

IRSNINSTITUT
DE RADIOPROTECTION
ET DE SÛRETÉ NUCLÉAIRE*Faire avancer la sûreté nucléaire*

Occupation du sol par télédétection dans les 100 km autour du site de Fukushima Daiichi (Japon)

PRP-ENV/SESURE/2016-06

Pôle radioprotection, environnement, déchets et
crise

Service d'étude et de surveillance de la radioactivité dans
l'environnement

Demandeur	GGp-ENV - V1.101.2 - 03.10.04.01				
N° action/processus de rattachement	R4				
<p>Occupation du sol par télédétection dans les 100 km autour du site de Fukushima Daiichi (Japon)</p> <p>J.-M. Métivier - J. Maxant *</p> <p>LEREN</p> <p>Rapport Pôle radioprotection, environnement, déchets et crise</p> <p>Service d'étude et de surveillance de la radioactivité dans l'environnement</p> <p>PRP-ENV/SESURE/2016-</p> <p>* : Service Régional de Traitement et de Télédétection (Université de Strasbourg)</p>					
	Réservé à l'unité		Visas pour diffusion		
	Auteur(s)	Vérificateur	Chef du SESURE	Directeur de l'environnement	Directeur Général de l'IRSN *
Noms	J.-M. MÉTIVIER	M.-A. GONZE	M. BAUDRY	J-C. GARIEL	J. REPUSSARD
Dates					
Signatures					

DIFFUSION : Libre Interne Limitée

*si nécessaire

LISTE DES PARTICIPANTS

Nom	Organisme
J.-M. MÉTIVIER	IRSN
J. MAXANT	SERTIT
B. ALLENBACH	SERTIT

PROVISOIRE

RESUME

Pour répondre aux besoins de plusieurs projets, il est apparu nécessaire de réaliser une occupation du sol dans les 100 km autour du site de Fukushima daiichi. Pour cela, il a été fait l'acquisition de 6 scènes SPOT-5 (résolution 10 m) pour 2 dates et 3 scènes LANDSAT-8 (résolution 30 m) également asynchrones. Les données multispectrales ont été traitées et classées au moyen de l'algorithme SVM (Support Vector Machine) selon une nomenclature à 8 thèmes (urbain continu, urbain diffus, rizière, sol nu, végétation basse [prairie], végétation haute [forêt à feuilles caduques et persistantes], surface d'eau) dont la précision globale est d'environ 90%.

Cette première étude à l'IRSN apparaît prometteuse pour décrire l'environnement de sites d'études et suivre son évolution, surtout en sachant que le programme Copernicus mettra à disposition gratuitement une observation de la Terre d'une résolution de 10 m avec une revisite d'une semaine environ.

ABSTRACT

To answer the needs of multiple projects, it appeared necessary to produce a land use within 100 km of the Fukushima Daiichi site. For this, it was acquired 6 SPOT-5 scenes (resolution 10 m) for 2 times and 3 scenes Landsat-8 (30 m resolution) also asynchronous. The multispectral data were processed and classified using the algorithm has SVM (Support Vector Machine). The nomenclature includes 8 themes (continuous urban, urban diffuse, rice, bare soil, low vegetation [meadow], high vegetation [leafy forest deciduous and evergreen], water surface). The accuracy is about 90%.

To describe the study sites environment and identify its evolution, this first study IRSN appears promising, especially knowing that the Copernicus program will provide a free Earth observation with a resolution of 10 m with a revisit about a week.

MOTS-CLES

Occupation du sol, Fukushima daichi, télédétection multispectrale, SPOT-5, LANDSAT-8, Classification SVM, validation (précision), SENTINEL-2 (Copernicus).

TABLE DES MATIERES

1 INTRODUCTION	5
2 L'EXISTANT	6
3 LA NOMENCLATURE RECHERCHEE	8
4 LES BASES DE DONNEES UTILISEES	9
5 LES IMAGES SPOT-5.....	10
6 LES IMAGES LANDSAT-8	11
7 LA MISE EN ŒUVRE METHODOLOGIQUE	14
7.1 LES TRAITEMENTS METHODOLOGIQUES.....	14
7.2 LA CLASSIFICATION DE L'OCCUPATION DU SOL.....	15
7.2.1 LE Traitement des donnees spot-5	15
7.2.2 LE Traitement des donnees LANDSAT-8.....	18
8 L'OCCUPATION DU SOL FINALE	20
9 LA VALIDATION THEMATIQUE DES RESULTATS	20
10 CONCLUSION ET PERSPECTIVES	27

PROVISoire

LISTE DES FIGURES

Figure 1 : La localisation de la zone d'étude.....	5
Figure 2 : L'existant	7
Figure 3 : Emprise des scènes SPOT-5	10
Figure 4 : Les images SPOT-5 du 22 juillet 2014 et du 20 novembre 2013	11
Figure 5 : Emprise des scènes LANDSAT-8	12
Figure 6 : Schéma de production de la classification SPOT-5	16
Figure 7 : Origine des classifications SPOT-5	17
Figure 8 : Schéma de production de la classification LANDSAT-8.....	19
Figure 9 : Origine des classifications LANDSAT-8	19
Figure 10 : Occupation du sol finale	21
Figure 11 : Occupation du sol finale (Zoom).....	22
Figure 12 : Les points de validation.....	23

LISTE DES TABLEAUX

Tableau 1 : Nomenclature de l'occupation du sol.....	8
Tableau 2 : Avantages/inconvénients des données SPOT-5	9
Tableau 3 : Avantages/inconvénients des données LANDSAT-8.....	9
Tableau 4 : Caractéristiques spectrales des images SPOT-5.....	10
Tableau 5 : Caractéristiques spectrales des images LANDSAT-8	12
Tableau 6 : Caractéristiques des images LANDSAT-8 disponibles.....	13
Tableau 7 : Résultats de la validation de la classification SPOT-5	24
Tableau 8 : Résultats de la validation de la classification LANDSAT-8.....	24
Tableau 9 : Pourcentage de sur-détection et de sous-détection pour chaque classe SPOT-5	25
Tableau 10 : Pourcentage de sur-détection et de sous-détection pour chaque classe LANDSAT-8	25
Tableau 11: Valeurs de précision de l'utilisateur et du producteur pour chaque classe SPOT-5	26
Tableau 12 : Valeurs de précision de l'utilisateur et du producteur pour chaque classe LANDSAT-8.....	27
Tableau 13 : Table d'interprétation du coefficient KHAT (Cohenov Kappa).....	27

Bibliographie

Annexe

1 INTRODUCTION

Aucune base de données géographiques d'occupation du sol n'étant satisfaisante pour nos besoins de modélisation radioécologique environnementale (AMORAD/CONTI/ERO & CYCL, EDOFU ...), il a été décidé d'en réaliser une par traitement de données satellitaires SPOT-5 (champ proche 50 km) et LANDSAT-8 (champ plus lointain 50-100km) : Figure 1.

L'objectif est de mettre au point une «méthode semi-automatique» pour cartographier l'occupation du sol (thèmes : urbain continu, urbain diffus, rizière, sol nu, végétation basse [prairie], végétation haute [forêt à feuilles caduques et persistantes], surface d'eau) autour du site de Fukushima au Japon.

Ce travail a été demandé sous forme de prestation au SERTIT (Service Régional de Traitement d'Image et de Télédétection) de l'Université de Strasbourg. Ce rapport s'inspire largement de celui rédigé par le prestataire.

Ce travail s'inscrit dans le cadre de la fiche GGP environnement V1-101.2 - Action C.

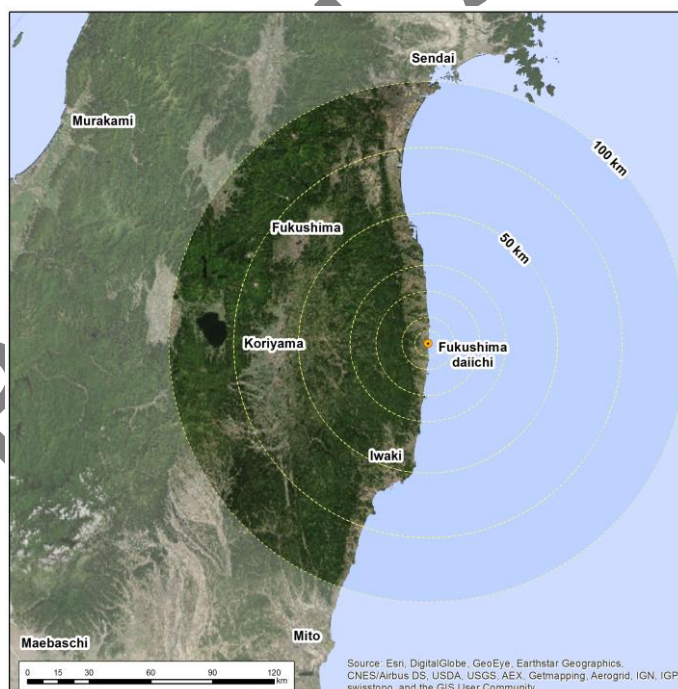


Figure 1 : La localisation de la zone d'étude

2 L'EXISTANT

Plusieurs bases de données géographiques d'occupation du sol sont disponibles mais ne répondent pas à nos besoins et ne sont pas toujours référencées ni millésimées (Figure 2). On en identifie plusieurs :

- Occupation du sol réalisée par la JAXA (Japan Aerospace Exploration Agency) nommée ALOS (Advanced Land Observing Satellite)

La nomenclature est satisfaisante avec 8 classes (Urbain, rizière, surface agricole, prairie, sol nu, forêt à feuilles caduques, forêts à feuilles persistantes, surface d'eau), le millésime n'est pas indiqué (probablement 2008), l'UMC est un pixel de 30 m de côté, l'homogénéité entre les scènes n'est pas respectée (voir scène dans le coin haut à gauche). La précision globale est de 80%. Il semble que les zones forestières aient été surestimées.

- Occupation du sol réalisée par le CEA LSCE à partir d'images SPOT4 et SPOT5 diachronique

La nomenclature est de 5 classes (Urbain, surface agricole, forêt, réseau routier majeur, surface d'eau). Le millésime, l'UMC et la précision ne sont pas indiqués. La zone est limitée aux bassins versants de la Mano et de la Nitta.

Occupation du sol réalisée par l'IRSN/LM2E en mars 2011 à partir d'une image LANDSAT-7TM du 4 Avril 2004

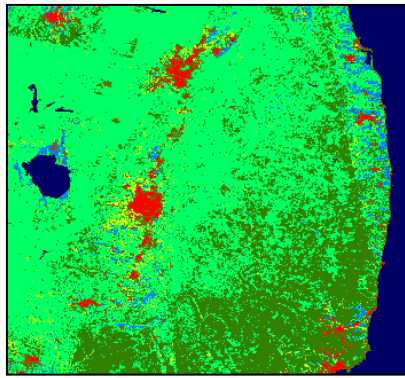
La nomenclature est de 4 classes (Urbain, surface agricole dont rizière, forêt, surface d'eau). L'UMC est un pixel de 30 m de côté. La précision n'est pas indiquée.

- Occupation du sol réalisée par l'Université Toulon (J. Caniac) en 2013 à partir d'images Landsat-7TM

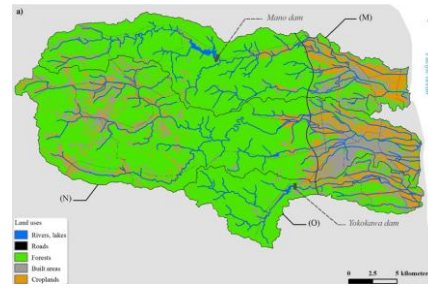
La nomenclature est de 5 classes (Urbain, surface agricole dont rizière, forêt à feuilles persistantes, forêt à feuilles caduques ou prairies, surface d'eau). L'UMC est un pixel de 30 m de côté. Le millésime (probablement les dernières scènes exploitables LANDSAT-7TM, 2005) et la précision ne sont pas indiqués.

- Occupation du sol Kondo (Kondo est le seul qualificatif que nous possédons)

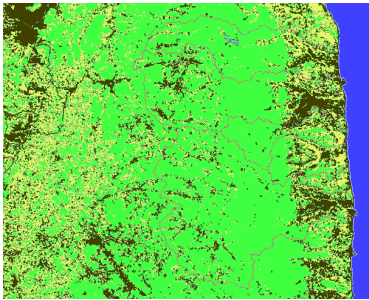
La nomenclature est de 7 classes (Urbain, surface agricole, rizière, prairie ou sol nu, forêt, surface d'eau, autre). On ne sait pas comment a été réalisée cette occupation du sol. L'UMC, le millésime et la précision ne sont pas indiqués.



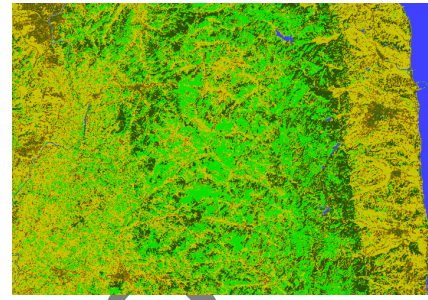
ALOS



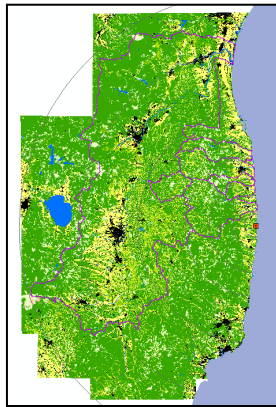
CEA-LSCE



IRSN/LM2E



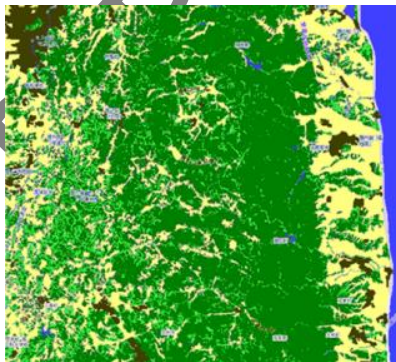
TOULON (J. Caniac)



KONDO



10 m



GLOBAL land cover

Figure 2 : Bases de données géographiques existantes

- Occupation du sol 10m
La nomenclature est de 5 classes (Urbain, surface agricole, rizière, prairie ou sol nu, forêt à feuilles persistantes, forêt à feuilles caduques, surface d'eau). On ne connaît pas comment a été réalisée cette occupation du sol. L'UMC, le millésime et la précision ne sont pas indiqués. La zone est limitée aux bassins versants (Mano, Nitta, Otta, Ukedo [Étude ERO]).
- Occupation du sol réalisée par le NGCC (Centre National Chinois de la Géomatique) nommée GLOBAL Land cover.
Pour la zone d'étude, la nomenclature est de 5 classes (Urbain, surface agricole dont rizière, prairie, sol nu, forêt, surface d'eau), le millésime n'est pas indiqué (probablement 2008), l'UMC est un pixel de 30 m de côté,.
- Occupation du sol réalisée par le CEA/LSCE à partir d'images Pléiades (en cours).

3 LA NOMENCLATURE RECHERCHEE

Les éléments du sol à cartographier sont définis dans la nomenclature ci-après (tableau 1).

Cette nomenclature répond aux conditions suivantes : couverture de la totalité du territoire (partition), homogénéité sémantique, hiérarchisation logique et faisabilité technique, classe SYMBIOSE.

La vision biophysique, associée aux capacités de la télédétection est l'axe technique principal de l'étude.

Tableau 1 : Nomenclature de l'occupation du sol

Niveau 1	Niveau 2	Classe	Classe SYMBIOSE	Commentaires	Libellé
Urbain	Continu	11	C11	<i>Caractérisés par la présence humaine</i>	<i>Surface anthropisée</i>
	Diffus	12	C12	<i>Différenciation par critère morphologique</i>	<i>Surface anthropisée diffuse (ou autres)</i>
Surface agricole	Rizière	21	C4	<i>Analyse diachronique nécessaire</i>	<i>Rizière</i>
Végétation basse	Prairie	22	C5	<i>Prairie principalement</i>	<i>Prairie</i>
Sol nu		23		<i>Rizière sèche, carrière, etc.</i>	<i>Sol nu (labours)</i>
Végétation haute	Feuilles caduques	31	C81	<i>Différenciation par analyse diachronique</i>	<i>Forêt à feuilles caduques</i>
	Feuilles persistantes	32	C82	<i>Différenciation par analyse diachronique</i>	<i>Forêt à feuilles persistantes</i>
Eau		41	C9-C10	<i>Continental</i>	<i>Eau continentales</i>
		42	C11	<i>Estuaire</i>	<i>Estuaire</i>
		43	C12	<i>Maritime</i>	<i>Océan Pacifique</i>

4 LES BASES DE DONNEES UTILISEES

Ce sont essentiellement des scènes satellitaires composées de plusieurs contenus spectraux issus de deux capteurs, SPOT-5 et LANDSAT-8, de caractéristiques différentes. Les avantages et inconvénients de chaque satellite sont présentés dans les tableaux 2 et 3.

Les images SPOT-5 ont été acquises au moyen du programme ISIS (Incitation à l'utilisation Scientifique des Images SPOT). Ce dernier vise à promouvoir l'utilisation de l'imagerie satellitaire SPOT auprès de la communauté scientifique en lui permettant d'acquérir des données satellitaires à un tarif préférentiel (200€ l'image au lieu de 700€). Pour ce faire, il faut identifier si la zone recherchée est disponible et sans trop de couverture nuageuse, monter un dossier et en attendre la réponse, ce qui a pris dans notre cas 6 mois.

Les images LANDSAT-8 ont été directement chargées à partir du site de USGS (United States of Geological Survey) : www.usgs.gov/pubprod/

Tableau 2 : Avantages/inconvénients des données SPOT-5

SPOT-5	
Avantages	Inconvénients
La résolution (10 mètres en mode multispectral pour SPOT-5) permet une caractérisation précise du territoire	Couverture de 60*60 km pour une scène. Zone d'étude non couverte intégralement, besoin de plusieurs scènes (hétérogénéité temporelle)
Préparation à l'arrivée des données Sentinel-2 : même résolution	Faible nombre d'images disponibles. Impossible de s'affranchir complètement des nuages
	Fourniture des images via le programme ISIS : faible coût des données mais cependant pas gratuites, délai, soumis à l'acceptation du dossier, montage du dossier.
	SPOT 4 désorbité en 2013 et SPOT 5 en fin de mission. Besoin de changer de capteur dans le cas d'une éventuelle mise à jour future.

Tableau 3 : Avantages/inconvénients des données LANDSAT-8

LANDSAT-8	
Avantages	Inconvénients
Une scène (185 par 185 km) suffit à couvrir à peu de choses près la totalité de la zone d'étude	Résolution grossière : 30 mètres en mode multispectral. Les éléments dont la taille est inférieure ou égale à la résolution ne seront pas détectés.
L'emprise couverte est la même pour chaque image de date différente	
Offre d'images suffisante pour s'affranchir des nuages	
Données gratuites et disponibles sans délai	
Acquisitions systématiques et assurées pour d'éventuelles mises à jour futures (satellite en début de mission, lancé le 11 février 2013)	

5 LES IMAGES SPOT-5

Le tableau 4 donne les caractéristiques spectrales des images acquises par le satellite SPOT-5.

Tableau 4 : Caractéristiques spectrales des images SPOT-5

Type d'image	Résolution (m)	Fauchée (km)	Bande spectrale	Longueur d'onde (µm)
Multispectrale	10	60 x 60	1 - Vert (V)	0.50 – 0.59
			2 - Rouge (R)	0.61 – 0.68
			3 - Proche Infrarouge (PIR)	0.78 – 0.89
			4 - Moyen Infrarouge (MIR)	1.58 – 1.75

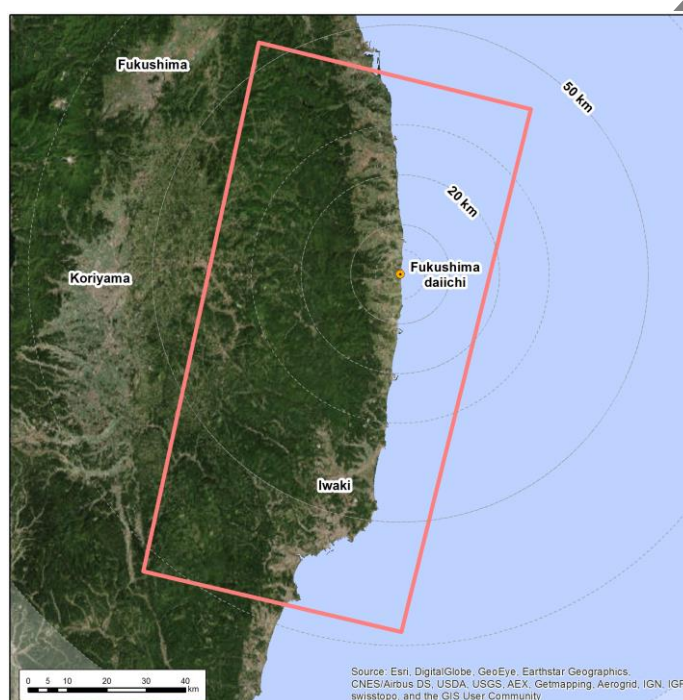


Figure 3 : Emprise des scènes SPOT-5

Deux couvertures asynchrones composées chacune de trois images multispectrales SPOT-5 couvrent la majorité de la zone de 50 km autour du site de Fukushima daiichi (Figure 3). Ces deux couvertures sont disponibles avec une proportion de nuages relativement faible, elles datent du 20 novembre 2013 et du 22 juillet 2014.

Les rizières sont identifiables. Elles apparaissent irriguées pour la scène estivale et sont soit végétalisées, soit labourées dans celle de novembre. Le niveau de chlorophylle des feuilles est bien plus bas en novembre, ce qui permet la différenciation feuilles persistantes/feuilles caduques.

Comme le montre la Figure 4, afin de compléter les espaces manquants sur ces images ou ceux occultés par la présence de nuages dans l'une ou l'autre des deux images, l'utilisation d'images LANDSAT-8 a été nécessaire. Les deux couvertures disposent d'une proportion de nuage faible mais non négligeable comme on peut le voir sur la figure 3. Ces secteurs non exploitables ont été traités différemment par rapport aux secteurs exploitables dans les deux images.

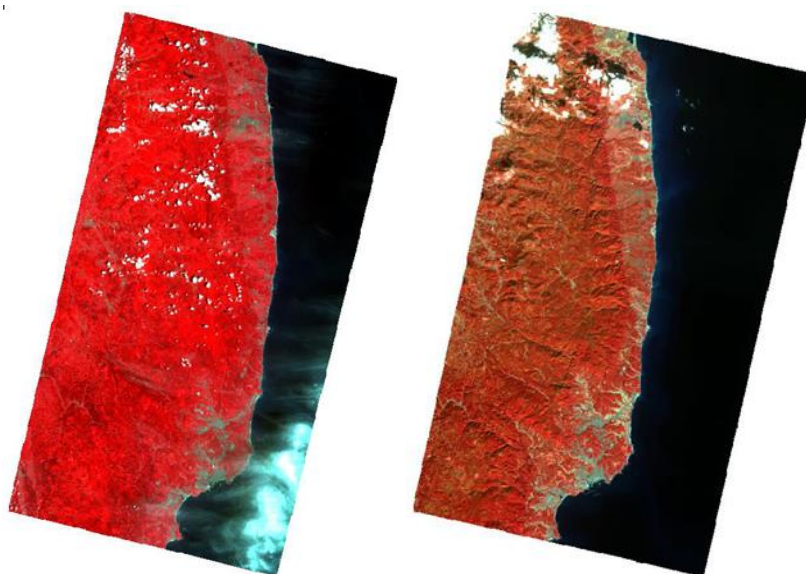


Figure 4 : Les images SPOT-5 du 22 juillet 2014 et du 20 novembre 2013 (restituées en fausse couleur)

6 LES IMAGES LANDSAT-8

Le tableau 5 donne les caractéristiques spectrales des images acquises par le satellite SPOT-5.

Une fouille d'images exploitables a été réalisée dans les archives LANDSAT-8 disponibles. Cinq scènes ont été sélectionnées couvrant le secteur d'étude (Figure 5). Le tableau 6 présente ces données en détaillant l'intérêt de chacune ainsi que les problèmes éventuels en vue de leur traitement ultérieur.

Les images les plus récentes sont favorisées dans le choix afin de produire une occupation du sol la plus récente possible. La scène du 6 août 2015 est toutefois difficilement exploitable en raison d'un couvert nuageux trop important. Celle du 13 avril 2014 est également écartée car le manteau neigeux occulte une grande partie du paysage. Trois images LANDSAT-8 sont finalement retenues pour les traitements (Tableau 6):

- celle du 17 septembre 2013, indispensable pour la détection des rizières et faiblement nuageuse ;
- celle du 2 mai 2015, très récente et faiblement nuageuse ;
- celle du 31 mai 2014, en complément car absence de nuages dans le secteur non-exploitable dans les deux autres.

La distinction forêt à feuilles persistantes/forêt à feuilles caduques est rendue difficile par le compromis à faire pour choisir dans le jeu de données une image suffisamment hivernale pour que les feuilles ne soient pas visibles et une image dont le couvert neigeux ne soit pas trop important pour ne pas masquer les forêts d'altitude.

Tableau 5 : Caractéristiques spectrales des images LANDSAT-8

Type d'image	Résolution (m)	Fauchée (km)	Bande spectrale	Longueur d'onde (µm)
Multispectrale	30	185 x 185	1 - Bleu côtier (B)	0.43 - 0.45
			2 - Bleu (B)	0.53 - 0.59
			3 - Vert (V)	0.78 - 0.89
			4 - Rouge (R)	0.64 - 0.67
			5 - Proche Infrarouge (PIR)	0.85 - 0.88
			6 - Moyen Infrarouge (MIR-1)	1.57 - 1.65
			7 - Moyen Infrarouge (MIR-2)	2.11 - 2.29
			8 - Nuage Cirrus	1.36 - 1.38

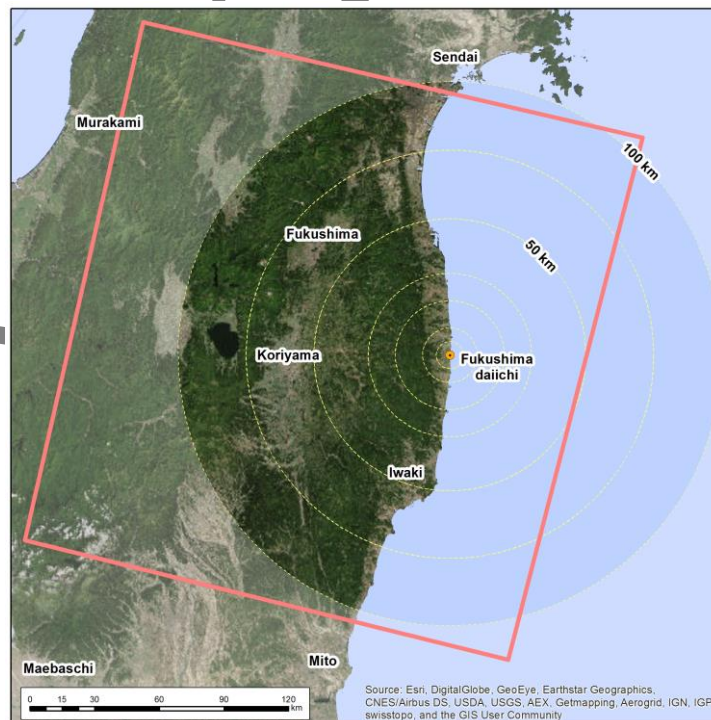







Figure 5 : Emprise des scènes LANDSAT-8

Tableau 6 : Caractéristiques des images LANDSAT-8 disponibles

Date d'acquisition	Aperçu	Intérêts	Problèmes Inconvénients
17 septembre 2013		Proportion faible de nuage Rizières végétalisées, réponse spectrale caractéristique	Présence de nuages Date d'acquisition
13 avril 2014		Quasi-absence de nuages Date à laquelle le débournement n'a pas encore eu lieu, seuls les conifères sont végétalisés et se distinguent des feuillus	Manteau neigeux très important : occupation du sol masquée dans les reliefs notamment dans la partie de l'image non couverte par SPOT-5
31 mai 2014		Proportion faible de nuage	Présence de nuages Manteau neigeux sur les reliefs élevés
2 mai 2015		Quasi-absence de nuages Date d'acquisition	Manteau neigeux très important, occupation du sol masquée dans les reliefs
6 août 2015		Acquisition très récente	Proportion importante de nuage

 Scènes non utilisées

7 LA MISE EN ŒUVRE METHODOLOGIQUE

Tout travail exploitant les traitements de la télédétection doit débuter par la mise en place de la base de données, qui consiste en des traitements géométriques et radiométriques (pour atténuer les effets atmosphériques).

Aucune correction géométrique n'a été réalisée dans cette étude car les données SPOT-5 et LANDSAT-8 ont été reçues orthorectifiées, une vérification du calage géométrique a toutefois été effectuée afin de garantir la cohérence géométrique de la pile d'images (stack).

7.1 LES TRAITEMENTS METHODOLOGIQUES

L'algorithme de conversion en valeurs de réflectance exo-atmosphérique est utilisé afin de normaliser les valeurs radiométriques des pixels, traitement indispensable lorsque plusieurs images doivent être analysées. Pour convertir les valeurs de nos images en réflectance, elles subissent un traitement "Top of Atmosphere" (TOA) :

$$\rho_{toa} = \frac{\pi \times L \times d^2}{ESUN \times \cos\theta_s} \text{ avec } L = \frac{CN}{G}$$

ρ_{toa}	Réflectance de l'ensemble objet+atmosphère
L	Luminance
CN	Compte numérique
G	Gain d'étalonnage absolu du capteur (fourni par les entêtes des fichiers images pour chaque bande)
d	Distance Terre-Soleil (en unité astronomique)
ESUN	Moyenne des éclairagements exo-atmosphériques solaire
θ_s	Angle zénithal solaire

L'USGS distribue les données LANDSAT-8 corrigées des effets atmosphériques. Seules les images SPOT-5 sont converties en valeurs de réflectance exo-atmosphérique.

7.2 LA CLASSIFICATION DE L'OCCUPATION DU SOL

Le traitement des images SPOT-5 et LANDSAT-8 est opéré par classification supervisée en s'appuyant sur l'algorithme SVM (Support Vector Machine). Il s'agit d'une méthode supervisée d'apprentissage statistique, considérée comme le «classifieur» actuel le plus performant [Devadas and al., 2012; Mathur and al., 2008; Huang and al., 2002; Melgani and Bruzzone, 2004]. L'algorithme SVM est intégré dans la librairie open-source Orfeo Toolbox (OTB) développée par le CNES (Centre National d'Etudes Spatiales).

L'enjeu pour certaines classes de l'occupation du sol est de traiter des images multi-temporelles. Plutôt que de traiter chaque donnée indépendamment, les signatures spectrales multi-temporelles peuvent être directement analysées par un seul processus. Pour cela il est nécessaire d'empiler les bandes de toutes les images pour obtenir une seule donnée concaténée sur laquelle peut être réalisée la classification supervisée. C'est un moyen immédiat et sans traitement intermédiaire de prendre en compte les caractéristiques spectrales multi-temporelles [Castrence and al., 2014]. Cette méthode est notamment utilisée lorsqu'il y a calcul d'indice (NDVI, IARR, Brillance ...)

L'ensemble de la nomenclature est «visé directement» excepté l'urbain, d'abord détecté par classification supervisée puis fractionné en deux classes par application d'un critère de taille afin de distinguer l'urbain diffus de l'urbain continu.

La sélection des échantillons pour l'apprentissage et l'entraînement du classifieur est sans aucun doute une des phases les plus cruciales de la classification supervisée. En effet les techniques de sélection des échantillons sont très nombreuses et complexes mais il n'existe cependant pas de «règles absolues». La définition des échantillons requiert un niveau d'expertise élevé et peut être fastidieux et long lorsque de grandes surfaces, comme dans cette étude, sont mises en jeu.

7.2.1 LE TRAITEMENT DES DONNEES SPOT-5

La présence de nuages et d'ombre dans chaque image SPOT-5 a dû être prise en compte pour les traitements. Non seulement il n'est pas possible d'exploiter ces pixels pour déterminer une occupation du sol, mais ces pixels non exploitables ne peuvent également pas être isolés et apportent de la confusion à la classification. En effet, leur signature spectrale se mélange respectivement soit avec les pixels à forte réflectance (nuage, urbain, sable), soit avec ceux à faible réflectance (ombre de nuage, ombre du relief, eau, végétation haute).

Un traitement spécifique doit donc être réalisé pour les pixels non exploitables afin de ne pas perturber la classification des pixels exploitables. L'isolement de ces portions d'images se fait par l'extraction d'un masque des zones claires et d'un masque des zones sombres dans les deux images. La production de ces quatre couches constitue l'étape 1 de la chaîne de traitement mise en place et illustrée par la figure 6.

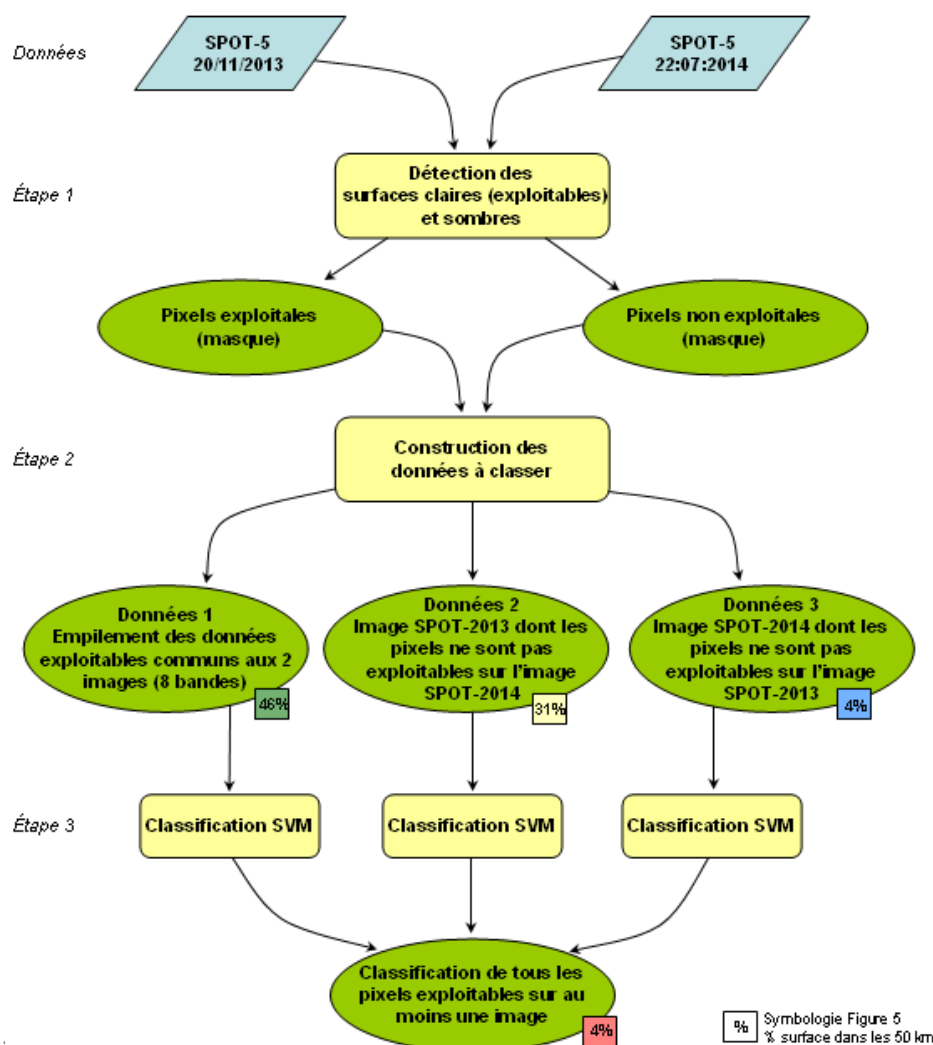


Figure 6 : Schéma de production de la classification SPOT-5

La seconde étape est de construire les données à renseigner en entrée du classifieur. La donnée n°1 est la donnée principale car elle correspond uniquement à des pixels exploitables dans les deux images et s'affranchit du problème des nuages et des ombres (pixels exclus du traitement). Les deux images SPOT-5 sont alors traitées conjointement (après empilement des deux données pour n'en constituer qu'une seule de huit bandes) par classification supervisée après la phase de sélection des échantillons. La première classification obtenue a confirmé l'intérêt de la méthode mais un ajustement des échantillons fut nécessaire pour converger vers un résultat optimal.

Dans le cas de pixels non exploitables dans l'une ou l'autre donnée, la classification supervisée ne peut être réalisée sur les deux images conjointement. La lecture multi-temporelle étant impossible, les deux images sont alors traitées individuellement pour compléter la classification principale. La donnée de novembre 2013 est privilégiée par rapport à la donnée de juillet 2014 afin d'assurer une qualité haute de détection des rizières. Cependant certaines zones ne sont pas classifiées car certains pixels ne sont pas exploitables ni dans l'image du 20 novembre 2013, ni dans l'image du 22 juillet 2014. Les surfaces claires (nuages, neige, urbain) communes aux deux images couvrent un très grand territoire et nécessitent d'être classifiées afin de ne pas perdre de l'information exploitable (urbain) à 10 mètres de résolution, au risque d'apporter de la confusion avec la classification des nuages et du manteau neigeux.

Les classifications obtenues sont ensuite compilées pour obtenir la classification finale dont la qualité varie en fonction de la qualité de chaque donnée en entrée. La figure 7 ci-dessous illustre l'emprise de chaque classification d'origine dans la classification finale.

Concernant les zones non traitées, la classification issue des traitements SPOT-5 est complétée par la classification issue des traitements Landsat-8 ré-échantillonnée à 10 mètres.

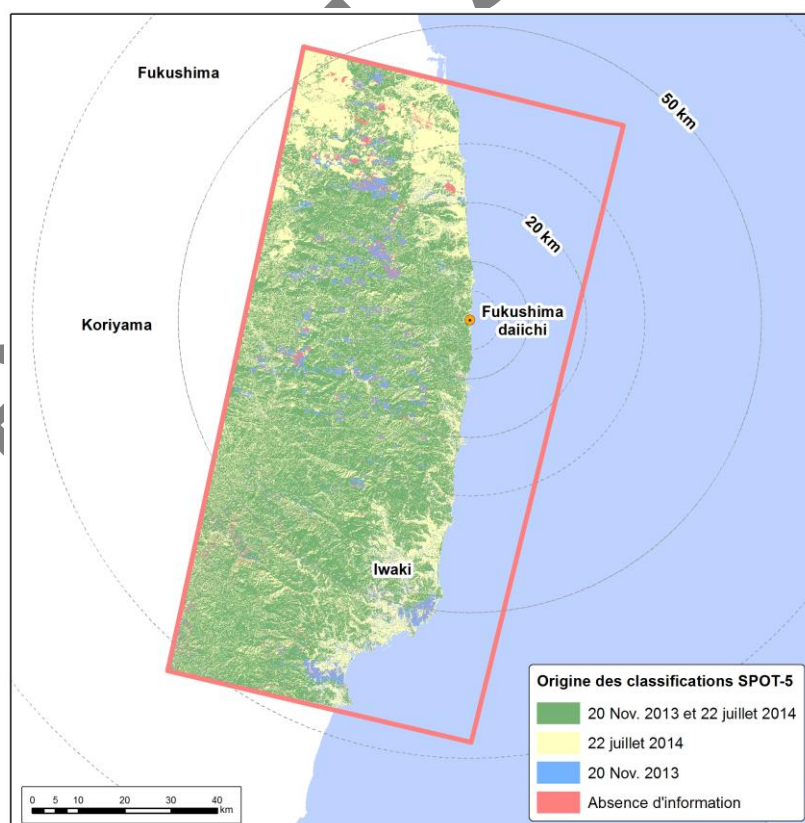


Figure 7 : Origine des classifications SPOT-5

7.2.2 LE TRAITEMENT DES DONNEES LANDSAT-8

La classification issue des images LANDSAT-8 a pour but de couvrir un plus large secteur autour du site de Fukushima Daiichi (rayon de 100 km) mais également de combler les zones classées en «no data» dans la classification SPOT-5. Les données LANDSAT-8 sont traitées indépendamment sur l'ensemble de l'emprise des images.

Classification n° 1

De même que pour les données SPOT-5, la présence de nuages a dû être prise en compte pour les traitements. La scène du 2 mai 2015 dispose d'un important couvert nuageux. La classification n°1 est le résultat de l'algorithme SVM exécuté conjointement sur les images acquises le 17 septembre et le 2 mai 2015 après empilement des 16 bandes (2 fois 8). C'est le résultat principal obtenu par traitement de données multi-temporelles avec une majorité des images exploitables et une caractérisation des rizières optimale. La distinction feuilles persistantes/feuilles caduques est facilitée par l'analyse diachronique. L'idéal aurait été de traiter une des deux images plus hivernale afin d'observer aucune végétation dans les feuillus. Cependant, il s'est avéré que les images d'hiver, ainsi que celles de début de printemps, se caractérisent par une couverture très importante de neige masquant la majorité des forêts.

Classification n° 2

La classification n° 1 doit être complétée dans les secteurs de nuage ou de neige composant les deux images en entrée. L'image LANDSAT-8 du 17 septembre 2013 est alors traitée individuellement pour combler les secteurs de neige de l'image du 2 mai 2015. De par sa date d'acquisition, la scène de 2013 permet de distinguer les rizières des sols nus, des forêts ou encore des prairies. La classification n°2 est le résultat de la classification supervisée SVM opérée sur cette image.

Classification n° 3

Les deux premières classifications cumulées doivent être complétées dans les secteurs de nuage de l'image acquise en 2013. Ces espaces sont comblés par la classification de l'image LANDSAT-8 acquise le 31 mai 2014, dont la couverture nuageuse n'intersecte pas les zones manquantes.

Classification finale

Comme l'illustre le schéma de la figure 8, les trois classifications sont ensuite empilées par ordre de priorité des classements 1, 2 et 3.

La qualité de la classification finale est donc liée à l'origine des classifications utilisées et donc aux images traitées en entrée. La figure 9 ci-dessous illustre l'emprise de chaque classification d'origine dans la classification finale.

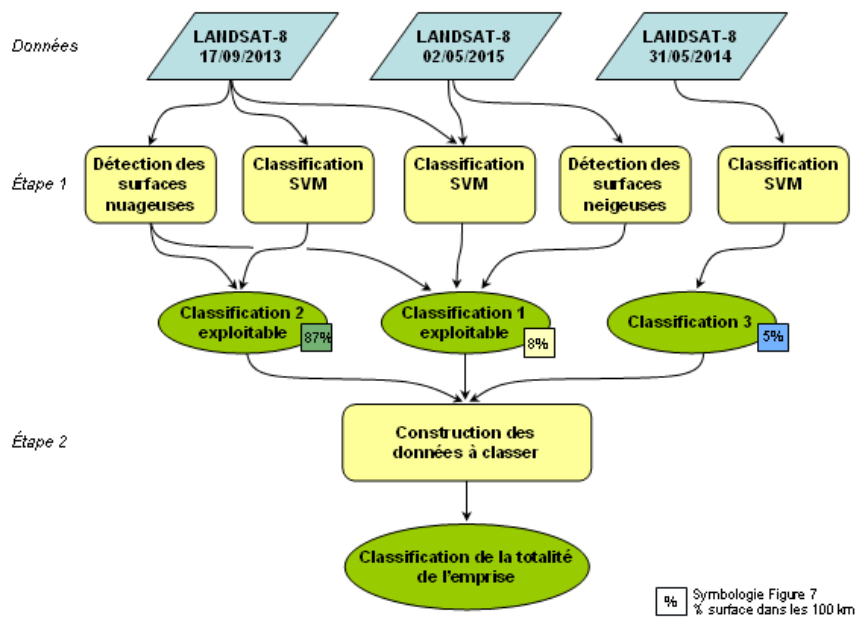


Figure 8 : Schéma de production de la classification LANDSAT-8

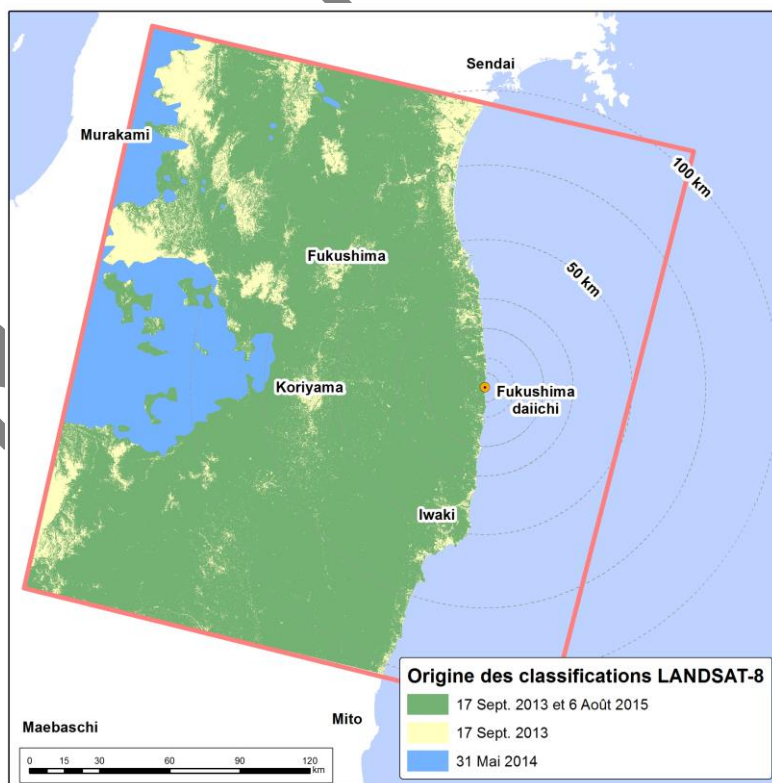


Figure 9 : Origine des classifications LANDSAT-8

8 L'OCCUPATION DU SOL FINALE

Le résultat final de la cartographie est illustré sur la figure 10. La figure 11 montre une extraction (zoom) avec confrontation d'une image aérienne et l'occupation du sol donnée par Open Street Map.

La limite entre les classifications issues des scènes SPOT-5 et LANDSAT-8 ne se remarque pas seulement par la taille des pixels mais également par une discontinuité thématique partielle. En effet, la résolution des images LANDSAT-8 ne permet pas une détection fine du linéaire urbain et des parcelles étroites de rizières dans les fonds étroits de vallée. Dans ce cas, la signature spectrale résultante de LANDSAT-8 est une agrégation des différentes signatures spectrales des différents objets composant le pixel alors que pour SPOT-5, le paysage se distingue «plus finement» (urbain, rizière, massif forestier ...).

Les surfaces nuageuses ou neigeuses communes aux deux images SPOT-5 se remarquent également par une correspondance de classes aberrantes. Une phase de nettoyage par photo-interprétation, en s'appuyant sur les images SPOT-5 et au besoin LANDSAT-8 et Google-Maps, s'est avérée nécessaire afin de corriger ces défauts prévisibles mais inévitables.

9 LA VALIDATION THEMATIQUE DES RESULTATS

Une validation a été réalisée pour évaluer la qualité thématique (précision de chaque classe (voir annexe).

Une distinction est faite pour évaluer la qualité de la classification issue de SPOT-5 (zone des 50 km) et la classification issue de LANDSAT-8 (figure 12). Deux validations ont donc été réalisées en prenant soin d'écarter la mer des espaces de validation. Pour obtenir une validation représentative du point de vue statistique, il est recommandé de sélectionner 50 points de manière aléatoire pour chaque classe [Congalton et Green, 2009]. Il est donc nécessaire de prendre 400 points au total pour valider chaque classification.

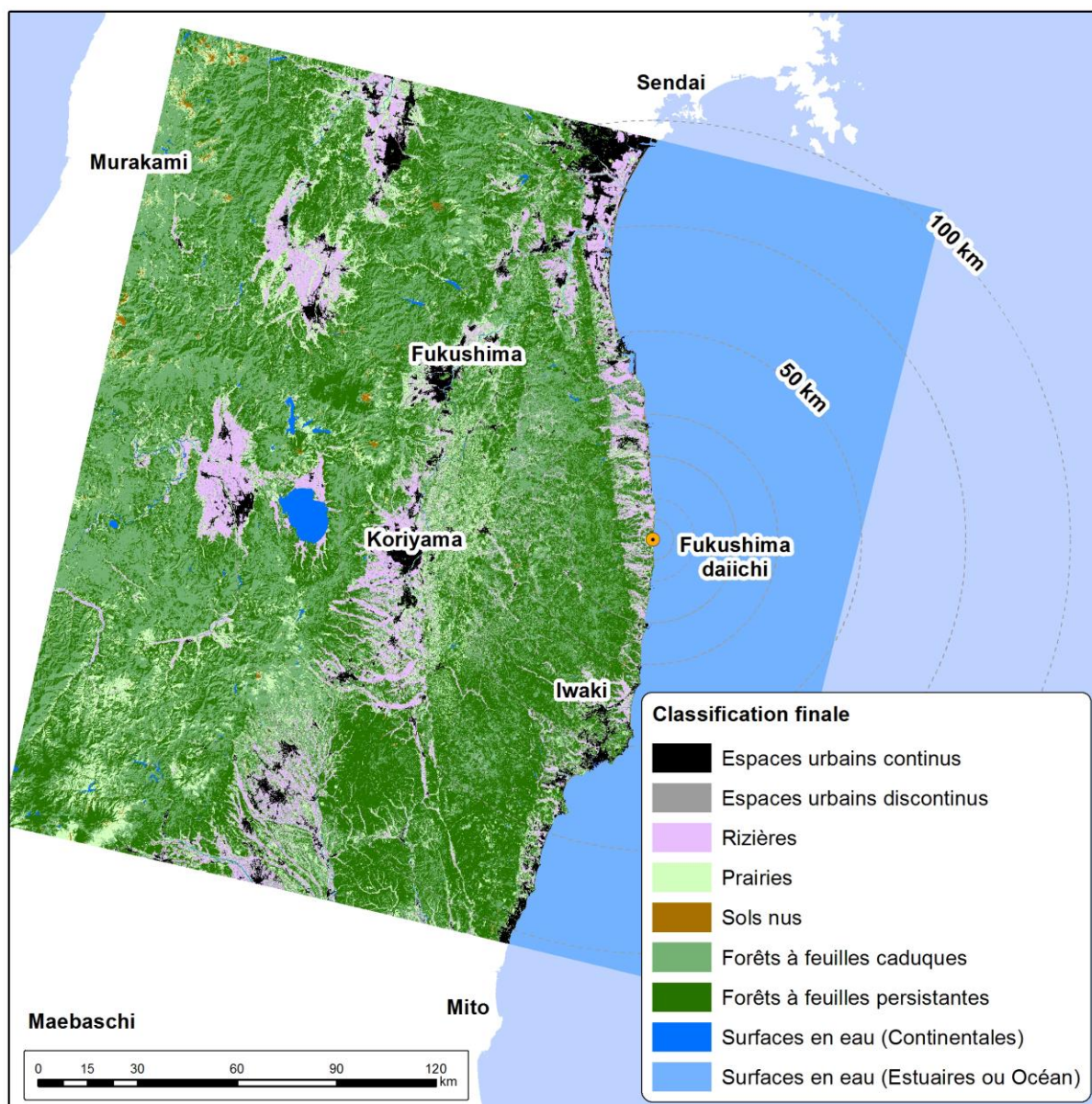
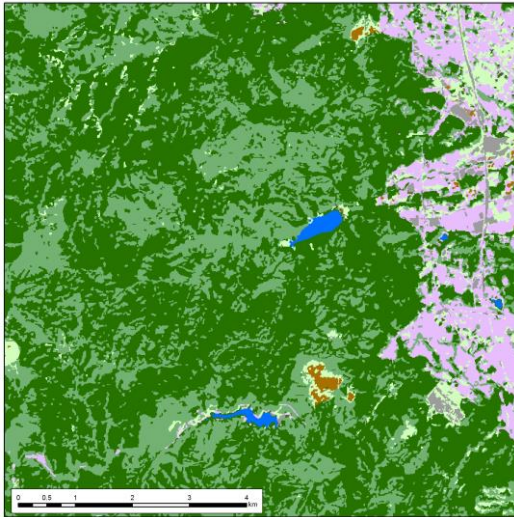


Figure 10 : Occupation du sol finale



Zoom

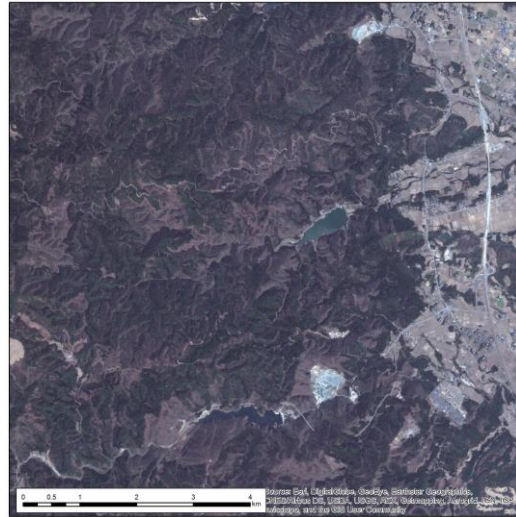
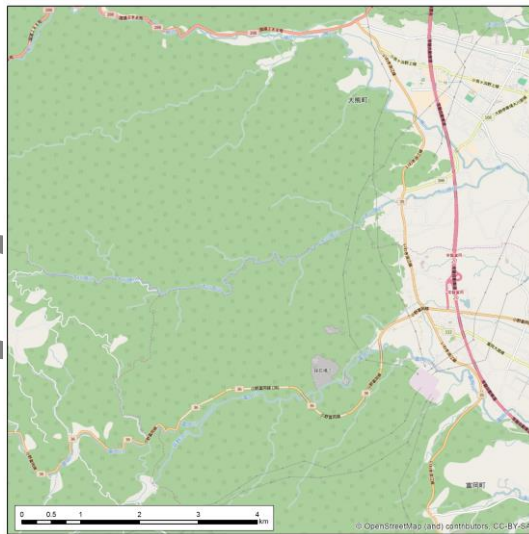


Image aérienne



Open Street Map

Figure 11 : Occupation du sol finale (Zoom)

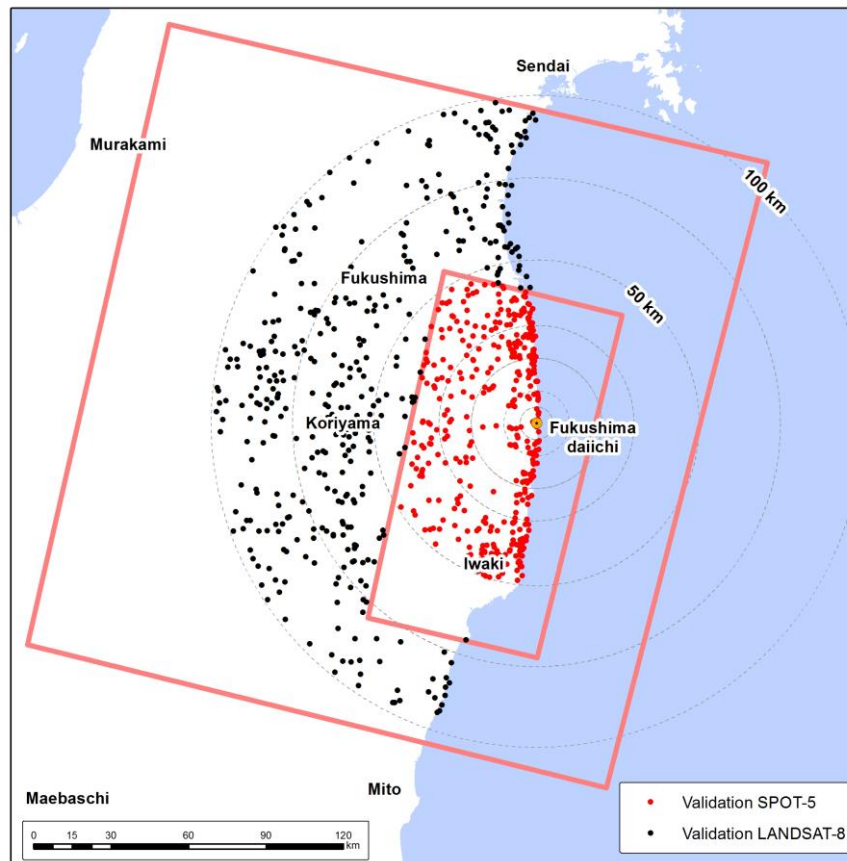



Figure 12 : Les points de validation

Les tableaux 7 et 8, appelés matrices de confusion*, sont un outil de base pour évaluer de manière statistique la précision des classifications. Dans ces tableaux, il faut savoir que les lignes correspondent aux valeurs des classes de la cartographie issue de la télédétection (qui doivent être validées) tandis que les colonnes sont les valeurs interprétées par photo-interprétation et qui constituent la référence. La photo-interprétation est opérée à partir des images sources utilisées pour la classification. Google Maps est également utilisé mais en gardant à l'esprit que la résolution de référence est la résolution des images utilisées pour la classification. Plus précisément ce ne sont pas directement les points qui sont analysés mais les surfaces circulaires autour des points de diamètre équivalent à la résolution du capteur. Par exemple un point classé en feuillus qui tombe sur le fond cartographique infra-métrique Google Maps sur un chemin à quelques mètres de la forêt n'est pas considéré comme faux si la forêt est majoritaire au sein du pixel à 10 ou 30 mètres. L'objectif de la validation est d'analyser la qualité de l'algorithme de classification automatique des pixels, nous ne pouvons aller au-delà de la limite de capacité des capteurs à discerner les objets.

* Une matrice de confusion (ou matrice de contingence) sert à évaluer la qualité d'une classification. Elle est obtenue en comparant des données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification. Elle ne doit pas être confondue avec la table de performance qui sert à évaluer l'homogénéité thématique des zones-test servant à réaliser une classification assistée.

Tableau 7 : Résultats de la validation de la classification SPOT-5


	11 Urbain continu	12 Urbain diffus	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute caduque	32V Vég. haute persistante	4 Eau
11 Urbain continu	47		1	1		1		
12 Urbain diffus		39	6		4			1
21 Rizière		1	43	4		2		
22 Vég. basse Prairie	1		3	42	1	3		
23 Sol nu			«3		47			
31 Vég. hte Caduque		1	1	1		45	2	
32 Vég. hte persistante						3	47	
4 Eau	1							49
Total	49	41	57	48	52	54	49	50

 Précision géographique

 Précision statistique

Tableau 8 : Résultats de la validation de la classification LANDSAT-8

	11 Urbain continu	12 Urbain diffus	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute F. caduque	32V Vég. haute F. persistante	4 Eau
11 Urbain continu	50							
12 Urbain diffus		43	3		4			
21 Rizière	1	2	45	2				
22 Vég. basse Prairie		1	2	39		8		
23 Sol nu	1		1	1	46		1	
31 Vég. hte F. caduque				2		47	1	
32 Vég. bas F. persistante						4	46	
4 Eau								50
Total	52	46	51	44	50	59	48	50

 Précision géographique


 Précision statistique

Tableau 9 : Pourcentage de sur-détection et de sous-détection pour chaque classe SPOT-5

En %	11 Urb. Cont.	12 Urb. Dif.	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute F. caduque	32V Vég. haute F. persistante	4 Eau
Sous détection	4.2	4.9	20.4	12.5	9.6	16.7	4.1	7.5
Sur détection	6	22	14	16	6	10	6	2

Tableau 10 : Pourcentage de sur-détection et de sous-détection pour chaque classe LANDSAT-8

En %	11 Urbain continu	12 Urbain diffus	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute F. caduque	32V Vég. haute F. persistante	4 Eau
Sous détection.	3.8	6.5	11.8	9.3	9.8	20.3	4.2	0
Sur détection	0	14	10	22	8	6	8	0

Les tableaux 9 et 10 présentent les taux de sur-détection et de sous-détection pour les huit classes analysées. Ces valeurs sont importantes pour donner une idée de la qualité des résultats pour chaque classe plutôt que de manière générale. La sous-détection correspond à un oubli dans la classification obtenue grâce à la télédétection par rapport à la réalité, alors que la sur-détection est une détection trop importante par rapport à la référence.

En complément des taux de sous-détection et de sur-détection, les valeurs complémentaires de précision du producteur et de l'utilisateur sont utilisées pour compléter l'évaluation de la qualité de la classification (tableaux 11 et 12). En considérant par exemple la classe de conifères issue de SPOT-5, la précision du producteur correspond au taux de conifères correctement identifié dans la classification par rapport à l'ensemble des conifères effectivement présents dans la zone d'étude. Cela signifie que dans cette classification près de 96% des conifères ont été identifiés. Toujours pour la même classe, la précision de l'utilisateur correspond cette fois au pourcentage de la zone classée en conifère qui est effectivement du conifère dans la réalité (et pas une autre classe d'occupation du sol). Ici, ce taux est de 94%.

Concernant SPOT-5, la précision du producteur est supérieure à 90% pour 5 classes d'occupation du sol : urbain continu, urbain diffus, conifères, eau et sol nu. L'algorithme SVM a privilégié l'attribution de ces pixels à ces classes au détriment des 3 autres classes qui affichent un taux de sous-détection plus important : feuillus, prairies et rizières. La précision de l'utilisateur est supérieure ou égale à 90% pour 5 classes également : urbain continu, conifères, feuillus, eau et sol nu. Cumulés, les deux taux de précision permettent d'estimer la qualité individuelle de chaque classe. La signature spectrale de l'urbain, des conifères, de l'eau et du sol nu permet de distinguer ces différents éléments sans trop de difficultés tandis que l'algorithme peine à discriminer les feuillus, les prairies et les rizières car les signatures spectrales se confondent dans certains cas.

Concernant LANDSAT-8, la précision du producteur est supérieure à 90% pour 6 classes d'occupation du sol : urbain continu, urbain diffus, conifères, prairies, eau et sol nu. L'algorithme SVM a privilégié l'attribution de ces pixels à ces classes au détriment des 2 autres classes qui affichent un taux de sous-détection plus important : feuillus et rizières. La précision de l'utilisateur est supérieure ou égale à 90% pour 6 classes également : urbain continu, conifères, feuillus, rizières, eau et sol nu. L'espace de confusion principal se situe au niveau des feuillus et des prairies, il semble que les prairies ont été légèrement sur-détectés au détriment des feuillus (précision du producteur et précision de l'utilisateur inversées). Les rizières sont mieux détectées par rapport à la classification SPOT-5, la différence s'explique par les dates des images utilisées correspondant à des périodes du cycle de culture plus ou moins identifiables par télédétection.

Afin de renforcer la confiance dans les résultats, il est également recommandé de pousser plus loin l'analyse statistique. Diverses méthodes peuvent être utilisées, partant toutes de la matrice de confusion qui est un élément de base indispensable pour effectuer ces analyses. La méthode choisie ici est de procéder à une analyse Kappa, basée sur le calcul d'un coefficient (appelé KHAT ou Cohenov Kappa) qui permet de mesurer le degré de corrélation entre la classification obtenue par télédétection et la réalité sur le terrain. La valeur de ce coefficient représente la concordance plus ou moins forte entre les deux sources de données. Le coefficient est toujours compris entre -1 et 1 (accord maximal). Habituellement, pour interpréter la valeur obtenue, on utilise le barème du tableau 13 [Landis et Koch, 1977].

Cette grille de lecture ne prend pas en compte le fait que l'interprétation du kappa doit être modulée par le nombre de classes. En effet, le kappa aura toujours tendance à être plus faible pour un grand nombre de classes possibles, que pour seulement deux classes à attribuer [Bakeman and al., 1997].

A partir des matrices de confusion (tableaux 7 et 8), les valeurs de KHAT obtenues sont 0.75 et 0.77 respectivement pour les classifications SPOT-5 et LANDSAT-8, indiquant une forte concordance entre la donnée de référence et les classifications, ce qui apporte une bonne crédibilité quant à la précision de la classification globale.

Tableau 11: Valeurs de précision de l'utilisateur et du producteur pour chaque classe SPOT-5

En %	11 Urb. Cont.	12 Urb. Dif.	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute F. caduque	32V Vég. haute F. persistante	4 Eau
Précision producteur	95.9	95.1	79.6	87.5	90.4	83.3	95.9	92.5
Précision utilisateur	94	88	86	84	94	90	94	98

Tableau 12 : Valeurs de précision de l'utilisateur et du producteur pour chaque classe LANDSAT-8

En %	11 Urb. Cont.	12 Urb. Dif.	21 Rizière	22 Vég. basse Prairie	23 Sol nu	31 Vég. haute F. caduque	32V Vég. haute F. persistante	4 Eau
Précision producteur	96.2	93.5	88.2	90.7	90.2	79.7	95.8	100
Précision utilisateur	100	86	90	78	92	94	92	100

Tableau 13 : Table d'interprétation du coefficient KHAT (Cohenov Kappa)

KHAT	INTERPRÉTATION
< 0	Grand désaccord
0.00 – 0.20	Accord très faible
0.21 – 0.40	Accord faible
0.41 – 0.60	Accord moyen
0.61 – 0.80	Accord fort
0.81 – 1.00	Accord presque parfait

10 CONCLUSION ET PERSPECTIVES

La télédétection a permis de répondre au besoin qui était de réaliser une occupation du sol récente à partir de scènes satellitaires multispectrales avec évaluation de la précision des classes thématiques pour une surface de plus de 30.000 km² autour du site nucléaire de Fukushima Daiichi. Cette occupation du sol servira de support aux besoins de modélisation radioécologique environnementale (AMORAD/CONTI/ERO & CYCL, EDOFU ...). La résolution de 10 mètres pour le champ «proche» (zone des 50 km, 6.000 km²) a été réalisée par traitement de scènes SPOT-5 ; la résolution de 30 mètres pour un champ dit «lointain» (zone des 50 km - 100 km, 6.000 km²) a été réalisée par traitement de scènes LANDSAT-8.

Les images SPOT-5 ont été obtenues par le programme ISIS qui vise à promouvoir l'utilisation de l'imagerie spatiale SPOT auprès de la communauté scientifique. Le faible nombre d'archives disponibles n'a pas permis d'avoir la main sur les paramètres d'acquisition suivants : dates d'acquisition et couverture nuageuse. La richesse du catalogue LANDSAT-8 facilite l'utilisation d'imagerie récente et acquise à des périodes d'acquisitions choisies, cependant la qualité des images sélectionnées n'est pas idéale car la zone est rarement complètement dégagée d'un point de vue météorologique et comprend des régions montagneuses enneigées, notamment en hiver et au printemps ce qui empêche une bonne analyse diachronique (distinction feuilles caduques/feuilles persistantes et phénologie des rizières).

Après une phase de traitement radiométrique (scènes SPOT-5), les données ont été traitées pour chaque capteur par classification supervisée en s'appuyant sur l'algorithme SVM. La dimension multi-temporelle des images SPOT-5 et des images LANDSAT-8 est intégrée

directement au coeur de la méthode supervisée d'apprentissage statistique. Ainsi, toutes les bandes multi-temporelles des images (par capteur) sont empilées dans une seule image à partir de laquelle sont sélectionnés les échantillons qui vont entraîner le classifieur SVM. La présence de nuages et d'ombres dans les deux bases de données d'images a complexifié le processus de production. En effet, celui-ci a dû être parallélisé pour ne pas perturber les espaces exploitables tout en optimisant au maximum le traitement des informations dans les espaces difficilement exploitables.

Les classifications SPOT-5 et LANDSAT-8 ont été soumises à un processus de validation statistique. Les résultats indiquent une précision globale d'environ 90% pour SPOT-5 et 92% pour LANDSAT-8, ce qui montre un accord presque parfait avec la réalité. Une analyse statistique approfondie décrit un fort degré de corrélation entre la classification obtenue et l'analyse par photo-interprétation. La qualité de classification est évidemment variable selon les classes et reflète la confusion entre certaines classes dont les signatures spectrales se chevauchent : c'est le cas des forêts à feuilles caduques, des rizières et des prairies qui sont végétalisées tout ou partie du cycle phénologique ou du cycle de culture. La confusion est moindre concernant la classification des rizières à partir d'images LANDSAT-8 car les dates des images utilisées correspondent à des périodes du cycle de culture et la réponse spectrale se démarque du reste du paysage.

La qualité de la classification est optimale par rapport à la qualité des images en entrées du processus de traitement. Une amélioration pourrait néanmoins être possible par utilisation de données plus adaptées, notamment pour les données SPOT-5. Les dates d'acquisition et la couverture nuageuse sont des paramètres qui ont une influence très importante sur la qualité de l'occupation du sol produite.

Avec le projet Copernicus, la mise en orbite récente du premier satellite SENTINEL-2A (10 m de résolution comme les images SPOT-5) et le lancement du second (SENTINEL-2B) pour l'année 2016, cette télédétection ouvre des perspectives intéressantes pour ce type d'étude. En effet, l'acquisition d'images gratuites, de résolution équivalente à SPOT-5, à 10 puis 5 jours d'intervalle de visite pour chaque point de la surface terrestre émergée va rapidement permettre de constituer une base de données suffisamment riche pour s'affranchir des problèmes décrits précédemment, à savoir obtenir une pile d'images exploitables à 100%, acquises aux périodes optimales pour la caractérisation des objets changeants (occupation du sol sur étagère récente). À l'Institut de saisir l'opportunité d'être présent dès la constitution de cette base de données pour l'intégration éventuelle des données d'observation de la Terre dans des processus applicatifs de gestion environnementales. En effet, la télédétection répond au besoin de connaissance des territoires en apportant une matrice paysagère à proximité plus ou moins immédiate des sites nucléaires français et étrangers pour n'importe quel endroit susceptible d'être impacté par des radionucléides.

A l'heure actuelle, l'IRSN n'utilise pas encore issues de l'observation de la Terre avec les moyens satellitaires. Cette étude en représente la première étape exploratoire et s'inscrit dans une approche innovante et prometteuse.

PROVISoire

BIBLIOGRAPHIE

Bakeman R., Quera V., McArthur D., Robinson B.F., 1997. Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods*, (2) :357-370

Castrence M., Nong D.H., Tran C.C., Young L., Fox J. 2014. Mapping urban transitions using multi-temporal landsat and DMSP-OLS night-time lights imagery of the red river delta in Vietnam. *Remote Sensing Land 2013*, ISSN 2073-445X

Congalton R.G., Green K., 2009. *Assessing the accuracy of Remotely Sensed Data - Principles and Practices*, 2nd Edition, CRC Press

Devadas R., Denham R.J., Pringle M., 2012. Support vector machine classification of object-based data for crop mapping, using multi-temporal landsat imagery. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXIX-B7, 2012 XXII ISPRS Congress, 25 August - 01 Septembre 2012, Melbourne, Australie

Huang C., Davis L.S., Townshend J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23, pp. 725-749

Landis J.R., Koch G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, (33) :159-174

Mathur A., Foody G., 2008. Crop classification by a support vector machine with intelligently selected training data for an operational application. *International Journal of Remote Sensing*, 29 (8). pp. 2227-2240. ISSN 0143-1161

Melgani F., Bruzzone L., 2004. Classification of Hyperspectral Remote Sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, vol. 42, n° 8, août 2004

Précisions et type d'erreur en Télédétection

Précision du producteur appelée aussi erreur de déficits ou erreur d'omission (producer accuracy)

Probabilité pour un pixel de la classe C_i d'être bien classé en C_i

$$P_p = (X_{ii}/C_i) * 100 \quad \text{où } C_i = \sum x_{ki}$$

C'est le rapport entre le nombre de pixels bien classés et le nombre total de pixels de la classe de la réalité terrain

Complément à 100% : Erreur de commission

Précision de l'utilisateur appelée aussi erreur de d'excédents ou erreur de commission (user accuracy)

Probabilité pour un pixel classé en C_i d'appartenir à C_i sur le terrain

$$P_u = (X_{ii}/L_i) * 100 \quad \text{où } L_i = \sum x_{ij}$$

C'est le rapport entre le nombre de pixels bien classés et le nombre total de pixels de la classe d'après l'image classée

Complément à 100% : Erreur d'omission

Erreur statistique

Proportion de pixels classés en C_k par rapport au nombre de pixels réellement en C_k sur le terrain

$$EA = (X_{k.}/X_{.k}) * 100$$

Erreur de commission moyenne

Moyenne des EC sur les n classes

Erreur d'omission moyenne

Moyenne des EO sur les n classes

Précision géographique globale

Somme de la diagonale

$$PGG = [(\sum X_{ij}) / X_{..}] * 100$$