

# Analyse de l'incertitude et de la précision thématique de classifications GEOBIA d'une image WorldView-2

François Messner<sup>1</sup>, Jeannine Corbonnois<sup>1</sup>, Fanny Stella Tchitouo Ntanzou<sup>2</sup>

1 : Laboratoire UMR ESO Le Mans, Université du Maine, France

2 : ESGT-CNAM, Le Mans, France

## Résumé

L'évaluation de la précision des cartes thématiques produites par télédétection est une finalité de tout processus de classification modélisant le paysage. Reposant traditionnellement sur la matrice de confusion, elle peut être complétée par des méthodes alternatives plus à même de prendre en compte le biais relatif à la sélection des échantillons d'apprentissage, ainsi que par l'emploi d'approches représentant spatialement l'incertitude inhérente aux classifications. Une telle démarche est adoptée dans cet article, en évaluant la précision à l'aide des estimateurs du Maximum de Probabilité a Posteriori, puis en déterminant, pour chaque unité de carte, des mesures d'incertitude : l'entropie  $\alpha$  quadratique, la divergence de Kullback-Leibler et un indice de concordance qualitatif. Ces traitements sont analysés et comparés selon 3 classifieurs, Random Forest, C5.0 et l'Analyse Discriminante Linéaire et selon 4 stratégies de classification : classifieurs seuls, classifieurs avec procédure de bagging, classifieurs avec procédure d'apprentissage actif et classifieurs avec procédure d'apprentissage actif et de bagging. Les résultats obtenus soulignent la complémentarité des estimateurs de précision pour mettre en évidence un biais dans l'évaluation de la précision ou dans la détermination des probabilités a posteriori, et justifie la prise en considération des indices d'incertitude comme source d'informations sur la distribution spatiale des erreurs de cartographie.

Mots clés : Classification, incertitude, évaluation de la précision thématique, GEOBIA, apprentissage actif, WorldView-2

## Abstract

*The accuracy assessment of thematic maps produced by remote sensing is a goal of any classification process modeling the landscape. Traditionally based on the confusion matrix, it may be supplemented by alternative methods more likely to take into account the relative bias in the selection of training samples and by using approaches spatially representing the inherent classifications uncertainty. Such an approach is used in this article, assessing accuracy using Maximum posterior probability estimators, then determining uncertainty measurements for each map unit: entropy  $\alpha$  quadratic, Kullback-Leibler divergence and a qualitative concordance index. These treatments are analyzed and compared for 3 classifiers, Random Forest, C5.0 and Linear Discriminant Analysis and according to 4 strategies: only classifiers, classifiers with bagging, classifiers with active learning and classifiers with active learning and bagging. The results underline the complementary precision of estimators to highlight bias in assessing the accuracy or determining the posterior probabilities, and justify the consideration of the indices of uncertainty as a source of information on spatial distribution of mapping errors.*

**Keywords:** Classification, uncertainty, thematic accuracy assessment, GEOBIA, active learning, WorldView-2

## 1. Introduction

Pour répondre au mieux aux perturbations, voire aux bouleversements, qu'implique le changement global en matière d'environnement, il est avéré que la connaissance et la classification de l'enveloppe superficielle de la Terre est une nécessité. Dans cette optique, les procédures de cartographie automatique numérique, utilisant notamment comme source de données la télédétection, sont de plus en plus fréquemment employées, et ce dans divers domaines des sciences de la Terre tels que la biogéographie (Corbane et al., 2014 ; Smith et al., 2014), l'écologie (Boyd et Foody, 2011, Nagendra et al., 2013), la pédologie (Minasny et McBratney, 2015 ; Mulder et al., 2011), la géomorphologie (Bishop et al., 2012), la lithologie (van der Meer et al., 2012 ; Cracknell et Reading, 2014)...

Cependant, ces classifications sont des modélisations de paysages naturels, donc complexes, et elles entraînent inévitablement des erreurs et/ou des incertitudes liées à la nature des entités cartographiées ainsi qu'aux données et méthodes employées pour y parvenir (Rocchini et al., 2013b). Dans ce sens, l'incertitude réfère à toutes notions de flou et d'ambiguïté concernant la distinction entre les éléments cartographiés (Rocchini et Ricotta, 2007). Depuis l'avènement de l'utilisation des techniques statistiques, de nombreux travaux ont été effectués pour estimer cette incertitude ou évaluer la précision des classifications de l'occupation du sol. Rocchini et al. (2013b) en ont constitué une synthèse et suggèrent certaines bonnes pratiques à prendre en compte. Cet article souhaite mettre en application quelques-unes de ces recommandations en proposant une méthode d'estimation et de représentation spatiale de

l'incertitude thématique d'une classification supervisée de l'occupation du sol par approche GEOBIA et stockée dans une base de données spatiales. En complément, elle s'accompagne d'une évaluation de la précision de la classification, prenant en compte toutes les unités de cartes, et définie selon la méthode décrite par Steele (2005). Cette évaluation est comparée selon 3 classifieurs, Random Forest, arbre de décision boosté C5.0 et analyse discriminante linéaire (LDA), et selon 4 stratégies de classification supervisée : classifieurs seuls avec jeu d'apprentissage initial, classifieurs avec procédure de bagging, classifieurs avec procédure d'apprentissage actif (Active Learning) et classifieurs avec procédure d'active learning et de bagging.

### 1.1 Classification supervisée d'images de Télédétection

La classification automatique est une branche de la reconnaissance de formes (pattern recognition) dont l'objectif est de regrouper des objets en différentes catégories, identifiés par une étiquette/label, en fonction de leurs caractéristiques statistiques. En télédétection, l'objet est une unité de carte, c'est-à-dire un pixel ou un polygone, qui est caractérisé par des valeurs d'attributs statistiques. Ces attributs peuvent être des réflectances spectrales pour les images de télédétection optique ou des mesures d'émission dans l'infrarouge thermique, des valeurs de retrodiffusion pour les images RADAR, des indicateurs de texture ou même des descripteurs du terrain tels que l'altitude, la pente ou l'orientation (Tso et Mather, 2009). Lors de l'étape de classification, l'espace spectral, par extension l'espace des attributs, représentant la position des objets définis par leurs caractéristiques statistiques est partitionné selon des règles déterminées par un classifieur pour aboutir à un regroupement par classes (Tso et Mather, 2009). L'approche supervisée suppose que l'utilisateur possède une connaissance préalable des différentes catégories thématiques de l'image (les classes représentées par des étiquettes) dont il se servira pour déterminer l'échantillon d'apprentissage nécessaire au classifieur. Pour chaque observation, une valeur d'appartenance, c.-à-d. de probabilité, peut être calculée selon la règle mathématique de l'algorithme de classification. Pour plus de clarté et pour faciliter la compréhension de la suite de cet article, reprenons la formulation mathématique proposée par Steele (2005) : Soit une population  $P$  d'unités de cartes composée de  $c$  classes identifiées par des étiquettes/labels allant de 1, ...,  $c$ . Un élément de  $P$  est une paire  $x = (t, y)$  consistant en un vecteur de covariable  $t$  représentant la position de l'élément dans l'espace des attributs, et une étiquette de classe d'occupation du sol  $y$ . L'échantillon d'apprentissage de taille  $n$  collecté à partir de  $P$  est désigné par  $X = \{x_1, \dots, x_n\}$ . La probabilité a posteriori que  $x \in P$  appartient à la classe  $c$ , étant donné les valeurs des attributs du vecteur covariable  $t$ , est notée  $P(y = c|t)$ .

Notons  $\eta$  la règle de classification appliquée sur l'échantillon d'apprentissage  $X$ , et  $\eta(t)$  la prédiction du label  $y$  en appliquant la règle de classification pour tout élément  $x \in P$ .

La probabilité que l'observation  $x = (t, y)$  soit correctement classée dans la classe  $c$ , par la règle  $\eta$  et étant donné  $t$  est le maximum de probabilité a posteriori :

$$P[\eta(t) = y|t] = \max_c P(y = c|t).$$

Littéralement, cette équation signifie que l'unité de carte est affectée à l'étiquette ayant la plus grande valeur d'appartenance, c'est-à-dire le maximum de probabilité a posteriori parmi toutes les classes disponibles.

Une autre propriété des probabilités a posteriori est l'information « floue » qu'elles fournissent, permettant d'estimer des niveaux variables d'appartenance pour plusieurs catégories thématiques (Gopal et Woodcock, 1994 ; Rocchini et al., 2013b). Tout classifieur reposant sur une fonction d'appartenance, qu'il relève de la logique floue proprement dite (ex : fuzzy rule based), de l'apprentissage automatique (ex : arbre de décision automatique, réseaux de neurones, svm), ou de la statistique bayésienne (ex : maximum de vraisemblance) produit, pour chaque objet, une valeur d'appartenance floue  $\mu_c$ , comprise entre 0 et 1 (Rocchini et al., 2013b, Tso et Mather, 2009). Une restriction existe toutefois dans l'utilisation de l'appartenance floue pour les 2 dernières familles de classifieurs par le fait que la somme des probabilités a posteriori doit être égale à 1 pour chaque unité de carte.

Comme l'a démontré Zadeh (1965), l'appartenance floue, de par sa nature non-exclusive, apporte une information particulièrement utile pour évaluer l'incertitude et la confusion liées à la variabilité entre les classes (mixité) ou pour indiquer d'éventuelles erreurs de classifications de différentes origines (Rocchini, 2010).

### 1.2 Classification orienté-objet (GEOBIA)

Depuis le début des années 2000, et l'apparition des capteurs produisant des images à Très Haute Résolution Spatiale (THRS), l'analyse d'image par approche orientée-objet, désignée sous les acronymes OBIA ou GEOBIA, s'est considérablement développée (Blaschke, 2010). Blaschke et al. (2014) affirment même qu'elle constitue un nouveau paradigme qui a révolutionné l'analyse d'images de télédétection. Par le fait qu'elle regroupe, de manière automatisée, des pixels voisins ayant des valeurs d'attributs similaires (spectral, textural, formes...), en segment ou objet, elle est en effet incontestablement adaptée aux images THRS pour lesquelles la taille des pixels est inférieure à celle des objets cartographiés (Blaschke et al., 2014). L'approche OBIA peut être décomposée en 2 grandes étapes : la segmentation qui correspond à la partition de l'image en segments et la classification thématique de ces objets, parfois appelée « objectification » (Blaschke et al., 2014). La procédure de classification

peut s'effectuer soit selon l'induction de règles hiérarchiques appliquées aux attributs des objets de l'image et définies selon la connaissance experte de l'utilisateur, soit selon une démarche classique de classification supervisée, pour laquelle certains segments font office de jeu d'apprentissage (Dronova et al., 2012 ; Gosh et Joshi, 2014), ou non supervisée, pour laquelle les objets sont regroupés thématiquement a posteriori, consécutivement à l'application d'algorithmes de regroupement. Depuis peu, il est également à noter que la connaissance experte peut être conceptualisée par l'intermédiaire des ontologies qui facilitent la reproductibilité des classifications (Arvor et al., 2013). Un autre grand intérêt des méthodes GEOBIA réside dans leurs relations avec le monde du SIG, notamment les traitements vectoriels. Les segments sont par définition des polygones caractérisés par une géométrie à laquelle peut être associée une table d'attributs. De plus, les relations topologiques entre les polygones constituent une information essentielle qui peut être employée durant le processus de classification ou en second lieu pour effectuer des corrections post-classifications à l'aide d'outils adaptés (Blaschke et al., 2014).

### **1.3 Evaluation de la précision thématique de la classification**

Toute classification de l'occupation du sol est incomplète si elle ne s'accompagne pas d'une évaluation de sa précision (Tso et Mather, 2009). Dans ce sens, la précision exprime le degré d'exactitude de la classification thématique vis-à-vis de la réalité du paysage et au contraire, l'erreur de classification relève d'un certain écart entre cette classification et la réalité (Foody, 2002). Stehman et Czaplewski (1998) distinguent 3 grandes étapes fondamentales pour l'évaluation de la précision de la classification : le plan d'échantillonnage choisi pour sélectionner les échantillons de référence utilisés lors de la validation, l'élaboration d'un protocole d'attribution d'une étiquette thématique de référence pour chaque unité d'échantillonnage, l'estimation synthétique de la précision par la production d'estimateurs statistiques. Ces mêmes auteurs préconisent d'utiliser un plan d'échantillonnage probabiliste tel qu'un échantillonnage aléatoire, aléatoire stratifié ou systématique pour éviter de biaiser trop fortement l'évaluation de la classification. Une ou plusieurs étiquettes peuvent être attribuée à l'unité d'échantillonnage, qui elle-même, peut

correspondre soit à un pixel, un groupe de pixels, un point ou un polygone. Enfin, l'analyse de la comparaison entre le résultat de classification et les échantillons de référence est effectuée et résumée à travers la traditionnelle matrice de confusion. Cette dernière permet le calcul de différentes métriques indicatrices de la précision. Par exemple, l'indice de précision globale indiquant la proportion d'échantillons tests bien classés ou le coefficient kappa de Cohen corrigeant l'indice de précision globale du biais liant une bonne affectation seulement à l'effet du hasard (Foody, 2002). Cependant, ces mesures présentent certaines limites et peuvent être imprécises si le plan d'échantillonnage n'est pas probabiliste. De plus, elles nécessitent le plus souvent d'avoir accès à des échantillons de validation provenant de sources extérieures à l'image, comme des relevés effectués sur le terrain ou l'accès à d'autres cartes ou images de référence (Foody, 2002 ; Stehman et Czaplewski, 1998, Steele, 2005). Un autre inconvénient de ces types d'indicateurs persiste dans l'absence de représentation spatialisée de la qualité de la classification. Ceci peut être comblé par l'estimation de mesures d'incertitude pour chaque unité de carte. Rocchini et al. (2013b) et Foody (2002) évoquent plusieurs exemples de travaux illustrant cette application. Par extension, il est également judicieux d'obtenir des estimateurs synthétiques prenant en compte la totalité des unités cartographiques (Steele 2005), ce que ne permettent pas les estimateurs classiques (précision globale, Kappa) qui sont par convention restreints aux localités de validation, ou d'apprentissage dans le cas d'une validation interne.

Face à ces limites, Steele et al. (2003) et Steele (2005) proposent une méthode alternative d'estimation de la précision thématique de la classification basée sur l'utilisation du maximum de probabilité a posteriori (MPP) calibré selon différentes stratégies. Ces résultats montrent que la méthode est adaptée pour tout plan d'échantillonnage de validation et qu'elle offre également la possibilité de s'affranchir de l'emploi de ce type d'échantillons externes en utilisant directement, par l'intermédiaire d'une validation croisée, les échantillons d'apprentissage pour estimer la précision de la classification. En prime, les indicateurs de précision sont estimés à partir de la population statistique, c'est-à-dire l'exhaustivité des unités de cartes (pixels ou polygones).

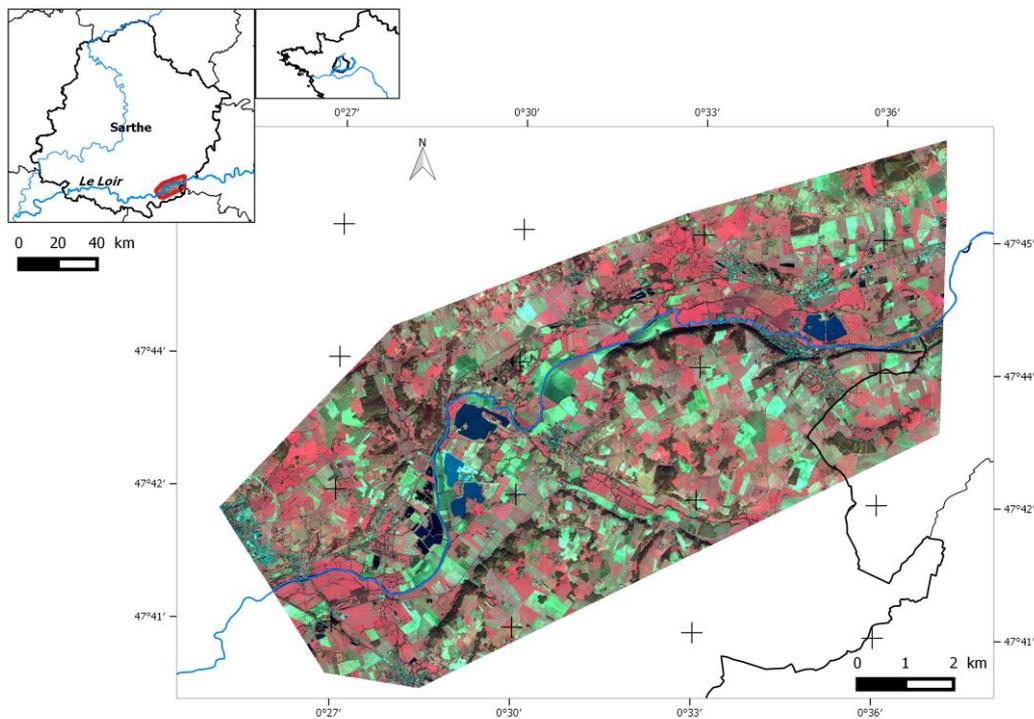


Figure 1 : Localisation de la zone d'étude

## 2. Site et Données

### 2.1 Site d'étude

La zone d'étude (Figure 1) est localisée sur un territoire de fond de vallée de la rivière du Loir, un affluent de la Sarthe et située dans le département du même nom. Ce fond de vallée comprend des formes variées et une occupation du sol contrastée, ce qui en fait un lieu privilégié pour l'analyse de la morphologie fluviale. Plus précisément, le contexte initial qui a motivé l'acquisition de cette image consistait à vérifier comment l'occupation du sol réagit aux microtopographies qui agissent en particulier sur l'état d'humidité des sédiments et la circulation de l'eau en surface.

### 2.2 Image WorldView-2

L'image initiale utilisée pour ce travail est une portion de scène WorldView-2, composée d'une image multispectrale à 1,8 m de résolution spatiale et d'une image panchromatique à 46 cm de résolution spatiale. Ces images ont été acquises le 04/03/2011 à 11h28 GMT. Elles couvrent une superficie de 97 km<sup>2</sup> et sont fournies géoréférencées (traitement ortho-ready de niveau LV2A) dans le système UTM 31 Nord de datum WGS84.

L'image multispectrale est composée de 8 bandes spectrales comprise dans la gamme spectrale de 400-450 nm (Bleu côtier), 450-510 nm (Bleu), 510-580 nm (Vert), 585-625 nm (Jaune), 630-690 (Rouge), 705-745 nm (Red edge), 770-895 nm (Proche Infrarouge 1) et 860-1040 nm (Proche Infrarouge 2).

## 3. Méthodes

L'essentiel de la méthode appliquée dans ce travail est résumé à travers la Figure 2. Les différentes étapes présentées dans ce schéma sont expliquées dans les parties suivantes.

### 3.1 Prétraitement - Calibration des images.

Préalablement à la classification, des étapes de prétraitements sont nécessaires afin d'obtenir des images présentant suffisamment de fidélité radiométrique et de précision géographique. Premièrement, une calibration en valeurs de luminance énergétique (radiance en anglais) a été effectuée avec le logiciel ENVI en appliquant les gains de calibration indiqués dans le fichier de métadonnées. Ceci a pour but de passer des comptes numériques, qui sont des valeurs de luminance relative, à des valeurs de luminance absolue correspondant à une grandeur physique qui est fonction du flux d'énergie émis ou réfléchi en direction du capteur (Tso et Mather, 2009 ; Schowengert, 2007). Cette étape est réalisée pour chacune des images, multispectrale et panchromatique. Les 2 images calibrées sont ensuite fusionnées (Pan-Sharpning) par la méthode de Gram-Schmidt (Laben et Brower, 2000), disponible dans ENVI, pour aboutir à une image multispectrale à 50 cm de résolution spatiale. Selon plusieurs auteurs (Du et al., 2013; Yuhendra et al., 2012) l'algorithme Gram-Schmidt préserve suffisamment l'information radiométrique des bandes multispectrales. L'image fusionnée est ensuite calibrée en valeurs de réflectance au sol par l'intermédiaire d'une correction atmosphérique. La méthode QUAC (Quick Atmospheric Correction,

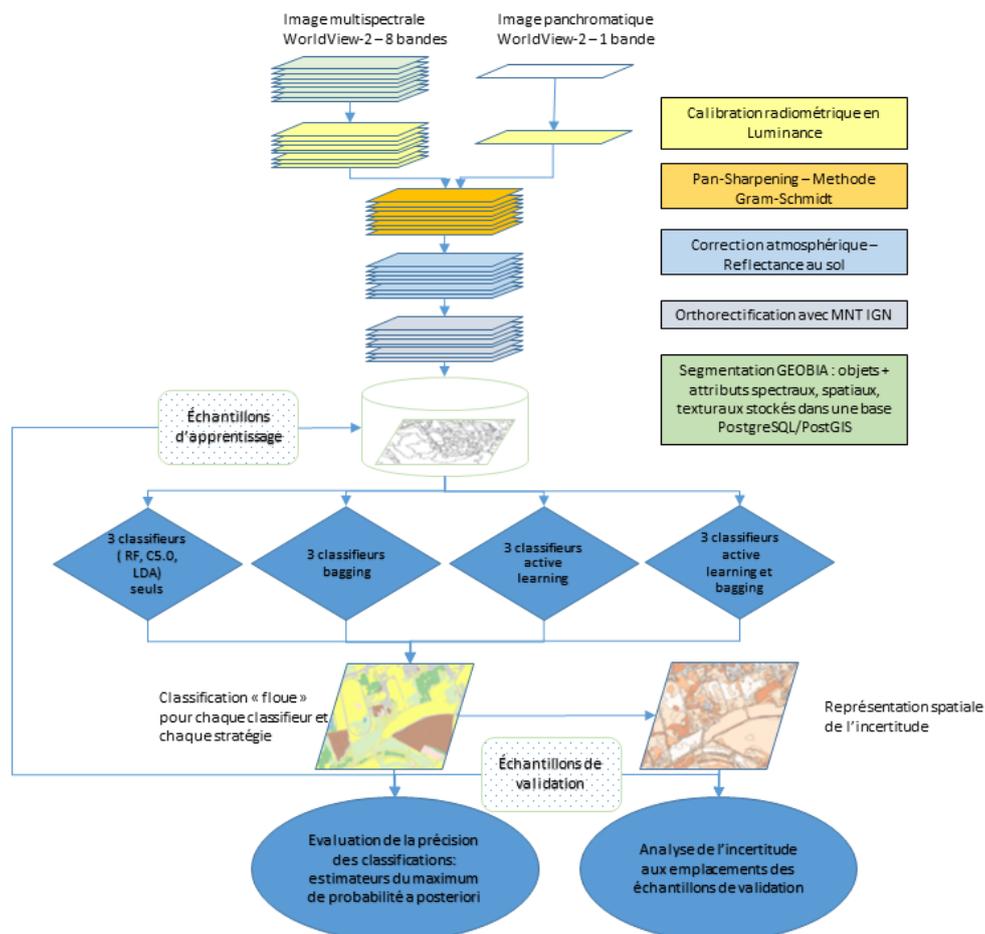
Bernstein et al., 2005) disponible dans le logiciel ENVI a été retenue du fait de la simplicité et la rapidité de sa mise en œuvre et également parce qu'elle est reconnue pour sa précision dans des contextes où l'occupation du sol est suffisamment variée (Bernstein et al., 2005; Zengeya et al., 2012).

La précision géométrique a été améliorée par l'orthorectification de l'image des réflectances au sol. En complément des coefficients polynomiaux propres à la géométrie d'acquisition de l'image, cette étape a

nécessité l'utilisation d'un Modèle Numérique de Terrain au pas de 25m fourni par l'IGN (Institut Géographique National) pour corriger au mieux les effets de la topographie. Le résultat final de ces traitements est une image multispectrale à 50 cm de résolution spatiale, orthorectifiée, calibrée en réflectance au sol et géoréférencée dans le système Lambert 93–RGF93 en utilisant un ré-échantillonnage par la méthode des plus proches voisins.

cultures	terres agricoles cultivées portant de la végétation
eau	rivière, étangs, plans d'eau
feuillus	bois de feuillus
haie, buisson	haies, buissons, ripisylves
peupleraie	plantations de peupliers
prairie	pâturages, pelouses naturelles
résineux	bois de résineux (très peu présent)
sol nu	terres agricoles labourées, non couvertes par la végétation
urbain	bâtiments, parkings, routes,

**Table 1:** Définition des classes d'occupation du sol.



**Figure 2 :** Schéma d'ensemble résumant la méthode

### 3.2 Segmentation orienté-objet (GEOBIA) de l'image

De nombreuses méthodes de segmentation d'images de télédétection sont disponibles dans des logiciels propriétaires comme par exemple Ecognition, ENVI Feature Extraction, Erdas Imagine Feature Analyst ... ou bien en tant que modules de logiciels libres tels que Orfeo Toolbox, SAGA, QGIS via les extensions supplémentaires Orfeo Toolbox et SAGA ...

Dans cet article, la procédure de segmentation a été effectuée sous ENVI 5.0 avec le module Feature Extraction et appliquée sur l'image WorldView-2 à 50 cm de résolution spatiale. Elle débute par une segmentation (Jin, 2012) employant un algorithme de ligne de Partage des Eaux (Roerdink et Meijster, 2001), adaptable selon un paramètre d'échelle, puis se poursuit par un regroupement des segments selon des critères de ressemblance radiométrique et de voisinage. Bien que potentiellement discriminant, aucun indice spectral (NDVI, NDWI, indices de brillance...) n'a été calculé et pris en compte lors de la segmentation. Seules les bandes spectrales de l'image ont été segmentées. Une approche empirique a été employée pour déterminer les paramètres optimaux de segmentation (Duro et al., 2012). Pour chaque objet, des attributs d'ordre spatial, spectral et textural sont estimés. Les attributs spatiaux, au nombre de quatorze, renseignent sur la forme, et la surface des segments. Les attributs spectraux correspondent aux minimum, maximum, moyenne et écart-type des pixels contenus dans chaque objet et calculés pour chaque bande de l'image. 32 attributs spectraux ont été produits. Enfin, les attributs texturaux réfèrent à la détermination d'indices de texture (gamme, moyenne, variance et entropie), pour chaque objet, moyennés selon une taille de fenêtre de voisinage. Appliqués pour chaque bande spectrale, il en résulte 32 attributs texturaux. Au total, 78 attributs sont générés pour 445 029 objets et serviront en tant que variables prédictives lors de l'étape de classification.

### 3.3 Préparation des classifications supervisées.

La couche de polygones résultant du processus de segmentation est importée dans une base de données relationnelle spatiale PostgreSQL/PostGIS (PostgreSQL, 2015). Cette base de données libre permet de stocker des volumes importants, facilite la modification des données par plusieurs utilisateurs simultanément, est plus rapide que les logiciels SIG standards pour effectuer certains traitements (par exemple les requêtes spatiales), et est facilement interopérable avec des logiciels libres de SIG (Quantum GIS) ou de statistiques (R).

532 observations d'apprentissage réparties selon 9 classes d'occupation du sol sont sélectionnées, de manière non probabiliste, par photo-interprétation sur l'image WorldView-2 fusionnée. La définition des classes d'occupation du sol retenues est présentée dans la Table 1.

En vue de l'évaluation de la précision thématique, 150 observations ont également été sélectionnées par un échantillonnage aléatoire simple. Pour chaque observation, l'occupation du sol de référence a été renseignée par photo-interprétation et connaissance du terrain sur l'image WorldView-2 fusionnée.

### 3.4 Classifieurs

#### 3.4.1 Analyse Discriminante Linéaire (LDA)

L'analyse discriminante linéaire est une méthode de classification dont l'objectif est de trouver la combinaison linéaire de variables explicatives permettant de séparer au mieux les classes d'apprentissage. Pour un fonctionnement optimal, elle requiert que la distribution statistique des classes d'apprentissage soit normale, et elle est fondée sur l'hypothèse que les matrices de covariance entre chaque classe sont identiques. Il est à noter que ces 2 conditions sont rarement rencontrées dans la réalité (Steele, 2005). La détermination des probabilités a posteriori pour LDA dépend du théorème de Bayes. Ceci requiert le calcul des probabilités a priori (estimées à partir de la répartition des observations de chaque classe parmi l'ensemble des échantillons d'apprentissage) et de la fonction de densité de probabilité qui suppose que la distribution statistique de chaque groupe est normale et détermine la moyenne de chaque groupe, ainsi que sa matrice de covariance. Ce classifieur paramétrique est utilisé depuis longtemps par la communauté de télédétection, et a fourni des résultats très satisfaisants pour la classification d'espèces, que ce soit lorsqu'il est appliqué à des images THRS WorldView-2 (Immitzer et al., 2012 ; Pu et Landry., 2012) ou bien à des données hyperspectrales (Clark et al., 2005).

#### 3.4.2 Random Forest (Forêt Aléatoire).

Random Forest, appelé Forêt aléatoire en français, est un algorithme de classification et de régression supervisée, issu des techniques d'apprentissage automatique (machine learning). Il repose sur le principe des arbres de décisions auquel a été adjoint une méthode d'ensemble comme le bagging. Schématiquement, la technique du bagging appliquée dans Random Forest consiste à répéter un grand nombre de fois une sélection aléatoire avec remise sur le jeu de données d'apprentissage et à établir des règles de décisions (un arbre) sur chaque échantillon ainsi créé. L'ensemble des arbres (la forêt) est ensuite combiné par l'intermédiaire d'un vote majoritaire aboutissant à la prédiction (Breiman, 2001). Random Forest est un classifieur non paramétrique, c'est à dire ne formulant pas d'hypothèse sur la distribution statistique des données (distribution normale). Pendant la classification, chaque arbre aléatoire définit une probabilité a posteriori selon la répartition d'affectation des classes au niveau des feuilles de l'arbre. La probabilité a posteriori finale correspond à la moyenne

des probabilités de l'ensemble des arbres pris en compte dans la construction de la forêt (Criminisi et al., 2011 ; Loosvelt et al., 2012).

Il est actuellement fréquemment employé pour diverses applications thématiques de la télédétection, ce qui est justifié par la précision de ses performances notamment lorsqu'il est appliqué dans le cadre d'une approche GEOBIA, pour la classification des paysages agricoles (Duro et al., 2012), l'extraction des zones urbaines sur des image THRS (Belgiu et Dragut 2014) ou la discrimination des espèces forestières (Immitzer et al., 2012).

### 3.4.3 Arbre de décision C5.0

L'arbre de décision C5.0 ou See5 (Quinlan 1996), est une amélioration du classifieur C 4.5 développé par John Ross Quinlan (1993). Les principales modifications apportées à ce classifieur supervisé non paramétrique résultent dans la création d'arbres de décision plus petits, une meilleure sélection des attributs les plus significatifs et l'ajout de la technique d'ensemble du boosting lors de la création des règles de partitionnement des données. La méthode de boosting appliquée dans C 5.0, dérivée de l'algorithme Adaboost.M1 (Freund et Schapire, 1996), attribue des poids d'autant plus forts aux échantillons d'apprentissage que ceux-ci sont mal classés lors de l'évaluation de la performance interne du classifieur issue de la construction de chaque arbre. Ces poids sont ajustés de nouveau à chaque itération de construction, puis l'ensemble des arbres est combiné, par un vote pondéré selon le degré d'ajustement, pour aboutir au classifieur final (Quinlan, 1996 ; Tso et Mather, 2009). Les probabilités a posteriori sont générées selon un procédé sensiblement similaire à celui de Random Forest, qui agrège les probabilités créées par l'ensemble des arbres construits pendant le processus de boosting puis les normalise afin que la somme des probabilités d'une observation soit égale à 1.

Des travaux récents ont montré l'intérêt de C5.0 pour des classifications par approche GEOBIA, dans un contexte de cartographie des cultures et des pâturages (Esch et al., 2014), de la cartographie des perturbations de nature industrielle sur l'environnement (Powers et al., 2015) ou dans la mise en place d'un service web de classifications de l'occupation du sol à différentes échelles spatiales (Huth et al., 2012).

Ces 3 modèles de classifieurs ont été implémentés depuis le logiciel R (R Core Team, 2015), en utilisant le package caret (Kuhn et al., 2015) et les packages respectifs aux classifieurs : package randomForest (Liaw et Wiener, 2002) pour Random Forest, package C50 (Kuhn et al., 2014) pour C5.0 et package MASS (Venables et Ripley, 2002) pour LDA.

Le package caret propose un environnement unique pour tester différents algorithmes et optimiser leur paramétrage par l'intermédiaire d'une validation croisée. Une validation croisée de type k-plis (k-fold)

avec 10 échantillons répétés 10 fois, a été appliquée sur le jeu d'apprentissage pour évaluer les paramètres des classifieurs selon leurs performances vis-à-vis de l'indice de précision global et de l'indice kappa.

## 3.5 Stratégies de classification

### 3.5.1 Apprentissage actif

L'apprentissage actif ou active learning (AL) est une procédure d'échantillonnage itérative et interactive appliquée à l'apprentissage automatique supervisé pour laquelle de nouvelles observations préalablement non étiquetées sont sélectionnées et ajoutées au jeu initial afin d'en améliorer la généralisation et la représentativité, et ainsi favoriser l'augmentation de la précision du classifieur (Settles, 2010 ; Tuia et al., 2011b ; Tuia et al., 2011a ; Crawford et al., 2013). Les observations (pixels, polygones) ajoutées sont celles qui sont considérées comme les plus incertaines par le modèle c'est-à-dire celles qui répondent le plus favorablement à un critère d'évaluation statistique, ou heuristique, qui les classera selon leur degré d'incertitude. Les heuristiques les plus communes sont principalement mesurées par l'ambiguïté entre les probabilités a posteriori, la distance vis-à-vis d'une frontière de décision ou le désaccord entre un comité de classifieurs (Stumpf et al., 2013 ; Tuia et al., 2011b ; Rajan et al., 2008). L'apprentissage actif est interactif dans la mesure où les observations classées par l'heuristique sont étiquetées par un utilisateur (désigné également sous le terme d'oracle), généralement un photo-interprète, puis intégrées dans le processus de classification supervisée. Il est également itératif car la sélection des nouvelles observations peut être répétée plusieurs fois jusqu'à ce que les conditions d'arrêt soient satisfaites.

La structure du package « activelearning » (Ramey, 2012) de R a été employée et modifiée pour développer la procédure d'apprentissage actif. 30 nouveaux échantillons, parmi lesquels la moitié a été choisie selon une requête par la méthode du bagging sur le classifieur Random Forest (query by bagging) et l'autre moitié selon une requête par comité des 3 classifieurs (query by committee), ont été étiquetés par photo-interprétation puis ajoutés au jeu d'apprentissage initial. La sélection des échantillons s'est effectuée en mode batch (lot d'échantillons choisis) d'une seule itération. L'heuristique d'incertitude retenue pour la méthode par bagging correspond à l'entropie de Shannon déterminée à partir des probabilités a posteriori, tandis que la divergence de Kullback-Leibler identifie les observations les plus incertaines pour le comité de classifieurs.

### 3.5.2 Bagging

Le bagging ou bootstrap aggregation est une technique d'agrégation aléatoire employée en apprentissage automatique pour améliorer l'ajustement des modèles de classifieurs, tout en évitant le sur-ajustement

(Breiman, 1996 ; Tso et Mather, 2009). Schématiquement, le but est d'entraîner un ensemble de plusieurs classifieurs appartenant à une même famille sur des sous-ensembles d'échantillons d'entraînement construits par la méthode du bootstrap. L'échantillon bootstrap est construit en sélectionnant de manière aléatoire et avec remise les observations d'entraînement à partir du jeu d'apprentissage initial. Le bagging est considéré comme une méthode apportant une amélioration de la précision et de la stabilité, notamment quand elle est appliquée à des classifieurs tels que les arbres de décisions.

Dans ce travail, 25 sous-ensembles (paquets) bootstrap ont été créés à l'aide du package caret du logiciel R, afin d'obtenir les probabilités à posteriori résultant de la moyenne sur ces 25 paquets.

### 3.6 Evaluation de la précision de la classification et mesures d'incertitude

#### 3.6.1 Evaluation de la précision de la classification

Pour chaque classifieur, et pour chaque stratégie de classification, la précision thématique de la classification a été estimée selon 2 approches distinctes. Premièrement, les estimateurs classiques tels que l'indice de précision global et l'indice Kappa ont été générés à partir des matrices de confusion établies soit sur le jeu d'apprentissage par l'intermédiaire d'une validation croisée, soit sur les échantillons de validation, indépendants de l'apprentissage et sélectionnés aléatoirement. Alternativement, mais toujours sur les mêmes échantillons (validation croisée et zones tests indépendantes), les estimateurs basés sur le maximum de probabilité à posteriori (Steele et al., 2003, mais lire plutôt Steele et al., 2005) sont retenus pour évaluer la précision des classifications. Ils reposent sur l'utilisation des probabilités a posteriori telles que décrites dans l'introduction de cet article. 2 indices basés sur la moyenne du maximum de probabilité a posteriori pour l'ensemble de la population des unités de cartes sont proposés. Pour plus de clarté, nous reprenons le développement proposé par Steele (2005) pour expliquer leurs formulations.

L'estimateur du maximum de probabilité a posteriori calibré est défini comme suit :

$$\alpha_M(X) = N^{-1} \sum_{i=1}^N \hat{P}_c[\eta(\mathbf{t}_i) = y_i | \mathbf{t}_i]$$

(1)

Où  $\hat{P}_c[\eta(\mathbf{t}_i) = y_i | \mathbf{t}_i]$  correspond à la prédiction du maximum de probabilité a posteriori calibré pour chaque unité de carte  $\mathbf{t}_i$ .

Cet estimateur peut également être décliné pour chaque classe. Il constitue alors la moyenne du maximum de probabilité a posteriori calibré et calculé pour chaque classe.

$$\alpha_M^g(X) = \frac{N^{-1} \sum_{i=1}^N \hat{P}_c[\eta(\mathbf{t}_i) = y_i] \Psi[\eta(\mathbf{t}_i) = g]}{\sum_{i=1}^N \Psi[\eta(\mathbf{t}_i) = g]}$$

où  $\Psi[\eta(\mathbf{t}_i) = g]$  constitue la réponse binaire (1 ou 0) vis-à-vis de la réalisation de l'événement d'affectation à la classe considérée.

Pour réduire le biais des maximums de probabilité a posteriori, et a fortiori des estimateurs, il est nécessaire d'effectuer une calibration. Cette transformation est réalisée en calculant le coefficient de proportionnalité qui ajustera au mieux la probabilité de bonne ou mauvaise affectation,  $P[\eta(\mathbf{t}) = y | \mathbf{t}]$ , avec le maximum de probabilité estimé  $\hat{P}[\eta(\mathbf{t}) = y | \mathbf{t}]$ , (Steele, 2005). Ceci est formalisé de la manière suivante :

$$P[\eta(\mathbf{t}) = y | \mathbf{t}] = \beta \hat{P}[\eta(\mathbf{t}) = y | \mathbf{t}] \quad (1)$$

Où  $\beta$  est le coefficient de proportionnalité ou paramètre de calibration.

La résolution de cette régression linéaire amène à déterminer l'estimateur des moindres carrés de  $\beta$ ,  $b_0$  tel que  $b_0 = \sum p_i \psi_i / \sum p_i^2$  en choisissant  $p_i = \hat{P}[\eta(\mathbf{t}_i) = y_i | \mathbf{t}_i]$  et  $\psi_i = \Psi[\eta(\mathbf{t}_i) = y_i]$  qui correspond à la comparaison binaire (1 ou 0) entre l'étiquette prédite et celle de validation.

Les propriétés des probabilités a posteriori établissent que les probabilités ne peuvent être négatives et que la somme des probabilités des classes doit être égale à 1. Ceci implique que la valeur minimale du maximum de probabilité pour une classe équivaut à l'inverse du nombre de classe ( $1/c$  ou  $c^{-1}$ ) et contraint le modèle de calibration à avoir pour origine la paire  $(c^{-1}, c^{-1})$ . Il en résulte une nouvelle formulation de l'estimateur des moindres carrés de  $\beta$  tel que  $b = \frac{\sum (p_i - c^{-1})(\psi_i - c^{-1})}{\sum (p_i - c^{-1})^2}$  (2)

Le maximum de probabilité calibré est ainsi déterminé pour chaque unité de carte  $x = (\mathbf{t}, y)$  par la formule suivante :

$$\hat{P}_c[\eta(\mathbf{t}) = y | \mathbf{t}] = \begin{cases} b \hat{P}[\eta(\mathbf{t}) = y | \mathbf{t}] + c^{-1}(1-b) & \text{si } b \hat{P}[\eta(\mathbf{t}) = y | \mathbf{t}] + c^{-1}(1-b) \leq 1 \\ 1 & \text{si } b \hat{P}[\eta(\mathbf{t}) = y | \mathbf{t}] + c^{-1}(1-b) > 1 \end{cases} \quad (3)$$

Les échantillons servant à la calibration proviennent soit de zones de validation, généralement collectés de manière aléatoire, soit en utilisant les échantillons d'apprentissage par l'intermédiaire d'une validation croisée.

#### 3.6.2 Estimation et représentation spatiale de l'incertitude

L'approche floue de classification produit, pour chaque polygone, un vecteur des valeurs d'appartenance (ici, les probabilités a posteriori) (Rocchini et al., 2013b, Foody, 2002). L'information fournie par ces valeurs permet d'accéder à une mesure de l'incertitude présente dans chaque unité de carte, pouvant être représentée spatialement pour produire des cartes d'erreurs ou d'« ignorance » (Rocchini et al., 2013b ; Löw et al., 2015 ; Rocchini, 2010 ; Rocchini et al., 2011). Dans ce travail, les probabilités a posteriori obtenues pour les 9 classes d'occupation du sol sont extraites pour être ordonnées selon les 4 plus grandes valeurs et accompagnées par les étiquettes de classes correspondantes. Ce choix s'explique par la constatation de valeurs de probabilités faibles au-delà

de cette limite, ne reflétant pas une confusion manifeste. Il en résulte, pour chaque segment, la création d'une table attributaire renseignant sur l'incertitude quantitative liée aux valeurs numériques des probabilités, et sur l'incertitude qualitative des étiquettes de classe les plus probables. Ceci constitue des champs facilement interrogeables depuis un outil SIG.

Cependant, cette quantité d'information est difficilement interprétable et il est préférable d'avoir recours à des mesures issues de la théorie de l'information pour la résumer. De nombreuses mesures d'entropie, dérivées de l'entropie de Shannon, existent pour quantifier l'incertitude de la détermination paysagère (Gonçalves et al., 2010) ou la diversité des espèces (Rocchini et al., 2013a ; Rocchini et al., 2015). Dans cette continuité, l'entropie  $\alpha$ -quadratique (Pal et Bedzek, 1994), mise en application dans le domaine de la télédétection par Fauvel et al. (2006), Löw et al. (2013), Löw et al. (2015) et Giacco et al. (2010) a été choisie. Elle est calculée comme suit :

$$H_{\alpha}(x) = \frac{1}{n \cdot (2^{-2\alpha})} * \sum_{i=1}^n p_i^{\alpha}(x) (1 - p_i(x))^{\alpha} \quad (4)$$

Où  $p_i(x)$  est la probabilité d'appartenance à la classe  $i$ ,  $n$  est le nombre de classes (9 classes) et  $\alpha$  est un exposant qui détermine le comportement de la mesure d'incertitude. Plus  $\alpha$  est proche de 1, plus  $H_{\alpha}(x)$  est sensible à la présence de valeurs de probabilités équivalentes parmi les classes, chaque élément  $p_i(x)$  ayant approximativement le même degré d'entropie. Inversement, plus  $\alpha$  est proche de 0, plus la mesure d'entropie est sélective (Löw et al., 2015). Le compromis idéal, appliqué ici, consiste à choisir une valeur  $\alpha$  de 0.5. Si un objet possède un maximum de probabilité convergeant vers une seule classe ( $p_i(x) = 1$ ), alors  $H_{\alpha}(x)$  sera proche de 0. Tandis que si la distribution des probabilités est répartie uniformément selon les classes,  $H_{\alpha}(x)$  tendra vers 1.

L'entropie  $\alpha$ -quadratique est calculée pour chaque classifieur et pour chaque stratégie.

L'incertitude entre les différentes classes peut également être complétée par une mesure de l'incertitude entre les classifieurs. En effet, pour une même localité géographique (unité de carte), les résultats de prédiction peuvent différer entre les classifieurs, et il peut être utile de cartographier ce désaccord. La divergence de Kullback-Leibler (Kullback et Leibler, 1951), qui est une mesure de dissimilarité entre deux distributions de probabilités, a été employée. Elle est utilisée en tant qu'heuristique pour exprimer l'incertitude dans une procédure d'apprentissage actif (Rajan et al., 2008), pour comparer la richesse en biodiversité modélisée par des images de textures produites à partir d'images satellites (Convertino et al., 2012), ou, plus généralement en tant qu'outil pour faciliter l'analyse des données en écologie de la biodiversité (Burnham et Anderson, 2001, Ricotta et Anand, 2006).

Considérant 2 distributions de probabilités discrètes  $P$  et  $Q$ , la divergence de Kullback-Leibler de  $Q$  par rapport à  $P$  est définie par :

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Dans ce travail, le calcul de la divergence de Kullback-Leibler correspond à la moyenne de la divergence entre la distribution de probabilités a posteriori de chaque classifieur (RF, C 5.0, LDA) et la moyenne des distributions de probabilités a posteriori des 3 classifieurs (comité), qui est utilisée comme référence. Bien que cette distance heuristique ne soit pas symétrique, c'est-à-dire que la divergence de  $Q$  par rapport à  $P$  n'est pas égale à la divergence de  $P$  par rapport à  $Q$ , la divergence exprimée ici, n'est calculée que dans un seul sens car elle est jugée suffisante pour estimer l'incertitude.

En définitive, la divergence de Kullback-Leibler est calculée pour chacun des classifieurs, vis-à-vis de la moyenne des probabilités a posteriori des 3 classifieurs, et ces traitements sont répétés pour chaque stratégie de classification. La formulation précise de l'heuristique utilisée dans ce travail (Settles, 2009) est exprimée ici :

$$x_{KL}^* = \operatorname{argmax}_x \frac{1}{C} \sum_{c=1}^C D(P_{\theta^{(c)}} || P_C) \quad (5)$$

Où

$$D(P_{\theta^{(c)}} || P_C) = \sum_i P_{\theta^{(c)}}(y_i|x) \log \frac{P_{\theta^{(c)}}(y_i|x)}{P_C(y_i|x)}$$

Ici  $\theta^{(c)}$  représente un classifieur particulier, et  $C$  représente le comité dans son ensemble.  $P_C(y_i|x) = \frac{1}{C} \sum_{c=1}^C P_{\theta^{(c)}}(y_i|x)$  exprime ainsi la probabilité « consensuelle » que  $y_i$  soit l'étiquette correcte.

L'entropie et la divergence de Kullback-Leibler sont des mesures applicables aux distributions de probabilités a posteriori, c'est-à-dire aux quantités « floues » des prédictions. Il est utile d'obtenir un indice qui s'applique également aux résultats qualitatifs des prédictions de l'occupation du sol. Un indice simple nommé Indice de Concordance (IC), construit de manière empirique, a ainsi été créé pour synthétiser la concordance entre les étiquettes prédites par les différents classifieurs pour chaque unité spatiale, et ce pour les 4 premiers ordres de classes les plus probables. La comparaison thématique est effectuée entre chaque combinaison de classifieur (RF vs. C5.0, RF vs. LDA, C5.0 vs. LDA) puis est traduite par une valeur de 1 si la comparaison est vraie (concordance, même classe prédite) ou par 0 si la comparaison est fautive (désaccord, classes différentes prédites). La somme des 3 comparaisons est ensuite réalisée, ce qui aboutit à 3 valeurs possibles : 0 (aucune concordance), 1 (une combinaison concorde), et 3 (concordance entre les 3 combinaisons, même occupation du sol prédite par les 3 classifieurs). Afin de ne pas accorder la même importance en fonction de l'ordre des probabilités, des poids différents sont appliqués pour chaque rang. Ces poids sont ajustés de manière à ce que la somme

pondérée de la concordance entre les 3 classifieurs d'un même rang reste inférieure à la somme pondérée de la concordance entre 2 classifieurs d'un ordre supérieur. Les poids sont résumés dans la Table 2.

	Label proba.max	Label 2 <sup>de</sup> proba.	Label 3 <sup>ième</sup> proba.	Label 4 <sup>ième</sup> proba.	Indice de concordance
Valeur max. (sans poids)	3	3	3	3	
Valeur intermédiaire (sans poids)	1	1	1	1	
Valeur minimale (sans poids)	0	0	0	0	
Valeur du poids	1	0,25	0,075	0,0225	
Valeur max. (avec poids)	3	0,75	0,225	0,0675	4,0425
Valeur intermédiaire (avec poids)	1	0,25	0,075	0,0225	1,3475
Valeur minimale (avec poids)	0	0	0	0	0

**Table 2 :** Détermination de l'Indice de Concordance

## 4. Résultats

### 4.1 Analyse visuelle des résultats de classification selon le type de classifieurs

Similairement à Duro et al. (2012), une analyse visuelle des résultats de classification a été réalisée. Sur la Figure 3, le territoire affiché sur chaque imagerie est représentatif de la cartographie du paysage sur l'intégralité de la zone d'étude et l'on peut considérer que le comportement individuel de chaque classifieur est, à quelques exceptions près, relativement identique quel que soit la stratégie de classification. L'exemple présenté sur cette figure est la stratégie avec active learning et bagging.

Cette appréciation visuelle permet de mettre en évidence des similitudes mais également quelques différences dans la cartographie de l'occupation du sol produit par chaque algorithme.

D'une manière générale, les classes se répartissent sensiblement aux mêmes lieux pour chaque classifieur, ce qui est particulièrement vérifié pour les classes eau et haie-buissons. Pour les autres groupes, il existe des différences de localisation qui peuvent traduire la confusion entre les classes. A cette époque de l'année (début Mars) et pour cette situation géographique (Ouest de la France), les cultures portant de la végétation sont ainsi difficilement distinguables des prairies quand elles présentent une couverture végétale conséquente, tandis qu'elles pourront être confondues avec le sol labouré nu quand le stade de développement végétal est à son commencement. Globalement, la distinction entre Peupleraie et Feuillus diffère pour LDA, comparé aux résultats obtenus avec RF et C5.0. La connaissance du lieu nous permet d'affirmer que dans le cas des peupleraies, LDA a tendance à se tromper. Au contraire, concernant l'environnement urbain, il apparaît que LDA a tendance à mieux discerner les surfaces imperméables

(bâtiments et routes) des jardins ou espaces verts environnants, que ne le font RF et C5.0 qui regroupent souvent l'ensemble au sein de la même classe urbaine.

### 4.2 Evaluation de la précision de classification

La Table 3 résume l'évaluation statistique globale des classifications en fournissant différents estimateurs calculés pour les 4 stratégies. Il s'agit premièrement de l'indice de précision global ( $PG_{CV}$ ) et du Kappa ( $Kappa_{CV}$ ) obtenus par validation croisée à partir des échantillons d'apprentissage, qui constituent des estimateurs positivement biaisés au sens de Steele (2005) car appliqués sur des observations non échantillonnées selon un protocole probabiliste et sélectionnées préférentiellement grâce à leur facilité d'identification (Stehman et Czaplewski, 1998). Puis des indices de précision globale ( $PG_{VAL}$ ) et Kappa ( $Kappa_{VAL}$ ) obtenus à partir d'un protocole probabiliste et indépendant de l'apprentissage, et de fait considérés comme non biaisés. Ensuite les coefficients de calibration  $\beta_{CV}$  et  $\beta_{VAL}$  sont déterminés respectivement à partir du jeu d'apprentissage et du jeu de validation indépendant. Ils sont accompagnés par les estimateurs du maximum de probabilité a posteriori calibrés selon la validation croisée sur les échantillons d'apprentissage ( $\alpha_{M_{CV}}$ ) et les références de validation ( $\alpha_{M_{VAL}}$ ). Enfin, les estimateurs de biais,  $bias_{CV}$  et  $bias_{VAL}$ , correspondant à la différence entre un estimateur du maximum de probabilité à posteriori et l'indice de précision calculé sur le jeu de validation :  $bias_{CV} = (\alpha_{M_{CV}} - PG_{VAL}) \times 100$ ,  $bias_{VAL} = (\alpha_{M_{VAL}} - PG_{VAL}) \times 100$ .

L'analyse de ces résultats est effectuée en examinant, conjointement, les variations des estimateurs de précisions selon le type de classifieur et les différentes stratégies de classification, ainsi que l'évolution de leur biais selon l'échantillonnage choisi pour les générer.

Du point de vue général, les indices de précision globale et Kappa déterminés par validation croisée présentent des valeurs élevées (entre 0,86 et 0,96) quel que soit le classifieur ou la stratégie. Ceci semble logique compte tenu du biais dans la sélection des observations d'apprentissage, et est d'autant plus confirmé par la légère baisse de ces indices lors de l'emploi de l'apprentissage actif qui incorpore des objets plus incertains dans l'entraînement du modèle.

La valeur du coefficient de calibration  $\beta$ , approximé par l'estimateur contraint des moindres carrés b, présente une faible décroissance selon l'ordre établi des classifieurs (RF > C5.0 > LDA). Ce comportement est observé quelque-soit le type d'échantillonnage. Il est ainsi mis en évidence que RF présente la meilleure adéquation entre le maximum de probabilité, avec l'étiquette de classe qui en résulte, et le label de l'échantillon de référence.

La stratégie qui offre les meilleurs résultats concernant les valeurs de  $\beta_{CV}$  est le bagging, excepté pour le classifieur C5.0 pour lequel la stratégie active learning

suivie du bagging arrive en tête. Même si la différence reste relativement marginale avec les autres stratégies, on peut considérer que l'incorporation de la diversité aléatoire dans la sélection de l'apprentissage par bagging, a permis une meilleure capacité d'affinité entre les probabilités a posteriori et l'évènement de comparaison des étiquettes.

Concernant  $\beta_{VAL}$ , les valeurs les plus élevées sont obtenues avec la stratégie d'active learning. L'ajout d'observations plus incertaines, ou tout simplement la prise en compte d'un nombre un peu plus important d'échantillons d'apprentissage, a permis une meilleure généralisation de la prédiction des probabilités a posteriori facilitant la concordance avec les zones tests sélectionnées aléatoirement. Il est cependant étonnant de constater que la stratégie active learning + bagging fournit la plus faible valeur de calibration pour le classifieur Random Forest (une analyse opposée est observé pour LDA). Ceci peut éventuellement supposer que l'application du bagging sur un modèle Random Forest entraîné avec des observations plus incertaines est redondante, car ce classifieur repose par définition sur une approche de bagging, et que ceci contribue à faire diminuer l'adéquation de calibration. La comparaison entre les deux coefficients  $\beta_{CV}$  et  $\beta_{VAL}$  met en évidence un biais dépendant de la nature de l'échantillonnage, qui se révèle différemment selon le type de classifieur et la stratégie de prédiction. Ainsi, RF apparaît être moins sensible au biais de calibration (coefficient de même ordre de grandeur pour les 2 approches) que ne le sont C5.0 et surtout LDA. L'emploi des classifieurs seuls produit le plus faible biais de calibration pour Random Forest tandis que la stratégie d'active learning tend à diminuer le biais de cet estimateur pour les 2 autres classifieurs.

Enfin, l'analyse de biais<sub>CV</sub> et biais<sub>VAL</sub> nous permet de dégager certaines tendances sur le comportement prédictif des classifieurs et leur sensibilité aux stratégies.

Pour l'approche de calibration par validation croisée, les stratégies de classification affichant le plus faible biais sont celles incorporant une procédure d'apprentissage actif. Il s'agit de l'active learning + bagging pour le classifieur RF et LDA, de l'active learning pour C5.0. Dans le cas de la calibration par le jeu de validation indépendant, les stratégies affichant le moins de biais sont respectivement le bagging pour Random Forest, le classifieur seul pour C5.0 et l'active learning pour LDA. LDA est pratiquement insensible au biais quand l'estimateur MPP est calibré à partir des observations de validation. Au contraire, il s'agit du classifieur le plus fortement biaisé quand l'évaluation est effectuée par validation croisée. Ceci démontre sa forte dépendance vis-à-vis de la sélection des observations d'apprentissage. A l'opposé, son comportement est peu affecté par le choix de la stratégie de classification car ses résultats sont dans le même ordre de grandeur pour chaque stratégie. Ceci n'est pas observé pour C5.0 et RF qui présentent le biais de validation le plus

important pour les 2 stratégies employant l'apprentissage actif.

Ce biais observé est négatif (Table 3) indiquant que les observations d'apprentissage incertaines issues de l'active learning, de par leur procédure de sélection, tendent à biaiser négativement l'estimateur MPP quand celui-ci est déterminé de manière indépendante. Ceci corrobore les résultats obtenus par Steele (2005) et rejoint la conclusion de l'analyse du coefficient  $\beta$  montrant que l'apprentissage actif induit un maximum de probabilité plus faible (plus incertain) qui diminue la force de l'adéquation de calibration et donc la valeur de l'estimateur de précision. Généralement, RF présente la plus faible différence entre biais<sub>CV</sub> et biais<sub>VAL</sub>, ce qui démontre sa robustesse face à des jeux d'apprentissage de taille différente et plus ou moins biaisés (Mellor et al., 2015). C5.0 occupe la position intermédiaire concernant ce paramètre tandis que LDA termine la marche, en répondant de manière fort différente selon le type de sélection des zones d'entraînement.

Enfin, un dernier point à relever concerne la proximité entre les valeurs de  $\alpha_{MVAL}$  et celles du coefficient Kappa de validation pour les stratégies employant l'active learning et appliqués aux classifieurs Random Forest et C5.0. Ceci ne se retrouve pas pour les autres stratégies, ni pour le classifieur LDA.

#### 4.3 Analyse de l'incertitude aux emplacements de validation

L'analyse de l'incertitude consiste en une confrontation entre les étiquettes de classes de validation et les différents indices d'incertitude (entropie  $\alpha$ -quadratique, indice de Kullback-Leibler, indice de concordance qualitatif). Cette confrontation est organisée et analysée selon le degré de concordance entre les prédictions des 3 classifieurs et le label de validation. Par exemple, une concordance de validation de 3 signifie que pour la localisation de l'observation de validation, les 3 prédictions de l'étiquette de probabilité maximale des classifieurs correspondent chacune à l'étiquette de classe de la zone test. La concordance de validation diminue selon le nombre de classifieurs correspondant à l'étiquette de validation (3, 2, 1, 0). Pour chaque degré de concordance, la moyenne des indicateurs d'incertitude est calculée. Les résultats sont présentés dans la Table 4. La Figure 4 montre un exemple de la représentation spatialisée des indicateurs d'incertitude. Pour justifier de la signification statistique des comparaisons entre moyennes, des tests statistiques ont été effectués. Ils diffèrent selon la distribution statistique des données. Dans le cas d'une distribution normale, le test de student a été retenu, tandis qu'une absence de normalité entraîne l'emploi du test de la somme des rangs de Wilcoxon. Un exemple est présenté Table 5.

#### 4.3.1 Variation de l'entropie $\alpha$ -quadratique

Naturellement, l'entropie  $\alpha$  quadratique diffère selon le type de classifieur, ceci du fait de leurs propriétés différentes dans l'établissement des probabilités a posteriori (Niculescu-Mizil et Caruana, 2005). Cependant, à première vue, des tendances générales apparaissent, quel que soit le modèle choisi, montrant une différenciation selon le degré de concordance et la stratégie adoptée. En effet, pour les stratégies employant les classifieurs seuls et le bagging, l'incertitude associée à l'entropie varie en fonction de la concordance de validation selon l'ordre suivant : l'incertitude est la plus faible pour une concordance de 3, puis elle croît pour une concordance de 1, puis de 2 et devient la plus élevée quand aucune des prédictions ne correspond à la classe test. Globalement, une plus grande variation peut être observée pour les 2 stratégies employant l'apprentissage actif. La concordance de 3 présente toujours l'entropie la plus basse mais la plus forte incertitude n'est plus observée pour la concordance de 0 mais plutôt pour celle de 1. Ce comportement est toutefois seulement constaté pour RF et C5.0, LDA conservant, généralement, le même ordre d'entropie pour toutes les stratégies. L'intensité des valeurs d'entropie est également révélatrice de variations selon la stratégie de classification. Prenons l'exemple de la comparaison entre la stratégie du bagging et celle de l'active learning et bagging. Pour les classifieurs C5.0 et LDA, cette dernière a globalement augmenté l'incertitude des catégories dont la concordance est justement considérée comme incertaine (concordance de 1 ou 2), alors que pour RF, l'effet inverse se produit, cette stratégie caractérisant une diminution de l'entropie par rapport à l'approche appliquant uniquement le bagging. Néanmoins, l'interprétation de l'entropie  $\alpha$  quadratique selon le degré de concordance reste assez instable pour certaines stratégies dans la mesure où la comparaison des moyennes entre la catégorie 1 et 2 n'est pas toujours catégoriquement significative (valeur  $p > 0.05$ ). Il faut toutefois préciser que cet effet n'est pas remarqué pour les concordances de 3.

#### 4.3.2 Variation de la divergence de Kullback-Leibler.

L'analyse de la variation de la divergence de Kullback-Leibler selon le rang de concordance indique une évolution logique (Table 4) corroboré à une réelle signification statistique (Table 5). Quelle que soit la stratégie de classification, la divergence la plus faible est observée lorsque les 3 classifieurs ont prédit l'étiquette de classe correcte. A l'opposé, excepté la stratégie apprentissage actif et bagging, une concordance de 1 (1 seul classifieur sur 3 a prédit la bonne classe) est associée à la plus grande divergence. Dans la même continuité, la concordance de 2 présente également des valeurs relativement élevées, mettant en évidence la variabilité des distributions de probabilités entre les classifieurs pour ces unités où les prédictions s'accordent difficilement

avec la vérité terrain. Dans le cas où l'ensemble des modèles a prédit la mauvaise classe (concordance de 0), le désaccord est légèrement moins prononcé que pour les concordances de 1 ou 2, ce qui indiquerait une certaine régularité des classifieurs à se tromper pour les mêmes types d'objets.

#### 4.3.3 Variation de l'indice de concordance qualitatif.

Du point de vue mathématique, la formulation empirique et relativement approximative de cet indice ne permet pas une analyse détaillée des valeurs qu'il peut prendre ainsi qu'une comparaison objective selon les stratégies mais, néanmoins, il reflète bien la capacité de doute ou d'affirmation des classifieurs vis-à-vis des observations. Ainsi, la confrontation avec la concordance de validation nous permet de vérifier si les classifieurs ont tendance à être plutôt appropriés et consensuels, prudents voire contradictoires, ou présomptueux voire trompeurs.

Les qualificatifs d'appropriés et de consensuels sont évidents lorsque l'indice de concordance est confronté à la concordance de validation de 3. En effet, ce niveau indique que les 3 classifieurs ont prédit la bonne classe, et que la valeur minimale prise par l'indice ne peut être que 3. Quand seulement 2 classifieurs sur 3 ont prédit la bonne classe, l'indice de concordance, qui ne peut ici être inférieur à 1 et supérieur ou égal à 3, marque l'opposition d'un classifieur par rapport aux deux autres et dans ce cas, ce désaccord traduit son inexactitude à l'égard de la zone test. En revanche, pour une concordance de validation de 1, un seul classifieur, prudent mais inspiré, prédit la juste étiquette. A de très rares exceptions près, son indice de concordance marque la dissemblance entre chaque classifieur (indice inférieur à 1) mais dans la majorité des cas, il est supérieur à 1 exprimant la clairvoyance d'un seul algorithme contre les 2 autres. L'indice de concordance peut également manifester le caractère « présomptueux » des classifieurs dans le cas où 2 de ceux-ci s'accordent sur l'étiquette de prédiction bien qu'aucun d'entre eux ne soit fidèle à la réalité du paysage (concordance de validation de 0). L'indice de concordance peut même être carrément trompeur dans la mesure où les 3 classifieurs concordent sur le label de classe (indice supérieur ou égal à 3), quand bien même aucun d'entre eux ne prédit la bonne classe. Ce dernier point insiste sur l'erreur systématique qui peut être associée à certaines unités cartographiques, pour lesquelles le risque d'erreur est constant du fait de la confusion intrinsèque à l'objet ou bien parce que les échantillons d'apprentissage ne sont pas suffisamment représentatifs de ce type d'objet.

## 5. Discussion

### 5.1 Estimateurs du maximum de probabilité à posteriori

Plusieurs auteurs (Foody, 2002 ; Comber et al., 2012) ont déjà alerté sur l'insuffisance de la seule prise en

compte des indicateurs de la matrice de confusion, qui au demeurant n'est pas indicatrice de la distribution spatiale des erreurs. Il est d'ailleurs pratiquement impossible de fournir une mesure unique de la précision qui soit adaptée à toute configuration de classification (Foody, 2002). Les résultats présentés dans ce travail, confirment qu'il est judicieux de combiner plusieurs estimateurs de nature différente pour obtenir la meilleure évaluation globale. Il a également été vérifié que les estimateurs proposés par Steele sont relativement peu biaisés et ce, quel que soit la stratégie d'échantillonnage (Steele, 2005 ; Rocchini et al., 2013b). Étonnamment, cette approche est très peu utilisée par la communauté scientifique alors qu'elle présente des avantages indéniables, notamment quand elle est analysée conjointement avec un indicateur issue d'une validation externe probabiliste. En effet, au-delà de la simple comparaison avec le jeu de données de référence, elle permet à la fois d'estimer le biais inhérent à l'échantillonnage mais renseigne également sur la capacité du classifieur à fournir des probabilités a posteriori représentatives de la confusion inter-classes pour chaque observation. Ainsi, la relative constance des coefficients de calibration  $\beta_{CV}$  et  $\beta_{VAL}$  pour les classifieurs RF et C5.0 indique que leurs probabilités a posteriori présentent un certain degré de confiance et de robustesse. Pour LDA, la variabilité de ces coefficients peut signaler une imprécision dans l'estimation des probabilités a posteriori. L'examen de ces dernières a montré qu'elles avaient tendance à se rapprocher de 1, comme Niculescu-Mizil et Caruana (2005) le soulignent de manière générale pour les classifieurs utilisant les règles de Bayes. Ce surajustement associé à la prise en compte de la distribution a priori implique ainsi un biais positif pour le plan d'échantillonnage d'apprentissage (observé pour les valeurs de  $\beta_{CV}$  et  $\alpha_{M CV}$ ), qui n'est corrigé que par l'emploi d'une stratégie plus probabiliste tel que peut l'être l'échantillonnage aléatoire simple (Stehman et Czaplewski, 1998).

D'un autre côté, on peut s'interroger sur le bénéfice apporté par l'apprentissage actif dans l'évaluation de la précision de la classification à l'aide des estimateurs du maximum de probabilité a posteriori. Apporte-t-il plus de « bruit » ou sert-il à mieux ajuster la détermination des probabilités a posteriori et donc l'incertitude entre les classes ? A première vue, les résultats obtenus pour la stratégie d'active learning appliquée aux classifieurs RF et C5.0, aurait tendance à confirmer la seconde hypothèse, montrant une meilleure capacité de généralisation ( $PG_{VAL}$  élevé) alliée à une plus grande considération du caractère flou des unités spatiales ( $\alpha_{M CV}$  et  $\alpha_{M VAL}$ ).

Cependant, la pertinence des observations sélectionnées par les heuristiques d'apprentissage actif utilisées dans ce travail est discutable. L'expérience nous a montré qu'un petit nombre d'observations restent assez difficiles à étiqueter objectivement par photo-interprétation. Tuia et Muñoz-Marí (2012), proposent une solution pour ce problème, en appliquant

une procédure active qui filtre les observations incertaines et propose seulement à l'utilisateur celles qui sont les plus fiables à labelliser.

La méthode pour estimer les indicateurs du MPP pourrait également être approfondie, notamment par l'utilisation d'approches alternatives pour déterminer le coefficient de calibration par régression. Steele (2005) avait déjà évoqué une possible amélioration en proposant de calculer ces coefficients à partir de régressions non linéaires, ou variant selon des groupes spécifiques. Dans cette continuité, la calibration par une régression géographiquement pondérée, (Comber et al., 2012), fluctuant selon l'emplacement géographique et plus à même de prendre en compte l'autocorrélation spatiale, pourrait être testée.

## 5.2 Indicateurs d'incertitude

Complémentairement aux estimateurs synthétiques, ce travail a montré que les indices d'incertitude apportent un bénéfice d'informations avec, dans l'ensemble, des valeurs qui sont représentatives de la précision ou de la confusion observées aux emplacements de validation. Le stockage de ces indices, sous forme d'attributs des couches SIG de la base PostgreSQL/PostGIS (PostgreSQL, 2015), associés aux rangs des valeurs et labels de probabilité a posteriori, offre ainsi la possibilité d'identifier plus efficacement les polygones difficiles à classer. Il est également à noter que la méthode décrite ici n'est pas uniquement applicable à une démarche GEOBIA, elle peut être utilisée pour une approche pixel. Dans ce cas, les propriétés d'incertitudes seront représentées sous forme d'une pile de bandes. La liaison avec le monde du SIG est cependant facilitée avec la méthode GEOBIA dans la mesure où les relations spatiales entre entités voisines sont plus explicites (Blaschke et al., 2014). Par prolongement, ceci justifie l'intégration des segments dans une base PostgreSQL/PostGIS (Argyridis et Argialas, 2015). En effet, les outils topologiques disponibles dans ce SGBD pourraient permettre d'effectuer des traitements post-classifications pour corriger des erreurs de cartographie de l'occupation du sol ou pour mieux mettre en évidence les connexions écologiques du paysage.

En dernier lieu, la similarité thématique des différentes classes constitue une information qui peut également être prise en considération pour évaluer l'incertitude. Dans la même logique d'idée que l'indice de similarité thématique proposé par Costa et al. (2015), on peut imaginer une pondération des classes selon un degré de ressemblance thématique préalablement défini et contextualisé par l'utilisateur de la carte. Au lieu d'être exprimé selon la proportion surfacique des classes à l'intérieur de l'objet (Costa et al., 2015) qui nécessite une classification de l'occupation du sol de référence, elle pourrait être définie selon l'ordre des étiquettes de classes fournies par les probabilités a posteriori. Par exemple, le trio haie-buisson – feuillus – peupleraie (pour les 3 premiers rangs) aurait une pondération de similarité plus élevée que celle caractérisée par la suite

feuillus – sol nu – urbain. Cette approche prendrait encore plus d'intérêt si le protocole de labellisation exprimait lui aussi une information sur la classe d'occupation du sol principale mais également sur les classes subalternes présentes dans l'unité spatiale d'échantillonnage (Stehman et Czaplewski, 1998).

### 5.3 Perspectives générales

En définitive, dans une perspective d'amélioration complète de la classification/prédiction, pour que celle-ci se conforme au mieux à la réalité paysagère tout en respectant fidèlement l'incertitude qui peut résulter de cette démarche, de nombreuses améliorations omises dans ce travail peuvent être envisagées. Tout d'abord, la segmentation, essentielle dans l'approche GEOBIA, peut être optimisée afin d'épouser au mieux la forme des entités constituants le milieu d'analyse (Drăgut et al., 2014). Ceci peut supposer que la précision et l'incertitude géométrique des objets soient évaluées. Une procédure de sélection des variables les plus pertinentes peut également être réalisée (Li et al., 2016 ; Gosh et Joshi, 2014). Elle réduit la dimensionnalité des attributs tout en évitant la prise en compte des données trop bruitées ou redondantes. Par la suite, la fusion des classifieurs pourrait constituer un perfectionnement futur, tenant compte de l'éventuelle complémentarité des algorithmes. Par exemple, la fusion peut s'effectuer selon une méthode supervisée employant un méta-classifieur (Clinton et al., 2015), ou par une démarche non supervisée pour laquelle la décision de fusion s'opère à l'aide de la théorie des ensembles flous (Fauvel et al., 2006) ou grâce à la théorie de l'évidence de Dempster-Shafer (Ouerghemmi et al., 2017) associant en prime l'estimation de l'incertitude. D'un autre côté, la connaissance experte découlant du processus de classification peut être mieux formalisée par l'intermédiaire des ontologies GEOBIA floues (Argyridis et Argialas, 2015). Ce type de développement incorporant les relations spatiales floues entre segments tend à mieux relier la connaissance symbolique « experte » à l'information numérique fournie par l'image. Finalement, un raisonnement analogue pourrait être étendu aux traitements post-classification pour corriger les erreurs de classification.

## 6. Conclusion

Ce travail présente la classification par approche GEOBIA d'une image WorldView-2, selon 3 classifieurs, Random Forest, arbre de décision C5.0 boosté et Analyse Discriminante Linéaire (LDA) et selon 4 stratégies de classification supervisée : classifieurs seuls, classifieurs avec procédure de bagging, classifieurs avec procédure d'apprentissage actif et classifieurs avec apprentissage actif et bagging. L'accent est mis ici sur l'évaluation de la précision de la classification à l'aide des estimateurs du maximum de probabilité a posteriori (Steele, 2005) pour laquelle il a

été montré un faible biais quand les estimateurs étaient calibrés à partir de zones de validation sélectionnées de manière probabiliste. Pour ce type d'échantillonnage et d'estimateurs, les classifieurs RF et C5.0 sont néanmoins biaisés vis-à-vis de l'indice de précision globale (mais très peu vis-à-vis de l'indice Kappa) quand une stratégie incorporant l'apprentissage actif est appliquée. Ce comportement n'est pas observé pour LDA. Toutefois, LDA présente la plus forte sensibilité au type d'échantillonnage choisi pour la calibration et sa précision est particulièrement surestimée quand il est entraîné à partir d'observations non sélectionnées aléatoirement. Dans un second temps, l'analyse des indicateurs d'incertitude, quantitatifs comme l'entropie  $\alpha$  quadratique ou la divergence de Kullback-Leibler déterminés à partir des distributions de probabilité a posteriori, ou qualitatifs tel que l'indice de concordance empirique, a confirmé leurs correspondances avec la précision ou la confusion observées aux emplacements de validation. Ceci justifie pleinement l'emploi de ces représentations spatialisées pour compléter l'évaluation de la qualité des classifications de l'occupation du sol.

## 7. Remerciements

Les auteurs tiennent à remercier Patrick Messner pour ses conseils sur l'utilisation de PostgreSQL/PostGIS et Patricia Messner pour les corrections apportées à la rédaction de ce document. Ils remercient également les deux relecteurs de la revue RFPT pour leurs suggestions améliorant la qualité de cet article.

	$PG_{CV}$	$Kappa_{CV}$	$\beta_{CV}$	$\beta_{VAL}$	$\alpha_{M CV}$	$\alpha_{M VAL}$	$PG_{VAL}$	$Kappa_{VAL}$	$biais_{VAL}$	$biais_{CV}$
Random Forest	0,96	0,93	1,03	1,02	0,71	0,71	0,69	0,61	2,20	2,82
C5.0	0,96	0,93	1,00	0,86	0,76	0,65	0,67	0,58	-1,36	9,69
LDA	0,94	0,90	0,96	0,72	0,92	0,69	0,67	0,58	1,84	25,10
Random Forest bagging	0,96	0,93	1,08	1,04	0,73	0,72	0,70	0,62	1,61	3,22
C 5.0 bagging	0,95	0,92	1,05	0,86	0,78	0,65	0,67	0,58	-1,59	11,73
LDA bagging	0,94	0,90	0,97	0,73	0,92	0,69	0,67	0,59	1,74	24,58
Random Forest AL	0,93	0,89	1,03	1,07	0,63	0,65	0,71	0,63	-6,08	-7,86
C5.0 AL	0,93	0,88	1,00	0,91	0,67	0,61	0,67	0,60	-6,31	-0,66
LDA AL	0,91	0,87	0,95	0,76	0,89	0,71	0,69	0,61	1,17	19,51
Random Forest AL + bagging	0,93	0,89	1,03	0,92	0,66	0,59	0,67	0,60	-7,90	-0,99
C5.0 AL + bagging	0,93	0,88	1,05	0,91	0,69	0,60	0,67	0,59	-6,46	2,16
LDA AL + bagging	0,91	0,87	0,95	0,76	0,89	0,70	0,69	0,60	1,80	20,09

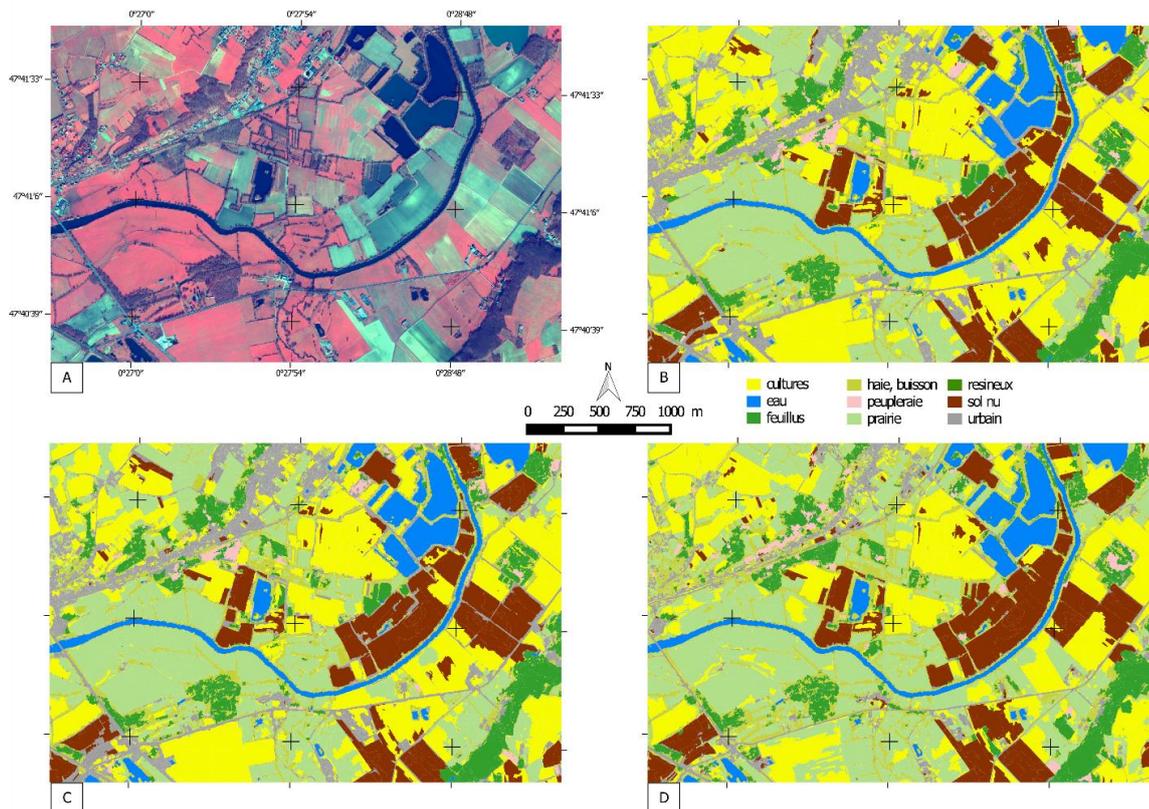
**Table 3 :** Résultats des estimateurs de précision selon les différentes stratégies de classification

	Indicateurs d'incertitudes	concordance de validation			
		3	2	1	0
classifieurs seuls	moy. $H_{\alpha}$ RF	0,29	0,38	0,37	0,41
	moy. $H_{\alpha}$ C.5.0	0,17	0,24	0,19	0,26
	moy. $H_{\alpha}$ LDA	0,02	0,09	0,07	0,11
	moy. $x_{kL}^*$	0,10	0,28	0,37	0,25
	moy. IC	3,36	1,26	1,11	2,46
classifieurs avec bagging	moy. $H_{\alpha}$ RF	0,30	0,38	0,37	0,42
	moy. $H_{\alpha}$ C.5.0	0,19	0,27	0,19	0,29
	moy. $H_{\alpha}$ LDA	0,02	0,06	0,06	0,10
	moy. $x_{kL}^*$	0,16	0,37	0,44	0,32
	moy. IC	3,33	1,24	1,12	2,17
classifieurs avec AL	moy. $H_{\alpha}$ RF	0,31	0,42	0,42	0,41
	moy. $H_{\alpha}$ C.5.0	0,19	0,26	0,33	0,27
	moy. $H_{\alpha}$ LDA	0,03	0,10	0,06	0,14
	moy. $x_{kL}^*$	0,13	0,33	0,38	0,28
	moy. IC	3,34	1,25	1,26	2,74
classifieurs avec AL et bagging	moy. $H_{\alpha}$ RF	0,25	0,36	0,36	0,37
	moy. $H_{\alpha}$ C.5.0	0,21	0,29	0,33	0,28
	moy. $H_{\alpha}$ LDA	0,04	0,13	0,08	0,15
	moy. $x_{kL}^*$	0,12	0,35	0,34	0,24
	moy. IC	3,36	1,23	1,29	2,69

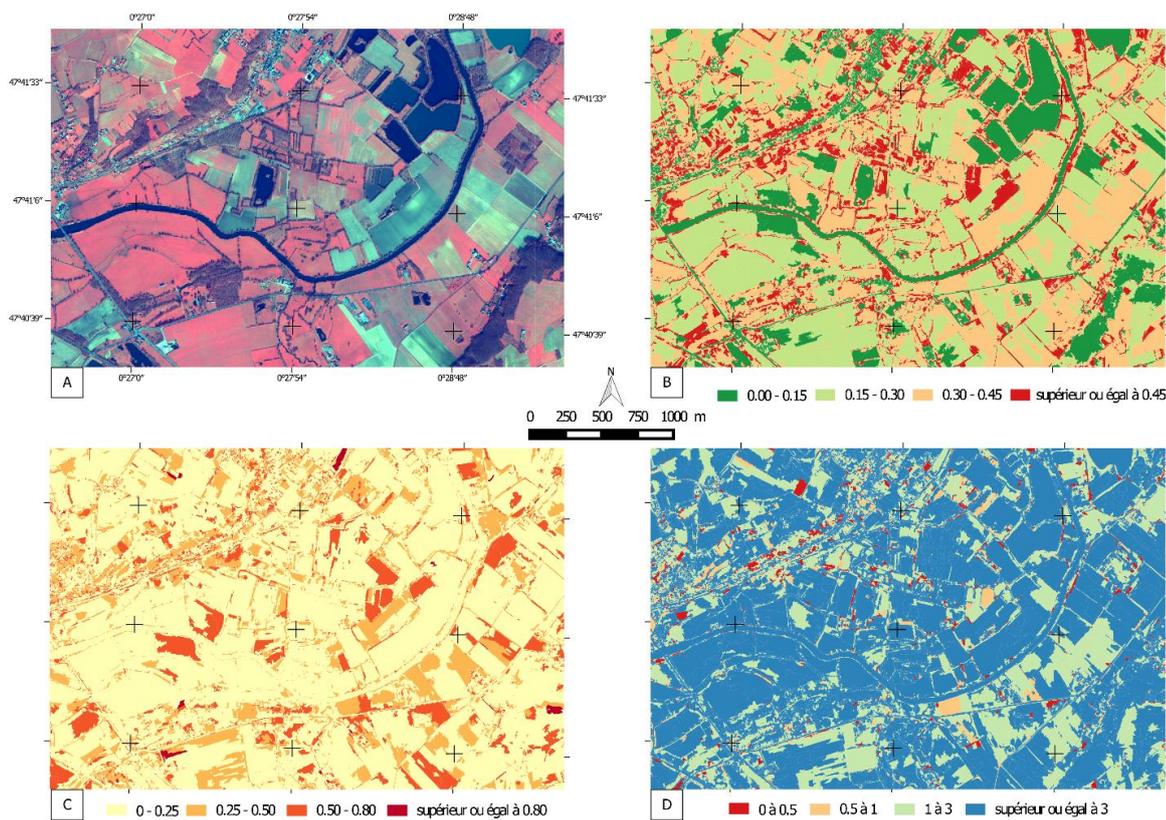
**Table 4 :** Confrontation des indicateurs d'incertitude avec la concordance de validation. Les valeurs des indicateurs d'incertitude résultent de la moyenne de ces indicateurs aux emplacements des observations de validation.

Valeur de la concordance	0	1	2
1	test de la somme des rangs de Wilcoxon W = 151, p-value = 0.04893 hypothèse alternative : la vraie différence entre les échantillons n'est pas égale à 0		
2	test de la somme des rangs de Wilcoxon W = 173, p-value = 0.03959 hypothèse alternative : la vraie différence entre les échantillons n'est pas égale à 0	t-test de Student données: x et y t = -0.17819, df = 28, p-value = 0.8599 hypothèse alternative: la vraie différence entre les moyennes n'est pas égale à 0. Intervalle de confiance à 95 %: -0.1461774 0.1227810 Estimations des échantillons: moyenne de x moyenne de y 0.3366011 0.3482993	
3	test de la somme des rangs de Wilcoxon W = 2184, p-value = 2.647e-05 hypothèse alternative : la vraie différence entre les échantillons n'est pas égale à 0	test de la somme des rangs de Wilcoxon W = 1038, p-value = 1.517e-05 hypothèse alternative : la vraie différence entre les échantillons n'est pas égale à 0	test de la somme des rangs de Wilcoxon W = 1224, p-value = 8.33e-07 hypothèse alternative : la vraie différence entre les échantillons n'est pas égale à 0

**Table 5 :** Exemple de tests statistiques de comparaison des moyennes aux emplacements de validation. Le tableau présenté ici correspond à la divergence de Kullback-Leibler pour la stratégie apprentissage actif et bagging. La valeur des moyennes est donnée dans le tableau 4.



**Figure 3 :** Résultat de classifications pour la stratégie apprentissage actif et bagging sur une partie de la zone d'étude : A) Composition colorée de l'imagerie (R : PIR 2, V : Rouge, B : Bleu) ; B) Classification Random Forest ; C) Classification C5.0 ; D) Classification LDA.



**Figure 4** : Exemple de représentation spatiale des indicateurs d'incertitudes (la stratégie choisie ici est l'active learning et bagging). A) Composition colorée de l'imagerie (R : PIR 2, V : Rouge, B : Bleu) ; B) Entropie  $\alpha$ -quadratique pour le classifieur Random Forest ; C) Divergence de Kullback-Leibler ; D) Indice de Concordance qualitatif.

## Références

- Argyridis, A., and Argialas, D.P., 2015. A Fuzzy Spatial Reasoner for Multi-Scale GEOBIA Ontologies. *Photogrammetric Engineering & Remote Sensing* 81, 491–498.
- Arvor, D., Durieux, L., Andrés, S., and Laporte, M.-A., 2013. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 82, 125–137.
- Belgiu, M., and Drăguț, L., 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, 67–75.
- Bernstein, L.S., Adler-Golden, S.M., Sundberg, R.L., Levine, R.Y., Perkins, T.C., Berk, A., Ratkowski, A.J., Felde, G., and Hoke, M.L., 2005. Validation of the QUick atmospheric correction (QUAC) algorithm for VNIR-SWIR multi- and hyperspectral imagery. pp. 668–678.
- Bishop, M.P., James, L.A., Shroder Jr., J.F., and Walsh, S.J., 2012. Geospatial technologies and digital geomorphological mapping: Concepts, issues and research. *Geomorphology* 137, 5–26.
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65, 2–16.
- Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., et al., 2014. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 180–191.
- Boyd, D.S., and Foody, G.M., 2011. An overview of recent remote sensing and GIS based research in ecological informatics. *Ecological Informatics* 6, 25–36.
- Breiman, L., 1996. Bagging Predictors. *Machine Learning* 24, 123–140.
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32.
- Burnham, K.P., and Anderson, D.R., 2001. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildl. Res.* 28, 111–119.
- Clark, M.L., Roberts, D.A., and Clark, D.B., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment* 96, 375–398.
- Clinton, N., Yu, L., and Gong, P., 2015. Geographic stacking: Decision fusion to increase global land cover map accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing* 103, 57–65.
- Comber, A., Fisher, P., Brunsdon, C., and Khmag, A., 2012. Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment* 127, 237–246.
- Convertino, M., Mangoubi, R.S., Linkov, I., Lowry, N.C., and Desai, M., 2012. Inferring Species Richness and Turnover by Statistical Multiresolution Texture Analysis of Satellite Imagery. *PLOS ONE* 7, e46616.
- Corbane, C., Lang, S., Pipkins, K., Alleaume, S., Deshayes, M., García Millán, V.E., Strasser, T., Vanden Borre, J., Toon, S., and Michael, F., 2015. Remote sensing for mapping natural habitats and their conservation status – New opportunities and challenges. *International Journal of Applied Earth Observation and Geoinformation* 37, 7–16.
- Costa, H., Foody, G.M., and Boyd, D.S., 2015. Integrating User Needs on Misclassification Error Sensitivity into Image

- Segmentation Quality Assessment. *Photogrammetric Engineering & Remote Sensing* 81, 451–459.
- Cracknell, M.J., and Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences* 63, 22–33.
- Crawford, M.M., Tuia, D., and Yang, H.L., 2013. Active Learning: Any Value for Classification of Remotely Sensed Data? *Proceedings of the IEEE* 101, 593–608.
- Criminisi, A., Shotton, J., and Konukoglu, E., 2011. E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning.
- Drăguț, L., Csillik, O., Eisank, C., and Tiede, D., 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing* 88, 119–127.
- Dronova, I., Gong, P., Clinton, N.E., Wang, L., Fu, W., Qi, S., and Liu, Y., 2012. Landscape analysis of wetland plant functional types: The effects of image segmentation scale, vegetation classes and classification methods. *Remote Sensing of Environment* 127, 357–369.
- Du, P., Liu, S., Xia, J., and Zhao, Y., 2013. Information fusion techniques for change detection from multi-temporal remote sensing images. *Information Fusion* 14, 19–27.
- Duro, D.C., Franklin, S.E., and Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118, 259–272.
- Esch, T., Metz, A., Marconcini, M., and Keil, M., 2014. Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. *International Journal of Applied Earth Observation and Geoinformation* 28, 230–237.
- Fauvel, M., Chanussot, J., and Benediktsson, J.A., 2006. Decision Fusion for the Classification of Urban Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing* 44, 2828–2838.
- Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80, 185–201.
- Freund, Y., and Schapire, R.E., 1996. Experiments with a New Boosting Algorithm. In *Proceedings of the thirteenth international conference on machine learning*, (Morgan Kaufmann), pp. 148–156.
- Ghosh, A., and Joshi, P.K., 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation* 26, 298–311.
- Giacco, F., Thiel, C., Pugliese, L., Scarpetta, S., and Marinaro, M., 2010. Uncertainty Analysis for the Classification of Multispectral Satellite Images Using SVMs and SOMs. *IEEE Transactions on Geoscience and Remote Sensing* 48, 3769–3779.
- Gonçalves, L.M.S., Fonte, C.C., Júlio, E.N.B.S., and Caetano, M., 2012. The application of uncertainty measures in the training and evaluation of supervised classifiers. *International Journal of Remote Sensing* 33, 2851–2867.
- Gopal, S., and Woodcock, C., Boston U., 1994. Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets. *Photogrammetric Engineering and Remote Sensing*; (United States) 60:2.
- Huth, J., Kuenzer, C., Wehrmann, T., Gebhardt, S., Tuan, V.Q., and Dech, S., 2012. Land Cover and Land Use Classification with TWOPAC: towards Automated Processing for Pixel- and Object-Based Image Classification. *Remote Sensing* 4, 2530–2553.
- Immitzer, M., Atzberger, C., and Koukal, T., 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing* 4, 2661–2693.
- Jin, X., filed Nov. 14, 2007, and issued Sept. 4, 2012. Segmentation-based image processing system. U.S. Patent 8,260,048.
- Kuhn, M., Weston, S., and Coulter, N.. C code for C5.0 by R. Quinlan, 2014. C50: C5.0 Decision Trees and Rule-Based Models. R package version 0.1.0-21. <http://CRAN.R-project.org/package=C50>
- Kuhn, M., 2015. caret: Classification and Regression Training. R package version 6.0-41 <http://CRAN.R-project.org/package=caret>
- Kullback, S., and Leibler, R.A., 1951. On Information and Sufficiency. *Ann. Math. Statist.* 22, 79–86.
- Laben, C.A., and Brower, B.V., 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening.
- Li, M., Ma, L., Blaschke, T., Cheng, L., and Tiede, D., 2016. A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *International Journal of Applied Earth Observation and Geoinformation* 49, 87–98.
- Liaw, A., and Wiener, M., 2002. Classification and Regression by randomForest. *R News* 2(3), 18–22.
- Loosvelt, L., Peters, J., Skriver, H., Lievens, H., Van Coillie, F.M.B., De Baets, B., and Verhoest, N.E.C., 2012. Random Forests as a tool for estimating uncertainty at pixel-level in SAR image classification. *International Journal of Applied Earth Observation and Geoinformation* 19, 173–184.
- Löw, F., Michel, U., Dech, S., and Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing* 85, 102–119.
- Löw, F., Knöfel, P., and Conrad, C., 2015. Analysis of uncertainty in multi-temporal object-based classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, 91–106.
- Mellor, A., Boukir, S., Haywood, A., and Jones, S., 2015. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, 155–168.
- Minasny, B., and McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, Part B, 301–311.
- Mulder, V.L., de Bruin, S., Schaepman, M.E., and Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. *Geoderma* 162, 1–19.
- Nagendra, H., Lucas, R., Honrado, J.P., Jongman, R.H.G., Tarantino, C., Adamo, M., and Mairota, P., 2013. Remote sensing for conservation monitoring: Assessing protected areas, habitat extent, habitat condition, species diversity, and threats. *Ecological Indicators* 33, 45–59.
- Niculescu-Mizil, A., and Caruana, R., 2005. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning*, (ACM), pp. 625–632.
- Ouergemmi, W., A. Le Bris, N. Chehata, et C. Mallet. 2017. A Two-Step Decision Fusion Strategy: Application to Hyperspectral and Multispectral Images for Urban Classification ». *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42W1 (mai):167-74. <https://doi.org/10.5194/isprs-archives-XLII-1-W1-167-2017>.

- Pal, N.R., Bezdek, J.C., 1994. Measuring fuzzy uncertainty. *IEEE Trans. Fuzzy Syst.*, 2, pp. 107–118
- PostgreSQL Global Development Group, 2015. PostgreSQL 9.4.0 documentation, URL: <http://www.opengeospatial.org/standards/>, Open Geospatial Consortium.
- Powers, R.P., Hermosilla, T., Coops, N.C., and Chen, G., 2015. Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. *International Journal of Applied Earth Observation and Geoinformation* 34, 51–57.
- Pu, R., and Landry, S., 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment* 124, 516–533.
- Quinlan, J.R., 1993. C4.5: Programs for Machine Learning (Morgan Kaufmann).
- Quinlan, J.R., 1996. Bagging, Boosting, and C4.5. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence and Eighth Innovative Applications of Artificial Intelligence Conference, AAAI 96, IAAI 96*, Portland, Oregon, August 4-8, 1996, Volume 1, W.J. Clancey, and D.S. Weld, eds., AAAI Press / The MIT Press), pp. 725–730.
- R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rajan, S., Ghosh, J., and Crawford, M.M., 2008. An Active Learning Approach to Hyperspectral Data Classification. *IEEE Transactions on Geoscience and Remote Sensing* 46, 1231–1242.
- Ramey, J.A., 2012. *activelearning: An R Package to label unlabeled data for classification*. Fred Hutchinson Cancer Research Center, Baylor University.
- Ricotta, C., and Anand, M., 2006. Spatial complexity of ecological communities: Bridging the gap between probabilistic and non-probabilistic uncertainty measures. *Ecological Modelling* 197, 59–66.
- Rocchini, D., and Ricotta, C., 2007. Are landscapes as crisp as we may think? *Ecological Modelling* 204, 535–539.
- Rocchini, D., 2010. While Boolean sets non-gently rip: A theoretical framework on fuzzy sets for mapping landscape patterns. *Ecological Complexity* 7, 125–129.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., and Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography* 35, 211–226.
- Rocchini, D., Delucchi, L., Bacaro, G., Cavallini, P., Feilhauer, H., Foody, G.M., He, K.S., Nagendra, H., Porta, C., Ricotta, C., et al., 2013. Calculating landscape diversity with information-theory based indices: A GRASS GIS solution. *Ecological Informatics* 17, 82–93.
- Rocchini, D., Foody, G.M., Nagendra, H., Ricotta, C., Anand, M., He, K.S., Amici, V., Kleinschmit, B., Förster, M., Schmidlein, S., et al., 2013. Uncertainty in ecosystem mapping by remote sensing. *Computers & Geosciences* 50, 128–135.
- Rocchini, D., Hernández-Stefanoni, J.L., and He, K.S., 2015. Advancing species diversity estimate by remotely sensed proxies: A conceptual review. *Ecological Informatics* 25, 22–28.
- Roerdink, J. and Meijster, A., 2001. The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. *Fundamenta Informaticae*, 41(1-2), pp.187-228.
- Schowengerdt, R.A., 2006. *Remote Sensing, Third Edition: Models and Methods for Image Processing* (Orlando, FL, USA: Academic Press, Inc..
- Settles, B., 2009. *Active learning literature survey*. Computer Sciences Technical Report 1648, University of Wisconsin–Madison
- Smith, A.M.S., Kolden, C.A., Tinkham, W.T., Talhelm, A.F., Marshall, J.D., Hudak, A.T., Boschetti, L., Falkowski, M.J., Greenberg, J.A., Anderson, J.W., et al., 2014. Remote sensing the vulnerability of vegetation in natural terrestrial ecosystems. *Remote Sensing of Environment* 154, 322–337.
- Steele, B.M., Patterson, D.A., and Redmond, R.L., 2003. Toward estimation of map accuracy without a probability test sample. *Environmental and Ecological Statistics* 10, 333–356.
- Steele, B.M., 2005. Maximum posterior probability estimators of map accuracy. *Remote Sensing of Environment* 99, 254–270.
- Stehman, S.V., and Czaplewski, R.L., 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64, 331–344.
- Stumpf, A., Lachiche, N., Malet, J.P., Kerle, N., and Puissant, A., 2014. Active Learning in the Spatial Domain for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* 52, 2492–2507.
- Tso, B. and Mather, P.M., 2009. *Classification methods for remotely sensed data*. Boca Raton, FL: CRC Press.
- Tuia, D., Pasolli, E., and Emery, W.J., 2011. Using active learning to adapt remote sensing image classifiers. *Remote Sensing of Environment* 115, 2232–2242.
- Tuia, D., Volpi, M., Copa, L., Kanevski, M., and Munoz-Mari, J., 2011. A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification. *IEEE Journal of Selected Topics in Signal Processing* 5, 606–617.
- Tuia, D., and Munoz-Mari, J., 2013. Learning user's confidence for active learning. *Geoscience and Remote Sensing, IEEE Transactions on* 51, 872–880.
- Van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., Smeth, J.B. de, and Woldai, T., 2012. Multi- and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation* 14, 112–128.
- Venables, W. N. & Ripley, B. D., 2002. *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Yuhendra, Alimuddin, I., Sumantyo, J.T.S., and Kuze, H., 2012. Assessment of pan-sharpening methods applied to image fusion of remotely sensed multi-band data. *International Journal of Applied Earth Observation and Geoinformation* 18, 165–175.
- Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338–353.
- Zengeya, F.M., Mutanga, O., and Murwira, A., 2013. Linking remotely sensed forage quality estimates from WorldView-2 multispectral data with cattle distribution in a savanna landscape. *International Journal of Applied Earth Observation and Geoinformation* 21, 513–524