types de données, qui peuvent être appliquées pour évaluer la diversité *alpha*. Lorsqu'on mesure des traits morphologiques, des paramètres statistiques tels que la variance et le coefficient de variation peuvent être calculés pour déterminer les niveaux de diversité phénotypique dans un site donné (voir la Section 5.2). De plus, si l'on fait appel à une analyse multi-variée, on obtient des distances entre individus dans un espace multi-varié (c'est-à-dire des distances euclidiennes). Cette information peut ensuite être utilisée pour regrouper des individus semblables en vue d'analyses de richesse ultérieures. Dans le cas des marqueurs moléculaires, outre l'analyse fondée sur la richesse allélique décrite précédemment, des paramètres génétiques spécifiques (comme l'hétérozygotie ou la distribution d'allèles courants à l'échelle locale) peuvent aussi être utilisés pour décrire les différences en diversité intra-spécifique entre sous-unités.

La diversité *beta* est complémentaire à la diversité *alpha*. La diversité *beta* se focalise sur la divergence au sein des espèces, allélique ou relative aux traits, entre les différentes sous-unités de la zone d'étude. Par exemple, pour comprendre la structure spatiale de la diversité génétique, on peut organiser les sous-unités de la zone étudiée en groupes de composition similaire (voir la Section 5.3). Dans la Section 5.4 (sélection d'une réserve naturelle), l'analyse combine des mesures de diversités *alpha* et *beta* pour classer les sites par ordre de priorité pour la conservation. Dans le processus de sélection d'aires protégées, on met le plus souvent l'accent sur la conservation d'un nombre maximum d'espèces préservées (ou d'allèles : voir

Petit *et al.*, 1998). Il est important de réaliser que si l'on focalise les mesures de conservation uniquement sur les sites les plus riches en diversité, on peut en arriver à ignorer des espèces menacées que l'on ne trouve que dans des sites généralement pauvres en diversité (comme les écosystèmes de montagne, à la diversité réduite mais où les espèces sont uniques et absentes d'autres écosystèmes).

Généralement, les études de biodiversité évaluent l'état de menace des espèces ou la diversité génétique à un moment donné. La détection de changements dans la dynamique de la biodiversité demande des collectes répétées sur les mêmes sites, mais ce sujet dépasse le cadre de ce manuel.

5.1. Richesse des espèces

De nombreuses analyses de diversité sont axées sur le niveau de l'espèce. Comme indiqué plus haut, la richesse est la méthode la plus simple pour évaluer la diversité (*alpha*). Cette section décrit comment réaliser ce type d'analyse.

PROGRAMMES ET FICHIERS DE DONNÉES À UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel 	Fichiers de données : <p>Répertoire 5.1 <i>Species Diversity</i> (Diversité des espèces)</p> <ul style="list-style-type: none"> • <i>Vasconcellea species</i> (shp, shx, dbf) • <i>Latin America countries</i> (shp, shx, dbf)

5.1.1. Comment réaliser une analyse spatiale de la richesse spécifique

L'analyse ci-dessous utilise des données issues d'une étude de la diversité du genre *Vasconcellea* (Scheldeman *et al.*, 2007). Le genre *Vasconcellea* contient 21 espèces, qui sont toutes liées à la papaye commune (*Carica papaya*) et ses populations naturelles sont distribuées à travers toute l'Amérique latine. Du fait leur capacité à s'adapter à la haute altitude, les espèces sont souvent connues sous le nom de 'papaye d'altitude'. Tandis que certaines espèces sont cultivées pour leurs fruits (surtout dans les régions andines), d'autres sont utilisées comme source de matériel génétique dans des programmes de croisement de la papaye commune [par exemple pour favoriser des traits comme la tolérance au froid ou la résistance au virus « ringspot » de la papaye (PRSV-p)]. De plus, certaines espèces ont une aire de distribution étendue (on trouve *V. cundinamarcensis* du Costa Rica à la Bolivie), tandis que d'autres, comme *V. palandensis*, ont des aires de distribution limitées et sont en danger d'extinction. La conservation de ce genre est importante du fait de son potentiel à la fois pour la production fruitière et l'amélioration de la papaye, et ces efforts de conservation bénéficient grandement de l'information sur la distribution et la diversité du genre.

Avant de se lancer dans l'analyse, il est important de rappeler qu'il faut avoir un nombre suffisant d'observations dans une cellule de raster afin de réaliser des analyses fiables de la diversité. La crédibilité et l'exactitude du résultat final dépendra de la qualité de la stratégie d'échantillonnage, bien que le choix de la taille de cellule utilisée dans l'analyse (idéalement, la taille de cellule devrait être définie lors de la formulation de la stratégie d'échantillonnage) influe aussi sur la qualité du résultat. Si l'on effectue une analyse avec des cellules trop petites, le raster résultant générera une carte de haute résolution avec des résultats de valeur limitée,

puisque chaque cellule contiendra probablement trop peu de points de présence (souvent un seul) pour que l'on puisse détecter une structure spatiale de la diversité des espèces. Au contraire, si les cellules sont trop grandes, elles auront un nombre suffisant d'observations, mais la carte sera de mauvaise résolution, ce qui compliquera son interprétation et son utilisation.

L'analyse suivante utilisera le nombre d'espèces comme mesure de la diversité. L'analyse sera conduite au niveau régional (Amérique latine) et le raster utilisera une cellule de 1 x 1 degré (111 x 111 km à l'équateur, voir le tableau dans le Chapitre 2). Ici, vous apprendrez à utiliser le menu *Analysis (Point to Grid)* dans DIVA-GIS pour réaliser une analyse de diversité interspécifique sur le genre *Vasconcellea*.

Point to Grid

Etapes :

1. Commencez par visualiser deux couches : *Vasconcellea data* et *Latin America Countries*. Puis sélectionnez la couche des espèces de *Vasconcellea* et allez sur *Analysis/Point to Grid/Richness*.

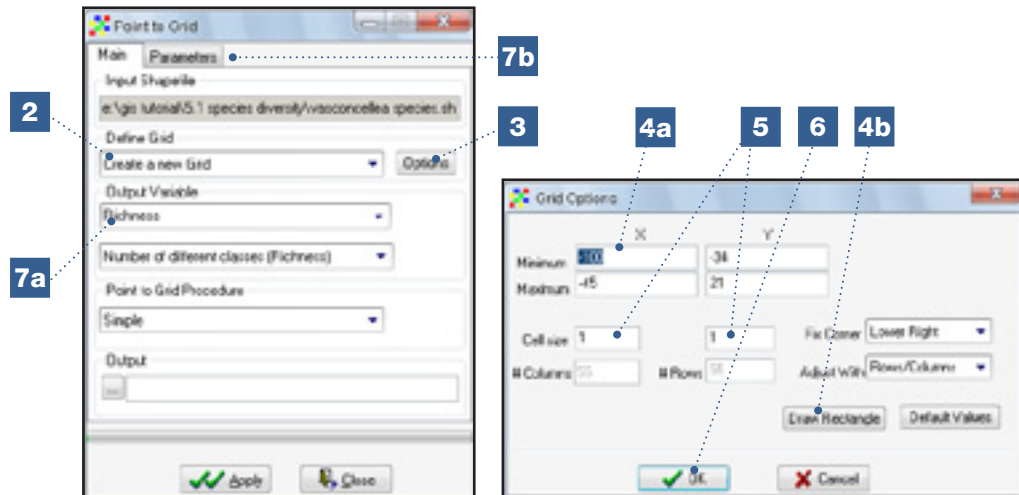


2. Ensuite, définissez les propriétés du raster qui sera utilisé pour l'analyse. Les dimensions de la zone d'étude ainsi que la résolution (taille de la cellule) doivent être déterminées. Dans la fenêtre *Point to Grid*, allez sur l'option *Define Grid* et sélectionnez *Create a New Grid* (option par défaut).
3. Cliquez sur la fenêtre *Options* pour définir les propriétés du raster : l'origine et l'étendue de la zone d'étude ainsi que la résolution (taille des cellules).
4. La zone d'étude peut être définie en utilisant l'une des options suivantes :
 - a. En saisissant manuellement, dans la fenêtre *Options*, les valeurs pour l'axe des X et l'axe des Y (vous pouvez choisir de sélectionner les options par défaut) ; ou
 - b. En traçant l'étendue de la zone sur la carte à l'aide de l'outil *Draw Rectangle*.

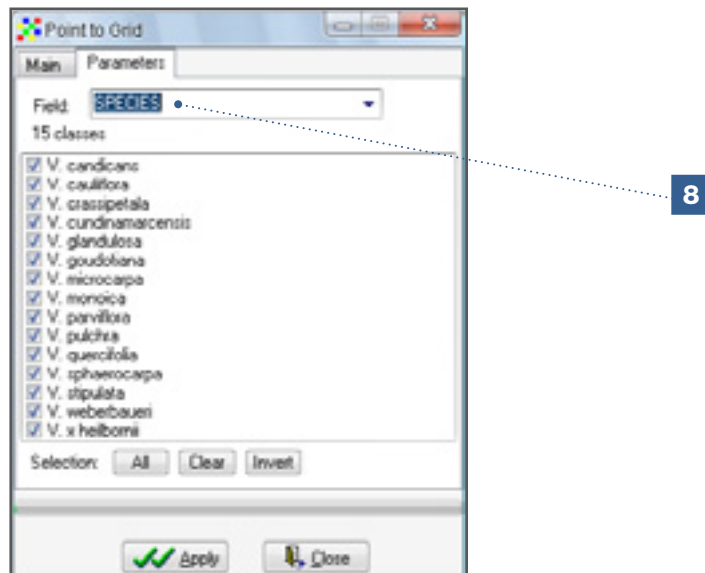
Les valeurs utilisées pour cette analyse sont : Min X : -100 ; Max X : -45 et Min Y : 34 ; Max Y : 21. Ce sont les valeurs par défaut lorsque la couche *Vasconcellea* est sélectionnée au début de l'analyse. Une autre origine de maillage générera un résultat légèrement différent.

5. La taille de cellule utilisée dans cette analyse sera d'un (1) degré (valeur par défaut dans *Cell Size*), ce qui équivaut à 111 km à l'équateur.
6. Cliquez sur *OK* pour accepter les valeurs.

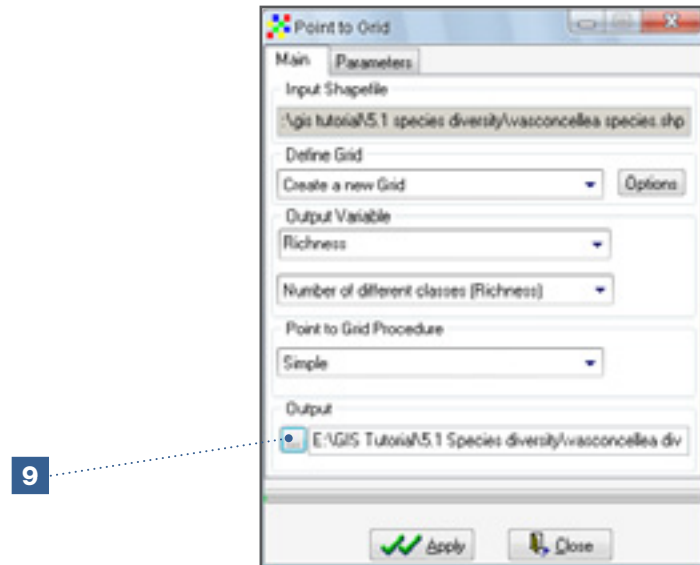
7. Retournez à la fenêtre *Point to Grid* pour sélectionner le type d'analyse à réaliser. Ici, nous entreprendrons une analyse de richesse des espèces.
 - a. Dans les listes déroulantes de *Outputs Variable*, sélectionnez d'abord la variable de sortie : *Number of different classes (Richness)*
 - b. Cliquez sur l'onglet *Parameters* pour sélectionner les unités.



8. Dans l'onglet *Parameters*, sous l'option *Field*, indiquez le paramètre que vous souhaitez analyser. Pour cette analyse, le paramètre sera *Species*, afin d'analyser la richesse des espèces. Vous pouvez exclure certaines espèces de l'analyse (en décochant les cases devant chaque espèce). Dans cette analyse, on inclura toutes les espèces.



9. Cliquez sur l'onglet *Main* et sélectionnez le bouton à gauche de la case *Output* (le bouton avec l'ellipse). Saisissez le nom du fichier à sauvegarder et son chemin d'accès. Enfin, cliquez sur *Apply*. Le raster résultant montrera le nombre d'espèces observées dans chaque cellule.



Les résultats de l'analyse (après avoir déplacé les couches) montre que les cellules situées au Sud de l'Equateur et au centre de la Colombie contiennent jusqu'à neuf espèces différentes de *Vasconcellea*, indiquant que, globalement, l'Equateur et la Colombie possèdent une diversité de ce genre supérieure à celle existant dans d'autres pays d'Amérique latine. Les données utilisées dans cette analyse correspondent à un sous-ensemble de l'information utilisée dans l'étude conduite par Scheldeman *et al.* (2007) et, de ce fait, les résultats sont légèrement différents des résultats donnés dans l'article cité. Dans l'étude complète, les espèces rares endémiques du Sud de l'Equateur sont aussi incluses, contribuant ainsi à des niveaux de diversité élevés dans cette zone.

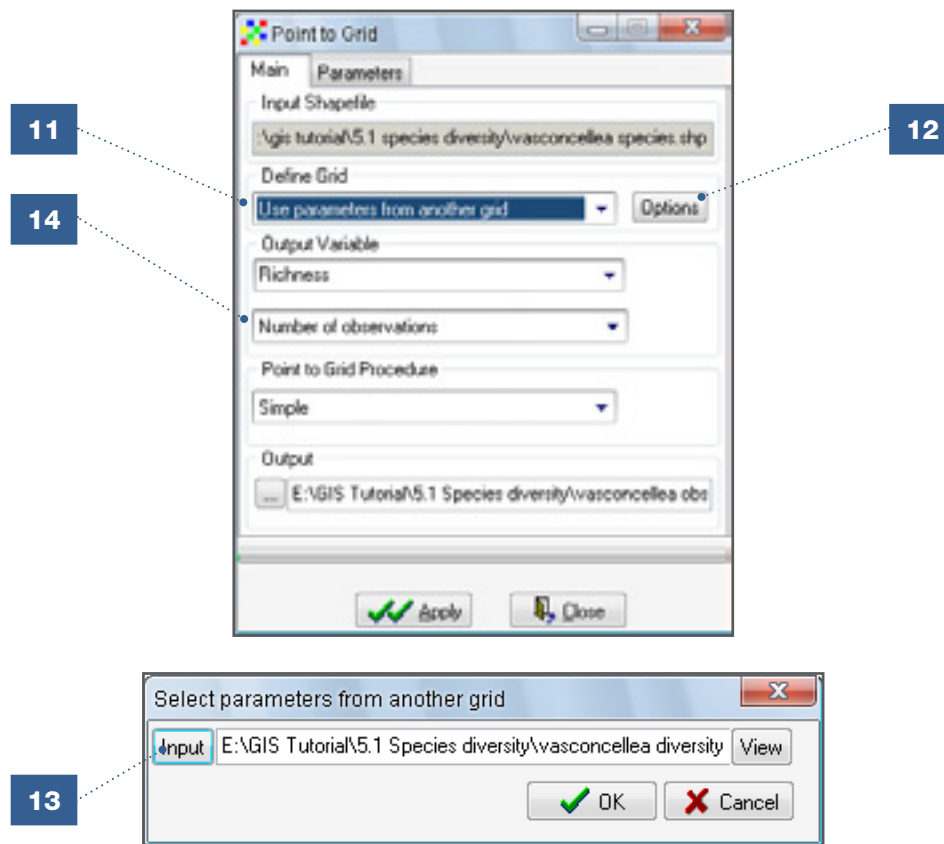
Travail personnel : Utilisez les outils de DIVA-GIS pour afficher une légende pertinente de la richesse des espèces (voir l'Analyse 3.1.2).

Pour réaliser cet exercice, vous devrez connaître la valeur de chaque cellule, qui peut être déterminée en cliquant sur le bouton *Information* ou en utilisant la flèche, avec l'information affichée sur la barre d'état de la carte (voir le Chapitre 3). Le résultat final devrait être une carte similaire à celle présentée ci-dessous.



10. Vérifiez le nombre d'observations dans chaque cellule. Pour comparer les résultats de cette analyse avec ceux de l'analyse précédente, il est important d'utiliser les mêmes définitions de raster (voir le Chapitre 3). Allez sur *Analysis/Point to Grid/Richness* (voir l'Etape 1).
11. Dans l'option *Define Grid*, sélectionnez *Use parameters from another grid*.
12. Cliquez sur le bouton *Options* pour sélectionner le raster dont vous souhaitez utiliser les paramètres (celui qui a été créé au cours des étapes précédentes).
13. A l'aide du bouton *Input*, sélectionnez le fichier raster de richesse (*.grd) créé dans les étapes ci-dessus ; cliquez sur *OK*.

14. Retournez à la fenêtre *Point to Grid* et sélectionnez *Number of observations* dans la seconde fenêtre située sous *Output Variable*.



15. Bien que cette analyse soit axée sur le nombre d'observations différentes par cellule (qui est indépendant de l'unité de diversité observée que vous êtes en train d'analyser), un champ doit toujours être indiqué dans la fenêtre *Parameters*. Cette analyse utilisera le champ *Species*, comme à l'Etape 8.

16. Indiquez un nom approprié pour le fichier (cliquez sur le bouton à gauche de la case *Output*) et cliquez sur *Apply*.

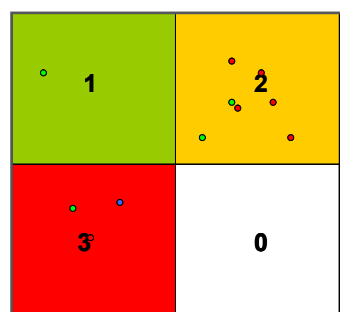


Le résultat (après avoir édité la légende du raster comme expliqué dans le Chapitre 3) de cette analyse révèle que la plupart des observations sont issues du Sud de l'Equateur. Il s'agit d'une situation typique dans laquelle les zones connues pour posséder de hauts niveaux de diversité sont souvent les sites préférés des botanistes et des chercheurs, ce qui entraîne des efforts plus intenses d'échantillonnage et un plus grand nombre d'observations. Un tel échantillonnage inégal prévaut souvent dans les parcs nationaux près des grandes villes ou dans les zones d'endémisme élevé. C'est un problème courant connu sous le nom d'*échantillonnage biaisé*¹. Ce problème peut être partiellement traité avec la méthode de *raréfaction*, exposée dans la Section 5.3, mais l'utilisation de cette méthode conduit à perdre certaines observations. Dans la Section 6.4, on explique comment améliorer les études de diversité en utilisant la modélisation de la distribution d'espèces. L'option *Point to Grid* de DIVA-GIS contient aussi plusieurs estimateurs de richesse afin de pallier, partiellement, à ce problème. Ces outils sont aussi utiles pour estimer le nombre supplémentaire d'espèces que l'on peut trouver dans chaque unité géographique de mesure, et qui n'ont pas encore été observées du fait du sous-échantillonnage. Consultez la Section 6.2.2 de la Version 5.2 du Manuel de DIVA-GIS si vous souhaitez avoir plus d'information sur ces estimateurs (http://www.diva-gis.org/docs/DIVA-GIS5_manual.pdf). La meilleure solution, cependant, est d'éviter de tels biais en s'assurant de faire, autant que possible, un échantillonnage homogène.

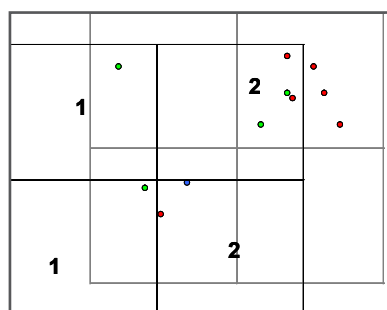
Effets du changement d'origine d'un raster sur le résultat d'une analyse *Point to Grid*

Une analyse avec un raster à partir de points d'indices (*Point to Grid Analysis*) prend en compte les observations trouvées dans chaque cellule du raster (c'est-à-dire que le calcul de la richesse s'appuie sur le nombre d'unités de diversité observées, par exemple nombre d'espèces, dans chaque cellule).

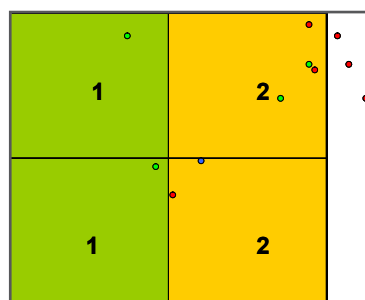
La définition du raster influe évidemment sur le résultat de l'analyse. Les sections qui suivent expliquent comment les différences entre les tailles de cellules des rasters (la résolution) influent sur le résultat final. Outre la taille de ses cellules, le raster est aussi défini par son origine (les valeurs minimum et maximum X et Y saisies dans *Grid Options*). Voici une illustration simple de cet effet sur un raster de quatre cellules.



Richesse selon le raster 1



Utilisation de raster de différentes origines pour les mêmes données



Richesse selon le raster 2

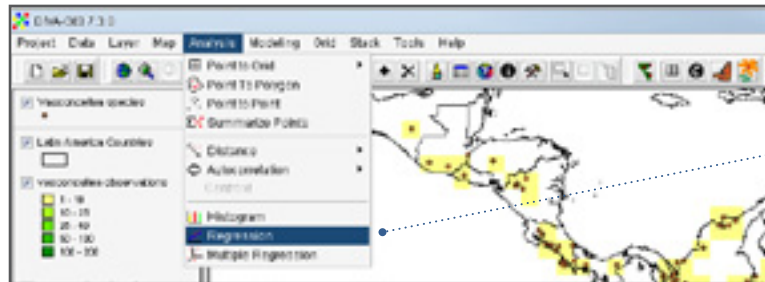
Du fait du faible nombre de cellules du raster, l'effet est exagéré dans cet exemple. Les résultats d'un échantillonnage plus complet seront moins influencés par un changement des propriétés du raster.

¹ Pour plus d'information sur l'échantillonnage biaisé, vous pouvez vous référer à Hijmans *et al.* (2000).

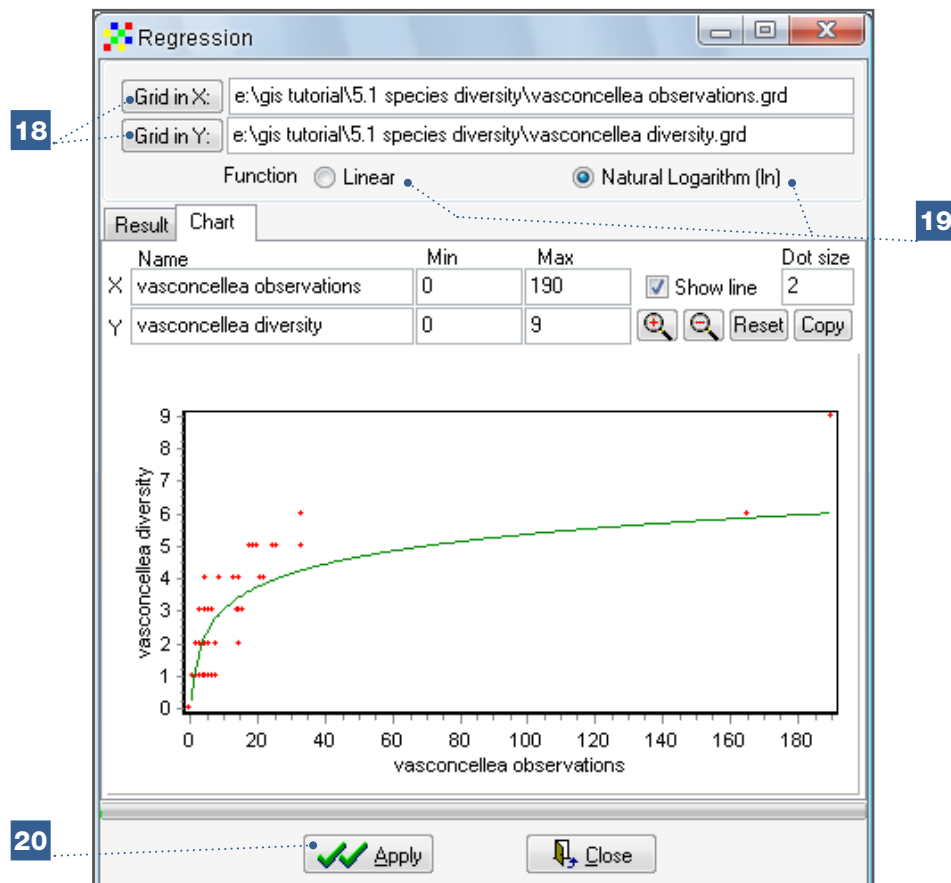
Visualisation de la courbe d'accumulation d'espèces pour évaluer s'il existe un échantillonnage biaisé

Etapes :

17. Une analyse de régression vous permet de visualiser le phénomène souligné dans l'encadré précédent. Dans le *Menu*, allez sur *Analysis/Regression*.



18. Sélectionnez les couches sur la richesse spécifique et sur le nombre d'observations que vous avez générées récemment.
19. Vous pouvez choisir entre une régression linéaire et une régression logarithmique naturelle. Une régression linéaire est assez simple, tandis qu'une régression logarithmique est complexe sur le plan mathématique mais peut mieux représenter des biais dans l'échantillonnage du fait qu'elle prend en compte la stabilisation (mise à niveau) des courbes d'accumulation d'espèces au fur et à mesure que l'échantillonnage (nombre d'observations) s'accroît.
20. Cliquez sur *Apply*.

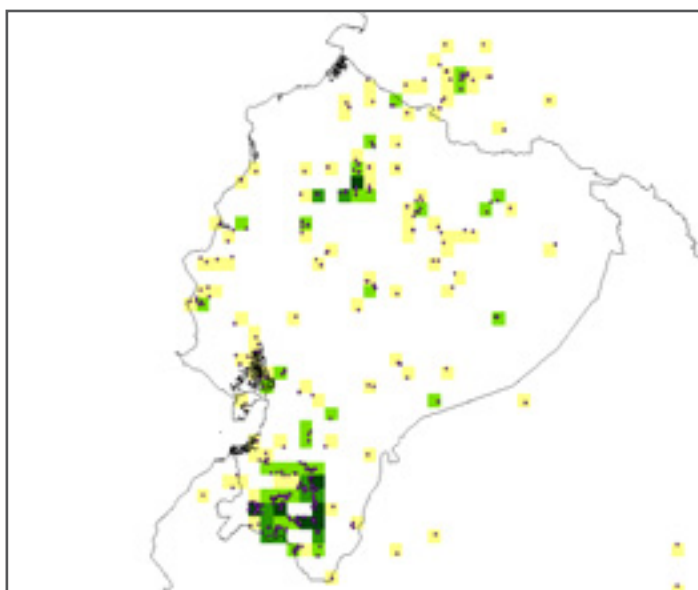


Le graphe dans l'onglet *Chart* confirme qu'il y a une relation entre le nombre d'échantillons utilisés et le nombre d'espèces observées (il est probable qu'un plus grand nombre de points d'échantillonnage conduise à une richesse plus élevée). Au fur et à mesure que le nombre d'observations (points d'échantillonnage) s'accroît, le graphique commence à se stabiliser, ce qui entraîne des questions à la fois sur la stratégie d'échantillonnage et sur la taille de cellule utilisées dans l'analyse (utiliser de plus grandes cellules et inclure plus d'observations serait plus approprié). Ce n'est que dans le cas où l'utilisateur connaît bien les données et qu'il est certain qu'elles proviennent d'un échantillonnage exhaustif et relativement homogène, que les résultats obtenus ci-dessus peuvent être considérés comme fiables et donc utiles. Les résultats doivent être interprétés avec prudence si l'échantillonnage n'a été réalisé que partiellement ou s'il a été fragmenté, ou encore si l'origine des données est inconnue. Dans de tels cas, les résultats peuvent refléter une représentation invalide et biaisée de la réalité. Comme il a été mentionné dans l'introduction du Chapitre 4, de nombreuses analyses spatiales de la biodiversité sont partiellement ou totalement fondées sur des données compilées à partir d'herbiers ou de banques de gènes, qui reflètent souvent un échantillonnage non systématique et inégal (Hijmans *et al.*, 2000 ; Chapman, 2005).

Travail personnel : Analysez minutieusement la diversité en Equateur avec un raster de 10 minutes (cellules d'approximativement 18 km) (seulement en Equateur, à l'aide de l'option *Draw Rectangle*). Indice : Quelle sera la taille de cellule souhaitable ?

Note

Vos résultats doivent être légèrement différents de ce qui suit puisqu'ils dépendent de l'origine du raster. Comme vous analysez des résultats seulement pour l'Equateur, ignorez le message d'alerte qui indique qu'il y a des points en dehors du raster sélectionné.



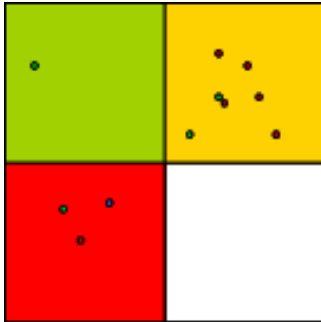
Utiliser l'option *Circular Neighborhood* pour l'analyse de richesse

Jusqu'ici, les résultats des analyses ont été largement dépendants de la définition du raster, particulièrement de la taille de la cellule, ou résolution. Une taille réduite engendre une plus haute résolution (niveau de détail) mais risque de passer à côté de structures spatiales (par exemple lorsque les cellules sont si petites que chacune ne contient qu'une observation).

La situation peut être améliorée avec l'option *Circular Neighborhood*, qui prend en compte la diversité dans les zones adjacentes. Avec cette option, chaque cellule reçoit la valeur de la diversité trouvée à l'intérieur d'un cercle d'un diamètre spécifié centré sur la cellule, au lieu de la valeur de la diversité trouvée dans la seule cellule.

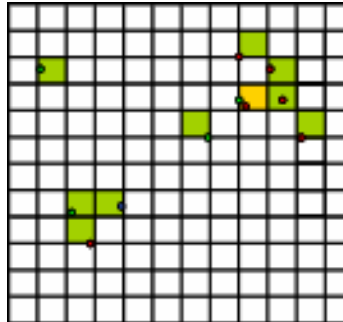
Le concept de voisinage circulaire

Les diagrammes ci-dessous représentent le concept de voisinage circulaire, où les cercles de couleurs différentes indiquent des richesses d'espèces différentes : le vert indique une valeur de richesse d'une espèce, le jaune de deux et le rouge de trois espèces. Quand on utilise cette méthode, les résultats près des marges de la distribution des points peuvent être faussés par l'assignation d'une diversité plus basse aux frontières dans le cas d'un échantillonnage incomplet (par exemple aux frontières d'un pays).



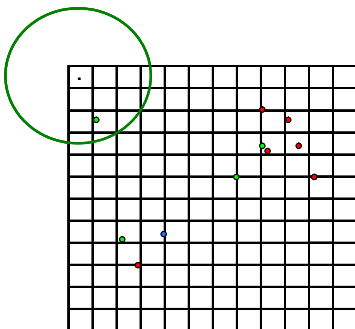
Richesse dans une cellule d'un (1) degré

On observe clairement une diversité plus grande dans la cellule en bas à gauche ; mais le raster est de faible résolution.

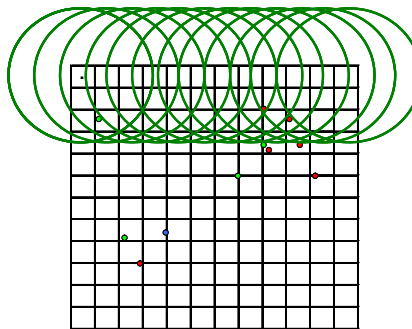


Richesse dans des cellules de 10 minutes

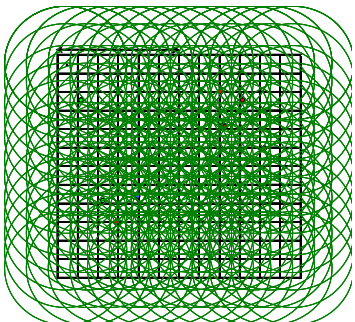
Dans ce cas, le raster est plus détaillé (haute résolution), mais la structure spatiale de la diversité a été perdue.



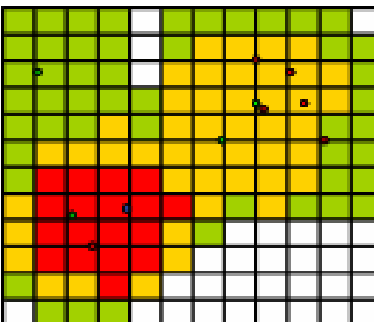
Le voisinage circulaire consiste à utiliser un cercle de diamètre fixe centré sur chaque cellule du raster pour assigner une valeur à la cellule ; la valeur de la cellule en haut à gauche est un (1), du fait qu'il n'y a qu'une espèce à l'intérieur du cercle indiqué.



Le voisinage circulaire se répète pour chaque cellule. Dans cet exemple, le cercle a un diamètre d'un (1) degré, la même taille que celle de la cellule dans le premier exemple.

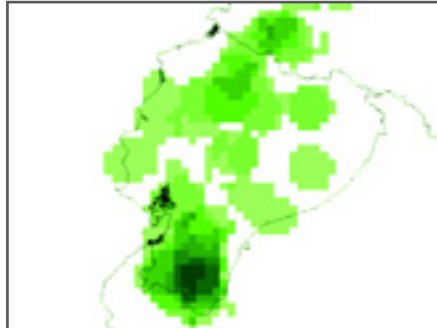


Comme cela est illustré, chaque observation contribue à la valeur de plusieurs cellules.

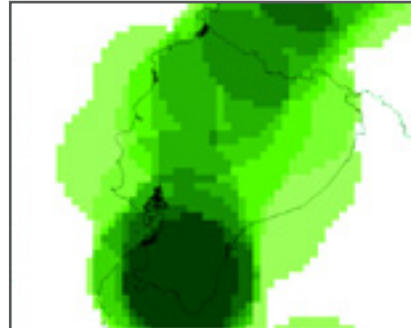


Richesse avec des cellules de 10 minutes et un voisinage circulaire d'un (1) degré. Cette méthode permet de maintenir un haut degré de résolution sans perdre la structure de la diversité.

Il peut être difficile de définir le rayon optimal du cercle (le rayon idéal est souvent obtenu par tâtonnements). L'exemple ci-dessous illustre le fait que lorsqu'un cercle devient trop grand (à droite), les résultats de l'analyse sont peu utiles, le niveau de détail nécessaire pour les interpréter ayant été perdu.



Analyse de richesse Point to Grid avec une taille de cellule de 10 minutes et un voisinage circulaire d'un (1) degré.

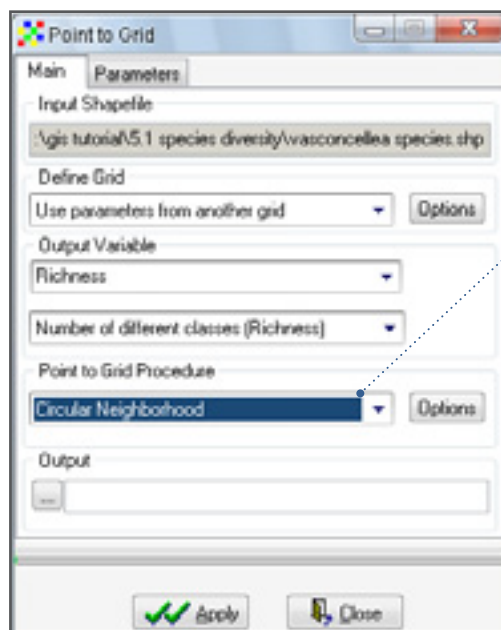


Analyse de richesse Point to Grid avec une taille de cellule de 10 minutes et un voisinage circulaire de trois (3) degrés.

Pour observer l'effet de l'option *Circular Neighborhood* et les changements entraînés dans le résultat précédent, l'analyse de richesse va maintenant être répétée avec des cellules de 10 minutes et un cercle de voisinage d'un (1) degré.

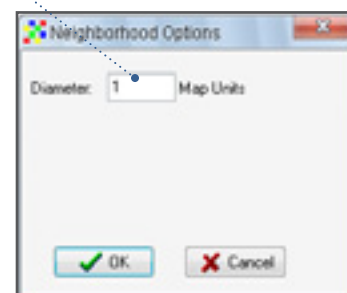
Étapes :

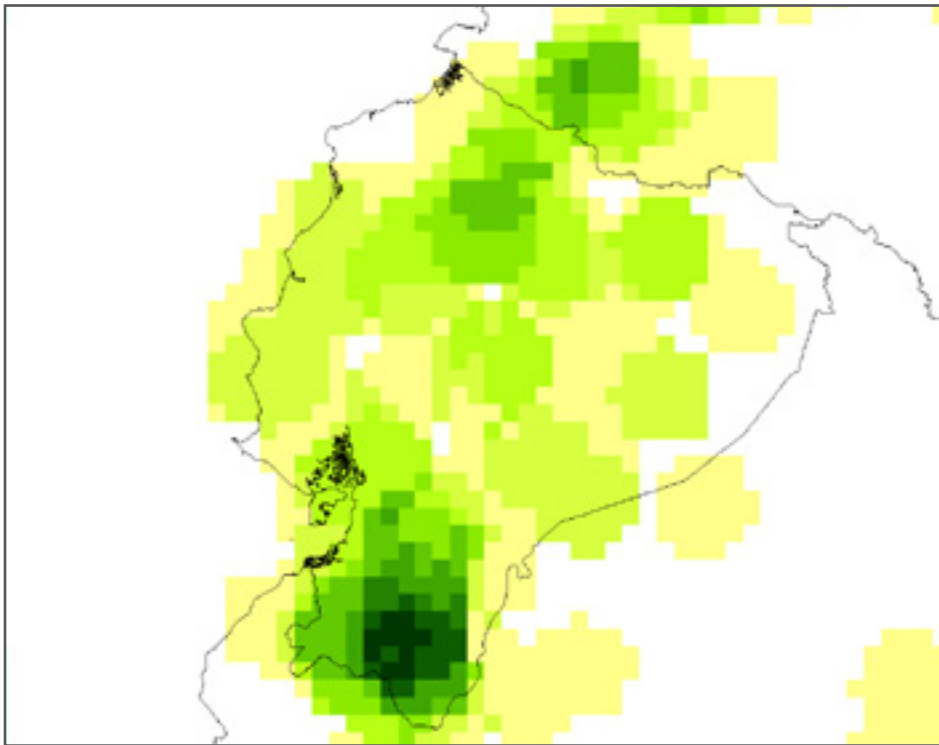
21. Dans la fenêtre *Point to Grid Procedure*, sélectionnez l'option *Circular Neighborhood*.
22. Utilisez les paramètres du raster généré au cours de l'analyse précédente (la taille des cellules du raster est 0,1666). Cliquez sur *Options* à côté de *Circular Neighborhood*. La fenêtre *Neighborhood Options* apparaît. Saisissez un (1) comme *Map Unit* pour indiquer la taille désirée [dans le cas de cartes basées sur des coordonnées latitude/longitude, l'unité de la carte est un (1) degré]. Répétez les étapes de l'analyse précédente.



21

22





Utiliser l'option *Circular Neighborhood* vous donne des structures spatiales plus précises pour la diversité existante, ainsi qu'une résolution assez élevée.

Travail personnel : Observez ce qui se passe si vous appliquez un grand cercle de voisinage (*Circular Neighborhood*) (par exemple 5 *Map Units*).

5.2. Analyse de la diversité intra-spécifique basée sur des données phénotypiques

Les études de la diversité génétique, y compris des structures spatiales de la diversité génétique, sont fréquemment basées sur des données de marqueur moléculaire (voir la Section 5.3). Cependant, les données phénotypiques, particulièrement les données morphologiques, peuvent constituer une autre source indirecte d'information sur la diversité génétique. Les données phénotypiques issues d'un individu unique varient en fonction soit du génotype (G), soit de l'environnement (E), ou d'une combinaison des deux (effet GxE). Certains traits, comme la couleur des fleurs, ne sont pas influencés par l'environnement. Lorsqu'on utilise des données basées sur une caractérisation *in situ*, conduite dans des lieux hétérogènes sur le plan environnemental, il est recommandé de se focaliser sur ces traits. Lorsqu'on travaille sur d'autres traits, une option permettant de minimiser la variation déterminée par l'environnement consiste à effectuer la caractérisation en dehors du site originel de collecte et dans des conditions environnementales contrôlées et uniformes (caractérisation *ex situ*), que ce soit dans le même lieu géographique (par exemple des champs expérimentaux) ou dans un environnement contrôlé (comme dans des serres). Bien que les effets de l'environnement puissent encore jouer un rôle dans la caractérisation *ex situ*, en particulier dans des champs expérimentaux – par exemple lorsque les individus observés sont détenteurs de traits qui les rendent mieux adaptés aux conditions du site choisi – la caractérisation dans les champs expérimentaux sera plus pertinente que celle que l'on obtiendrait *in situ*.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION

Programmes :	Fichiers de données :
<ul style="list-style-type: none"> • DIVA-GIS • Programme statistique (optionnel) 	<ul style="list-style-type: none"> • Répertoire 5.2 <i>Diversity - Phenotypic data</i> (Diversité – données phénotypiques) • <i>Manihot ex situ</i> (<i>shp, shx, dbf</i>) (Données de caractérisation du manioc <i>ex situ</i> à Ucuyali) • <i>PER_Adm0 ; PER_Adm1 ; PER_Adm2</i> (<i>shp, shx, dbf</i>) (Données administratives du Pérou) • <i>Peru_towns</i> (<i>shp, shx, dbf</i>) (Municipalités du Pérou) • <i>PER_roads</i> (<i>shp, shx, dbf</i>) (Données sur les routes du Pérou) • <i>PER_water_areas_dcw ; PER_water_lines_dcw</i> (<i>shp, shx, dbf</i>) (Données sur les rivières et étendues d'eau du Pérou)

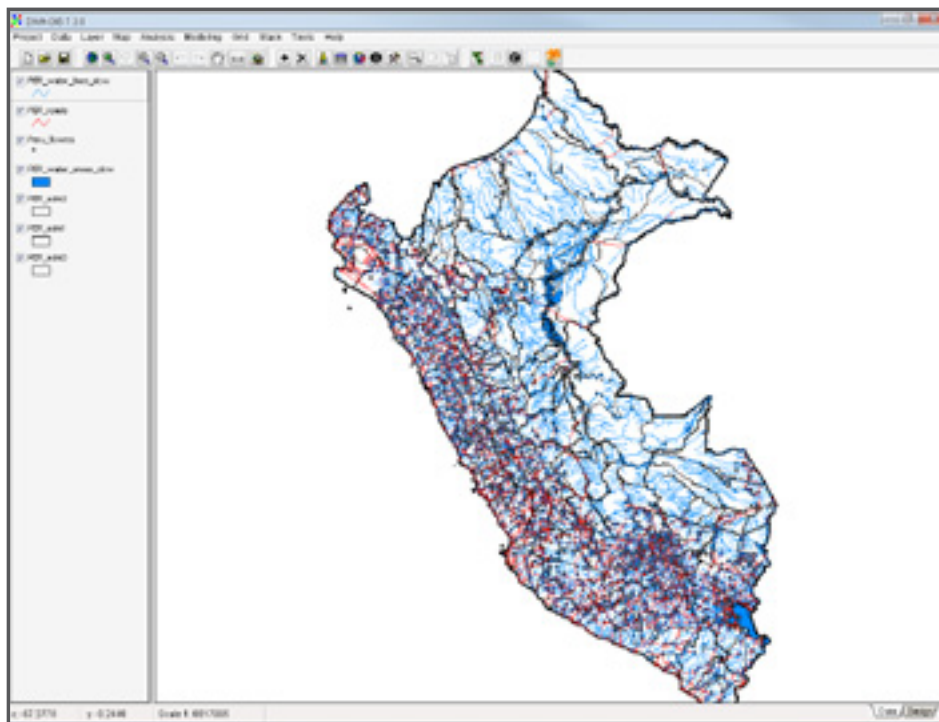
5.2.1. Comment réaliser une analyse de diversité spatiale en utilisant des données phénotypiques

Nous donnons ci-dessous un exemple d'analyse spatiale fondée sur des données phénotypiques issues d'une caractérisation morphologique *ex situ* (dans le même champ expérimental), combinées à des données de passeport (information sur le site où le matériel a été originellement récolté). Les données utilisées dans cette analyse proviennent d'une étude de la diversité et de l'érosion génétique du manioc (*Manihot esculenta*) en Amazonie péruvienne (Province d'Ucayali) (Willemen *et al.*, 2007). Dans cette section, vous apprendrez à utiliser le menu *Analysis (Point to Grid)* de DIVA-GIS pour conduire une analyse de diversité intraspécifique fondée sur des données morphologiques.

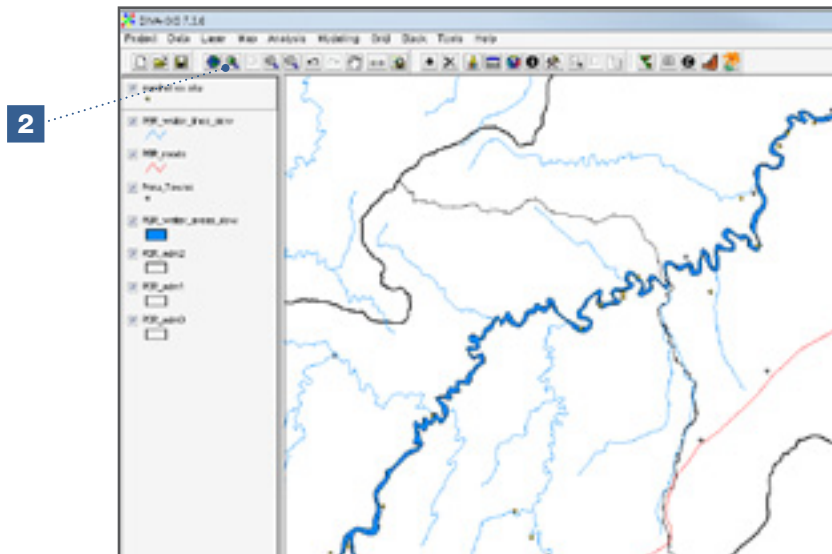
Point to Grid

Etapes :

1. Commencez par examiner les options de DIVA-GIS présentées dans le Chapitre 3 (Eléments de base). Créez une carte du Pérou qui montre trois niveaux administratifs : pays (épaisseur de ligne = 3), province (épaisseur de ligne = 2), départements et villes (épaisseur de ligne = 1). Ajoutez les couches correspondant aux rivières et aux étendues d'eau (en bleu), aux routes (en rouge) et aux villages (en gris). La carte résultante devrait ressembler à ceci :

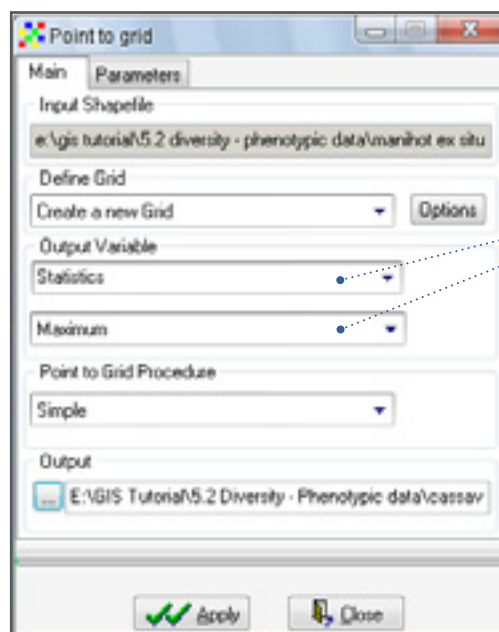


2. Ajoutez maintenant la couche correspondant aux données de caractérisation du manioc (*Manihot ex situ.shp*) et zoomez sur ces données à l'aide de l'option *Zoom to theme*.

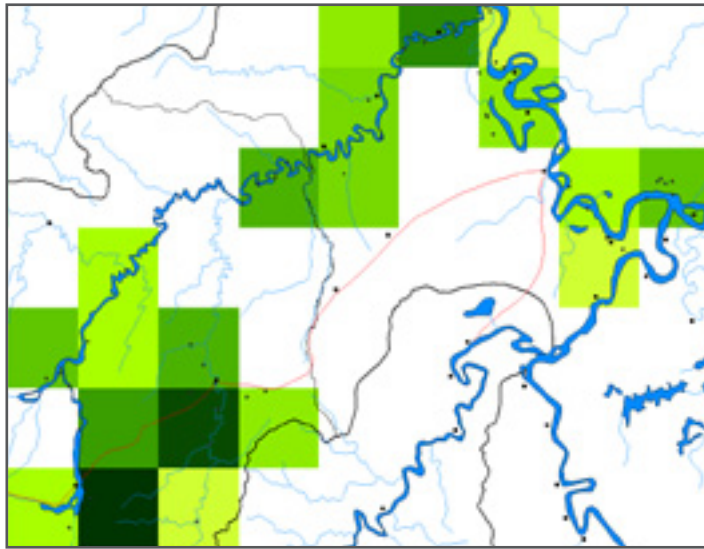


3. Explorez les valeurs de la caractérisation *ex situ* et identifiez les zones où l'on a trouvé les racines de manioc les plus lourdes. On utilisera dans cet exercice le paramètre correspondant au poids moyen de la racine fraîche (*Fresh Root Weight, FRW*). Sélectionnez la couche de caractérisation du manioc et allez sur les options *Analysis/Point to Grid/Statistics*.

Dans le cadre *Define Grid*, en utilisant le bouton *Options*, définissez, dans l'option *Cell Size*, un raster de 10 minutes (taille de cellule = 0,1666). Ce raster sera utilisé pour toutes les analyses de cette section. Dans la fenêtre *Output Variable*, sélectionnez l'option *Maximum*. Dans la fenêtre *Parameters*, sélectionnez le paramètre à analyser (dans le cas présent : *FRW*). Pour finir, donnez un nom au fichier ainsi généré. Changez la légende sur la carte pour améliorer son affichage et son interprétation.

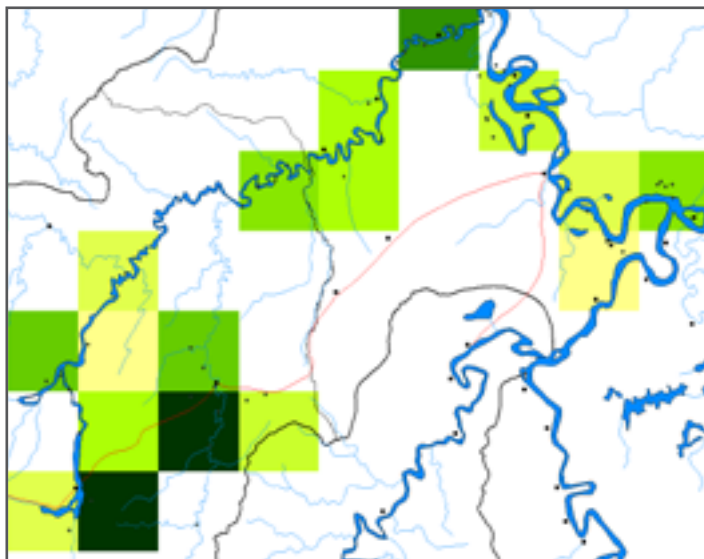


Le résultat (après modification de la légende) montre l'emplacement des zones où l'on trouve les individus qui présentent les racines les plus lourdes².



Les accessions de manioc qui ont les racines les plus lourdes se trouvent principalement dans le Sud-est de la zone étudiée. Cette analyse fournit une information sur une caractéristique spécifique (dans le cas présent, le poids de la racine fraîche) mais elle n'offre pas d'information sur la diversité dans la zone étudiée. Pour obtenir cette information, vous devez vous intéresser à la plage de variation du paramètre analysé (FRW).

4. Dans *Analysis/Point to Grid/Statistics*, sélectionnez le paramètre *Range*. Utilisez le même raster que dans l'analyse précédente (FRW). Notez que la légende utilisée précédemment n'est plus applicable du fait que le paramètre calculé (plage de variation au lieu de maximum) a été changé et que les dimensions sont différentes. Ajustez la légende aux nouveaux résultats.

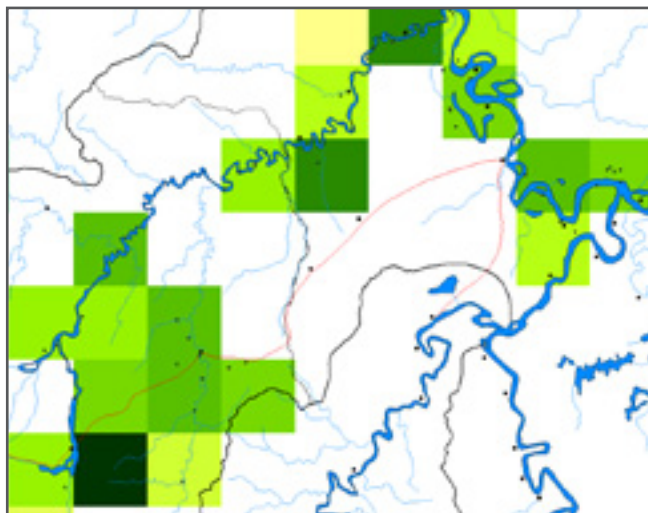


² Les résultats peuvent être légèrement différents de la carte ci-dessus, du fait qu'ils dépendent de l'origine du raster. Cette analyse a utilisé un raster aux caractéristiques suivantes : Min X : -78 ; Max X : -71 / Min Y : -12 ; Max Y : -5.

La carte résultante révèle que les cellules correspondant aux sites où l'on trouve le manioc avec les racines les plus lourdes, correspondent aussi aux cellules qui présentent la plus grande variation du poids. Bien que la plage de variation donne une idée de la variabilité d'un paramètre, elle a aussi deux inconvénients : d'abord, la plage de variation ne prend en compte que les valeurs extrêmes (et non la distribution statistique des valeurs observées) ; et ensuite, la valeur du résultat final dépendra des unités de mesure (dans le cas présent, le gramme). Combiner ces résultats avec ceux tirés de l'analyse d'autres traits, comme la longueur du pétiole ou le nombre de lobes des feuilles, peut s'avérer délicat, mais n'en reste pas moins important pour comprendre la diversité globale (qui demande d'inclure plusieurs traits différents).

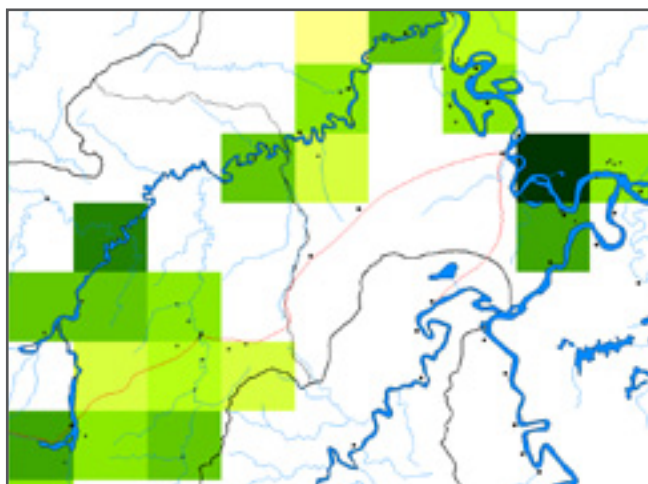
Il est plus facile de combiner des analyses dans lesquelles on utilise des paramètres sans dimensions (par exemple, le coefficient de variation [CV]), qui, contrairement aux autres, s'attache à la distribution d'un paramètre autour de sa moyenne.

5. Dans les fenêtres *Analysis/Point to Grid/Statistics*, sélectionnez le paramètre *Coefficient of variation*.



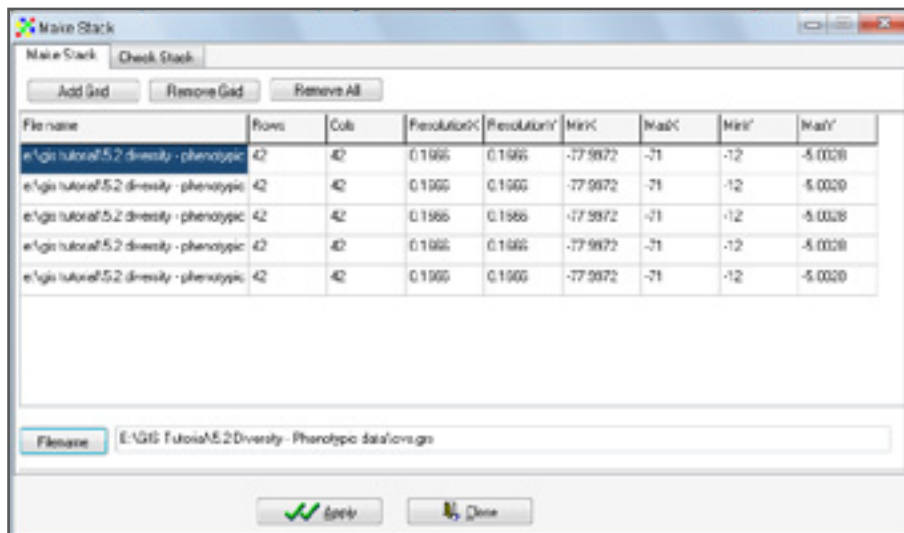
Cette carte permet de mieux visualiser la diversité du paramètre FRW dans les différentes cellules.

6. Analysez le CV d'un autre trait : la longueur du pétiole (en anglais, *Length of Leaf Petiole*, LLP). Utilisez les mêmes propriétés de raster que dans l'analyse précédente afin de pouvoir comparer les résultats de ces deux analyses.

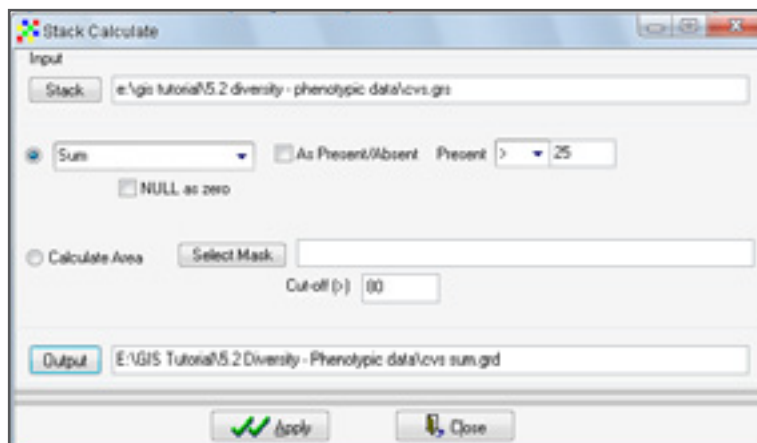


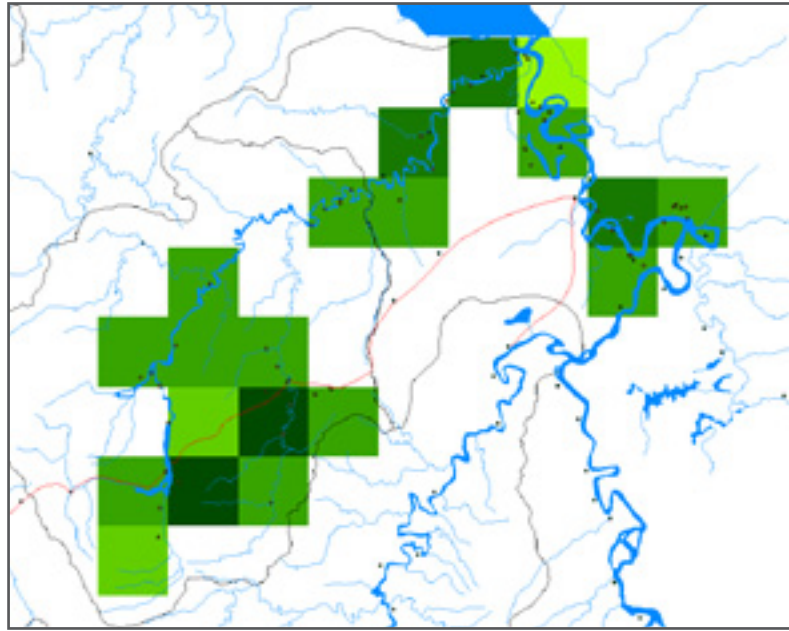
Les zones qui présentent la plus grande variabilité du paramètre LLP sont différentes de celles qui ressortent du paramètre FRW. Comment peut-on alors établir où sont les sites qui présentent la plus grande diversité de manioc ? La solution consiste à combiner les différents traits. Comme mentionné plus haut, à la différence de la plage de variation, l'utilisation du *Coefficient of Variation* permet de combiner plusieurs paramètres, ce que requiert une telle analyse.

7. Cinq traits sont maintenant inclus dans la comparaison : les deux traits analysés précédemment, FRW et LLP; et trois nouveaux paramètres : le nombre de lobes foliaires (en anglais, *Number of Leaf Lobes*, NLL), la distance entre nœuds (en anglais, *Distance Between Nodes*, DBN) et les jours de récolte (en anglais, *Days to Harvest*, DH). Trois couches de CV supplémentaires doivent être créées.
8. Les couches générées doivent maintenant être combinées en une couche unique, en calculant le CV moyen des cinq couches. Allez à *Stack/Make Stack* et ajoutez les cinq rasters correspondant aux CVs calculés, à l'aide du bouton *Add Grid*. Assignez un nom au *stack* et cliquez sur *Apply*. La combinaison de différents rasters en *stack* n'est possible que si les rasters partagent tous les mêmes propriétés, comme souligné dans le Chapitre 3.



9. Allez à la fenêtre *Stack/Calculate* et calculez le raster moyen qui intègre les cinq rasters du *stack*.



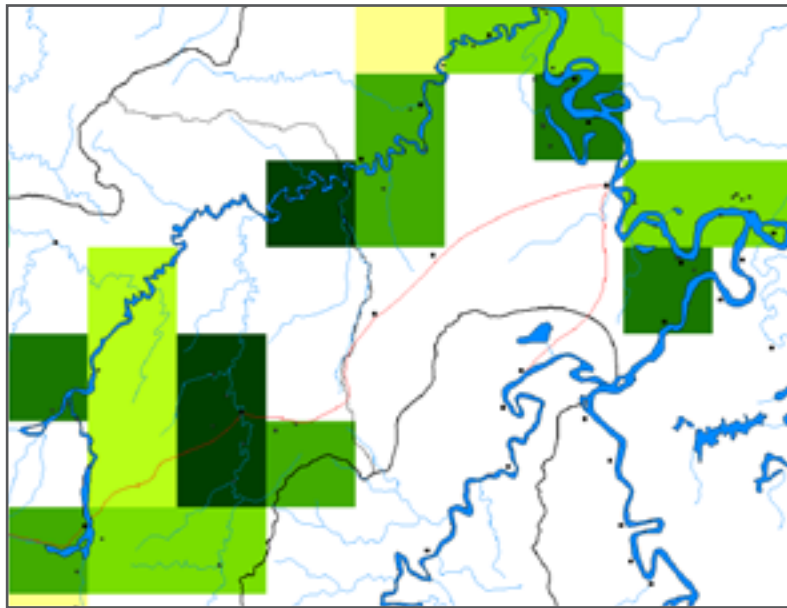


La combinaison des couches correspondant à cinq caractères morphologiques différents produit une visualisation plus forte de la diversité.

Ceci est évidemment une analyse de diversité simplifiée. Idéalement, il faudrait prendre en compte un plus grand nombre de traits quantitatifs et qualitatifs. Pour travailler avec un plus grand nombre de variables, particulièrement des variables qualitatives, il faut utiliser une statistique multi-variée, qui requiert des logiciels statistiques supplémentaires. Comme ceci dépasse le cadre de ce manuel, vous pouvez consulter les ouvrages de Grum *et al.* (2007), Kindt *et al.* (2005) et Mathur *et al.* (2008) pour de plus amples informations.

Les données relatives aux traits phénotypiques et/ou à la présence d'allèles (reflétant respectivement le phénotype et le génotype d'un individu) peuvent être utilisées comme indicateurs de la diversité dans les différentes sous-unités géographiques de la zone étudiée, ainsi que pour comparer, directement (allèles) ou indirectement (traits), les similarités génétiques entre individus. Ces similarités peuvent être utilisées dans des analyses multi-variées, comme l'analyse de groupes. Pour illustrer cette manière de faire, le fichier *manihot ex situ.shp* inclut les résultats d'une analyse de groupes. La variable *Cluster* montre à quel groupe chaque individu appartient, sur la base des similarités entre traits. Cette information permet de conduire une analyse de diversité comme celle illustrée dans la Section 5.1, mais en utilisant, ici, le nombre de groupes comme unité de diversité, à la place des différentes espèces. Dans ce cas, la diversité est définie par le nombre de groupes différents trouvés sur le site.

Travail personnel : Réalisez une analyse de richesse simple à l'aide des résultats de l'analyse de groupes (variable : *Cluster*), avec le même raster que dans les analyses précédentes.



Il y aura des différences entre les résultats des analyses fondées sur le coefficient de variation et ceux produits par les analyses multi-variées. L'analyse de groupes met en œuvre une méthodologie de statistique multi-variée plus complexe, et manipule un plus grand nombre de variables (Willemen *et al.*, 2007). Néanmoins, les tendances de diversité révélées dans les résultats de ces deux types d'analyse sont les mêmes : on trouve une diversité faible autour de la ville de Pucallpa (au bord de la rivière, là d'où part la route vers Lima) et un niveau élevé de diversité dans la zone centrale près de la route vers Lima (coin en bas à gauche de la carte). La cellule qui présente la plus grande diversité est la même dans les deux analyses. Comme indiqué dans la Section 5.1, la technique du voisinage circulaire peut aussi être utilisée pour de telles analyses.

5.3. Analyse de diversité intra-spécifique fondée sur les données de marqueurs moléculaires

Comme noté plus haut, les données phénotypiques et morphologiques peuvent être utilisées dans les analyses de diversité intra-spécifique ; toutefois, l'environnement qui règne là où a été effectuée la caractérisation influera toujours sur les résultats. La composition allélique ou les séquences de gènes des individus observés ne sont pas influencées par des tels facteurs environnementaux (un changement de conditions climatiques, comme une année humide par rapport à une année sèche, ne modifie pas la composition des paires de bases de l'ADN, mais il affectera en revanche l'apparence phénotypique, comme la taille des feuilles ou la croissance). C'est pourquoi les marqueurs moléculaires sont des instruments de choix pour conduire une analyse de diversité intra-spécifique.

Bien que les marqueurs moléculaires ne soient pas, pour la plupart, directement reliés à des gènes fonctionnels, on peut s'attendre à ce qu'un haut niveau de diversité mesuré avec ces marqueurs indique aussi une forte diversité des gènes utiles. La section suivante présente une analyse spatiale de base effectuée sur des données de marqueurs moléculaires dans laquelle on utilise des microsatellites (SSRs en anglais), un marqueur co-dominant à l'utilisation répandue. Il faut cependant noter que tout type de données de marqueur moléculaire, comme le polymorphisme de longueur des fragments amplifiés (AFLP en anglais), peut être utilisé pour mener une analyse de diversité, pourvu que l'on sache identifier de manière unique chaque variation dans la composition de l'ADN.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION

Programmes :

- DIVA-GIS
- Excel

Fichiers de données :

- Répertoire 5.3 *Diversity - Molecular marker data* (Diversité – données de marqueurs moléculaires)
- SSR cherimoya rand column (shp, shx, dbf)
- Latin America Countries (shp, shx, dbf)

5.3.1. Comment réaliser une analyse spatiale de diversité à base de données de marqueurs moléculaires

Les données utilisées dans l'analyse suivante viennent du projet CHERLA, qui a réalisé une étude importante de la diversité de cherimoya dans sa région d'origine (pour plus d'information, voir <http://www.eelm.csic.es/proyecto/cherla/index.html>). Etant donné que les résultats finaux de cette étude n'ont pas encore été publiés, nous n'utiliserons qu'un sous-ensemble des données, avec des coordonnées modifiées de manière aléatoire pour recréer un scénario hypothétique.

Les principes de l'analyse spatiale de diversité à base de données de marqueurs moléculaires sont très proches de ceux d'une analyse au niveau spécifique. Dans cette analyse, nous utilisons des données de marqueurs microsatellites. Les SSRs sont des séquences dans lesquelles se répètent des paires de bases hautement variables et distribuées de manière uniforme dans l'ensemble du génome (De Vicente *et al.*, 2004a sur la base du travail de Hajeer, Worthington et John [eds] 2000). L'analyse de SSR s'intéresse aux différences de longueur des régions du microsatellite (habituellement pas associées aux gènes fonctionnels, c'est-à-dire neutres). Ces différences de longueur, que l'on désigne plus loin comme des allèles différents, sont les unités de diversité observées dans cette analyse.

Pour réaliser une analyse spatiale, les données de marqueurs moléculaires doivent être formatées de façon à ce que chaque allèle inclue une information de géoréférencement. Dans l'analyse suivante, chaque allèle doit être formaté comme suit : code du microsatellite + poids des paires de bases (par exemple : SSR1-293). Parcourez l'exemple *SSRs cherimoya rand column.dbf* pour vous familiariser avec le format en tableau sous Excel.

ID	Name	SSR1-293
1	SSR1-293	1218 5544-293
2	SSR1-293	1218 5544-293
3	SSR1-293	1218 5544-293
4	SSR1-293	1218 5544-293
5	SSR1-293	1218 5544-293
6	SSR1-293	1218 5544-293
7	SSR1-293	1218 5544-293
8	SSR1-293	1218 5544-293
9	SSR1-293	1218 5544-293
10	SSR1-293	1218 5544-293
11	SSR1-293	1218 5544-293
12	SSR1-293	1218 5544-293
13	SSR1-293	1218 5544-293
14	SSR1-293	1218 5544-293
15	SSR1-293	1218 5544-293
16	SSR1-293	1218 5544-293
17	SSR1-293	1218 5544-293
18	SSR1-293	1218 5544-293
19	SSR1-293	1218 5544-293
20	SSR1-293	1218 5544-293
21	SSR1-293	1218 5544-293
22	SSR1-293	1218 5544-293
23	SSR1-293	1218 5544-293
24	SSR1-293	1218 5544-293
25	SSR1-293	1218 5544-293
26	SSR1-293	1218 5544-293
27	SSR1-293	1218 5544-293
28	SSR1-293	1218 5544-293
29	SSR1-293	1218 5544-293
30	SSR1-293	1218 5544-293
31	SSR1-293	1218 5544-293
32	SSR1-293	1218 5544-293
33	SSR1-293	1218 5544-293
34	SSR1-293	1218 5544-293
35	SSR1-293	1218 5544-293

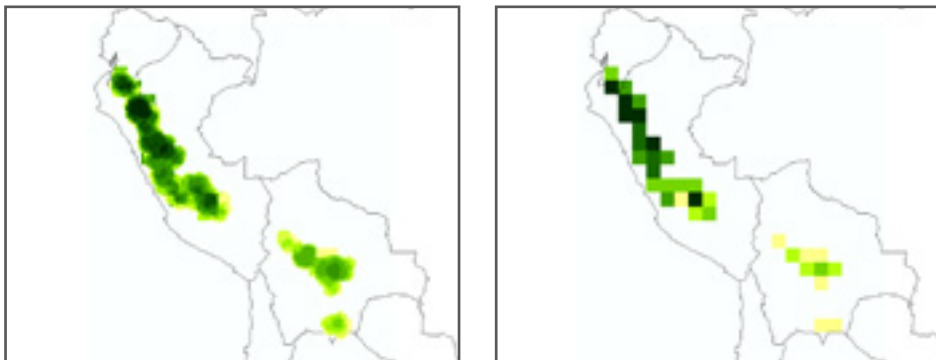
Dans cette section, vous apprendrez à :

- Utiliser l'option *Analysis (Point to Grid et Distance)* dans DIVA-GIS
- Utiliser l'option *Cluster* dans DIVA-GIS.

Travail personnel : Visualisez les sites de collecte de cherimoya inclus dans l'analyse de diversité (voir le Chapitre 3 – Éléments de base).



Travail personnel : Réalisez une analyse de richesse allélique, d'abord avec des cellules d'un (1) degré, puis avec des cellules de 10 minutes et un voisinage circulaire d'un (1) degré (voir Section 5.1). Regardez où se trouve le plus haut niveau de diversité.



Les rasters générés, montrés ci-dessus, indiquent que le Nord du Pérou est la région où l'on trouve le plus haut niveau de diversité (plus grand nombre d'allèles différents trouvés), tandis que la Bolivie se maintient au plus bas niveau.

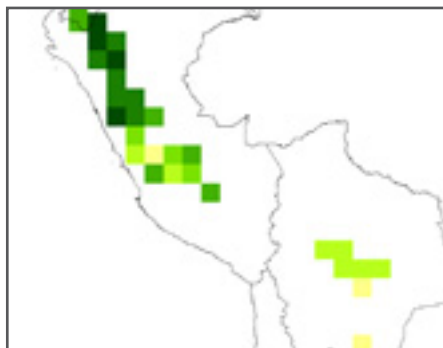
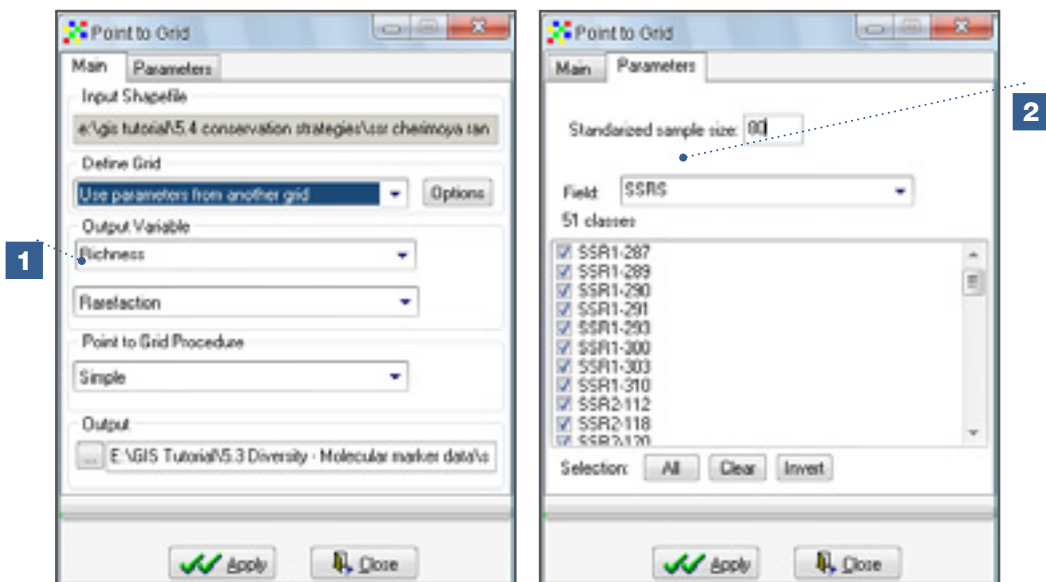
Méthode de raréfaction

Comme nous l'avons vu dans la première analyse (Analyse 5.1.1) une distribution inégale des observations peut avoir un impact significatif sur l'analyse de la richesse. La méthode de raréfaction a été développée pour comparer la richesse entre des cellules qui ont un nombre différent d'observations ou de spécimens (pour de plus amples informations, se référer à Petit *et al.* [1988] et Leberg [2002]). La méthode de raréfaction recalcule la richesse mesurée dans les différentes cellules comme si le même nombre standard d'observations avaient été effectuées dans chaque cellule. Seules les cellules qui contiennent un nombre d'observations

égal ou supérieur au nombre standard sont incluses dans l'analyse, celles qui en contiennent moins sont exclues. Le choix du nombre standard d'observations est un compromis entre le nombre de cellules incluses dans l'analyse et la richesse maximale à calculer. Par exemple, si l'on choisit un petit nombre d'observations, l'analyse inclura presque toutes les cellules ; mais la diversité maximale par cellule sera basse puisqu'elle ne peut pas dépasser le nombre d'échantillons. Inversement, si l'on choisit un nombre élevé d'observations, le nombre de cellules possédant au moins ce nombre d'observations (essentiel pour être incluses dans le calcul de raréfaction) sera réduit, limitant de ce fait le nombre de cellules pour lesquelles une valeur sera calculée. L'analyse suivante est fondée sur 8 microsatellites qui ont permis d'utiliser 80 observations d'allèles comme taille fixe d'échantillonnage (équivalent à 10 individus homozygotes ou 5 individus hétérozygotes).

Etapes :

1. Sélectionnez la couche qui contient les données moléculaires et allez à *Analysis/Point to Grid/Richness*. Dans le cadre *Output Variable*, sélectionnez *Rarefaction*. Utilisez le même fichier raster que dans l'analyse de richesse avec des cellules d'un (1) degré (voir Travail personnel).
2. Dans l'onglet *Parameters*, marquez les SSRs avec une *Standardized Sample Size* de 80.



Comme on peut l'observer, la différence de diversité entre le Nord du Pérou et la Bolivie (où une grande partie de l'échantillonnage a été effectué) devient encore plus évidente dans cette analyse.

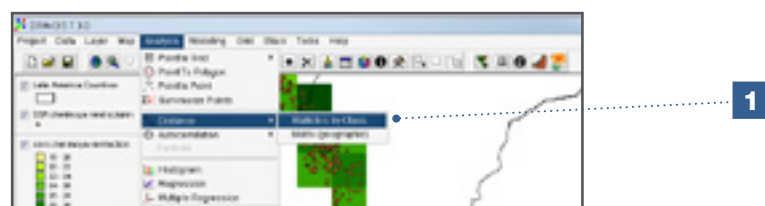
Distribution géographique des allèles individuels

L'analyse précédente s'est focalisée sur le nombre total d'allèles. Maintenant, nous allons nous intéresser à la distribution d'allèles spécifiques. A partir de leurs fréquences et de leurs distributions géographiques, on peut identifier différents types d'allèles. Parmi ceux-ci, les allèles communs localement sont les plus importants d'un point de vue conservation du fait qu'ils indiquent une adaptation aux conditions locales et que leur distribution limitée les rend plus vulnérables à la disparition que les allèles plus largement distribués. Pour une plus ample information sur les différents types d'allèles, vous pouvez vous référer à Frankel *et al.* (1995).

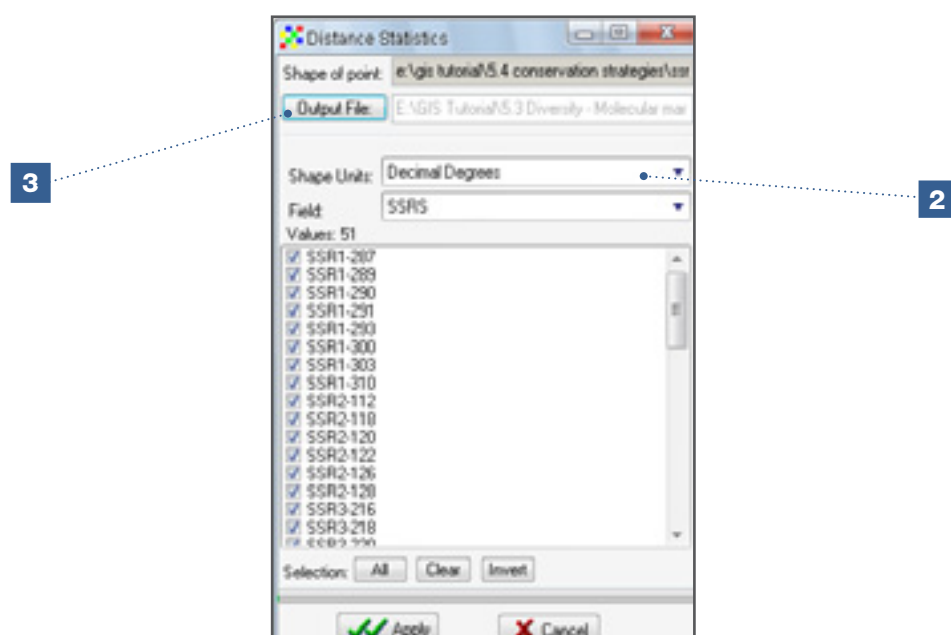
Fréquence	Distribution
Fréquent (5-10 %)	Local
Rare	Local
Fréquent	Répandu
Rare	Répandu

Etapes :

1. Sélectionnez la couche qui contient les données de marqueurs moléculaires. Allez à *Analysis/Distance/Statistics by Class*.



2. Sélectionnez les SSRs pour le calcul des distances géographiques.
3. Assignez un nom au fichier dBase IV (*.dbf) qui va être généré. Les résultats peuvent aussi être sauvegardés dans un fichier texte (*.txt).



SSRs	N	MAXD	MIND	AVGD
SSR1-290	1	0	0	0
SSR1-303	1	0	0	0
SSR3-300	1	0	0	0
SSR4-147	1	0	0	0
SSR6-150	2	90	0	45
SSR1-289	10	99	0	49.5
SSR1-287	2	112	0	56
SSR7-208	20	268	0	134
SSR2-126	8	536	0	272
SSR1-300	23	568	0	286
SSR5-163	63	847	0	423.5
SSR7-191	7	891	0	445.5
SSR5-182	35	962	0	481
SSR2-118	23	1004	0	502
SSR4-143	48	1090	0	545

Quand vous ouvrez le fichier dBase IV (*.dbf) ou le fichier texte (*.txt) sous Excel, un résumé du nombre de fois où chaque allèle est observé (N) et la distance maximale entre allèles (MAXD) seront affichés. Cette distance vous donne une idée des zones géographiques couvertes par chaque allèle. Les paramètres restants, distance minimale (MIND) et moyenne (AVGD), sont moins importants dans ce cas.

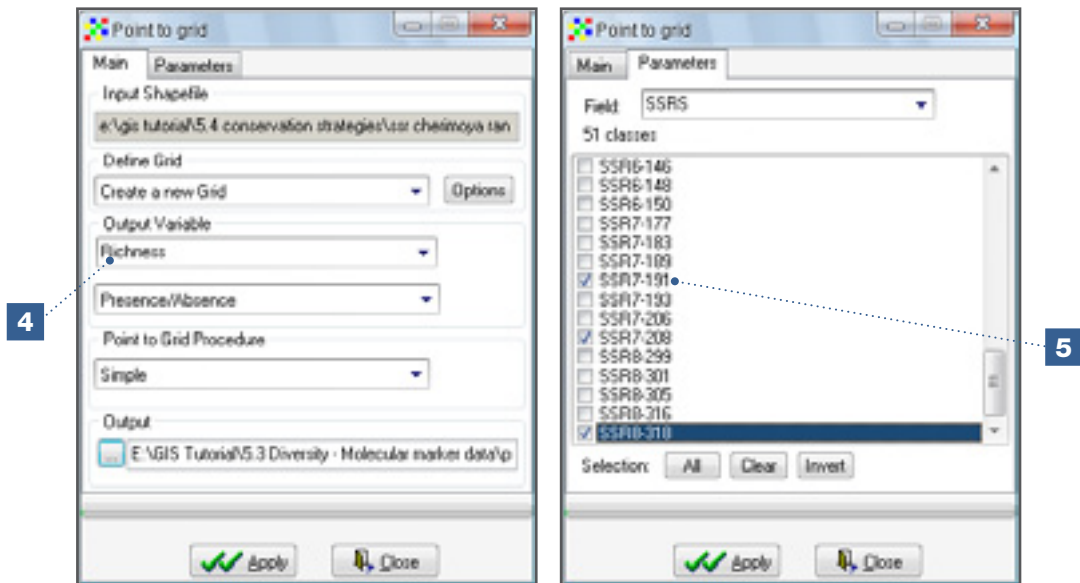
Pour visualiser l'information présentée sous Excel dans un format plus organisé, modifiez les distances de mètres à kilomètres, éliminez les décimales et triez les données par ordre croissant en se basant sur les valeurs de distance maximale (MaxD). Différentes classes d'allèles seront alors révélées, comme on peut l'observer dans le tableau ci-dessous, réduit à titre d'illustration.

SSRs	N	MaxD(km)
SSR1-290	1	0
SSR1-303	1	0
SSR3-300	1	0
SSR4-147	1	0
SSR6-150	2	90
SSR1-289	10	99
SSR1-287	2	112
SSR7-208	20	268
SSR2-126	8	536
SSR1-300	23	568
SSR5-163	63	847
SSR7-191	7	891
SSR5-182	35	962
SSR2-118	23	1004
SSR4-143	48	1090

SSR5-166	81	1177
SSR1-291	89	1182
SSR3-220	31	1192
SSR7-177	53	1222
SSR4-145	67	1234
SSR7-206	43	1251
SSR5-146	101	1294
SSR5-164	114	1337
SSR8-318	5	1646
SSR7-193	89	2158
SSR4-156	55	2226
<i>tableau incomplet</i>		

- Le premier groupe inclut les allèles qui n'ont été observés qu'une fois (surlignés en jaune). Ces allèles sont considérés comme uniques bien que le nombre de leur occurrence dépende largement du nombre d'échantillons pris dans une zone donnée. Comme l'intensité de l'échantillonnage a une forte influence sur l'identification d'allèles uniques, ce type d'allèle est considéré comme moins intéressant que ceux du groupe suivant.
 - Les allèles les plus intéressants pour identifier les zones de diversité élevée sont ceux que l'on observe de manière répétée dans une petite zone : par exemple, les allèles SSR1-289 et SSR7-208 (surlignés en vert). Ils sont ce que l'on appelle des allèles communs au niveau local. La probabilité que ce phénomène apparaisse de manière répétée et dans une zone limitée est beaucoup plus basse que la probabilité d'observer un allèle déterminé une fois seulement, étant donné les différences dans la stratégie d'échantillonnage. Les populations qui contiennent un nombre significatif d'allèles communs localement peuvent être placées en priorité dans les mesures de conservation du fait qu'elles contiennent des allèles (et très probablement aussi des gènes utiles) que l'on ne trouve pas ailleurs et leur observation dépend moins des différences dans l'intensité de l'échantillonnage.
 - Un autre groupe d'allèles est constitué par ceux que l'on trouve dans une zone relativement grande mais avec une fréquence faible (par exemple SSR7-191 et SSR8-318, surlignés en orange). Ces observations peuvent dénoter un échantillonnage incomplet.
 - Enfin, la majorité des allèles se trouvent couramment dans toute la zone d'étude. On les appelle allèles communs, et ils ne contribuent guère à l'analyse de diversité.
4. Regardez de plus près la distribution des allèles d'intérêt : SSR1-289, SSR7-191, SSR7-208 et SSR8-318. La première étape consiste à sélectionner la couche des SSRs dans *Analysis/Point to Grid/Richness*. Cette fois-ci, sélectionnez *Presence/Absence* comme *Output Variable*. Utilisez des rasters avec des cellules d'un (1) degré.

5. Dans l'onglet *Point to Grid*, sélectionnez uniquement les SSRs d'intérêt (voir l'Etape 4).



Contrairement à la plupart des analyses réalisées avec DIVA-GIS, les résultats de l'analyse de présence et d'absence (*Presence/Absence Analysis*) ne sont pas affichés automatiquement. Pour visualiser les résultats, ouvrez la couche et allez dans le sous-répertoire qui porte le nom assigné au fichier. Les résultats seront affichés de la manière suivante :

Cas N° 1 :

Allèles communs dans une zone réduite



SSR1-289
(10 observations, Distance max. 99 km)

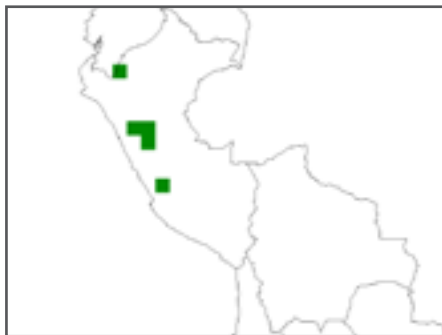


SSR7-208
(20 observations, Distance max. 268 km)

Si on compare ces deux cartes avec la carte de diversité générée dans l'analyse précédente, on s'aperçoit que les allèles communs au niveau local se situent dans les zones qui présentent une richesse allélique plus élevée, ce qui confirme ce qui a été observé plus haut, à savoir que le Nord du Pérou est une zone de haute diversité pour le cherimoya.

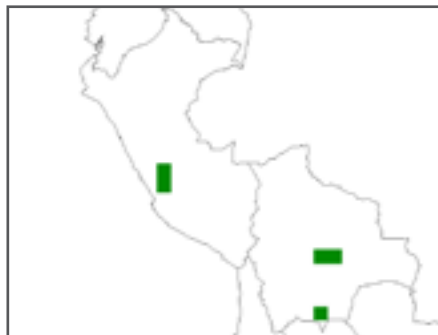
Cas N° 2 :

Allèles rares (SSR7-191 et SSR8-318) ayant une large distribution



SSR7-191

(7 observations, Distance max. 891 km)



SSR8-318

(5 observations, Distance max. 1646 km)

L'allèle SSR7-191 se trouve dans différentes zones du Nord du Pérou mais on ne le trouve pas dans la zone de diversité la plus élevée (où la présence d'allèles uniques indique un échantillonnage assez complet), ce qui suggère que cette situation ne résulte pas d'un échantillonnage incomplet mais est attribuable aux lois de la probabilité, qui font que (bien que la probabilité soit faible), avec un plan d'échantillonnage aléatoire comportant suffisamment d'observations, cette situation va se produire (un peu comme lorsqu'on tire à « pile ou face » plusieurs fois et que la pièce ne tombe jamais sur « face »). Une autre explication de l'occurrence atypique de cet allèle peut être qu'il y a eu une adaptation locale, étant donné que l'allèle SSR7-191 se trouve dans la même zone que l'allèle SSR7-208 (voir la carte ci-dessus), et est donc commun dans le centre de diversité. La présence dominante de l'allèle SSR7-208, qui remplace l'allèle SSR7-191, peut indiquer un processus d'adaptation locale. Néanmoins, comme l'analyse est fondée sur des marqueurs neutres (qui, selon toute vraisemblance, n'interviennent pas dans l'expression génétique), une telle explication devrait être donnée avec prudence et il faut poursuivre les recherches pour comprendre la distribution particulière des allèles dans cet endroit.

Le cas de l'allèle SSR8-318 est unique. La zone où l'on trouve cet allèle recouvre le centre du Pérou et la Bolivie ; de ce fait, il y a une forte probabilité que l'allèle soit aussi présent dans l'espace entre ces deux zones, c'est-à-dire le Sud du Pérou, une zone n'a pas été prospectée. La présence dans cette zone signifierait que cet allèle est plus commun que la carte ne l'indique. Tel quel, le résultat suggère que l'échantillonnage a été incomplet. Il indique aussi qu'il pourrait y avoir de nouveaux allèles en dehors de la zone de diversité la plus élevée (le Nord du Pérou). Comme indiqué dans la Section 5.4, ces considérations sont importantes et utiles lorsque l'on formule des stratégies de conservation.

Analyse de groupes (*clusters*)

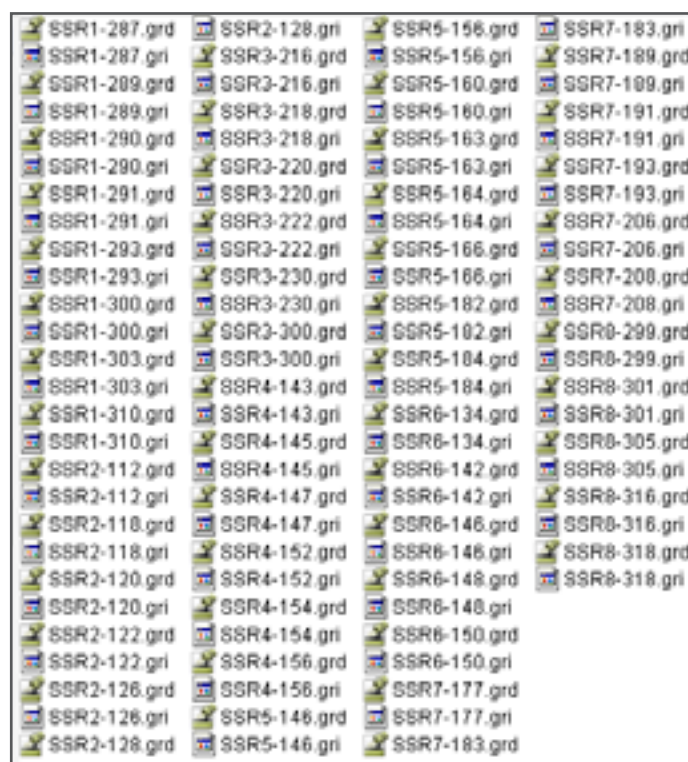
Les analyses de richesse réalisées jusqu'ici prennent en compte le nombre d'allèles dans une cellule de raster (diversité *alpha*), mais elles ignorent les variations de composition parmi les différentes cellules (diversité *beta*). Si deux cellules peuvent avoir une richesse similaire, elles peuvent aussi afficher une composition en allèles complètement différente. Comment la diversité peut-elle être analysée dans ce cas ? Le programme DIVA-GIS inclut un outil d'analyse de groupes pour évaluer les différences de diversité entre deux cellules de raster. Dans l'analyse présentée ci-dessous, cet outil est mis en œuvre pour analyser les différences de composition allélique apportant ainsi plus de profondeur à la connaissance de la structure génétique existant au sein de la distribution géographique de l'espèce.

De nombreux programmes informatiques ont été développés pour analyser la composition allélique de différentes populations. L'un d'entre eux, *Structure* (<http://pritch.bsd.uchicago.edu/structure.html>) assigne des génotypes à des groupes sur la seule base des fréquences

alléliques, indépendamment de l'unité géographique dans la zone d'étude (cellules raster) dans laquelle ils se trouvent ou de toute autre définition *a priori* des populations. Pour plus d'information sur *Structure*, vous pouvez vous référer à Pritchard *et al.* (2000). L'analyse suivante explique comment entreprendre une analyse de groupes dans DIVA-GIS en utilisant les allèles comme unité observée de diversité.

Etapes :

1. Effectuer une cartographie de différents allèles est la base de toute analyse spatiale de la composition allélique. Répétez le processus de cartographie conduit aux Etapes 4 et 5 de la section précédente (Distribution géographique d'allèles individuels – *Analysis/Point to Grid/Richness* avec l'option *Presence/Absence*) sans sélectionner d'allèle spécifique. Créez un raster qui montre la distribution de chaque allèle, en utilisant des cellules d'un (1) degré. Le résultat se trouve dans le répertoire nouvellement créé, qui contient les rasters de distribution pour chacun des 51 allèles.



Quelques exemples :

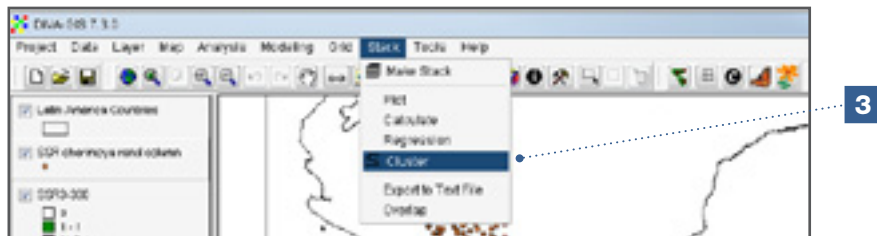


Un allèle commun : SSR1-310

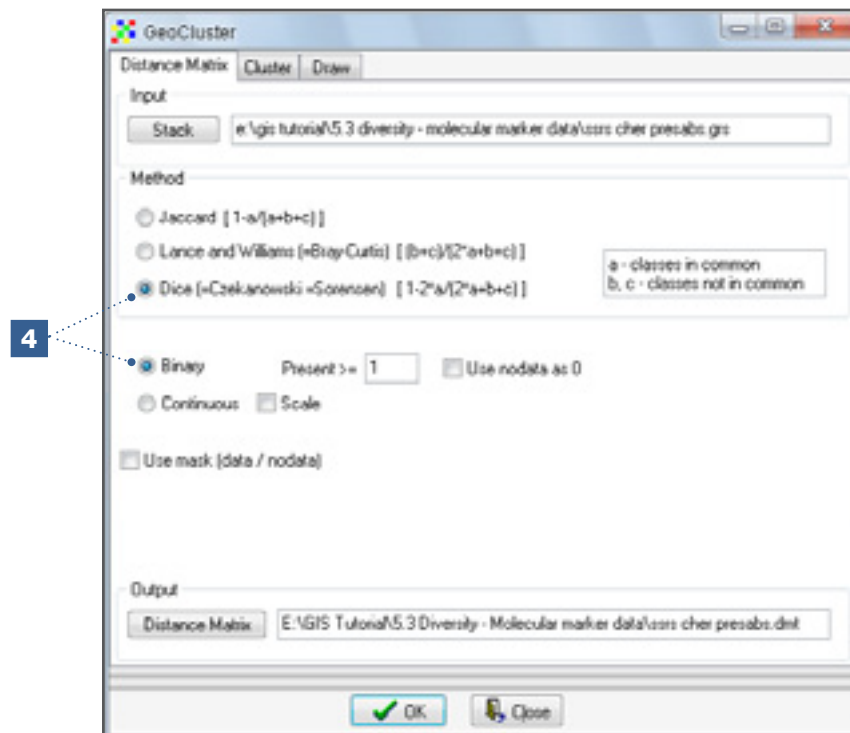


Un allèle rare : SSR3-300

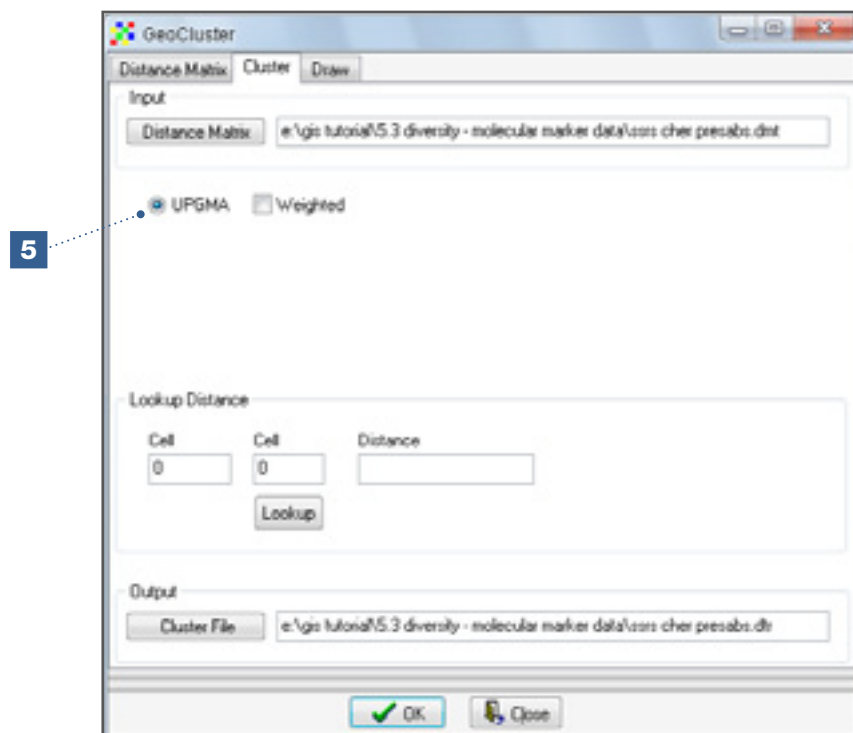
2. Avant de lancer l'analyse de cluster, constituez un regroupement de rasters de présence/absence pour les 51 allèles générés à l'étape précédente (voir l'Analyse 5.2.1, Etape 8). Ce fichier constituera un *stack* (regroupement de rasters ayant les mêmes propriétés) et fournira la base de l'analyse suivante.
3. Allez à *Stack/Cluster* et sélectionnez le *stack* que vous venez de créer.



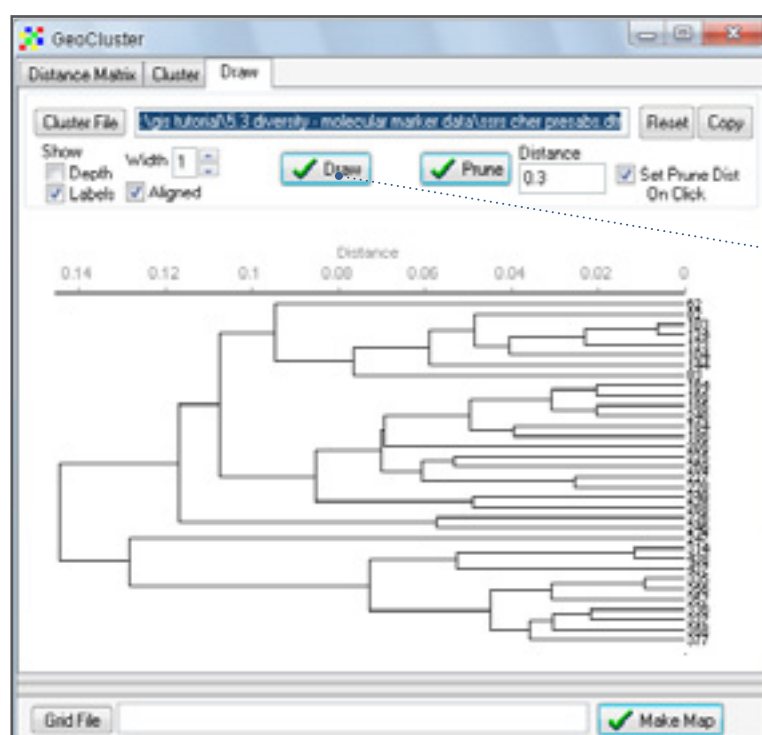
4. La prochaine étape consiste à calculer la distance entre différentes cellules de raster en se fondant sur la composition allélique trouvée dans chaque cellule. Comme les données utilisées sont binaires [valeurs de un (1) pour la présence d'un allèle et de zéro (0) pour l'absence d'allèles], laissez sélectionnée l'option par défaut (*Binary*). Trois méthodes sont couramment utilisées pour calculer les distances. Dans cette analyse, la méthode *Dice* sera utilisée pour calculer la matrice des distances. Elle est aussi connue sous le nom de *Nei-Li coefficient* et peut être utilisée avec des données de marqueurs co-dominants comme celles des marqueurs microsatellites (voir De Vicente *et al.*, 2004b). Utilisez le même nom de fichier que celui assigné au fichier du *stack* et cliquez sur *OK* pour continuer.



5. L'étape suivante consiste à réaliser une analyse de groupes fondée sur le regroupement des cellules séparées par les plus petites distances (c'est-à-dire ayant la composition allélique la plus proche). Utilisez la méthode *UPGMA* (méthode par défaut) pour grouper les cellules. Cliquez sur *OK* pour continuer (là encore, il est recommandé d'utiliser le même nom de fichier).



6. Dans l'onglet *Draw*, cliquez sur le bouton *Draw*. Un dendrogramme (arbre) sera affiché avec les cellules similaires, c'est-à-dire celles qui présentent une composition allélique similaire.

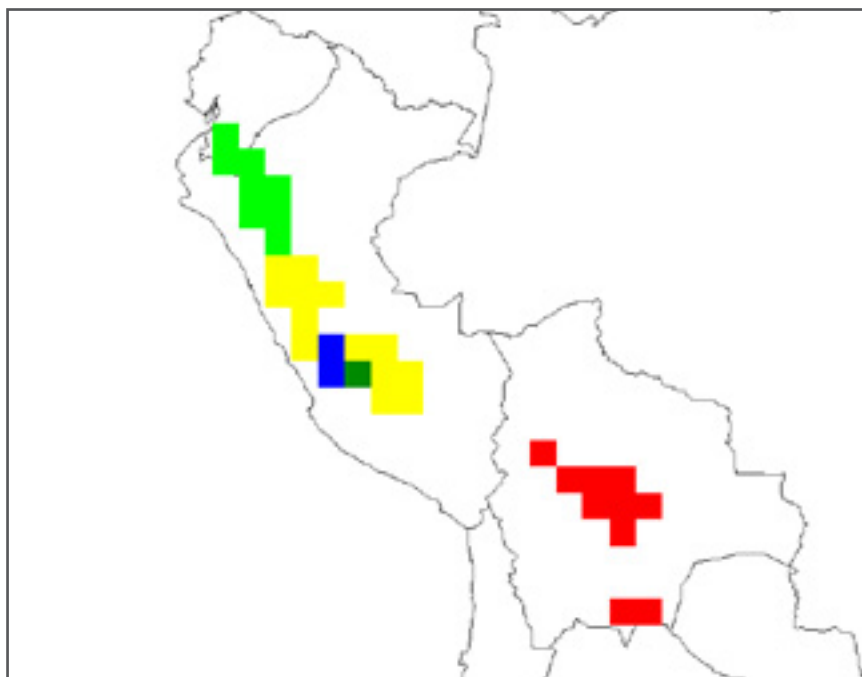


7. Les cellules peuvent être groupées selon leur niveau de similarité (indiqué par des distances courtes). On peut faire cela en élaguant le dendrogramme. Cliquez sur le dendrogramme à la distance que vous voulez utiliser pour grouper les cellules, puis cliquez sur le bouton *Prune* (par exemple, à une distance de 0,1). Vous pouvez aussi saisir la valeur de distance désirée pour le regroupement des cellules, dans la fenêtre *Distance*.



8. Cliquez sur le bouton *Prune* pour couper l'arbre à la distance sélectionnée. Le regroupement de cellules similaires résultant sera affiché sous forme d'un arbre portant moins de branches (le dendrogramme complet peut être réaffiché à l'aide du bouton *Draw*). Lorsqu'on élague avec une distance de 0,1, on obtient cinq groupes.
9. Pour visualiser le résultat, assignez un nom au raster dans la fenêtre *Grid File*, puis cliquez sur *Make Map*.





Après regroupement des cellules similaires en cinq groupes différents, on observe trois grandes zones, dont chacune a une composition allélique différente ; la première couvre le Nord du Pérou et le Sud de l'Equateur, la seconde le centre du Pérou et la troisième se situe en Bolivie.

On peut poursuivre l'exploration de la structure génétique de la distribution du cherimoya en utilisant d'autres points d'élague mais il faut savoir que les différences peuvent ne plus être évidentes lorsque le nombre de groupes se réduit. Par exemple, si l'on élague avec une distance de 0,12, il ne reste plus que trois groupes.



La séparation entre la plupart des populations du Pérou et de l'Equateur a disparu du fait de l'élitage, ce qui confirme que les cherimoya observés en Bolivie sont différents de ceux trouvés en Equateur et au Pérou. Bien que l'analyse précédente ait fait apparaître que la Bolivie présente de faibles niveaux de diversité, la composition allélique des individus observés dans le pays est bien distincte de celle que l'on trouve dans les autres zones étudiées.

10. Lorsqu'il y a trop de groupes, il est aussi difficile d'identifier des structures dans la composition allélique. L'élitage du dendrogramme à une distance de 0,085 génère 11 groupes d'allèles.



Le fait d'avoir trop de groupes, comme l'illustre la carte ci-dessus, complique l'interprétation des résultats. Malgré le nombre élevé de groupes, les individus de Bolivie constituent toujours un groupe homogène, distinct des autres.

5.4. Implications pour la formulation de stratégies de conservation

Les trois analyses précédentes se sont focalisées sur la mise en œuvre d'analyses spatiales pour détecter des zones de haute diversité (diversité *alpha*) et, à un moindre niveau, pour comprendre les différences de diversité entre les zones (diversité *beta*). Comprendre l'extension et la distribution de la diversité est cruciale pour concevoir des stratégies de conservation efficaces et appropriées. Cette compréhension est aussi vitale pour identifier les sites clés où l'on peut conduire des activités de conservation *ex situ* et *in situ* (zones prioritaires pour la collecte et la protection), surtout du fait que, la plupart du temps, les ressources allouées à la conservation sont modestes, ce qui limite la possibilité de conduire des opérations de conservation dans de nombreuses zones. En combinant des analyses de diversité *alpha* et *beta*, on peut optimiser l'allocation des ressources afin de conserver le maximum de diversité possible.

Par exemple, une zone présentant la diversité *alpha* la plus élevée est habituellement choisie en priorité comme premier site de conservation *in situ* (si l'on fait abstraction des contraintes logistiques) mais le choix de la zone de seconde priorité est une question plus complexe à résoudre, dans le cas où il reste des ressources disponibles. La zone de seconde priorité ne doit pas nécessairement être celle qui présente le deuxième plus haut niveau de diversité *alpha*, car il se peut qu'une grande part de cette diversité soit déjà conservée dans la première zone. Dans ce cas, il faut prendre en compte la diversité *beta* et se focaliser sur les zones où l'espèce ou la composition allélique est différente de celle trouvée dans la première zone.

Cette notion de complémentarité est prise en compte dans DIVA-GIS. Les activités de conservation cherchent généralement à traiter le plus grand nombre possible d'espèces, mais elles peuvent aussi cibler la conservation d'une espèce particulière. Dans ce cas, les allèles sont utilisés comme unité de diversité observée pour la définition des zones de conservation *in situ* des gènes. Ce concept est présentée dans l'analyse suivante consacrée au cherimoya.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel 	Fichiers de données : <p>Répertoire 5.4 <i>Conservation Strategies</i> (Stratégies de conservation)</p> <ul style="list-style-type: none"> • <i>SSR cherimoya rand column (shp, shx, dbf)</i> • <i>Latin America Countries (shp, shx, dbf)</i> • <i>Protected Areas Latin America (shp, shx, dbf)</i>

5.4.1. Comment identifier les zones prioritaires pour la conservation *in situ* ou pour la collecte de matériel génétique

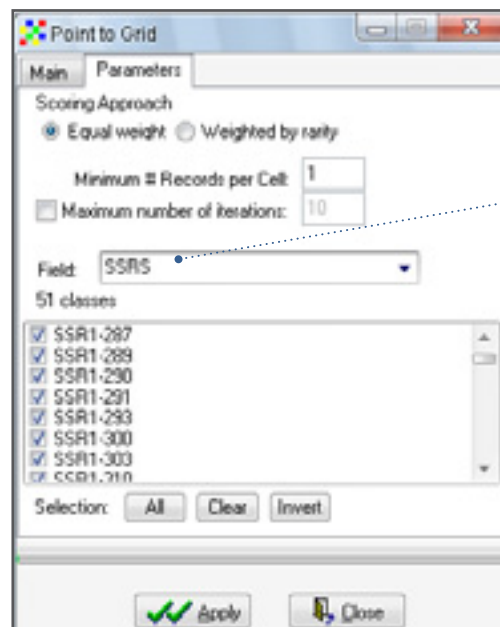
Avec cette analyse, vous apprendrez à utiliser l'option *Analysis (Point to Grid et Reserve Selection)* de DIVA-GIS pour aider à définir des zones de conservation prioritaires. La procédure *Reserve Selection* utilise un algorithme d'optimisation qui a été développé à l'origine pour minimiser la surface nécessaire pour conserver la diversité des plantes à fleurs en Afrique du Sud (voir Rebelo et Siegfried, 1992). Dans cette analyse, nous utiliserons cet algorithme pour définir le nombre minimum d'unités géographiques nécessaires pour conserver toute la diversité génétique (mesurée au moyen de marqueurs moléculaires) et pour identifier, par ordre d'importance, les unités géographiques prioritaires pour la conservation. Les données utilisées dans l'analyse précédente le seront ici aussi. Le même type d'analyse peut être conduit également au niveau espèces.

Etapes :

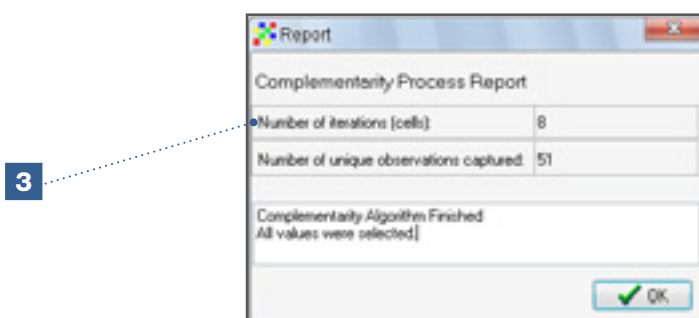
1. Récupérez les données de marqueurs moléculaires de cherimoya pour débiter l'analyse et sélectionnez l'option *Analysis/Point to Grid*. Pour cette analyse, sélectionnez l'option *Reserve Selection*, ainsi que l'option *Complementarity*. Utilisez un raster avec des cellules d'un (1) degré (de préférence l'un des rasters utilisés dans les analyses précédentes).



2. Chaque allèle doit recevoir le même poids (option par défaut), mais vous pouvez aussi donner plus de poids aux allèles ou espèces rares, si nécessaire.

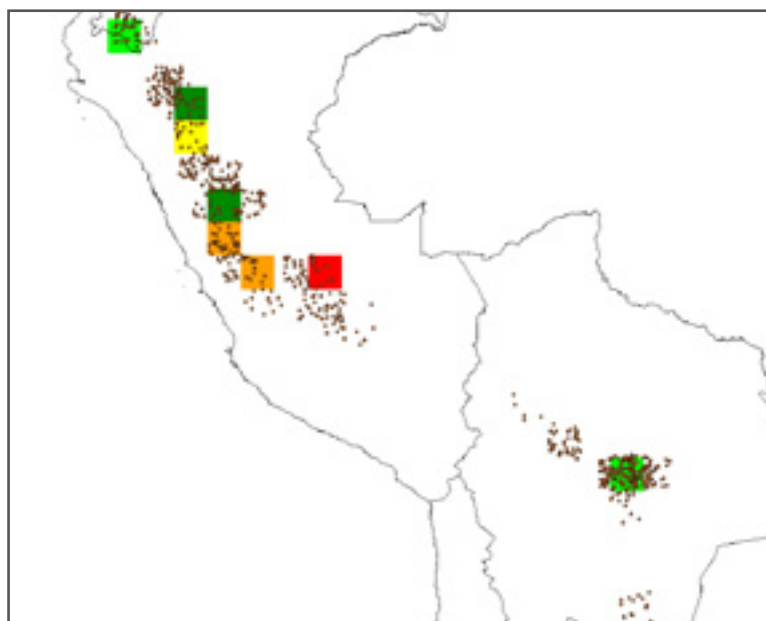


3. Le premier résultat affiché est un rapport sur le nombre de cellules qu'il faut sélectionner (huit) pour conserver tous les allèles et le nombre d'observations uniques (les 51 allèles) générées par l'analyse.



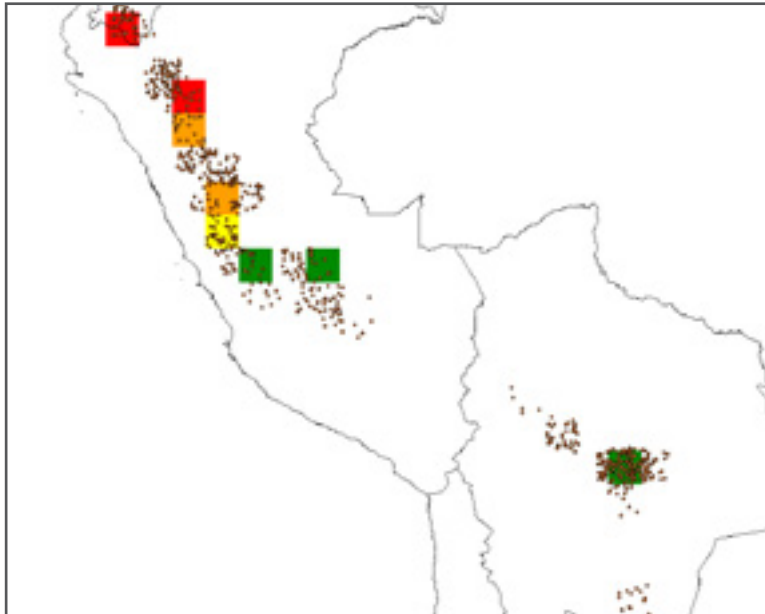
4. Ensuite, cliquez sur OK dans la fenêtre *Report*. Trois rasters intitulés *Sequence*, *Classes* et *Additional Classes* seront affichés. Avant de visualiser chaque résultat, améliorez la légende à l'aide de l'option *NoData transparent* pour élargir le nombre de classes.

▪ *Séquence*



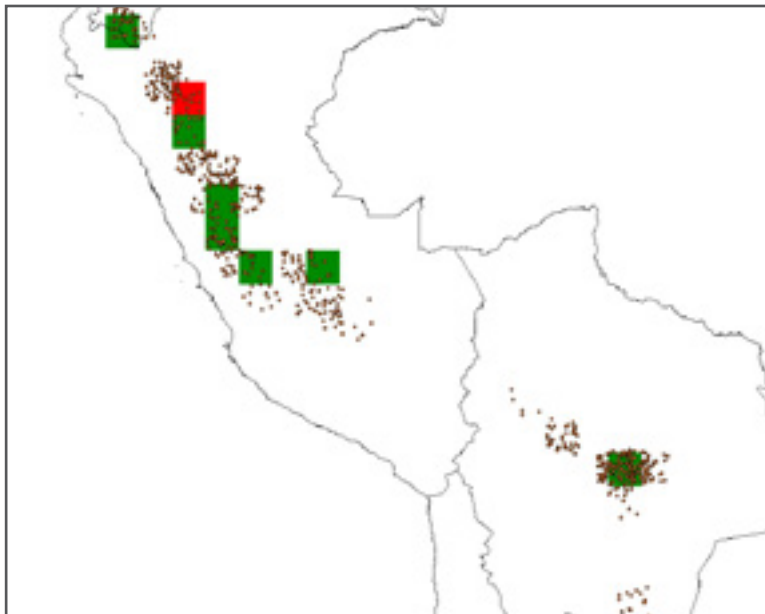
Ce raster correspond au résultat le plus important : la séquence de sélection de cellules. En d'autres termes, le raster indique les cellules prioritaires pour la conservation (la cellule avec une valeur d'un [1] est la plus importante). En sélectionnant la couche et en plaçant le curseur sur chaque cellule, on peut voir les valeurs dans la partie basse de la carte, dans la barre d'état. Ici, les cellules prioritaires sont, par ordre d'importance : 1) la cellule située dans le Nord du Pérou ; 2) la cellule située un peu plus au Sud ; 3) la cellule située en Bolivie et 4) la cellule située dans le Sud de l'Equateur. Cette analyse souligne le fait que les cellules n'ont pas été sélectionnées sur la seule base de la plus haute diversité, mais aussi sur la base des différences de composition allélique (comme l'illustre l'inclusion d'une cellule de Bolivie, où la diversité n'est pas très élevée).

- *Classes*



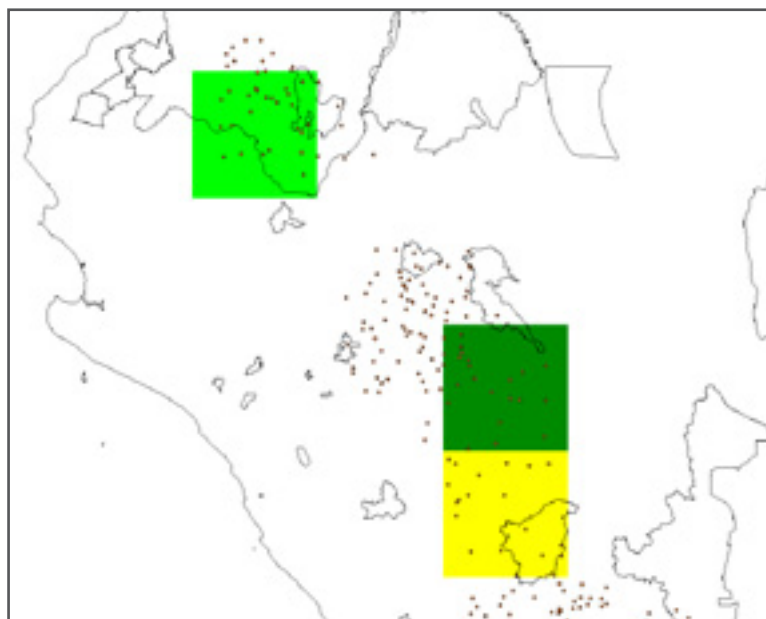
Ce raster correspond à une analyse de richesse, il ne montre que l'information (richesse) des cellules sélectionnées dans l'analyse précédente. A l'aide du curseur, on peut vérifier que la cellule sélectionnée comme ayant la plus haute priorité correspond à une cellule ayant le plus haut niveau de diversité dans le groupe (39 allèles). Note : dans cette analyse, la cellule de seconde priorité n'est pas celle qui présente le deuxième plus haut niveau de diversité.

- *Classes supplémentaires*



Ce raster révèle le nombre d'allèles apportés par chaque cellule supplémentaire, il prend en compte le fait que toutes les cellules sélectionnées couvrent les 51 allèles. La première cellule apporte le plus grand nombre d'allèles (39 allèles).

5. Pour achever l'analyse, ajoutez la couche des aires protégées (en utilisant le fichier *Protected Areas Latin America.shp*) et analysez le statut actuel de conservation des cellules prioritaires en vous fondant sur la couche *Sequence*.



La dernière étape révèle qu'il y a des aires protégées dans chacune des cellules, ce qui donne la première indication sur la conservation *in situ* du cherimoya (dans une situation réelle, il faudrait des analyses plus détaillées). Les résultats de l'analyse semblent indiquer que la diversité du cherimoya est actuellement partiellement conservée puisqu'on trouve des aires protégées dans toutes les cellules prioritaires. Mais la cellule de plus haute priorité a une mauvaise couverture en aires protégées. Une analyse à plus petite échelle permettrait d'affiner les conclusions en indiquant si les accessions de cherimoya observées qui présentent les allèles les plus importants sont effectivement inclus dans ces aires protégées.

Références

- Chapman AD 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. [en ligne] Disponible à l'adresse : http://www.gbif.org/orc/?doc_id=1229. Lien vérifié en Juin 2012.
- De Vicente MC, Lopez C, Fulton T. 2004a. Molecular marker learning module volume 1: Using Molecular Marker Technology in : Studies on Plant Genetic Diversity: Learning module. IPGRI et Cornell University.
- De Vicente MC, Lopez C, Fulton T. 2004b. Molecular marker learning module volume 2: Genetic diversity analysis with molecular marker data: Learning module. IPGRI et Cornell University.
- Frankel OH, Brown AHD, Burdon JJ. 1995. The conservation of plant biodiversity. Cambridge University Press, Cambridge, Royaume Uni.
- Grum M, Atieno F. 2007. Statistical analysis for plant genetic resources: clustering and indices

- in R made simple. Handbooks for Genebanks No. 9. Bioversity International, Rome, Italie.
- Hajeer A, Worthington J, John S, editors. 2000. SNP and microsatellite genotyping: Markers for genetic analysis. In: Biotechniques: Molecular laboratory methods series. Eaton Publishing, Manchester, Royaume Uni.
- Hijmans RJ, Garrett KA, Huamán Z, Zhang DP, Schreuder M, Bonierbale M. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology* 14(6):1755-1765.
- Kindt R, Coe R. 2005. Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi.
- Leberg PL. 2002. Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology* 11:2445-2449.
- Mathur PN, Muralidharan K, Parthasarathy VA, Batugal P, Bonnot F. 2008. Data Analysis Manual for Coconut Researchers. Bioversity Technical Bulletin No. 14. Bioversity International, Rome, Italie.
- Petit RJ, El Mousadik A, Pons O. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12:844-855.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959. [en ligne] Disponible à l'adresse : <http://www.genetics.org/content/155/2/945.full.pdf>. Lien vérifié en Juin 2012.
- Rebello AG, Sigfried WG. 1992. Where should nature reserves be located in the Cape Floristic Region, South Africa? Models for the spatial configuration of a reserve network aimed at maximizing the protection of diversity. *Conservation Biology* 6(2):243-252.
- Scheldeman X, Willems L, Coppens D'eeckenbrugge G, Romeijn-Peeters E, Restrepo MT, Romero Motoche J, Jimenez D, Lobo M, Medina CI, Reyes C, Rodriguez D, Ocampo JA, Van Damme P, Goetghebeur P. 2007. Distribution, diversity and environmental adaptation of highland papaya (*Vasconcellea* spp.) in tropical and subtropical America. *Biodiversity and Conservation* 16(6):1867-1884.
- Willems L, Scheldeman X, Soto Cabellos V, Salazar SR, Guarino L. 2007. Spatial patterns of diversity and genetic erosion of traditional cassava (*Manihot esculenta* Crantz) cultivation in the Peruvian Amazon: an evaluation of socio-economic and environmental indicators. *Genetic Resources and Crop Evolution* 54(7):1599-1612.

Chapitre 6

Modélisation et analyse de la distribution d'espèces

La niche écologique est un concept fréquemment utilisé dans les études de conservation de la biodiversité. Ce concept a été appliqué avec succès pour définir les sites prioritaires pour la conservation *in situ* d'espèces sauvages et dans la planification des missions de collecte de ressources génétiques liées aux plantes cultivées (par exemple, Jarvis *et al.*, 2005 ; Scheldeman *et al.*, 2007).

Si différentes définitions de la niche écologique ont été formulées, le concept se réfère essentiellement à l'espace environnemental qu'occupe une espèce dans des conditions naturelles (Pullam, 2000). La définition couramment acceptée, fournie par Hutchinson (1957), fait une distinction entre une niche fondamentale et une niche effective. Une niche fondamentale est la plage des conditions environnementales dans lesquelles une espèce peut théoriquement exister, tandis qu'une niche effective est définie par la combinaison des interactions négatives (comme la compétition et la prédation) qui restreignent la présence d'une espèce et des interactions positives (comme la facilitation) qui peuvent étendre la plage environnementale dans laquelle une espèce est capable de se développer.

Les systèmes d'information géographique comme DIVA-GIS incluent la capacité de modélisation des niches écologiques à partir des données environnementales disponibles sur les sites où l'espèce a été observée (points de présence). Il existe déjà des bases de données qui fournissent des informations détaillées sur le climat (fondées sur l'interpolation de données collectées dans le monde entier par des stations de mesure climatiques) comme WorldClim (Hijmans *et al.*, 2005), mais la disponibilité de données concernant d'autres facteurs environnementaux pertinents, comme les variables édaphiques, est encore limitée. De ce fait, de nombreux outils de SIG fournissent une valeur approximative de la niche écologique en utilisant des variables climatiques connues sous le nom d'« enveloppe climatique » (Guarino *et al.*, 2002).

Les analyses de niche écologique d'espèces sauvages et de leurs ressources génétiques s'appliquent à l'aide de points de présence sous différentes conditions, dont les plus importantes sont les suivantes :

6. L'espèce devrait être en équilibre avec son environnement ; en d'autres termes, les plages environnementales sont restreintes par la compétition et la prédation, et non par des limitations de dispersion.
7. Les variables environnementales disponibles (comme les variables climatiques) utilisées dans la modélisation sont des facteurs abiotiques déterminants de la distribution naturelle de l'espèce.
8. On ne doit pas prendre en compte les points de présence d'individus cultivés dans des plantations, ou issus de collecte de terrain ou de jardins botaniques situés dans un environnement étranger à la niche effective de l'espèce.

En pratique, l'une ou plusieurs des conditions ci-dessus ne sont souvent pas satisfaites ; mais la modélisation de la distribution d'espèces n'en reste pas moins utilisable pour approcher la niche effective et la distribution naturelle d'une espèce. En tant que telle, la modélisation de la distribution d'espèces est utile pour établir les priorités dans les activités de conservation.

Le concept de niche écologique peut aussi être utilisé pour identifier les zones agro-écologiques idéales pour faire pousser des espèces données d'arbres ou de plantes cultivées.

Ecocrop (FAO, 2007), qui est inclus dans DIVA-GIS, de même que Homologue (Jones *et al.*, 2005) sont des exemples de programmes de modélisation que l'on peut utiliser pour réaliser de telles analyses. L'identification de zones de production optimales est une tâche plus complexe que l'identification de l'aire naturelle de distribution, du fait de l'interaction dynamique entre les pratiques culturales et les facteurs environnementaux. Certaines pratiques culturales permettent de cultiver une espèce dans un environnement situé en dehors de sa niche effective, voire même de sa niche fondamentale, si des ressources additionnelles sont disponibles, y compris irrigation et fertilisation.

Enfin, il importe de ne pas ignorer la relation entre la variation génétique et l'environnement. Des populations d'espèces sauvages ou de plantes cultivées sont capables d'évoluer localement et de s'adapter ainsi aux environnements de sites spécifiques. Cela se traduit par des différences d'expression phénotypique entre individus de populations différentes, même lorsqu'ils sont plantés ensemble sur un même site. Des expériences de type Génotypes x Environnement (GxE) permettent d'identifier les écotypes et les variétés les mieux appropriés pour une zone agroécologique donnée. Ce type d'analyse est pertinent pour sélectionner le matériel génétique prometteur adapté à des aires spécifiques, mais cette application ne sera pas présentée ici car elle dépasse le cadre de ce manuel.

Ce chapitre se consacre à l'utilisation de la modélisation pour prédire la distribution d'espèces et à la manière dont ses résultats peuvent contribuer à définir les sites prioritaires pour la conservation, les études d'impact du changement climatique et la collecte de matériel génétique.

6.1. Analyse de la niche effective d'une espèce

Comme cela a été mentionné dans l'introduction à ce chapitre, plusieurs logiciels de SIG, y compris DIVA-GIS, incluent des modèles simplifiés de distribution d'espèces fondés sur les données climatiques relatives aux points de présence d'individus ou de groupes d'individus d'une espèce.

Si l'on ne dispose que d'un nombre limité de points de présence pour une espèce donnée, il se peut que les données environnementales correspondantes ne représentent pas sa niche effective de manière adéquate. Ceci peut conduire à une interprétation très erronée des résultats. C'est pourquoi il importe de s'assurer que les points de présence utilisés dans une modélisation sont de bonne qualité (voir le Chapitre 4) et en quantité suffisante pour obtenir un modèle fiable de la distribution de l'espèce. Il n'y a pas de norme quant au nombre minimal de points requis, car cela dépend souvent de la nature de l'espèce. Pour des espèces rares ou qui ont une niche restreinte, il se peut que l'on ne dispose que d'un petit nombre de points de présence. Mais dans ces cas-là, même un nombre de points réduit peut être hautement représentatif de la niche. Pour ces raisons, le nombre minimal de points requis pour réaliser une modélisation crédible de la distribution d'une espèce ne peut pas faire l'objet de directives strictes. Mais il existe quelques exemples de nombres de points utilisés pour des espèces données : a) des études conduites par Scheldeman *et al.* (2007) ont utilisé un minimum de 10 points pour des espèces rares de *Vasconcellea* avec une distribution restreinte ; b) le projet MAPFORGEN – qui a évalué la distribution naturelle de 100 espèces originaires d'Amérique latine – a utilisé un minimum de 20 points de présence ; et c) Van Zonneveld *et al.* (2009a) ont travaillé avec un minimum de 50 points de présence pour deux espèces de pin ayant une large distribution géographique à travers l'Asie du Sud-est.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Excel 	Fichiers de données : <p>Répertoire 6.1 <i>Realized niche</i> (Niche effective)</p> <ul style="list-style-type: none"> • <i>Vasconcellea_4species</i> (shp, shx, dbf) <p>Pour cette analyse, vous devez avoir importé dans DIVA-GIS des données climatiques de WorldClim à 2,5 minutes de résolution (voir la Section 2.2)</p>

6.1.1. Comment analyser et comparer les niches effectives de différentes espèces

Les espèces de *Vasconcellea* sont adaptées à différents environnements. Pour évaluer ces différences, les données climatiques des sites de présence respectifs de chaque espèce peuvent être extraites à l'aide de DIVA-GIS (voir le Chapitre 3). Si l'on connaît la capacité adaptative de chaque espèce aux différentes conditions climatiques, on pourra déterminer l'intérêt potentiel de l'inclure dans des programmes d'amélioration (par exemple, la tolérance au froid de la papaye) ainsi que d'en faire une culture commerciale, en tenant compte des zones agro-écologiques optimales. L'identification des limites de la niche écologique de l'espèce fournit aussi une information clé pour la conservation ; par exemple, des espèces que l'on trouve sous une plage de climats restreinte seront vraisemblablement plus vulnérables aux changements climatiques que celles qui ont une large niche climatique.

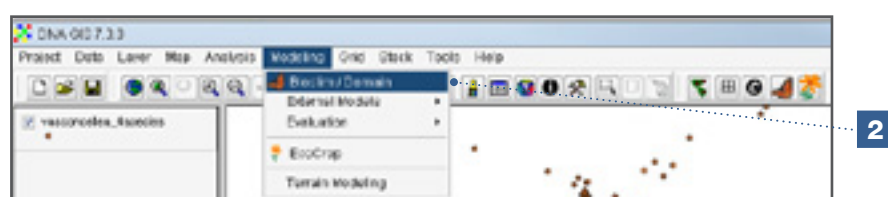
La section suivante présente comment déterminer la niche des espèces suivantes de *Vasconcellea* : *V. cundinamarcensis*, *V. microcarpa*, *V. quercifolia* et *V. parviflora*. Cette section explique également comment utiliser le menu *Modeling* de DIVA-GIS pour analyser les niches climatiques et comparer ensuite les niches de différentes espèces avec Excel.

Utilisation du menu *Modeling* de DIVA-GIS pour examiner la niche effective d'une espèce

Etapes :

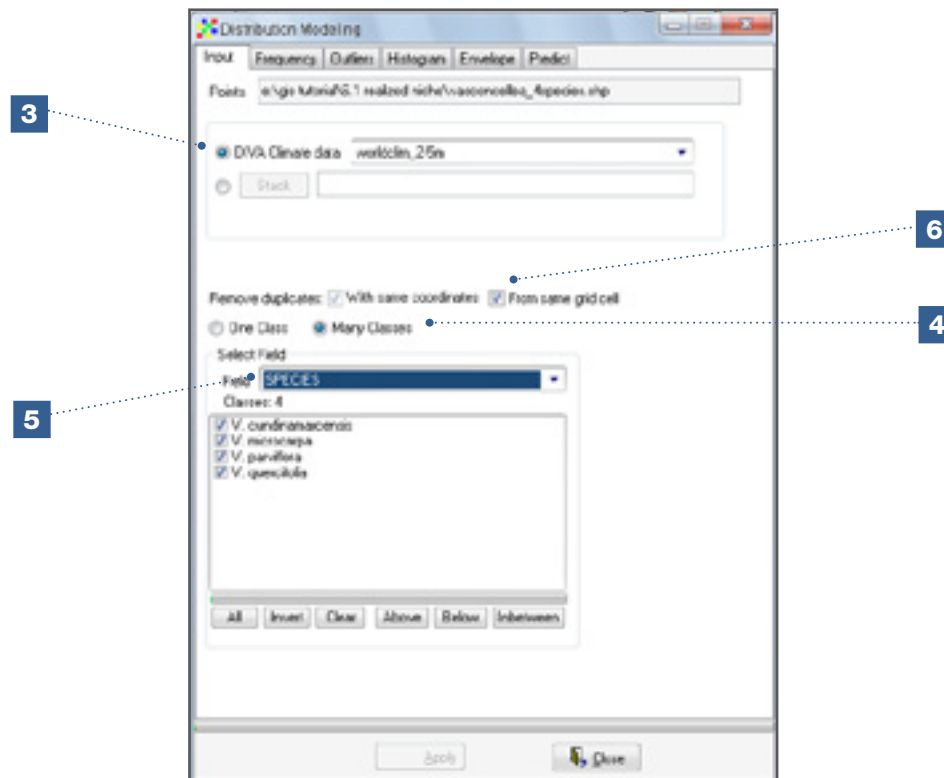
Après avoir importé les variables climatiques de BioClim dans DIVA-GIS (voir le Chapitre 2), on peut identifier la niche effective d'une espèce à partir des points de présence de l'espèce et des données climatiques correspondantes. On peut afficher la niche effective de différentes manières avec DIVA-GIS : en visualisant les fréquences des différents paramètres climatiques à l'aide d'histogrammes ; ou bien en affichant une niche climatique à deux dimensions selon deux paramètres climatiques. DIVA-GIS propose ces options d'analyse dans le menu *Modeling*.

1. Ouvrez le fichier *Vasconcellea_4species.shp* dans DIVA-GIS et garder le fichier comme couche sélectionnée.
2. Allez sur *Modeling/Bioclim/Domain* pour ouvrir la fenêtre *Distribution modeling*.



3. Dans la fenêtre *Distribution modeling*, sur l'onglet *Input*, allez sur *DIVA Climate data* et sélectionnez la base de données climatique à utiliser. Cette analyse utilisera des données climatiques WorldClim à 2,5 minutes de résolution (fichier : *worldclim_2-5min*). Il s'agit du fichier qui a été importé dans DIVA-GIS au Chapitre 2.
4. Sélectionnez l'option *Many Classes* pour distinguer les différentes classes au sein du fichier vecteur (*.shp), c'est-à-dire les différentes espèces, géotypes et pays. Cette option doit être sélectionnée puisque cette analyse porte sur quatre espèces différentes.
5. Dans la fenêtre *Field*, sélectionnez le paramètre qui sera utilisé pour définir les différentes classes. Dans le cas présent, sélectionnez *Species*. La liste complète des espèces de *Vasconcellea* sera alors affichée.

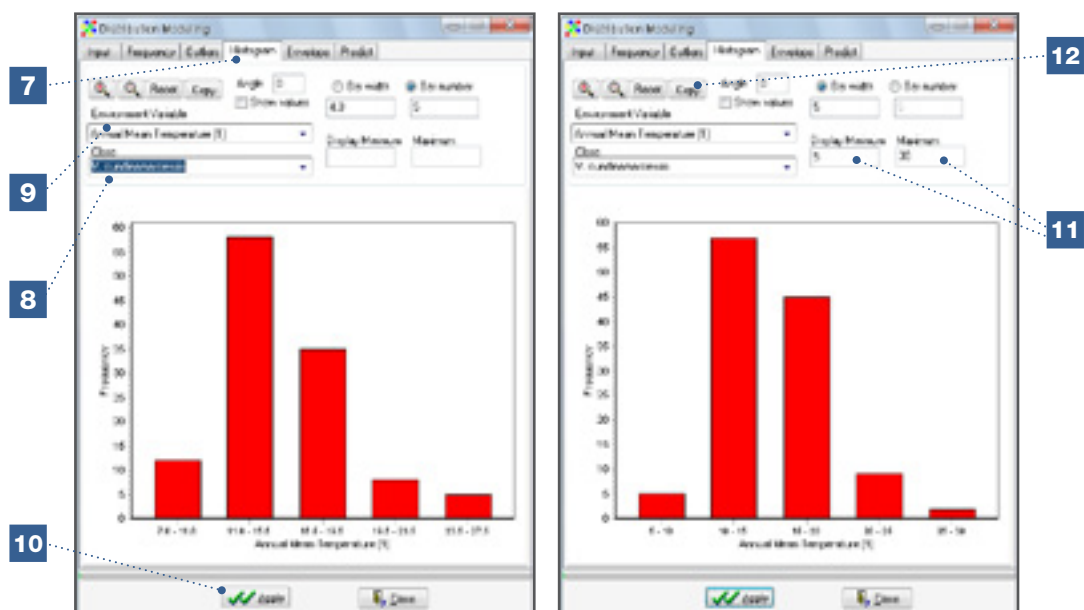
6. Indiquez si vous souhaitez ou non retirer les points de présence en double dans la même cellule du raster climatique. Pour cette analyse, sélectionnez l'option *Remove duplicates : From same grid cell*. Cela évitera que les cellules qui contiennent de nombreuses observations ne contribuent de manière disproportionnée à la définition de la niche effective.



Le menu inclut cinq types d'analyses (cinq onglets différents) dont deux seront expliqués dans cette section : *Histogram* et *Envelope*. Les options d'analyse *Frequency* et *Outliers* ont déjà été décrites au Chapitre 4. L'option *Predict* modélise la distribution potentielle d'espèce à l'aide de l'algorithme BioClim. Dans ce manuel cependant, la modélisation met en œuvre une méthode et un logiciel différents (Maxent) ; c'est pour cette raison que l'option *Predict* n'est pas expliquée plus en détail.

7. L'outil *Histogram* construit des histogrammes de fréquence qui traduisent la distribution d'une espèce, en fonction de différentes variables climatiques, sur des intervalles de valeurs de votre choix.
8. Sélectionnez l'espèce pour laquelle vous souhaitez générer un histogramme. Pour cette analyse, sélectionnez *V. cundinamarcensis*.
9. Sélectionnez ensuite la variable climatique que vous souhaitez utiliser pour visualiser la fréquence de la plage climatique dans laquelle on a observé l'espèce. Dans le cas présent, sélectionnez *Annual Mean Temperature*.
10. Cliquez sur *Apply* pour afficher l'histogramme.
11. La largeur et le nombre de barres de l'histogramme, de même que les valeurs minimale et maximale, peuvent être modifiées. Dans cette analyse, on choisit 5°C et 30°C comme valeurs minimale et maximale afin de générer cinq barres à des intervalles de cinq degrés Celsius.

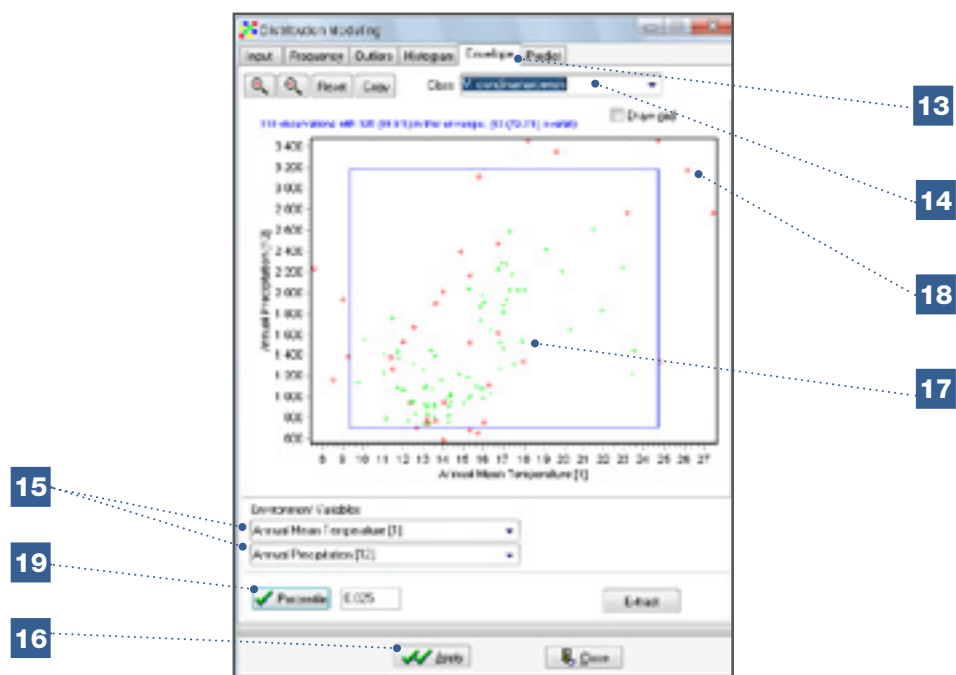
12. L'histogramme peut être copié et collé directement dans un document à l'aide de l'option *Copy*.



Enveloppe climatique

13. L'outil *Envelope* vous permet de visualiser une niche bidimensionnelle fondée sur deux variables climatiques.
14. Sélectionnez les espèces dont vous souhaitez générer la niche bidimensionnelle. Intéressez-vous aux différentes espèces de *Vasconcellea*, en particulier à *V. cundinamarcesis*.
15. Sélectionnez les deux variables climatiques sur lesquelles sera fondée la niche climatique ; Ce sont, pour cette analyse, *Annual Mean Temperature* et *Annual Precipitation*.
16. Cliquez sur *Apply* pour afficher la niche bidimensionnelle.
17. Les points verts situés à l'intérieur du rectangle bleu de la niche climatique représentent les points de présence dont le profil climatique entre dans les plages de variation des 19 variables climatiques de BioClim.
18. Les points rouges représentent les points de présence dont le profil climatique sort des plages de BioClim pour au moins une ou plusieurs des valeurs de ses 19 variables climatiques. Les points rouges situés dans le rectangle bleu correspondent aux points de présence dont le profil climatique est dans les plages de BioClim pour les deux variables sélectionnées (*Annual Mean Temperature* et *Annual Precipitation*) mais hors des plages pour au moins l'une des 17 autres variables.

19. Les limites de la niche bidimensionnelle sont, par défaut, les centiles 0,025 (les 2,5% les plus bas des points) et 0,975 (les 2,5% les plus haut des points), ce qui signifie que 95% des points de présence ont été pris en compte pour établir la niche bidimensionnelle (5% des points sont considérés comme aberrants). La largeur de la niche peut être ajustée en modifiant la valeur du centile. Selon ce qui vous intéresse, vous pouvez soit rétrécir la niche pour déterminer le domaine environnemental du cœur de la distribution de l'espèce, soit inversement l'élargir pour déterminer les conditions extrêmes dans lesquelles on peut encore trouver l'espèce (il est recommandé de ne faire cela qu'après avoir vérifié la qualité des données, étant donné que les valeurs atypiques peuvent avoir une influence significative sur la niche écologique lorsque l'on inclut tous les points avec une valeur centile de 0).



20. La fenêtre principale de DIVA-GIS affichera tous les points (en jaune) qui correspondent aux points verts de l'analyse *Envelope*. Les points qui conservent leur couleur originale (dans ce cas le bleu) correspondent à ceux dont le profil climatique sort des plages de BioClim pour l'une au moins des 19 variables climatiques.



Comparaison des niches effectives de différentes espèces dans Excel

DIVA-GIS permet d'exporter les données climatiques associées aux points de présence à l'aide de l'option *Extract values by points* (voir le Chapitre 3). Les données peuvent ensuite être analysées dans des feuilles de calcul Excel pour une meilleure visualisation.

Travail personnel : Extrayez les variables bioclimatiques de la base de données *WorldClim* avec une résolution de 2-5 minutes, à l'aide de l'option *Data/Extract Values by Points/From Climate Data* pour tous les points de présence du fichier *Vasconcellea_4species.shp* (voir l'Analyse 3.1.4).

Étapes :

1. On peut ouvrir dans Excel le fichier texte (*.txt) des données climatiques extraites. Les colonnes BIO1 à BIO19 correspondent aux 19 variables BioClim dont les valeurs correspondent à chaque point de présence.

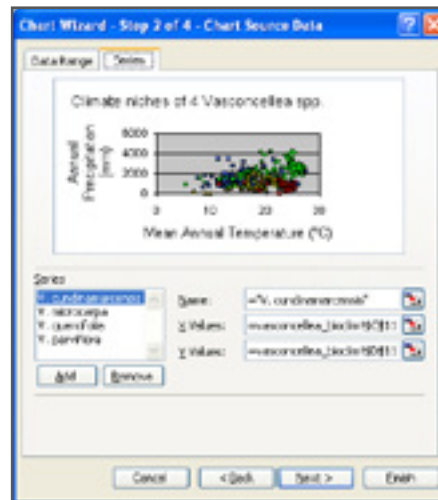
Species	Latitude	Longitude	Country	BIO1	BIO2	BIO3	BIO4	BIO5	BIO6	BIO7	BIO8	BIO9	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19
V. cundinamaricensis	11.8	-75.4177	Colombia	Antioquia	1	1	-75.4177	8.3636	11.8	10.21807	83.07524	47.99007	21.5	9.2	12.9	15.9	15.4	15.4	15.4	15.4	15.4	15.4
V. microcarpa	11.8	-75.4177	Colombia	Antioquia	2	2	-75.4177	8.3636	11.8	10.21807	83.07524	47.99007	21.5	9.2	12.9	15.9	15.4	15.4	15.4	15.4	15.4	15.4
V. quercifolia	11.8	-75.4177	Colombia	Antioquia	3	3	-75.4177	8.3636	11.8	10.21807	83.07524	47.99007	21.5	9.2	12.9	15.9	15.4	15.4	15.4	15.4	15.4	15.4
V. parviflora	11.8	-75.4177	Colombia	Antioquia	4	4	-75.4177	8.3636	11.8	10.21807	83.07524	47.99007	21.5	9.2	12.9	15.9	15.4	15.4	15.4	15.4	15.4	15.4

2. Pour construire des graphiques représentant les niches climatiques bidimensionnelles, il est recommandé de réorganiser les colonnes de la manière suivante : Colonne A : *ID* ; Colonne B : *SPECIES* ; Colonne C : la variable climatique de l'axe des X du graphique (dans le cas présent : *BIO1, Annual mean temperature*) ; Colonne D : la variable climatique de l'axe des Y du graphique (dans le cas présent : *BIO12, Annual precipitation*).

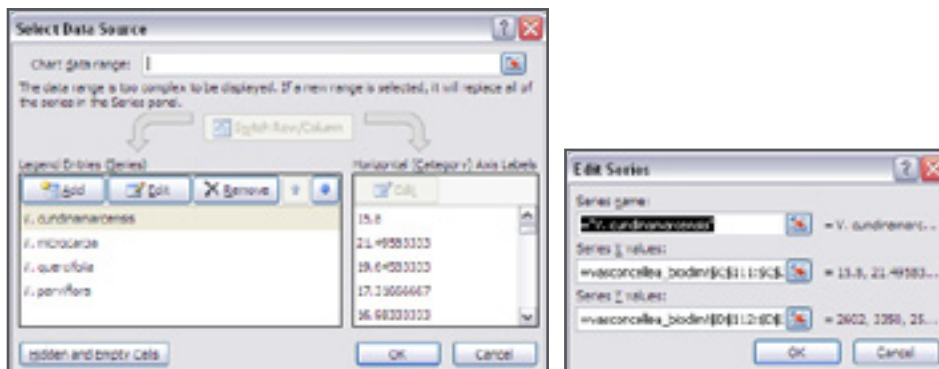
ID	Species	BIO1	BIO12
1	V. cundinamaricensis	11.8	10.21807
2	V. microcarpa	11.8	10.21807
3	V. quercifolia	11.8	10.21807
4	V. parviflora	11.8	10.21807

3. A l'aide d'Excel, vous pouvez maintenant créer un graphique de distribution des différentes espèces, selon la température moyenne annuelle en X et la précipitation annuelle en Y. Ceci vous permet de visualiser les différences entre espèces dans l'espace à deux dimensions de la niche. Le graphique généré dans cette analyse montre les plages climatiques pour *V. cundinamaricensis*, *V. microcarpa*, *V. quercifolia* et *V. parviflora*.

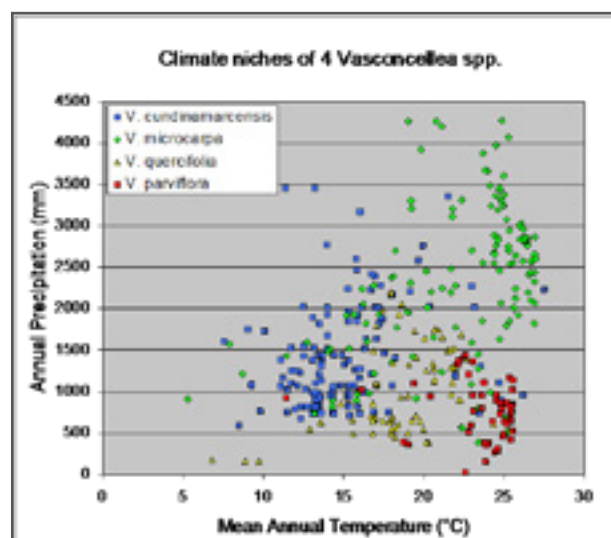
4. Sélectionnez les valeurs de température (valeurs en X) et de précipitation (en Y) pour chaque espèce comme des séries de valeurs distinctes.
 - a. Sélection d'une série dans Excel 2003.



- b. Sélection d'une série dans Excel 2007.



Résultat final :



Le graphique représentant les niches bidimensionnelles des différentes espèces de *Vasconcellea* devrait ressembler au graphique ci-dessus. On observe clairement des différences en précipitation et en température moyenne annuelles entre les niches effectives. Les niches de *V. cundinamarcensis* et *V. microcarpa* sont plus étendues que celles de *V. quercifolia* et *V. parviflora*, ce qui indique que ces espèces se sont adaptées à une plus grande variété d'environnements. La niche effective de *V. quercifolia* se limite à des zones tempérées avec une pluviosité annuelle modérée (c'est une espèce typique des vallées inter-andines de Bolivie), et la niche de *V. parviflora* se limite aux environnements chauds et secs (zones côtières du Nord du Pérou et de l'Equateur), ce qui indique que l'espèce est bien adaptée à des endroits où l'on trouve un haut degré de stress environnemental.

De manière analogue, on peut utiliser une niche multi-variée intégrant les 19 variables climatiques pour décrire les différences entre espèces. Des analyses multi-variées comme l'Analyse en Composante Principale (ACP) sortent du cadre de ce manuel, mais elles fournissent certainement une information supplémentaire qui ne peut être obtenue avec seulement deux dimensions. La pertinence d'un outil comme DIVA-GIS tient notamment à ce qu'il permet d'extraire les données des 19 variables climatiques pour les points de présence, données qui peuvent ensuite être analysées plus avant à l'aide de programmes statistiques et des logiciels comme Genstat (<http://www.vsnr.co.uk/software/genstat>) ou R (<http://www.r-project.org/>).

6.2. Modélisation de la distribution potentielle d'une espèce

Il arrive souvent que les données de présence disponibles ne couvrent pas la totalité de l'aire de distribution naturelle d'une espèce. Les programmes de modélisation de la distribution d'espèces comme Maxent (Philips *et al.*, 2006) permettent d'approcher la distribution dans sa globalité et sont des outils pratiques pour identifier les zones où l'on a des chances de trouver une espèce. Les résultats de l'analyse de modélisation de la distribution d'espèces peuvent être utilisés dans différentes analyses spatiales combinées, par exemple pour évaluer l'impact du changement climatique sur la distribution d'espèces (un thème qui est présenté dans la prochaine section), identifier les zones de collecte (ceci est expliqué dans la Section 6.4) ou les zones appropriées et la production forestière (comme indiqué dans l'introduction de ce chapitre).

Les programmes de modélisation de la distribution d'espèces identifient, comme zones d'occurrence possible d'une espèce, les sites où les conditions environnementales sont comparables à celles où l'espèce a déjà été observée. Pour identifier ces zones de distribution potentielle, les données requises comprennent les points de présence de l'espèce et les rasters des variables environnementales qui couvrent la zone étudiée. On commence par définir une niche à partir des valeurs environnementales qui correspondent aux points de présence utilisés dans l'analyse. Ensuite, on calcule, pour chaque cellule du raster dans la zone d'étude, les similarités entre les valeurs environnementales dans une cellule donnée et celle de la niche de l'espèce modélisée. Avec cette information, le modèle calcule la probabilité de la présence de l'espèce dans chaque cellule du raster. Bien que la prochaine analyse soit fondée sur des données climatiques, Maxent et d'autres programmes de niche permettent aussi d'inclure d'autres types de variables dans le modèle (comme des variables édaphiques).

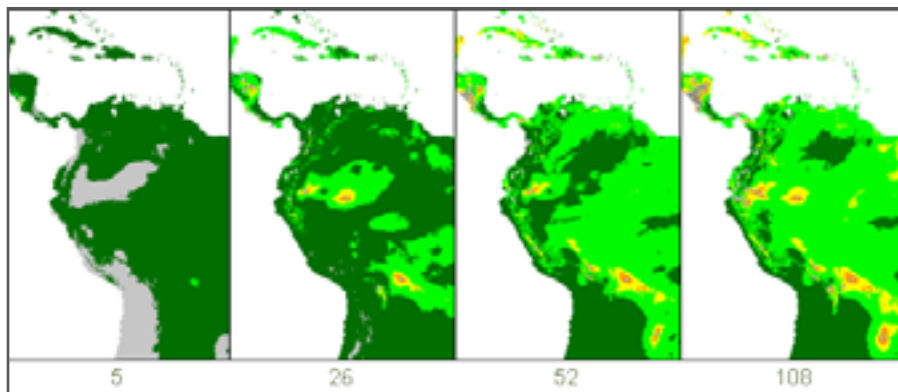
Comme cela a déjà été mentionné, le menu *Modeling* de DIVA-GIS inclut l'option *Predict* (voir la Section 6.1) qui propose d'intégrer deux programmes de modélisation de distribution d'espèces - *Bioclim* et *Domain* - qui diffèrent de Maxent. Chacun de ces programmes utilise des méthodes statistiques différentes pour estimer la niche effective et calculer la probabilité d'occurrence de l'espèce dans chaque cellule du raster. De ce fait, leurs résultats ont des chances d'être différents. Maxent calcule la niche effective et la probabilité d'occurrence d'une

espèce à l'aide d'un algorithme reposant sur le maximum d'entropie (Philips *et al.*, 2006). Comme, en comparaison avec les autres programmes, Maxent a bien fonctionné dans les évaluations (Elith *et al.*, 2006 ; Hernandez *et al.*, 2006), il est le programme privilégié dans ce manuel pour les analyses de modélisation de distribution d'espèces.

Il faut bien réaliser que quand une zone géographique présente des conditions environnementales favorables pour une espèce, ceci ne veut pas dire nécessairement que l'on trouve effectivement cette espèce dans cette zone. Des limitations de dispersion dues au système de reproduction de l'espèce et à des barrières géographiques peuvent empêcher une espèce d'occuper toutes les aires géographiques qui présentent des conditions environnementales similaires à sa niche effective. Il est aussi vrai qu'une espèce peut être absente de zones où elle pourrait se trouver, si son habitat naturel a été altéré par une interférence humaine.

L'importance de la quantité des données de points de présence

Comme il a été mentionné dans l'introduction de la Section 6.1, il est crucial de disposer d'un nombre suffisant de points de présence pour que la modélisation produise des résultats solides. L'illustration ci-dessous montre un exemple de cartes de distribution potentielle de *Carica papaya* générées dans DIVA-GIS à l'aide de l'outil de modélisation *BioClim*. Les résultats de la distribution potentielle s'affinent au fur et à mesure que le nombre de points utilisés au cours du processus de modélisation augmente (de 5 à 108 points de présence). Au delà de 50 points, la prédiction de la distribution potentielle se stabilise et ne change plus de manière significative, même si l'on continue à ajouter d'autres points de présence.



PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION

Programmes :	Fichiers de données :
<ul style="list-style-type: none"> • DIVA-GIS • Excel • Maxent et Java 	<p>Répertoire 6.2 <i>Potential distribution</i> (Distribution potentielle)</p> <ul style="list-style-type: none"> • <i>pkesiya.csv</i> • <i>seacountries (shp, shx, dbf)</i> • Répertoire <i>wclim_sea_2-5min</i> (fichiers asc)

6.2.1. Comment modéliser la distribution naturelle potentielle d'une espèce végétale

Pinus kesiya est une espèce dominante des forêts de pins naturelles en Asie du Sud-est et elle est économiquement importante. Les observations effectuées dans de nombreux pays indiquent que cette espèce a une distribution étendue ; les observations disponibles sont toutefois dispersées. Dans certains pays, on ne dispose que d'un ou deux points de présence. La modélisation de la distribution de l'espèce permet d'identifier l'aire potentielle où cette espèce pourrait se trouver de

façon naturelle (van Zonneveld *et al.*, 2009a). Cette analyse va modéliser la distribution potentielle de *P. kesiya*. Vous apprendrez comment utiliser Maxent pour modéliser la distribution potentielle d'une espèce et comment visualiser les résultats obtenus par Maxent dans DIVA-GIS.

La distribution observée de *P. kesiya*

La carte, ci-dessous, montre les points de présence disponibles pour *P. kesiya*. Dans certains pays, comme le Laos et la Birmanie (en gris foncé) les points de présence de cette espèce sont rares (un et deux points, respectivement, dans ce jeu de données). Maxent applique un modèle prédictif sur la base de ces points de distribution observée, pour identifier les zones où cette espèce pourrait potentiellement se trouver. Après avoir achevé l'analyse, il apparaît que, malgré le nombre limité de points disponibles, de larges zones dans ces pays ont une forte probabilité de présence de *P. kesiya*.



Utiliser Maxent pour modéliser la distribution naturelle potentielle d'une espèce

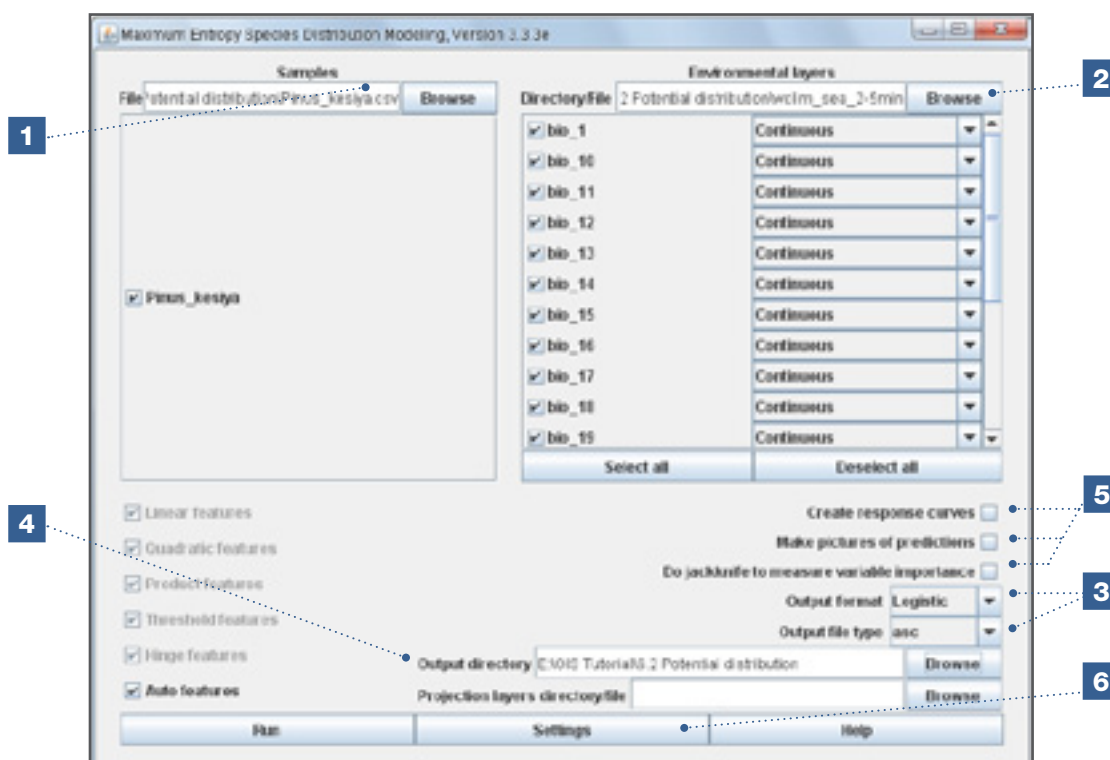
Etapes :

1. Dans Maxent, importer dans la fenêtre *Samples* le fichier *pkesiya.csv* qui contient les points de présence.

Note

De nombreux fichiers seront générés durant cette analyse. Nommez et sauvegardez minutieusement ces fichiers afin qu'ils puissent être facilement retrouvés, car ils vont être fréquemment utilisés dans cette section.

2. Dans Maxent, importez les couches climatiques au format ASCII (*.asc) en utilisant l'option *Environmental layers*. Dans cette analyse, les rasters ont une résolution de 2,5 minutes, qui représente les valeurs des 19 variables *BioClim* dans la zone d'étude (Asie du Sud-est). Voir le Chapitre 2 pour plus de détails.
3. Maxent génère un raster de la distribution potentielle. A l'aide de l'option *Output file type*, on peut sélectionner différents types de formats de fichiers résultats mais il est recommandé d'utiliser le fichier de type ASCII (*.asc).
4. Dans la fenêtre *Output directory*, sélectionnez l'emplacement (chemin d'accès) où doivent être sauvegardés les résultats de la modélisation.
5. Décochez l'option *Make pictures of predictions* et maintenez toutes les autres options par défaut. Pour de plus amples informations sur ces options, nous vous invitons à vous référer au Manuel de Maxent (Philips, 2009).
6. Allez sur l'onglet *Settings* pour modifier les paramètres.



Note

Pour que Maxent traite les rasters environnementaux, tous les rasters doivent avoir les mêmes propriétés, résolution et coordonnées des coins ou sommets (vertices). Si ce n'est pas le cas, Maxent génèrera une erreur et ne pourra pas fonctionner.

Maxent n'utilise que l'information présente dans les trois premières colonnes (voir 2.1.3). Si on inclut d'autres colonnes (*Fields*) dans le fichier CSV (*.csv), le signe *Visual warnings* s'affichera automatiquement. Dans ce cas, cochez OK dans chaque champ ou bien sélectionnez l'option *Suppress similar visual warnings*.

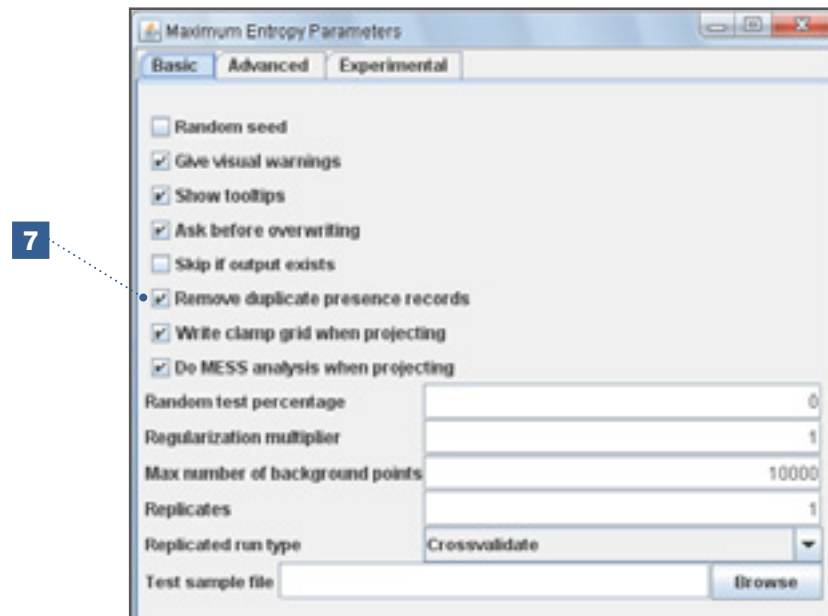
Comme cela a été mentionné à l'Etape 3, Maxent nous permet aussi de choisir des fichiers raster de résultats au format *.grd. L'avantage d'utiliser des fichiers *.grd, plutôt que des fichiers raster au format ASCII (*.asc), est que ceux-ci peuvent être ouverts directement dans DIVA-GIS. Néanmoins, on a quelquefois observé des erreurs dans Maxent en utilisant des fichiers de sortie *.grd. C'est pourquoi dans ce manuel nous recommandons d'utiliser des fichiers de sortie au format ASCII.

Maxent lira tous les fichiers raster qui sont dans le répertoire sélectionné comme *Directory File* dans l'option *Environmental layers*. Lorsqu'il y a des fichiers raster (*.grd) supplémentaires dans le répertoire *Directory File*, ils sont automatiquement inclus dans la liste des couches environnementales. Ceci peut entraîner des erreurs ou des résultats non souhaités et arrive quand des fichiers ASCII sont créés dans DIVA-GIS à partir de fichiers *.grd (voir le Chapitre 3) et stockés dans le même répertoire que les fichiers raster originaux (*.grd). On peut éviter de telles erreurs en sauvegardant dans un répertoire séparé les fichiers ASCII utilisés par l'analyse, ou en décochant tous les fichiers raster non pertinents dans la liste des couches environnementales dans Maxent.

Maxent est capable de conduire une analyse simultanée sur plusieurs espèces lorsque les points de présence de chaque espèce sont sauvegardés dans le même fichier CSV (*.csv). Voir la Section 6.4.

L'option *Settings* vous permet de modifier les conditions dans lesquelles Maxent génère un modèle de distribution potentielle. Un paramètre pertinent pour cette analyse est l'option *Remove duplicate presence records*, qui se trouve dans l'onglet *Basic*. Vous trouverez de plus amples informations sur *Settings* dans le manuel de Maxent (Philips, 2009).

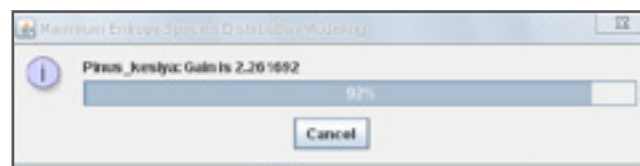
7. L'option *Remove duplicate presence records* permet d'éliminer des données analysées les points de présence dupliqués dans une même cellule de raster afin de réduire le biais d'échantillonnage, qui favoriserait, sinon, les conditions climatiques des sites où l'échantillonnage s'est le plus concentré. Pour cette analyse, assurez-vous que cette option est sélectionnée (cochée) et que les autres paramètres de configuration de base correspondent à ce que montre la capture d'écran ci-dessous.



8. Après avoir modifié l'option *Settings*, revenez à la fenêtre principale. Cliquez sur *Run* pour lancer le calcul de l'aire potentielle de l'espèce.

Comment faire tourner Maxent

Lorsque vous appuyez sur le bouton *Run*, une petite fenêtre affiche la progression et décrit les étapes franchies. Une fois que les couches environnementales ont été chargées et que le processus d'initialisation s'est achevé, une fenêtre affichera la progression dans le développement du modèle :

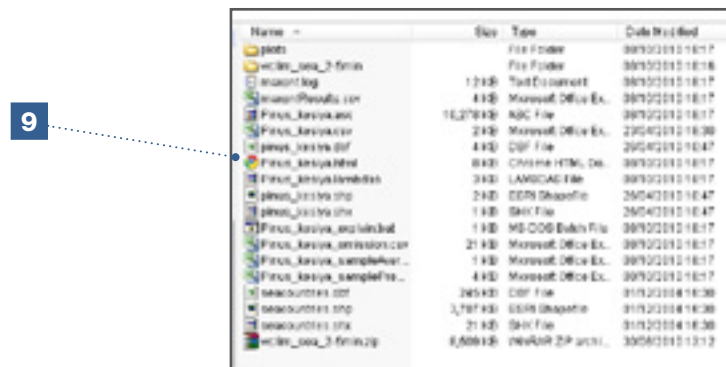


Le gain est étroitement lié à la déviance, une mesure de précision d'ajustement. Plus le gain est élevé, plus la distribution prédite pour l'occurrence d'une espèce est discriminante par rapport à une distribution aléatoire. Par exemple, si le gain vaut deux (2), cela indique que la vraisemblance moyenne de la présence est $\exp(2) = 7,4$ fois plus élevée que dans un pixel quelconque. Pour de plus amples informations, nous vous invitons à consulter l'article de référence sur Maxent (Philips *et al.*, 2006) et le manuel de Maxent (Philips, 2009).

Résultats de Maxent

Les résultats sont sauvegardés dans le répertoire sélectionné dans *Output Directory*. L'un des fichiers sauvegardés est un document HTML qui résume tous les résultats. L'analyse des paramètres les plus importants est décrite brièvement ci-dessous. Pour des plus amples informations sur l'analyse, voir Anderson (2003) et Philips *et al.* (2006).

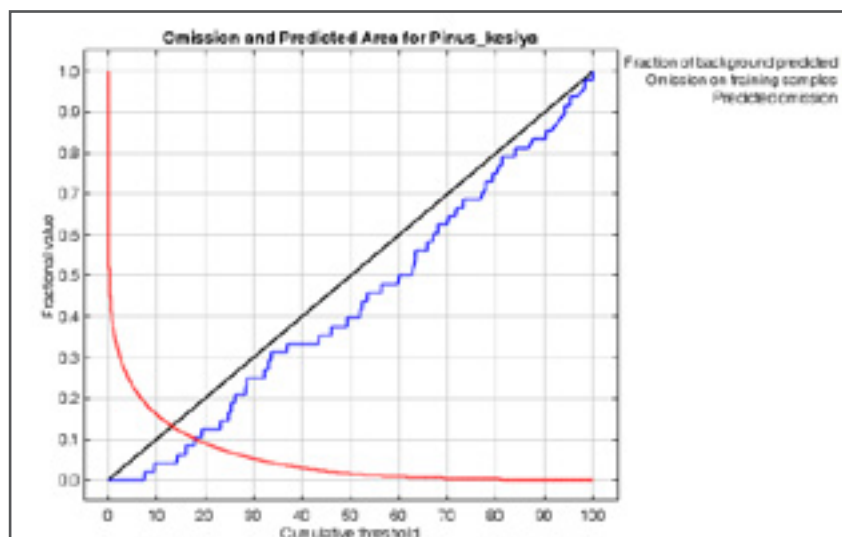
9. Ouvrez le document HTML *Pinus_kesiya* dans votre navigateur web (comme Firefox, Safari, Chrome ou Internet Explorer).



10. Le « taux d'omission » est une variable statistique indicatrice de la performance du modèle. Le graphique *Omission and Predicted Area* présente trois courbes :

- *Omission on training samples* (courbe bleue) montre les sous-ensembles des points de présence situés en dehors de l'aire potentielle modélisée par Maxent (*Fraction value*) par les valeurs du seuil de probabilité (basse et haute) qui délimitent l'aire prédite (*Cumulative threshold*). *Training samples* est synonyme de 'points de présence'.
- *Fraction of background predicted* (courbe rouge) montre les sous-ensembles de points aléatoires de la zone d'étude qui ont été inclus dans l'aire prédite selon divers seuils cumulatifs *Cumulative thresholds*.
- *Predicted omission* (courbe noire) est une courbe de référence.

Si la courbe bleue (*Omission on training samples*) est située beaucoup plus bas que la courbe de référence (*Predicted omission*), le modèle généré souffre probablement d'un 'surajustement' (*overfitting*) du fait de la dépendance entre les points (Phillips, 2009).

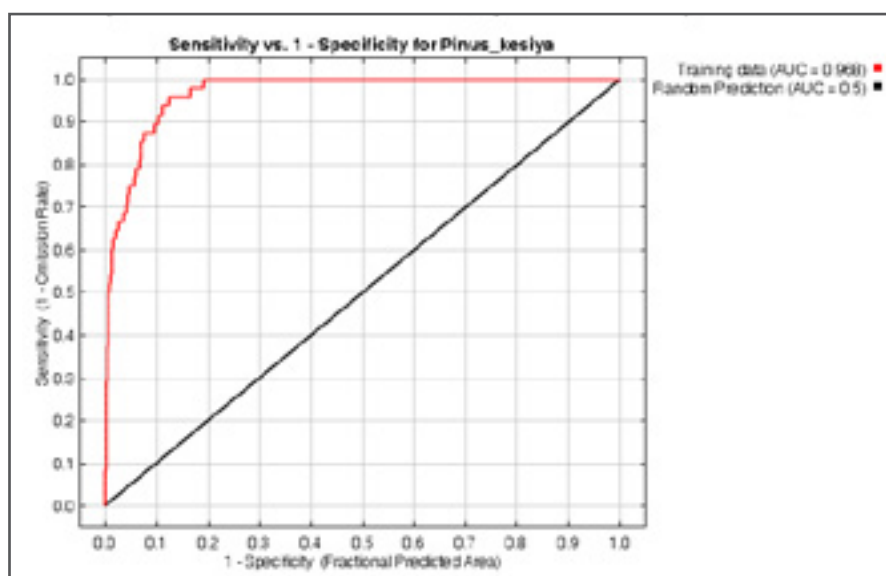


11. L'un des paramètres utilisé pour évaluer la capacité prédictive d'un modèle généré par Maxent est la Surface Incluse sous la Courbe (en anglais : AUC) de la courbe *Receiver Operating Characteristic* (ROC). Le sous-ensemble de la zone prédite (*Fractional predicted area*) sur l'axe X fait référence au sous-ensemble de l'aire totale de l'étude où l'espèce est modélisée, alors que la sensibilité (*Sensivity*) sur l'axe Y fait référence à la proportion de points de présence dans l'aire d'occurrence prédite par rapport au nombre total de points

de présence (Philipps, 2009). L'AUC est égal à la vraisemblance qu'un point de présence choisi au hasard soit situé dans une cellule du raster avec une plus grande probabilité d'occurrence de l'espèce qu'un point d'absence choisi au hasard. La modélisation de la distribution d'espèce dans ce manuel ne prend en compte que les points de présence. Afin de pouvoir quand même calculer l'AUC, Maxent remplace les points d'absence par des points de la zone d'étude générés de manière aléatoire. L'AUC peut alors être interprété comme la vraisemblance qu'un point de présence choisi au hasard soit situé dans une cellule du raster avec une plus grande probabilité d'occurrence de l'espèce qu'un point généré aléatoirement (Philipps *et al.*, 2006).

Un modèle généré par Maxent atteint le plus haut pouvoir prédictif lorsque l'AUC a une valeur de 1. Ce serait le cas si tous les points de présence avaient une plus haute probabilité que les points générés aléatoirement. En pratique, aucun AUC ne sera inférieur à 0,5. Dans ce cas, les cellules du raster où l'on trouve des points de présence n'ont pas une probabilité supérieure à celle des cellules du raster où l'on trouve des points générés aléatoirement, et le modèle généré par Maxent n'a aucun pouvoir prédictif. Cette situation est reflétée dans le graphique par la courbe de référence *Random Prediction*. Araújo *et al.* (2005) recommandent l'interprétation suivante de l'AUC pour les modèles générés : Excellente si $AUC > 0,90$; Bonne si $0,80 > AUC < 0,90$; Acceptable si $0,70 > AUC < 0,80$; Mauvaise si $0,60 > AUC < 0,70$; Invalide si $0,50 > AUC < 0,60$. Dans le cas de cette analyse, l'AUC est de 0,963. Pour de plus amples informations sur l'AUC et la courbe ROC, nous vous prions de vous référer à Fawcett (2006).

Lorsque la taille de l'aire d'occurrence prédite (modélisée), comparée à l'aire totale de l'étude, est faible, les valeurs élevées d'AUC ne reflètent pas nécessairement la bonne performance du modèle, mais peuvent résulter des caractéristiques statistiques de l'AUC.



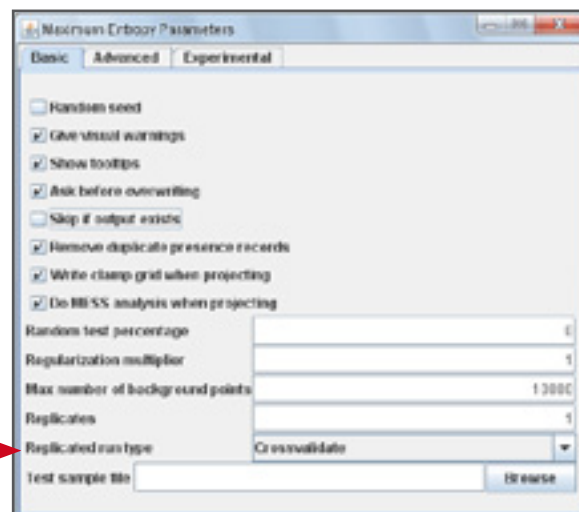
Le tableau de l'étape 12 illustre comment on peut utiliser des seuils pour délimiter l'aire potentielle de distribution d'une espèce (ici, *Pinus kesiya*). Trois seuils représentent la probabilité minimale qu'une espèce soit présente dans l'environnement d'une cellule spécifique. Ce concept suppose que les sites qui ont une probabilité supérieure au seuil ont des conditions climatiques appropriées pour l'espèce, tandis que l'espèce ne se trouverait pas dans les sites en dessous du seuil. Plus le seuil est élevé, plus les aires de distribution potentielle sont réduites. Il n'y a pas de valeur de seuil normalisée et l'utilisateur doit définir ce paramètre (pour plus d'information sur les seuils pour la délimitation des aires de distribution potentielle et des niches, voir Lui *et al.* [2005]).

12. Dans cette analyse, on utilisera le seuil de *10 percentile training presence* : la probabilité que 90% des points de présence tombent dans cette aire potentielle. Les 10% restant, qui tombent en dehors de l'aire potentielle, sont ceux qui ont un environnement atypique, qui n'entre pas dans les limites de la niche effective. Selon le format de sortie choisi (*Output format*), sélectionnez le format *Cumulative* ou le format *Logistic threshold* (le format de sortie par défaut est *Logistic threshold*). Pour cette analyse, sélectionnez le format *Logistic threshold* (voir la flèche rouge).

Cumulative threshold	Logistic threshold	Description	Fraction of predicted area	Training omission rate
1.900	0.019	Fixed cumulative value 1	0.377	0.900
2.960	0.049	Fixed cumulative value 5	0.232	0.900
10.090	0.091	Fixed cumulative value 10	0.160	0.942
3.483	0.073	Maximum training presence	0.190	0.900
17.725	0.134	10 percentile training presence	0.162	0.903
17.715	0.158	Equal training sensitivity and specificity	0.162	0.164
14.256	0.128	Maximum training sensitivity plus specificity	0.124	0.942
3.954	0.040	Balance training omission, predicted area and threshold value	0.254	0.900
17.495	0.152	Entropy of thresholded and original distributions	0.164	0.903

Validation de la robustesse du modèle

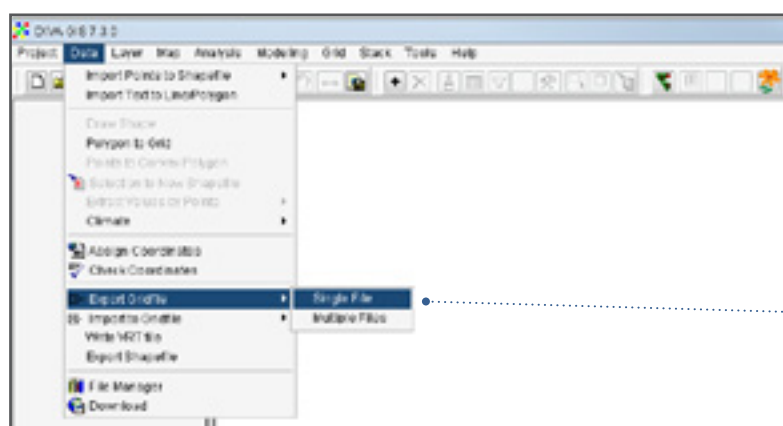
On peut valider la robustesse du modèle élaboré par Maxent en utilisant l'une des méthodes disponibles dans l'option *Replicate* run type de l'onglet *Basic*. (Ce manuel ne traite pas dans le détail de la validation des modèles ; pour de plus amples informations consultez Araújo *et al.* [2005] et Philips *et al.* [2006]). La robustesse ou capacité du modèle à être transposé est importante lorsque l'on veut prédire des aires de distribution potentielle en dehors de la distribution observée ou que l'on utilise divers scénarios climatiques, tels que les prédictions de la distribution d'espèces dans des conditions climatiques futures.



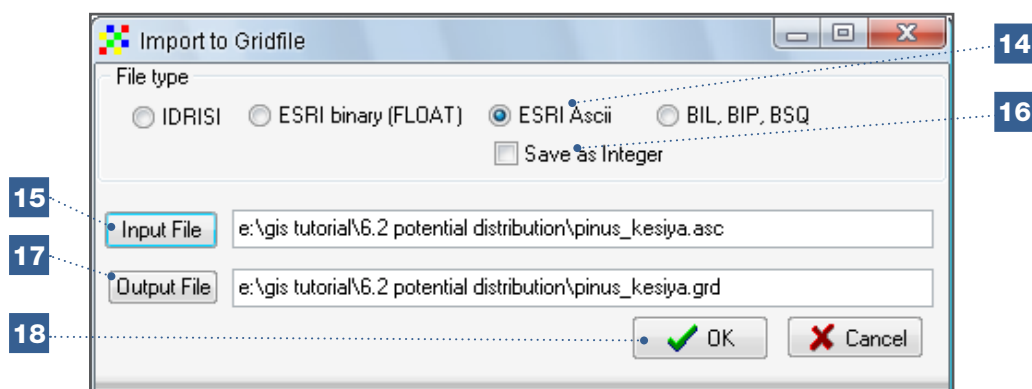
Visualisation des résultats de Maxent dans DIVA-GIS

Le raster de la distribution potentielle de *P. kesiya* généré par Maxent est au format ASCII (*.asc) et on peut le trouver dans le répertoire de sortie (le même répertoire où est stocké le fichier HTML). Afin de visualiser et modifier ces fichiers, il faut les importer dans DIVA-GIS.

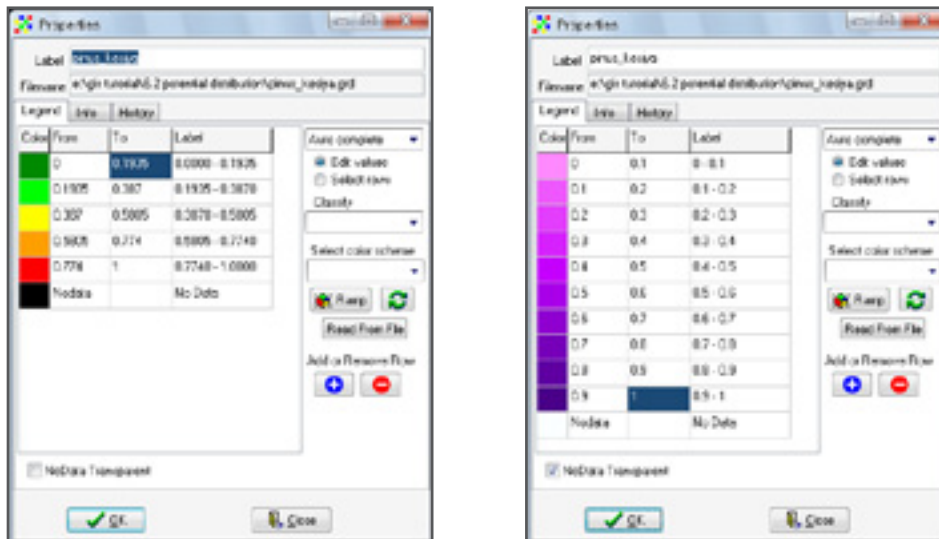
13. Ouvrez l'option *Import to Grid/Single File* dans le menu *Data* pour importer les rasters au format ASCII (*.asc) dans DIVA-GIS (voir le Chapitre 3).



14. Dans la fenêtre *File type*, sélectionnez *ESRI ASCII*.
15. Sélectionnez le fichier raster au format ASCII (*.asc) dont vous avez besoin pour générer un fichier raster au format *.grd. Pour cette analyse, sélectionnez le fichier ASCII généré par Maxent : *Pinus_kesiya.asc*.
16. L'option *Save as Integer* vous permet de choisir ou non de générer un raster dont les valeurs sont arrondies en entiers dans la légende de DIVA-GIS. Pour cette analyse, le raster n'aura pas de valeurs entières, il est donc important de NE PAS COCHER cette case.
17. Dans *Output File*, sélectionnez le dossier dans lequel vous allez stocker le fichier raster (*.grd) généré par DIVA-GIS. Sauvegardez le fichier raster dans ce même dossier en sélectionnant l'option *Same as input*.
18. Cliquez sur *OK* pour lancer le processus.

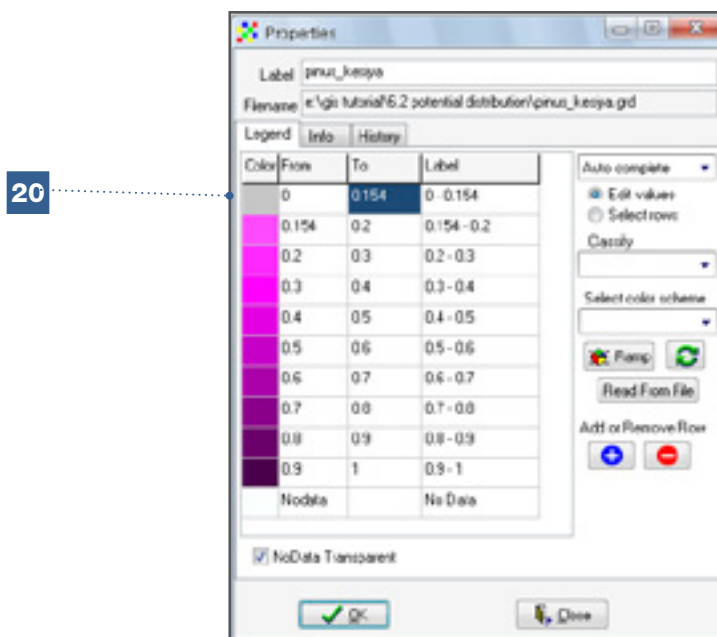


19. Après avoir importé le raster généré par Maxent dans DIVA-GIS, vous pouvez ouvrir le fichier, le visualiser et le modifier. La visualisation peut être améliorée en adaptant la légende proposée par défaut, comme décrit au Chapitre 3. Les étapes suivantes présentent les types de modifications qui peuvent être effectuées ; le choix des couleurs et la progressivité de l'échelle peuvent être adaptés aux préférences de l'utilisateur.

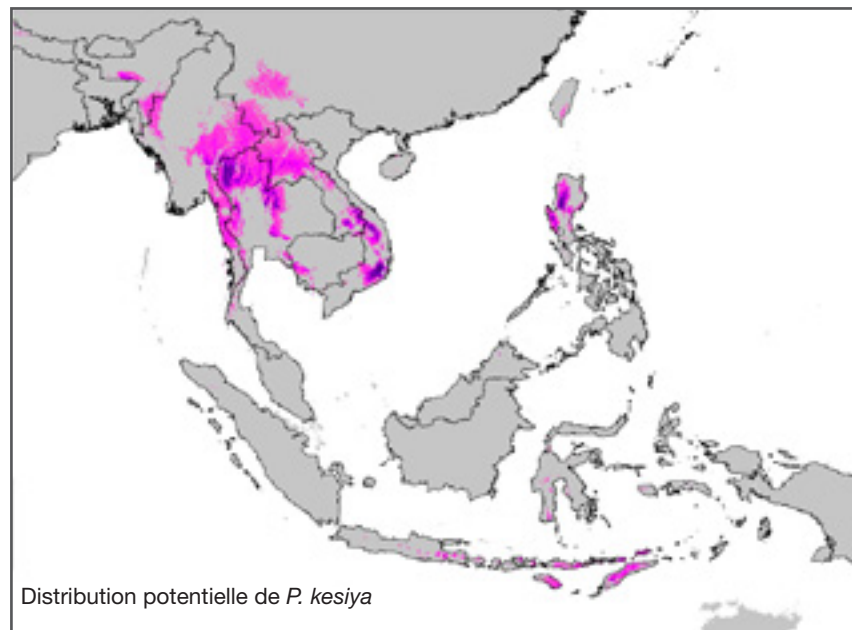


Visualisation du seuil de délimitation des aires de distribution potentielle

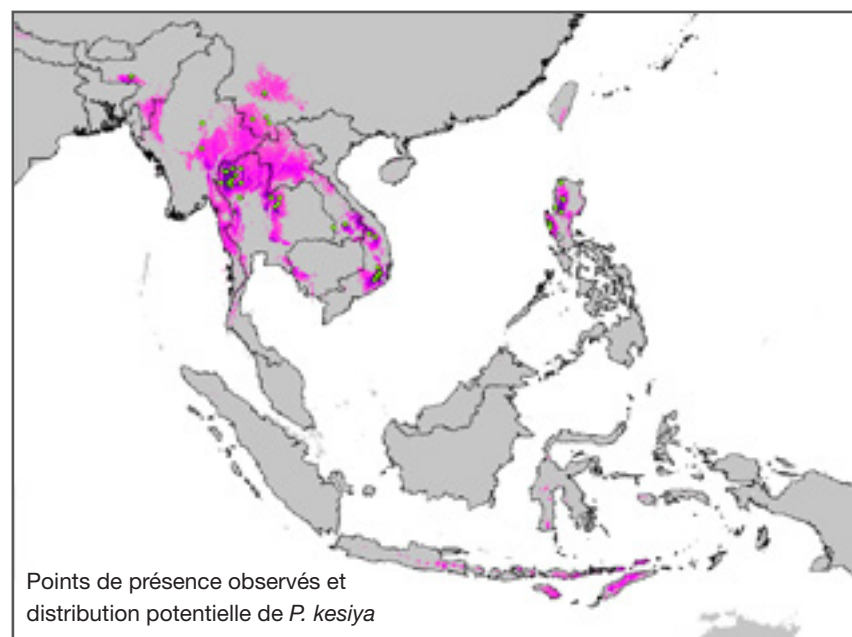
20. Dans cette analyse, on utilisera le seuil de *10 percentile training presence*, trouvé dans le tableau des seuils généré par Maxent (comme expliqué à l'Etape 11). La valeur du seuil pour *P. kesiya*, donnée par le tableau des seuils, est de 0,154. Allez à la légende du raster et créez une nouvelle classe s'étageant de zéro (0) jusqu'à la valeur du seuil ; sélectionnez une couleur neutre pour cette nouvelle classe.



21. Ajoutez une couche d'unités administratives (*seacountries.shp*) pour faciliter la localisation des aires de distribution potentielle.

Résultat final :

Après avoir modifié la carte de la distribution potentielle de *P. kesiya* (comme présenté dans le Chapitre 3), elle devrait ressembler à l'image ci-dessus. Certains pays, comme le Laos et la Birmanie, n'ont pas beaucoup de points de présence disponibles. Maxent révèle que ces pays possèdent des zones étendues où l'on pourrait trouver cette espèce de pin. La modélisation de la distribution d'espèces réalisée par Maxent complète la distribution observée, et dans cet exemple, fait ressortir les lacunes d'observation. Ceci signifie que les aires vacantes nécessitent une étude approfondie pour déterminer la localisation des populations de pins et définir les stratégies de conservation pour cette espèce (cette question sera traitée plus en détail dans la Section 6.4).



Il est très probable que l'espèce n'occupe pas toute l'aire de distribution potentielle présentée par le modèle parce que la dispersion de l'espèce est limitée par son système de reproduction

et par la présence de barrières géophysiques et climatiques. Dans cette illustration, l'archipel indonésien comprend plusieurs zones présentant des conditions climatiques similaires à la niche effective de *P. kesiya* ; néanmoins, ces zones ne sont pas incluses dans la distribution naturelle de l'espèce du fait des limitations signalées ci-dessus. Si on le souhaite, on peut limiter les aires de distribution potentielle en utilisant une zone tampon de rayon fixe autour de chaque point de présence (voir Willis *et al.*, 2003). Cette approche peut refléter de manière adéquate les facteurs limitant la dispersion liés au système de reproduction de l'espèce, mais elle ne résoudra pas le problème des barrières géophysiques et climatiques, qui peuvent être plus proches des points de présence représentés que l'on ne s'y attend. On peut aussi consulter la littérature ou contacter des experts de l'espèce étudiée et comparer ensuite l'information ainsi recueillie avec celle fournie par le modèle de distribution potentielle.

Il se peut que l'on trouve dans la littérature la mention de zones où l'on trouve une espèce, même si le modèle n'y prédit pas son occurrence. Cela peut arriver dans les cas où l'environnement qui règne sur les points de présence dont on dispose n'est pas tout à fait représentatif de la niche climatique, ce qui fait qu'une partie de l'aire de distribution potentielle échappe au modèle. Par exemple, l'ouvrage Taxonomie des plantes de GRIN (USDA, ARS, National Genetic Resources Programme, 2009) rapporte la présence de *P. kesiya* au Bhoutan et dans la province chinoise de Xizang ; toutefois, le modèle généré par Maxent ne prédit pas la présence de l'espèce dans ces zones. Pour pallier ce défaut, il faut une étude plus détaillée ou faire appel à l'expertise locale.

Il faut donc noter que la modélisation de la distribution d'espèces, la littérature scientifique et les données fournies par les experts, sont des sources d'information complémentaires. La combinaison de ces trois sources aide à se faire une idée complète de la distribution naturelle d'une espèce donnée.

Comparaison entre la distribution naturelle potentielle de *P. kesiya* selon Maxent et la distribution naturelle d'après la littérature



La distribution naturelle de *P. kesiya*, selon la Taxonomie des plantes de GRIN (USDA, ARS, National Genetic Resources Programme, 2009), est définie par les limites administratives surlignées sur la carte en gris foncé. Les pays inclus dans la distribution naturelle sont la Chine (Xizang, Yunnan), le Bhoutan, l'Inde, le Laos, la Birmanie, la Thaïlande, le Vietnam et les Philippines (Luzon).

6.3. Modélisation de l'impact du changement climatique sur la distribution des espèces

Le changement climatique mondial est chaque jour plus perceptible (IPCC, 2007). En conséquence, les aires géographiques correspondant aux biomes, aux écosystèmes et aux niches écologiques des espèces changent et cela aura probablement un impact sur la distribution naturelle de nombreuses espèces.

La modélisation de la distribution d'espèces peut être utilisée pour fournir une évaluation rapide de l'impact potentiel du changement climatique sur la distribution des écosystèmes et des espèces qui les composent. Le procédé consiste à détecter des changements dans la distribution d'espèces, en comparant les aires de distribution potentielle sous le climat actuel (sur la base des conditions climatiques au niveau des points de présence) avec les aires de distribution potentielle fondées sur les préférences climatiques actuelles d'une espèce dans des conditions climatiques futures. Les aires de distribution potentielle futures sont identifiées à l'aide de couches climatiques fondées sur les projections des modèles de circulation générale en anglais (GCM en anglais). Ces modèles sont générés par des institutions de recherche sur le climat dans divers pays (<http://www.ipcc-data.org/>) qui prédisent les conditions climatiques futures selon divers scénarios d'émission de gaz à effet de serre développés par le Groupe d'experts Intergouvernemental sur l'Evolution du Climat (IPCC en anglais). Pour de plus amples informations, visitez le site http://www.ipcc.ch/publications_and_data/publications_and_data.shtml.

Note

Assurez-vous que lorsque vous développez une carte de la distribution potentielle d'une espèce, vous n'utilisez pas les conditions climatiques futures prédites pour les points de présence actuels de cette espèce. Les aires de distribution potentielles, tant actuelles que futures, doivent se fonder sur la niche climatique de l'espèce, qui, elle, est calculée à partir des données climatiques actuelles.

6.3.1. Comment évaluer l'impact du changement climatique sur la distribution des espèces

Bien qu'il existe aujourd'hui plusieurs projets sur la conservation *in situ* du pin en Asie du Sud-est, la zone occupée par *Pinus kesiya* a diminué dans les dernières décennies. Dans de nombreux sites où on le trouve encore, il est menacé, de façon permanente, par des pratiques non durables d'extraction de résine et de coupe de bois et il sera aussi vraisemblablement menacé par les effets du changement climatique. La modélisation de la distribution d'espèces, tout comme les modèles climatiques, peuvent aider à identifier les populations les plus menacées et à déterminer où la conservation des ressources génétiques requiert les mesures les plus urgentes (van Zonneveld *et al.*, 2009a). Dans cette analyse, nous allons explorer l'impact possible du changement climatique sur la distribution de *P. kesiya*.

Dans cette analyse, vous apprendrez comment utiliser Maxent pour prédire la distribution potentielle d'une espèce dans des conditions climatiques actuelles et futures et pour examiner, à l'aide de DIVA-GIS, l'impact du changement climatique sur une espèce.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION

Programmes :

- DIVA-GIS
- Maxent et Java

Fichiers de données :

- Répertoire 6.3 *Climate change* (Changement climatique)
- *pkesiya.csv*
 - *seacountries (shp, shx, dbf)*
 - *Carpeta wclim_sea_2-5min (asc files)*
 - *Carpeta gcm_sea_2-5min (asc files)*

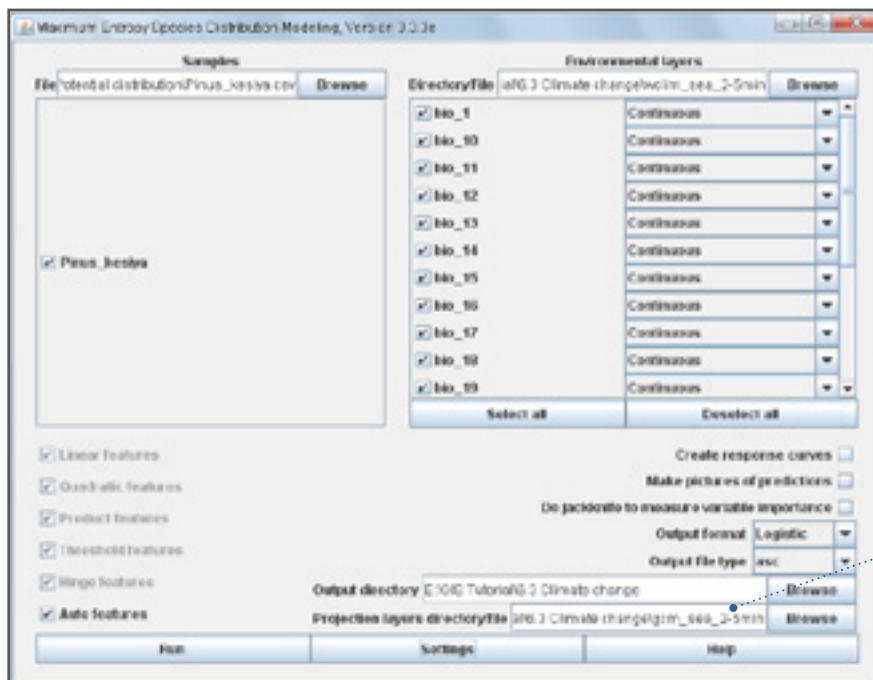
Cette analyse utilisera des projections climatiques pour l'année 2050, selon le scénario d'émissions A2, calculées à l'aide de trois GCM différents : *CCCMA*, *HADCM3* et *CSIRO*. Chaque modèle donne une projection du climat futur légèrement différente. Les prédictions des aires de distribution potentielle des espèces varieront donc selon le modèle utilisé dans Maxent. Ici, nous utiliserons la moyenne des trois GCM sélectionnés.

Note

De nombreux fichiers vont être générés au cours de cette analyse. Nommez et sauvegardez avec soin ces fichiers pour pouvoir les retrouver aisément ensuite car ils seront utilisés fréquemment dans cette section.

Etapes :

1. Pour faire tourner le modèle de Maxent, suivez les Etapes 1 à 8 présentées dans l'Analyse 6.2.1. Puis, dans la fenêtre *Projection layers directory file*, importez, dans la base de données, les valeurs des variables climatiques dans les conditions futures (*gcm_sea_2-5min*).



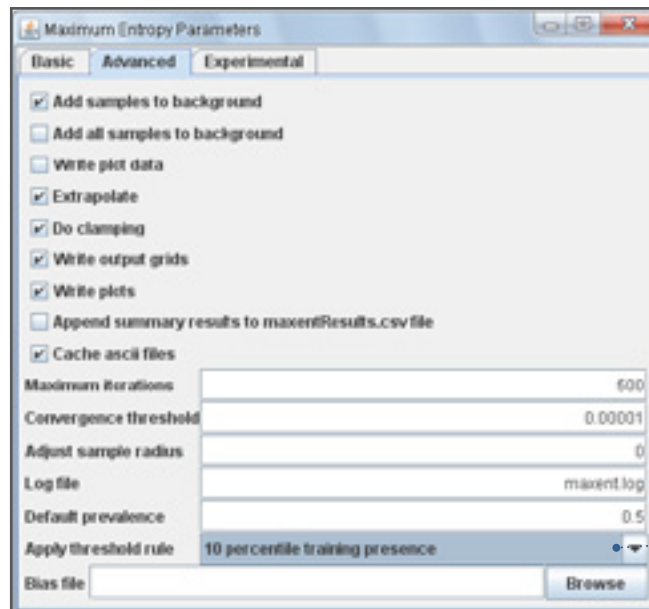
Note

Pour que Maxent traite les fichiers raster environnementaux des conditions actuelles et futures (dans le cas présent, *wclim_2-5min* et *gcm_sea_2-5min*), tous les fichiers doivent avoir les mêmes paramètres en termes de propriétés du raster, de résolution et de position des coins.

Appliquer la règle du seuil pour créer des rasters binaires

Maxent peut aussi générer des rasters binaires de présence (1) et d'absence (0) des aires de distribution potentielle. Ce format est utile lorsqu'on compare et combine des couches de distribution potentielle des espèces, comme nous allons le faire dans cette analyse. Les aires de distribution potentielle prédites selon les conditions actuelles et les conditions futures projetées vont être comparées pour permettre d'identifier les zones de distribution sur lesquelles le changement climatique aura un impact fort, ainsi que celles où l'impact sera moindre, et les nouvelles aires de présence naturelle des espèces à l'avenir.

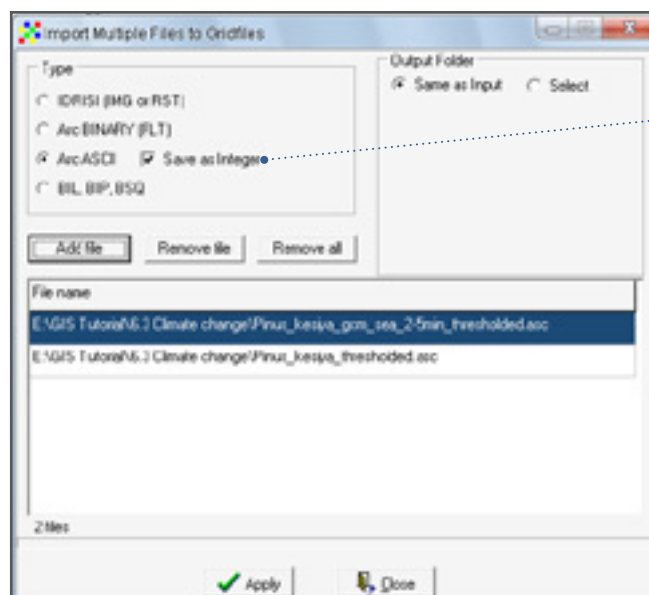
- Pour générer les rasters binaires dans Maxent, allez au menu *Advanced settings* et, dans l'option *Applied threshold rule*, sélectionnez un seuil qui conditionnera la délimitation de l'aire de distribution potentielle. Ce seuil (*10 percentile training presence*) est le même que celui qui a été visualisé manuellement dans DIVA-GIS dans l'Etape 21 de l'Analyse 6.2.1. Lancez l'analyse dans Maxent.



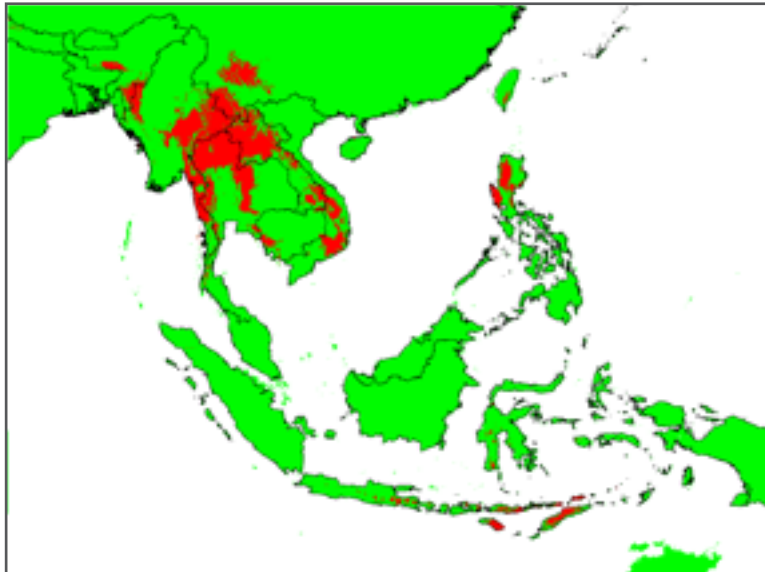
Examiner l'impact du changement climatique à l'aide de DIVA-GIS

Dans la première partie de cette section, nous avons généré dans Maxent des rasters binaires au format ASCII (*.asc) correspondant au climat actuel et à la projection future. Ces rasters vont maintenant être importés dans DIVA-GIS comme cela a été fait dans les Etapes 13 à 19 de l'Analyse 6.2.1.

- Ouvrez DIVA-GIS et importez les rasters de présence/absence des aires de distribution potentielle (*_thresholded.asc) sous forme de fichiers rasters à valeurs entières (ces fichiers peuvent être trouvés dans le répertoire de sorties). Cochez l'option *Save as integer* dans la fenêtre *Import Multiple Files to Gridfiles*.



4. Après avoir fermé la fenêtre *Import Multiple Files to Gridfiles*, ouvrez les rasters avec DIVA-GIS. La carte de présence/absence générée pour *P. kesiya* pour les conditions climatiques actuelles (*Pinus_kesiya_thresholded.grd*) devrait ressembler à l'illustration suivante.



Superposition des rasters des aires de distribution potentielle actuelles et futures

Une manière pratique de déterminer l'impact du changement climatique sur la distribution de groupes d'espèces consiste à superposer les rasters des aires de distribution potentielle actuelles et futures. Il est recommandé d'utiliser des rasters binaires pour construire des cartes faciles à interpréter ; mais, pour une analyse plus détaillée, il est aussi possible d'utiliser des rasters qui visualisent les aires de distribution potentielle de manière graduelle avec des probabilités croissantes. Dans cette analyse, on superpose les rasters binaires des zones de distribution potentielle actuelles et futures de *P. kesiya*.

La superposition des rasters binaires peut conduire à quatre situations pour chaque cellule :

- i. *Zones à fort impact* : zones où une espèce peut se trouver dans les conditions climatiques actuelles mais qui ne seront plus adaptées à l'avenir.
- ii. *Zones hors de la niche effective* : zones qui ne sont adaptées, ni dans les conditions climatiques actuelles, ni dans les conditions futures (telles qu'elles sont prédites).
- iii. *Zones à faible impact* : zones où l'espèce peut se trouver, aussi bien dans les conditions climatiques actuelles que dans les conditions futures.
- iv. *Zones nouvellement adaptées* : zones où une espèce pourrait se trouver à l'avenir, mais qui ne sont pas adaptées actuellement.

Chaque raster binaire a deux valeurs possibles : présence (1) et absence (0). Lorsque ces deux valeurs sont additionnées ou soustraites, les seuls résultats possibles pour les cellules sont : moins un (-1), zéro (0) et un (1). Mais cela pose problème car il faut une quatrième valeur pour représenter les quatre situations décrites ci-dessus. Le tableau suivant illustre ce problème : soustraire les rasters donne des cellules qui ont la même valeur dans la seconde et la troisième situations (ii – zones en dehors de la niche effective et iii – zones à faible impact), puisque dans les deux cas il n'y a pas de changement (d'où la valeur de zéro [0] après soustraction).

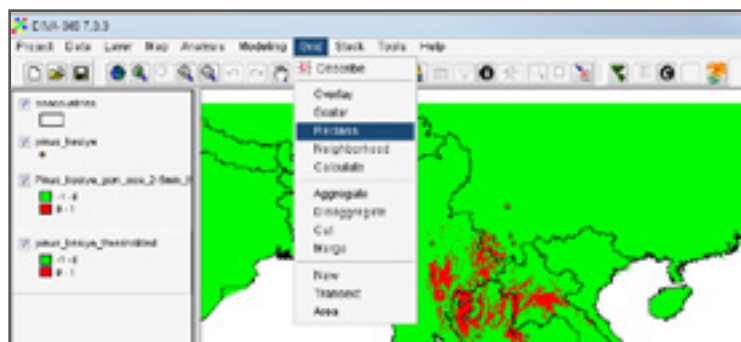
Situation	Raster des zones de distribution potentielle future (valeur de cellule)	Raster des zones de distribution potentielle actuelle (valeur de cellule)	Résultat après soustraction des rasters (valeur de cellule)
(i) Zones à fort impact	0	1	-1
(ii) Zones hors de la niche effective	0	0	0
(iii) Zones à faible impact	1	1	0
(iv) Zones nouvellement adaptées	1	0	1

Pour résoudre ce problème dans DIVA-GIS, allez à l'option *Reclass* et changez la valeur de cellule des zones de distribution potentielle de un (1) à deux (2) pour l'un des deux rasters (voir l'Analyse 3.1.3). Le tableau suivant reflète comment ce changement de valeur de cellule conduit à une valeur de cellule différente dans les quatre situations.

Situation	Raster des zones de distribution potentielle future (valeur de cellule)	Raster des zones de distribution potentielle actuelle (valeur de cellule)	Résultat après soustraction des rasters (valeur de cellule)
(i) Zones à fort impact	0	1	-1
(ii) Zones hors de la niche effective	0	0	0
(iii) Zones à faible impact	2	1	1
(iv) Zones nouvellement adaptées	2	0	2

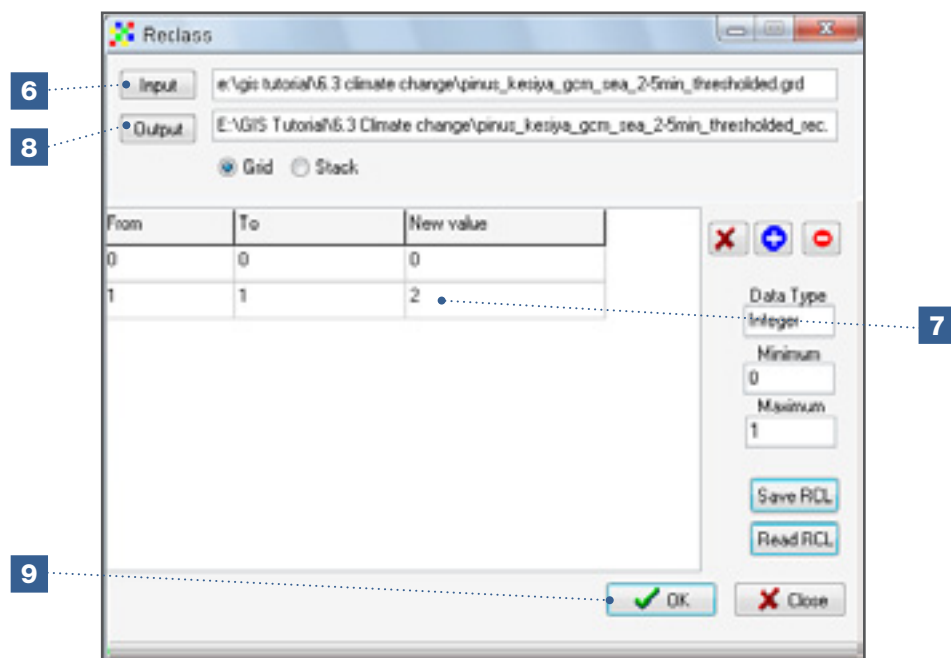
Pour la prochaine analyse, changez la valeur des zones de distribution potentielle de un (1) à deux (2) pour les rasters de la distribution future prédite pour *P. kesiya*.

- Allez à *Grid/Reclass* pour changer la valeur de cellule du raster des zones de distribution potentielle pour le climat futur (*Pinus_kesiya_gcm_sea_2-5min_thresholded.grd*) de un (1) à deux (2).

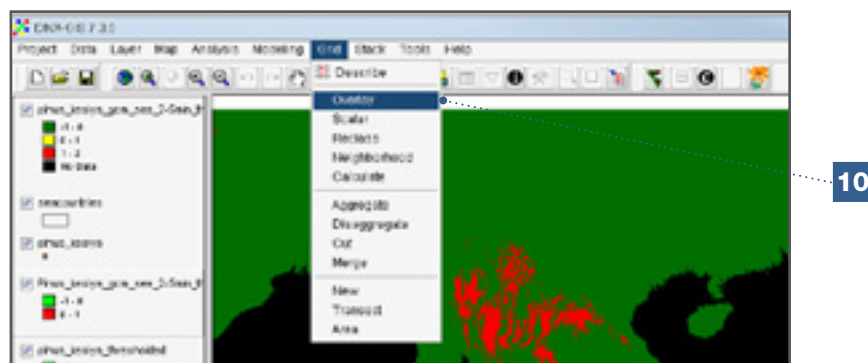


- Dans *Input*, sélectionnez le raster pour lequel vous allez changer les valeurs de cellules.

7. Pour éviter toute confusion durant cette modification, changez les valeurs des cellules dans la colonne *From* de moins un (-1) à zéro (0) et de zéro (0) à un (1). Ensuite changez les valeurs de cellules de un (1) à deux (2) dans la colonne *New*.
8. Indiquez où le raster contenant les valeurs modifiées sera sauvegardé. Sauvegardez le raster sous un nom différent. Normalement, on fait cela en ajoutant le suffixe (rec).
9. Sauvegardez le raster contenant les valeurs modifiées.

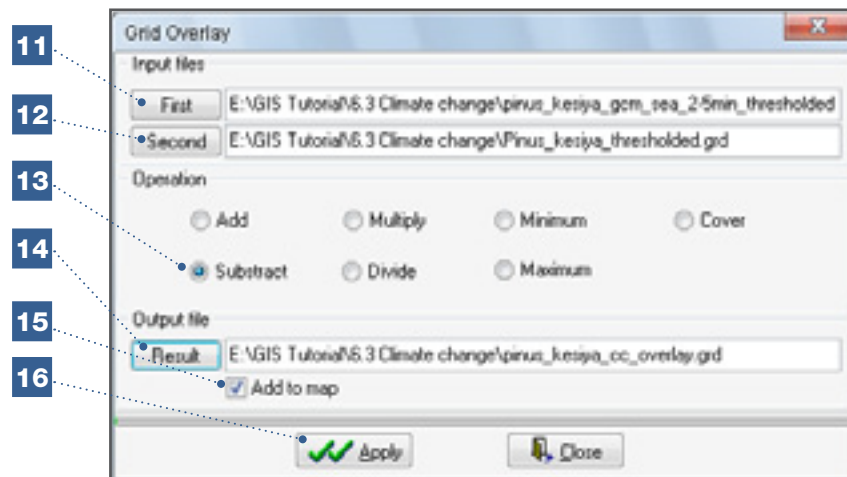


10. Après avoir changé les valeurs, allez à *Grid/Overlay* et superposez les rasters des zones de distribution potentielle actuelles et futures.



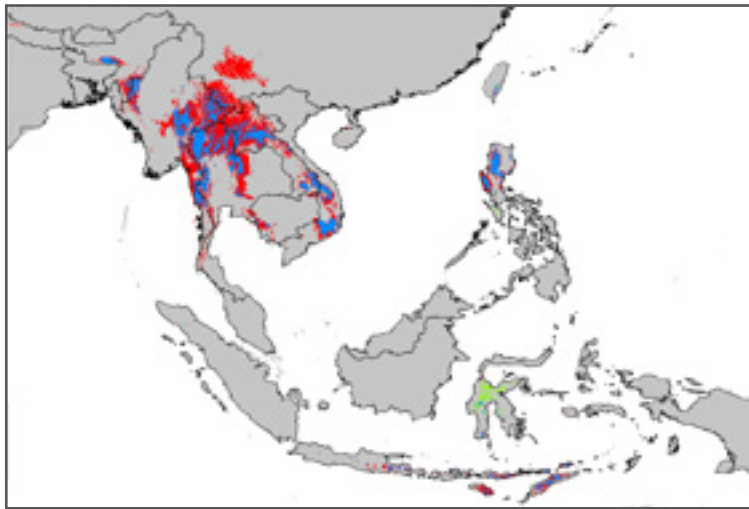
11. Dans l'onglet *First*, sélectionnez le raster binaire des zones de distribution potentielle futures.
12. Dans l'onglet *Second*, sélectionnez le raster binaire des zones de distribution potentielle actuelles.
13. Dans *Operation*, sélectionnez l'option *Subtract* pour soustraire les valeurs des deux rasters.

14. Dans l'onglet *Result*, indiquez le raster qui contient le résultat de la soustraction des rasters superposés.
15. Sélectionnez l'option *Add to map*. Une fois que vous aurez cliqué sur *Apply*, le nouveau raster s'ouvrira automatiquement dans DIVA-GIS.
16. Cliquez sur *Apply* pour lancer les calculs.



17. Une fois que le raster a été ouvert dans DIVA-GIS, la légende et les étiquettes peuvent être modifiées. Pour visualiser aisément les quatre situations, il est recommandé de modifier la légende comme illustré dans l'image suivante.





Après avoir édité la carte (comme expliqué au Chapitre 3), le résultat devrait ressembler à la carte ci-dessus. Les résultats révèlent que le changement climatique aura un impact particulièrement important sur les populations de *P. kesiya* dans la province chinoise du Yunnan (du fait d'une augmentation prévisible de la saisonnalité) et aura, plus généralement, un impact sur les zones de basse altitude (puisque'il est prédit que celles-ci vont devenir trop chaudes).

En tant que telles, ces zones peuvent être considérées comme prioritaires pour la mise en place de stratégies de conservation, y compris les stratégies conçues pour collecter le matériel végétal dans le but d'assurer la conservation *ex situ* des ressources génétiques avant que les populations existantes ne disparaissent du fait du changement climatique. Les stratégies de conservation *in situ* pour protéger les populations dont il est prédit qu'elles subiront un fort impact pourraient se focaliser sur l'amélioration de la connectivité entre fragments de populations afin d'assurer un flux de gènes adaptatifs. Une autre option serait d'aider à la migration des espèces menacées vers des zones nouvellement adaptées (colorées en vert sur la carte ci-dessus). On s'attend à ce que, du fait du changement climatique, les zones de l'île indonésienne de Sulawesi et de l'île Mindoro des Philippines deviennent adaptées à la présence de *P. kesiya* vers 2050. Une autre stratégie de conservation *in situ* pourrait être d'accroître l'effort pour conserver les populations dans les zones à faible impact, où les modèles prédisent que l'espèce survivra en 2050, en les protégeant des menaces d'origine anthropique.

Ces prédictions ne traitent que de l'impact potentiel. Il se peut que les modèles surestiment l'impact du changement climatique dans le cas où l'espèce est capable de s'adapter à une large gamme de conditions climatiques. Plusieurs espèces d'arbres ont justement un haut degré de variabilité génétique et ils peuvent, de ce fait, tolérer des climats très divers. Des essais multisites, conduits sur l'espèce de pin *P. kesiya*, ont montré que cette espèce s'adapte bien à une grande variété de climats et qu'elle s'adaptera vraisemblablement aussi aux nouvelles conditions climatiques, même si les études de modélisation de la distribution de l'espèce prédisent que les conditions climatiques futures seront inadéquates (van Zonneveld *et al.*, 2009b). Il faut noter que les caractéristiques édaphiques, la compétition, les prédateurs ainsi que d'autres facteurs, influent aussi sur la présence d'une espèce et constituent des limitations supplémentaires à la distribution actuelle de l'espèce et à ses déplacements à venir. Néanmoins, du fait que le climat est considéré comme le facteur qui aura la plus grande influence sur les aires de distribution futures, les modèles utilisés pour prédire les effets du changement climatique ne se sont pas encore penchés sur ces autres facteurs et ne les ont pas encore intégrés.

En dépit de leurs limitations, les modèles de niche sont considérés comme un outil utile pour établir une première appréciation de l'impact potentiel du changement climatique sur la distribution d'espèces (Pearson et Dawson, 2003).

6.4. Identification des lacunes dans les collections d'espèces de plantes sauvages

Comme mentionné dans la section 6.2, une des utilisations des cartes de distribution potentielle est d'identifier les lacunes dans les données de distribution des espèces. Une lacune fait référence à un endroit où les modèles de distribution d'espèces prédisent que les espèces peuvent être présentes, mais hors des lieux où ont été collectés des spécimens ou du matériel génétique d'espèces sauvages ou cultivées. Les lacunes peuvent indiquer que les collections des banques de gènes ou les herbiers manquent d'accessions et de spécimens en provenance de ces zones (par exemple Jarvis *et al.*, 2005; Scheldeman *et al.*, 2007). Il est également possible que des observations et des études locales des espèces dans ces zones soient disponibles, mais que l'information ne se soit pas largement disséminée, ni incluse dans des initiatives telles que la plate-forme mondiale sur la biodiversité (GBIF, sigle en anglais (<http://www.gbif.org/>) qui encouragent l'utilisation de ces données entre le grand public (voir section 2.3).

Il est également possible, cependant, qu'une espèce n'existe tout simplement pas dans la région que le modèle a prédit, en raison des limitations de la dispersion. Un cas comme celui-ci est étudié à la Section 6.2 où il a été prédit que le *Pinus kesiya* serait présent dans plusieurs îles du Sud-est de l'Indonésie, mais il n'a pas été trouvé de façon naturelle, dans ces zones.

Il est également possible que certaines espèces aient disparu de la région en raison de la déforestation, l'abattage sélectif ou d'autres pressions anthropiques sur le milieu naturel. Par exemple, les forêts de *Polylepis* en Bolivie et au Pérou sont actuellement très fragmentées. Sur la base des études de niche écologique, il a été démontré que ces forêts ont une large distribution dans la région andine de ces pays (Fjeldså, 2002).

6.4.1. Comment identifier des lacunes possibles dans les collections

Dans la section 5.1, on a identifié les domaines de la diversité observée des espèces de *Vasconcellea* en Amérique latine. En utilisant la modélisation de la distribution des espèces, ainsi que la diversité observée, on peut générer une carte de la diversité potentielle de ces espèces. En comparant les cartes de la diversité observée et potentielle, il est possible d'identifier les zones où on peut trouver plusieurs espèces de *Vasconcellea*, où il existe actuellement peu ou pas d'enregistrement des observations. Ces lacunes sont des zones d'intérêt particulier pour les missions de collecte de matériel génétique (Scheldeman *et al.*, 2007). L'analyse qui suit illustre comment utiliser DIVA-GIS pour générer une carte de la diversité potentielle et comparer les lacunes entre la diversité observée et le potentiel d'une espèce.

PROGRAMMES ET FICHIERS DE DONNEES A UTILISER DANS CETTE SECTION	
Programmes : <ul style="list-style-type: none"> • DIVA-GIS • Maxent et Java 	Fichiers de données : <p>Répertoire 6.4 <i>Gap analysis</i> (Analyse des lacunes)</p> <ul style="list-style-type: none"> • Vasconcellea.csv • Vasconcellea species (shp, shx, dbf) • Folder wclim_ams_5min (asc files))

Dans les sections précédentes, nous avons exploré des rasters de zones de distribution de la présence naturelle d'une seule espèce. Il est également possible de générer des rasters de la diversité potentielle en se basant sur les niches effectives de diverses espèces, en utilisant un groupe de rasters binaires avec les zones de distribution potentielle pour chaque espèce.

Note

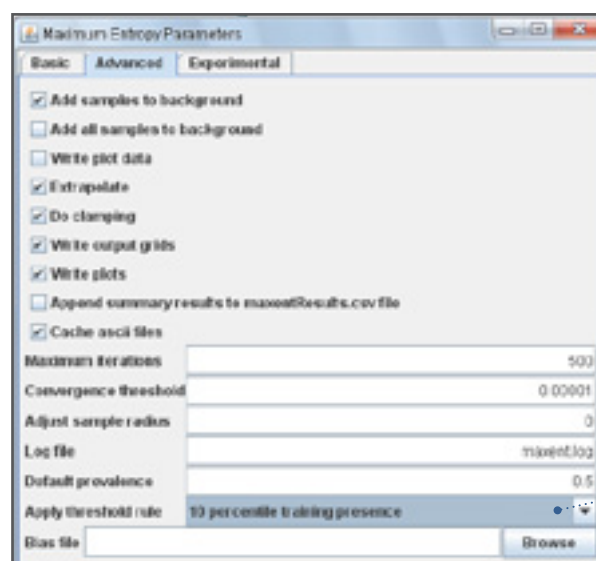
Pour que Maxent traite les fichiers raster environnementaux des conditions actuelles et futures (dans le cas présent, *wclim_2-5min* et *gcm_sea_2-5min*), tous les fichiers doivent avoir les mêmes paramètres en termes de propriétés du raster, de résolution et de position des coins ou sommets.

Etapas :

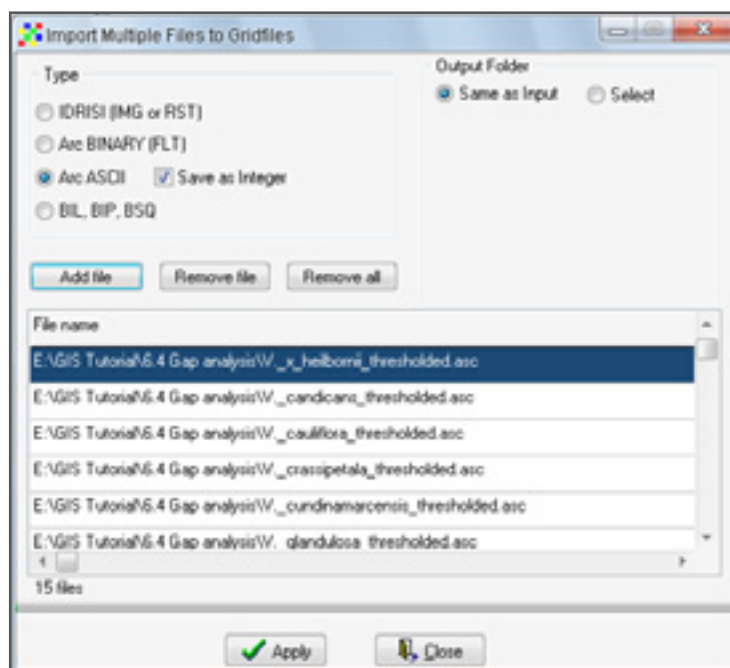
1. Utilisez Maxent pour générer des modèles de distribution naturelle potentielle des diverses espèces de *Vasconcellea*, en suivant l'explication donnée aux Etapes 1 à 8 de l'Analyse 6.2.1, où l'on a modélisé la distribution potentielle d'une espèce. Maxent peut aussi effectuer l'analyse pour plusieurs espèces en même temps si l'on a sauvegardé les points de présence de toutes les espèces dans le même fichier CSV (*.csv). C'est ici le cas pour les 15 espèces de *Vasconcellea* (fichier *Vasconcellea.csv*). Pour cette analyse, les modèles seront générés pour les espèces de *Vasconcellea* en se fondant sur les rasters des variables BioClim de 5 minutes de résolution pour l'Amérique latine et les Caraïbes ; ils se trouvent dans le répertoire *WCLIM_AMS_5min*.



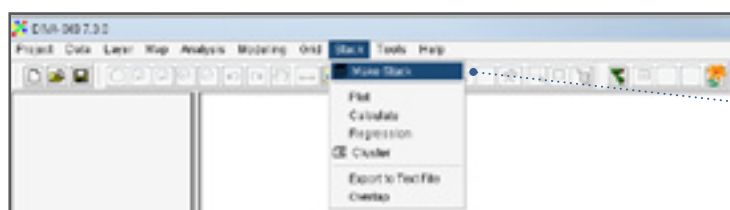
2. Pour générer les rasters binaires des aires de distribution potentielle à l'aide de Maxent, on procède de manière analogue à ce qui a été fait dans l'Etape 2 de l'Analyse 6.3.1 : dans la fenêtre *Advanced settings*, dans la case *Apply threshold rule*, sélectionnez l'option *10 percentile training presence*.



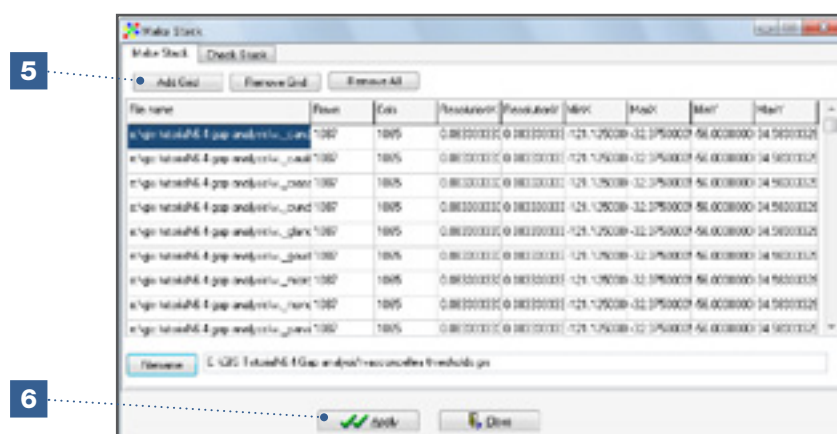
3. Importez les rasters binaires des aires de distribution potentielle dans DIVA-GIS, comme expliqué aux Etapes 13 à 19 de l'Analyse 6.2.1 (en utilisant *Import Multiple Files to Gridfile*).



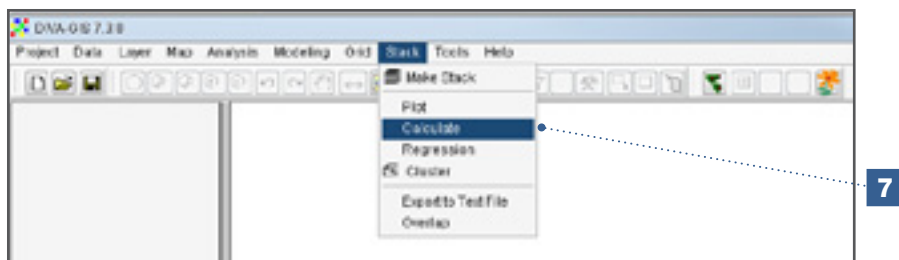
4. Dans le menu *Stack*, constituez un groupe des rasters binaires (*_thresholded.grd) en sélectionnant l'option *Make Stack*.



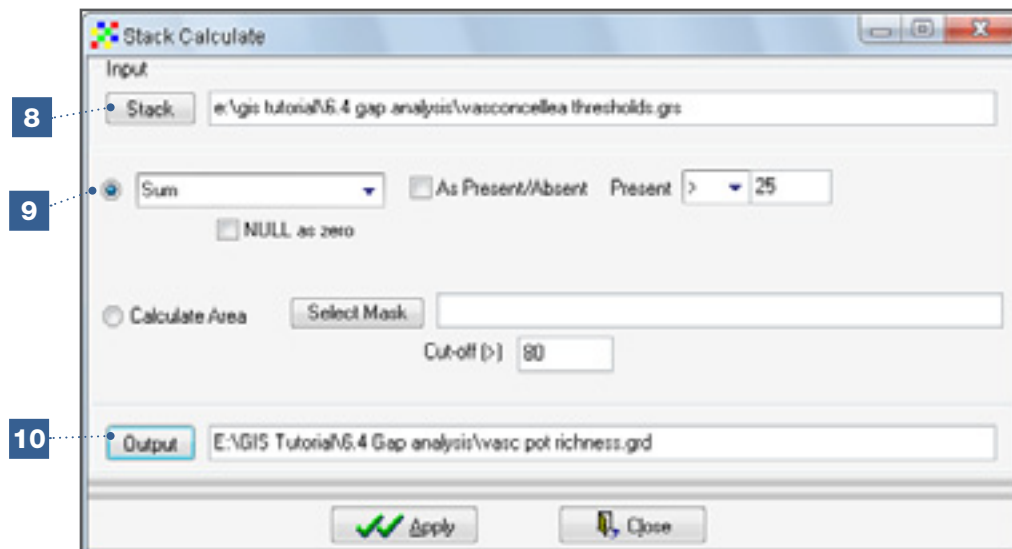
5. Indiquez les rasters que vous souhaitez inclure dans le groupe. Pour cette analyse, utilisez les fichiers de rasters binaires des espèces de *Vasconcellea* (*_thresholded.grd).
6. Assignez un nom au groupe, cliquez sur *Apply* puis sur *Close*.

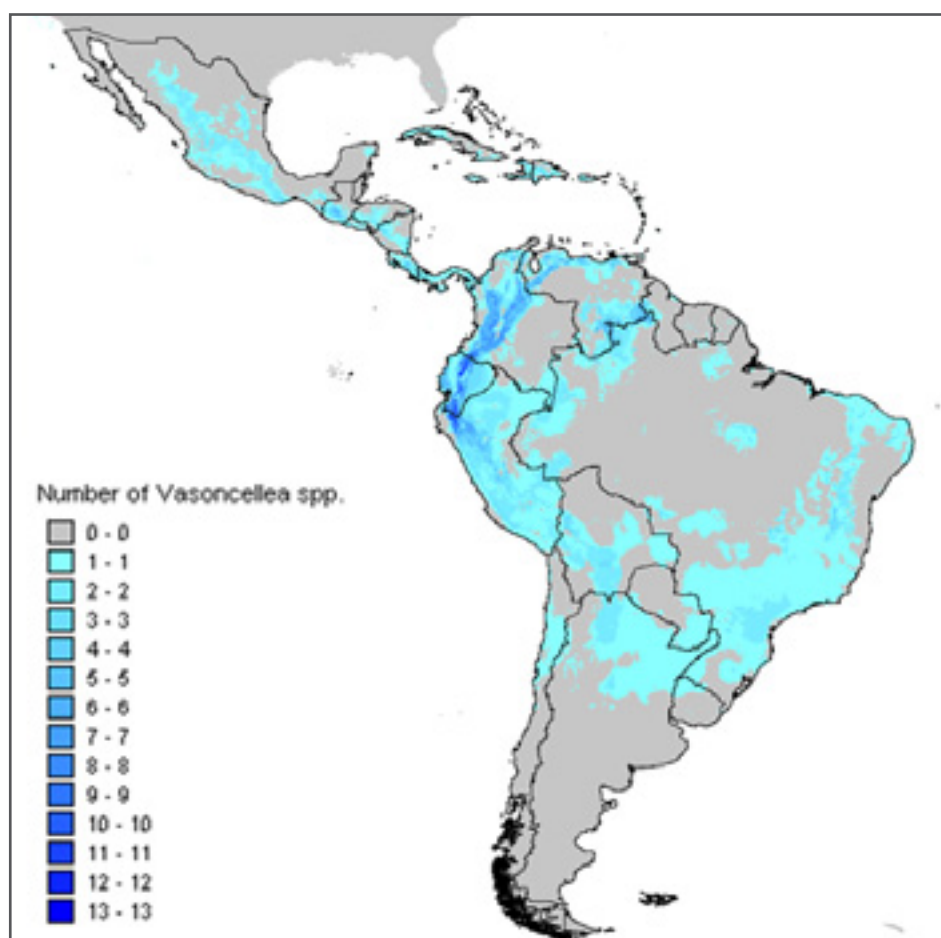


7. Pour estimer la richesse potentielle des espèces de *Vasconcellea* en Amérique latine, allez sur *Stack/Calculate* ; cette fonction additionne les rasters binaires représentant les aires potentielles de ces espèces.



8. Sélectionnez le groupe de rasters sur lequel vous souhaitez effectuer les calculs. Dans le cas présent, nous sélectionnons le groupe qui a été constitué à l'Etape 6.
9. Sélectionnez l'option *Sum* (addition) pour additionner les rasters binaires des aires potentielles de distribution des espèces de *Vasconcellea*.
10. Dans l'onglet *Output*, indiquez le nom du fichier raster qui sera généré.
11. Laissez les autres options à leurs valeurs par défaut.
12. Cliquez sur *Apply* pour lancer les calculs.





Après avoir édité la carte (comme expliqué au Chapitre 3), la carte de la diversité potentielle des espèces de *Vasconcellea* en Amérique latine devrait ressembler à celle qui est présentée ci-dessus. La plus grande zone pour la diversité attendue se trouve dans la partie Nord de la région andine, en particulier entre l'Equateur et le Pérou, et entre l'Equateur et la Colombie.

Analyse de lacunes dans DIVA-GIS

Pour identifier les lacunes dans la distribution des espèces, on compare les rasters de richesse potentielle et observée des espèces de *Vasconcellea*. Pour commencer l'analyse des lacunes, vous aurez besoin :

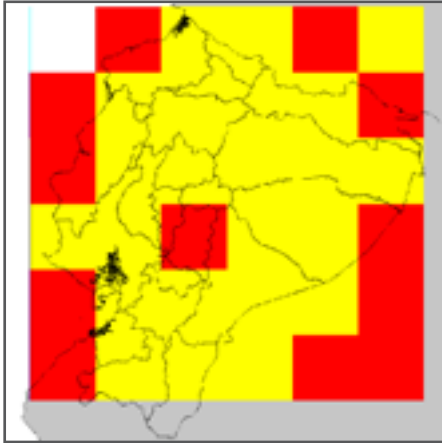
- Du raster des aires de distribution naturelle potentielle de l'espèce.
- D'un raster de la distribution observée de l'espèce qui ait les mêmes propriétés que celui de la distribution potentielle.

Taille de cellule des rasters

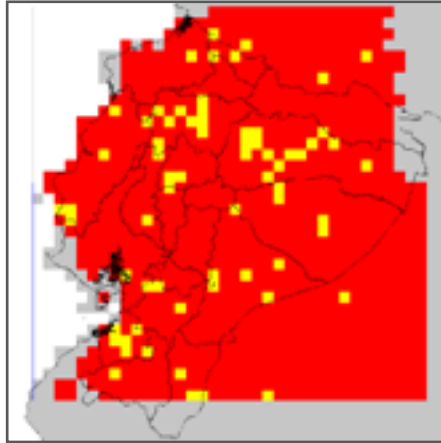
Nous avons déjà examiné l'importance de la taille des cellules pour détecter la structuration spatiale dans la Section 5.1. La taille des cellules a aussi son importance pour identifier les lacunes. Une taille réduite sera utile pour détecter des lacunes à une échelle locale, mais si l'on maintient une petite taille de cellules pour l'étude des lacunes sur une large zone d'étude (comme un pays ou une région), seul un nombre limité de cellules contiendront des points de présence car la plupart des cellules n'auront pas fait l'objet d'un échantillonnage. Cela conduira évidemment à identifier de nombreuses lacunes présumées et compliquera le choix des sites prioritaires pour des collectes complémentaires. Pour les études à un niveau national ou régional,

[...suite page suiv.]

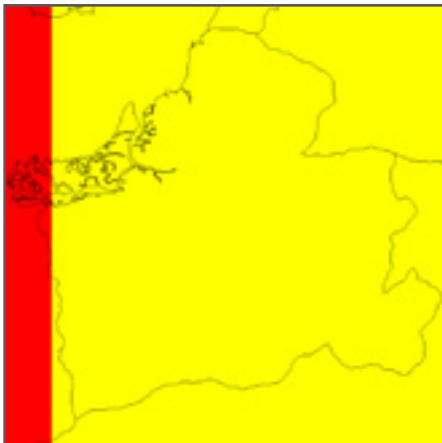
une plus grande taille de cellule est préférable, mais c'est à l'analyste de choisir la taille optimale pour son étude. Les images ci-dessous montrent une analyse de lacune en Equateur pour *Vasconcellea microcarpa* effectuée avec plusieurs tailles de cellules différentes. Les figures illustrent le fait que les rasters de 10 minutes de résolution peuvent être utiles pour l'analyse de lacunes à l'échelle locale (comme la province), tandis qu'une résolution moins fine (1 degré) est plus appropriée pour l'analyse au niveau national.



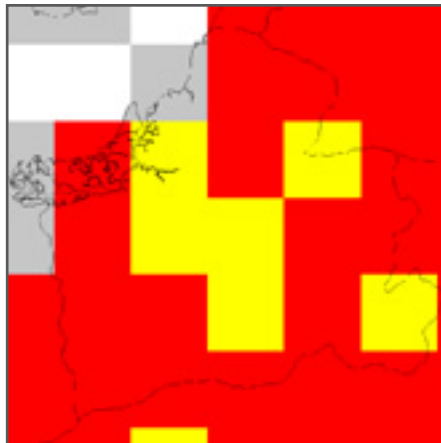
Analyse de lacune au niveau national
avec 1 degré de résolution



Analyse de lacune au niveau national
avec 10 minutes de résolution



Analyse de lacune au niveau provincial
avec 1 degré de résolution



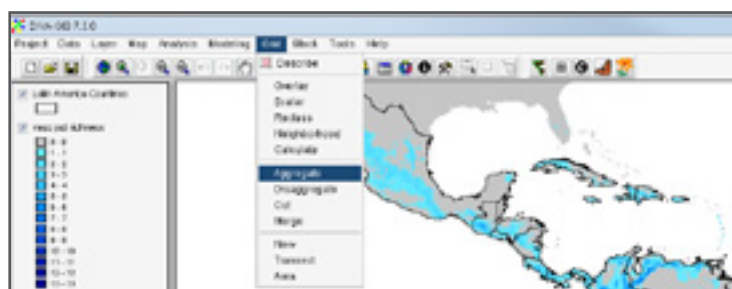
Analyse de lacune au niveau provincial
avec 10 minutes de résolution

- Jaune : zones montrant une présence potentielle où l'espèce a effectivement été observée
- Rouge : aires de distribution potentielle, où l'espèce n'a pas été observée
- Gris : zones situées hors de la distribution observée et potentielle de l'espèce.

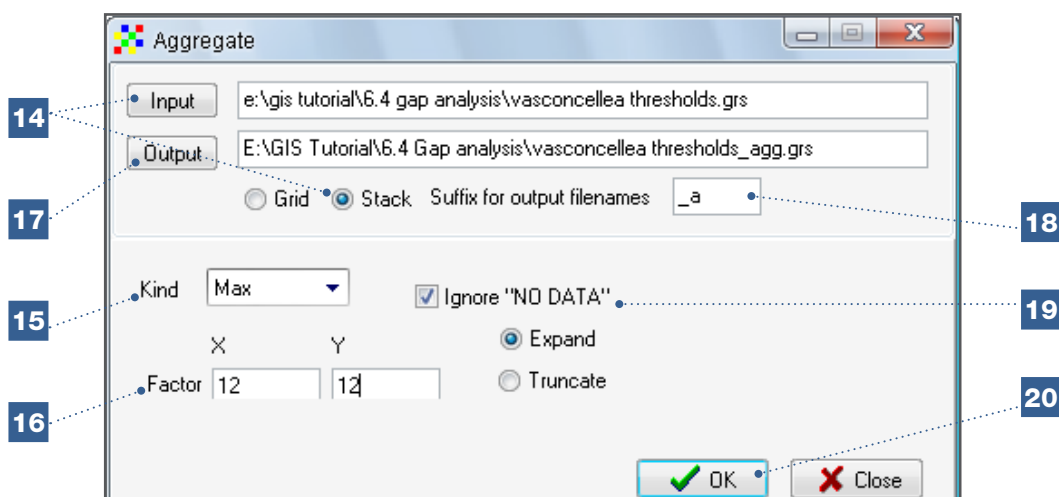
Au vu de la capacité limitée d'identifier efficacement les lacunes lorsqu'on utilise des rasters de haute résolution, il est important d'accroître la taille des cellules des rasters binaires par rapport aux rasters de la diversité potentielle de *Vasconcellea* générés précédemment.

Etapas (poursuite de la section précédente) :

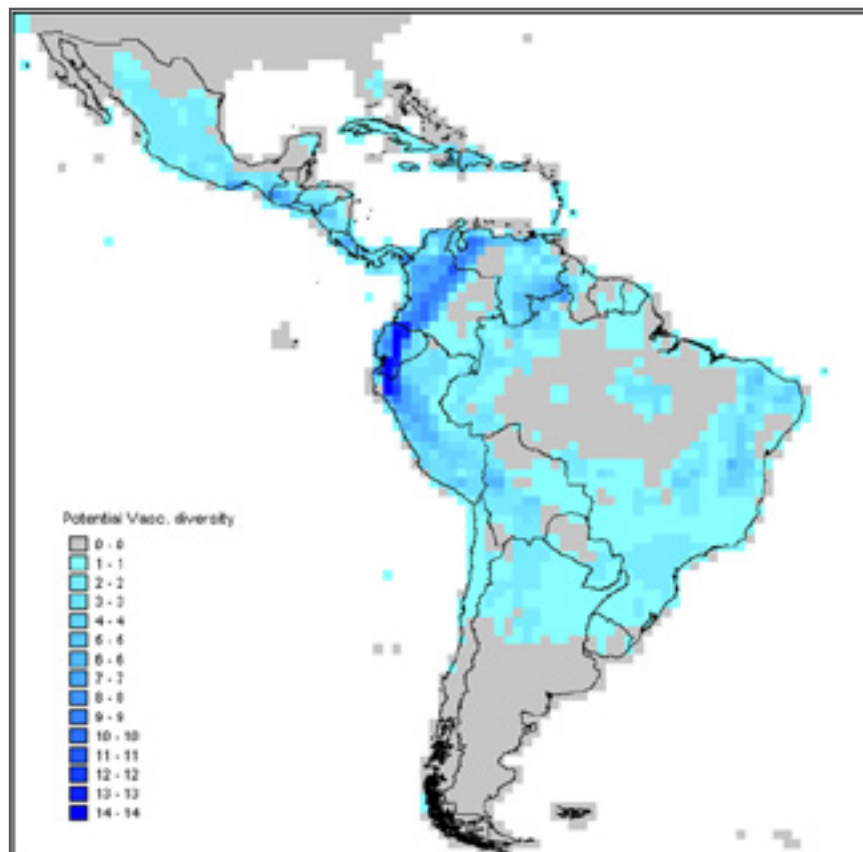
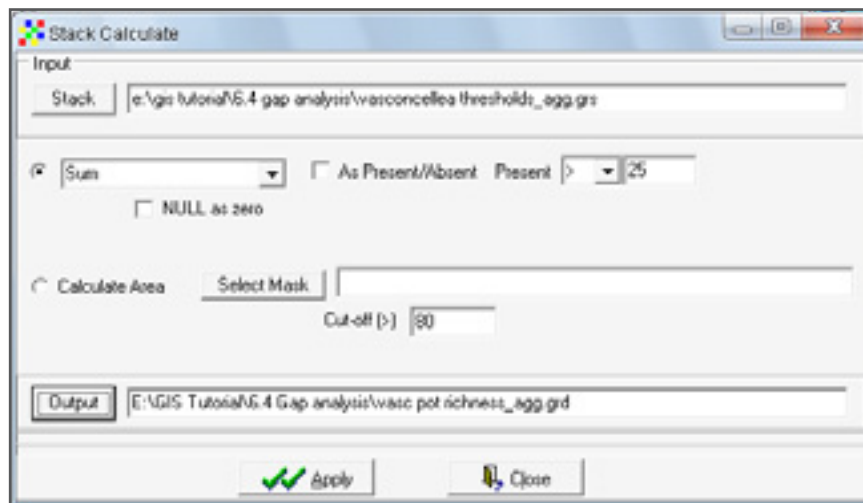
13. Dans le menu *Grid*, sélectionnez l'option *Aggregate*.



14. Dans la case *Input*, sélectionnez le groupe de rasters dont il faut accroître la taille de cellule. Dans cette analyse, sélectionnez le groupe contenant les rasters binaires (*_thresholded) des aires de distribution potentielle de chaque espèce de *Vasconcellea* ; c'est le groupe généré à l'Etape 4.
15. Dans l'option *Kind*, sélectionnez le type de calcul que vous souhaitez mettre en œuvre pour combiner les valeurs des cellules plus petites lorsque vous les fusionnez en une cellule plus grande. Pour cette analyse, sélectionnez l'option *Max* afin que la cellule agrégée utilise la valeur maximale des cellules qui la composent. Si au moins une des cellules composantes a une valeur d'un (1), la cellule agrégée aura aussi une valeur d'un (1), indiquant que l'espèce a été observée dans cette cellule.
16. Dans la case *Factor*, indiquez jusqu'où vous voulez étendre la taille de la cellule du raster. Dans cet exemple, la résolution de 5 minutes sera convertie en une résolution d'un (1) degré. Pour faire cela, les cellules doivent être élargies d'un facteur de 12 (*Factor of 12*) dans les deux directions (puisque 60 minutes équivalent à un [1] degré).
17. Indiquez dans la case *Output* le nom du nouveau fichier raster.
18. Vous avez la possibilité d'ajouter un suffixe aux noms des rasters agrégés. Dans cette analyse, utilisez l'option par défaut : *_a*.
19. Gardez les options *Ignore « NO DATA »* et *Expand* sélectionnées.
20. Cliquez sur *OK* pour lancer le calcul.



21. Après avoir augmenté la taille de cellules de tous les rasters du groupe, répétez l'opération à l'aide de l'option *Calculate* dans le menu *Stack* (Étapes 7 à 12 ci-dessus) pour déterminer combien d'espèces de *Vasconcellea* sont potentiellement présentes dans chaque cellule.



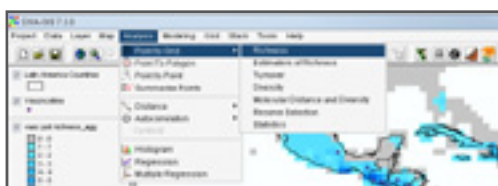
Après quelques modifications et un travail d'édition, la carte de diversité potentielle des espèces de *Vasconcellea*, avec les cellules agrégées, devrait ressembler à la carte ci-dessus.

Bien qu'il puisse sembler inefficace de générer d'abord les rasters binaires de distribution potentielle dans Maxent en utilisant des données climatiques détaillées (petite taille de cellule) pour les regrouper ensuite dans une cellule de plus grande taille, plutôt que de générer directement les rasters binaires d'occurrence potentielle avec des cellules plus grandes (avec des données climatiques de moindre résolution), il est recommandé

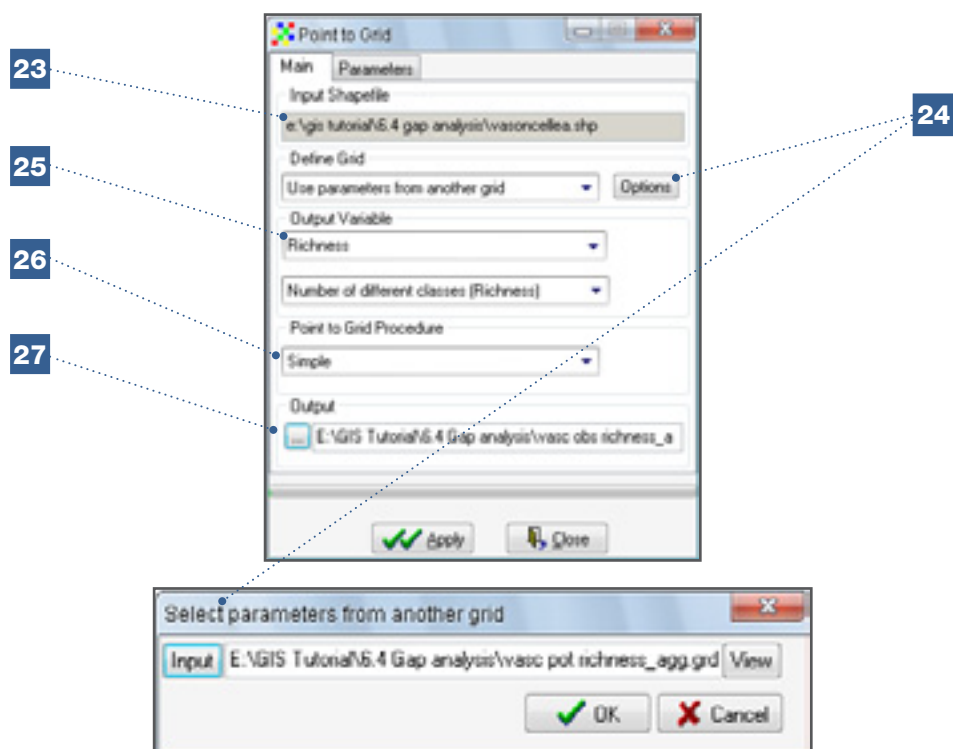
de procéder en deux temps comme décrit ici du fait que, dans Maxent, les données climatiques de haute résolution génèrent habituellement des prédictions plus exactes de la distribution potentielle que des données de moindre résolution.

Développement du raster de la richesse observée des espèces de *Vasconcellea* (voir aussi la Section 5.1)

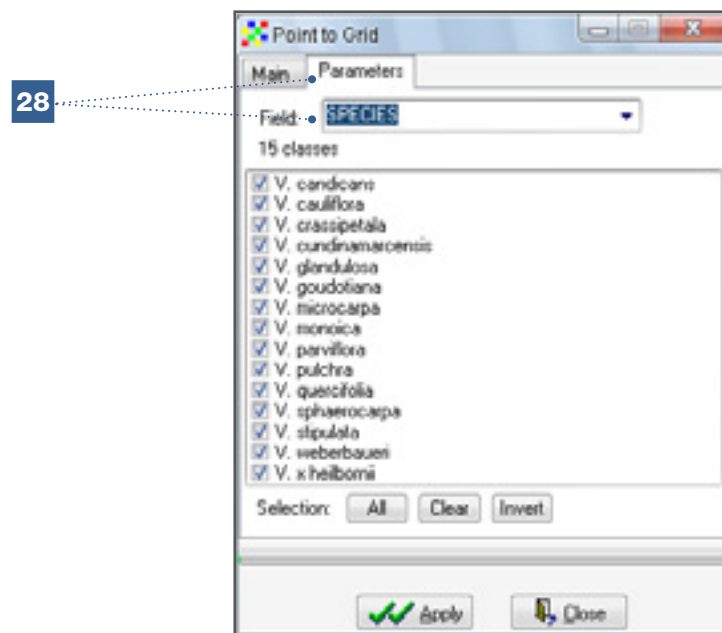
22. Sélectionnez, dans la légende, le fichier de type vecteur (*.shp) qui contient les points de présence d'une espèce ou d'une groupe d'espèces à analyser. Dans cette analyse, on sélectionnera les espèces de *Vasconcellea* (*Vasconcellea species.shp*).
23. Dans le menu *Analysis*, allez à *Point to Grid/Richness*.



24. Dans la fenêtre *Define Grid*, sélectionnez l'option *Use parameters from another grid* pour assurer que les paramètres du raster de la diversité observée (celui que vous êtes en train de créer) sont équivalents à ceux du raster de la diversité potentielle. Pour ce faire, dans *Options*, sélectionnez le raster de richesse potentielle créé au cours des étapes précédentes.
25. Sélectionnez *Richness* et *Number of different classes (Richness)* comme variables de sortie (*Output Variable*).
26. Cochez l'option *Simple* dans la case *Point to Grid Procedure*.
27. Sélectionnez le bouton à gauche de la case *Output (...)* pour indiquer le nom et l'emplacement du raster résultant.



28. Dans le menu *Parameters*, allez sur la fenêtre *Field* et sélectionnez l'élément pour lequel vous souhaitez réaliser une analyse de diversité. Dans le cas présent, sélectionnez *SPECIES*.



29. Cliquez sur *Apply* pour lancer les calculs.



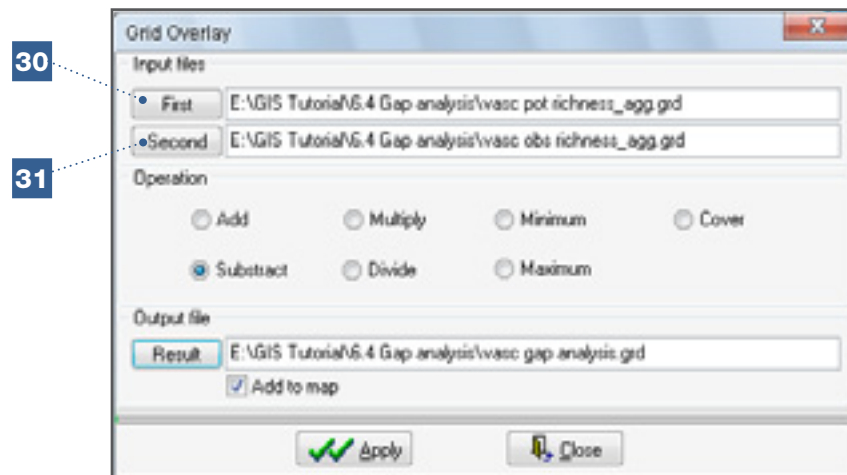
Après avoir achevé ce processus, la carte de la diversité observée des espèces de *Vasconcellea* devrait ressembler à la version non éditée de la carte de la diversité créée à la Section 5.1.

Identification de lacunes par comparaison des distributions potentielle et observée

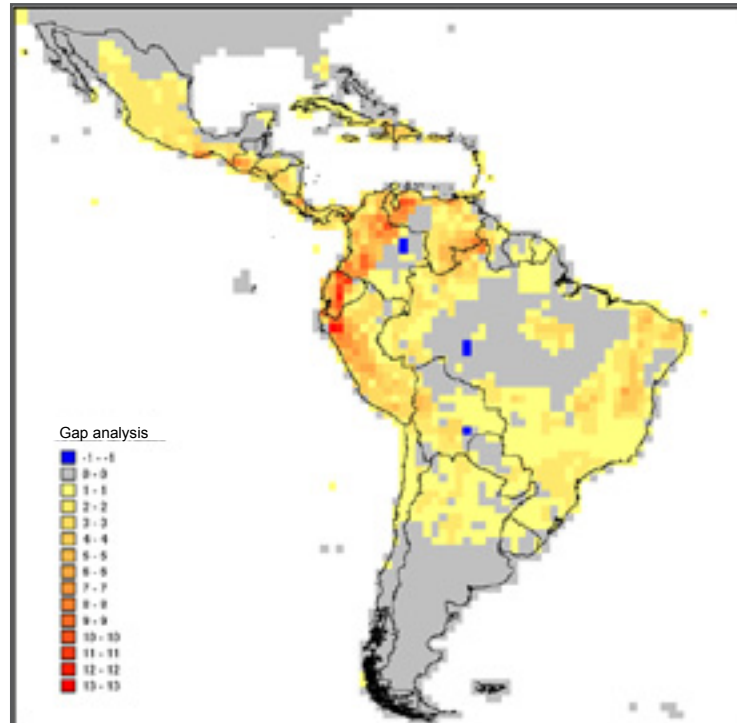
Après avoir généré les rasters des diversités potentielle et observée avec les paramètres indiqués, allez à l'option *Grid Overlay* et soustrayez les valeurs des deux rasters (comme dans les Etapes 10 à 17 de l'Analyse 6.3.1).

30. Dans la fenêtre *First*, sélectionnez le raster de la diversité potentielle.

31. Dans la fenêtre *Second*, sélectionnez le raster de la diversité observée.



Résultat final :



La carte de la diversité observée des espèces de *Vasconcellea* devrait ressembler à celle qui est présentée ci-dessus. La couleur rouge intense indique les zones où aucune collecte n'a encore été effectuée, même si ces localités sont adaptées à beaucoup d'espèces de *Vasconcellea*. En fait, en plus de l'espèce ou des deux espèces observées dans ces cellules,

12 ou 13 autres espèces devraient aussi s'y trouver. Ces emplacements sont situés dans le Nord du Pérou et dans la zone de transition équatoriale entre les régions andine et amazonienne. D'autres aires de distribution potentielle pour la collecte se trouvent en Colombie et au Venezuela. Certaines cellules, indiquées en bleu, ont une valeur de moins un (-1), ce qui signifie que l'on a observé plus d'espèces de *Vasconcellea* dans ces zones que prédit par le modèle de diversité potentielle.

Note

Cette section présente comment conduire une analyse de lacunes pour plusieurs espèces, mais on peut aussi réaliser une analyse de lacunes pour une seule espèce. Dans ce cas, vous devez superposer les couches de la richesse potentielle et observée d'une seule espèce (en prenant en compte le fait que la carte finale ne montrera que les lacunes de collecte et les zones où l'on a effectué des observations atypiques).

Références

- Anderson RP, Lew D, Peterson AT. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modeling* 162:211–232.
- Araújo MB, Pearson RG, Thuiller W, Erhard M. 2005. Validation of species–climate impact models under climate change. *Global Change Biology* 11:1504–1513.
- Elith, J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton J.Mc, Townsend C, Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
- FAO. 2007. Ecocrop [on line]. Available from: <http://ecocrop.fao.org/ecocrop/srv/en/home>. Lien vérifié en juin 2012.
- Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27:861– 874.
- Fjeldså J. 2002. *Polylepis* forests Vestiges of a vanishing ecosystem in the Andes. *Ecotropica* 8:111–123.
- Guarino L, Jarvis A, Hijmans RJ, Maxted N. 2002. Geographic Information Systems (GIS) and the Conservation and Use of Plant Genetic Resources. In: Engels JMM, Ramanatha Rao V, Brown AHD, Jacson MT, éditeurs. Managing plant genetic diversity. International Plant Genetic Resources Institute (IPGRI) Rome, Italie. pp. 387–404.
- Hernandez PA, Graham CH, Master LL, Albert DL. 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* 29:773–785.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Hutchinson GE. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:415–427.

- IPCC. 2007. Climate Change 2007: Synthesis Report. Cambridge University Press, New York, USA.
- Jarvis A, Williams K, Williams BD, Guarino L, Caballero PJ, Mottram G. 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendt) in Paraguay. Genetic Resources and Crop Evolution 52: 671–682.
- Jones PG, Díaz W, Cock JH. 2005. Homologue: A computer system for identifying similar environments throughout the tropical World. Version Beta a.0. Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombie.
- Liu C, Berry, PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385–393.
- Pearson RG, Dawson TP. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography.12:361–371.
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modeling 190:231–259.
- Philips J. 2009. A Brief Tutorial on Maxent [on line]. Available from: <http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>. Lien vérifié en Juin 2012.
- Puliam HR. 2000. On the relationship between niche and distribution. Ecology Letters 3:349–361.
- Scheldeman X, Willems L, Coppens D'eeckenbrugge G, Romeijn-Peeters E, Restrepo MT, Romero Motoche J, Jimenez D, Lobo M, Medina CI, Reyes C, Rodriguez D, Ocampo JA, Van Damme P, Goetghebeur P. 2007. Distribution, diversity and environmental adaptation of highland papaya (*Vasconcellea* spp.) in tropical and subtropical America. Biodiversity and Conservation 16(6):1867–1884.
- USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, Maryland. Available from: <http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?28462>. Lien vérifié en juin 2012.
- van Zonneveld M, Koskela J, Vinceti B, Jarvis A. 2009a. Impact of climate change on the distribution of tropical pines in Southeast Asia. Unasylva 60 (231/232):24–28.
- van Zonneveld M, Jarvis A, Dvorak W, Lema G, Leibing C. 2009b. Climate change impact predictions on *Pinus patula* and *Pinus tecunumanii* populations in Mexico and Central America. Forest Ecology and Management 257(7):1566–1576.
- Willis F, Moat J, Paton A. 2003. Defining a role for herbarium data in Red List assessments: a case study of *Plectranthus* from eastern and southern tropical Africa. Biodiversity and Conservation 12:1537–1552.

Annexe

Liste des commandes en anglais et leurs équivalents en français

Liste des commandes en anglais utilisées dans le manuel, avec leurs équivalents en français	
Commandes utilisées en anglais	Equivalent en français
<i>Absence</i>	Absence
<i>Add file</i>	Ajouter un fichier
<i>Add Layer</i>	Ajouter une couche
<i>Additional Classes</i>	Classes additionnelles
<i>Advanced</i>	Avancé
<i>Aggregate</i>	Agréger
<i>All</i>	Tous
<i>Altitude</i>	Altitude
<i>Analysis</i>	Analyse
<i>Annual Precipitation</i>	Précipitations annuelles
<i>Applied threshold rule</i>	Règle d'application de seuil
<i>Apply</i>	Appliquer
<i>Attributes</i>	Attributs
<i>Auto complete</i>	Saisie automatique
<i>Basic colors</i>	Couleurs de base
<i>Basic Elements</i>	Éléments de base
<i>Binary</i>	Binaire
<i>Blue</i>	Bleu
<i>Browse</i>	Naviguer
<i>Calculate</i>	Calculer
<i>Cell Size</i>	Taille de la cellule
<i>Chart</i>	Graphique
<i>Check Coordinates</i>	Vérifier les coordonnées
<i>Circular Neighborhood</i>	Voisinage circulaire
<i>Classes</i>	Classes
<i>Clear Selection</i>	Annuler la sélection
<i>Climate</i>	Climat
<i>Climate Data to Map</i>	Cartographie des données climatiques
<i>Climate Envelope</i>	Niche climatique
<i>Close</i>	Fermer
<i>Cluster</i>	Cluster (ou groupe)
<i>Coefficient of Variation</i>	Coefficient de variation
<i>Columns</i>	Colonnes

Commandes utilisées en anglais	Equivalent en français
<i>Commas</i>	Virgules
<i>Complementarity</i>	Complémentarité
<i>Copied formulas</i>	Formules copiées
<i>Copy</i>	Copier
<i>Country</i>	Pays
<i>Create a New Grid</i>	Créer une nouvelle grille (ou raster)
<i>Cumulative</i>	Cumulatif
<i>Cut</i>	Couper
<i>Data</i>	Données
<i>decimal degrees</i>	Degrés décimaux
<i>Degrees</i>	Degrés
<i>Define Grid</i>	Définir la grille (ou raster)
<i>Design</i>	Habiller
<i>Disaggregate</i>	Désagréger
<i>Distance</i>	Distance
<i>Distribution modeling</i>	Modélisation de la distribution
<i>Domain</i>	Domaine
<i>Draw</i>	Dessiner
<i>Draw rectangle</i>	Dessiner un rectangle
<i>Edit</i>	Editer
<i>Edit Values</i>	Editer les valeurs
<i>Environmental layers</i>	Couches environnementales
<i>Excel spreadsheet</i>	Feuille de calcul Excel
<i>Expand</i>	Agrandir
<i>Export Gridfile</i>	Exporter le fichier raster
<i>Export Project</i>	Exporter le projet
<i>Export Shapefile</i>	Exporter le fichier vecteur
<i>Extract Value by Points</i>	Extraire la valeur de chaque point
<i>Factor</i>	Facteur
<i>Field</i>	Champ
<i>Fields of Shape of Point</i>	Champs de forme d'un point
<i>Fields of Shape of Polygon</i>	Champs de forme d'un polygone
<i>File Extract Values by Points</i>	Archiver les valeurs extraites par points
<i>Final</i>	Final
<i>Find</i>	Chercher
<i>Finish</i>	Terminer
<i>Folder</i>	Répertoire
<i>Formulas</i>	Formules
<i>Frequency</i>	Fréquence
<i>From</i>	De
<i>From Climate data</i>	A partir des données climatiques
<i>From Excel</i>	A partir d'Excel
<i>From Grid or Stack</i>	A partir d'un raster ou d'un stack
<i>From same grid cell</i>	A partir de la même cellule du raster
<i>Go to download website</i>	Aller au site de téléchargement
<i>Green</i>	Vert
<i>Grid</i>	Grille (raster)

Commandes utilisées en anglais	Equivalent en français
<i>Grid Options</i>	Options de grille (raster)
<i>Grid Overlay</i>	Superposition de grilles (rasters)
<i>Group by</i>	Grouper par
<i>Highlight</i>	Souligner, mettre en exergue
<i>Histogram</i>	Histogramme
<i>History</i>	Histoire
<i>I accept the agreement</i>	J'accepte l'accord
<i>Import Multiple Files to Gridfile</i>	Importer plusieurs fichiers dans un fichier grille (raster)
<i>Import Points</i>	Importer des points
<i>Import to Gridfile</i>	Importer dans un fichier grille (raster)
<i>Importing observation data</i>	Importation des données d'observation
<i>Information</i>	Information
<i>Input files</i>	Fichiers d'entrée
<i>Install</i>	Installer
<i>Integration Tab</i>	Option (onglet) d'intégration
<i>International</i>	International
<i>Latitude</i>	Latitude
<i>Layer</i>	Couche
<i>LEFT</i>	Gauche
<i>Legend</i>	Légende
<i>Lines</i>	Lignes
<i>Logistic threshold</i>	Seuil logistique
<i>Longitude</i>	Longitude
<i>Main</i>	Principal
<i>Make CLM files</i>	Faire des archives CLM
<i>Make Map</i>	Faire une carte
<i>Make pictures of predictions</i>	Faire des images de prédictions
<i>Make Stack</i>	Former un stack (groupe de rasters)
<i>Make valid ESRI grid names</i>	Valider les noms de rasters ESRI
<i>Manual</i>	Manuel
<i>Many Classes</i>	De nombreuses classes
<i>Map</i>	Carte
<i>Map Units</i>	Unités cartographiques
<i>Maximum temperature</i>	Température maximale
<i>Mean Annual Temperature</i>	Température moyenne annuelle
<i>Mean Monthly Temperature Range</i>	Ecart moyen de température mensuel
<i>Menu bar</i>	Barre de menu
<i>Merge</i>	Fusion de plusieurs rasters en un seul fichier
<i>Minimum temperature</i>	Température minimale
<i>Minutes</i>	Minutes
<i>Modeling</i>	Modélisation
<i>Multiple Files</i>	Fichiers multiples
<i>Multiply</i>	Multiplier
<i>Neighborhood Options</i>	Options de voisinage
<i>Next</i>	Suivant
<i>NoData Transparent</i>	Cellules transparentes = absence de données

Commandes utilisées en anglais	Equivalent en français
<i>Notepad</i>	Bloc notes
<i>Number of Observations</i>	Nombre d'observations
<i>OK</i>	D'accord
<i>Omission</i>	Omission
<i>Open</i>	Ouvrir
<i>Options</i>	Options
<i>Outliers</i>	Valeurs atypiques
<i>Output File</i>	Fichier de sortie ou de résultats
<i>Overlay</i>	Superposer
<i>Pan To</i>	Déplacer vers
<i>Parameters</i>	Paramètres
<i>Paste special</i>	Collage spécial
<i>Period</i>	Période
<i>Places</i>	Lieux
<i>Point to Grid</i>	Créer un raster à partir d'un fichier de points
<i>Points</i>	Points
<i>Points do not match relations</i>	Les points ne correspondent pas aux liens
<i>Precipitation</i>	Précipitations
<i>Predict</i>	Prédire
<i>Predicted Area</i>	Aire prédite
<i>Presence</i>	Présence
<i>Project</i>	Projet
<i>Projection</i>	Projection
<i>Properties</i>	Propriétés
<i>Prune</i>	Elaguer un graphique (par exemple un dendrogramme)
<i>Ramp</i>	Gradient
<i>Range</i>	Ecart ou intervalle de valeurs
<i>Rarefaction</i>	Raréfaction
<i>Read From File</i>	Lire à partir d'un fichier
<i>Read from Layer</i>	Lire à partir d'une couche
<i>Reclass</i>	Reclassifier
<i>Rectangle</i>	Rectangle
<i>Red</i>	Rouge
<i>Regression</i>	Régression
<i>Remove duplicate presence records</i>	Supprimer les enregistrements doublons
<i>Remove duplicates: From same grid cell</i>	Supprimer les doublons : de la même cellule du raster
<i>Remove Layer</i>	Supprimer la couche
<i>Rename Master</i>	Renommer des jeux de fichiers
<i>Replace</i>	Remplacer
<i>Reset Legend</i>	Restaurer la légende
<i>Richness</i>	Richesse
<i>RIGHT</i>	Droite
<i>Run</i>	Fonctionner, lancer, faire fonctionner
<i>Same as Input</i>	Identique à l'entrée
<i>Samples</i>	Echantillons

Commandes utilisées en anglais	Équivalent en français
<i>Save (as)</i>	Sauvegarder (sous)
<i>Save as Integer</i>	Sauvegarder au format entier
<i>Save to Shapefile</i>	Sauvegarder au format vecteur
<i>Seconds</i>	Secondes
<i>Select by query</i>	Sélectionner par recherche
<i>Select by values</i>	Sélectionner par valeur
<i>Select Color Scheme</i>	Sélectionner la configuration des couleurs
<i>Select Features</i>	Sélectionner les caractéristiques
<i>Select Records</i>	Sélectionner les enregistrements
<i>Select Setup Language</i>	Sélectionner la langue du système
<i>Selection to New Shapefile</i>	Sélection vers un nouveau fichier de vecteurs
<i>Sequence</i>	Séquence
<i>Settings</i>	Configuration
<i>Single</i>	Individuel
<i>Single File</i>	Fichier individuel
<i>Solid Fill</i>	Remplissage solide
<i>Stack</i>	Groupe de rasters
<i>Standardized Sample Size</i>	Taille d'échantillon standardisée
<i>Start data</i>	Données de départ
<i>Startup Menu</i>	Menu de démarrage
<i>Statistic</i>	Statistique
<i>Status bar</i>	Barre d'état
<i>Sum</i>	Somme
<i>Suppress similar visual warnings</i>	Supprimer les avertisseurs visuels similaires
<i>Symbol</i>	Symbole
<i>Table</i>	Table
<i>Temporary Places</i>	Lieux temporaires
<i>To</i>	Vers
<i>Toolbar</i>	Barre d'outils
<i>Tools</i>	Outils
<i>Unique</i>	Unique
<i>Use parameters from another grid</i>	Utiliser des paramètres d'un autre raster
<i>Values</i>	Valeurs
<i>Variable</i>	Variable
<i>Welcome</i>	Bienvenue
<i>Worksheet</i>	Feuille de calcul
<i>Zoom in</i>	Zoom avant
<i>Zoom out</i>	Zoom arrière
<i>Zoom To</i>	Zoom vers
<i>Zoom To active layer</i>	Zoom vers la couche active
<i>Zoom To full extent</i>	Zoom pour visualiser toutes les couches
<i>Zoom Tools</i>	Outils de zoom

Bioversity International est le nom sous lequel opèrent l'Institut international des ressources phytogénétiques (IPGRI) et le Réseau international pour l'amélioration de la banane et de la banane plantain (INIBAP). Bioversity est un membre du Consortium du CGIAR. Le CGIAR est un partenariat mondial de recherche agricole pour un futur sans faim. Cette recherche est menée par 15 centres, membres du Consortium du CGIAR, en étroite collaboration avec des centaines de partenaires, www.cgiar.org

Bioversity International
Via dei Tre Denari, 472/a,
00057 Maccarese, Rome, Italie
Tél. : (39) 0661181

© Bioversity International, 2012

ISBN 978-92-9043-917-2