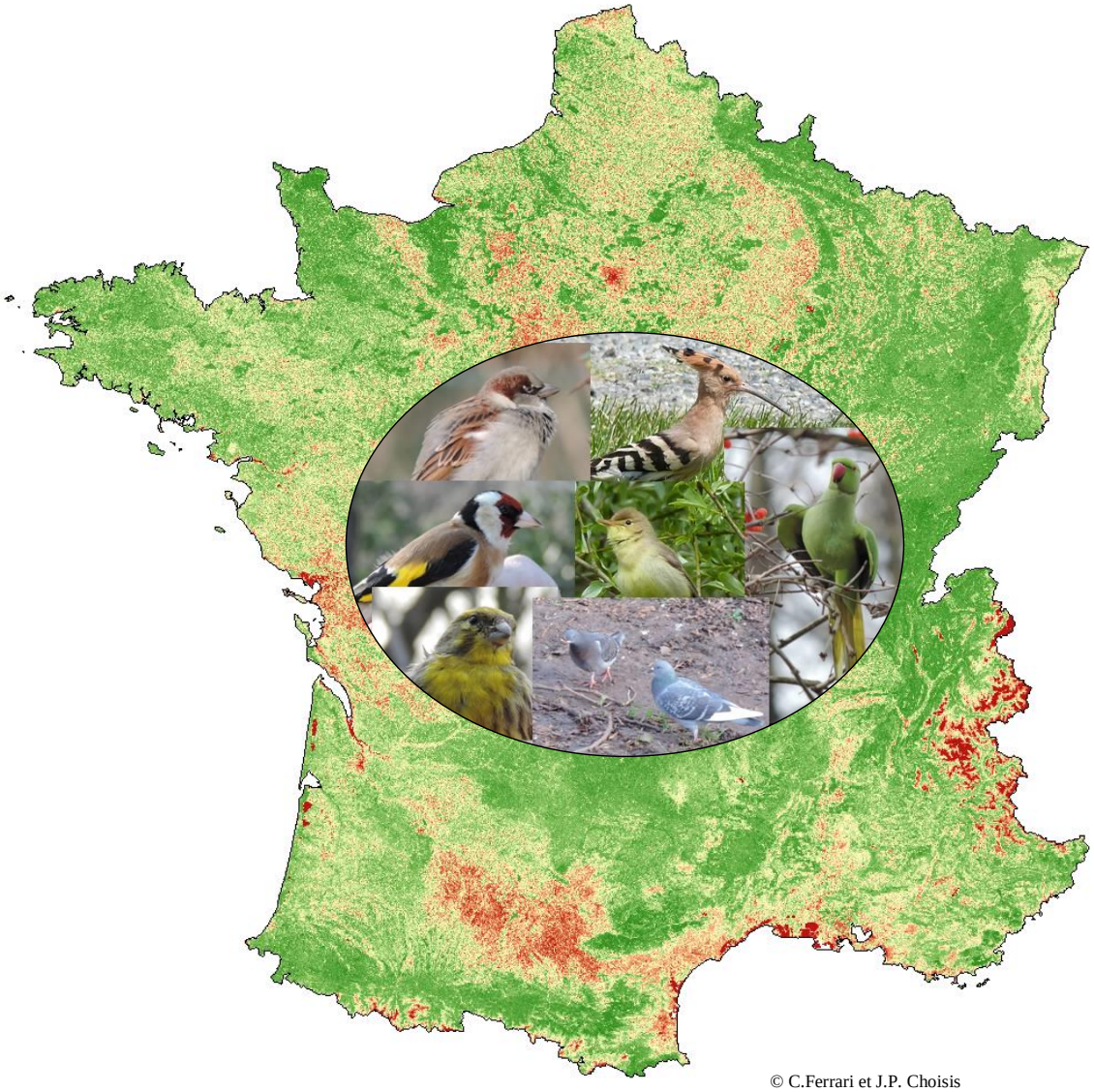
Évaluation de la diversité aviaire à large échelle à partir de séries temporelles d'images satellites



© C.Ferrari et J.P. Choisis

Mémoire de certificat de spécialité : Gestion des Milieux Naturels

Solenne LEFEVRE 2013 – 2014

Tuteur enseignant : C. Piedallu
Maîtres de stage : D. Sheeren et S. Bonthoux

Sources de l'illustration de la page de couverture

Carte de France : NDVI issu du produit MOD13Q1 de MODIS, synthèse du 14 septembre 2010

Photographies :
En haut à gauche : Moineau domestique, C.Ferrari
En haut à droite : Huppe fasciée, J.P. Choisis
Au centre à gauche : Chardonneret élégant, C. Ferrari
Au centre et milieu : Hypolaïs polyglotte, J.P. Choisis
A droite : Perruche à collier, C. Ferrari
En bas à gauche : Serin cini, C.Ferrari
En bas au milieu : Pigeons biset, C.Ferrari

AgroParisTech, Institut des
Sciences et Industries du Vivant
et de l'Environnement

DYNAFOR, Unité mixte de
recherche de l'Institut National
de Recherche Agronomique

Évaluation de la diversité aviaire à large échelle à partir de séries temporelles d'images satellites

Mémoire de certificat de spécialité : Gestion des Milieux Naturels

Solenne LEFEVRE 2013 – 2014

Tuteur enseignant : C. Piedallu
Maîtres de stage : D. Sheeren et S. Bonthoux

Résumé

Aujourd'hui, on assiste à l'émergence d'une nouvelle approche dans l'utilisation de la télédétection pour construire des modèles espèces/paysage : contrairement à la démarche classique qui consiste à établir des cartes d'occupation des sols, cette démarche alternative consiste à construire des variables explicatives directement issues d'une représentation paysagère continue.

Dans cette étude, nous étudions à l'échelle de la France l'effet de la représentation paysagère obtenue par télédétection sur la modélisation des relations oiseaux/habitat. D'abord, nous comparons la performance de modèles intégrant deux représentations différentes de l'état et la diversité des habitats à un instant donné : représentation discrète (CORINE Land Cover) versus images non classées (indices de végétation notamment). Ensuite, nous comparons les résultats de ces modèles avec d'autres intégrant une représentation fonctionnelle des habitats. Cette représentation est obtenue à partir d'une classification de la signature temporelle des habitats issue d'une série d'images NDVI traduisant la phénologie de la végétation.

Les relevés d'oiseaux sont issus du programme national STOC (Suivi Temporel des Oiseaux Communs) à partir desquels sont calculées plusieurs variables réponses (richesse de groupes d'oiseaux spécialistes ou généralistes). Les images utilisées correspondent aux synthèses NDVI à 16 jours du capteur MODIS (250 m de résolution spatiale).

Les résultats obtenus montrent un impact important de la représentation paysagère sur la performance des modèles. On observe notamment que les images non classées offrent toujours un pouvoir explicatif plus élevé que les données classées (gain de déviance expliquée de 10 à 15% selon la date de l'image et la variable réponse considérée). La représentation fonctionnelle des habitats fournit également des modèles de meilleure performance explicative que ceux issus de CORINE Land Cover. L'utilisation de séries temporelles d'images satellites non classées apparaît comme une source de donnée spatiale pertinente pour prédire les patrons de communautés d'oiseaux sur de larges étendues et à moindre coût.

Abstract

Nowadays, a new approach to use remote sensing to build is being used: unlike the traditional way that involves the establishment of landscape or habitat maps, this alternative approach consists in calculating spatial descriptors directly from non-classified images.

In this work, we assess the effect of landscape representation obtained by remote sensing on bird-habitat models. First, we compare the performance of models integrating two different representations of habitat patterns at a given time: discrete representation (CORINE Land Cover) versus continuous representation (vegetation index). Then we compare the results of these models with other models integrating a functional representation of habitats. This representation is obtained from a temporal classification through NDVI time-series data reflecting the vegetation phenology.

Bird data were obtained from the STOC national program (French Breeding Bird Survey) from which several response variables were calculated (several species richness and functional metrics). The used images correspond to 16 days syntheses NDVI obtained with the MODIS sensor (250m spatial resolution).

The results show a significant impact of landscape representation on model performances. Unclassified images (i.e. continuous representation) always offer a higher explanatory power than classified data (explained deviance difference of 10-15% depending on the date of the image and the response variable considered). The landscape functional representation also provides models with better explanatory performances than those built from CORINE Land Cover. Time series unclassified satellite imagery appears as a relevant source of spatial data to predict patterns of bird communities over large areas and at a low cost.

Remerciements

Je remercie chaleureusement David Sheeren et Sébastien Bonthoux de m'avoir suivie tout au long de ce stage, et guidé sur ce projet passionnant, à l'interface entre les outils informatiques et les petits oiseaux ! Merci pour tous vos précieux conseils, votre disponibilité et votre gentillesse.

Merci aussi à toute l'équipe de Dynafor de m'avoir accueillie dans votre laboratoire. Merci en particulier à Aude, Romain, Pierre-Alexis et Carole d'avoir été disponibles et soutenant en période de concours. Merci à Hugo, Anthony, Tiffany, Pauline, Thomas, Bruno, Clara, … d'avoir supporté mes fou-rires et mon exubérance, et d'avoir participé à rendre ce stage inoubliable.

Je remercie le Museum National d'Histoire Naturelle, et en particulier Frédéric Jiguet grâce à qui nous avons pu travailler sur les données oiseaux, indispensables à cette étude.

Je remercie le collectif PAYOTE, co-financeur de ce stage.

Je remercie Christophe Sausse pour nos échanges instructifs sur ce projet.

Je souhaite également remercier Jean-Philippe Choisis de m'avoir fait participer à des relevés d'oiseaux, et de m'avoir donné confiance en mes compétences ornithologiques. Merci pour ces belles observations, parfois inattendues, ainsi que pour tes photos !

Je remercie Christian Piedallu d'avoir accepté d'être mon tuteur pour ce stage, et d'avoir été présent lorsque j'en avais besoin.

Je remercie Anne Jolly pour sa participation à ma soutenance.

Merci aussi à Clara Ferrari pour ses photos !

Enfin, merci également aux oubliés de cette liste qui ne sont certainement pas les moins importants…

Table des matières

Introduction 5
1. Contexte et présentation de l'étude 5
1.1. Etat de l'art 5
1.2. Choix du modèle oiseaux 13
1.3. Problématique et objectifs 14
2. Matériel et méthodes 16
2.1. Les données de biodiversité aviaire 16
2.1.1. Description des données utilisées 16
2.1.2. Calcul des variables réponses 17
2.2. Les données d'observation de la terre 18
2.2.1. Description des données utilisées 18
2.2.2. Interprétation des variables brutes 20
2.2.3. Prise en compte de la qualité des données 22
2.2.4. Construction des variables d'état 25
2.2.5. Construction de classes de fonctionnement phénologique 26
2.3. Les données d'occupation du sol 27
2.4. Les données de stratification 29
2.5. Construction des modèles 31
2.5.1. Démarche générale 31
2.5.2. Impact de la date sur la performance des modèles 32
2.5.3. Prise en compte de l'effet région 32
2.5.4. Complémentarité entre types de variables pour une même date 33
2.5.5. Complémentarité entre dates pour un même type de variables 34
2.5.6. Classification fonctionnelle 34
2.5.7. Impact de la représentation spatiale 34
3. Résultats 35
3.1. Impact de la date sur la performance des modèles 37
3.2. Prise en compte de l'effet région 39
3.3. Complémentarité entre types de variables pour une même date, et complémentarité entre dates pour un même type de variables 42
3.4. Classification fonctionnelle 45
3.5. Impact de la représentation spatiale 48
4. Discussion 50
4.1. Analyse des résultats obtenus 50
4.2. Limites de l'étude 53
4.3. Apports de l'étude et perspectives 55
Références bibliographiques 57
ANNEXES 62

Table des Figures

Figure 1 : Approche traditionnelle de l'étude de la biodiversité à partir d'images satellites : modélisation à partir d'une carte d'occupation du sol utilisée comme proxy 6
Figure 2 : Approche émergente de l'étude de la biodiversité à partir d'images satellites 9
Figure 3 : Evolution annuelle du NDVI pour différents types d'occupation du sol 10
Figure 4 : Exemple de métriques phénologiques (Source : Eklundha et Jönssonb, 2012) 12
Figure 5 : Evolution des abondances des populations d'oiseaux communs 14
Figure 6 : Répartition des carrés STOC suivis au moins une fois depuis 1989 16
Figure 7 : Diagramme de classes UML représentant la structure de la base de données Oiseaux développée 17
Figure 8 : Photographie aérienne de la zone présentée sur les figures 9 à 12 (source : BDORTHO, IGN 2010) 20
Figure 9 : Extrait des couches de LSTD aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite) 20
Figure 10 : Extrait des couches de LSTN aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite) 21
Figure 11 : Extrait des couches de NDVI aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite) 21
Figure 12 : Extrait des couches d'EVI aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite) 21
Figure 13 : Impact de la fiabilité des données sur les valeurs de NDVI (synthèse du 19/12/2010) 23
Figure 14 : Variabilité de la qualité des données de NDVI et EVI selon la date 23
Figure 15 : Moyenne et écart-type de la fiabilité des données du produit MOD13Q1 à l'échelle de la France pour l'année 2010 24
Figure 16 : Présentation d'un carré STOC (N°45-1293) en lien avec la synthèse EVI (26 juin 2010). 25
Figure 17 : Carte d'occupation des sols français (CORINE Land Cover 2006, légende simplifiée) 28
Figure 18 : Présentation des trois cartes de régionalisation de la France 30
Figure 19 : Cartes prédictives de la répartition de la richesse en espèces forestières 36
Figure 20 : Pourcentages de déviance expliquée des modèles des différentes variables réponses à partir du NDVI (à gauche) et de l'EVI (à droite) aux différentes dates 38
Figure 21 : Pourcentages de déviance expliquée des modèles des différentes variables réponses à partir du LSTD (à gauche) et du LSTN (à droite) aux différentes dates 38
Figure 22 : Répartition des résidus du modèle : RS.Fores ~ MoyNDVI177 + EcaNDVI177 39
Figure 23 : Répartition des résidus du modèle : RS.Fores ~ EcaNDVI177 + MoyNDVI177 + reg_sylv + reg_sylv:EcaNDVI177 + reg_sylv:MoyNDVI177 (notation de R) 41
Figure 24 : D^2 des modèles de RS.Fores (à gauche) et RS.Agri (à droite) 43
Figure 25 : D^2 modèles de RS.Urb (à gauche), RS.Gene (au centre) et RS.Tot (à droite) 43
Figure 26 : D^2 des modèles du CTI (à gauche) et du CSI (à droite) 44
Figure 27 : Pourcentage de déviance expliquée et différence entre les AIC des modèles construits à partir d'un nombre différents de clusters pour chaque variable réponse 45
Figures 28 : Evolution moyenne du NDVI pour chacun des 11 clusters 46
Figure 29 : Exemple de classification des pixels des carrés STOC n°77-0007 et n°77-0008, en lien avec la classification de CORINE Land Cover 47
Figure 30 : Répartition des clusters des pixels des carrés STOC 48
Figure 31 : Pourcentage de déviance expliquée des modèles issus du VCF (b.2) et de l'information « présence de forêt » de CORINE Land Cover (c.1) 48
Figure 32 : Pourcentage de déviance expliquée des modèles issus de différentes représentations spatiales pour chaque variable réponse 49
Figure 33 : Représentation du triangle des corrélations entre indices de végétation et LST (source : Petropoulos et al., 2009) 52
Figure 34 : Répartition des carrés STOC conservés pour l'étude parmi ceux relevés en 2010 55

Table des tableaux

Tableau 1 : Caractéristiques techniques des données satellites utilisées 19
Tableau 2 : Liste des variables explicatives d'état 26
Tableau 3 : Noms et descriptions des variables explicatives issues de CORINE Land Cover 28
Tableau 4 : Corrélation de Pearson entre les variables de NDVI, EVI, LSTD et LSTN pour la date 065 (6 mars 2010) 33
Tableau 5 : Moyenne, minimum et maximum des variables réponses sur les carrés STOC 35
Tableaux 6 : Forme des courbes obtenues entre variables réponses et explicatives 36
Tableau 7 : Dates correspondants aux meilleurs modèles pour chaque type de variable explicative, et par variable réponse 39
Tableau 8 : Pourcentage de déviance expliquée des modèles avec régionalisation 40
Tableau 9 : Moyenne de la corrélation de Spearman (rho) et de l'erreur quadratique moyenne (RMSE) des meilleurs modèles avec régions 42
Tableau 10 : AIC des modèles construits à partir d'un nombre différent de clusters 45

Introduction

Cette étude a été réalisée dans le cadre d'un stage de 6 mois terminant une année de formation de niveau Master 2 en Gestion des Milieux Naturels (AgroParisTech, centre de Nancy), de mars à août 2014. Elle s'est effectuée à l'Institut National de Recherche Agronomique (INRA) de Toulouse, dans une unité mixte de recherche, DYNAFOR. Cet institut de recherche, fondé en 1946, est sous la double tutelle du ministère de la Recherche et du ministère de l'Agriculture. Les activités de DYNAFOR sont centrées sur la gestion durable des ressources forestières et de l'espace rural dans le cadre de l'écologie du paysage. Un des objectifs principaux de DYNAFOR est de comprendre et de modéliser les relations entre des processus écologiques et des patrons spatiaux. Ce stage s'inscrit également dans le réseau PAYOTE. Il s'agit d'un collectif qui regroupe des chercheurs et des ingénieurs de l'INRA et de l'INRIA travaillant sur la modélisation de paysages agricoles pour la simulation et l'analyse de processus écologiques et environnementaux.

Ainsi, cette étude interdisciplinaire a mobilisé des connaissances en écologie (écologie du paysage, écologie des oiseaux), mais aussi des compétences en géomatique et en télédétection.

1. Contexte et présentation de l'étude

1.1. Etat de l'art

Aujourd'hui, de nombreux taxons sont en déclin à l'échelle internationale. Ce déclin a un coût important pour la société humaine (Cardinale et al., 2012). Afin de pouvoir mettre en œuvre des plans de conservation, voire de restauration de la biodiversité, une étape préliminaire majeure est d'estimer cette biodiversité, et d'évaluer la répartition géographique des différentes espèces.

Le moyen le plus direct de connaître la biodiversité sur une zone est de réaliser des relevés sur le terrain sur des échantillons judicieusement choisis. Bien que cette méthode soit de loin la plus précise, elle présente un certain nombre de limites (Pfeifer et al., 2012) :

- Les relevés sont coûteux en temps et en main d'œuvre.
- Les méthodes d'échantillonnage peuvent être subjectives et induire un biais dans l'analyse statistique des résultats.
- Certaines zones ne peuvent pas être échantillonnées car difficiles d'accès pour des raisons géographiques (zones isolées) ou politiques (instabilité politique de certains endroits).

Avec la multiplication des satellites et des capteurs disponibles, l'utilisation de la télédétection est un moyen alternatif aux relevés de terrain en fort développement. Le terme de télédétection a été introduit officiellement dans la langue française en 1973 et est défini comme l'« ensemble des connaissances et techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par des mesures effectuées à distance, sans contact matériel avec ceux-ci. » (JORF du 11 décembre, 1980). Cette définition très large a été reprécisée par le Centre Canadien de cartographie et d'observation de la Terre de la manière suivante :

« La télédétection est l'ensemble des techniques qui permettent, par l'acquisition d'images, d'obtenir de l'information sur la surface de la Terre (y compris l'atmosphère et les océans), sans contact direct avec celle-ci. La télédétection englobe tout le processus qui consiste à capter et enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchi, à traiter et analyser l'information qu'il représente, pour ensuite mettre en application cette information. » (Centre Canadien de Télédétection.).

Il existe deux façons principales d'utiliser la télédétection pour étudier la biodiversité. La première, que nous qualifions d'approche directe, est la plus intuitive. Elle consiste à détecter directement certains organismes individuels. De nouveaux capteurs satellites à très haute résolution spatiale (résolution « hyperspatiale » < 5 m) permettent aujourd'hui d'identifier directement certains organismes de taille corporelle élevée tels que les éléphants ou les baleines (Leyequien et al., 2007). De même, de nouveaux capteurs hyperspectraux découpent le spectre électromagnétique en bandes plus fines qui permettent la détection de signatures spectrales caractéristiques de certaines espèces arborées ou de communautés végétales (Gillespie et al., 2008; Turner et al., 2003).

La deuxième approche est indirecte et se base sur l'interprétation de paramètres environnementaux pour expliquer les patrons de diversité biologique (Gottschalk et al., 2005). C'est cette approche « habitat » qui est privilégiée pour la plupart des espèces. De manière traditionnelle, elle suit le schéma de la figure 1 : les images satellites sont utilisées pour construire une cartographie d'occupation du sol, ce qui revient à classer l'information. Des relevés de biodiversité sont effectués en parallèle. A partir de la carte précédemment réalisée, des variables descriptives du paysage (composition et configuration) sont extraites pour chaque aire d'échantillonnage. Ces variables permettent de créer un modèle de présence d'une espèce ou d'une communauté d'espèces en fonction des variables paysagères. Les variables paysagères sont calculables sur toute la zone étudiée, et non seulement sur les aires d'échantillonnage. Cela permet donc par projection du modèle de prédire la biodiversité à large échelle.

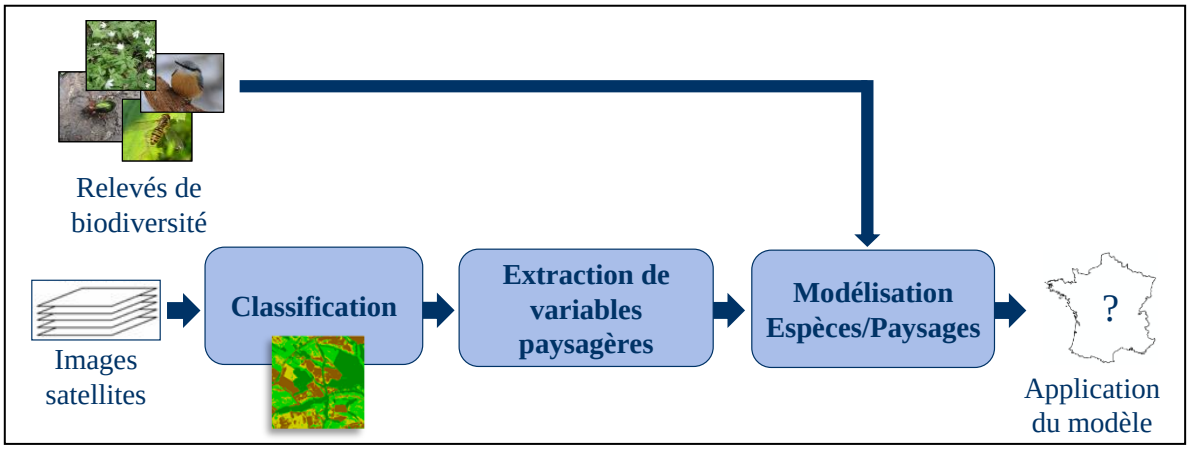


Figure 1 : Approche traditionnelle de l'étude de la biodiversité à partir d'images satellites : modélisation à partir d'une carte d'occupation du sol utilisée comme proxy

Il existe de nombreuses variantes possibles dans les modèles, à la fois par l'existence de plusieurs indicateurs de biodiversité, par la variété des images satellites disponibles, ainsi que par la façon dont les variables sont calculées.

Présentation des différents indicateurs de biodiversité et des différents types de diversité

La diversité biologique présente plusieurs facettes. Traditionnellement, elle est assimilée à la diversité taxonomique, c'est-à-dire à la diversité d'espèces observées pour un taxon donné. Elle présente l'avantage majeur d'être facilement mesurable. Il existe également deux autres caractéristiques importantes de la biodiversité : la diversité phylogénétique et la diversité fonctionnelle. La diversité fonctionnelle reflète la diversité des traits morphologiques, physiologiques et écologiques au sein des communautés biologiques et complète l'analyse du fonctionnement des écosystèmes par rapport à une approche taxonomique. La diversité phylogénétique, quant à elle, reflète l'histoire évolutive d'une communauté (Devictor et al., 2010; Meynard et al., 2011).

Pour l'étude de communautés d'espèces, un certain nombre d'indicateurs ont été mis en place par les chercheurs afin de mieux comprendre les tendances observées. L'indicateur le plus couramment utilisé pour évaluer la diversité taxonomique est la richesse spécifique, qui représente le nombre d'espèces différentes observées dans une communauté donnée. Les indices de Shannon (Shannon, 1948) et de Simpson (Simpson, 1949) prennent également en compte la distribution des individus au sein de ces espèces : l'indice de Shannon (H) permet d'exprimer la diversité à partir du nombre d'espèces et de l'abondance des individus au sein de chacune de ces espèces tandis que l'indice de Simpson (D) donne la probabilité que deux individus sélectionnés aléatoirement dans un milieu donné soient de la même espèce.

$$H = -\sum_{i=1}^{S} p_i \ln(p_i) \quad \text{et} \quad D = \sum_{i=1}^{S} \frac{n_i\,(n_i-1)}{N\,(N-1)}$$

Où S est le nombre total d'espèces, p_i correspond à l'abondance proportionnelle ou pourcentage d'importance de l'espèce : $p_i = n_i/N$ – n_i étant nombre d'individus d'une espèce dans l'échantillon et N le nombre total d'individus de toutes les espèces dans l'échantillon.

Pour évaluer la diversité fonctionnelle, différents indicateurs existent. Deux indicateurs sont courants pour la diversité fonctionnelle des communautés d'oiseaux (Bonthoux et al., 2013; Devictor et al., 2008a; Jiguet et al., 2012; Princé et al., 2013) : le *community trophic index* (CTI) et le *community specialisation index* (CSI). Ces deux indices sont calculés sur une surface en sommant respectivement le *species trophic index* (STI) et le *species specialisation index* (SSI) de chaque oiseau présent pondéré par son abondance dans la zone considérée. Pour une parcelle codée j, le CSI et le CTI sont donnés par :

$$CSI = \frac{\sum_{i=1}^{n} a_{ij}(SSI_i)}{\sum_{i=1}^{n} a_{ij}} \quad \text{et} \quad CTI = \frac{\sum_{i=1}^{n} a_{ij}(STI_i)}{\sum_{i=1}^{n} a_{ij}},$$

Où n est le nombre total d'espèces échantillonnées, a_{ij} est l'abondance des individus de l'espèce i sur la parcelle j, SSI_i est l'indice de spécialisation de l'espèce i et STI_i son indice trophique.

Le STI d'un oiseau représente son positionnement dans la chaîne trophique. Il est déterminé par la proportion de consommation dans le régime alimentaire de chaque espèce de trois catégories d'aliments : des végétaux, des invertébrés et des vertébrés, avec un poids respectif de 1, 2 et 3. Plus le STI d'un oiseau est élevé, plus l'oiseau est positionné haut dans la chaîne trophique. Le CTI donne donc une évaluation du niveau moyen de la communauté étudiée dans la chaîne trophique.

Le SSI d'un oiseau représente son niveau de spécialisation par rapport à un habitat. Plus un oiseau est présent exclusivement dans un type d'habitat donné, plus son SSI est élevé. Le CSI représente donc le niveau de spécialisation à l'habitat de la communauté étudiée.

Présentation des types de données disponibles en télédétection

Les capteurs enregistrent l'intensité d'un signal au sein d'un intervalle de longueur d'onde (la bande spectrale), d'une largeur définie au sein du spectre électromagnétique. Il existe deux principaux types de capteurs (Turner et al., 2003) :

- Les capteurs passifs : ils mesurent le signal atteignant le détecteur.
- Les capteurs actifs : ils émettent d'abord une pulsation et mesure ensuite l'énergie envoyée en retour au détecteur.

Les capteurs passifs, qui sont les plus communément utilisés pour l'observation des surfaces continentales, enregistrent l'énergie réfléchie sur les longueurs d'onde du visible, du proche et du moyen infrarouge. Les images dérivées sont exploitées visuellement ou statistiquement pour identifier les objets. Les capteurs actifs, insensibles à la couverture nuageuse, sont plus couramment utilisés pour mesurer la structure de la végétation ou décrire la topographie du terrain. Les deux principaux capteurs actifs sont le système *light detection and ranging* (LIDAR) qui opère dans le visible et le

proche infrarouge, et le système *radio detection and ranging* (RADAR) qui émet des radiations de plus grande longueur d'onde (Gillespie et al., 2008; Turner et al., 2003).

Les différents capteurs passifs diffèrent selon plusieurs critères, dont les trois plus importants sont les suivants :

- La résolution spatiale : la taille des pixels au sol fournis par le capteur,
- La résolution spectrale : la plus petite largeur des bandes spectrales enregistrées,
- La résolution temporelle : la période de temps entre deux enregistrements d'une même scène.

En pratique, il y a souvent un compromis à trouver entre ces trois résolutions : des données à haute résolution spatiale fournissent un niveau de détail très élevé mais ne couvrent qu'une étendue limitée. Leur coût est également important et elles sont très volumineuses ce qui augmente également les temps de traitement. Des chercheurs ont montré qu'il était préférable d'améliorer la résolution spectrale sur la résolution spatiale pour les applications concernant la biodiversité (Pfeifer et al., 2012; Rocchini et al., 2010).

Les images issues des capteurs passifs sont utilisées pour construire des cartes d'occupation du sol. Des variables paysagères sont ensuite calculées (Fahrig et al., 2011). Elles sont de différentes natures et dépendent des objectifs de l'étude. Il peut s'agir par exemple du pourcentage de prairies, du pourcentage de cultures céréalières, de la longueur des haies, etc. Ces variables traduisent la *composition* du paysage. L'hétérogénéité de composition est également souvent analysée et calculée en appliquant l'indice de diversité de Shannon aux différentes catégories paysagères (Bonthoux et al., 2013). D'autres variables caractérisant la *configuration* du paysage et son hétérogénéité sont également utilisées. Ces variables sont relatives à la complexité des formes et à l'agencement spatial des éléments paysagers (dispersion, agrégation, distribution).

Malgré les grandes avancées permises par l'approche présentée précédemment (figure 1), l'étape de classification des données satellites apparaît limitante pour plusieurs raisons (Rocchini et al., 2010) :

- Le temps de traitement pour la classification est important.
- Il peut y avoir des erreurs de classification introduisant potentiellement un biais dans les modèles. Gottschalk et al. (2005) ont montré qu'un changement de 5% dans la précision de la classification peut causer des différences significatives dans le niveau de l'index de qualité d'un habitat.
- La classification est souvent focalisée sur des catégories d'occupation du sol et non sur l'habitat spécifique de l'espèce étudiée. Les classes choisies *a priori* sont par ailleurs entachées d'une vision antropocentrée qui peut être différente de celles des espèces. Ainsi, souvent, la faune sauvage répond à l'hétérogénéité du paysage de manière différente de celle qu'on imagine, par exemple en fonction de l'étage inférieur de la végétation ou de la structure d'une forêt plutôt que de son âge ou du type de forêt (Laurent et al., 2005). Il a été montré que cette disparité entre les perceptions humaines et celles de la faune sauvage aboutit à des classifications d'habitat donnant de moins bons résultats lors de la modélisation (Mimet et al., 2014).
- Plus généralement, par l'action même de la classification qui simplifie la représentation du paysage, la majeure partie de l'hétérogénéité spectrale et spatiale intrinsèque aux images est perdue. Or cette hétérogénéité intra-habitat peut contenir une information non négligeable pour la construction des modèles (St-Louis et al., 2006). Par exemple, les zones de transitions sont remplacées par des limites précises, omettant potentiellement des zones d'interaction espèce-habitat (Duro et al., 2014).

Aussi, une approche émergente consiste aujourd'hui à considérer directement les patrons spectraux et leur hétérogénéité comme proxy de la diversité (figure 2) (Rocchini et al., 2010; St-Louis et al.,

2006; Wood et al., 2013). La réflectance des données brutes, ou les différents indices de végétation dérivés sont en effet le reflet de propriétés physiques voire fonctionnelles des écosystèmes (productivité primaire, indice foliaire, teneur en eau…). Les images non classées peuvent donc potentiellement jouer le rôle d'indicateurs indirects de la biodiversité.

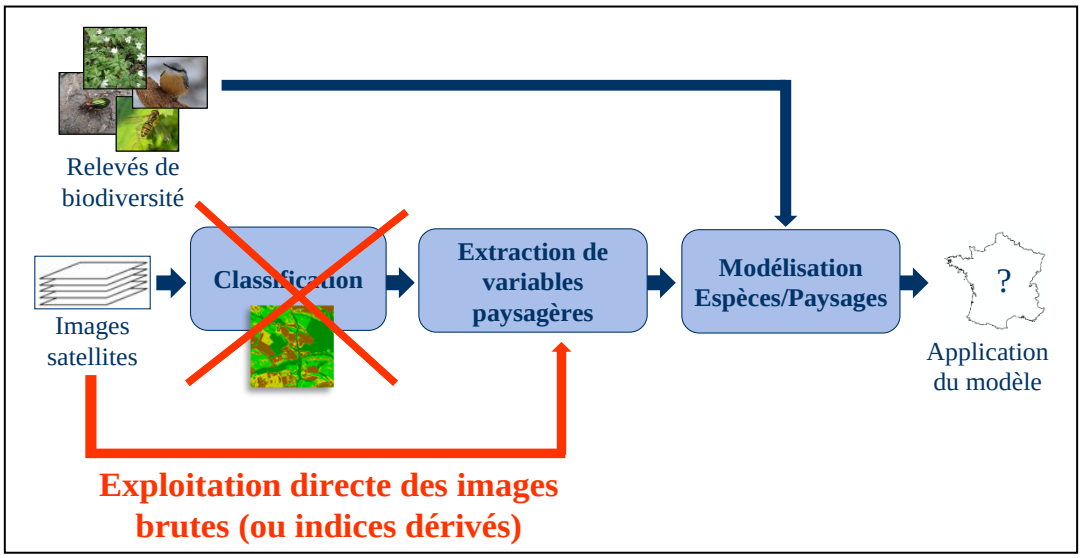


Figure 2 : Approche émergente de l'étude de la biodiversité à partir d'images satellites

S'il semble que cette nouvelle approche permette de créer de meilleurs modèles, peu d'études ont à ce jour effectué de comparaison directe entre les deux méthodologies (Duro et al., 2014; Sheeren et al., 2014).

Présentation des types de variables issues des données de télédétection pour la construction de modèles espèces/paysages

Les variables paysagères extraites des images brutes des capteurs passifs (les seuls utilisés dans cette étude) peuvent prendre en compte plusieurs types d'hétérogénéité : l'hétérogénéité spectrale, temporelle et spatiale.

L'hétérogénéité spectrale :

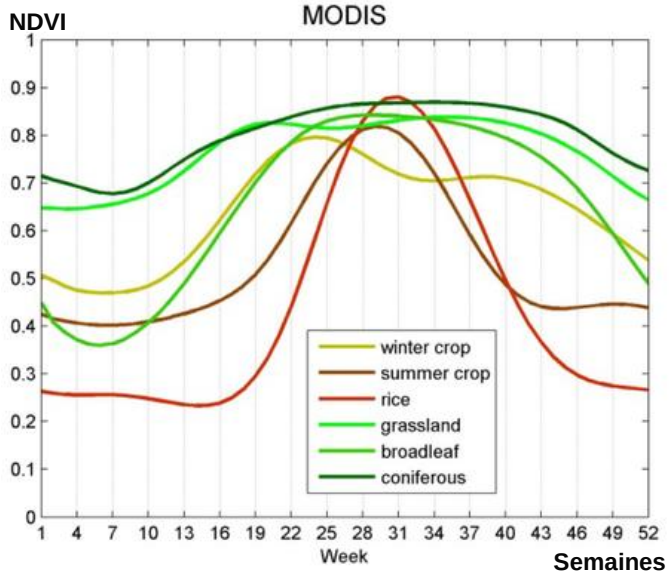
La variation spectrale ou le ton des images est difficile à prendre en compte : il existe de fortes corrélations extensives inter-bandes. Aussi, une seule bande spectrale est souvent utilisée dans les modèles (Laurent et al., 2005). Or plusieurs bandes spectrales apportent souvent une information complémentaire. Afin de contourner ce problème, deux principales méthodes sont actuellement utilisées (Rocchini et al., 2010) :

- l'utilisation de techniques de réduction des données telles que l'ACP,
- la construction d'indices ou neocanaux (par combinaison de bandes spectrales).

Des indices de végétation se sont révélés particulièrement pertinents dans la construction de modèles, et notamment le *Normalized Difference Vegetation Index* (NDVI). Cet indice est très classique en télédétection et est calculé à partir des bandes rouge (R) et proche infra-rouge (PIR) :

$$\text{NDVI} = \frac{PIR - R}{PIR + R}$$

Le NDVI, proposé par Rouse et al. (1973) est sensible à la biomasse photosynthétiquement active, corrélé à l'indice de surface foliaire (LAI), relié à la quantité de radiation active photosynthétiquement absorbée, et à la productivité primaire nette et ce, à plusieurs échelles spatiales (Sellers, 1987; Seto et al., 2004). Son intérêt a également été démontré sur des modèles de biodiversité en milieu urbain (Bino et al., 2008). De manière générale, une plus grande productivité primaire peut être reliée à une plus grande quantité de nourriture, et une plus grande variance de NDVI entre les pixels peut être reliée à une plus grande variabilité d'habitats (Pettorelli et al., 2011). De plus, le NDVI ne représente pas seulement l'état mais aussi le fonctionnement de la végétation. En effet, comme l'illustre la figure 3, le NDVI varie dans le temps et son cycle annuel et est différent selon les types de milieu.



Source : Atzberger *et al*., 2014

Figure 3 : Evolution annuelle du NDVI pour différents types d'occupation du sol

Au-delà d'être relié à la végétation, le NDVI est également utilisé pour construire des modèles relatifs à des espèces herbivores et non herbivores, la production primaire d'une région influençant toute sa variété de nourriture (Pettorelli et al., 2011).

Il existe de nombreux autres indices, dont notamment l'*Enhanced Vegetation Index* (EVI) (Pettorelli et al., 2005) :

$$\text{EVI} = 2{,}5 * \frac{PIR - R}{(PIR + C_1 * R - C_2 * B + L}$$

Où C_1, C_2 et L sont des coefficients de correction des conditions atmosphériques (dans des conditions standards : C_1 = 6, C_2 = 7,5 et L = 1), et PIR, R et B la réflectance corrigée atmosphériquement (ou partiellement corrigée) de la surface, respectivement dans les bandes proche infrarouge, rouge et bleue.

L'EVI fournit une information complémentaire sur les variations spatio-temporelles de la végétation, en minimisant des problèmes de contaminations présents dans le NDVI. Globalement, le NDVI répond mieux aux variations du rouge et est plus sensible à la chlorophylle. L'EVI est plus sensible aux variations du proche infra-rouge et répond plus aux variations structurelles de la canopée. De plus, l'EVI présente l'avantage de ne pas saturer comme le NDVI dans les forêts tropicales ou autres zones très riches en végétation (biomasse importante) (Pettorelli et al., 2005).

L'hétérogénéité spatiale :

Les variables basées sur la variation spatiale ou texture peuvent être de premier ou de second ordre. Les variables de premier ordre sont des résumés statistiques, comme la moyenne ou l'écart-type, calculés pour chaque pixel de l'image dans une fenêtre de voisinage (fenêtre de convolution). Les variables de second ordre, également calculées pour une fenêtre définie, prennent en compte la distribution spatiale et les dépendances des valeurs spectrales : il s'agit par exemple de l'homogénéité, de l'entropie et de la dissimilarité (Culbert et al., 2009, 2012). Elles sont plus longues et plus complexes à calculer (Coburn and Roberts, 2004).

Le choix de la taille de la fenêtre sur laquelle sont effectués les calculs d'hétérogénéité spatiale est fondamental, notamment pour le NDVI (Leyequien et al., 2007). Pour les oiseaux, Morelli et al. (2013) ont obtenus de meilleurs modèles pour un rayon de 125 m autour des relevés, soit une fenêtre de 4,9 ha. Bino et al. (2008), quant à eux, ont montré que la relation entre le NDVI et la richesse spécifique aviaire est maximale pour une échelle de 14,8 ha. Cela peut être relié à la taille de leur territoire (Laurent et al., 2005). A large échelle (comme celle d'un pays), les données à moyenne résolution spatiale (> 60 m) ont déjà prouvé leur efficacité (Albright et al., 2011). De manière similaire, on peut noter que dans le cas de l'approche traditionnelle, la résolution de la classification impacte également fortement la qualité des modèles obtenus (Lawler et al., 2004).

L'hétérogénéité temporelle :

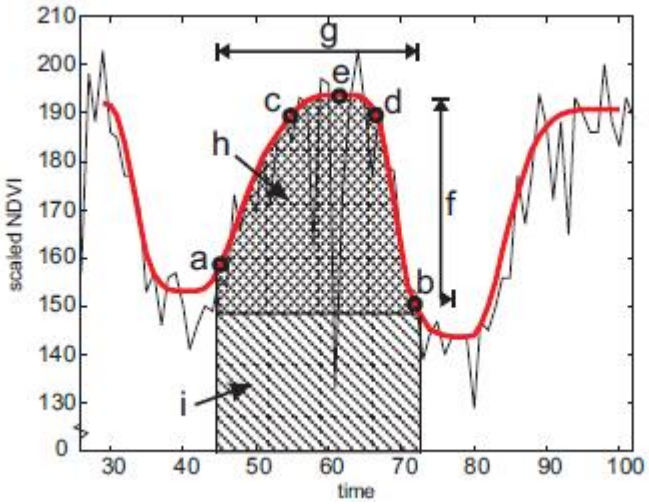
Généralement, en Europe, pour les oiseaux, les variables de productivité primaire comme le NDVI sont issus d'images prises vers le mois de juin, qui correspond à la période d'expansion maximale de la végétation, et qui est cohérent avec la période des relevés de biodiversité (Bacaro et al., 2011; Bailey et al., 2004).

Or, toutes les mesures d'hétérogénéité varient avec la phénologie (Culbert et al., 2009). Le choix de la date des images est donc fondamental. A ce jour, peu de chercheurs se sont penchés sur cette question pour construire des modèles oiseaux/paysages. On peut néanmoins noter que les résultats de Sheeren et al. (2014) suggèrent qu'il vaut mieux utiliser des images NDVI d'automne pour expliquer les patrons des communautés d'oiseaux pendant la période de nidification (mai-juin). En automne, les cultures d'été ont été récoltées et le NDVI discrimine mieux les cultures des éléments boisés que pendant l'été. Cette étude s'est focalisée sur une région de France et les résultats doivent maintenant être généralisés dans d'autres contextes paysagers.

L'utilisation de variables issues de données brutes présente deux principales limites :

- le temps de calcul pour certaines mesures d'hétérogénéité spatiale (indices du second ordre),
- la difficulté d'interprétation écologique *a priori* : les modèles construits ne permettent pas forcément d'améliorer les connaissances sur l'écologie des oiseaux, mais ils peuvent aider à d'autres applications où la qualité de la carte obtenue est plus importante, comme pour planifier des actions de conservation (Culbert et al., 2012; Gottschalk et al., 2005).

De plus, des variables calculées à partir de l'hétérogénéité temporelle permettent la prise en compte de la phénologie de la végétation. Il en existe de nombreuses, qualifiées de métriques phénologiques, telles que (Hurlbert and Haskell, 2003; Pettorelli et al., 2005) : le NDVI intégré, le taux de croissance ou décroissance du NDVI, les dates de début et de fin du pic de la saison de croissance, la durée de la saison de croissance, la date du maximum annuel du NDVI, les valeurs du NDVI à une date donnée, la différence entre le NDVI de juin et le NDVI minimum mensuel (mesure absolue du pulse saisonnier), le ratio entre ce pulse saisonnier et le NDVI de juin (proportion de la production en juin par rapport au mois le moins productif). Certaines sont illustrées en figure 4.



Points (a) and (b) mark, respectively, start and end of the season. Points (c) and (d) give the 80 % levels. (e) displays the point with the largest value. (f) displays the seasonal amplitude and (g) the seasonal length. Finally, (h) and (i) are integrals showing the cumulative effect of vegetation during the season.

Figure 4 : Exemple de métriques phénologiques (Source : Eklundha et Jönssonb, 2012)

Récemment, une autre manière de prendre en compte la phénologie de la végétation a été développée à partir du NDVI. Il s'agit non pas de dériver des variables à partir de la signature temporelle du NDVI mais plutôt de classer directement les pixels en fonction de cette signature. On obtient alors une classe pour chaque pixel qui traduit le type de milieu mais aussi son fonctionnement. L'algorithme de classification est adapté à la classification de courbes (données longitudinales). Il ne s'agit plus ici d'une approche fondée sur la donnée brute puisque les données sont classées, mais la sémantique de la classe intègre cette fois le fonctionnement phénologique des couverts et leur diversité. Tout récemment exploitée par Hurley et al. (2014), cette technique prometteuse est peu présente dans la littérature.

Les modèles espèces/paysages présentés précédemment se basent sur quelques principales hypothèses :

- La présence des espèces peut être prédite par des composants de leur habitat qui sont détectables spectralement (Laurent et al., 2005).
- L'hypothèse de la variabilité spectrale : hypothèse selon laquelle une variabilité spectrale plus importante devrait correspondre à une diversité spécifique plus élevée. Elle est liée à l'hypothèse de différence de niche selon laquelle la richesse spécifique des oiseaux est plus élevée lorsqu'il y a une plus grande hétérogénéité de végétation (Bacaro et al., 2011; Morelli et al., 2013).
- La théorie espèces-énergie : théorie selon laquelle les sites avec davantage d'énergie disponible sont capables de contenir plus d'individus, et donc plus d'espèces. Cela peut être lié soit directement à une augmentation de la disponibilité en nourriture, ou à une température plus élevée qui réduit la charge thermorégulatrice des individus, leur permettant ainsi d'allouer plus d'énergie à la croissance et la reproduction (Duro et al., 2014; Hurlbert and Haskell, 2003).
- L'hypothèse de productivité prédit que lorsque les ressources sont abondantes et fiables, les espèces deviennent plus spécialisées et donc moins compétitives, ce qui permet d'avoir davantage d'espèces pour une même surface (Wu et al., 2014).

En télédétection indirecte, trois variables principales sont utilisées : les variables de production primaire, les variables de climat, et les variables de structure des habitats et de topographie (Turner et al., 2003). Ficetola et al. (2014) a également montré que les variables abiotiques telles que le climat

sont plus importantes pour les modèles à large échelle, tandis qu'à plus fine échelle, ce sont surtout celles représentant les habitats, le paysage, la végétation ou les interactions biotiques qui priment. Culbert et al. (2012), pour sa part, a montré qu'à large échelle, c'est le type de paysage ou d'habitat qui compte le plus, tandis qu'à fine et moyenne échelle c'est la structure de l'habitat qui est plus efficace à expliquer les patrons de richesse spécifique des oiseaux. Ces résultats concordent avec ceux trouvés par Cord et Rödder (2011).

1.2. Choix du modèle oiseaux

Dans cette étude, le taxon étudié est celui des oiseaux. C'est un groupe particulièrement bien connu et étudié par les amateurs et les professionnels, et pour lequel un suivi structuré a été mis en place depuis 1989 par le Muséum National d'Histoire Naturelle (MNHN) à l'échelle nationale. En particulier, le MNHN coordonne un programme de Suivi Temporel des Oiseaux Communs (STOC). C'est un programme de sciences participatives dans lequel des ornithologues amateurs et éclairés sont mis à contribution pour collecter un grand nombre de données difficile à obtenir par les moyens techniques classiques des laboratoires de recherche. Il se compose de deux volets complémentaires :

- le STOC-EPS, Echantillonnages Ponctuels Simples, qui est conçu pour évaluer les variations spatiales et temporelles de l'abondance des populations nicheuses d'oiseaux communs. Il est basé sur des points d'écoute. Ce type de comptage compte parmi les méthodes les plus appropriées pour la construction des modèles (Gottschalk et al., 2005).
- le STOC-Capture, qui vise à étudier les variations dans la survie des adultes et dans le succès de la reproduction, deux importants paramètres démographiques. Cette approche est basée sur la capture et la recapture de passereaux nicheurs à l'aide de filets japonais.

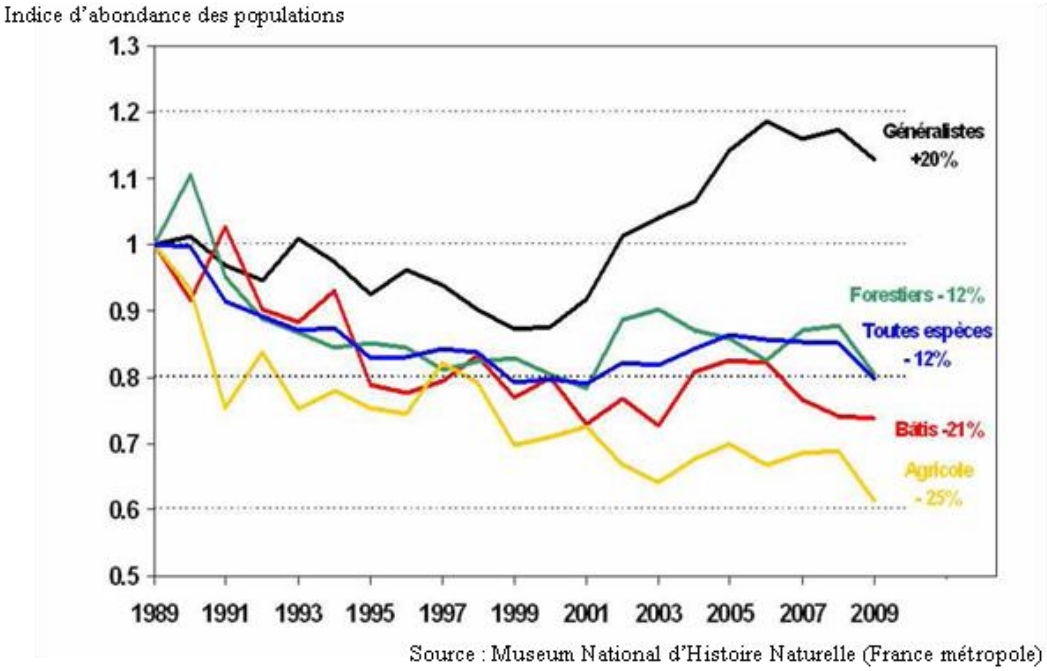
Les principaux atouts de ce programme de sciences participatives sont les suivants (Jiguet et al., 2012) :

- Il permet la mise à disposition de très grandes quantités de données, et sur de larges étendues d'espace et de temps.
- L'obtention de ces données aurait été difficilement possible avec les moyens financiers de la recherche.

Les données obtenues ont déjà permis de vérifier qu'en France les oiseaux sont particulièrement sensibles aux changements globaux : changement climatique et modifications d'occupation des sols.

- Le changement climatique affecte les patrons de communauté d'oiseaux (Jiguet et al., 2007; Julliard et al., 2004), ce qui est cohérent avec ce qui a été observé à l'échelle internationale (Both et al., 2006; Naef-Daenzer et al., 2012).
- La structure des habitats a un fort impact sur les assemblages d'oiseaux à fine échelle, et à plus large échelle, elle modifie les patrons spatiaux de richesse spécifique (St-Louis et al., 2006).

De plus, les chercheurs ont pu montrer que les oiseaux spécialistes et généralistes répondent de manière différente au changement climatique et à la fragmentation des habitats (Clavel, 2007; Devictor et al., 2008b). Ainsi, comme l'illustre la figure 5, on assiste aujourd'hui à une diminution des oiseaux spécialistes au profit des oiseaux généralistes (Davey et al., 2012). Les oiseaux spécialistes présentent également une tendance à se généraliser et les communautés à s'homogénéiser (Barnagaud et al., 2011; Devictor et al., 2008a).



Les indices sont arbitrairement fixés à 1 en 1989.

Figure 5 : Evolution des abondances des populations d'oiseaux communs

Les chercheurs ont ainsi montré que les changements globaux affectent déjà à l'échelle continentale de nombreuses espèces d'oiseaux, notamment par leurs générations courtes et ce, de manière à la fois positive et négative. De plus, ce taxon est bien compris biologiquement et utilise une large étendue de niches. Facilement identifiable visuellement et auditivement, il existe des techniques standardisées pour son suivi(Bacaro et al., 2011; Prins et al., 2005; Seto et al., 2004; St-Louis et al., 2006). En ce sens, les oiseaux sont considérés comme de bons indicateurs de l'évolution de la biodiversité (Gregory et al., 2009).

Pour résumer, les oiseaux ont été choisis pour notre étude parce qu'ils forment un taxon bien connu à l'échelle nationale, pour lequel on dispose de données sur la France entière, et qui est un bon indicateur de la biodiversité.

1.3. Problématique et objectifs

Comme l'a présenté l'état de l'art, l'approche émergente suivie aujourd'hui pour la réalisation de modèles prédictifs de richesse en oiseaux se base sur l'utilisation de variables explicatives directement issues des images brutes de la télédétection. Cependant, peu d'études ont comparé explicitement ces modèles à ceux issus de cartographie d'occupation du sol (Duro et al., 2014).

Par ailleurs, l'impact et la prise en compte de l'hétérogénéité spectrale et de l'hétérogénéité spatiale ont fait l'objet de nombreuses recherches, mais à ce jour, l'hétérogénéité temporelle et l'effet « date » n'ont été que peu étudiés.

Aussi, dans le cadre de mon stage, la problématique posée est la suivante :

Quel est l'apport de séries temporelles d'images satellites pour expliquer et prédire les patrons de communauté d'oiseaux à l'échelle de la France ?

Plusieurs grandes questions découlent de cette problématique :

Une représentation continue du paysage, à travers l'utilisation de variables directement issues d'images satellitaires, permet-elle de mieux expliquer les patrons d'oiseaux que les données discrètes d'utilisation du sol ?
Y a-t-il un effet de la date des images satellites sur la qualité des modèles ? Est-ce que la prise en compte de plusieurs dates permet d'obtenir des modèles plus performants ?

Actuellement, les capteurs à très haute répétitivité temporelle sont peu nombreux. Celui fournissant les données avec le plus haut niveau de détail est le capteur MODIS (*Moderate-resolution Imaging Spectroradiometer*), capable de produire tous les 16 jours une synthèse NDVI de résolution spatiale de 250 m. C'est pour cette raison, en plus de leur disponibilité et de leur facilité d'accès, que nous avons retenu ces données satellites MODIS pour notre étude.

Notre étude étant réalisée à large échelle, les variables les plus adaptées concernent donc la production primaire et le climat (cf. section 1.1). Le NDVI est un bon estimateur de la production primaire. En ce qui concerne le climat, MODIS fournit également le produit *Land Surface Température* (LST), qui est fortement lié aux données climatiques. Cet indice n'a été que peu utilisé dans la littérature à ce jour, mais il a déjà permis de construire de bons modèles en lien avec le NDVI (Cord et Rödder, 2011). Dans cette étude, nous souhaitons explorer l'apport de cet indice en apportant des réponses aux questions suivantes :

Les variables issues du LST permettent-elles de construire des modèles aussi performants que les variables issues de classification d'occupation du sol ?
Le LST et le NDVI portent-ils une information complémentaire ?

La présence de série temporelle de NDVI permet de prendre en compte la phénologie de la végétation. Dans cette étude, nous nous proposons d'explorer l'apport de la classification fonctionnelle des pixels à partir de ces séries temporelles :

La classification fonctionnelle du paysage selon le fonctionnement de la végétation permet-elle d'obtenir des modèles plus performants que ceux fondés sur des variables d'état à une date donnée ?

Par ailleurs, les données Oiseaux du MNHN sont disponibles à l'échelle de la France. Si quelques travaux ont été réalisés à une échelle nationale, ils se basent principalement sur des cartographies d'habitats (Foody, 2004; Fuller et al., 2007; Lawler et al., 2004; Rittenhouse et al., 2012), et non pas sur l'utilisation des images brutes. Ils permettent toutefois de montrer que la généralisation des modèles à large échelle est faisable, plusieurs auteurs insistant néanmoins sur l'importance d'une approche par écorégion. Sur cet aspect, la question que nous posons est :

Est-il plus pertinent de construire un seul modèle national sans stratification, ou faut-il prendre en compte l'effet des régions dans notre modèle global ?

2. Matériel et méthodes

2.1. Les données de biodiversité aviaire

2.1.1. Description des données utilisées

Les relevés d'oiseaux utilisés pour cette étude sont ceux du STOC-EPS présentés en section 1.2. Son protocole consiste à assigner à chaque participant du programme un carré de 4 km^2 (2 x 2 km) à étudier, qui est sélectionné de manière aléatoire dans un rayon de 10 km autour d'une zone de son choix. Ce mode de sélection permet d'assurer que les relevés couvrent une sélection représentative d'habitats, notamment en ce qui concerne les zones urbaines, les zones d'agriculture intensive et les zones de forêts (Devictor and Jiguet, 2007). La répartition des carrés STOC en France est présentée dans la figure 6.

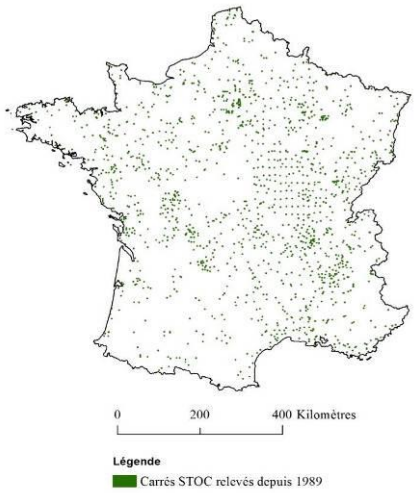


Figure 6 : Répartition des carrés STOC suivis au moins une fois depuis 1989

Chaque année, deux relevés sont effectués au printemps pour chaque carré, une fois avant et une fois après le 8 mai, en respectant quatre à six semaines d'intervalle, de manière à améliorer la probabilité de détection des espèces et diminuer le nombre de fausses absences. Dans chaque carré, l'observateur répartit lui-même 10 points d'écoute, séparés d'au moins 300 m, et de manière à ce que tous les types d'habitat présents dans le carré (parmi les milieux urbains, les forêts, les zones agricoles, les zones humides et surfaces en eau) soient représentés dans leurs proportions respectives. Sur chaque point, le volontaire enregistre tous les individus vus ou entendus pendant une période de 5 minutes.

Les observations sont répétées chaque année, par le même observateur, sur les mêmes points d'écoute et si possible aux mêmes dates (Julliard and Jiguet, 2002). La description détaillée du protocole peut être téléchargée sur le site de Vigie-Nature (Muséum National d'Histoire Naturelle).

Par ailleurs, au cours de mon stage, j'ai eu la chance de participer au relevé effectué sur un carré STOC avec un volontaire de l'étude, ingénieur de recherche dans l'unité d'accueil DYNAFOR (J.-P. Choisis). Les observations ont été réalisées les 3 mai et 7 juin 2014 sur le carré STOC n° 31-0507. L'objectif de la sortie n'était pas d'exploiter les résultats des relevés mais de mieux percevoir la manière dont les données sont acquises sur le terrain.

Les données sont mises à disposition par le MNHN. Pour chacune des deux sessions d'observation, le nombre d'individus observés sur chaque point d'écoute a été sommé pour obtenir un nombre d'individu par carré, par année de relevé, par espèce, et par session d'observation. Selon les espèces et les distances de migration des individus, la date d'arrivée sur site de nidification peut varier. Aussi, le nombre d'individus considérés présents par carré, par année de relevé et par espèce est celui

de la session au cours de laquelle le plus d'individus a été observé. C'est cette valeur que contient le fichier de données fourni par le MNHN. Les données fournies vont de 2001 – année à laquelle le programme STOC-EPS a été relancé – à 2010.

Pour cette étude, nous avons conservé les relevés les plus récents disponibles, c'est-à-dire ceux de 2010. Les carrés STOC étudiés sont donc ceux pour lesquels un relevé a été réalisé en 2010 : cela concerne 1091 carrés.

2.1.2. Calcul des variables réponses

Les données fournies par le programme STOC ne sont pas exploitables directement : elles nécessitent d'être mises en forme et les variables réponses doivent être calculées. Aussi, nous avons procédé à un certain nombre de traitements importants avant de construire les modèles :

(a). structuration des données brutes dans une base de données Microsoft Access,

(b). filtrage et regroupement des espèces en plusieurs communautés d'oiseaux,

(c). calcul de variables réponse exprimant la diversité taxonomique et fonctionnelle.

(a). Afin de faciliter l'exploitation de la grande quantité de données, nous avons créé une base de données dont le modèle conceptuel est présenté dans la figure 7.

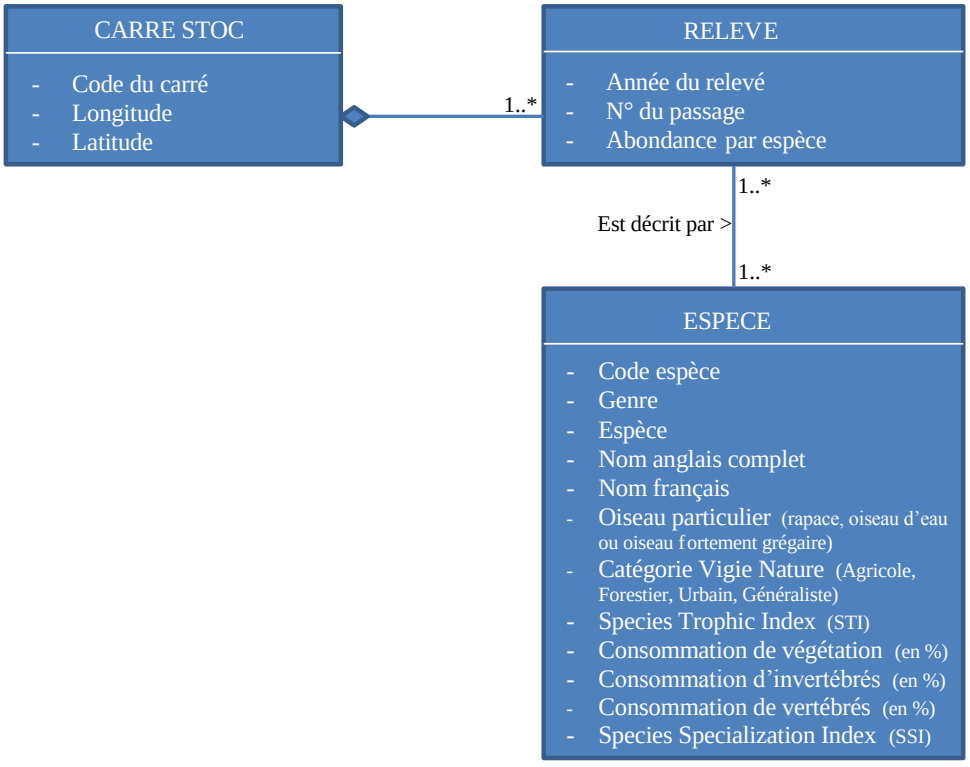


Figure 7 : Diagramme de classes UML représentant la structure de la base de données Oiseaux développée

(b). Un filtre s'est avéré nécessaire sur les données oiseaux. En effet, les rapaces et les oiseaux d'eau ne sont pas bien échantillonnés avec le protocole STOC : les rapaces ont un territoire très vaste qui va au bien au-delà des points d'écoute, et les oiseaux d'eau, de par leur besoin d'accès à des habitats variés, se déplacent également beaucoup. Afin de ne pas engendre de nombreuses fausses absences, nous avons décidé de retirer ces oiseaux de l'étude, de manière similaires à d'autres études menées sur les données STOC (Barnagaud et al., 2011; Bonthoux et al., 2013). Prins et al. (2005) a

également montré que les modèles étaient plus performants pour des espèces ayant un espace vital de taille moyenne et n'entretenant pas de relations complexes aux habitats.

En revanche, nous n'avons pas réalisé de filtre particulier dans le choix des carrés STOC de 2010 :

- L'étude se focalisant sur les différents groupes d'oiseaux spécialistes et généralistes, tous les habitats, même urbains, sont pertinents. La sélection aléatoire assure bien la présence des différents habitats dans les proportions réelles.
- Un autre filtre aurait pu être réalisé sur les carrés dont le relevé est effectué pour la première fois, afin de diminuer l'effet observateur. Cependant, il a été décidé de conserver le maximum de données disponibles, en faisant l'hypothèse que la quantité importante de données présente diminue l'impact de l'effet observateur.

Parce que différents groupes d'espèces répondent différemment aux gradients paysagers, les espèces ont été regroupées en plusieurs catégories : espèces **spécialistes des milieux agricoles**, espèces **spécialistes des milieux forestiers**, espèces **spécialistes des milieux bâtis**, e**spèces généralistes**, et une catégorie créée par la somme des quatre précédentes.

Ces regroupements sont issus du Centre de Recherches sur la Biologie des Populations d'Oiseaux (CRBPO) et détaillés sur le site de Vigie Nature (Muséum National d'Histoire Naturelle)[1]. Quelques modifications ont été apportées :

- conformément au filtre précédent, les rapaces (Faucon crécerelle, Buse variable) et oiseaux d'eau (Vanneau huppé) ont été retirés des listes,
- deux autres espèces ont également été retirées : particulièrement grégaires, leur abondance est que peu liée à l'habitat dans lequel elles se trouvent. Il s'agit du Corbeau freux et de la Corneille noire(Bonthoux et al., 2013).

L'annexe 1 fournit les listes complètes des espèces de chaque groupe.

(c). Les premières variables réponses calculées sont des indicateurs taxonomiques : pour chacune des cinq catégories, nous avons comptabilisé le nombre d'espèces différentes présentes dans chaque carré STOC. Cela correspond à la *richesse spécifique* en oiseaux agricoles (RS.Agri), en oiseaux forestiers (RS.Fores), en oiseaux urbains (RS.Urb), en oiseaux généralistes (RS.Gene) et à la richesse spécifique totale (RS.Tot).

L'abondance totale de ces différentes catégories a aussi été considérée mais nous avons décidé de seulement présenter les résultats par rapport à la richesse car le protocole des points d'écoute est *a priori* plus adapté aux présence-absence (notamment difficulté d'estimer de manière fiable l'abondance des espèces en forêt). Egalement à cause de la plus faible fiabilité du protocole pour les abondances, l'indice de diversité de Shannon n'a pas non plus été calculé. On peut toutefois noter que Levin et al. (2009) ont montré que la diversité de Shannon était corrélée à la richesse spécifique des oiseaux forestiers et non forestiers.

De plus, les données fournies ont également permis de calculer des indicateurs fonctionnels pour chaque carré STOC : le CSI et le CTI évoqués en section 1.1.

2.2. Les données d'observation de la terre

2.2.1. Description des données utilisées

Les données spatiales utilisées pour modéliser les patrons des communautés d'oiseaux à l'échelle de la France sont des images satellites issues du capteur MODIS (*Moderate-resolution Imaging Spectroradiometer*). Ce capteur est embarqué à bord des satellites EOS AM-1 (Terra) et EOS PM-1

[1] http://vigienature.mnhn.fr/page/produire-des-indicateurs-partir-des-indices-des-especes-habitat

(Aqua) mis en orbite respectivement en 1999 et 2002. Ce capteur présente la particularité de fournir des séries temporelles d'images sur toute la surface du globe avec une très haute fréquence temporelle (1 à 2 jours). Ces images sont disponibles depuis 2000 et distribuées gratuitement sur le site MODIS Web (National Aeronautics and Space Administration). Nous avons utilisé les données de 2010, de manière à ce qu'elles concordent avec les relevés de biodiversité.

Différents produits MODIS existent. Trois d'entre eux ont été retenus (tableau 1) :

- Le produit MOD13Q1 : il s'agit de synthèses d'images à 250m de résolution spatiale incluant deux indices de végétation : le NDVI et l'EVI. Ces données sont produites à partir d'images journalières de réflectance de surface corrigées des effets atmosphériques et acquises sur une période de 16 jours. Les synthèses sont élaborées à partir de la méthode dite *CV-MVC* (*Constraint View – Maximum Value Composite*) qui retient la valeur maximale de l'indice de végétation pour les valeurs d'angle de visée minimale (Huete et al., 1999).
- Le produit MOD11A2 : il s'agit d'une synthèse d'images à 1km de résolution spatiale qui fournit une information continue sur la température de surface de jour et de nuit (*Land Surface Temperature and Emissivity - LST*). La température de surface des objets au sol est estimée à partir de mesures effectuées dans l'infra-rouge thermique en appliquant un algorithme de type *split-window* (Wan et al., 2002). Pour ce produit, les synthèses sont élaborées à partir d'images acquises sur une période de 8 jours.
- Le produit MOD44B : ce produit est une image déjà classée, contrairement aux données précédentes. Il fournit pour chaque pixel le pourcentage de couverture arborée. Il est élaboré à partir des synthèses d'images à 16 jours en appliquant une procédure de classification supervisée standardisée, par arbre de décisions (Hansen et al., 2002). Le produit final est mono-date, à 250m de résolution spatiale, et mis à jour chaque année.

Caractéristiques techniques	**Couche NDVI et EVI**	**Couche de LST**	**Couche de VCF**
Nom de couche	MOD13Q1	MOD11A2	MOD44B
Format de couche	Raster (hdf)		
Dalles utilisées	h18v04, h18v03 et h17v04 (France métropolitaine)		
Résolution spatiale	250 m	1 000 m	250 m
Nombre de pixels présents sur la France	7 616 000	1 505 000	7 616 000
Résolution temporelle	Synthèse à 16 jours	Synthèse à 8 jours	Synthèse annuelle
Nombre d'images pour l'année 2010	23	46	1
Nom et descriptif des couches contenant les données utilisées	- Pour le NDVI : *250m 16 days NDVI* ; valeur du NDVI x 1000 - Pour l'EVI : *250m 16 days EVI* ; valeur de l'EVI x 1 000	- Pour le LST de jour : *LST_Day_1km* ; température de jour en Kelvin - Pour le LST de nuit : *LST_Night_1km* ; température de nuit en Kelvin	*Percent Tree Cover* ; pourcentage de couverture arborée (entre 0 et 100), ou 200 sur les zones d'eau
Nom des couche(s) contenant l'information qualité	*250m 16 days pixel reliability summary QA*	- Pour le LST de jour : *QC_Day* - Pour le LST de nuit : *QC_Night*	*Quality*

Tableau 1 : Caractéristiques techniques des données satellites utilisées

Plusieurs séries temporelles d'images sont donc utilisées dans cette étude. Selon les cas, toutes ces données ou une partie seulement sont introduites dans les modèles oiseaux/paysage élaborés.

2.2.2. Interprétation des variables brutes

Afin de mieux illustrer la signification des images retenues du capteur MODIS, nous avons réalisé un focus sur une zone de la France présentant les valeurs de quelques pixels de ces images pour des éléments paysagers variés. Les figures 8 à 12 présentent les données de NDVI, EVI, LSTD et LSTN aux dates du 6 mars 2010 (couche 065), du 26 juin (177) et du 14 septembre (257).

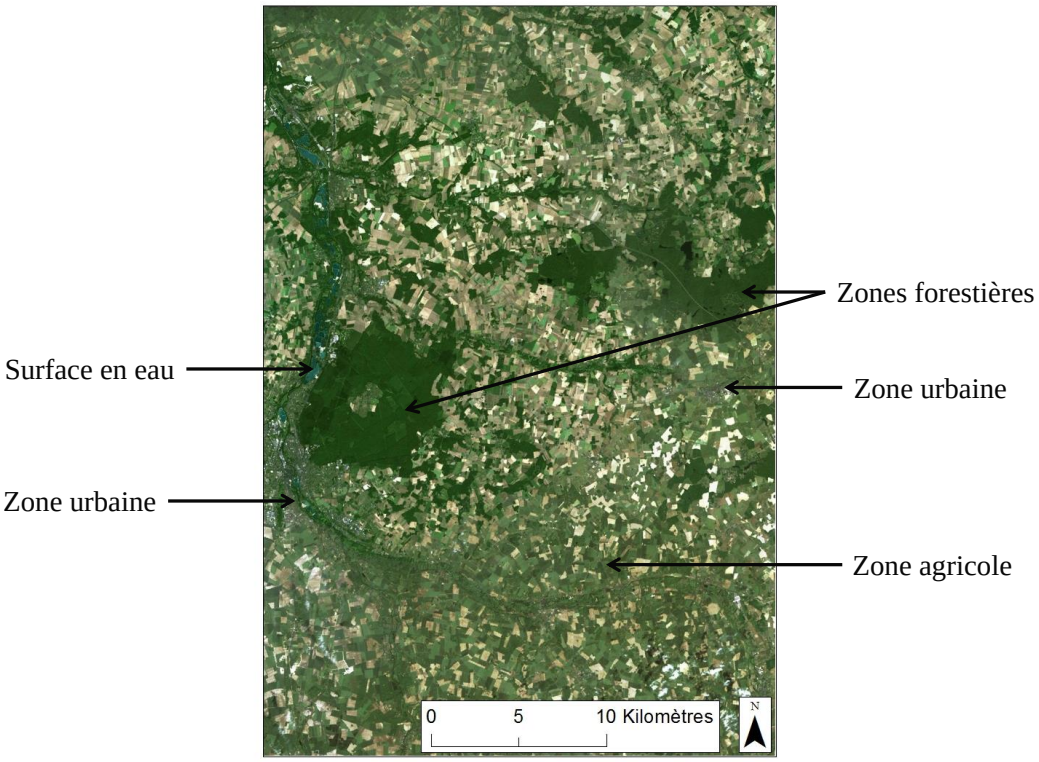


Figure 8 : Photographie aérienne de la zone présentée sur les figures 9 à 12 (source : BDORTHO, IGN 2010)

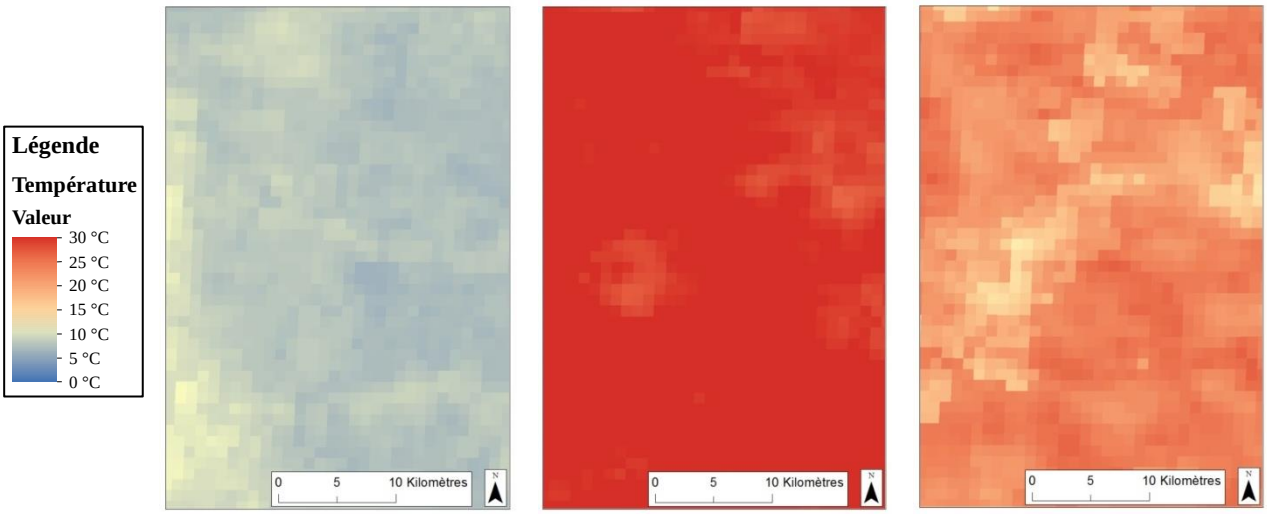


Figure 9 : Extrait des couches de LSTD aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite)

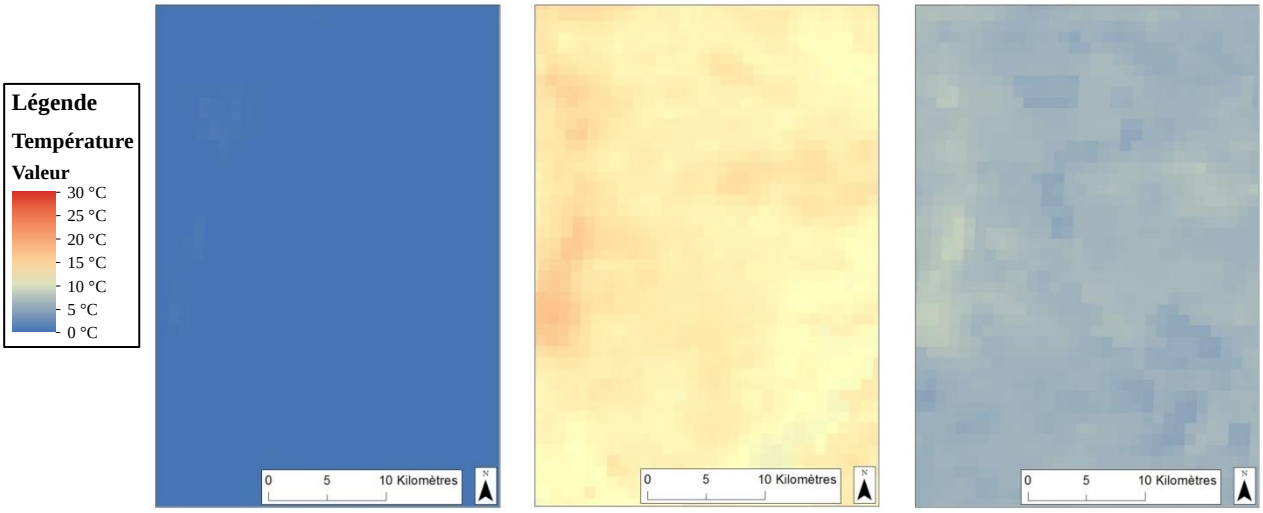


Figure 10 : Extrait des couches de LSTN aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite)

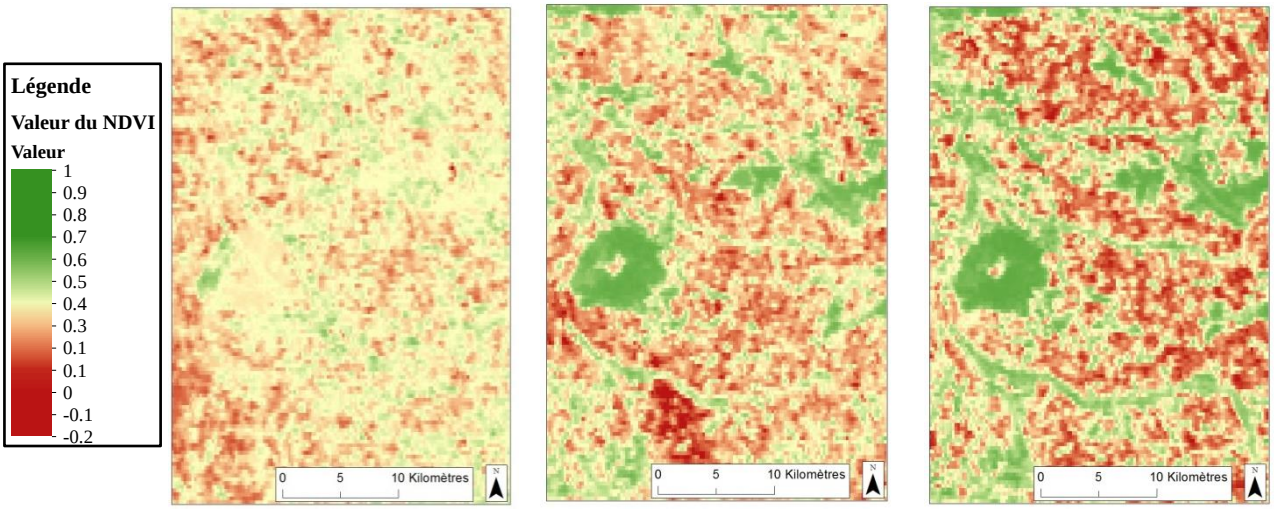


Figure 11 : Extrait des couches de NDVI aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite)

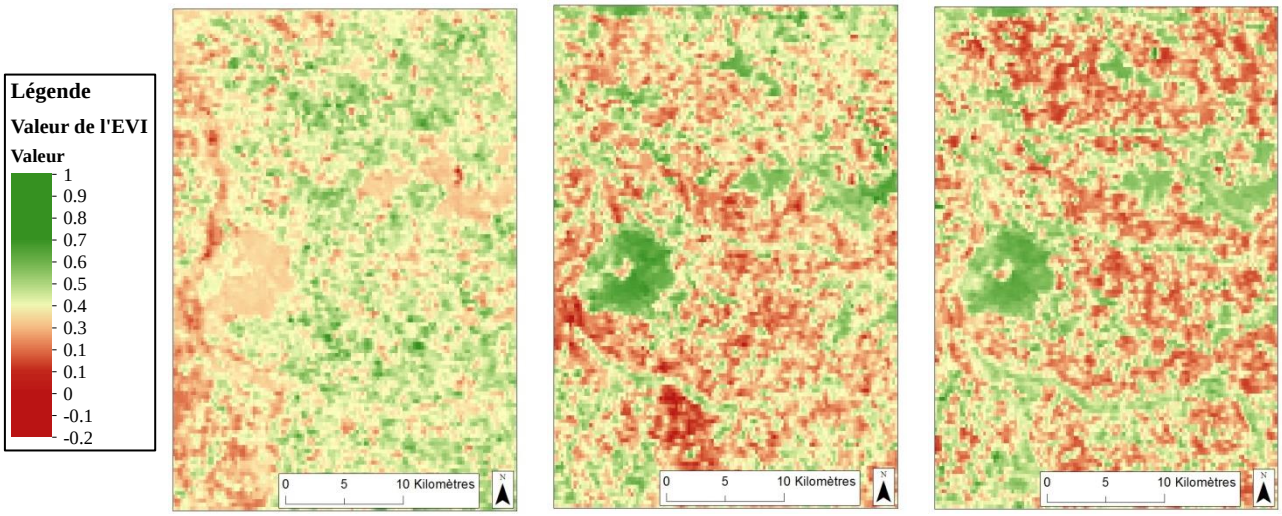


Figure 12 : Extrait des couches d'EVI aux dates du 6 mars (à gauche), du 26 juin (au milieu) et du 14 septembre (à droite)

En ce qui concerne les couches de température de surface, on constate que globalement, pour une date donnée, le LSTN fournit des valeurs beaucoup plus homogènes entre les milieux, tandis que le LSTD est beaucoup plus discriminant. Le LSTD montre des différences entre les habitats :

- les zones forestières sont surtout marquées par des températures plus fraiches en été (le milieu fermé joue un rôle tampon sur la température) ;
- les zones urbaines présentent souvent une température plus élevée que les autres habitats, quel que soit la saison ;
- les milieux ouverts sont plus froid en hiver et plus chaud en été.

Les différences entre le NDVI et l'EVI sont relativement faibles. L'EVI est plus sensible à la structure de la canopée alors que le NDVI est plus sensible à la vigueur de la végétation. L'EVI semble discriminer moins fortement les habitats, pour une date donnée. On peut cependant noter que les surfaces en eau ressortent mieux avec l'EVI, notamment en hiver.

On observe que :

- les zones forestières donnent un indice (NDVI ou EVI) faible en hiver, et très fort en été ;
- les surfaces en eau donnent un indice très faible ;
- les zones agricoles donnent des indices très hétérogènes selon les parcelles et les périodes de l'année.

2.2.3. Prise en compte de la qualité des données

Les conditions atmosphériques et météorologiques sont très variables à l'échelle de la France et dans l'espace d'une année. Ces conditions (présence de nuages notamment) impactent fortement la qualité des images obtenues, celle-ci est donc très variable selon la date et l'emplacement. Aussi, il est nécessaire de réaliser un filtre des données, de manière à ne conserver que des carrés STOC sur lesquels la qualité des images est bonne.

L'information sur la qualité des pixels dans les images est fournie avec les produits MODIS. Cette information est spécifique aux séries temporelles d'images : habituellement, elle n'est pas nécessaire car l'acquisition de l'image est programmée pour avoir en général moins de 5 % de nuages. Dans le cas des séries temporelles, l'acquisition des images est systématique, quelle que soit la couverture nuageuse. Il est donc important de fournir à l'utilisateur une information sur la qualité des données à l'échelle du pixel.

Qualité des données de MOD13Q1 (NDVI/EVI)

Pour étudier l'effet de la date des images sur la performance des modèles, nous souhaitons sélectionner quelques dates correspondant à des états différents de la végétation et de l'occupation du sol (ex : passage d'une culture en labour). Afin d'éviter tout biais dans les modèles, nous souhaitons que les dates sélectionnées correspondent à des synthèses contenant un maximum de pixels de bonne qualité. En effet, la qualité des données influe fortement sur les valeurs de NDVI et EVI présentes dans les images. Ainsi, sur la figure 13, les zones nuageuses ou recouvertes de neige ou de glace présentent une valeur particulièrement élevée de NDVI, valeur qui n'est pas directement liée à la végétation. A l'inverse, des valeurs particulièrement faibles sont observées au niveau de la Bretagne, et sous forme de bande. Ces valeurs ne doivent pas être prises en compte dans la construction des modèles.

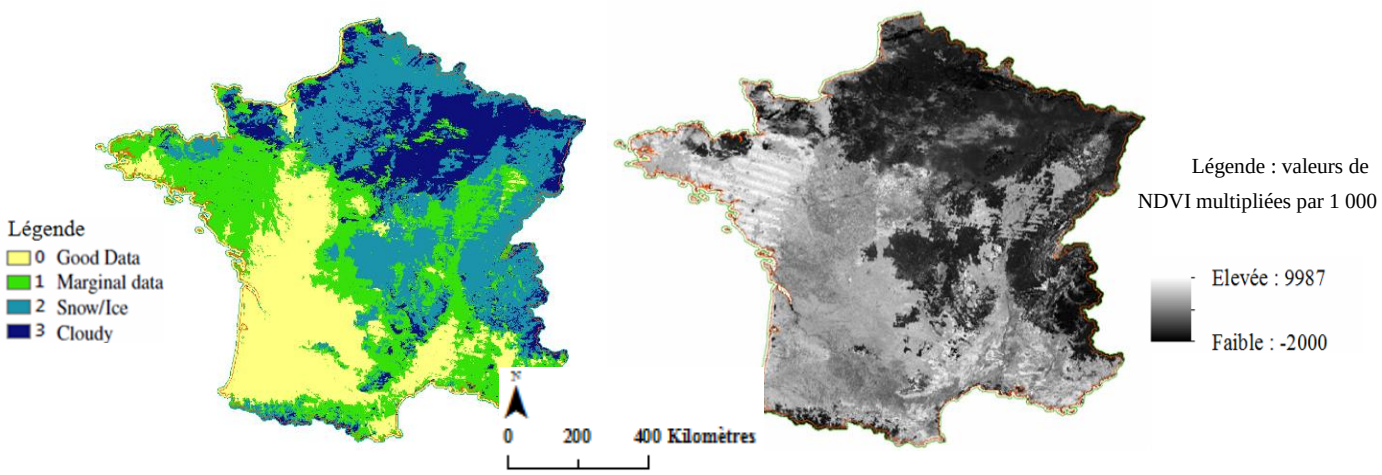
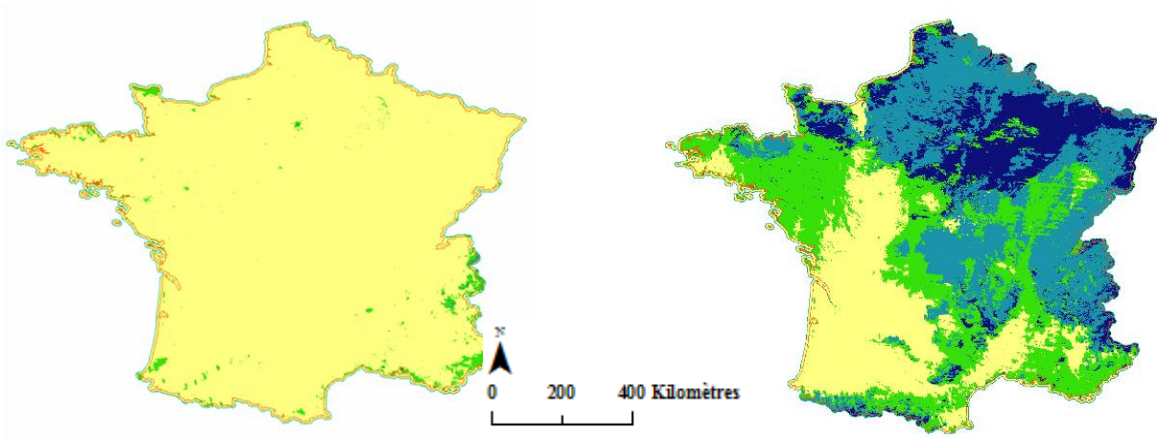


Figure 13 : Impact de la fiabilité des données sur les valeurs de NDVI (synthèse du 19/12/2010)

Dans le produit MOD13Q1, l'indice de qualité de l'EVI est le même que celui du NDVI. Il est contenu dans la couche « pixel reliability » : elle fournit par pixel et par synthèse à 16 jours une valeur entre 0 et 3. Plus cette valeur est proche de 0, meilleure est la qualité. Selon les dates des synthèses, la qualité de la couche peut très fortement varier, comme l'illustre la figure 14.



Fiabilité des pixels du produit MOD13Q1 pour la synthèse du 30/09/2010

Fiabilité des pixels du produit MOD13Q1 pour la synthèse du 19/12/2010

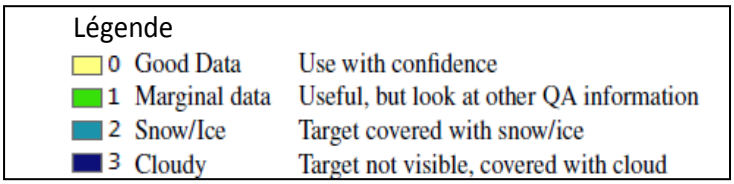


Figure 14 : Variabilité de la qualité des données de NDVI et EVI selon la date

Afin de choisir un ensemble de dates présentant un bon compromis entre qualité de données et répartition sur l'année, nous avons construit un graphique (figure 15) montrant la valeur moyenne de l'indice de qualité des différents produits MODIS de 2010 sur la France entière.

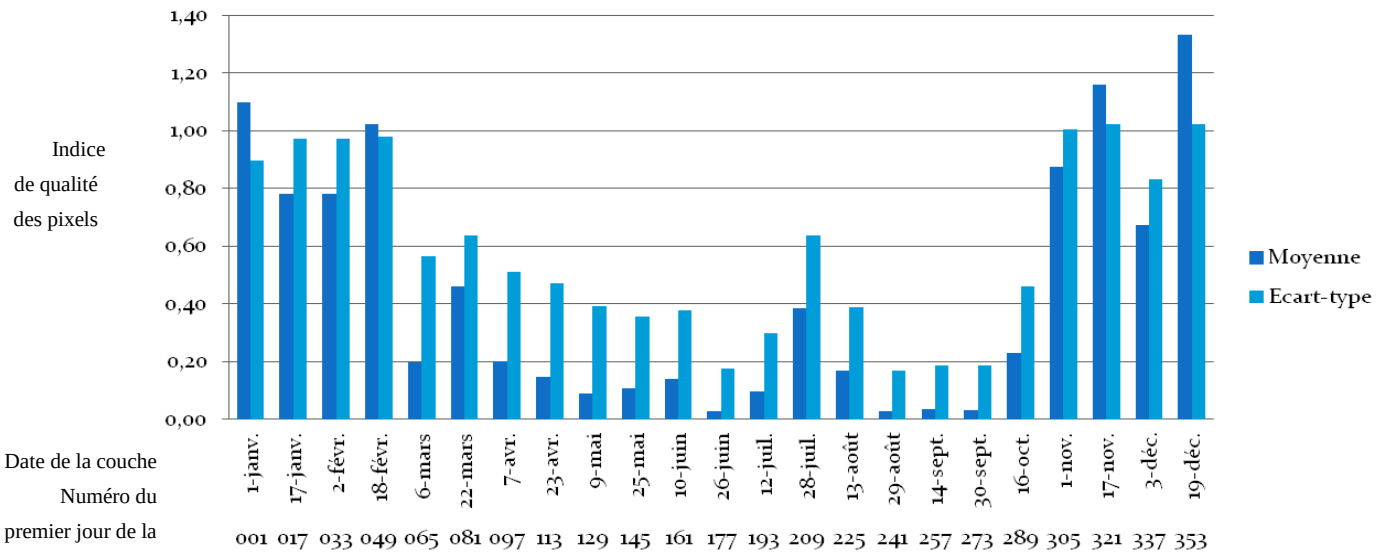


Figure 15 : Moyenne et écart-type de la fiabilité des données du produit MOD13Q1 à l'échelle de la France pour l'année 2010

En hiver, une seule couche présente une qualité acceptable, celle du 6 mars. C'est elle qui est retenue. Quatre autres couches sont de très bonne qualité : celles du 26 juin, puis les trois dates consécutives du 29 août, 14 septembre et 30 septembre. La date du 26 juin est la plus proche de celle des relevés STOC, elle est donc conservée pour cette raison. Les dates de fin d'été / début automne sont également sélectionnées car (1) elles correspondent approximativement aux dates de mises en labour d'une partie des cultures (variables selon les régions). Cela permet de distinguer les cultures des surfaces en herbe dans les images NDVI et EVI ; (2) cette période a déjà été identifiée comme la plus appropriée pour prédire des patrons de communautés d'oiseaux à partir du NDVI (Sheeren et al., 2014).

Finalement, cinq couches présentant la meilleure qualité sont retenues pour les données NDVI et EVI issues du produit MOD13Q1 de 2010 :

- la couche 065 du 6 mars 2010
- la couche 177 du 26 juin 2010
- la couche 241 du 29 aout 2010
- la couche 257 du 14 septembre 2010
- la couche 273 du 30 septembre 2010

Sur chacune de ces cinq dates, il reste au sein des carrés STOC des pixels pour lesquels les valeurs de NDVI et d'EVI ne sont pas fiables. Ne sont conservés dans la suite de l'analyse que les carrés STOC sur lesquels 100 % des pixels sont de bonne qualité pour les cinq dates, soit 904 carrés sur un total de 1 091.

Qualité des données de MOD11A2 (LST)

Après avoir sélectionné les carrés STOC ne contenant que des pixels de bonne qualité aux dates choisies (synthèses NDVI/EVI à 16 jours), nous avons vérifié la qualité des pixels correspondant pour le produit LST (synthèses à 8 jours, pixel de 1 km de résolution). Dans ce produit, un indice de qualité est fourni pour la température de jour, et un autre indice est fourni pour les températures de nuit. La qualité est considérée comme acceptable si l'erreur moyenne de la température de surface est inférieure à 2°C.

Parmi les carrés STOC précédemment conservés, nous avons éliminé ceux contenant au moins un pixel de mauvaise qualité dans l'une des synthèses à 8 jours et ceci, pour les températures de jour et de nuit. Toutefois, pour une période de temps définie par les synthèses NDVI/EVI à 16 jours, deux synthèses LST à 8 jours sont disponibles. Dans ce cas, nous avons retenu le carré STOC si la température de surface de l'une des deux synthèses à 8 jours était de bonne qualité.

A cette étape, 887 carrés STOC ont été conservés pour l'analyse.

Qualité des données de MOD44B (VCF)

Le produit VCF annuel est également accompagné d'une information qualité à l'échelle du pixel. Les données de qualité précisent le nombre de données d'entrée utilisées (synthèses à 16 jours) pour lesquelles la qualité est mauvaise (présence de nuages, ombres portées, aérosol important, angle zénithal de plus de 45°). Nous avons considéré la qualité comme étant satisfaisante à l'échelle du carré STOC si le nombre moyen de données d'entrée de mauvaise qualité pour chacun des pixels est strictement supérieur à 4, chaque carré STOC contenant entre 48 et 66 pixels VCF.

Finalement 867 carrés STOC communs aux trois produits MODIS sont conservés.

2.2.4. Construction des variables d'état

Une fois la qualité des images prise en compte et les carrés STOC sélectionnés, nous avons construit les variables explicatives à partir des images. Dans un premier temps, nous nous sommes focalisés sur des variables d'état, qui donnent des informations sur l'état de la végétation et sur les patrons paysagers à une date donnée. Pour cela, des indicateurs statistiques sont calculés sur les pixels situés dans chaque carré STOC, comme l'illustre la figure 16. Un pixel est considéré à l'intérieur d'un carré si son centre est dans le carré.

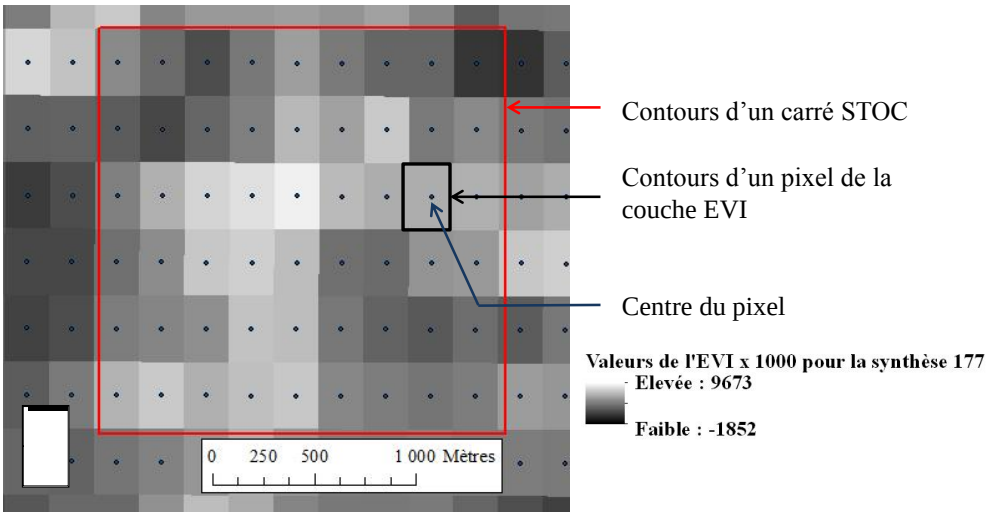


Figure 16 : Présentation d'un carré STOC (N°45-1293) en lien avec la synthèse EVI (26 juin 2010)

Sur chaque carré, nous avons pu calculer la moyenne, le minimum, le maximum, la médiane, l'amplitude et l'écart-type du LSTD, du LSTN, de l'EVI et du NDVI pour chacune des cinq dates sélectionnées précédemment (cf. 2.2.3.), ainsi que du VCF. Cependant, plusieurs de ces variables sont corrélées. Nous avons donc décidé de ne garder que la moyenne et l'écart-type pour la suite de l'étude : la moyenne donne une information sur l'habitat dominant, tandis que l'écart-type représente

l'hétérogénéité spatiale des habitats présents. Ce choix est également conforté par la littérature (St-Louis et al., 2006). De plus, les autres variables sont fortement corrélées à ces deux variables.

On obtient la liste de variables présentée dans le tableau 2 :

Produit MODIS	Type de données	Type de variable explicative	Nom des variables
MOD13Q1	NDVI	Moyenne	MoyNDVI067, MoyNDVI177, MoyNDVI241, MoyNDVI257, MoyNDVI273
		Ecart-type	EcaNDVI067, EcaNDVI177, EcaNDVI241, EcaNDVI257, EcaNDVI273
MOD13Q1	EVI	Moyenne	MoyEVI067, MoyEVI177, MoyEVI241, MoyEVI257, MoyEVI273
		Ecart-type	EcaEVI067, EcaEVI177, EcaEVI241, EcaEVI257, EcaEVI273
MOD11A2	LSTD	Moyenne	MoyLSTD067, MoyLSTD77, MoyLSTD241, MoyLSTD257, MoyLSTD273
		Ecart-type	EcaLSTD067, EcaLSTD177, EcaLSTD241, EcaLSTD257, EcaLSTD273
MOD11A2	LSTN	Moyenne	MoyLSTN067, MoyLSTN77, MoyLSTN241, MoyLSTN257, MoyLSTN273
		Ecart-type	EcaLSTN067, EcaLSTN177, EcaLSTN241, EcaLSTN257, EcaLSTN273
MOD44B	VCF	Moyenne	MoyVCF
		Ecart-type	EcaVCF

Les chiffres dans les noms des variables correspondent au n^{ème} jour de l'année : 067 correspond à la fin de l'hiver, 177 au début du printemps et 241,257 et 273 à la fin du printemps/début de l'automne.

Tableau 2 : Liste des variables explicatives d'état

A noter :

Pour le VCF, les valeurs à 200 – c'est-à-dire les zones d'eau – ont été mises à 0, de manière à rester cohérent sur la signification de cette donnée, le pourcentage de couverture arborée, qui est donc de 0% lorsqu'on est sur une zone d'eau.

Concernant le LST, les produits sont fournis par semaine et non par quinzaine. Afin de rester homogène avec les données de NDVI/EVI, les calculs de moyenne et d'écart-type se sont basés sur un produit construit comme la moyenne des deux synthèses de 8 jours.

Le logiciel ArcGis 10.2 a été utilisé pour réaliser le traitement des données.

2.2.5. Construction de classes de fonctionnement phénologique

L'objectif est d'évaluer si la prise en compte de l'évolution temporelle de la végétation au cours de l'année (i.e. phénologie) est plus pertinente pour expliquer les patrons des communautés d'oiseaux que simplement considérer les métriques de végétation à une seule date donnée. Pour étudier cela, nous avons utilisé les séries temporelles du NDVI.

Pour établir des classes de fonctionnement phénologique, une méthode de classification non supervisée a été mise en œuvre : l'algorithme K-means, adapté aux données fonctionnelles ou

longitudinales (Genolini and Falissard, 2011). L'algorithme prend en entrée les signatures temporelles de NDVI des pixels, et il regroupe les signatures similaires.

Le nombre de pixels sur la France entière est supérieur à 7 millions, ce qui nécessite un temps de calcul très important. Il a donc été décidé de restreindre la quantité de pixels à classer, et de ne garder que les pixels présents dans les 867 carrés STOC utilisés dans les modèles précédents : cela assure un minimum de qualité aux données, tout en gardant un échantillonnage représentatif des habitats français.

L'utilisateur doit fixer un certain nombre de *paramètres* dans l'application de cet algorithme. Un des paramètres les plus sensibles pour ce type de classification non supervisée est le nombre de clusters à créer. Un certain nombre minimal de groupes peut être estimé *a priori* : nous souhaitons distinguer *a minima* les prairies, les zones sans végétation (zones urbaines ou eau), les forêts, les zones de cultures. Le nombre minimal de clusters à créer est donc posé à 4. Toutefois, l'objectif n'est pas ici de faire une classification d'occupation du sol mais bien de distinguer des classes de fonctionnement par catégorie d'occupation (différentes phénologies de cultures par exemple). Pour le nombre maximal de clusters, il s'agit donc de trouver un compromis entre un nombre suffisant de classes pour dépasser la simple distinction des couverts, et un grand nombre de classes qui peut engendrer un temps de traitement très important. Nous avons choisi un nombre maximal de clusters de 16.

Les principaux autres paramètres fixés sont les suivants :

- Le nombre de lancements de l'algorithme de classification avec différentes configurations de départ, pour chaque nombre de clusters : 5
- Le nombre maximal d'itérations si la convergence n'est pas atteinte : 100
- La fonction de calcul de la distance entre deux courbes : distance euclidienne
- La méthode spécifiant les conditions de départ : cette méthode est importante car un défaut majeur de ces algorithmes est le risque de converger vers un maximum local qui ne corresponde pas à la meilleure partition possible en termes d'homogénéité (MacQueen, 1967; Selim and Ismail, 1984). Ici, tous les pixels sont initialement assignés de manière aléatoire à un cluster, avec au moins un pixel dans chacun des clusters.

Une fois la classification réalisée, nous avons créé les variables explicatives en calculant le pourcentage de pixels appartenant à chaque classe de fonctionnement de végétation dans chacun des carrés STOC : Pourcent_A, Pourcent_B, Pourcent-C, Pourcent_D, etc.

2.3. Les données d'occupation du sol

Les modèles de biodiversité sont aujourd'hui plus couramment réalisés à partir de variables issues de cartographies d'occupation du sol - elles-mêmes étant réalisées par classification ou photo-interprétation des images brutes. Pour comparer ces modèles à ceux créés à partir de variables directement issues non classés (produits MODIS précédents), il est nécessaire de construire un modèle construit à partir de variables d'occupation du sol.

A l'heure actuelle, la seule base cartographique d'occupation des sols couvrant toute la France est celle de CORINE Land Cover. Il s'agit d'un projet européen lancé en 1985 et qui propose une base de données vectorielle librement accessible de l'occupation du territoire métropolitain français pour les années 1990, 2000 et 2006. Les données d'occupation des sols utilisées pour la suite de l'étude sont donc celles de 2006 (figure 17), date la plus proche des relevés de biodiversité (2010). La surface minimale de cette base est de 25 ha et sa résolution est inférieure à 25 m.

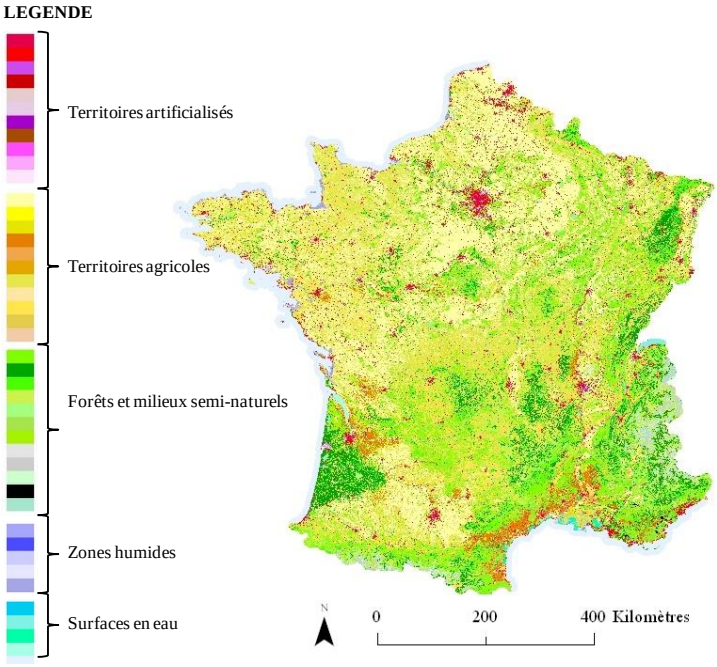


Figure 17 : Carte d'occupation des sols français (CORINE Land Cover 2006, légende simplifiée)

CORINE Land Cover possède une nomenclature hiérarchisée en trois niveaux. Dans l'idéal, tous les niveaux de la nomenclature sont conservés de manière à obtenir le modèle avec le meilleur ajustement possible aux données. Cependant, un trop grand nombre de variables implique également un fort risque de colinéarité, ce qui induit un biais dans les estimations des paramètres (Dormann et al., 2013). Il a donc été décidé de ne garder que le niveau 2 de la nomenclature, à l'exception des forêts pour lesquelles c'est le troisième niveau qui est retenu, cette information étant pertinente, notamment pour l'étude des oiseaux forestiers. Le tableau 3 récapitule les différentes variables conservées.

Nom des variables	Description de la variable
Part_11	Proportion surfacique du carré STOC contenant un habitat de nomenclature CORINE Land Cover 1.1. : Zones urbanisées
Part_12	Proportion de l'habitat 1.2. : Zones industrielles ou commerciales et réseaux de communication
Part_13	Proportion de l'habitat 1.3. : Mines, décharges et chantiers
Part_14	Proportion de l'habitat 1.4. : Espaces verts artificialisés, non agricoles
Part_21	Proportion de l'habitat 2.1. : Terres arables
Part_22	Proportion de l'habitat 2.2. : Cultures permanentes
Part_23	Proportion de l'habitat 2.3. : Prairies
Part_24	Proportion de l'habitat 2.4. : Zones agricoles hétérogènes
Part_311	Proportion de l'habitat 3.1.1. : Forêts de feuillus
Part_312	Proportion de l'habitat 3.1.2. : Forêts de conifères
Part_313	Proportion de l'habitat 3.1.3. : Forêts mélangées
Part_32	Proportion de l'habitat 3.2. : Milieux à végétation arbustive et/ou herbacée
Part_33	Proportion de l'habitat 3.3. : Espaces ouverts, sans ou avec peu de végétation
Part_41	Proportion de l'habitat 4.1. : Zones humides intérieures
Part_42	Proportion de l'habitat 4.2. : Zones humides maritimes
Part_51	Proportion de l'habitat 5.1. : Eaux continentales
Part_52	Proportion de l'habitat 5.2. : Eaux maritimes

Tableau 3 : Noms et descriptions des variables explicatives issues de CORINE Land Cover

Afin de comparer avec les modèles construits à partir du produit VCF de MODIS (cf. section 2.2.4), un modèle construit uniquement à partir de l'information « présence de forêt » est également réalisé. Pour cela, la variable utilisée correspond à la somme des variables Part_311, Part_312 et Part_313 : la variable Part_31.

2.4. Les données de stratification

Dans les modèles à construire à partir des données d'observation de la terre, au vu de l'échelle de travail (France entière), une régionalisation des effets observés peut être attendue (exemple : l'intensité de l'effet du NDVI pour expliquer la richesse des oiseaux forestiers peut varier en fonction des régions). Une stratification du territoire selon différents points de vue a donc été recherchée. Cette stratification peut aussi limiter la corrélation spatiale dans les résidus des modèles. Trois catégories de régions ont été retenues :

- des sylvo-éco-régions,
- des régions agricoles,
- des régions climatiques.

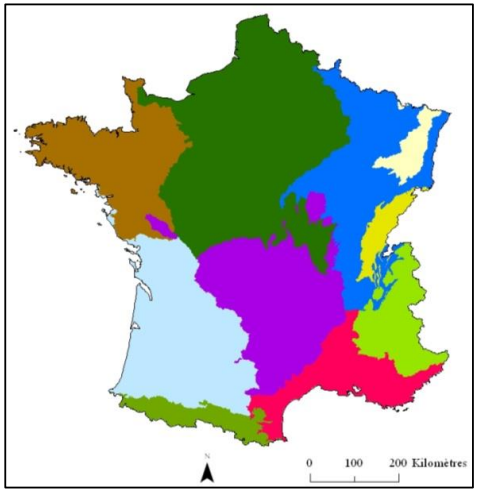
Les sylvo-éco-régions ont été récupérées auprès de l'Institut national de l'information géographique et forestière (IGN). Pour les régions agricoles, l'Institut de l'élevage a régionalisé la France en 7 grandes zones (Rouquette and Pflimlin, 1995). La carte des zones a pu être reconstituée en passant par les Petites Régions Agricoles qui sont chacune affectées à une zone. En ce qui concerne les régions climatiques, Joly et al. (2010) ont construit une carte typologique de huit climats en France.

Ces différentes sources sont présentées en figure 18.

Une région est affectée à chaque carré STOC. Lorsqu'un carré est localisé à l'interface entre plusieurs régions, c'est la région la plus représentée sur le carré qui est considérée. Trois variables qualitatives de régionalisation sont ainsi créées :

- le code de la région climatique des carrés STOC est donné par la variable **reg_clim** (11 modalités),
- le code de la région agricole est donné par la variable **reg_agri** (13 modalités : les zones 1, 2, 4, 5 et 6 sont subdivisées en 2),
- le code de la sylvo-écorégion est donné par la variable **reg_sylvo** (8 modalités).

Répartition des sylvoécorégions

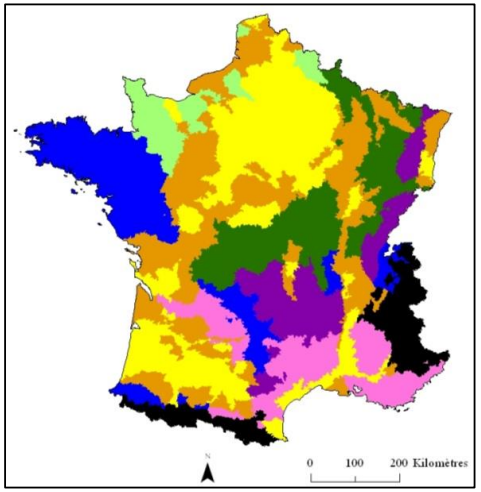


Légende :

- GRECO A : Grand Ouest cristallin et océanique
- GRECO B : Centre Nord semi-océanique
- GRECO C : Grand Est semi-continental
- GRECO D : Vosges
- GRECO E : Jura
- GRECO F : Sud-ouest océanique
- GRECO G : Massif Central
- GRECO H : Alpes
- GRECO I : Pyrénées
- GRECO J : Méditerranée

GRECO = Grande Région Ecologique

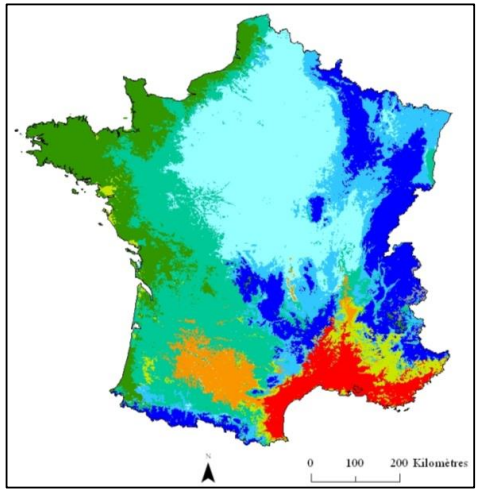
Répartition des régions agricoles



Légende :

- Zone 0 : Grandes cultures ou absence d'élevage
- Zone 1 : Cultures et élevages
- Zone 2 : Cultures fourragères (herbe et maïs)
- Zone 3 : Zone herbagère du Nord-Ouest
- Zone 4 : Zone herbagère du Centre et de l'Est
- Zone 5 : Zones pastorales
- Zone 6 : Montagnes humides
- Zone 7 : Haute-Montagne

Répartition des régions climatiques



Légende :

- Type 1 : les climats de montagne
- Type 2 : le climat semi-continental et le climat des marges montagnardes
- Type 3 :le climat océanique dégradé des plaines du Centre et du Nord
- Type 4 : le climat océanique tempéré
- Type 5 : le climat océanique franc
- Type 6 : le climat méditerranéen altéré
- Type 7 : le climat du bassin du Sud-Ouest
- Type 8 : le climat méditerranéen franc
- Hors interpolation

Figure 18 : Présentation des trois cartes de régionalisation de la France

2.5. Construction des modèles

Après avoir récupéré et mis en forme toutes les données nécessaires, nous avons élaboré plusieurs modèles visant à répondre aux objectifs fixés.

2.5.1. Démarche générale

Pour chacune des questions posées (cf. section 1.3.), différentes séries de modèles sont créées, chaque série comprenant des modèles pour les sept variables réponses relatives aux relevés oiseaux présentées au 2.1.2. Pour cela, les étapes suivantes sont réalisées :

(a). Choix des variables explicatives pertinentes pour répondre à la question posée. Ce point dépend de chaque objectif, l'idée étant de comparer des modèles judicieusement choisis pour répondre aux questions.

(b). Analyse des corrélations entre les variables explicatives choisies et retrait de certaines d'entre elles si nécessaire. L'analyse des corrélations se fait par l'étude du coefficient de corrélation de Pearson (R) : deux variables sont considérées corrélées si $R > 0,7$ (Dormann et al., 2013).

(c). Construction de modèles linéaires généralisés (GLM). Les variables RS.Agri, RS.Fores, RS.Urb, RS.Gene et RS.Tot sont des données discrètes de comptage. La distribution la plus adaptée est la distribution de Poisson. Les variables CSI et CTI sont continues, et leur distribution est gaussienne.

(d). Sélection éventuelle des variables dans les modèles. Une fonction permet, par ajout et suppression successives des différentes variables explicatives, de fournir le meilleur modèle. Les modèles sont comparés entre eux par le critère d'information d'Akaike (AIC, Akaike Information Criteria) qui indique la parcimonie du modèle en considérant le niveau d'ajustement du modèle aux données tout en prenant en compte le nombre de paramètres utilisés. C'est un indicateur de l'équilibre entre la précision d'un modèle et sa simplicité. Le modèle possédant le plus faible AIC est le meilleur modèle.

(e). Comparaison des modèles obtenus entre eux. Les modèles sont comparés entre eux également à partir du critère AIC. L'AICc n'a pas été analysé, le nombre d'échantillons étant suffisamment important (Hurvich and Tsai, 1989). Les modèles sont jugés significativement différents si la différence d'AIC est supérieure à 2 (Burnham and Anderson, 2008). On peut également noter que lorsque les modèles à comparer ont le même nombre de paramètres, cela revient à utiliser un indicateur d'ajustement des modèles aux données. L'indicateur utilisé dans ce cas est le pourcentage de déviance expliquée (ci-après noté D^2) : plus ce pourcentage est élevé, meilleur est l'ajustement aux données.

(f). Etude du pouvoir prédictif de modèles. Le D^2 ne donne qu'une information sur la performance explicative des modèles (i.e. ajustement des modèles aux données). Pour certains d'entre eux, la performance prédictive des modèles a également été calculée. Le principe est le suivant : pour chaque modèle, une sélection aléatoire de deux-tiers des carrés STOC (578) a été réalisée. Le modèle est construit à partir de ce groupe de calibration. Il est ensuite appliqué au groupe d'évaluation constitué des 289 données restantes. Les valeurs de biodiversité ainsi prédites sont comparées à celles qui ont été mesurées sur le terrain. Cette comparaison est réalisée de deux manières différentes :

- par mesure de corrélation de rang de Spearman (rho) entre les données calculées et les données mesurées : plus elle est élevée, meilleure est le pouvoir prédictif du modèle ;
- par calcul de l'erreur quadratique moyenne : plus elle est faible, meilleur est le pouvoir prédictif du modèle.

Afin d'améliorer la robustesse des résultats, toute cette procédure depuis la sélection aléatoire des 578 carrés STOC est réalisée 100 fois. On calcule ensuite pour chaque modèle une moyenne de la corrélation de Spearman (rho moyen) et une moyenne de l'erreur quadratique moyenne (RMSE moyen). Ces deux moyennes permettent de comparer entre eux la performance prédictive des modèles obtenus (Sheeren et al., 2014).

Pour l'analyse statistique des données, c'est le logiciel R qui est utilisé (3.0.2).

2.5.2. Impact de la date sur la performance des modèles

Pour répondre à cette question, quatre types de variables continues sont étudiés : le NDVI, l'EVI, le LST de jour (LSTD) et le LST de nuit (LSTN). L'objectif de cette partie est d'analyser pour chacune d'entre elles si une date parmi les cinq préalablement sélectionnées (cf. 2.2.3.) fournit des modèles explicatifs plus performants.

Les modèles construits ici sont donc simples, du format suivant (exemple du LSTD à la date 177 et la variable réponse de richesse spécifique en oiseaux agricoles) :

RS.Agri ~ EcaLSTD177 + MoyLSTD177

Cela permet de comparer pour chaque variable réponse et pour chaque type de variable les modèles réalisés à différentes dates.

Aucune corrélation n'a été observée entre les variables d'écart-type et de moyenne pour un type de variable et une date donnée.

2.5.3. Prise en compte de l'effet région

Les modèles précédents sont construits sans prendre en compte l'étendue importante de l'étude. Or plusieurs auteurs ont souligné l'importance d'une approche par écorégion (Lawler et al., 2004; Rittenhouse et al., 2012). Dans cette partie, on étudie l'apport de la stratification des données sur la performance des modèles.

(a). Effet région

Tout d'abord se pose la question de savoir si les régions ont un impact sur la répartition des différentes variables réponses. A ce stade, on utilise les trois types de stratifications présentés au 2.4. Les modèles sont similaires à ceux du 2.5.2., mais en rajoutant la variable région. Ainsi, le modèle suit le format suivant (exemple du LSTN pour la variable réponse CTI, à la date 241, et en utilisant les régions agricoles) :

CTI ~ MoyLSTN241 + EcaLSTN241 + reg_agri

Pour chaque variable réponse, seuls les modèles des meilleures dates sont étudiés. Pour ces modèles, la performance prédictive est également évaluée.

(b). Effet des interactions

Au-delà de l'impact des régions sur la biodiversité, il est également possible que l'effet NDVI, l'effet EVI, l'effet LSTD et l'effet LSTN varient en eux-mêmes selon les régions. Il est donc également pertinent de rajouter les termes des interactions entre les régions et les types de variables continues. Dans ce cas, les modèles prennent le format suivant (exemple de l'EVI, à la date 241, pour

la variable réponse de richesse spécifique en oiseaux généralistes, et en se basant sur la régionalisation climatique), suivant la notation de R :

RS.Gene ~ MoyEVI241 + EcaEVI241 + reg_clim + MoyEVI241 : reg_clim + EcaEVI241 : reg_clim

Sur ces modèles, la procédure stepAIC est réalisée pour sélectionner les meilleurs modèles.

2.5.4. Complémentarité entre types de variables pour une même date

Dans le cas où l'on disposerait des quatre types de variables (EVI, NDVI, LSTD et LSTN), la question se pose de savoir si pour une date donnée la combinaison de toutes les variables fournit de meilleurs modèles, et si oui, à quelle période de l'année.

L'analyse des corrélations de Pearson pour les variables de la date 065 (tableau 4) montre que la corrélation est uniquement élevée entre le NDVI et l'EVI. Les résultats sont similaires aux autres dates. Il n'y a pas de forte corrélation entre le LSTD et le LSTN. Des résultats similaires sont obtenus pour les autres dates. Dès lors, l'analyse de la complémentarité entre variables a été réalisée en procédant aux regroupements suivants : d'une part NDVI, LSTD et LSTN, et d'autre part EVI, LSTD et LSTN.

	EcaLSTD065	**MoyLSTD065**	**EcaLSTN065**	**MoyLSTN065**	**EcaNDVI065**	**MoyNDVI065**	**EcaEVI065**	**MoyEVI065**
EcaLSTD065	1	-0,29	0,18	-0,21	0,24	-0,26	0,11	-0,26
MoyLSTD065	-0,29	1	-0,33	0,56	-0,23	0,26	-0,25	0,25
EcaLSTN065	0,18	-0,33	1	-0,21	0,21	-0,26	0,25	-0,26
MoyLSTN065	-0,21	0,56	-0,21	1	-0,02	0,21	-0,19	0,1
EcaNDVI065	0,24	-0,23	0,21	-0,02	1	-0,26	0,6	-0,22
MoyNDVI065	-0,26	0,26	-0,26	0,21	-0,26	1	-0,24	0,81
EcaEVI065	0,11	-0,25	0,25	-0,19	0,6	-0,24	1	0,07
MoyEVI065	-0,26	0,25	-0,26	0,1	-0,22	0,81	0,07	1

Tableau 4 : Corrélation de Pearson entre les variables de NDVI, EVI, LSTD et LSTN pour la date 065 (6 mars 2010)

Les dix modèles ainsi construits sont de la forme suivante (exemple pour la date 065 et la variable réponse de richesse spécifique en oiseaux forestiers) :

RS.Fores ~ EcaNDVI065 + MoyNDVI065 + EcaLSTD065 + MoyLSTD065 + EcaLSTN65 + MoyLSTD065

et

RS.Fores ~ EcaEVI065 + MoyEVI065 + EcaLSTD065 + MoyLSTD065 + EcaLSTN65 + MoyLSTD065

La procédure de stepAIC est ici réalisée, et permet de conserver les modèles les plus parcimonieux.

2.5.5. Complémentarité entre dates pour un même type de variables

Dans le cas où seul un type de variable serait à disposition (ex : le NDVI), la question est alors de voir d'une part, si un modèle prenant en compte la combinaison de plusieurs dates est plus performant qu'un modèle simple, et d'autre part, si la complémentarité entre les dates est meilleure pour certains types de variables.

L'analyse des corrélations de Pearson montre que les corrélations entre variables d'un même type à différentes dates sont très faibles, à l'exception des trois dates qui se suivent d'août et septembre.

Les modèles construits sont de la forme suivante (exemple du type de variable NDVI et la variable réponse de richesse spécifique en oiseaux urbains) :

RS.Urb ~ EcaNDVI065 + MoyNDVI065 + EcaNDVI177 + MoyNDVI177 + EcaNDVI241 + MoyNDVI241 + EcaNDVI257 + MoyNDVI257 + EcaNDVI273 + MoyNDVI273

Pour ce point également, la procédure de stepAIC est réalisée.

2.5.6. Classification fonctionnelle

Les variables explicatives issues de la classification fonctionnelle (section 2.2.5), sont également utilisées pour construire des modèles : il s'agit d'évaluer l'apport de ces variables intégrant le fonctionnement phénologique de la végétation sur la performance des modèles. Quelque soient le nombre de clusters, les variables obtenues sont toujours corrélées à moins de 30%.

Aussi, les modèles construits sont du format suivant (exemple du cas 6 clusters pour la variable réponse du CSI) :

CSI ~ Pourcent_A + Pourcent_B + Pourcent_C + Pourcent_D + Pourcent_E + Pourcent_F

2.5.7. Impact de la représentation spatiale

L'objectif de cette partie est de comparer la performance explicative de modèles construits à partir d'images classées (données CORINE Land Cover) ou brutes (variables précédentes issues des produits MODIS).

Les différents modèles que nous avons construits sont les suivants :

(a) Les modèles construits à partir de **données brutes** (cf. 2.2.4.) : de manière à rester uniquement sur des données non classées, la stratification n'est pas prise en compte ici. Ce sont donc les modèles du 2.5.2 qui sont utilisés. Pour chaque type de variable explicative et variable réponse, seul le modèle réalisé à la date présentant la meilleure performance est conservé.
   (a.1) A partir du NDVI
   (a.2) A partir de l'EVI
   (a.3) A partir du LSTD
   (a.4) A partir du LSTN

(b) Les modèles construits à partir de **données ayant déjà subi un premier classement** :
   (b.1) A partir des variables explicatives fonctionnelles (cf. 2.5.6.)
   (b.2) A partir du produit VCF de MODIS (Vegetation Continuous Field) : Les deux variables de VCF utilisées sont sa moyenne et son écart-type par carré STOC. Ces variables sont corrélées à 39%. Cette valeur, bien que élevée, reste néanmoins suffisante pour construire un modèle explicatif (Dormann et al., 2013; Shmueli, 2010). Ce modèle se présente de la

façon suivante (exemple pour la richesse spécifique forestière) : RS.Fores ~ MoyVCF + EcaVCF

(c) Les modèles construits à partir de **cartographie d'occupation du sol** (données classées) : à partir des données de CORINE Land Cover (cf. 2.3.) : La corrélation entre ces variables est inférieure à 20 % pour la plupart d'entre elles, sauf pour 6 couples de variables, sans dépasser 43 %. Comme pour les variables de VCF, cette valeur reste donc acceptable pour construire des modèles explicatifs (Dormann et al., 2013).

(c.1) A partir de l'information « présence de forêt » uniquement : Ces modèles ont le format suivant (cas de la richesse en espèces forestières) : RS.Fores ~ Part_31

(c.2) A partir de toutes les données de CORINE Land Cover : Toujours pour le cas de la richesse forestière, ces modèles sont du type : RS.Fores ~ Part_11 + Part_12 + Part_13 + Part_14 + Part_21 + Part_22 + Part_23 + Part_24 + Part_311 + Part_312 + Part_313 + Part_32 + Part_33 + Part_41 + Part_42 + Part_51 + Part_52

Tout d'abord, nous comparons les modèles construits à partir de l'information « présence de forêts ». Cette information est donnée à la fois par le VCF et par CORINE Land Cover. Aussi, pour rester homogène, seule l'information sur la présence de forêts est récupérée de la base de CORINE Land Cover : il s'agit ici de comparer les modèles (c.1) et (b.2).

Dans un deuxième temps, les autres modèles sont comparés entre eux, il s'agit des modèles présentant les meilleurs résultats : (a.1), (a.2), (a.3), (a.4), (b.1) et (c.2).

Pour l'ensemble de l'étude, nous avons téléchargé et pré-traité plus de 200 images satellites et construit plus de 2 000 modèles ! La production des modèles a été réalisée de manière automatique à l'aide de scripts développés dans R.

3. Résultats

Sur les carrés STOC étudiés, moyenne, le minimum et le maximum des variables réponses sont présentées dans le tableau 5.

Variable réponse	Moyenne sur les carrés STOC de l'étude (n = 867)	Moyenne sur *tous* les carrés STOC de 2010 (n = 1091)	Minimum sur les carrés STOC de l'étude (n = 867)	Maximum sur les carrés STOC de l'étude (n = 867)
RS.fores	8,6	8,4	0	19
RS.agri	5,1	4,9	0	14
RS.Urb	6,7	6,6	0	13
RS.Gene	11,4	11,0	1	14
RS.Tot	31,8	30,8	5	49
CTI	13,3	13,3	12,5	15,2
CSI	0,64	0,67	0,41	1,45

Tableau 5 : Moyenne, minimum et maximum des variables réponses sur les carrés STOC

Les tableaux 6 présentent les formes des courbes obtenues : + correspond à une courbe croissante, - à une courbe décroissante.

Variable réponse	Moyenne du VCF
RS.fores	+
RS.agri	-
RS.Urb	-
RS.Gene	+
RS.Tot	+
CTI	-
CSI	-+

Date	Variable réponse	Moyenne du NDVI	Moyenne de l'EVI	Moyenne du LSTD	Moyenne du LSTN
6 mars 2010	RS.Fores	+	+-	-	NA
	RS.Agri	NA	+	+	NA
	RS.Urb	NA	NA	NA	NA
	RS.Gene	+	+	NA	NA
	RS.Tot	+	+	NA	-
	CTI	NA	+	+	+
	CSI	-	NA	NA	NA
26 juin 2010	RS.Fores	+	+	-	NA
	RS.Agri	+-	NA	+	NA
	RS.Urb	-	-	+	NA
	RS.Gene	NA	NA	+	NA
	RS.Tot	+	+	+	NA
	CTI	-	-	+	+
	CSI	-	-	+	+
14 septembre 2010	RS.Fores	+	+	-	-
	RS.Agri	-	NA	+	NA
	RS.Urb	-	-+	+-	-+
	RS.Gene	NA	+	+	-
	RS.Tot	NA	+	+	-
	CTI	-	-	+	+
	CSI	-	-	+	NA

Tableaux 6 : Forme des courbes obtenues entre variables réponses et explicatives

Les modèles construits ont permis de réaliser des cartes prédictives de la répartition des différentes variables réponses. A titre d'exemple, le cas de la variable de richesse spécifique en espèces forestières est présenté en figure 19 pour deux modèles réalisés.

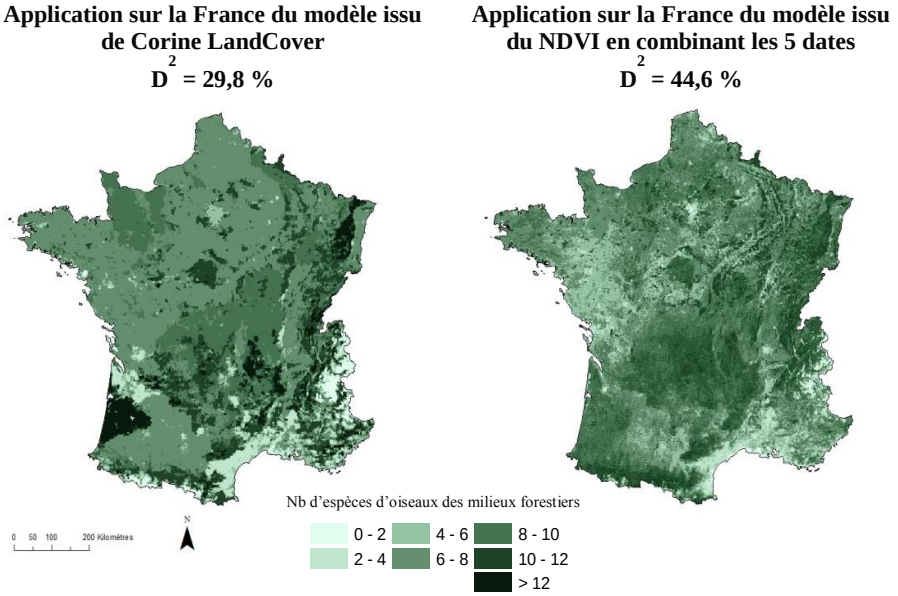


Figure 19 : Cartes prédictives de la répartition de la richesse en espèces forestières

3.1. Impact de la date sur la performance des modèles

Nous présentons ci-dessous (figures 20 et 21) la performance des modèles explicatifs réalisés à chacune des dates en utilisant le couple de variables *Moyenne* et *Ecart-type* pour chaque type de variable : *NDVI*, *EVI*, *LSTD*, *LSTN*. Les modèles ont été calculés sur chaque variable réponse (140 modèles au total). Ils sont comparés en se basant sur la valeur de l'AIC (voir graphiques en annexe 2).

Les dates des modèles les plus performants varient en fonction du type de variable explicative et de la variable réponse. On constate ainsi que les modèles de RS.Fores et du CTI sont très bien ajustés (pour ce type de modèle), leur pourcentage de déviance expliquée allant respectivement jusqu'à 39,7 % et 47,1 %. RS.Agri, RS.Urb et le CSI sont moyennement bien expliqués par les modèles : leur pourcentage de déviance expliquée va respectivement jusqu'à 11,4 %, 12,7 % et 20,1 %. Globalement, RS.Gene et RS.Tot donnent des modèles aux performances nettement plus faibles, avec une déviance expliquée atteignant au maximum respectivement 8,2 % et 9,6 %.

Pour RS.Fores, le CTI et le CSI, les modèles construits à partir du NDVI sont plus performants que ceux construits à partir de l'EVI. Les résultats pour les autres variables réponses dépendent des dates, mais on peut tout de même noter que les modèles construits à partir de l'EVI donnent la majorité du temps de meilleures performances. Le LSTD, quant à lui, donne de meilleurs modèles que le LSTN pour la majorité des variables réponses, à l'exception de RS.Gene et RS.Tot qui sont mieux expliquées par les modèles issus du LSTN. Généralement, le LSTD et le LSTN donne des modèles moins performants que ceux issus du NDVI ou de l'EVI. Néanmoins, les modèles issus du LSTD sont parfois meilleurs que ceux issus de l'EVI, et on peut également remarquer qu'ils expliquent particulièrement bien le CTI. Cependant, RS.Gene et RS.Tot ne sont expliqués qu'à moins de 1 %.

Pour les variables réponses RS.Fores, RS.Urb et CTI, les modèles issus de l'EVI et du NDVI donnent de meilleures performances avec la synthèse du 26 juin (la synthèse du 6 mars étant particulièrement mauvaise), et ceux issus du LSTD et du LSTN sont généralement meilleurs à la synthèse du 29 août. RS.Agri est moins bien expliqué par les modèles du 6 mars qu'aux autres dates pour les modèles construits à partir du NDVI et du LSTD. Pour RS.Gene et RS.Tot, les meilleurs modèles sont obtenus à la date du 6 mars, pour des variables explicatives issues respectivement de l'EVI et du LSTD. Le CSI donne systématiquement des modèles peu performants à la date du 6 mars, la performance aux autres dates dépendant des types de variables explicatives.

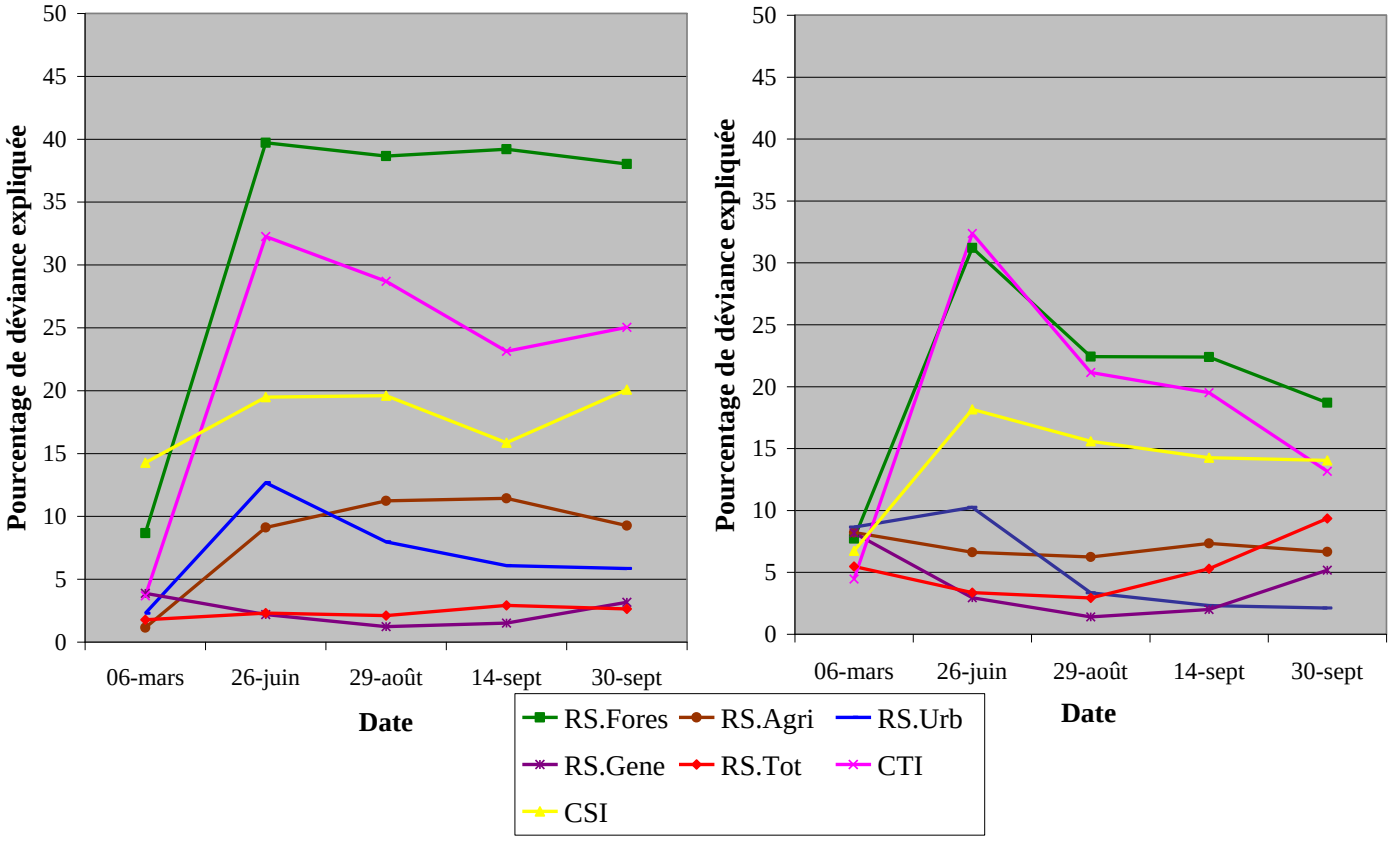


Figure 20 : Pourcentages de déviance expliquée des modèles des différentes variables réponses à partir du NDVI (à gauche) et de l'EVI (à droite) aux différentes dates

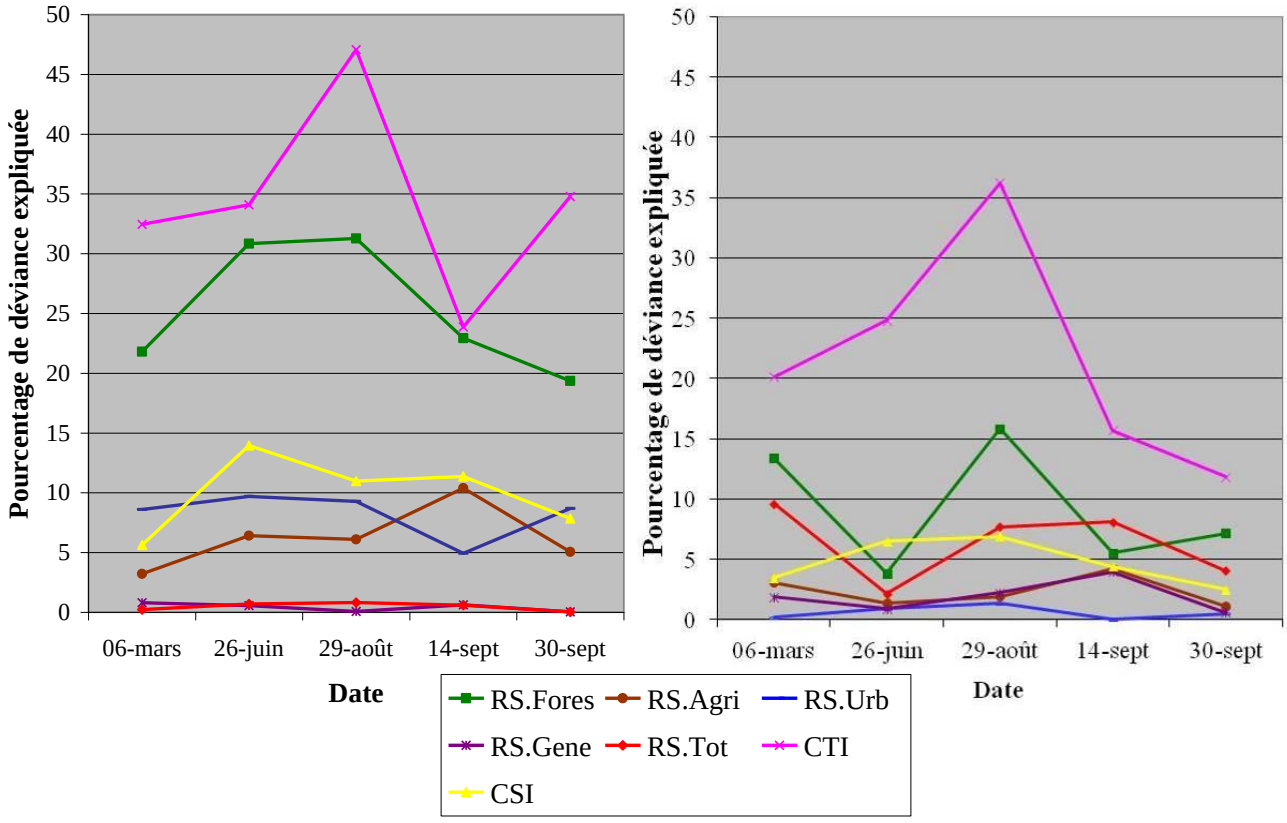


Figure 21 : Pourcentages de déviance expliquée des modèles des différentes variables réponses à partir du LSTD (à gauche) et du LSTN (à droite) aux différentes dates

A partir de ces résultats, le tableau 7 présente les meilleures dates sélectionnées par variable réponse et par type de variable explicative :

Var réponse \ Var explicative	NDVI	EVI	LSTD	LSTN
RS.Fores	**177**	177	241	241
RS.Agri	**257**	065	257	257
RS.Urb	177	177	177	**241**
RS.Gene	**065**	065	065	257
RS.Tot	257	273	241	**065**
CTI	177	177	**241**	241
CSI	**273**	177	177	241

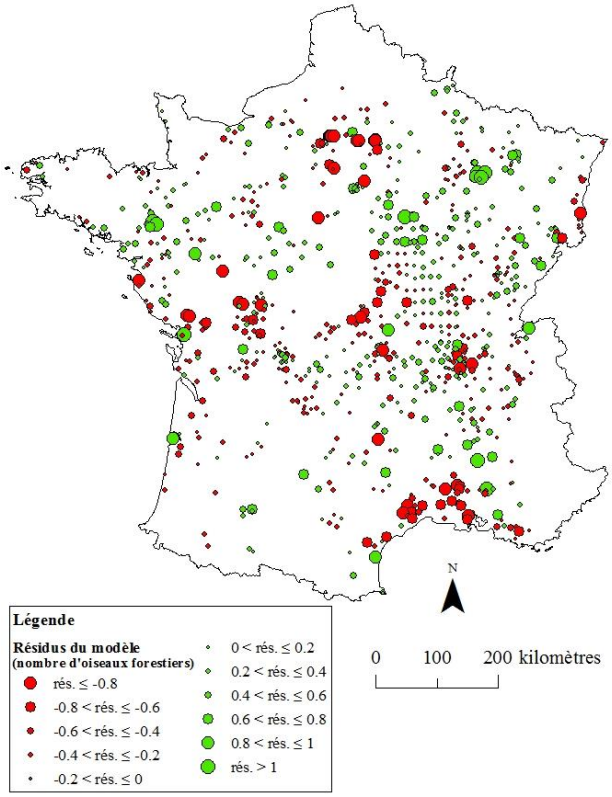
065 : début mars (en bleu) - 177 : fin juin (en vert) - 241, 257 et 273 : de fin août à fin septembre (en orange)
Les dates en gras correspondent aux synthèses donnant le meilleur modèle par variable réponse.

Tableau 7 : Dates correspondants aux meilleurs modèles pour chaque type de variable explicative, et par variable réponse

Les périodes de fin août à fin septembre donnent les meilleurs modèles pour la majorité des variables, à l'exception de RS.Gene et RS.Tot qui ont de meilleurs résultats sur la synthèse du 6 mars, et de RS.Fores qui est mieux expliqué sur la synthèse de fin juin.

3.2. Prise en compte de l'effet région

La représentation spatiale des résidus relatifs aux modèles explicatifs établis dans la section précédente montre que la distribution de ces résidus n'est pas aléatoire. Nous donnons un exemple en figure 22. Les résidus ne sont donc pas indépendants et l'hypothèse d'homoscédasticité des modèles linéaires généralisés n'est pas vérifiée. Ceci nous a conduit à introduire l'effet « région » dans les modèles afin de s'abstraire de l'autocorrélation spatiale qui augmente le risque d'erreur de type I et rend ainsi artificiellement les modèles plus performants (Dormann et al., 2007).



La grosseur des cercles est proportionnelle à la valeur absolue du résidu. Les cercles verts correspondent à des résidus positifs, et les cercles rouges à des résidus négatifs.

Les résidus ne sont pas répartis aléatoirement : ils sont globalement négatifs en région méditerranéenne, et positifs en Bretagne.

Figure 22 : Répartition des résidus du modèle : RS.Fores ~ MoyNDVI177 + EcaNDVI177

Les modèles prédictifs précédents sont repris en ajoutant un effet région simple, et les pourcentages de déviance expliquée obtenus sont présentés dans le tableau 8. L'annexe 3 donne les valeurs des AIC des meilleurs modèles.

Variables réponses	Variables explicatives	Pas de régionalisation	Régions agricoles	Régions climatiques	Sylvo-écorégions
RS.Fores	NDVI-177	39,72	42,92	**44,20**	42,71
	EVI-177	31,21	35,75	**37,22**	36,31
	LSTD-241	31,27	37,18	36,59	**38,07**
	LSTN-241	15,85	21,18	**26,38**	23,80
RS.Agri	NDVI-257	11,43	**24,66**	16,35	18,41
	EVI-065	8,22	**17,23**	11,94	15,44
	LSTD-257	10,39	**19,58**	15,28	12,01
	LSTN-257	4,26	10,71	9,53	**16,65**
RS.Urb	NDVI-177	12,67	**24,21**	15,07	18,96
	EVI-177	10,25	**21,3**	12,5	15,37
	LSTD-177	9,69	**19,91**	11,65	15,06
	LSTN-241	1,34	**6,73**	3,22	4,22
RS.Gene	NDVI-065	3,87	**14,13**	12,71	13,4
	EVI-065	8,19	15,61	14,62	**15,82**
	LSTD-065	0,8	**13,28**	11,04	11,97
	LSTN-257	3,97	**13,34**	10,91	11,29
RS.Tot	NDVI-257	2,91	**17,08**	9,94	13,07
	EVI-273	9,35	**18,39**	14,17	17,31
	LSTD-241	0,82	**14,86**	10,43	11,96
	LSTN-065	9,61	**16,67**	13,1	14,09
CTI	NDVI-177	32,25	52,06	53,42	**52,96**
	EVI-177	32,37	50,00	51,28	**51,47**
	LSTD-241	47,06	55,64	58,7	**59,11**
	LSTN-241	36,26	47,33	**48,63**	48,16
CSI	NDVI-273	20,08	**27,77**	26,77	26,12
	EVI-177	18,17	**26,92**	24,72	26,52
	LSTD-177	13,95	**24,95**	19,84	22,84
	LSTN-241	6,9	**15,41**	12,57	13,76

Les chiffres en vert correspondent à la valeur la plus élevée de la ligne.

Tableau 8 : Pourcentage de déviance expliquée des modèles avec régionalisation

Les modèles sont très fortement améliorés par l'ajout de la stratification, quel que soit le type de stratification considéré : le pourcentage de déviance expliquée des modèles de RS.Agri augmente de 13,23 % et RS.Urb de 11,54 % avec la variable NDVI. A l'exception de RS.Urb avec la variable LSTN, tous les pourcentages sont supérieurs à 10 % avec les régions agricoles et les sylvo-écorégions, tandis qu'en l'absence de stratification, 13 des 28 modèles présentés ont un pourcentage de déviance expliquée inférieur à 10, soit presque la moitié des modèles. Les modèles explicatifs construits à partir de LSTN présentent une amélioration particulièrement forte (+ 9,37 % pour la variable réponse RS.Gene). L'augmentation est un peu moins marquée pour RS.Fores qui donnait déjà des performances élevées (+ 4,5 % avec la variable NDVI).

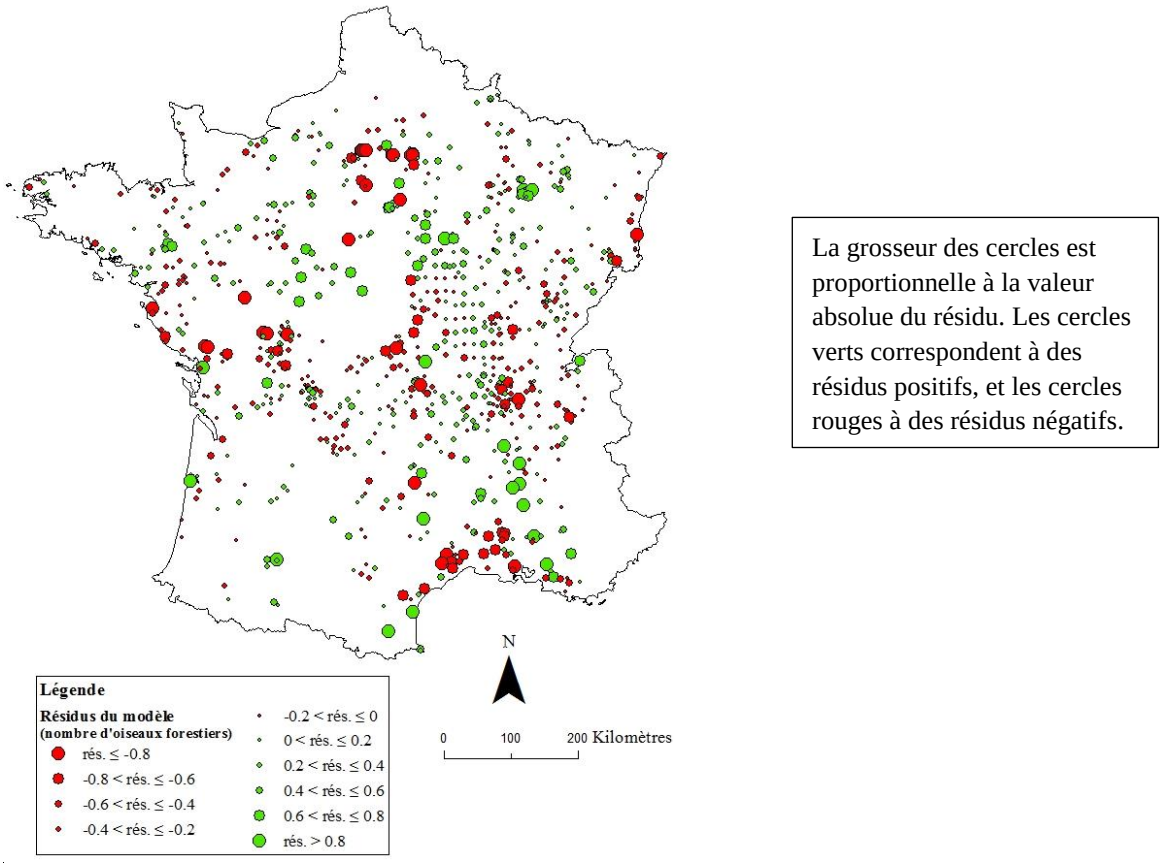
Dans la très majorité des cas, les régions agricoles donnent le modèle avec la plus grande déviance expliquée. On peut toutefois noter deux exceptions : pour RS.Fores, ce sont les régions climatiques qui améliorent le plus les modèles, et ce sont les sylvo-écorégions pour le CTI.

Comme expliqué dans la démarche méthodologique, nous avons ensuite complexifié les modèles précédents en ajoutant les interactions entre les variables explicatives et les différents types de

stratification. L'hypothèse sous-jacente est que les variables explicatives peuvent avoir un effet différent suivant la strate dans laquelle on se trouve.

La sélection des variables par la fonction stepAIC permet de constater que dans la majorité des cas, les interactions ne sont pas conservées dans le meilleur modèle. Ainsi, les meilleurs modèles sont généralement les modèles avec un effet région simple, sans les interactions.

A la suite de cette prise en compte des effets régions, nous pouvons de nouveau représenter la répartition spatiale des résidus (exemple de la figure 23).



La grosseur des cercles est proportionnelle à la valeur absolue du résidu. Les cercles verts correspondent à des résidus positifs, et les cercles rouges à des résidus négatifs.

Figure 23 : Répartition des résidus du modèle : RS.Fores ~ EcaNDVI177 + MoyNDVI177 + reg_sylv + reg_sylv:EcaNDVI177 + reg_sylv:MoyNDVI177 (notation de R)

On peut constater que l'autocorrélation spatiale, bien que moins marquée, reste présente. Des modèles spatiaux plus complexes devraient être mis en œuvre pour s'en affranchir (Bacaro et al., 2011; Dormann et al., 2007). La prise en compte plus poussée de cette autocorrélation spatiale sort du cadre de ce stage.

Nous avons aussi cherché à évaluer la performance prédictive des modèles avec un effet région simple, sans interaction. Les résultats obtenus sont présentés dans le tableau 9.

Variable réponse	Type de stratification	Données utilisées dans le modèle	Rho moyen	RMSE moyen
RS.Fores	Régions climatiques	NDVI-177	**0,687**	**2,86**
		EVI-177	0,618	3,08
		LSTD-241	0,614	3,10
		LSTN-241	0,487	3,41
RS.Agri	Régions agricoles	NDVI-257	**0,503**	**2,71**
		EVI-065	0,433	2,83
		LSTD-257	0,455	2,78
		LSTN-257	0,303	2,91
RS.Urb	Régions agricoles	NDVI-177	**0,479**	**2,85**
		EVI-177	0,414	2,86
		LSTD-177	0,461	**2,85**
		LSTN-241	0,247	3,09
RS.Gene	Régions agricoles	NDVI-065	0,323	1,91
		EVI-065	**0,392**	**1,88**
		LSTD-065	0,349	1,90
		LSTN-257	0,318	1,90
RS.Tot	Régions agricoles	NDVI-257	0,429	6,40
		EVI-273	**0,436**	**6,32**
		LSTD-241	0,356	6,49
		LSTN-065	0,379	6,42
CTI	Sylvo-écorégions	NDVI-177	0,654	0,26
		EVI-177	0,614	0,27
		LSTD-241	**0,711**	**0,24**
		LSTN-241	0,517	0,28
CSI	Régions agricoles	NDVI-273	0,480	**0,119**
		EVI-177	**0,502**	0,120
		LSTD-177	0,481	0,120
		LSTN-241	0,317	0,128

Les valeurs en gras correspondent aux meilleures valeurs par variable réponse

Tableau 9 : Moyenne de la corrélation de Spearman (rho) et de l'erreur quadratique moyenne (RMSE) des meilleurs modèles avec régions

La performance prédictive des modèles avec un effet région simple est très bonne pour toutes les variables réponses, y compris celles qui donnaient de moins bonnes performances explicatives, comme RS.Gene et RS.Tot. Les variables de type NDVI et EVI donnent de meilleurs résultats que le LSTD et le LSTN, à l'exception du CTI dont la corrélation de Spearman dépasse les 70 % avec le LSTD.

3.3. Complémentarité entre types de variables pour une même date, et complémentarité entre dates pour un même type de variables

Les modèles précédents sont améliorés en cumulant :

- d'une part l'information d'une même variable contenue sur plusieurs dates (cf. 2.5.5)
- d'autre part l'information contenue par différentes variables (EVI-LSTD-LSTN et NDVI-LSTD-LSTN) à une date donnée (cf. 2.5.4)

Les graphiques 24 à 26 présentent par variable réponse les pourcentages de déviance expliquée des modèles obtenus pour chacune des cinq dates prises indépendamment et pour les 5 dates ensemble,

ainsi que pour chacun des types de variables et pour la combinaison des variables ensemble. Les écarts d'AIC entre les modèles sont donnés à l'annexe 4.

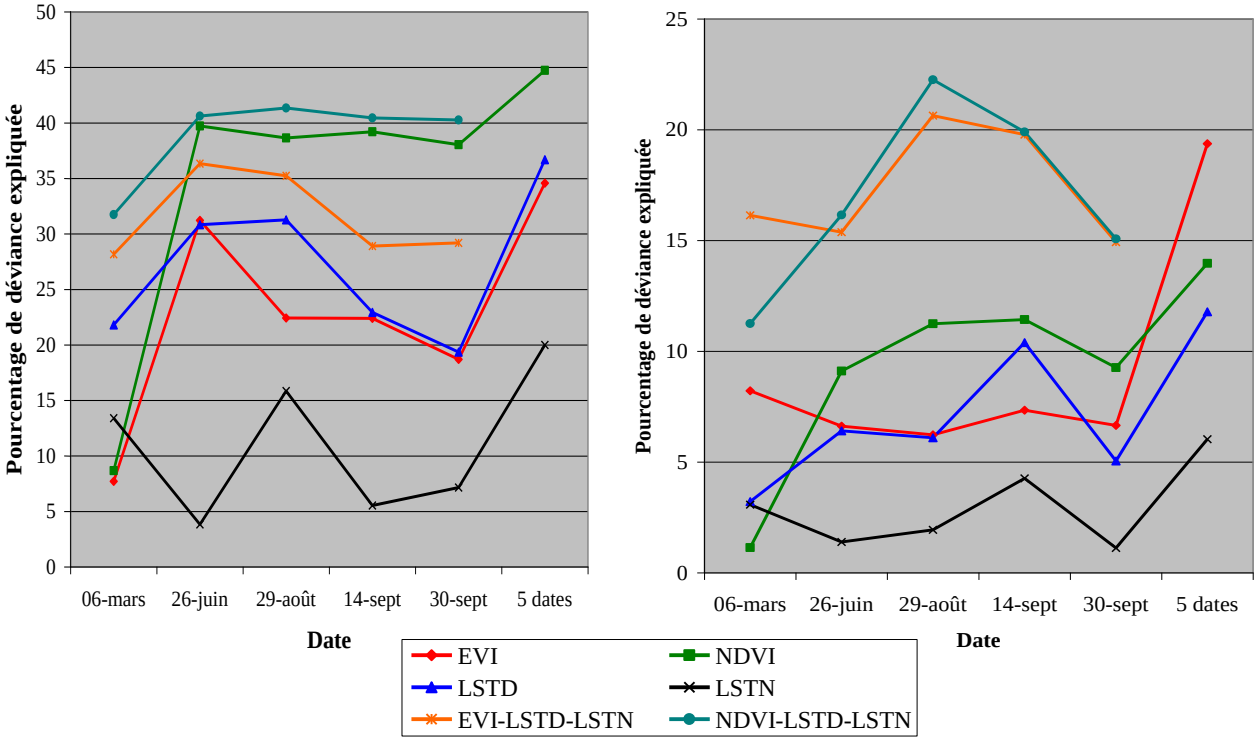


Figure 24 : D^2 des modèles de RS.Fores (à gauche) et RS.Agri (à droite)

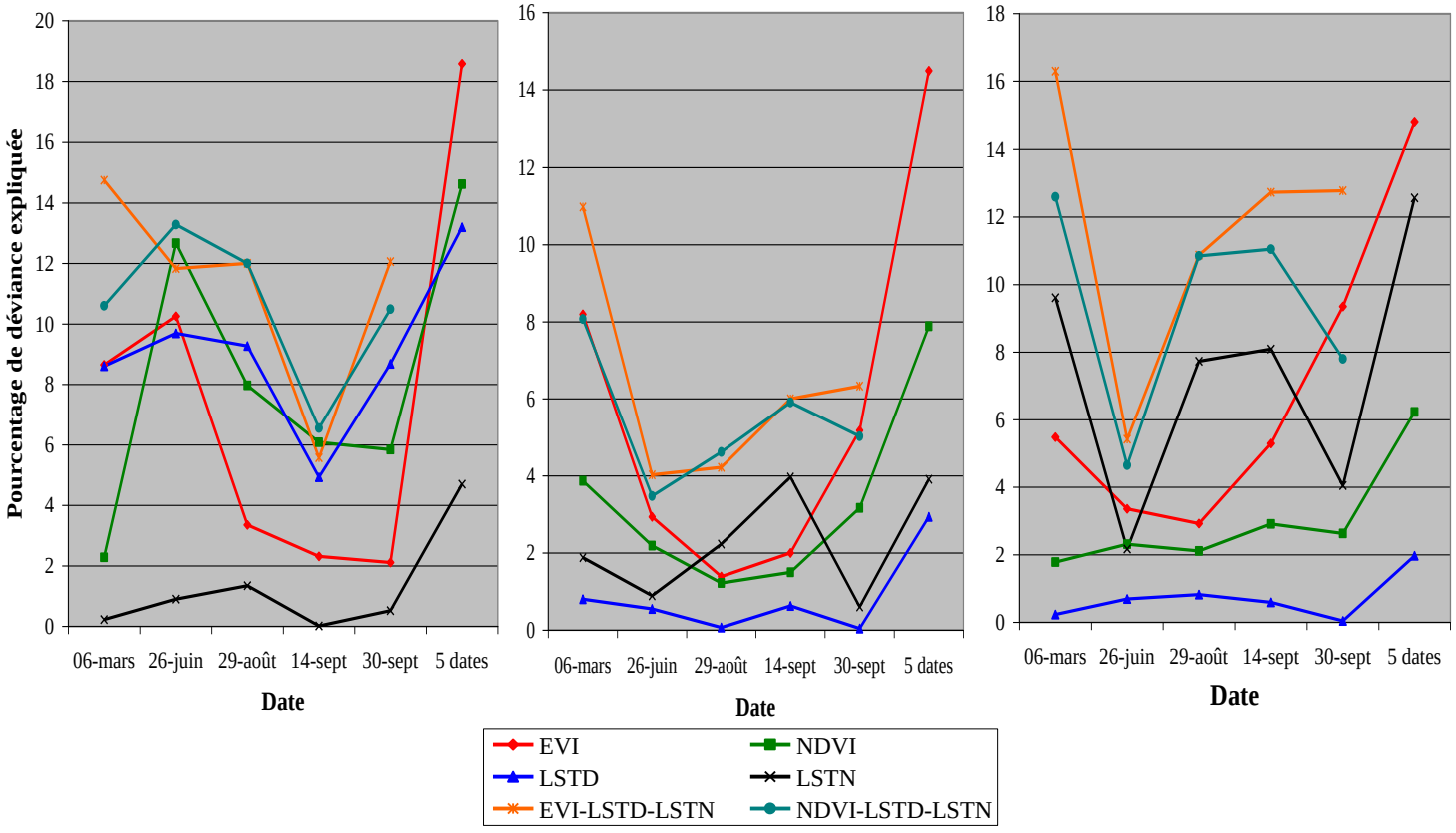


Figure 25 : D^2 modèles de RS.Urb (à gauche), RS.Gene (au centre) et RS.Tot (à droite)

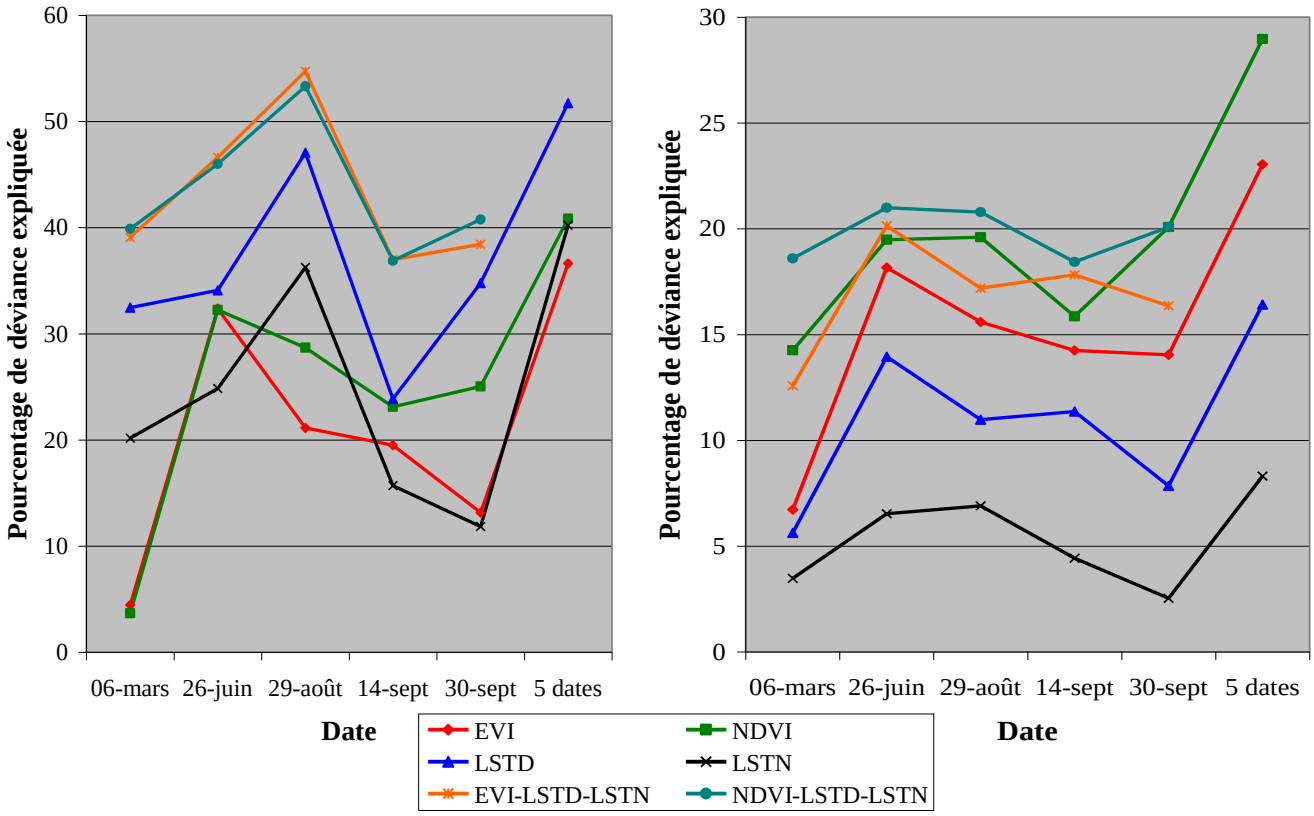


Figure 26 : D^2 des modèles du CTI (à gauche) et du CSI (à droite)

On constate que pour RS.Fores, RS.Urb, RS.Gene et le CSI, c'est un des modèles cumulant les dates qui donne le meilleur résultat, soit à partir de l'EVI, soit du NDVI. Pour les autres variables réponses (RS.Agri, RS.Tot et le CTI), le meilleur modèle est un de ceux cumulant plusieurs variables explicatives à une date donnée, respectivement pour les synthèses du 29 août, du 6 mars et du 29 août. A l'exception des modèles construits à partir de LSTN, les modèles cumulant les 5 dates donnent de meilleures performances que ceux réalisés pour une date donnée, quelle qu'elle soit.

En ce qui concerne les modèles issus de la combinaison de plusieurs variables à une date, les résultats diffèrent selon les variables réponses :

- Pour RS.Fores, RS.Urb, et le CSI, les modèles explicatifs issus de variables « NDVI-LSTD-LSTN » donnent de meilleures performances que ceux issus d'un seul type de variable. On peut cependant noter une exception pour le modèle issu du NDVI et la variable réponse CSI dont la performance n'est pas significativement différente ($\Delta AIC \leq 2$) du modèle issu de « NDVI-LSTD-LSTN » pour la synthèse du 30 septembre.
- Pour RS.Gene et RS.Tot, les modèles issus des variables « EVI-LSTD-LSTN » sont significativement meilleurs que ceux issus d'un seul type de variable ($\Delta AIC > 2$).
- Pour RS.agri et le CTI, les modèles issus des variables « NDVI-LSTD-LSTN » et ceux issus des variables « EVI-LSTD-LSTN » donnent tous les deux des performances plus élevées que ceux issus de variables simples ($\Delta AIC > 2$).

3.4. Classification fonctionnelle

Dans cette partie, nous présentons les résultats obtenus pour les modèles construits à partir des classes de fonctionnement phénologiques de la végétation issues des séries temporelles NDVI. Pour rappel, les modèles construits à partir de ce jeu de données intègrent comme variables explicatives les pourcentages de chaque classe obtenue, contenue dans les carrés STOC.

La figure 27 et le tableau 10 présentent l'évolution de la déviance expliquée et la valeur d'AIC selon le nombre de classes (clusters) fixés dans l'algorithme de classification utilisé, pour chaque variable réponse.

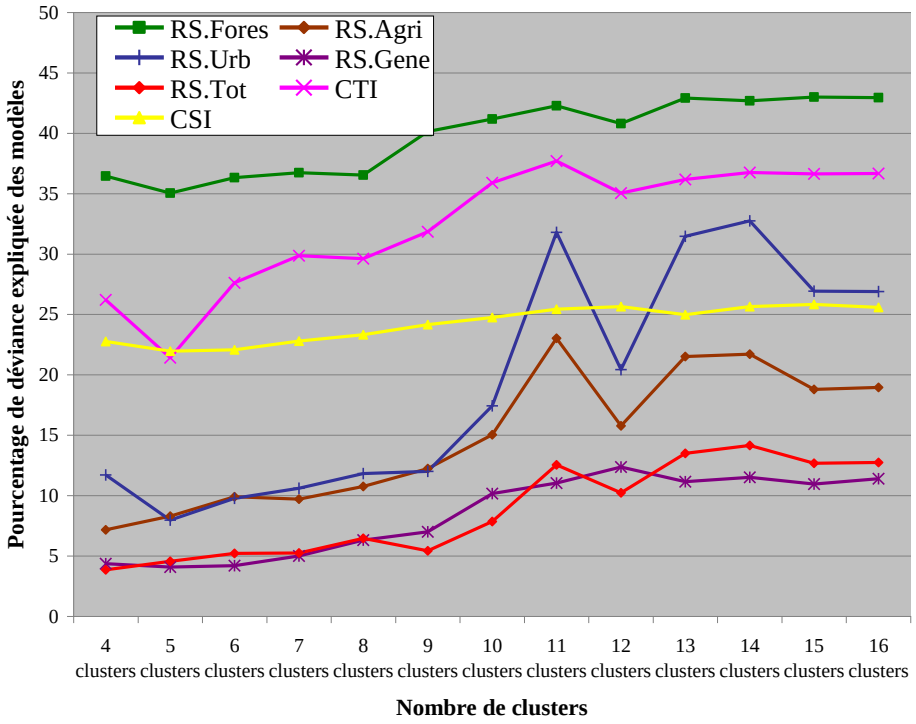


Figure 27 : Pourcentage de déviance expliquée et différence entre les AIC des modèles construits à partir d'un nombre différents de clusters pour chaque variable réponse

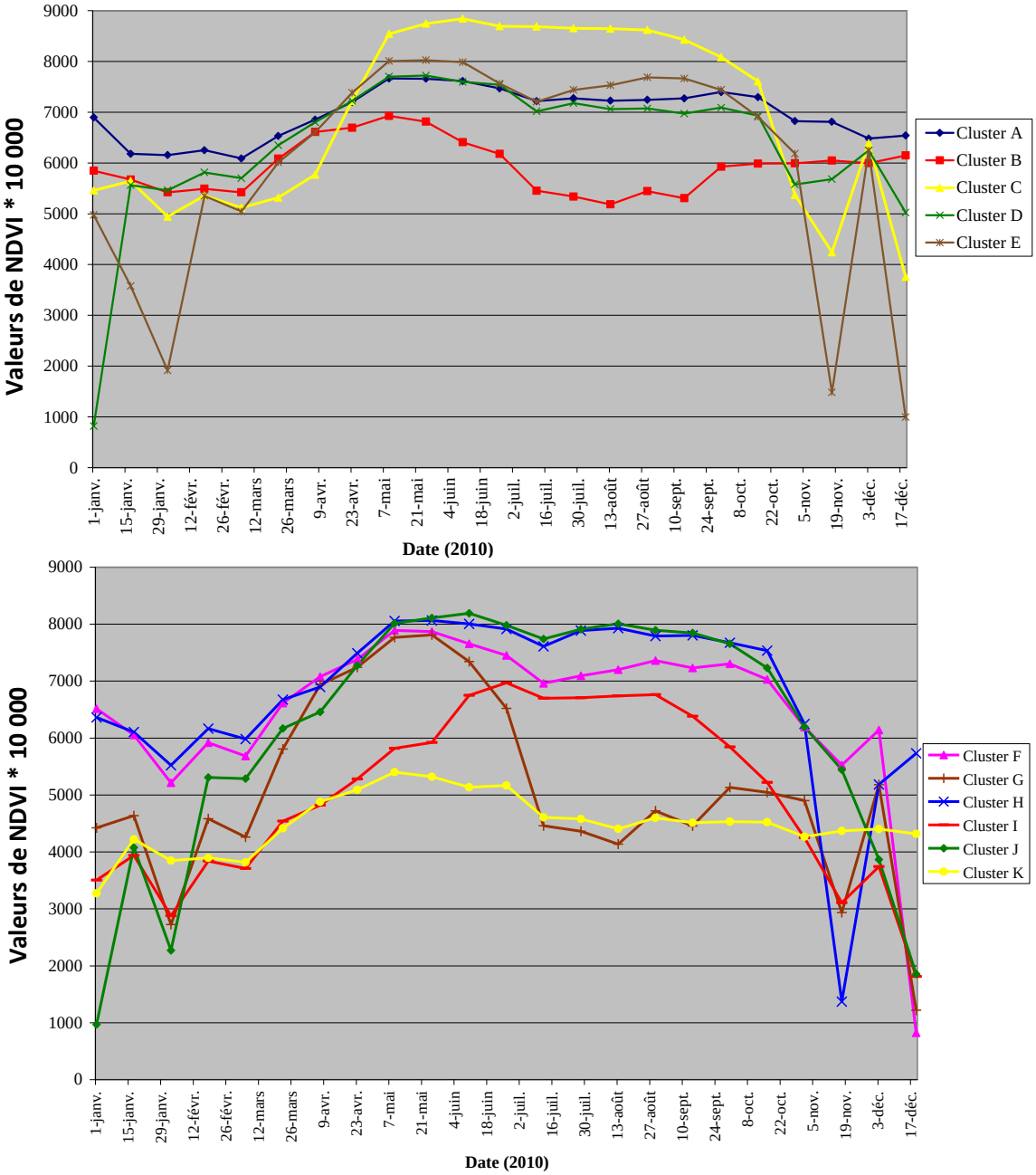
	4 clusters	5 clusters	6 clusters	7 clusters	8 clusters	9 clusters	10 clusters	11 clusters	12 clusters	13 clusters	14 clusters	15 clusters	16 clusters
RS.Fores	4477	4504	4483	4478	4484	4421	4404	4386	4415	**4379**	4385	4381	4384
RS.Agri	4521	4501	4473	4478	4460	4434	4383	**4233**	4373	4266	4264	4322	4321
RS.Urb	4524	4588	4560	4548	4530	4529	4442	4207	4396	4216	**4197**	4295	4297
RS.Gene	4058	4061	4063	4062	4059	4058	4048	4047	**4044**	4051	4051	4055	4056
RS.Tot	5939	5931	5924	5925	5910	5927	5894	5830	5865	5821	**5814**	5837	5837
CTI	519	575	506	481	486	460	409	**386**	424	411	405	408	410
CSI	-1203	-1192	-1191	-1197	-1201	-1209	-1213	**-1220**	**-1220**	-1210	-1216	-1216	-1211

Les chiffres en vert correspondent aux modèles présentant le plus petit AIC par variable réponse.

Tableau 10 : AIC des modèles construits à partir d'un nombre différent de clusters

On observe que les modèles basés sur une classification à 11, 13 ou 14 clusters présentent peu de différences. Pour faciliter l'interprétation des classes (la classification étant non supervisée), c'est le cas à 11 clusters qui est retenu dans la suite de l'étude.

Il s'agit maintenant d'identifier la signification de chacun des 11 clusters obtenus. Pour cela, nous avons d'abord tracé les courbes moyennes de NDVI de chaque cluster (figures 28). Nous avons ensuite interprété ces courbes en les confrontant aux différentes catégories d'occupation du sol de CORINE Land Cover par superposition des pixels classés sur la carte (figure 29). Bien que le choix d'utiliser CORINE ne soit pas le plus judicieux, nous avons opté pour celui-ci dans un souci de rapidité d'exécution.



Figures 28 : Evolution moyenne du NDVI pour chacun des 11 clusters

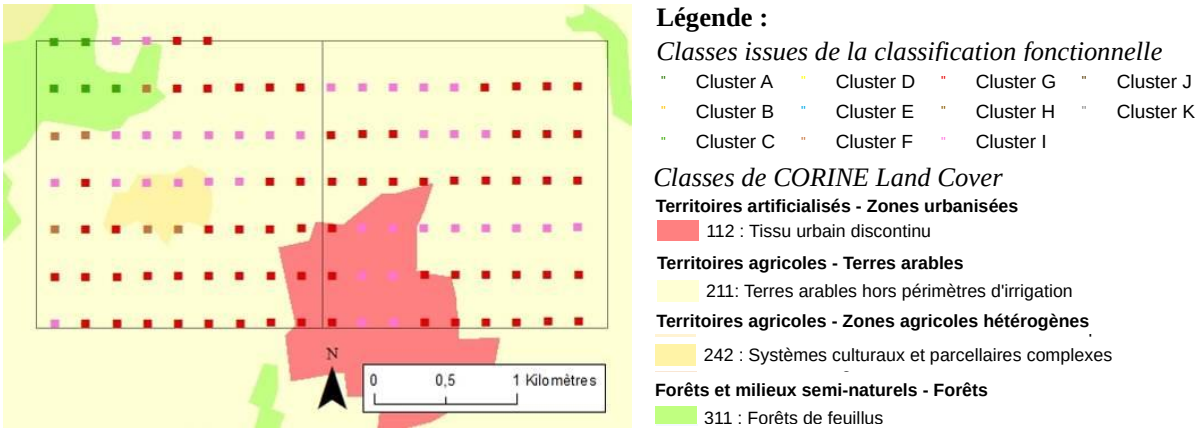


Figure 29 : Exemple de classification des pixels des carrés STOC n°77-0007 et n°77-0008, en lien avec la classification de CORINE Land Cover

Ces analyses permettent d'interpréter de manière assez précise certains clusters :

- Cluster A : Forêts de conifères, éventuellement forêts mixtes (le NDVI reste élevé en permanence, même en hiver) ;
- Cluster B : Prairies et zones de pâturage (le NDVI montre un maximum fin mai/début juin avant une diminution correspondant à la phase de sénescence ou à la fauche du début de l'été) ;
- Cluster C : Forêts de feuillus (le NDVI augmente au printemps, reste stable ensuite pour diminuer en automne, coïncidant avec la chute des feuilles) ;
- Cluster G : Cultures de printemps sur terre arable (le NDVI augmente fortement au printemps mais diminue ensuite brutalement en juin/juillet, coïncidant avec la période de récolte) ;
- Cluster I : Culture d'été sur terre arable (le NDVI a une allure proche de celle du cluster G mais atteint son maximum plus tardivement, au mois d'août, avant de dominer) ;
- Cluster K : milieu fortement urbanisé avec peu de végétation (le NDVI est plus faible et relativement constant). La correspondance avec CORINE montre également que ce cluster représente les zones de vignobles de la zone Méditerranéenne (l'effet sol expliquant les valeurs plus faibles de NDVI).

Pour les autres clusters, l'interprétation est plus délicate. En effet, la comparaison avec CORINE montre qu'il n'y a pas de correspondance unique avec une classe d'occupation du sol dominante. Les correspondances sont plurielles. Elles peuvent s'expliquer par le faible niveau de détail de la base CORINE (unité minimale de 25 ha). Elles peuvent aussi indiquer un fonctionnement phénologique similaire pour des catégories d'occupation du sol différentes. Ainsi, on observe :

- Cluster D : une correspondance majoritaire avec des prairies incluant quelques relations avec des zones forestières.
- Cluster E : une correspondance majoritaire avec des prairies également, ainsi qu'avec des terres arables et quelques zones forestières, généralement plus en altitude.
- Clusters F, H et J : une correspondance avec des classes très variées mêlant forêts, prairies et cultures.

A titre illustratif, nous donnons un aperçu de la répartition des différents clusters à l'échelle de la France pour chaque carré STOC étudié (figure 30).

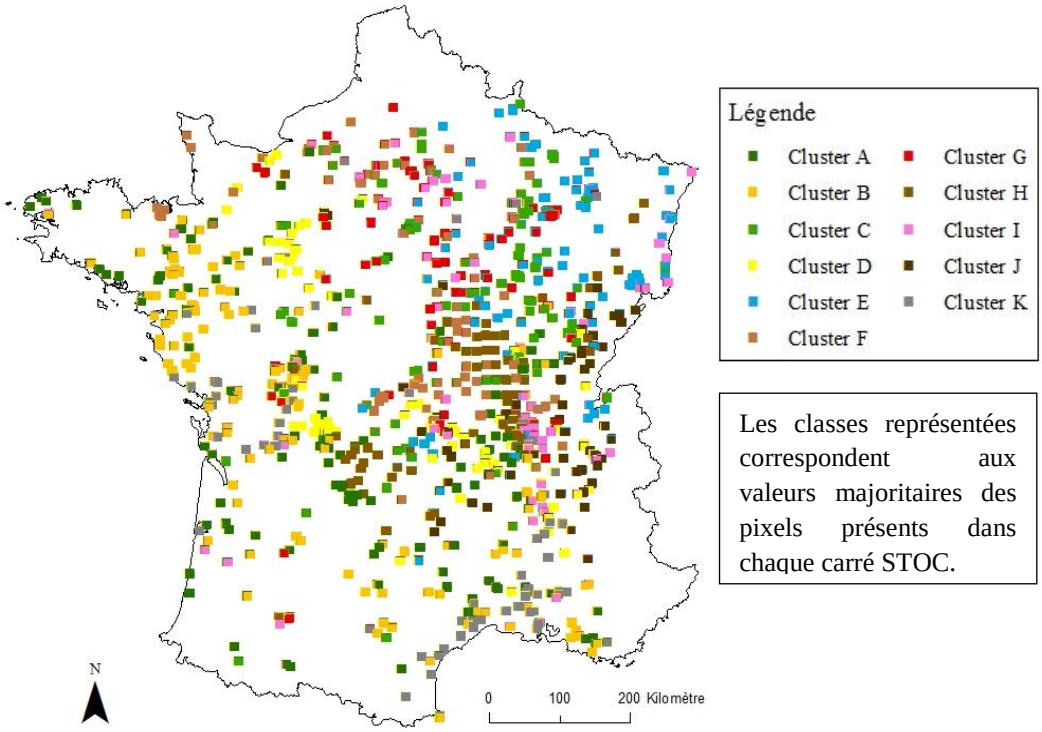


Figure 30 : Répartition des clusters des pixels des carrés STOC

3.5. Impact de la représentation spatiale

Dans un premier temps, nous nous focalisons sur la comparaison entre les modèles « forêts », c'est-à-dire les modèles construits à partir du VCF (b.2) et à partir de l'information « présence de forêt » de CORINE Land Cover (c.1). Pour chaque variable réponse, les modèles sont significativement différents : l'écart d'AIC est supérieur à 2. La figure 31 présente les pourcentages de déviance expliquée obtenus pour ces modèles.

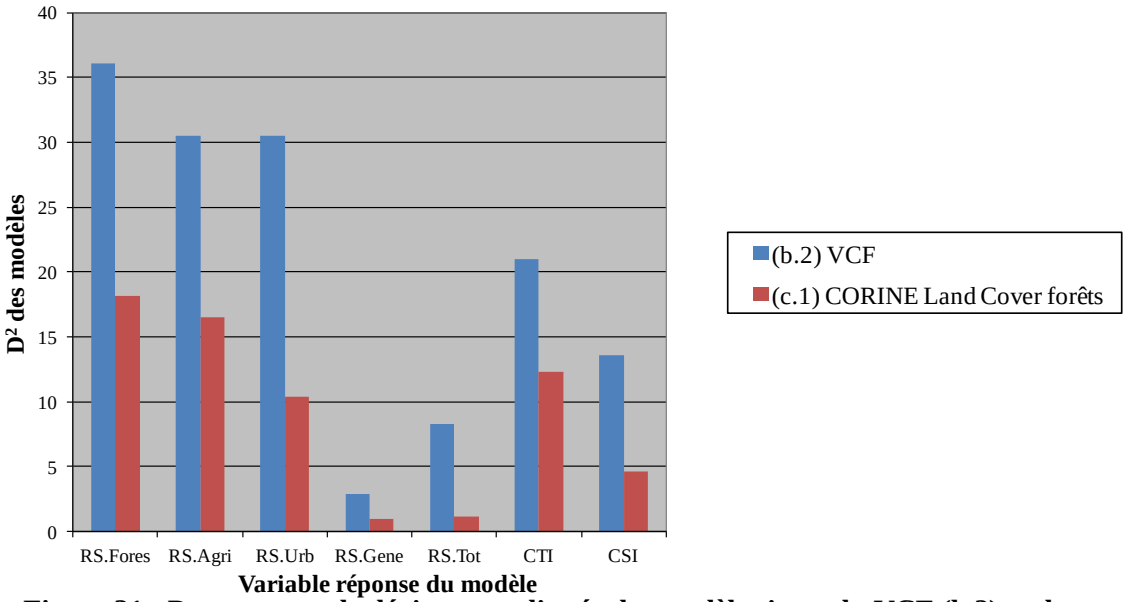


Figure 31 : Pourcentage de déviance expliquée des modèles issus du VCF (b.2) et de l'information « présence de forêt » de CORINE Land Cover (c.1)

On constate que quel que soit la variable réponse étudiée, le modèle issu du VCF a une performance explicative plus élevée que celle du modèle issu de CORINE Land Cover.

La figure 32 présente le pourcentage de déviance expliquée des autres modèles (comparaison données brutes versus classées). Pour une même variable réponse, les modèles sont statistiquement différents entre eux ($\Delta AIC > 2$), sauf les quatre couples de modèles suivants (cf. annexe 5) :

- Pour RS.Fores et CSI : les modèles (a.2) et (a.3),
- Pour RS.Gene : les modèles (a.1) et (a.4),
- Pour CTI : les modèles (a.4) et (b.1).

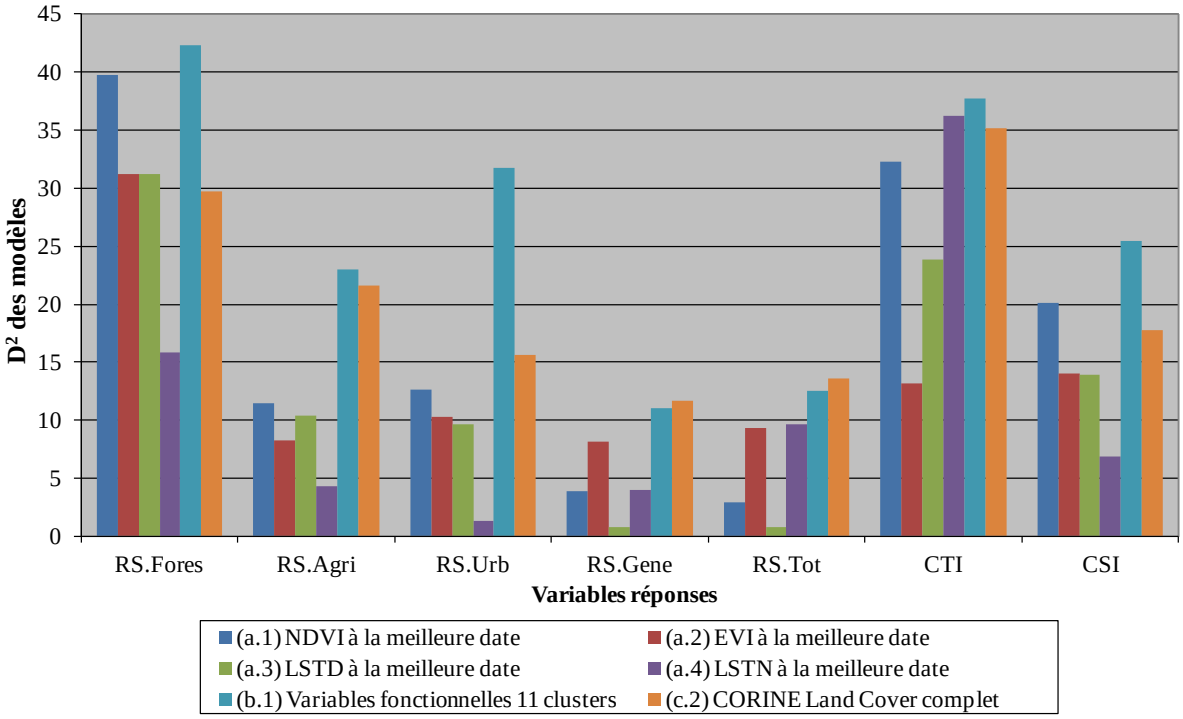


Figure 32 : Pourcentage de déviance expliquée des modèles issus de différentes représentations spatiales pour chaque variable réponse

Le modèle (b.1) issu des variables explicatives fonctionnelles donne presque systématiquement le meilleur pourcentage de déviance expliquée, à l'exception de RS.Gene et RS.Tot pour lesquels c'est CORINE Land Cover qui donne le meilleur modèle.

Parmi les variables issues des images brutes, ce sont les variables issues du NDVI qui donnent le modèle explicatif le plus performant, à l'exception des deux mêmes variables réponse RS.Gene et RS.Tot pour lesquelles ce sont respectivement les modèles issus de l'EVI et du LSTN qui donnent de meilleurs résultats.

Les modèles issus du LSTD donnent des performances du même ordre de grandeur que les modèles issus de l'EVI pour les variables réponses RS.Fores, RS.agri, RS.Urb et le CSI, et sont meilleures que pour les modèles issus du LSTN. En revanche, ce n'est pas le cas pour les variables RS.Gene, RS.Tot et CTI pour lesquelles les modèles issus du LSTN donnent de meilleures performances que ceux issus du LSTD.

Pour RS.Fores, le CTI et le CSI, les modèles issus de CORINE Land Cover donnent de moins bonnes performances que le meilleur des modèles issus des données continues (le meilleur modèle entre (a.1), (a.2), (a.3) et (a.4)). En revanche, CORINE Land Cover donne de meilleurs résultats pour les variables RS.Agri, RS.Urb, RS.Gene et RS.Tot.

4. Discussion

4.1. Analyse des résultats obtenus

Les résultats obtenus dans cette étude sont assez riches et variés, l'analyse se focalise sur les objectifs présentés en section 1.3.

D'une manière générale, on observe d'assez bonnes performances explicatives pour des modèles intégrant des données écologiques relatives à la présence-absence d'espèces. Globalement, les modèles pour les groupes d'espèces spécialistes sont plus performants que pour les groupes d'espèces généralistes. Une explication possible est qu'il est plus facile d'expliquer la niche des espèces très spécialisées à un habitat : lorsque l'habitat est présent, l'espèce est généralement présente. Les espèces généralistes au contraire peuvent se retrouver de manière plus aléatoire sur une grande variété d'habitats. Ce résultat est concordant avec les recherches de Cord et Rödder (2011).

Quel est l'impact de la date sur la performance des modèles ?

Les résultats présentés en section 3.3 montrent que la date influe effectivement sur les performances explicatives des modèles.

Pour le LSTD et le LSTN, c'est généralement la période de fin août qui donne les meilleurs modèles.

Cependant, les modèles réalisés pour la période de début septembre donnent des résultats très différents, et on observe souvent une forte baisse de la performance des modèles. Le LST semble donc variable en fonction de la météorologie, et en ce sens un indicateur moyennement robuste.

En ce qui concerne l'EVI et le NDVI, les résultats dépendent fortement des variables réponses considérées :

- Pour RS.Fores, RS.Urb et le CTI, c'est la période du mois de juin qui donne les meilleures performances explicatives.
- Pour RS.Agri et RS.Tot, c'est la période de fin aout/début septembre qui fournit les meilleurs modèles
- Pour RS.Gene, c'est à la période hivernale que les meilleurs modèles sont obtenus.
- Pour le CSI, les périodes de juin à fin aout donnent de bons résultats.

Pour les oiseaux forestiers et le urbains, c'est principalement la période hivernale qui est à éviter pour la construction de modèles : pendant cette période, les arbres n'ont pas de feuilles et le NDVI est moins adapté pour distinguer les zones forestières des zones de cultures, et pour distinguer le taux d'urbanisation d'une zone.

Pour la richesse agricole et totale, les résultats obtenus sont cohérents avec ceux de Sheeren et al. (2014), qui trouvent que la fin d'été/début d'automne est la période la plus adaptée pour corréler des images NDVI aux relevés d'oiseaux. En fin d'été, les cultures d'été sont récoltées et par conséquent, les milieux agricoles, en labour, sont plus facilement différentiables des éléments ligneux, les feuillus ne perdant leurs feuilles qu'un peu plus tard dans la saison. Cette distinction est essentielle, notamment pour les oiseaux agricoles. En revanche, les résultats diffèrent pour les oiseaux forestiers. Sheeren et al. (2014) observent que c'est également la période de fin d'été qui est la plus adaptée contrairement à nos résultats mettant en évidence la fin du mois de juin. Cela peut s'expliquer par le fait que les cultures d'été (maïs, tournesol) très présentes dans le paysage agri-forestier du Sud-Ouest de leur étude ne sont pas les seules présentes à l'échelle de la France. Il existe également des cultures d'hiver ou de printemps qui sont récoltées plus tôt dans la saison, permettant ainsi une bonne discrimination des éléments du paysage dès le début de l'été.

Faut-il prendre en compte l'effet des régions et des interactions dans notre modèle national ?

Pour la majorité des variables réponses, il n'y a pas d'interaction entre l'effet du NDVI, de l'EVI, du LSTD et du LSTN et les régions.

Cela suggère que l'effet du NDVI sur la richesse est généralisable à l'échelle de la France.

Cependant, les carrés STOC ont été filtrés sur la qualité des images satellites associées, et de nombreux relevés des zones montagnardes ont été supprimés. Aussi, la généralisation des modèles sur ces zones est à éviter.

Concernant la stratification, quel que soit sa nature, le type de strate est une variable explicative pertinente à prendre en compte dans les modèles. Ce résultat est assez logique puisque les communautés d'oiseaux sont différentes en fonction des zones biogéographiques françaises, à cause de différences historiques, climatiques ou autres.

Les écarts entre les différentes stratifications ne sont pas très élevés. On peut toutefois noter que les régions agricoles sont celles qui améliorent le plus les modèles, probablement en partie parce qu'elles présentent davantage de modalités (13 contre 8 pour les régions climatiques et 11 pour les sylvo-écorégions). Cela montre également que le type de paysage agricole joue un rôle important dans la répartition de la biodiversité.

Pour une date donnée, les variables sont-elles complémentaires dans la construction des modèles ?

Concernant les interactions, pour la majorité des variables réponses, il n'y a pas d'interaction entre l'effet du NDVI, de l'EVI, du LSTD et du LSTN avec les régions. Autrement dit, l'effet de ces variables explicatives ne varie pas suivant la nature de la région considérée.

Aussi, les variables EVI, LSTD et LSTN d'une part et NDVI, LSTD et LSTN d'autre part sont porteuses d'une information complémentaire.

Les combinaisons contenant l'EVI donnent de meilleurs résultats que celles contenant le NDVI pour RS.Gene et RS.Tot, et réciproquement pour RS.Fores, RS.Urb et le CSI. Cela est concordant avec les résultats obtenus pour les variables NDVI et EVI séparément.

Cette complémentarité entre les variables a également été observée par Venier et al. (2004), qui ont aussi obtenu de meilleurs modèles en utilisant des données combinées du climat et du paysage : ils concluent que le climat n'est pas qu'un substitut pour le paysage, car il améliore souvent les modèles d'espèces significativement plus que les modèles paysagers ne le pourraient.

La complémentarité entre le LST et les indices de végétation (NDVI en particulier) a déjà été bien étudiée précédemment (Carlson, 2007; Petropoulos et al., 2009). Le LST est faiblement sensible aux variations d'humidité dans des zones fortement recouvertes de végétation mais cette sensibilité augmente lorsque le taux de couverture de la végétation diminue, faisant apparaître un effet sol. L'évolution de l'indice de végétation, selon le taux de recouvrement de la végétation, fait alors émerger une répartition des couples de valeurs LST/NDVI sous la forme d'un trapèze (figure 33). Les deux indices LST/NDVI réunis dans nos modèles permettent par conséquent de mieux exprimer les caractéristiques des habitats auxquels sont sensibles les communautés d'oiseaux. Cette complémentarité est cependant à nuancer. En effet, Cord et Rödder (2011) ont conclu de leurs recherches que les données climatiques sont mieux adaptées à complémenter le NDVI que le LST. Néanmoins, leur travail est basé sur des espèces d'anoures mexicains et non sur des communautés d'oiseaux.

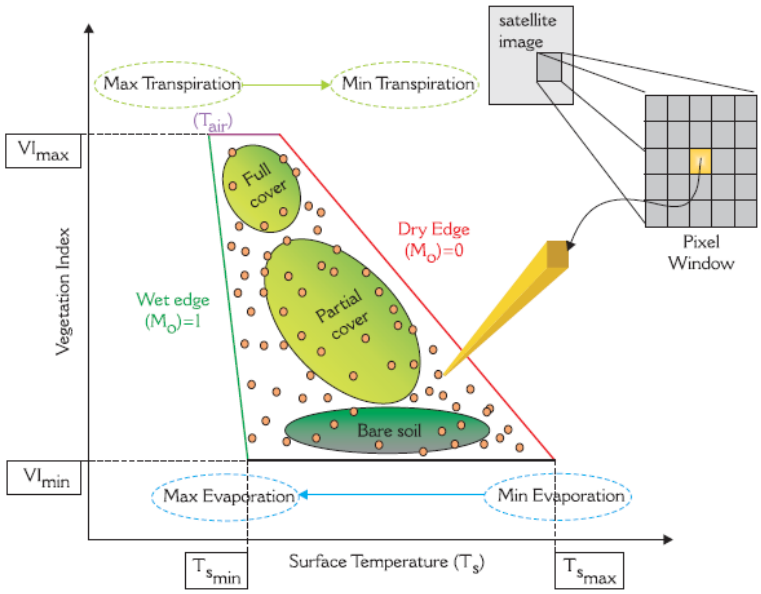


Figure 1 Summary of the key descriptors and physical interpretations of the T_s/VI feature space or 'scatterplot'

Source : Petropoulos et al., 2009

Figure 33 : Représentation du triangle des corrélations entre indices de végétation et LST (source : Petropoulos et al., 2009)

Pour une variable donnée, la prise en compte de plusieurs dates permet-elle d'améliorer la performance des modèles ?

A l'exception des modèles issus du LSTN pour certaines variables réponses, les modèles prenant en compte les 5 dates donnent de meilleures performances que les modèles à une date, même lorsque la quantité de variables est prise en compte (critère AIC, cf. annexe 4).

Pour les variables NDVI, EVI et LSTD, les différentes dates contiennent donc une information différente qui est complémentaire dans la construction des modèles.

Cela peut s'expliquer par l'échelonnage des processus de sélection d'habitat des oiseaux sur une bonne partie de l'année (Hutto, 1985).

Le LSTN ne donne pas systématiquement de meilleurs modèles lorsque les 5 dates sont prises en compte, ce qui suggère que son information portée par les différentes dates est similaire. En effet, les températures de nuit sont « lissées » par rapport aux températures de jour, et le LSTN semble principalement discriminer de grandes zones climatiques facilement distinguables à toutes les périodes de l'année.

Quel est l'apport de la classification fonctionnelle ?

Cette classification revient à prendre en considération plusieurs dates mais de manière intégrée et sans sélection *a priori* d'une période particulière. Les classes obtenues représentent des types d'éléments paysagers mais aussi leur fonctionnement phénologique. Les modèles obtenus sont plus performants que ceux réalisés à partir d'images prises à une seule période de l'année. En revanche, l'intégration des cinq dates rend les modèles issus d'images non classées globalement équivalents à ceux issus de cette classification (les différences observées dépendent des variables réponses étudiées).

Ces résultats confirment l'importance de la prise en compte de la phénologie des couverts pour expliquer les patrons de communautés d'oiseaux. Ils ouvrent également de nouvelles perspectives intéressantes dans la modélisation des patrons de communautés d'oiseaux à partir d'images satellites.

Quel est l'impact de la représentation spatiale (continue ou classée) utilisée pour créer les variables explicatives sur la performance des modèles ?

Le type de représentation spatiale (image non classée versus image classée) influe fortement sur la performance explicative des modèles.

D'une manière générale, à l'exception de la RS.Fores, les variables NDVI, EVI, LST fournissent des résultats moins performants que ceux obtenus avec CORINE Land Cover lorsque la stratification n'est pas prise en compte. En revanche, lorsque les strates sont introduites dans les modèles, la performance est systématiquement plus élevée avec les variables issues des images non classées. Ceci confirme la pertinence et l'intérêt de construire des modèles oiseaux/paysage à partir d'images satellitaires non classées.

Concernant la comparaison des produits classés (CORINE, produit VCF, classification fonctionnelle), on constate un résultat intéressant avec le VCF de MODIS. En fournissant pour chaque pixel un pourcentage de couverture forestière, le produit VCF montre des performances plus élevées que les modèles construits à partir de CORINE Land Cover n'intégrant que la présence ou l'absence de forêts. Pour les oiseaux spécialistes (RS.Fores, RS.Agri et RS.Urb), les modèles issus du VCF sont même plus performants que ceux issus de CORINE Land Cover prenant en compte toute la nomenclature. La comparaison entre les différents modèles montre aussi l'intérêt de la classification fonctionnelle des pixels (i.e. la classification de leur signature temporelle) : à l'exception de RS.Gene et RS.Tot, ces modèles présentent une performance explicative significativement plus élevée que les modèles issus de CORINE Land Cover intégrant toutes les classes. En particulier, pour les oiseaux des milieux urbains, la déviance expliquée est plus que doublée par rapport aux autres modèles.

Pour la plupart des variables réponses (à l'exception de RS.Fores et du CSI), les modèles issus de données classées donnent de meilleures performances explicatives, mais l'écart est faible et ne justifie peut-être pas le temps nécessaire à la création des cartes d'occupation du sol. De plus, lorsque la stratification est rajoutée aux modèles, les modèles issus de données brutes deviennent plus performants que ceux issus de données classées.

4.2. Limites de l'étude

Malgré les résultats intéressants obtenus, plusieurs limites de note étude peuvent être discutées.

Des limites dans les données oiseaux :

Lors de ma participation aux relevés oiseaux sur le terrain pour le carré STOC n° 31-0507, j'ai pu constater des difficultés dans la mise en œuvre du protocole STOC Vigie nature, concernant (i) l'échantillonnage des points d'écoute, (ii) les conditions de relevé et (iii) le protocole du relevé en lui-même.

(i). Les points d'écoute sont placés par chaque bénévole, et selon des méthodes qui peuvent être différentes :

- Soit l'échantillonnage est systématique : les points sont répartis aléatoirement dans le carré. Les résultats d'abondance peuvent ensuite mieux être analysés d'un point de vue statistique car cela diminue l'effet observateur.
- Soit l'échantillonnage est stratifié selon les habitats : tous les habitats présents sont représentés, et la richesse spécifique du carré est par conséquent plus fiable.

(ii). Les relevés doivent s'effectuer sous des conditions météorologiques optimales (sans pluie, avec peu de vent) pour optimiser la détection auditive des individus. Cela n'est pas toujours simple à respecter par des bénévoles n'ayant souvent que les week-ends disponibles, et le protocole imposant de garder des dates proches d'une année sur l'autre. De plus, la météo peut être amenée à évoluer au cours de la matinée d'échantillonnage (Bas et al., 2008). Par ailleurs, les relevés doivent être terminés quatre heures après le lever du soleil, ce qui n'est pas toujours le cas en pratique.

(iii). Les espèces vues ou entendues doivent être notées pendant cinq minutes. Cependant, lors des sessions de terrain, il est arrivé que des oiseaux soient observés en dehors des cinq minutes. Ces oiseaux ont tout de même été notés, l'objectif du relevé étant d'identifier les oiseaux nicheurs. Toutefois, tous les bénévoles ne procèdent pas de la même façon, et cette source d'hétérogénéité pourrait créer un biais dans les résultats. D'autres études proposent des temps d'observation plus longs (dix minutes, (Pidgeon et al., 2001)), ou davantage de sessions d'observation (trois passages, (Seto et al., 2004)), mais cela implique un engagement plus important de la part des bénévoles. De plus, la durée de comptage n'affecte pas les performances explicatives et prédictives des modèles de distribution des espèces (Bonthoux and Balent, 2012).

On peut également noter que la distance des détections par rapport au point de comptage est illimité. Or certaines espèces, comme le Coucou gris (*Cuculus canorus*) sont plus bruyantes et détectables à plus grande distance que d'autres. Cela augmente le risque de double comptage pour les individus de ces espèces.

Malgré ces limites, le protocole STOC présente néanmoins le très gros avantage d'être standardisé et généralisé à toute la France. La quantité de données disponibles est supposée gommer le bruit induit par les différents effets observateurs. De plus, c'est de plus le seul programme qui permet de couvrir une telle étendue spatiale et temporelle en France.

Des limites dans les données satellitaires :

Des limites d'ordre temporel ont également été identifiées au cours de l'étude.

(i). Les analyses sur les données classées ont été faites à partir de CORINE Land Cover. Or ces données datent de 2006. Il existe donc un écart de quatre ans avec les données oiseaux. Cet écart n'existe pas pour les données non classées puisque les images satellites utilisées datent de 2010. Aussi, une partie de la différence observée entre les performances des modèles classés et non classés peut être lié à ce décalage temporel. L'intérêt des données non classées réside dans le fait que, nécessitants moins de temps de traitement, elles peuvent être disponibles plus fréquemment, et les modèles peuvent être réactualisés chaque année.

(ii). L'étude a été réalisée sur l'année 2010, correspondant aux données oiseaux les plus récentes dont nous disposions. Cependant, un effet année peut être présent dans nos données. Aussi, il pourrait être plus pertinent de moyenner sur plusieurs années à la fois les données oiseaux et les celles issues des images satellites.

En particulier, la phénologie est particulièrement dépendante du climat. L'année 2010 a été marquée par plusieurs événements météorologiques marquants (Météo France) : la tempête Xynthia fin février, des épisodes pluvieux exceptionnels (dans le Var en juin et dans le Sud-Est en septembre), et un mois de décembre particulièrement froid et neigeux. Cette dernière anomalie se retrouve fortement dans les courbes de NDVI issues des classes fonctionnelles (cf. figures 28). Les courbes des clusters F et H notamment sont essentiellement différenciées par la présence d'épisodes neigeux en novembre ou en décembre. Moyenner sur plusieurs années aurait permis de prendre en compte la variation interannuelle de la saisonnalité de la végétation et les anomalies ponctuelles (Cord and Rödder, 2011; Gottschalk et al., 2005). Nous avons filtré les pixels sur les carrés STOC déjà conservés dans les autres modèles. Cela a permis d'homogénéiser avec les autres modèles, mais également d'assurer la présence d'au moins 5 dates de bonne qualité, mais aucune vérification sur la qualité des données aux autres dates n'a été réalisée par manque de temps.

Des limites sur la construction des modèles :

Les modèles construits sont de type GLM. Or nous avons vu que les résidus sont autocorrélés spatialement. En toute rigueur, il faudrait donc construire des modèles spatiaux prenant en compte la répartition géographique des carrés STOC.

De plus, la sélection des carrés STOC sur la qualité des données satellites a changé la répartition des données (cf. figure 34). Presque tous les carrés de haute montagne ont été retirés de l'étude. Les modèles ne s'appliquent donc pas sur ces régions.

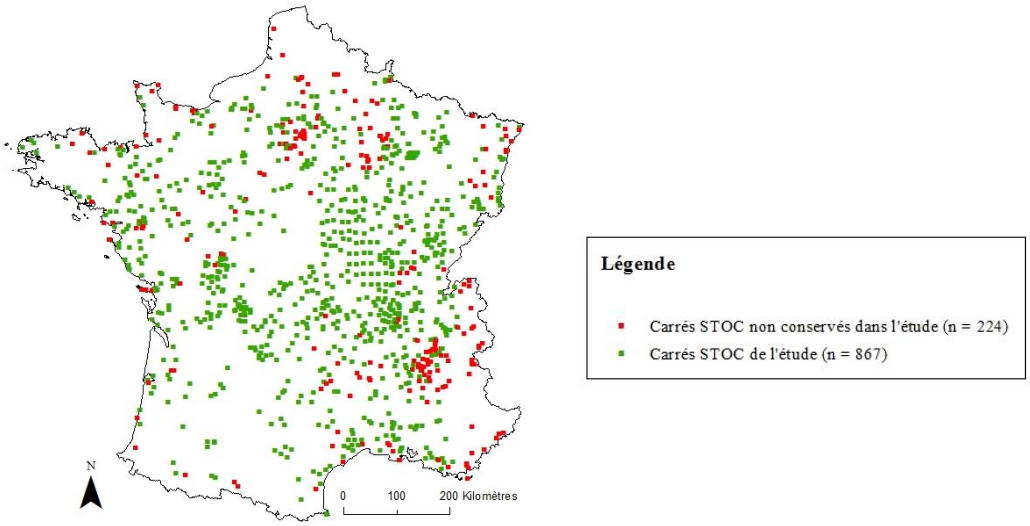


Figure 34 : Répartition des carrés STOC conservés pour l'étude parmi ceux relevés en 2010

4.3. Apports de l'étude et perspectives

Cette étude présente un certain nombre de nouveautés :

- Contrairement à l'approche traditionnelle, elle prend en compte le paysage de manière continue (images brutes NDVI, EVI, LST) et ne se base pas uniquement sur des cartes d'occupation du sol.
- Lorsque cette approche est suivie, la question du choix de la date peut fortement impacter les résultats des modèles. Ici, plusieurs dates sont considérées et nous avons comparé leurs effets sur les performances des modèles. Nous avons également exploré l'intérêt de prendre en compte l'ensemble du cycle phénologique en classant les pixels selon leur signature temporelle.
- Enfin, plusieurs types de données ont été utilisées, et en particulier le LST, tandis que la plupart des recherches se focalisent actuellement sur les indices de végétation.

De très nombreuses perspectives sont offertes par cette étude.

Tout d'abord, les limites observées suggèrent que l'on pourrait reprendre cette étude en travaillant sur des données moyennées, et en utilisant des modèles prenant en compte l'autocorrélation spatiale.

Au-delà des limites de l'étude, la très grande quantité de données ouvre la porte à de nombreuses analyses non explorées dans le cadre de stage. Quelques exemples sont listés ci-après :

- Nous nous sommes limités à des modèles linéaires généralisés (GLM). Nous pourrions évaluer la pertinence d'autres modèles et algorithmes d'apprentissage (GAM, arbres de décisions) permettant de modéliser des relations non linéaires.
- Nous avons également focalisé la prise en compte de la phénologie sur la classification fonctionnelle. Il pourrait être intéressant de rester sur une approche continue en calculant des métriques phénologiques telles que discutées en 1.1 (figure 4).
- L'effet date a été étudié à l'échelle de la France, mais la phénologie est fortement dépendante des régions, avec notamment des dates de fauches variables. Aussi, l'effet date pourrait être étudié régionalement.
- L'année étudiée est celle de 2010, une autre possibilité aurait été de comparer les résultats avec ceux d'une année antérieure afin d'étudier l'évolution des communautés.
- La performance prédictive des modèles construits pourrait aussi être analysée de manière temporelle, en appliquant nos modèles de 2010 aux années 2009 ou 2011.

Finalement, ces analyses ont été réalisées sur les oiseaux, dont la biologie est déjà bien connue en France. La méthodologie s'avérant concluante sur ce taxon, il pourrait être intéressant de l'appliquer à d'autres données de biodiversité moins connues ou moins documentées.

Références bibliographiques

Albright, T.P., Pidgeon, A.M., Rittenhouse, C.D., Clayton, M.K., Flather, C.H., Culbert, P.D., and Radeloff, V.C. (2011). Heat waves measured with MODIS land surface temperature data predict changes in avian community structure. Remote Sensing of Environment *115*, 245–254.

Bacaro, G., Santi, E., Rocchini, D., Pezzo, F., Puglisi, L., and Chiarucci, A. (2011). Geostatistical modelling of regional bird species richness: exploring environmental proxies for conservation purpose. Biodiversity and Conservation *20*, 1677–1694.

Bailey, S.-A., Horner-Devine, M.C., Luck, G., Moore, L.A., Carney, K.M., Anderson, S., Betrus, C., and Fleishman, E. (2004). Primary productivity and species richness: relationships among functional guilds, residency groups and vagility classes at multiple spatial scales. Ecography *27*, 207–217.

Barnagaud, J.Y., Devictor, V., Jiguet, F., and Archaux, F. (2011). When species become generalists: on-going large-scale changes in bird habitat specialization: Temporal trends in bird specialization. Global Ecology and Biogeography *20*, 630–640.

Bas, Y., Devictor, V., Moussus, J.-P., and Jiguet, F. (2008). Accounting for weather and time-of-day parameters when analysing count data from monitoring programs. Biodivers Conserv *17*, 3403–3416.

Bino, G., Levin, N., Darawshi, S., Van Der Hal, N., Reich-Solomon, A., and Kark, S. (2008). Accurate prediction of bird species richness patterns in an urban environment using Landsat-derived NDVI and spectral unmixing. International Journal of Remote Sensing *29*, 3675–3700.

Bonthoux, S., and Balent, G. (2012). Point count duration: five minutes are usually sufficient to model the distribution of bird species and to study the structure of communities for a French landscape. Journal of Ornithology *153*, 491–504.

Bonthoux, S., Barnagaud, J.-Y., Goulard, M., and Balent, G. (2013). Contrasting spatial and temporal responses of bird communities to landscape changes. Oecologia *172*, 563–574.

Both, C., Bouwhuis, S., Lessells, C.M., and Visser, M.E. (2006). Climate change and population declines in a long-distance migratory bird. Nature *441*, 81–83.

Burnham, K.P., and Anderson, D.R. (2008). Model selection and multimodel inference: a practical imformation-theoretic approach (New York, NY [u.a.]: Springer).

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., et al. (2012). Biodiversity loss and its impact on humanity. Nature *486*, 59–67.

Carlson, T. (2007). An overview of the "Triangle Method" for estimating surface evapotranspiration and soil moisture from satellite imagery. Sensors *7*, 1612–1629.

Centre Canadien de Télédétection. Ressources naturelles Canada : http://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/imagerie-satellitaire-produits/ressources-educatives/9364. Consulté le vendredi 1 août 2014.

Clavel, J. (2007). Spécialistes et Généralistes face aux changements globaux : Approches comportementales et évolutives. Université Pierre et Marie-Curie.

Coburn, C.A., and Roberts, A.C.B. (2004). A multiscale texture analysis procedure for improved forest stand classification. International Journal of Remote Sensing *25*, 4287–4308.

Cord, A., and Rödder, D. (2011). Inclusion of habitat availability in species distribution models through multi-temporal remote-sensing data? Ecological Applications *21*, 3285–3298.

Culbert, P.D., Pidgeon, A.M., St-Louis, V., Bash, D., and Radeloff, V.C. (2009). The impact of phenological variation on texture measures of remotely sensed imagery. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of *2*, 299–309.

Culbert, P.D., Radeloff, V.C., St-Louis, V., Flather, C.H., Rittenhouse, C.D., Albright, T.P., and Pidgeon, A.M. (2012). Modeling broad-scale patterns of avian species richness across the Midwestern United States with measures of satellite image texture. Remote Sensing of Environment *118*, 140–150.

Davey, C.M., Chamberlain, D.E., Newson, S.E., Noble, D.G., and Johnston, A. (2012). Rise of the generalists: evidence for climate driven homogenization in avian communities. Global Ecology and Biogeography *21*, 568–578.

Devictor, V., and Jiguet, F. (2007). Community richness and stability in agricultural landscapes: The importance of surrounding habitats. Agriculture, Ecosystems & Environment *120*, 179–184.

Devictor, V., Julliard, R., Clavel, J., Jiguet, F., Lee, A., and Couvet, D. (2008a). Functional biotic homogenization of bird communities in disturbed landscapes. Global Ecology and Biogeography *17*, 252–261.

Devictor, V., Julliard, R., and Jiguet, F. (2008b). Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos *0*, 080211051304426–0.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., and Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world: Spatial mismatch between diversity facets. Ecology Letters no–no.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D.W., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography *30*, 609–628.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography *36*, 27–46.

Duro, D.C., Girard, J., King, D.J., Fahrig, L., Mitchell, S., Lindsay, K., and Tischendorf, L. (2014). Predicting species diversity in agricultural environments using Landsat TM imagery. Remote Sensing of Environment *144*, 214–225.

Eklundha, L., and Jönssonb, P. (2012). TIMESAT 3.1 software manual. Lund University, Sweden.

Fahrig, L., Baudry, J., Brotons, L., Burel, F.G., Crist, T.O., Fuller, R.J., Sirami, C., Siriwardena, G.M., and Martin, J.-L. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes: Heterogeneity and biodiversity. Ecology Letters *14*, 101–112.

Ficetola, G.F., Bonardi, A., Mücher, C.A., Gilissen, N.L.M., and Padoa-Schioppa, E. (2014). How many predictors in species distribution models at the landscape scale? Land use versus LiDAR-derived canopy height. International Journal of Geographical Information Science 1–17.

Foody, G.M. (2004). Spatial nonstationarity and scale-dependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecology and Biogeography *13*, 315–320.

Fuller, R.M., Devereux, B.J., Gillings, S., Hill, R.A., and Amable, G.S. (2007). Bird distributions relative to remotely sensed habitats in Great Britain: Towards a framework for national modelling. Journal of Environmental Management *84*, 586–605.

Genolini, C., and Falissard, B. (2011). KmL: A package to cluster longitudinal data. Computer Methods and Programs in Biomedicine *104*, e112–e121.

Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., and Saatchi, S. (2008). Measuring and modelling biodiversity from space. Progress in Physical Geography *32*, 203–221.

Goetz, S., Steinberg, D., Dubayah, R., and Blair, B. (2007). Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. Remote Sensing of Environment *108*, 254–263.

Gottschalk, T.K., Huettmann, F., and Ehlers, M. (2005). Review article: Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. International Journal of Remote Sensing *26*, 2631–2656.

Gregory, R.D., Willis, S.G., Jiguet, F., Voříšek, P., Klvaňová, A., van Strien, A., Huntley, B., Collingham, Y.C., Couvet, D., and Green, R.E. (2009). An Indicator of the Impact of Climatic Change on European Bird Populations. PLoS ONE *4*, e4678.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., Dimiceli, C., and Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. Remote Sensing of Environment *83*, 303–319.

Huete, A., Justice, C., and Van Leeuwen, W. (1999). MODIS vegetation index (MOD13). Algorithm Theoretical Basis Document *3*, 213.

Hurlbert, A.H., and Haskell, J.P. (2003). The effect of energy and seasonality on avian species richness and community composition. The American Naturalist *161*, 83–97.

Hurley, M.A., Hebblewhite, M., Gaillard, J.-M., Dray, S., Taylor, K.A., Smith, W.K., Zager, P., and Bonenfant, C. (2014). Functional analysis of Normalized Difference Vegetation Index curves reveals overwinter mule deer survival is driven by both spring and autumn phenology. Philosophical Transactions of the Royal Society B: Biological Sciences *369*, 20130196–20130196.

Hurvich, C.M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. Biometrika *76*, 297–307.

Hutto, R.L. (1985). Habitat Selection in Birds (M.L. Cody).

Jiguet, F., Gadot, A.-S., Julliard, R., Newson, S.E., and Couvet, D. (2007). Climate envelope, life history traits and the resilience of birds facing global change. Global Change Biology *13*, 1672–1684.

Jiguet, F., Devictor, V., Julliard, R., and Couvet, D. (2012). French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. Acta Oecologica *44*, 58–66.

Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., and Wavresky, P. (2010). Les types de climats en France, une construction spatiale. Cybergeo : European Journal of Geography.

JORF du 11 décembre (1980). Arrêté du 28 novembre 1980 TERMINOLOGIE DE LA TELEDETECTION AEROSPATIALE.

Julliard, R., and Jiguet, F. (2002). Integrated monitoring of common bird species in France. Alauda (Dijon) *70*, 137–147.

Julliard, R., Jiguet, F., and Couvet, D. (2004). Evidence for the impact of global warming on the long–term population dynamics of common birds. Proc. R. Soc. Lond. B *271*, S490–S492.

Laurent, E.J., Shi, H., Gatziolis, D., LeBouton, J.P., Walters, M.B., and Liu, J. (2005). Using the spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns. Remote Sensing of Environment *97*, 249–262.

Lawler, J.J., O'Connor, R.J., Hunsaker, C.T., Jones, K.B., Loveland, T.R., and White, D. (2004). The effects of habitat resolution on models of avian diversity and distributions: a comparison of two land-cover classifications. Landscape Ecology *19*, 515–530.

Levin, N., McAlpine, C., Phinn, S., Price, B., Pullar, D., Kavanagh, R.P., and Law, B.S. (2009). Mapping forest patches and scattered trees from SPOT images and testing their ecological importance for woodland birds in a fragmented agricultural landscape. International Journal of Remote Sensing *30*, 3147–3169.

Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkönig, I.M.A., and Skidmore, A. (2007). Capturing the fugitive: Applying remote sensing to terrestrial animal distribution and diversity. International Journal of Applied Earth Observation and Geoinformation *9*, 1–20.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. (The Regents of the University of California),.

Météo France Bilan de l'année 2010. http://www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2010/bilan-de-lannee-2010. Consulté le 9 juillet 2014.

Meynard, C.N., Devictor, V., Mouillot, D., Thuiller, W., Jiguet, F., and Mouquet, N. (2011). Beyond taxonomic diversity patterns: how do α, β and γ components of bird functional and phylogenetic diversity respond to environmental gradients across France?: Multiple facets of diversity. Global Ecology and Biogeography *20*, 893–903.

Mimet, A., Maurel, N., Pellissier, V., Simon, L., and Julliard, R. (2014). Towards a unique landscape description for multi-species studies: A model comparison with common birds in a human-dominated French region. Ecological Indicators *36*, 19–32.

Morelli, F., Pruscini, F., Santolini, R., Perna, P., Benedetti, Y., and Sisti, D. (2013). Landscape heterogeneity metrics as indicators of bird diversity: Determining the optimal spatial scales in different landscapes. Ecological Indicators *34*, 372–379.

Muséum National d'Histoire Naturelle Vigie-Nature. Protocole STOC. http://vigienature.mnhn.fr/page/protocole. Consulté le 16 avril 2014.

Muséum National d'Histoire Naturelle Vigie-Nature. Produire des indicateurs à partir des indices des espèces. http://vigienature.mnhn.fr/page/produire-des-indicateurs-partir-des-indices-des-especes-habitat. Consulté le 5 juin 2014.

Naef-Daenzer, B., Luterbacher, J., Nuber, M., Rutishauser, T., and Winkel, W. (2012). Cascading climate effects and related ecological consequences during past centuries. Clim. Past *8*, 1527–1540.

National Aeronautics and Space Administration MODIS Web. http://modis.gsfc.nasa.gov/. Consulté de mars à août 2010.

Petropoulos, G., Carlson, T.N., Wooster, M.J., and Islam, S. (2009). A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture. Progress in Physical Geography *33*, 224–250.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., and Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution *20*, 503–510.

Pettorelli, N., Ryan, S., Mueller, T., Bunnefeld, N., Jedrzejewska, B., Lima, M., and Kausrud, K. (2011). The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. Climate Research *46*, 15–27.

Pfeifer, M., Disney, M., Quaife, T., and Marchant, R. (2012). Terrestrial ecosystems from space: a review of earth observation products for macroecology applications: Satellite products for macroecology. Global Ecology and Biogeography *21*, 603–624.

Pidgeon, A.M., Mathews, N.E., Benoit, R., and Nordheim, E.V. (2001). Response of Avian Communities to Historic Habitat Change in the Northern Chihuahuan Desert. Conservation Biology *15*, 1772–1789.

Princé, K., Lorrillière, R., Barbet-Massin, M., and Jiguet, F. (2013). Predicting the fate of French bird communities under agriculture and climate change scenarios. Environmental Science & Policy *33*, 120–132.

Prins, E., Petersen, B.S., Aunins, A., and Priednieks, J. (2005). Using Landsat TM and field data to produce maps of predicted bird densities in Latvian farmland. International Journal of Remote Sensing *26*, 1881–1891.

Rittenhouse, C.D., Pidgeon, A.M., Albright, T.P., Culbert, P.D., Clayton, M.K., Flather, C.H., Masek, J.G., and Radeloff, V.C. (2012). Land-Cover Change and Avian Diversity in the Conterminous United States: Land-Cover Change and Avian Diversity. Conservation Biology *26*, 821–829.

Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., et al. (2010). Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. Ecological Informatics *5*, 318–329.

Rouquette, J.L., and Pflimlin, A. (1995). Les grandes régions d'élevage : proposition de zonage pour la France. (Clermont-Ferrand),.

Rouse, J.W., Haas, R.H., Schell, J.A., and Deering, D.W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. NASA Special Publication *SP-351I*, 309–317.

Selim, S.Z., and Ismail, M.A. (1984). K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality. IEEE Transactions on Pattern Analysis and Machine Intelligence *PAMI-6*, 81–87.

Sellers, P.J. (1987). Canopy reflectance, photosynthesis and transpiration. II. The role of biophysics in the linearity of their interdependence. International Journal of Remote Sensing *6*, 1335–1372.

Seto, K.C., Fleishman, E., Fay, J.P., and Betrus, C.J. (2004). Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. International Journal of Remote Sensing *25*, 4309–4324.

Shannon, C.E. (1948). A mathematical theory of communication. The Bell System Technical Journal *77*, 379–423, 623–656.

Sheeren, D., Bonthoux, S., and Balent, G. (2014). Modeling bird communities using unclassified remote sensing imagery: Effects of the spatial resolution and data period. Ecological Indicators *43*, 69–82.

Shmueli, G. (2010). To Explain or to Predict? Statistical Science *25*, 289–310.

Simpson, E.H. (1949). Measurement of diversity. Nature *163*, 688.

St-Louis, V., Pidgeon, A.M., Radeloff, V.C., Hawbaker, T.J., and Clayton, M.K. (2006). High-resolution image texture as a predictor of bird species richness. Remote Sensing of Environment *105*, 299–312.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., and Steininger, M. (2003). Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution *18*, 306–314.

Venier, L.A., Pearce, J., McKee, J.E., McKenney, D.W., and Niemi, G.J. (2004). Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. Journal of Biogeography *31*, 315–331.

Wan, Z., Zhang, Y., Zhang, Q., and Li, Z. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. Remote Sensing of Environment *83*, 163–180.

Wood, E.M., Pidgeon, A.M., Radeloff, V.C., and Keuler, N.S. (2013). Image texture predicts avian density and species richness. PloS One *8*, e63211.

Wu, X., Lv, M., Jin, Z., Michishita, R., Chen, J., Tian, H., Tu, X., Zhao, H., Niu, Z., Chen, X., et al. (2014). Normalized difference vegetation index dynamic and spatiotemporal distribution of migratory birds in the Poyang Lake wetland, China. Ecological Indicators.

ANNEXES

Annexe 1 : Liste des espèces considérées dans chaque groupe

Oiseaux forestiers :

Nom français	Nom latin complet
Grimpereau des jardins	*Certhia brachydactyla*
Grimpereau des bois	*Certhia familiaris*
Grosbec casse-noyaux	*Coccothraustes coccothraustes*
Pic épeiche	*Dendrocopos major*
Pic mar	*Dendrocopos medius*
Pic noir	*Dryocopus martius*
Rougegorge familier	*Erithacus rubecula*
Mésange noire	*Parus ater*
Mésange huppée	*Parus cristatus*
Mésange nonnette	*Parus palustris*
Pouillot de Bonelli	*Phylloscopus bonelli*
Pouillot véloce	*Phylloscopus collybita*
Pouillot siffleur	*Phylloscopus sibilatrix*
Pouillot fitis	*Phylloscopus trochilus*
Pic cendré	*Picus canus*
Bouvreuil pivoine	*Pyrrhula pyrrhula*
Roitelet à triple bandeau	*Regulus ignicapillus*
Roitelet huppé	*Regulus regulus*
Sittelle torchepot	*Sitta europaea*
Fauvette mélanocéphale	*Sylvia melanocephala*
Troglodyte mignon	*Troglodytes troglodytes*
Grive musicienne	*Turdus philomelos*
Grive draine	*Turdus viscivorus*

Oiseaux agricoles :

Nom français	Nom latin complet
Alouette des champs	*Alauda arvensis*
Perdrix rouge	*Alectoris rufa*
Pipit rousseline	*Anthus campestris*
Pipit farlouse	*Anthus pratensis*
Linotte mélodieuse	*Carduelis cannabina*
Caille des blés	*Coturnix coturnix*
Bruant zizi	*Emberiza cirlus*
Bruant jaune	*Emberiza citrinella*
Bruant ortolan	*Emberiza hortulana*
Cochevis huppé	*Galerida cristata*
Pie-grièche écorcheur	*Lanius collurio*
Alouette lulu	*Lullula arborea*
Bruant proyer	*Emberiza calandra*
Bergeronnette printanière	*Motacilla flava*

Traquet motteux	*Œnanthe œnanthe*
Perdrix grise	*Perdix perdix*
Tarier des prés	*Saxicola rubetra*
Tarier pâtre	*Saxicola torquata*
Fauvette grisette	*Sylvia communis*
Huppe fasciée	*Upupa epops*

Espèces supprimées par rapport aux regroupements du CRBPO : la buse variable (*Buteo buteo*), le faucon crécerelle (*Falco tinnunculus*), le vanneau huppé (*Vanellus vanellus*) et le corbeau freux (*Corvus frugilegus*)

Oiseaux urbains :

Nom français	Nom latin complet
Pigeon ramier	*Columba palumbus*
Coucou gris	*Cuculus canorus*
Pinson des arbres	*Fringilla cœlebs*
Geai des chênes	*Garrulus glandarius*
Hypolaïs polyglotte	*Hippolais polyglotta*
Rossignol philomèle	*Luscinia megarhynchos*
Loriot d'Europe	*Oriolus oriolus*
Mésange bleue	*Parus caeruleus*
Mésange charbonnière	*Parus major*
Pic vert	*Picus viridis*
Accenteur mouchet	*Prunella modularis*
Fauvette à tête noire	*Sylvia atricapilla*
Merle noir	*Turdus merula*

Espèce supprimée par rapport aux regroupements du CRBPO : la corneille noire ou mantelée (*Corvus corone*)

Oiseaux généralistes :

Nom français	Nom latin complet
Martinet noir	*Apus apus*
Chardonneret élégant	*Carduelis carduelis*
Verdier d'Europe	*Carduelis chloris*
Choucas des tours	*Corvus monedula*
Hirondelle de fenêtre	*Delichon urbicum*
Hirondelle rustique	*Hirundo rustica*
Moineau domestique	*Passer domesticus*
Moineau friquet	*Passer montanus*
Rougequeue noir	*Phœnicurus ochruros*
Rougequeue à front blanc	*Phœnicurus phœnicurus*
Pie bavarde	*Pica pica*
Serin cini	*Serinus serinus*
Tourterelle turque	*Streptopelia decaocto*

Annexe 2 : Valeurs des AIC des modèles détaillés en 2.5.2 et 3.1

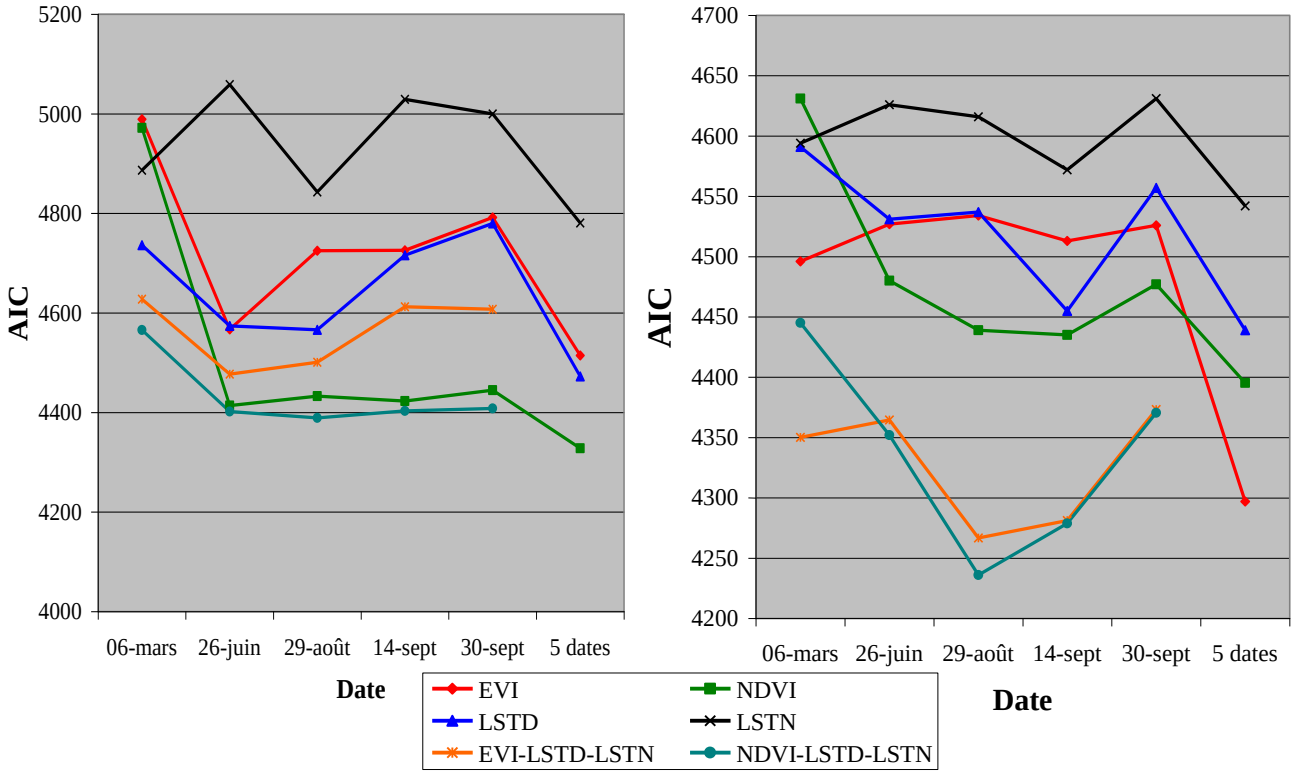
Catégories de variables explicatives		Variable réponse						
Type de variable	**Date**	**RS.Fores**	**RS.Agri**	**RS.Urb**	**RS.Gene**	**RS.Tot**	**CTI**	**CSI**
EVI	065	4989	4496	4571	4040	5912	739	-1043
	177	4567	4527	4544	4059	5942	439	-1157
	241	4725	4534	4658	4065	5948	572	-1130
	257	4726	4513	4675	4063	5915	590	-1116
	273	4792	4526	4679	4051	5857	656	-1114
LSTD	065	4736	4591	4572	4067	5986	438	-1033
	177	4574	4531	4554	4068	5980	417	-1113
	241	4566	4537	4560	4070	5978	227	-1084
	257	4716	4455	4632	4068	5981	542	-1088
	273	4780	4557	4570	4070	5989	408	-1054
LSTN	065	4887	4594	4710	4063	5854	583	-1014
	177	5059	4626	4699	4067	5959	531	-1042
	241	4843	4616	4691	4062	5880	388	-1045
	257	5029	4572	4713	**4056**	5875	630	-1022
	273	5000	4631	4705	4068	5932	669	-1005
NDVI	065	4972	4631	4676	4056	5964	746	-1116
	177	4414	4480	4504	4062	5957	441	-1171
	241	4433	4439	4582	4066	5960	485	-1172
	257	4423	4435	4613	4065	5948	550	-1133
	273	4445	4477	4617	4059	5952	528	-1177

Exemple de lecture pour la **valeur en rouge** : le modèle RS.Gene ~ MoyLSTN257 + EcaLSTN257 donne un AIC de 4056.

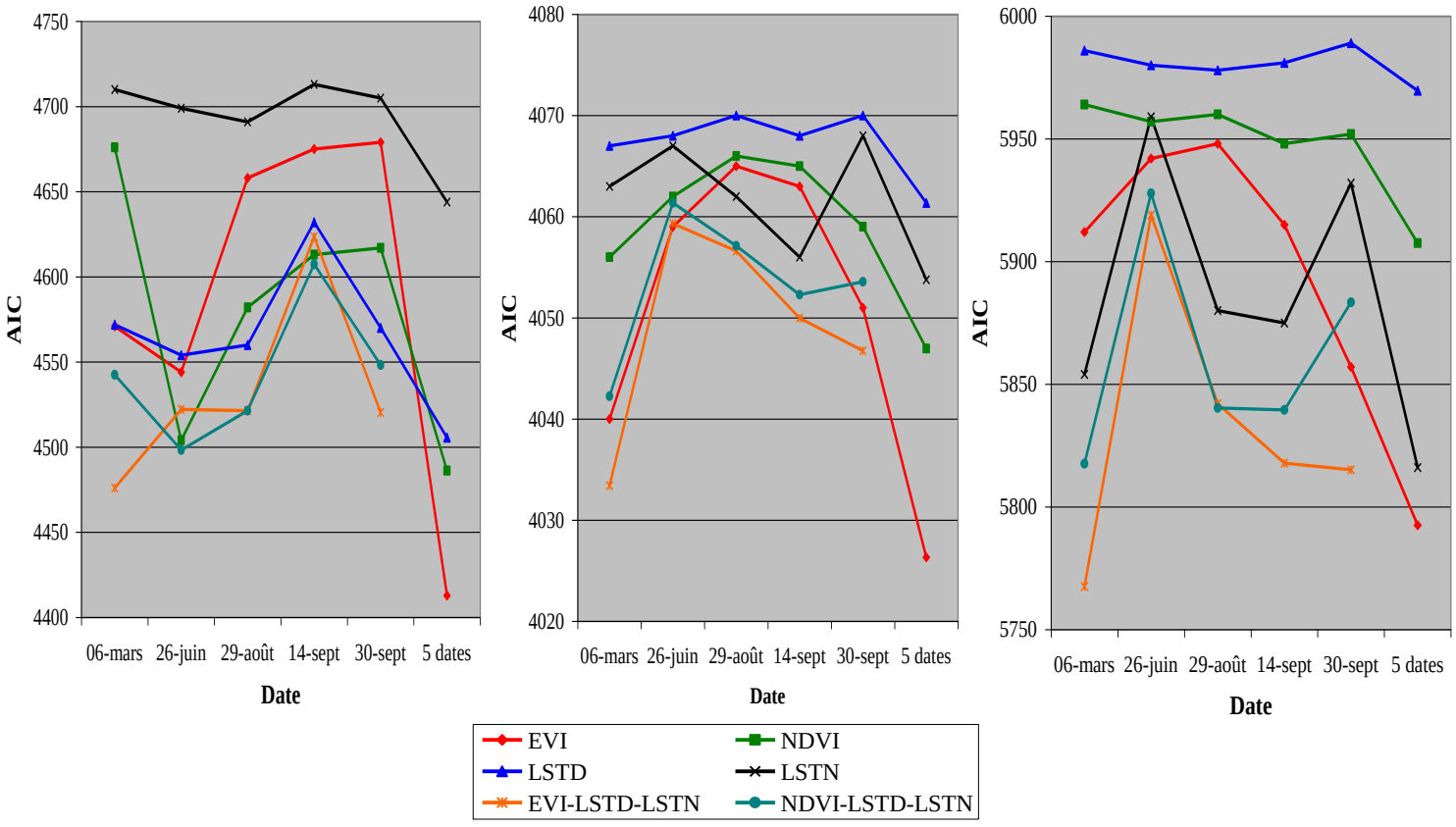
Annexe 3 : AIC des modèles avec effet région simple, présentés en sections 2.5.3 et 3.2

Variable réponse	Régions considérées	Date et type de variable explicative	AIC
RS.Fores	Régions climatiques	NDVI-177	4348
		EVI-177	4473
		LSTD-241	4484
		LSTN-241	4668
RS.Agri	Régions agricoles	NDVI-257	4209
		EVI-065	4349
		LSTD-257	4305
		LSTN-257	4473
RS.Urb	Régions agricoles	NDVI-177	4338
		EVI-177	4386
		LSTD-177	4409
		LSTN-241	4626
RS.Gene	Régions agricoles	NDVI-065	4042
		EVI-065	4036
		LSTD-065	4045
		LSTN-257	4045
RS.Tot	Régions agricoles	NDVI-257	5772
		EVI-273	5754
		LSTD-241	5804
		LSTN-065	5778
CTI	Sylvo-écorégions	NDVI-177	142
		EVI-177	169
		LSTD-241	21
		LSTN-241	226
CSI	Régions agricoles	NDVI-273	-1241
		EVI-177	-1231
		LSTD-177	-1208
		LSTN-241	-1104

Annexe 4 : Valeurs des AIC des modèles présentés en sections 2.5.4, 2.5.5 et 3.3



AIC des modèles de RS.Fores (à gauche) et RS.Agri (à droite)



AIC des modèles de RS.Urb (à gauche), RS.Gene (au centre) et RS.Tot (à droite)

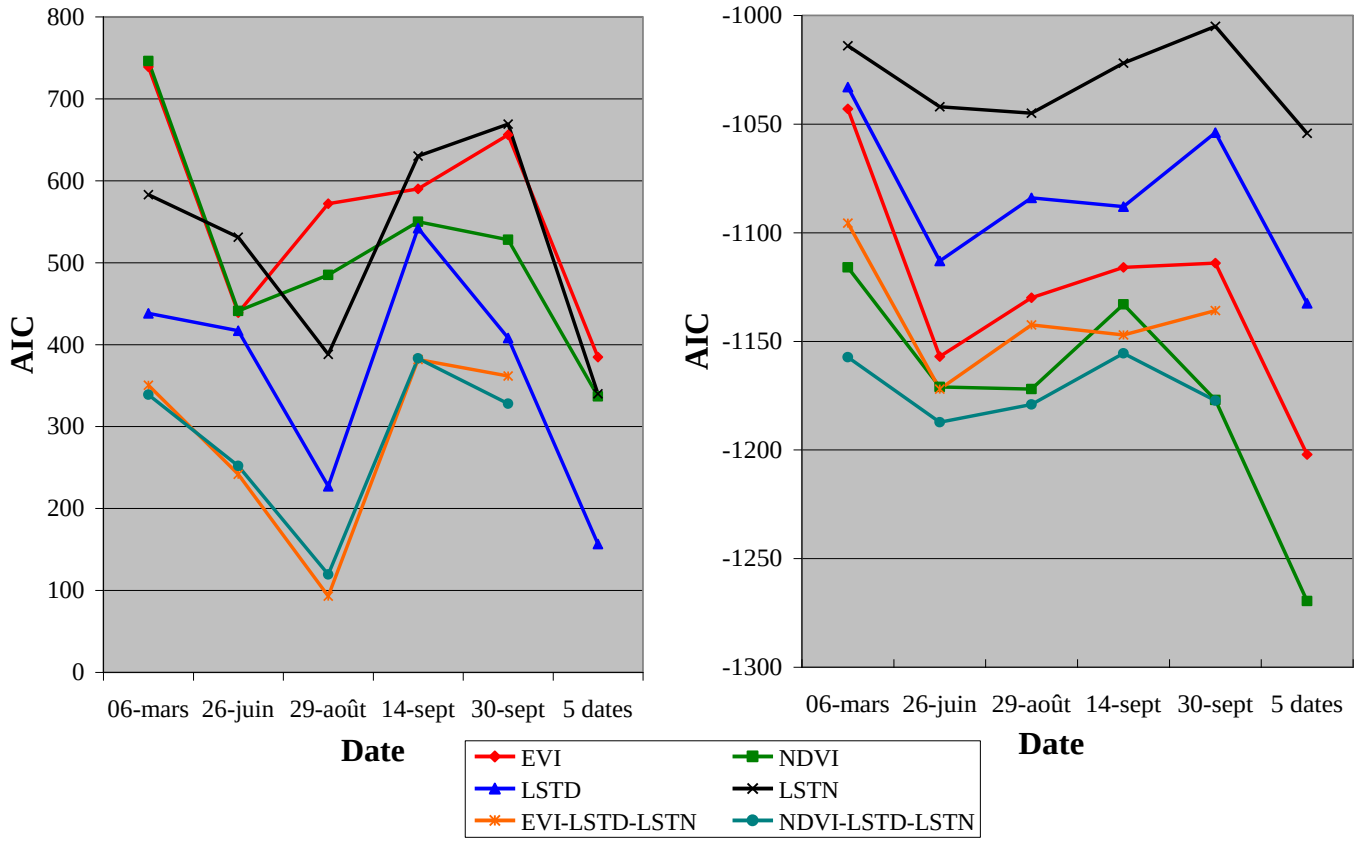


Figure X : AIC des modèles de CTI (à gauche) et du CSI (à droite)

Annexe 5 : AIC des modèles présentés en section 2.5.7 et 3.5

Type de modèle \ Variable réponse	RS.Fores	RS.Agri	RS.Urb	RS.Gene	RS.Tot	CTI	CSI
(a.1) NDVI à la meilleure date	4414	4435	4504	4056	5948	441	-1177
(a.2) EVI à la meilleure date	4567	4496	4544	4040	5857	656	-1114
(a.3) LSTD à la meilleure date	4566	4455	4554	4067	5978	542	-1113
(a.4) LSTN à la meilleure date	4843	4572	4691	4056	5854	388	-1045
(b.1) Variables fonctionnelles 11 clusters	4386	4233	4207	4047	5830	386	-1220
(b.2) VCF	4479	4074	4209	4060	5872	574	-1109
(c.1) CORINE Land Cover forêts	4800	4338	4540	4065	5971	662	-1026
(c.2) CORINE Land Cover complet	4605	4257	4469	4033	5814	415	-1133