

Développement de méthodes innovantes de cartographie de l'occupation du sol à partir de séries temporelles d'images haute résolution visible (NDVI)

Joana Roussillon

► To cite this version:

Joana Roussillon. Développement de méthodes innovantes de cartographie de l'occupation du sol à partir de séries temporelles d'images haute résolution visible (NDVI). Sciences de l'ingénieur [physics]. 2015. [dumas-01339899](#)

HAL Id: [dumas-01339899](#)

<https://dumas.ccsd.cnrs.fr/dumas-01339899>

Submitted on 30 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
ECOLE SUPERIEURE DES GEOMETRES ET TOPOGRAPHES

MEMOIRE

Présenté en vue d'obtenir

Le DIPLOME D'INGENIEUR CNAM

Spécialité : Géomètre et Topographe

par

Joana ROUSSILLON

« Développement de méthodes innovantes de cartographie de l'occupation du sol à partir de séries temporelles d'images haute résolution visible (NDVI) »

Soutenu le 07 Juillet 2015

Maître de stage : Vincent SIMONNEAUX

JURY

PRESIDENT : M. Jérôme VERDUN

Président de Jury

MEMBRES : Mme Joëlle NICOLAS-DUROY

Professeur référente

M. Laurent MOREL

Professeur de l' ESGT

Table des matières

I	Introduction	6
II	Contexte d'étude	8
1	Etat de l'art	8
1.1	Aperçu des méthodes classiques de classification	8
1.1.1	Méthodes ponctuelles ou objets	8
1.1.2	Classifications supervisées/non supervisées	8
1.1.3	Méthodes paramétriques/non paramétriques	9
1.2	Méthodes exploitant les séries temporelles NDVI	11
1.3	Evaluation de la classification	13
1.3.1	Méthodes de validation	13
1.3.2	Indices de performance	15
2	Présentation des données disponibles	16
2.1	Données satellitaires	16
2.1.1	Caractéristiques “brutes”	16
2.1.2	Corrections radiométriques	16
2.2	Données terrain	17
III	Méthodologie de développement	19
1	Présentation générale de la méthode	19
2	Démarche de développement	20
2.1	Visualisation et analyse des données à disposition	20
2.1.1	Construction graphique	20
2.1.2	Visualisation des profils par classe	21
2.1.3	Analyses transversales	27
2.2	Mise en place de la méthode	32
2.2.1	Extraction des profils de référence	32
2.2.2	Extraction des cycles de végétation	32
2.2.3	Critères de comparaison	37
2.2.4	Traitement post-comparaison	37
2.2.5	Méthode retenue	38
IV	Résultats	41
1	Exemples de visualisation de résultats	41
2	Analyse et typologies des données terrain	43
2.1	Typologie relative aux données de références	43
2.2	Typologie relative à la classification	43

3 Tests d'optimisation de la méthode	43
3.1 Mode de comparaison des résultats issus des phases 1 et 2	44
3.2 Valeurs des aires minimales	45
3.3 Influence de la base choisie pour le calcul d'aire	45
3.4 Regroupement de classes	47
3.5 Ajout de critères « experts »	49
4 Performance finale de la méthode retenue	50
4.1 Statistiques finales	50
4.2 Réponse aux mélanges de cultures	53
4.3 Influence du nombre de profils de référence stockés dans la base de données	54
5 Perspective d'améliorations dans le cadre d'études ultérieures	55
6 Rendus finaux	56
V Conclusion	60

Remerciements

Je souhaite remercier tous ceux qui ont pu contribuer de près ou de loin à la réalisation de ce travail.

Tout d'abord, j'adresse ma profonde reconnaissance à Vincent Simonneaux pour m'avoir acceptée en tant que stagiaire, puis pour m'avoir transmis avec autant de patience et d'humour un grand nombre de connaissances. Un grand merci pour tes explications claires, ta bonne humeur et ta disponibilité en toutes circonstances. J'ai apprécié la qualité de tes conseils, de ton encadrement, mais également l'autonomie que tu m'as laissée dans ce travail.

Je remercie plus en amont Vincent Ravalland, pour m'avoir transmis cette proposition de stage, ainsi que Nathalie Le Roi, Laurent Polidori et Vincent Simonneaux pour avoir participé à la bataille administrative m'ayant permis de commencer ce stage à temps.

Je remercie ensuite Joëlle Nicolas-Duroy, ma professeur référente au sein de l'ESGT, pour ses remarques pertinentes suite à une relecture détaillée du mémoire.

Merci à Clément Verne pour son aide informatique, et à Jordi Ingla pour les données qu'il a pu me fournir. Plus généralement, je suis reconnaissante à tous les membres du laboratoire pour leur accueil chaleureux.

Je tiens également à citer tous les autres stagiaires du CESBIO, et plus particulièrement Alice, Marina, Tom, Josselin, Mouna, Arthur, Elias, Mathias, Hassan et Erik pour les moments que nous avons pu partager, notamment lors de nos savoureuses pauses culinaires au CNES.

Enfin, je remercie l'École Nationale des Sciences Géographiques, l'École Supérieure des Géomètres et Topographes et l'Université de Calgary pour ces trois années de formation, qui furent d'une grande richesse sur tant de points.

Glossaire

- **Cycle végétatif** : Temps dont une culture annuelle a besoin pour accomplir son cycle annuel d'implantation, croissance et production de la partie récoltable.
- **Plante/culture annuelle** : Plantes qui ne vivent qu'une saison, qui naissent, fructifient et meurent dans le cours d'une année. Elles se distinguent des plantes vivaces.
- **Cultures permanentes** : Plantes dont la durée de vie est supérieure à deux saisons de croissance végétale et qui perdent leurs feuilles après chaque saison ou dont la croissance est continue.
- **Stress** : En biologie, l'ensemble des réponses à un organisme soumis à des pressions ou contraintes de la part de son environnement.
- **Stress hydrique** : Stress subi par une plante placée dans un environnement qui amène à ce que la quantité d'eau transpirée par la plante soit supérieure à la quantité qu'elle absorbe. Ce stress se rencontre notamment en période de sécheresse.
- **Phénologie** : Etude de l'apparition d'évènements périodiques (annuels le plus souvent) dans le monde vivant, déterminée par les variations saisonnières du climat.
- **Typologie** : Démarche méthodique consistant à définir ou étudier un ensemble de types, afin de faciliter l'analyse, la classification et l'étude de réalités complexes.
- **Complantation ou agriculture complaintée** : Mode d'exploitation traditionnel des terres où différentes cultures sont pratiquées en même temps sur le même terrain. Ce terme est utilisé pour les techniques d'agricultures traditionnelles fortes consommatrices de main-d'oeuvre.
- **Temps de revisite** : Temps écoulé entre deux acquisitions d'images.
- **Série temporelle** : Suite de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps.
- **Images panchromatiques** : Images obtenues à partir de l'enregistrement du rayonnement dans un unique intervalle de longueur d'onde situé dans le domaine du visible, entre 400 et 700 nm.

Liste des abréviations

- **ENVI** : Logiciel de traitement d'images de télédétection commercialisé par la société EXELIS.
- **ROI** : Region Of Interest, terme propre au logiciel de traitement d'images satellitaires ENVI, désignant un ensemble de pixel sélectionnés.
- **NDVI** : Normalized Difference Vegetation Index. Indice de végétation couramment utilisé en télédétection.
- **LAI** : Leaf Area Index. Indice de végétation utilisé en télédétection et permettant de mesurer la surface foliaire par unité de surface au sol.
- **SPOT** : Satellite Pour l'Observation de la Terre.
- **SAMIR** : Satellite Monitoring of IRrigation. Logiciel développé au CESBIO, dans le cadre du projet SudMed, et permettant une gestion spatialisée du bilan hydrique de zones irriguées.
- **UTM** : Universal Transverse Mercator. Projection cartographique.
- **DTW** : Dynamic Time Warping. Algorithme permettant de mesurer la similarité entre deux suites qui peuvent varier au cours du temps.
- **SVM** : Séparateur à Vaste Marge. Méthode de classification d'occupation du sol.
- **UA** : User Accuracy. Indice de performance d'une classification.
- **PA** : Producer Accuracy. Indice de performance d'une classification.
- **OA** : Overall Accuracy. Indice de performance d'une classification.
- **IDL** : Interactive Data Language. Langage de programmation propriétaire répandu dans le domaine de la télédétection.
- **ArcGIS** : Suite de logiciels d'information géographique développés par la société américaine Esri.
- **UMR** : Unité Mixte de Recherche.
- **ST** : Série Temporelle.
- **BISE** : Best Index Slope Extraction.

Première partie

Introduction

Dans les pays sud méditerranéens, l'agriculture, occupe entre 15 et 20% du territoire, emploie environ le quart de la population et consomme 80% des ressources en eau [5]. Cette consommation croissante, loin d'être optimisée, pose de nombreux problèmes sociaux. Les tensions entre usages concurrents (domestique et industriel) sont exacerbées, les inégalités sociales s'aggravent lorsque l'accès à l'eau devient difficile, de nombreuses dégradations environnementales s'observent (salinisation des terres, ...) et certains craignent le déclenchement futur de conflits armés.

Dans ce contexte global, nous nous intéressons plus particulièrement pour notre étude au bassin-versant de Merguillil, de la plaine alluviale de Kairouan, situé en Tunisie Centrale (Figure 1). Densément peuplée, cette zone abrite une activité agricole intense aux forts enjeux économiques : en Tunisie, l'agriculture emploie 16% de la population active (2011) [14] et représente 8,5% du PIB (2012). En climat semi-aride, les conditions météorologiques sont peu favorables à la production qui doit faire face aux aléas climatiques et au déficit de la balance agricole alimentaire. Les pluies sont rares et irrégulières, et on observe périodiquement des chutes de rendements agricoles liées aux sécheresses, si ce n'est parfois la perte totale des récoltes. En vue de combattre ces difficultés, la plaine alluviale est intensivement irriguée à partir de l'aquifère de Kairouan qui constitue la première ressource hydrique de la région. Il en résulte une surexploitation de la nappe phréatique dont le niveau est en baisse continue depuis 30 ans.

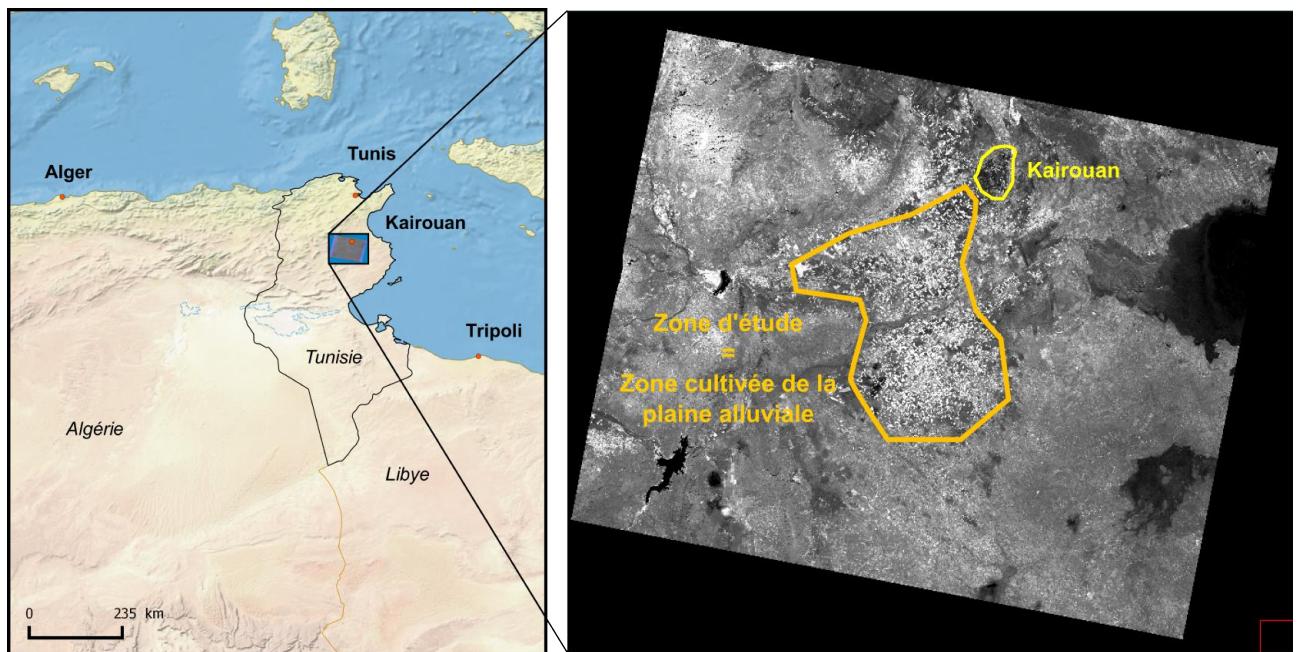


FIGURE 1 – Carte de localisation de notre zone d'étude

Le potentiel offert par les images satellites pour permettre une meilleure gestion des ressources et usages en eau n'est plus à démontrer [7], et s'avère d'une importance primordiale dans ce contexte actuel de pression sur la ressource. Le CESBIO (Centre d'Etudes Spatiales de la BIOsphère), situé à Toulouse, et l'UMR G-EAU de Montpellier travaillent sur l'apport de solutions à ces problèmes notamment grâce à la télédétection. Ces travaux s'inscrivent dans un cadre général de modélisation intégrée du bassin versant du Merguillil et reposent sur deux axes principaux. Le premier consiste au suivi du remplissage des lacs collinaires en relation avec le fonctionnement hydrologique du bassin-versant

amont du Merguellil et à l'alimentation de la nappe de Kairouan située en aval. Le second porte sur l'étude du bilan hydrique des cultures et la gestion de l'irrigation dans la plaine en aval du bassin en relation avec l'évolution de la nappe et les changements d'occupation du sol et de pratiques agricoles. Les séries temporelles permettent le suivi du développement des cultures, des pratiques d'irrigation et ainsi de quantifier par modélisation les pompages dans une nappe profonde en baisse continue. C'est dans ce cadre que s'inscrit ce travail de fin d'études.

Les modèles hydrologiques développés au CESBIO (Modèle SAMIR) n'intègrent pour l'instant qu'une information grossière concernant l'occupation du sol. Afin d'affiner l'étude de leur bilan hydrique, il devient primordial de connaître le plus précisément possible la nature des parcelles agricoles et ce si possible en temps quasi-réel. Les progrès observés dans le monde de l'imagerie spatiale satellitaire offrent de nombreuses perspectives d'améliorations du suivi de ces cultures en continu. En particulier, le premier satellite de Sentinel-2 (Sentinel-2A) vient d'être lancé le 23 juin 2015, et va révolutionner le monde de l'observation de la Terre en proposant gratuitement et en tout point du globe des images haute résolution (15 m) tous les 15 jours dans le domaine du visible. Cette fréquence d'acquisition des images associée à une haute résolution spatiale laisse entrevoir l'émergence de nouvelles méthodes de suivi, plus fiables, plus précises et quasi-instantanées.

A partir de ces considérations, ce travail de fin d'études vise à tester et à développer de nouvelles méthodes de classification d'occupation du sol adaptées à notre zone d'étude et aux séries temporelles prochainement disponibles. En effet, les méthodes actuelles s'avèrent peu efficaces en zone semi-aride du fait de la forte variabilité des cycles de développement des cultures, liées aux conditions climatiques et aux pratiques agricoles employées. Il devient nécessaire de proposer des méthodes robustes, adaptées à ce contexte, qui puissent rendre compte des différentes successions culturales (2 à 3 par an) observées sur les parcelles agricoles, et ce en temps quasi réel et continu, sans attendre l'acquisition d'un nombre minimal d'images. Enfin, l'enjeu reste également de parvenir à des méthodes aussi automatiques que possible, dont on souhaite limiter le recours aux observations de terrains.

Nous nous attacherons à proposer, développer puis tester une telle méthode, à l'aide d'une série temporelle SPOT simulant les futures données acquises par Sentinel-2.

Dans un premier temps, ce travail exploratoire devra s'appuyer sur une analyse détaillée des données dont nous disposons afin de faciliter leur interprétation puis la définition de critères d'identification. Ensuite, ceci permettra de développer un algorithme capable de détecter les changements d'occupation du sol, de manière la plus automatique possible. Enfin, seront étudiés dans une dernière partie les résultats obtenus par la méthode ainsi proposée. Y seront analysés sa performance, sa robustesse, ses limites et son potentiel en vue des applications envisagées.

Deuxième partie

Contexte d'étude

1 Etat de l'art

En télédétection, la classification consiste à effectuer la correspondance entre les éléments d'une scène de l'image (espace géographique) matérialisés par leurs valeurs radiométriques (espace des signatures spectrales), et des classes connues a priori ou non par un utilisateur [20]. Est défini par classe un ensemble d'éléments (pixels ou objets) dont les caractéristiques ont été jugées similaires, humainement ou mathématiquement [6]. La correspondance est réalisée à l'aide d'un « classifieur » : il s'agit d'une fonction ou de règles de décision permettant de transformer un ensemble de données appartenant à R^n en des classes discrètes appartenant à N . Ces classificateurs sont nombreux et variés. La partie qui suit vise à exposer les différentes méthodes existantes, leurs caractéristiques, avantages et limites. Elles seront analysées dans le cadre de notre application afin de s'en inspirer tout en comprenant pourquoi elles n'aboutissent pas à de bons résultats dans notre contexte d'étude.

1.1 Aperçu des méthodes classiques de classification

On distingue plusieurs catégories de méthodes, qui possèdent chacune leurs propres avantages et inconvénients. Aucune ne maximise la performance sur tous les plans ni dans toutes les situations.

1.1.1 Méthodes ponctuelles ou objets

Il est possible de traiter chaque pixel de l'image indépendamment des autres (méthodes ponctuelles) ou de considérer un pixel dans son contexte, spatial la plupart du temps, et de travailler sur des objets. Ceux-ci peuvent découler d'une segmentation effectuée au préalable sur l'image, pour aboutir à des surfaces spectralement homogènes. Théoriquement, c'est ce que nous cherchons à obtenir : un type de culture, donc un type d'occupation du sol, pour chaque parcelle agricole. Notre approche conceptuelle du problème est donc de travailler sur des objets. En pratique, nous choisirons de développer dans un premier temps une méthode ponctuelle afin de minimiser les traitements appliqués à l'image. Ceci réduit le risque d'éventuelles erreurs relatives au choix de certains paramètres et permet d'exploiter toute l'information brute contenue dans les données.

1.1.2 Classifications supervisées/non supervisées

Lors d'une classification non supervisée, le classifieur détermine automatiquement les différentes classes de l'image à partir des caractéristiques spectrales des pixels. L'utilisateur n'influence pas la classification, si ce n'est éventuellement pour définir le nombre de classes ou pour choisir d'en regrouper certaines, et laisse parler les données par elles-mêmes.

Au contraire, dans le cas d'une classification supervisée, l'utilisateur intervient pour imposer une typologie existante au travers d'une phase d'apprentissage. Celle-ci nécessite des pixels dits « d'entraînements », dont les couples spectre-classes sont disponibles grâce à des campagnes de terrains ou par photo-interprétation. Ils permettent « d'entraîner » le classifieur, à savoir de déterminer certains de ses paramètres pour obtenir une fonction de décision à l'aide de laquelle seront classés tous les autres pixels de l'image dont la nature est inconnue [8].

L'utilisation de classifications supervisées ou non dépend principalement de la précision thématique recherchée et de l'hétérogénéité spectrale des classes. Les classifications non supervisées sont surtout efficaces pour discriminer de grandes classes thématiques (forêt, eau, urbain, cultures) [20] ou sont encore utilisées lorsque la typologie est mal définie a priori et qu'il est difficile de choisir des parcelles

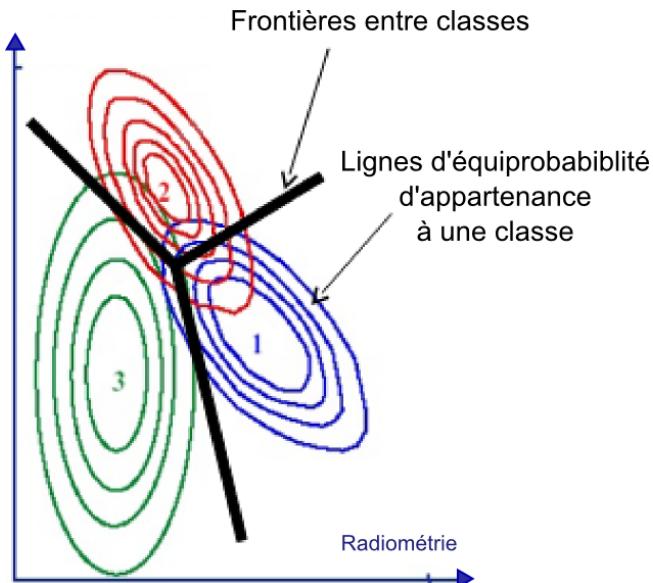


FIGURE 2 – Illustration d'une partition de l'espace à 2 dimensions obtenu par maximum de vraisemblance [9]

d'apprentissage qui donneront une partition correcte de la scène (classes hétérogènes car nuages irréguliers) [9]. Elles sont inappropriées pour notre étude dans laquelle nous cherchons à différencier un grand nombre de types de cultures à partir d'une typologie précise.

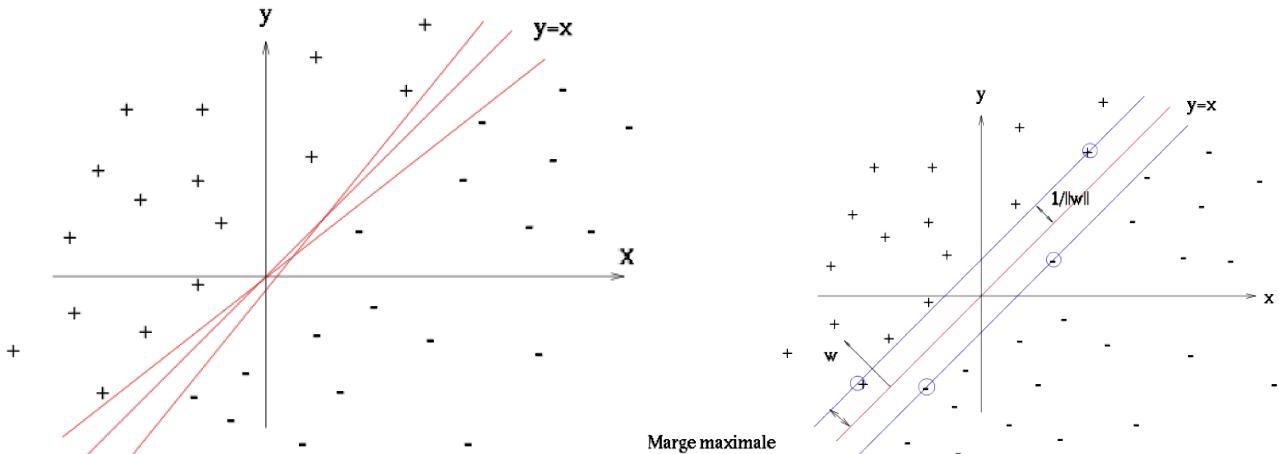
1.1.3 Méthodes paramétriques/non paramétriques

Se distinguent également les méthodes paramétriques des non-paramétriques. Dans le premier cas, la fonction discriminante suit une distribution statistique connue [6], et à chaque pixel est affecté un certain pourcentage de chance d'appartenir à une classe donnée. Peuvent être citées :

- La méthode du maximum de vraisemblance, basée sur une approche centroïde-distance. Celle-ci se rencontre aussi bien dans le cas des méthodes paramétriques que des non-paramétriques. Elle consiste en caractériser une classe par son centroïde, alors utilisé pour classer les pixels selon leur distance aux centroïdes de chaque classe. Dans le cas du maximum de vraisemblance, le calcul de probabilités gaussiennes effectué à partir de l'échantillon d'apprentissage aboutit à définir des frontières simples entre classes (Figure 2).
- SVM : Séparateur à Vastes Marges. Dans l'espace des signatures, il s'agit de séparer les classes grâce à des hyperplans dont les équations maximisent les « marges de séparation » entre classes (Figure 3), et donc le taux de bonne classification des échantillons d'apprentissage. Ces équations peuvent être définies grâce à des fonctions noyaux paramétriques, par exemple basées sur une gaussienne.

Au contraire, les méthodes non paramétriques ne considèrent que les relations entre individus, sans statistiques particulières, et la classification répond à une logique booléenne. Y figurent notamment les méthodes de seuillage et de minimisation de distance.

- Méthodes basées sur la minimisation de la distance : elles consistent en chercher la classe la plus proche de chaque pixel à partir d'un calcul de distance dont la notion de proximité dépend. Les algorithmes sont simples mais facilement sensibles au bruit. Ces méthodes restent pour la



Pour un ensemble de points linéairement séparables, il existe une infinité d'hyperplans séparateurs.

L'hyperplan optimal (en rouge) avec la marge maximale. Les échantillons entourés sont des vecteurs dits 'supports' (échantillons les plus proches de l'hyperplan)

FIGURE 3 – Illustrations de la recherche de l'hyperplan dans le cas de la méthode SVM [15]

plupart basées sur un modèle centroïde-distance, le centre de chaque classe correspondant à une moyenne ou à un barycentre de valeurs radiométriques. On peut ainsi citer les méthodes des parallélépipèdes, des k-plus proches voisins, de classification selon la distance de Mahalanobis et les méthodes itératives par k-means et par ISODATA.

— Méthodes de seuillage, aussi appelées arbres de décisions. Il s'agit de classer les pixels en fonction de seuils appliqués sur les valeurs des bandes spectrales ou de certains indices (NDVI, LAI, ..). Un expertise est nécessaire pour définir les seuils à établir, mais la classification est souvent efficace puisque les critères sélectionnés possèdent une réelle signification thématique (agrobiologique, ...).

Évidemment, quelques méthodes mixtes existent, combinant des approches paramétriques et booléennes. Par exemple, concernant la méthode du Random Forest, plusieurs arbres de décisions sont construits à partir de sous-ensembles de l'échantillon d'apprentissage. Lors de la classification, une classe est obtenue par chaque arbre, puis un vote majoritaire décide du résultat retenu pour chaque pixel [12].

Toutes les méthodes énumérées ci-dessus peuvent être aussi bien lancées à partir d'une image que de plusieurs, dans le cas de séries temporelles, afin de considérer un maximum d'informations. Dans ce cas, le nombre de dimensions est multiplié par le nombre d'images.

Dans notre cas d'étude, les cultures sont asynchrones donc les classes sont hétérogènes et difficilement séparables spectralement. L'utilisation d'une méthode de classification classique (et en particulier d'une méthode basée sur une approche centroïde-distance) à partir d'une seule image est difficile : des confusions apparaissent suite aux décalages phénologiques et de nombreuses classes sont nécessaires pour rendre compte de la complexité de la réalité (stade de développement, complantation) [9], sans parler de l'influence de la date de l'image à classer. L'exploitation d'une série de dates ne résout pas le problème pour ces méthodes puisqu'elles ne considèrent pas explicitement les cycles phénologiques des cultures et de nombreuses sous-classes restent nécessaires. Ainsi, dans une zone d'étude marocaine présentant des conditions plus ou similaires à la notre, des essais de classification par maximum de vraisemblance furent menés au CESBIO pour plusieurs configurations possibles de dates d'images. Dans tous les cas, la précision ne dépassa pas les 50% [7].

A partir de ces considérations, les méthodes classiques basées sur des approches centroïde-distance ou sur la définition d'hyperplans ne peuvent fonctionner. Il s'avère préférable d'utiliser des méthodes de seuillages définies à partir d'observations diachroniques.

1.2 Méthodes exploitant les séries temporelles NDVI

L'intérêt de l'utilisation des données NDVI multi-temporelles pour la cartographie d'occupation du sol a été démontré à plusieurs reprises [1,10].

Pour rappel, le NDVI (Normalized Difference Vegetation Index) est un indice de végétation construit à partir des bandes rouge et NIR (proche infrarouge) d'une image. Il exploite les caractéristiques en réflectance des végétaux, qui absorbent les longueurs d'ondes autour de 650 nm et réfléchissent plus ou moins celles dans le proche infrarouge (entre 700 et 900 nm) en fonction de leur activité chlorophyllienne. L'indice se définit comme :

$$NDVI = \frac{\rho_{ir} - \rho_r}{\rho_{ir} + \rho_r}$$

avec ρ_{ir} la réflectance dans la bande proche infrarouge et ρ_r la réflectance dans la bande rouge.

Les valeurs du NDVI sont comprises entre -1 et +1. Les valeurs négatives correspondent aux surfaces autres que les végétaux (neige, eau, nuages), et les valeurs avoisinant 0 correspondent aux sols nus pour lesquels les réflectances dans le rouge et le proche infrarouge sont du même ordre de grandeur. Les formations végétales présentent des valeurs positives, proportionnelles à la densité du couvert et à l'activité chlorophyllienne [2].

A partir d'une série temporelle d'images optiques, les valeurs NDVI d'un même pixel pour différentes dates peuvent être tracées en fonction du temps et permettre ainsi de rendre compte de l'évolution de la végétation.

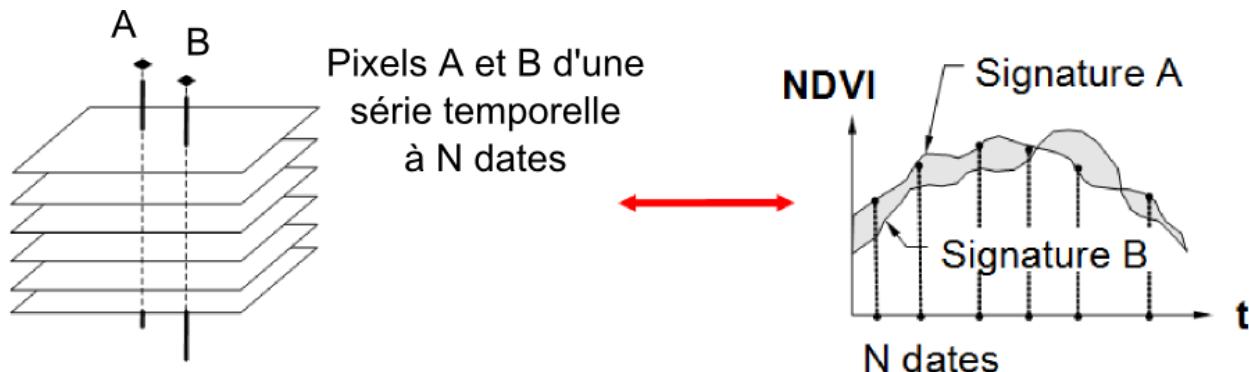


FIGURE 4 – Visualisation des signatures spectrales obtenues pour deux pixels à partir d'une série temporelle NDVI visible [9]

En ce qui concerne les cultures annuelles, le profil NDVI obtenu illustre, en théorie, les différents stades de développement de la plante (Figure 5 [10]) :

- (1) Début de croissance
- (2) Début de phase de maturité
- (3) Début de sénescence
- (4) Début de dormance.

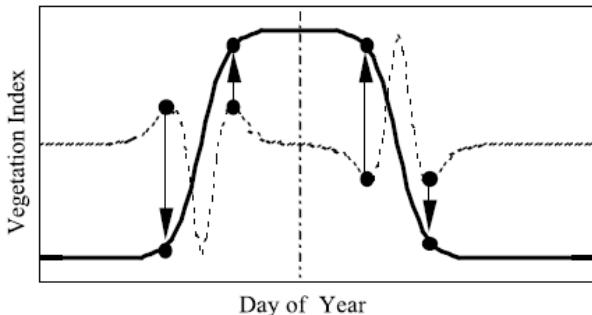


FIGURE 5 – Dates de transition remarquables d'un cycle de développement d'une culture annuelle, observées à l'aide d'un profil NDVI [10].

Le trait plein correspond à une série temporelle idéale d'un indice de végétation (pour nous le NDVI), et le trait pointillé sa dérivée seconde. Les points indiquent les différentes dates de transition remarquables.

La plupart des espèces végétales que l'on rencontre dans la nature présentent des cycles phénologiques relativement uniques. Elles bourgeonnent, atteignent leur stade de maturité puis de sénescence à approximativement la même date tous les ans, bien que des conditions météorologiques irrégulières puissent décaler ces cycles jusqu'à une trentaine de jours [1]. En agriculture, d'autres variables influencent les différentes dates de transitions observées, mais l'identification des natures de culture à partir de leurs séries temporelles NDVI a été démontrée à plusieurs reprises. Dans une étude consacrée aux grandes plaines centrales des États-Unis, Wardlow et Egbert [3] ont mené une combinaison d'analyses graphiques et statistiques sur des données NDVI issues d'une série temporelle MODIS (de 12 mois, avec 250 m de résolution spatiale et un temps de revisite de 16 jours) pour quelques 2000 champs cultivés. Les résultats obtenus démontrent la possibilité de détecter une unique signature multi-temporelle pour chacune des cultures majoritaires associées à certaines pratiques agricoles (cultures doubles, jachères, irrigation). Ces signatures sont cohérentes vis-à-vis de leurs caractéristiques phénologiques, et la plupart des classes sont spectralement séparables à un certain point de la période de croissance.

Furlan et al. [4] proposent une méthode de classification qui exploite la très haute résolution temporelle des images SPOT végétation en utilisant comme vecteurs caractéristiques les profils NDVI des images. Un profil temporel NDVI annuel est calculé pour chaque classe en moyennant les valeurs NDVI des zones d'entraînement, puis est utilisé comme base pour le processus de classification. Le profil NDVI lissé de chacun des pixels est comparé grâce à une distance de mesure (Coefficient de Jaccard et DTW) pour ceux disponibles pour chaque classe. L'utilisation d'une structure hiérarchique des classes basée sur le coefficient de Jaccard permet d'en regrouper certaines lorsque leur similitude limite leur distinction.

Cette méthode vise à permettre une mise à jour continue des cartes d'occupation du sol et s'inscrit dans une problématique similaire de celle à laquelle nous nous intéressons. Néanmoins, elle se base sur la représentation d'une classe par un unique profil moyen et n'est donc pas adaptée à la forte variabilité des cycles de développement des cultures que l'on peut rencontrer en Tunisie.

De précédents travaux menés au CESBIO [7], sur la plaine du Haouz au Maroc, ont aboutit à une méthode permettant de surmonter la difficulté liée à cette forte disparité des dates de développements. Elle s'appuie sur une approche arbre de décisions utilisant des critères appliqués aux profils NDVI des cultures. Après analyse visuelle des profils, 4 grandes classes thématiques sont retenues pour tenter d'être identifiées sur la base de critères phénologiques simples et robustes : sol nu, végétation annuelle, arbres sur sol nu et arbres sur culture annuelle. Les critères retenus pour la classification sont résumés

dans le tableau ci-dessous :

Conditions for NDVI values	Conditions for NDVI range	Class
All $NDVI < Sn$	No	Bare soil (SN)
Some $NDVI > Sn$ and some $NDVI < Sn$	No	Annual crops (H)
All $NDVI > Sn$	$Max - Min < Sr$ AND All $NDVI < Sa$	Trees on bare soil (ASN)
	$max - min > Sr$ or some $NDVI > Sa$	Trees + annual understory (AH)

TABLE 1 – Critères utilisés pour la classification des profils NDVI par la méthode développée au CESBIO [7], avec $Sn=0.18$ (seuil sol nu), $Sa=0.45$ (valeur de NDVI considérée comme maximale pour les arbres sur sol nu) et $Sr=0.15$ (amplitude minimale de NDVI pour les arbres sur cultures annuelles).

Les résultats obtenus (précision globale de 85 %) démontrent le potentiel de la méthode, et des séries temporelles en général, pour palier aux fortes disparités phénologiques par rapport auxquelles les autres méthodes plus classiques se révèlent peu performantes. En effet, des classifications par maximum de vraisemblance furent menées sur les mêmes jeux de données, pour plusieurs configurations d'images, et n'aboutirent à des résultats meilleurs que 50 % pour 4 grandes classes.

La performance de la méthode qu'ils proposent repose ici sur l'utilisation de critères robustes tels que la différence min/max qui permet de la rendre indépendante des dates d'images utilisées et de l'appliquer sur toute série temporelle présentant une distribution raisonnable des dates d'images.

Cependant, ces méthodes reposent toutes sur le même postulat qui consiste à ne considérer qu'une unique classe de culture par parcelle, ce qui est incompatible avec notre zone d'étude où on souhaite être capable d'identifier pour chaque pixel ses différentes successions culturelles. De plus, l'identification d'une culture à partir de sa série temporelle NDVI nécessite la connaissance des caractéristiques de son cycle phénologique [1] selon plusieurs facteurs (conditions météorologiques, pratiques agricoles, ...). Ainsi, il paraît intéressant de proposer une méthode de classification alternative basée sur une approche base de données : on espère pouvoir introduire un maximum d'expertise en intégrant des données relatives à différentes situations passées.

1.3 Evaluation de la classification

L'évaluation de la qualité de la classification constitue une étape clef de notre démarche dans la mesure où elle détermine la performance de la méthode développée ainsi que la significativité des résultats obtenus. Il convient de définir précisément les termes et méthodes qui seront employés par la suite.

1.3.1 Méthodes de validation

En ce qui concerne les méthodes de classifications supervisées, les données d'entraînements (échantillons d'apprentissage) fournissent des hypothèses que la classification extrapole [9]. Afin de valider cette extrapolation, un jeu de données indépendant des données d'entraînements doit être utilisé : les données de validation. Plusieurs méthodes existent pour définir cet échantillon à partir des données de référence [16].

Validation croisée

La validation croisée permet d'entraîner le classifieur puis d'estimer sa fiabilité grâce à différentes techniques d'échantillonages menées sur l'ensemble des données de référence. Parmi les différents procédés de validation croisée, on distingue :

- **Découpage des données** (Split validation). La plus facile à mettre en œuvre consiste à découper en deux échantillons distincts les données de référence. Le premier permet d'entraîner la classification (échantillon d'apprentissage) en estimant les paramètres du modèle, tandis que le second est utilisé pour évaluer la qualité du modèle ainsi généré. Cette méthode n'est efficace que lorsqu'on dispose d'un grand nombre de données : les données disponibles pour la construction du modèle sont réduites drastiquement ce qui aboutit parfois à une mauvaise estimation des paramètres.
- **K-fold validation.** Les échantillons d'apprentissage et de validation sont cette fois-ci tirés aléatoirement, et de taille différentes. L'ensemble de données de référence est divisé aléatoirement en k sous-ensembles D_j de même taille. L'estimation des paramètres de la classification est effectuée sur l'union des D_j avec j différent de i , tandis que la validation repose sur le sous-ensemble D_i . Tous les sous-ensembles générés sont successivement utilisés en tant qu'échantillons tests, alors que tous les autres permettent l'apprentissage de la classification. L'évaluation globale de la classification est calculée en moyennant les résultats obtenus pour chaque configuration. Cette méthode s'avère plus adaptée que la précédente dans le cas où l'on dispose de peu de données de références, puisqu'une proportion importante de ces données est utilisée pour entraîner le modèle. Cependant, le temps de calcul est élevé, et la méthode est sensible aux découpages effectués. Le choix du nombre de blocs k est primordial.
- **Leave-one-out.** Il s'agit d'un cas particulier de la k-fold validation dans lequel $k = n$, avec n le nombre de données de références disponibles. L'apprentissage se base sur $(n - 1)$ observations et la validation du modèle sur la n -ième observation. L'opération est répétée n -fois. L'avantage consiste ici à disposer de la configuration optimale d'entraînement.

Bootstrap

Dans ce cas, l'univers des données est exploré par tirages aléatoires d'échantillons de tailles variables. Pour plusieurs pourcentages p , sont tirés successivement de façon aléatoire n échantillons de taille $p\%$ des données de références. Chacun de ces échantillons est utilisé pour valider le modèle généré par le reste des données. Les résultats obtenus pour chaque pourcentage sont moyennés, et il en ressort l'influence du nombre de données utilisées pour l'apprentissage sur la qualité de la classification obtenue. Ensuite, du fait de l'utilisation de tirages aléatoires, toutes les données ne seront pas forcément exploitées au cours de la méthode, contrairement à la validation croisée, bien que l'on s'en rapproche pour un n très grand. Ceci permet de mieux rendre compte statistiquement de l'univers de nos données, en se rapprochant de la réalité, sans risque de biaiser les résultats.

Finalement, le choix des méthodes de validation utilisées dépend du nombre de données de références et des objectifs recherchés. Si certaines méthodes sont plus adaptées à la sélection de modèles en offrant la configuration optimale pour la phase d'entraînement (leave-one-out), d'autres permettent de souligner des aspects différents, comme l'influence du nombre de données utilisées pour l'entraînement grâce à la méthode du bootstrap.

Classement	Classe 1	...	Classe j	...	Classe c	Totaux
Réalité						
Classe 1	n_{11}		n_{1j}		n_{1c}	n_{1+}
...						
Classe i	n_{i1}		n_{ij}		n_{ic}	n_{i+}
...						
Classe c	n_{c1}		n_{cj}		n_{cc}	n_{c+}
Totaux	n_{+1}		n_{+j}		n_{+c}	N

TABLE 2 – Exemple théorique de matrice de confusion

1.3.2 Indices de performance

Lors du test de l'échantillon de validation, l'évaluation de la classification s'effectue classiquement grâce à une matrice de confusion. Son principe repose sur la comparaison des pixels (ou objets) de l'image issue de la classification et ceux de l'image servant de référence pour la validation [6].

De taille $nx n$, avec n correspondant au nombre de classes de la typologie employée, la matrice se présente la plupart du temps sous la forme d'un tableau à double entrées. Les colonnes présentent le résultat de la classification tandis que les lignes renseignent la vérité terrain. Les valeurs n_{ij} de la matrice correspondent aux occurrences des pixels classés en tant que classe j mais correspondant en réalité à la classe i (Figure 2).

Les totaux de chaque ligne et colonnes permettent de définir des indices d'évaluation qui apportent des informations quant à la qualité des résultats. Seront utilisés dans le cadre de notre étude :

- **La précision :** Pour une classe i donnée, elle correspond à la proportion d'éléments correctement classés par rapport au nombre d'éléments attribués à la classe i par le processus de classification. L'indice appelé User Accuracy (UA) estime la précision d'une classe par :

$$UA_i = \sum_{j=1}^c \frac{n_{ii}}{n_{ji}}$$

La précision globale d'une classification multi-classe s'obtient par la moyenne des précisions de chaque classe :

$$UA = \frac{1}{r} \sum_{i=1}^c UA_i$$

- **Le rappel :** pour une classe i donnée, il correspond à la proportion d'éléments correctement classés par rapport au nombre d'éléments appartenant à la classe i dans les données de référence. Le rappel d'une classe est estimé grâce à l'indice Producer Accuracy (PA) :

$$PA_i = \sum_{j=1}^c \frac{n_{ii}}{n_{ji}}$$

Le rappel global d'une classification multi-classes correspond alors à :

$$PA = \frac{1}{r} \sum_{i=1}^c PA_i$$

-
- Un autre indice s'attache à évaluer la performance globale de la classification, sans orientation particulière vis-à-vis d'une classe : le **Pourcentage de Pixels Correctement Classés** (PCC) ou Overall Accuracy (appelé 'précision globale' par la suite).

$$OA = \frac{\sum_{i=1}^c n_{ii}}{\sum_{i=1}^c \sum_{j=1}^c n_{ij}}$$

- **L'indice kappa** évalue la matrice de confusion par rapport à une matrice dont les valeurs seraient tirées aléatoirement. Une valeur de 0 indique une concordance similaire à celle obtenue aléatoirement, tandis qu'une valeur de 1 correspond à une exactitude parfaite. Pour un kappa=0.75, 75 % des pixels ne sont pas classés au hasard.

$$Kappa = \frac{\sum_{i=1}^c (N * n_{ii} - n_{+i} * n_{i+})}{N^2 - \sum_{i=1}^c (n_{+i} * n_{i+})}$$

2 Présentation des données disponibles

2.1 Données satellitaires

2.1.1 Caractéristiques “brutes”

La série temporelle utilisée pour le test et la validation de la méthode est constituée de 14 images acquises par SPOT 5 entre novembre 2012 et novembre 2013 (Figure 6). Les images sont de haute résolution spatiale : 5 m en panchromatique et 10 m en multi spectral. Etant donné que nous n'exploiteront que les valeurs NDVI des pixels, on se limitera à une résolution de 10 m. Ainsi, ces données sont particulièrement adaptées pour simuler les images acquises prochainement par Sentinel-2, dont la résolution spatiale sera de 15 m. La largeur minimale des parcelles agricoles de notre zone avoisine les 50 m, et cette résolution est suffisante pour tenter de produire une carte d'occupation du sol pertinente.

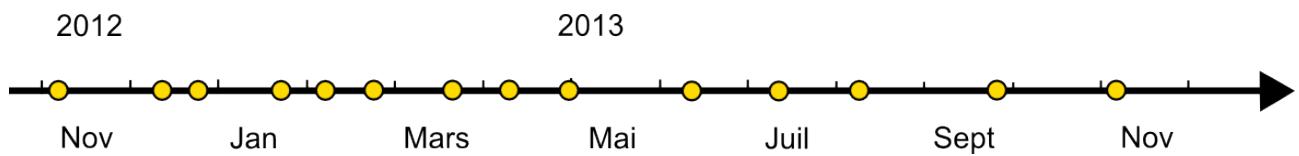


FIGURE 6 – Répartition temporelle des images de la série

2.1.2 Corrections radiométriques

A partir des données brutes enregistrées par les capteurs, qui mesurent la quantité d'énergie réfléchie par des objets, deux mêmes objets peuvent ne pas présenter les mêmes valeurs radiométriques en fonction de la date d'observation. En effet, la présence de gaz et de poussières dans l'atmosphère, qui peuvent réfléchir ou absorber certaines longueurs d'ondes, modifie les propriétés des rayonnements solaires incidents et réfléchis [18]. On rencontre deux phénomènes : les objets observés ne sont pas tous éclairés sous le même angle par la lumière solaire, et le rayonnement qu'ils réfléchissent est perturbé au cours de sa traversée vers l'atmosphère en direction du capteur. De plus, le traitement électronique du rayonnement reçu par les capteurs n'est pas idéal et nécessite un certain étalonnage. A partir de

ce constat, il est indispensable de procéder à des corrections radiométriques sur chacune des images acquises, afin de pouvoir comparer entre elles les différentes informations qu'elles contiennent.

Deux principaux types de méthodes existent. Les premières utilisent des modèles mathématiques et physiques qui décrivent les principales interactions des rayonnements incidents et réfléchis. Ils sont performants mais nécessitent de connaître l'état de l'atmosphère aux dates et lieux de prise de vue, ce qui limite leur utilisation. D'autres approches se basent sur l'utilisation de pixels de référence de l'image pour lesquels la radiométrie est connue [19].

Dans notre cas, les images ont été corrigées au CESBIO préalablement au stage par une méthode hybride. Un modèle physique de transfert radiatif est utilisé dans un premier temps pour éliminer les erreurs les plus grossières. Une méthode statistique de correction par invariant vient affiner les résultats. Celle-ci repose sur le postulat qui consiste en considérer que la radiométrie des invariants varie linéairement d'une date à l'autre. De là, après identification de pixels invariants, on peut calculer à partir de leur valeurs, par régression linéaire, des coefficients de correction à appliquer à chaque canal de chaque image.

2.2 Données terrain

Quatre campagnes de terrain ont eu lieu durant la période couverte par nos images. Elles permettent de récolter des informations de références pour initialiser puis valider les algorithmes. On dispose d'un fichier de polygones de type shapefile (ArcGIS, *.shp) par campagne, qui contient entre autres, pour chaque parcelle : un identifiant, une nature de culture, des observations complémentaires, des distances interrang et intrarang, le pourcentage de couverture, ainsi que la hauteur de végétation et le diamètre pour les arbres. Les différentes campagnes sont menées à travers toute la zone d'étude (Figure 7) dans le but d'intégrer une certaine variabilité spatiale. Plusieurs parcelles indépendantes sont observées par classe, dont le nombre varie en fonction de l'abondance de la classe. Seules celles respectant une taille minimale cohérente avec la résolution de l'image sont identifiées.

QGIS est utilisé pour créer un nouvel attribut dans les données de type shapefile. Celui-ci est obtenu par concaténation de l'identifiant unique de chaque parcelle observée, et de la nature de l'occupation du sol renseignée. Ceci permettra de faciliter le traitement des données pour la suite, dont la nature de culture apparaîtra directement dans le nom (Figure 7). Ensuite, chaque polygone est convertit en une ROI sous ENVI. Ceci permet d'accéder à des statistiques de données spectrales pour chaque parcelle identifiée, calculées par ENVI à partir de la moyenne des valeurs spectrales de chaque pixel la composant.

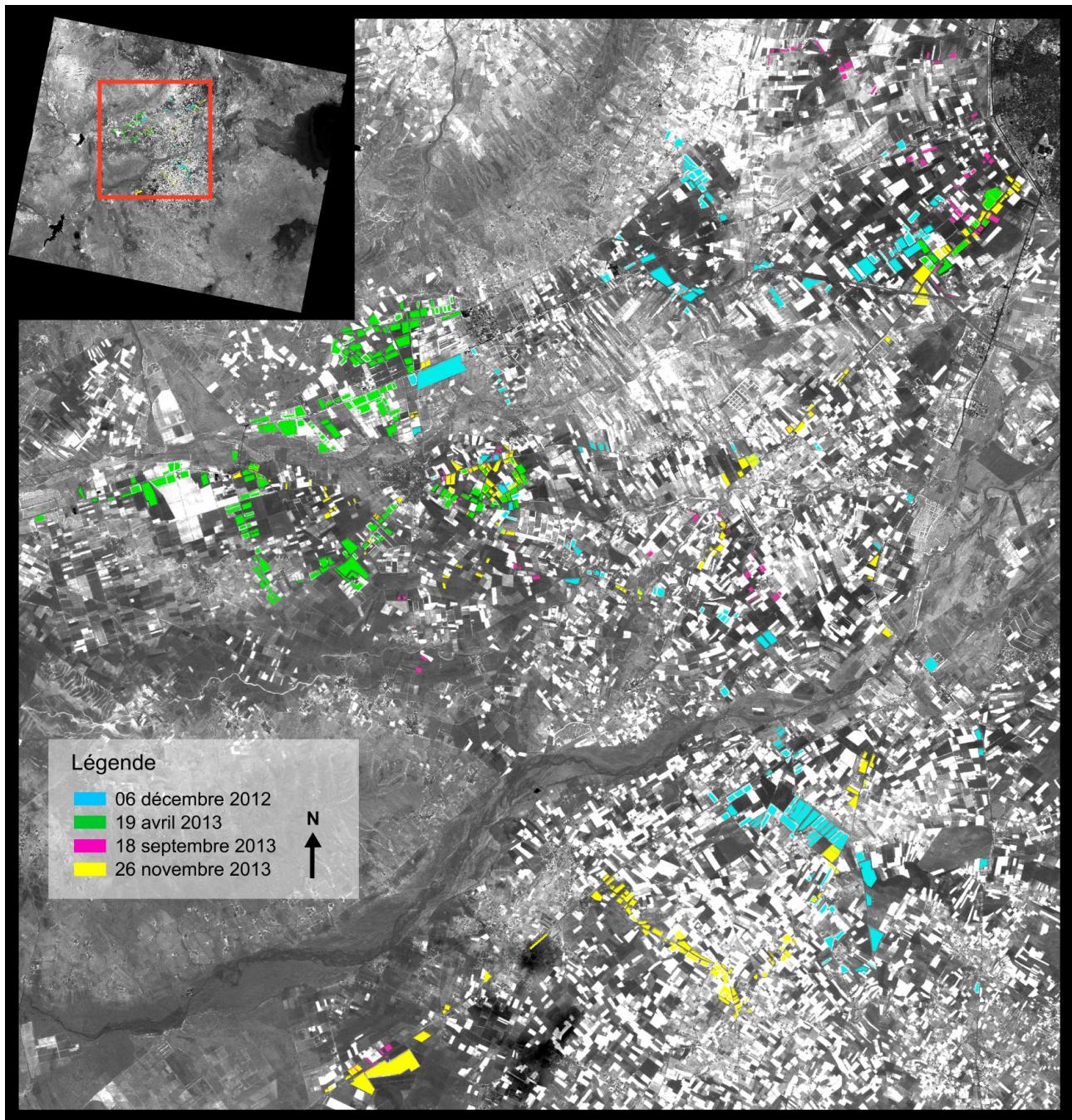


FIGURE 7 – Répartition spatiale des différentes campagnes de terrain

Troisième partie

Méthodologie de développement

Dans la partie précédente, nous avons insisté sur le fait que les méthodes classiques de classification d'occupation du sol, et en particulier les méthodes statistiques basées sur une approche centroïde-distance, sont inappropriées pour notre étude. De là, nous avons évoqué l'intérêt d'envisager des méthodes par seuillage sur la base de véritables critères biophysiques, et avons vu que si certaines méthodes existent pour exploiter le potentiel des séries temporelles NDVI, aucune n'est parfaitement adaptée à notre cas d'étude. Nous proposons donc dans la partie qui suit de tenter de développer une méthode qui puisse répondre à nos besoins. Nous présenterons tout d'abord le concept général visant à être mis en place, puis tenterons de le mettre œuvre à partir des données dont nous disposons, aussi bien satellitaires que de terrain.

1 Présentation générale de la méthode

Soit une série temporelle, de n images couvrant une période temporelle T , à classer. On cherche à obtenir pour chaque pixel les différentes occupations du sol qui se sont succédées au cours de cette période T (Figure 8).

On choisit pour cela de travailler sur les profils temporels NDVI qui présentent l'avantage de rendre compte de réels paramètres biophysiques. Le profil spectral NDVI de chacun des pixels est construit à partir de la série temporelle pour servir de base à la classification (1).

Il convient ensuite d'extraire pour chacun de ces profils les paramètres qui nous intéressent, à savoir les dates de début de croissance et de fin de sénescence des cultures qui s'y succèdent. Ceci permet une première segmentation temporelle du profil sous forme de cycles culturaux qu'il nous faut ensuite identifier (2).

Pour cela, une approche base de données est utilisée. Chaque cycle de développement de chaque profil est comparé à des données de référence en vue d'identifier sa nature (3). La comparaison repose aussi bien sur des critères de similitude de forme que de synchronisation temporelle. Il est indispensable de considérer non seulement l'allure du profil, caractéristique d'un type de culture, mais aussi d'étudier sa cohérence vis-à-vis du calendrier cultural (Figure 9).

Pour l'identification, on se base sur une collection de cycles de culture de référence. Ils doivent être suffisamment nombreux pour rendre compte de la variété des situations existantes : diversité de formes et de dates d'émergence au sein d'une même nature de culture suite à des conditions météorologiques variables ou à l'emploi de pratiques agricoles différentes. Idéalement, ces données de références sont issues de différentes campagnes de terrain menées sur plusieurs années. Une fois la base de données suffisamment étoffée et représentative, la méthode peut fonctionner de façon automatique sur plusieurs années sans recours à de nouvelles données d'apprentissage.

Finalement, après comparaison puis identification de chaque profil NDVI segmenté, on obtient une succession de nature de culture pour chaque pixel des images de la série temporelle à classer. Conceptuellement, cela revient à obtenir une superposition de cartes d'occupation du sol pour différentes dates.

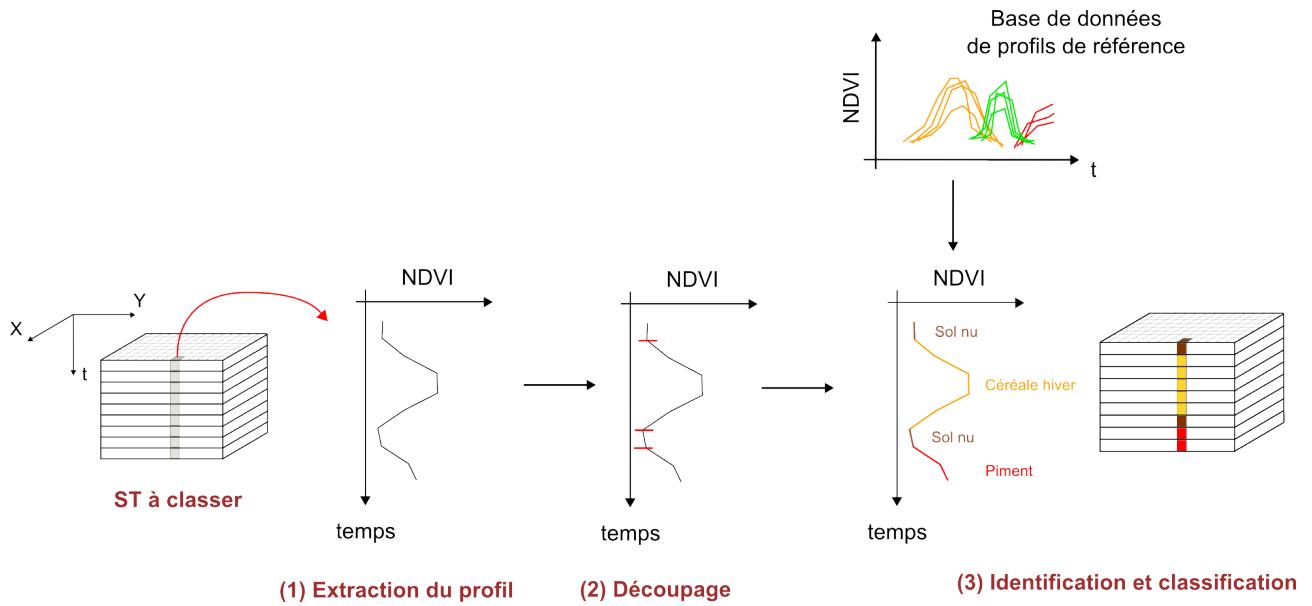


FIGURE 8 – Schéma du fonctionnement général de la méthode

2 Démarche de développement

L'enjeu du développement de la méthode est de la rendre la plus automatique possible afin de limiter le recours aux observations de terrain. On cherchera à s'appuyer sur des critères robustes pour assurer sa fiabilité en situations variées. Nous avons réalisé pour cela un travail d'exploration des profils visant à définir les critères sur lesquels sera basé l'algorithme.

2.1 Visualisation et analyse des données à disposition

2.1.1 Construction graphique

Le meilleur moyen d'acquérir une vue d'ensemble des données à notre disposition reste de toutes les visualiser. Si cette tâche apparaît fastidieuse et chronophage, elle se révèle indispensable pour mener à bien une réflexion précise.

ENVI 4.4 permet d'exporter sous forme de fichiers .txt le listing des valeurs NDVI de chaque date de la série temporelle pour chacune des ROI, ainsi que leur identifiant et a fortiori leur nature (cf partie 1.1). Un script que l'on développe sous Matlab permet de lire ce fichier et d'enregistrer les données dans une structure. Les valeurs NDVI de chacune des ROI sont interpolées linéairement et l'on représente à échelle graphique fixe les profils obtenus en vue de faciliter leur interprétation. Sur l'image du 21 janvier 2013, un masque a été appliqué pour indiquer la présence de nuages en affectant une valeur de NDVI=0 aux pixels concernés. Pour ces derniers, l'interpolation linéaire des valeurs de NDVI ne tient pas compte des valeurs de l'image de janvier.

Il faut noter que les valeurs de NDVI de nos images ont été ré-échelonnées : on leur ajoute 1, puis les multiplie par 100. Ainsi, les valeurs sont comprises entre 0 et 200, au lieu de l'être entre -1 et 1, ce qui permet de faciliter leur manipulation en évitant le type float et en les codant sur 1 octet.

Une ligne rouge symbolise le seuil NDVI=0.2 (donc 120 sur nos graphiques), considéré comme la limite entre sol nu et végétation active photosynthétiquement, et un marqueur (en vert) indique la date d'identification de la culture. Un regroupement des profils par classe vise ensuite à mettre en exergue leurs spécificités (Figure 10, 11, 12, 13 et 14).

Cet affichage permet d'analyser les différentes variations de NDVI observées. S'attacher à comprendre dans un premier temps l'évolution de chaque profil permet de faire émerger par la suite les

	janvier	fevrier	mars	avril	mai	juin	juillet	aout	septembre	octobre	novembre	decembre
Semi melon tunnel												
Melons avant saison		lab	plan			rec						
Semi maraichage saison												
melons			lab	plan		rec	rec					
Pastèque					lab	plan				rec		
Tomate été					lab	plan				rec		
Piments été						plan			rec			
Semi maraichage arrière saison												
Tomates arrière saison							lab	plan		rec		
Piments arrière saison							lab	plan		rec		
Fèves	rec	rec	rec						sem			
Oliviers	taille								rec	rec	taille	
Amandiers						rec						
Avoine en sec					rec				lab	sem		
Orge en sec					rec					lab	sem	
Blé en sec	sem					rec					lab	
Fourrages				coupe								
Paturage betail												

FIGURE 9 – Calendrier cultural ('lab' : labouré, 'plan' : planté, 'rec' : récolté, 'taille' : taillé, 'sem' : semé).

tendances globales qui se dessinent parmi les différentes espèces.

2.1.2 Visualisation des profils par classe

Nous nous attachons ici à analyser les profils de chaque classe, afin de tenter d'en dégager des caractéristiques propres à chacune d'entre elles.

Une cinquantaine de noms de classes apparaissent dans les données de terrain. Toutefois, certains correspondent à l'identification de mélanges de cultures difficilement exploitables (olivier + piment, ...) ou encore à des cultures très minoritaires pour lesquelles seulement quelques parcelles isolées sont identifiées. Nous nous concentrerons ici sur les cultures principales. Un calendrier cultural élaboré précédemment à partir d'observations de terrain nous aide à guider la réflexion (Figure 9).

- * **Piment** : A partir des profils NDVI des piments identifiés lors des campagnes de terrain 1, 3 et 4 (06/12/12, 19/09/13 et 26/11/13), on remarque que les dates de plantation s'étalement entre avril et septembre. De même, les dates de récoltes varient entre septembre et novembre. Ceci est cohérent avec le calendrier cultural et on retrouve la distinction qui avait été faite entre piments d'été et d'arrière saison. Outre des variations d'amplitude, l'allure des profils est globalement la même.
- * **Céréale d'hiver** : Elles semblent principalement plantées en janvier et récoltées en juin, même si on observe quelques décalages d'un ou deux mois concernant la date de semis.
- * **Fourrages** : Une plus grande variabilité dans la forme des profils s'observe par rapport aux classes 'piment' et 'céréale hiver'. Certains profils présentent une cycle de développement semblable aux céréales d'hiver, tandis que d'autres ne franchissent pas le seuil sol nu et gardent des valeurs élevées tout au long de l'année. Il est possible que l'on soit amené à distinguer par la suite le fourrage 'permanent' du fourrage 'annuel'.
- * **Oignons** : De même que pour les céréales d'hiver, les dates de semis varient d'un mois ou deux, et les profils présentent des pentes variables.
- * **Tomates** : Comme pour les piments, on distingue les tomates d'été, plantées aux alentours de mai et récoltées en juillet, des tomates d'arrière saison, plantées en septembre et récoltées en novembre.
- * **Pastèques** : Leurs cycles de développement s'avèrent particulièrement bien synchronisés par rapport aux autres types de cultures, entre fin avril et début juillet.

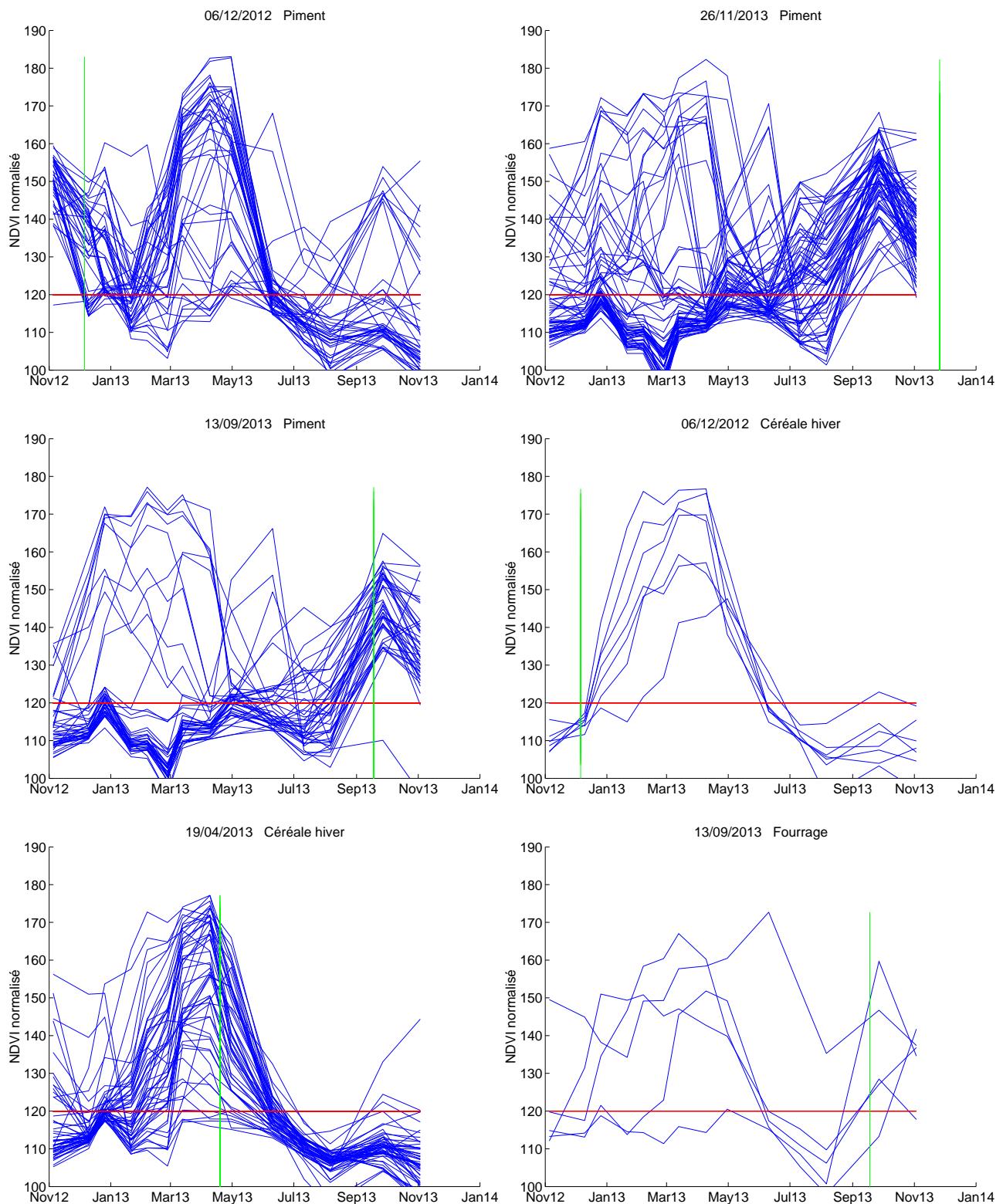


FIGURE 10 – Exemple de superposition de profils par classe et par campagne de terrain (1)

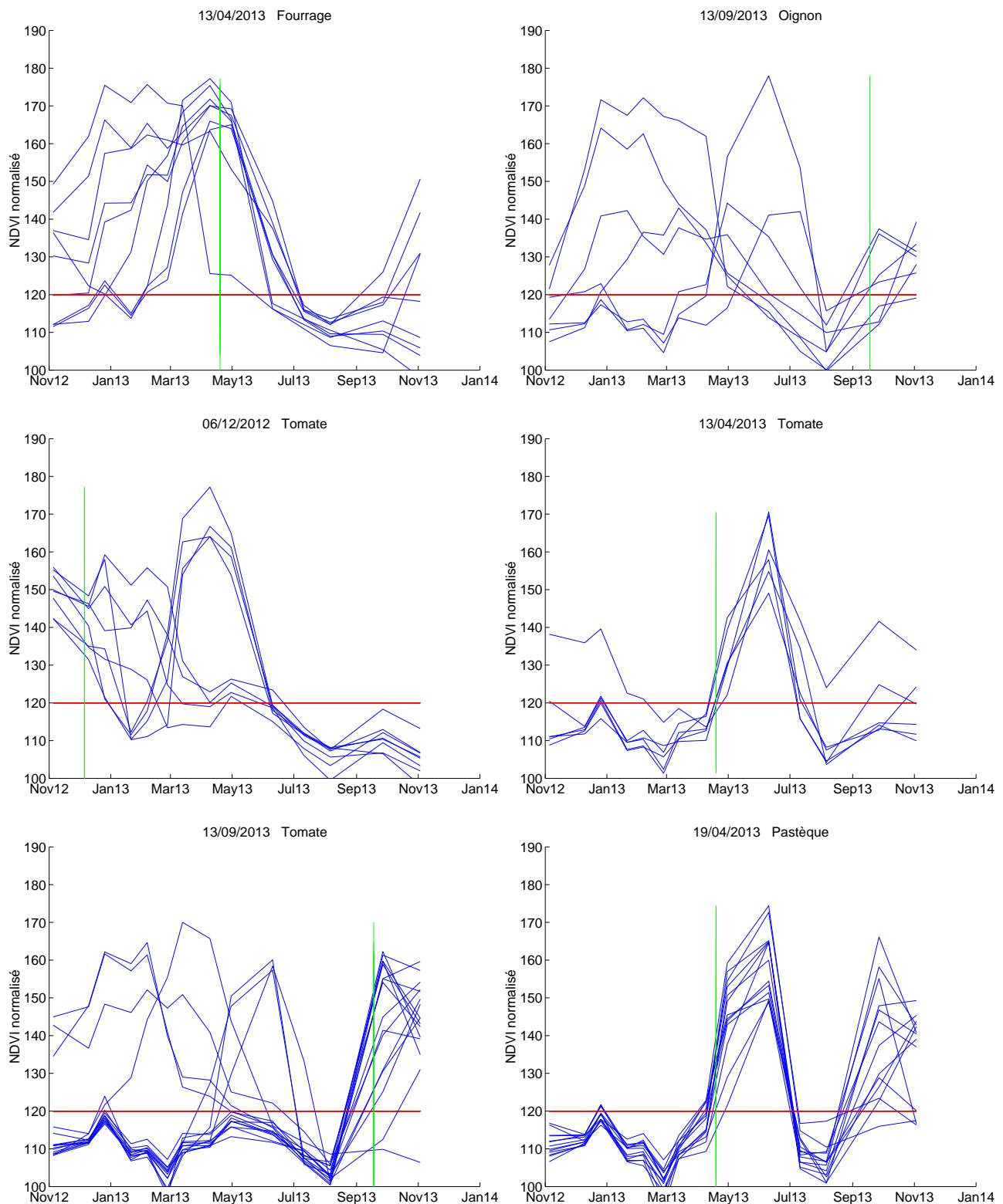


FIGURE 11 – Exemple de superposition de profils par classe et campagne de terrain (2)

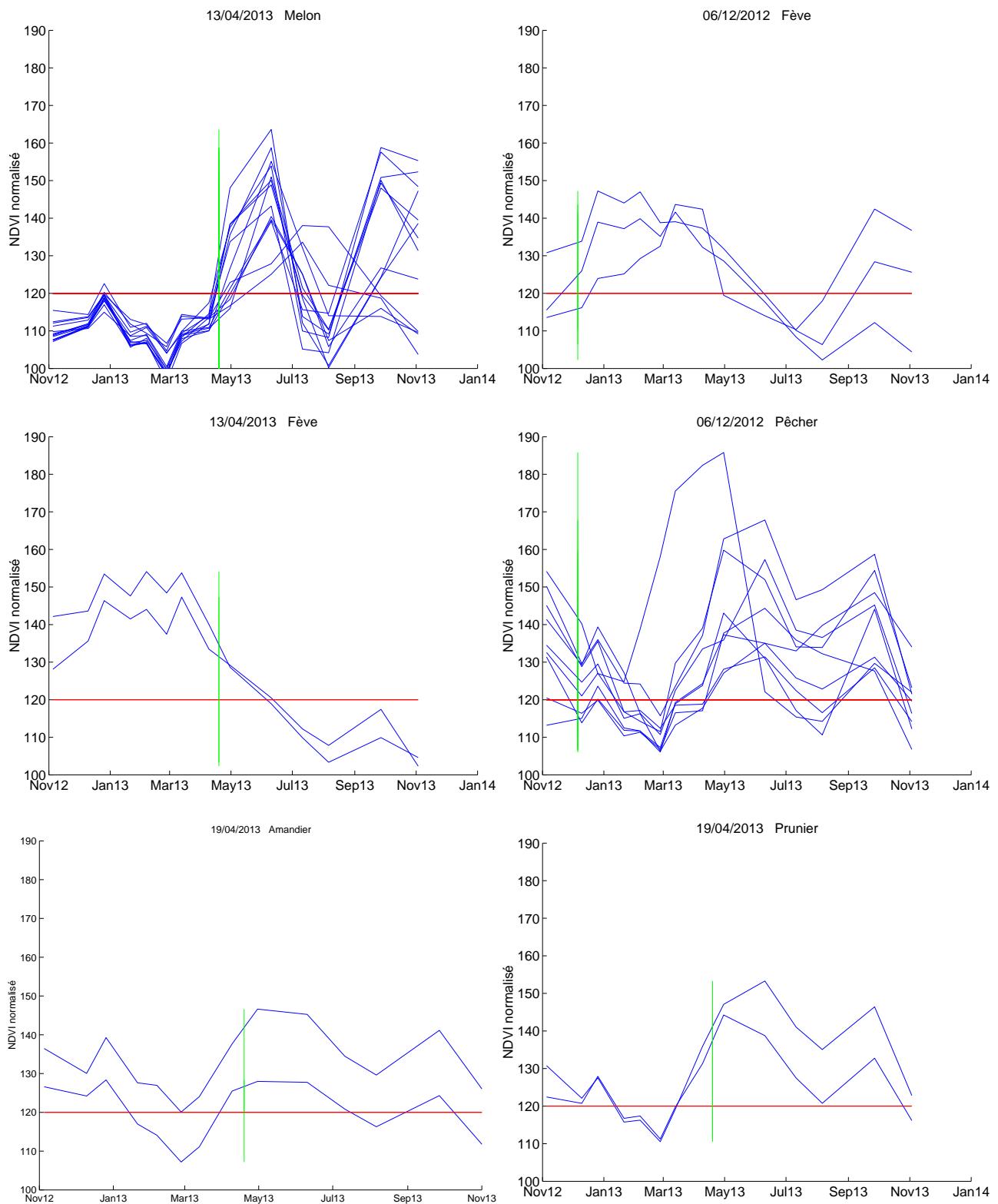


FIGURE 12 – Exemple de superposition de profils par classe et campagne de terrain (3)

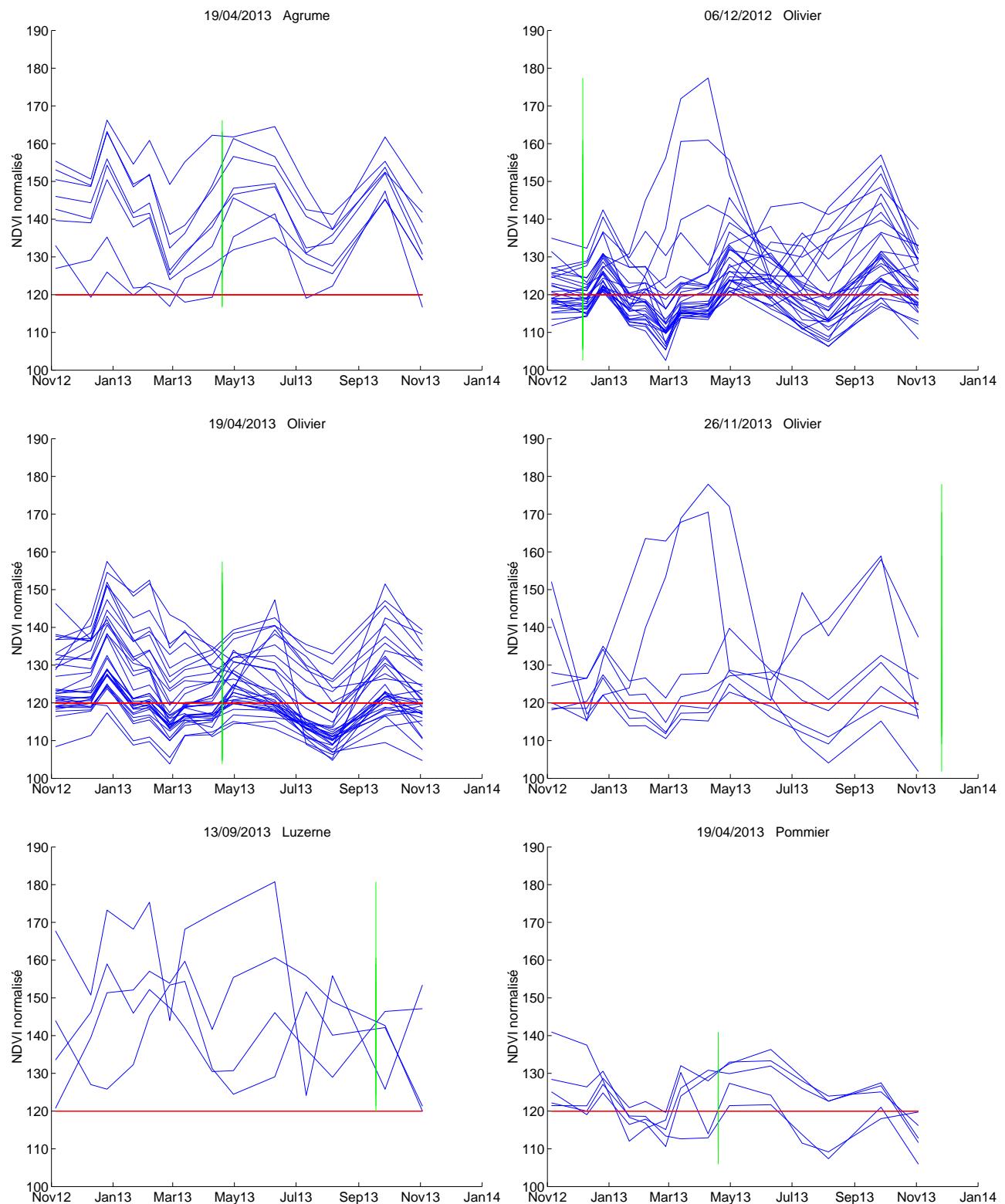


FIGURE 13 – Exemple de superposition de profils par classe et campagne de terrain (4)

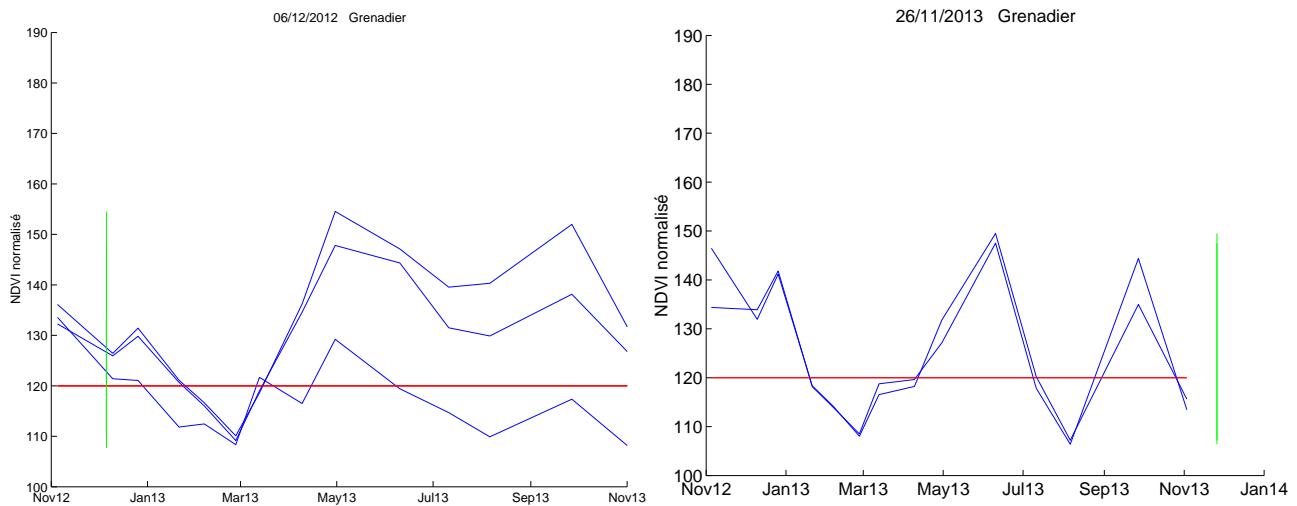


FIGURE 14 – Exemple de superposition de profils par classe et campagne de terrain (5)

- * **Melons** : Globalement, leur comportement rejoint celui des pastèques.
- * **Fèves** : Leur cycle cultural dure environ 6 mois, entre septembre et mars ou avril. L'allure du profil est caractéristique : un palier atteint autour de la valeur $NDVI=0.4$ succède pour plusieurs mois à la phase de croissance.
- * **Oliviers** : Globalement, on observe une tendance générale d'oscillation des profils autour d'une valeur moyenne proche de 0.2. Quelques profils dépassent la valeur de 0.6 et présente un véritable cycle de développement. Il peut s'agir de cultures plantées sous les rangs des oliviers et dont le signal détecte leur présence.
- * **Pêchers** : Mis à part un cycle de développement apparent qui relève d'une erreur d'identification, on remarque sur les données issues de la campagne de terrain de décembre 2012 un motif commun aux différents pêchers. Les valeurs de NDVI chutent en novembre pour franchir le sol de 0.2, puis remontent en mars pour osciller autour d'une valeur moyenne avoisinant les 0.3 entre mai et novembre. Cette allure s'explique par la chute des feuilles en automne suivie de la floraison printanière.
- * **Amandiers, pruniers, pommiers, grenadiers** : Les mêmes observations générales que pour les pêchers peuvent être faites. La valeur moyenne de NDVI atteinte en été par les pommiers semble légèrement plus faible que pour les autres espèces, et deux grenadiers identifiés en novembre 2013 présentent un comportement différent. Leur profil chute à une valeur de 0.1 en juillet, pour remonter à 0.4 en octobre. Ceci peut provenir de pratiques agricoles différentes (coupes, ..). De plus, les profils NDVI de ces deux parcelles de grenadiers se superposent presque en tous points : après vérification, ces parcelles s'avèrent être voisines géographiquement, et les chances sont fortes pour que les cultures soient menées par le même agriculteur. Les données ne sont donc pas indépendantes et l'on ne considère pour la suite que l'une des deux.
- * **Agrumes** : Les valeurs NDVI des profils oscillent autour de valeurs moyennes, supérieures à celles observées dans le cas des oliviers.
- * **Luzernes** : Les profils semblent totalement désynchronisés entre eux, même s'ils présentent quelques caractéristiques communes. Aucun d'eux ne franchit le seuil sol nu, et leurs valeurs oscillent plus ou moins périodiquement.

2.1.3 Analyses transversales

Distinction entre cultures annuelles et permanentes

On distingue facilement les signatures spectrales des cultures annuelles de celles des cultures permanentes (pêcher, prunier, olivier, agrume, ...).

Les cultures annuelles présentent une variation de NDVI rapide et de forte amplitude. Le cycle de développement, d'une période maximale de 5 mois, est précédé puis suivi d'une période de sol nu. Au contraire, les cultures permanentes affichent en général des motifs de variations pseudo-périodiques d'amplitudes plus faibles pour la majorité des classes. Ces nettes différences de signatures spectrales laissent envisager des traitements spécifiques à chacune de ces deux catégories de cultures lors de la définition de l'algorithme.

Au-delà de ces tendances générales, d'autres regroupements semblent se profiler parmi les cultures permanentes. Il ressort de l'analyse visuelle des profils des caractéristiques communes aux :

- **arbres à feuilles persistantes** : y figurent dans notre cas d'étude les oliviers, qui affichent peu de réponse en NDVI de part leur faible activité photosynthétique. Les variations de NDVI sont de très faibles amplitude dans la majorité des cas (inférieure à une variation de NDVI de 0.3) et oscillent autour d'une valeur moyenne. Des motifs périodiques apparaissent nettement mais il reste difficile de conclure avec certitude sur leur origine ce qui rend délicate leur exploitation.
- **arbres à feuilles caduques** : pour toutes les classes, le profil spectral rend nettement compte de leur cycle annuel (grenadier, pommier, prunier, amandier, pêcher, ...). La chute des feuilles observée durant l'automne se traduit par un passage du NDVI en dessous du seuil de sol nu, tandis que la reprise d'activité de la photosynthèse liée à la floraison ramène les valeurs à leur maximum. Se dégage une légère chute du profil en été, entre juin et août. Une fois encore, il est dangereux de conclure quant à sa cause. Plusieurs hypothèses sont envisageables :
 - (a) Les fortes températures recensées chaque année en été dans la zone peuvent provoquer un stress hydrique qui altère la réponse en NDVI de la plante.
 - (b) Une erreur persiste dans la correction des images en réflectance de notre série temporelle et un biais est observé pour les dates correspondant à la chute du signal (11/07/2013 et 06/08/2013). En effet, les corrections radiométriques appliquées en amont ne sont jamais parfaites et peuvent dans certains cas altérer légèrement les valeurs de NDVI observées. Concernant notre jeu d'images, des biais nets s'observent dans les valeurs entre les images du 10 décembre 2012 et du 06 février 2013 (Figure 15), où les profils présentent des variations de NDVI de même amplitude exactement (profils anormalement parallèles). Si l'influence des corrections radiométriques est ici évidente, elle l'est moins dans le cas de la chute de NDVI observée en été pour les arbres à feuilles caduques mais reste une hypothèse envisageable.
 - (c) Cette réponse spectrale des arbres rend compte d'un phénomène biologique commun à ces espèces.

Il est évident qu'une combinaison de plusieurs de ces trois facteurs n'est pas à exclure voire probable. Toutefois, ce phénomène ne peut pas être exploité directement pour caractériser ces arbres étant donné l'incertitude de son origine. Si les hypothèses (a) et (b) s'avéraient vraies, cette chute de NDVI dépendrait non seulement des conditions météorologiques de l'année étudiée mais également du jeu de données utilisé en entrée (par le biais des coefficients de correction radiométrique). On ne peut en extraire de critères robustes pour notre méthode.

Distinction inter et intra-classes

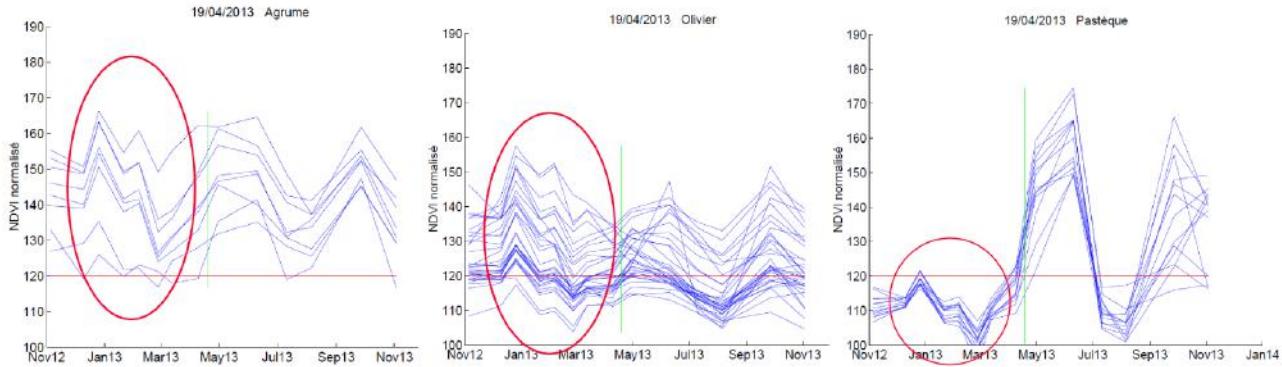


FIGURE 15 – Biais observés dans les valeurs de NDVI

Inter-classes Visuellement, la majorité des classes se distinguent facilement entre elles grâce à leurs dates de développement et leurs variations caractéristiques de NDVI (forme et amplitude du signal). Certaines confusions apparaissent concernant les couples de classes piment/tomate et melon/pastèque mais aussi concernant la nature exacte des arbres à feuilles caduques (pêcher, amandier, prunier, abricotier, pommier, grenadier). Des regroupements de classes seront sans doute nécessaires.

Intra-classes S'il est globalement possible de dégager des caractéristiques propres à chaque classe, une analyse plus détaillée des profils révèle une variabilité intra-classe à ne pas négliger. Ces disparités peuvent provenir de :

- **De la rugosité des surfaces** : Au-delà du simple pourcentage de couverture végétale, l'orientation spatiale des rangs de culture et leurs caractéristiques de surfaces influent le signal enregistré. Lorsque les surfaces sont rugueuses¹ la réflexion est diffuse et le rayonnement est réfléchi dans toutes les directions, avec généralement une direction privilégiée pour laquelle la réflexion apparaît plus importante [21]. Cette direction diffère selon la taille des cultures (hauteur sur largeur), l'orientation des rangs, l'orientation du capteur et du soleil. L'irrégularité de certains profils peut en découler (Figure 16).
- **De l'espacement des cultures au sol** : En télédétection, on observe successivement la strate arborée, la strate arbustive, la strate herbacée puis le tapis végétal et les résidus végétaux à la surface du sol. Les strates arborée et arbustive sont observées de manières différentes selon la densité du couvert [13] qui résulte de la distance inter-rangs et du diamètre des arbres. En effet, une espèce influe la réponse en NDVI si elle représente plus de 20 % de la couverture. Ensuite, plus le couvert est dense, moins la réponse spectrale sera affectée par la réflectance des sols et de la strate herbacée (adventice, ...). Les parcelles ROI 48 (Figure 17, gauche) et 50 (Figure 17, droite) de la campagne de terrain d'avril 2013 abritent des oliviers avec des distances inter-rangs de respectivement 10x10 et 3x3 m. Dans le permier cas, seule la réponse spectrale de la strate herbacée est détectée.
- **D'un mélange de cultures** (Figure 18). Il n'est pas rare d'observer dans la zone des parcelles qui abritent des mélanges de cultures. Pour la plupart du temps, il s'agit d'arbres plantés parmi des cultures annuelles afin d'optimiser les irrigations (Annexe 1). Cependant, comme expliqué précédemment, les différentes strates de végétation influent sur le signal enregistré par

1.

— Une surface est dite rugueuse lorsqu'elle présente des aspérités dont la taille est supérieure à la longueur d'onde du rayonnement incident

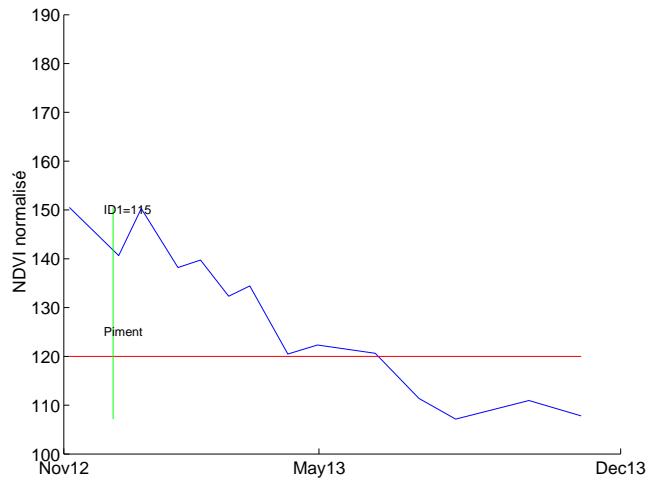


FIGURE 16 – Exemple de profil de piment dont les variations de NDVI relèvent potentiellement de sa rugosité

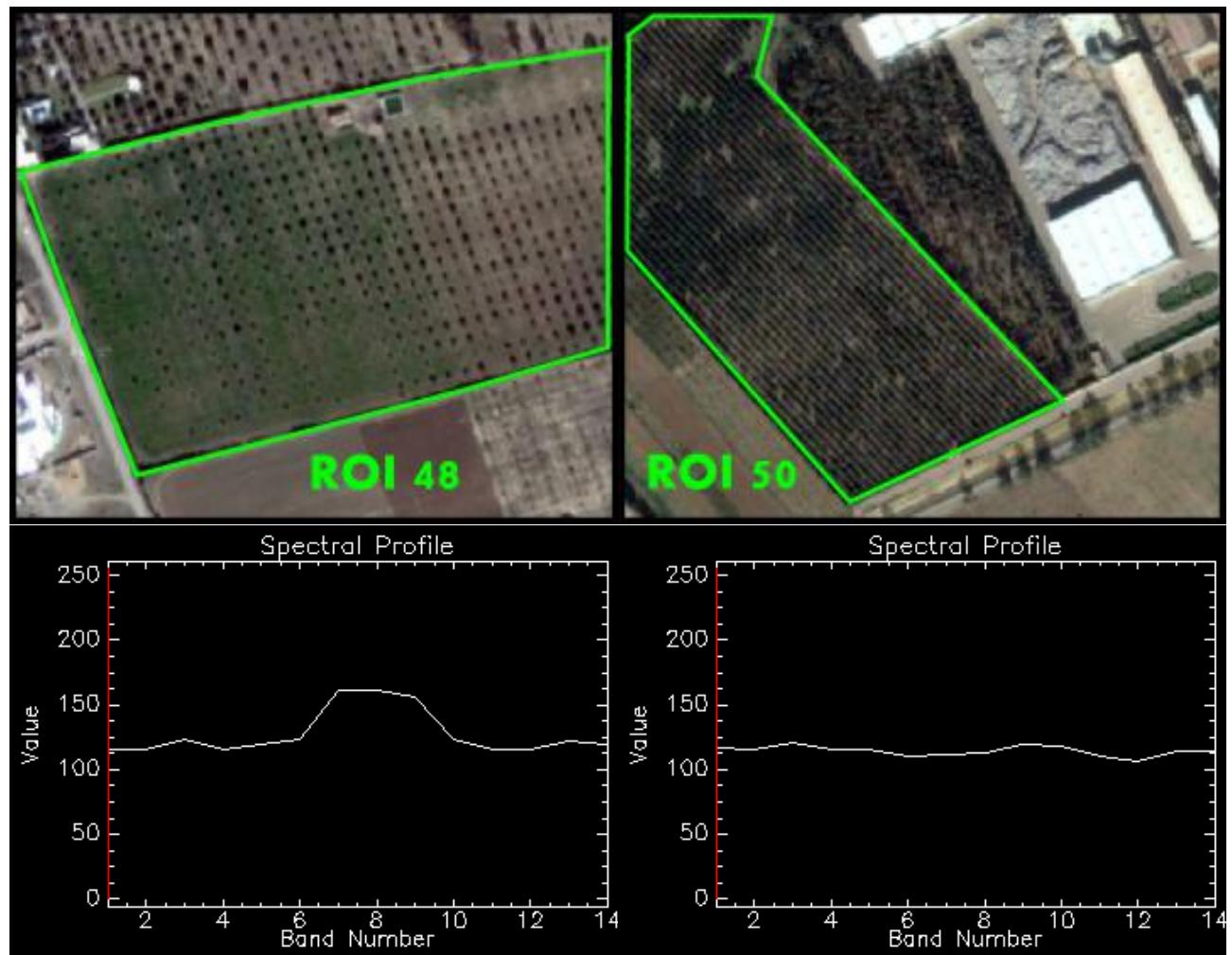


FIGURE 17 – Exemple de parcelles d'oliviers (haut) avec différentes distances inter-rangs, et profil NDVI associés (bas), visualisés sous ENVI

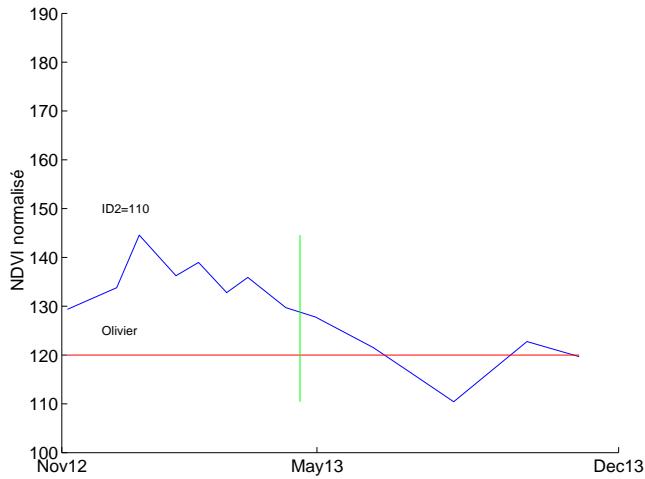


FIGURE 18 – Exemple de profil illustrant un mélange de signaux. ID2=110 signifie que l'on affiche le profil correspondant à la 110ème ROI de la campagne de terrain 2, identifiée à cette date comme parcelle d'oliviers.

le capteur de télédétection. Il en résulte un mélange de signaux difficilement exploitable. Dans la plupart des cas, les observations de terrain précisent les mélanges observés par des noms de classe doubles (ex : olivier + piment). Toutefois, il peut arriver de rencontrer des données identifiées comme une seule classe, mais dont il est évident que le signal est perturbé par d'autres couches de végétation. Ainsi, sur la figure 18, nous observons un profil de NDVI censé être d'olivier, mais dont l'allure peut s'apparenter à un signal de culture annuelle (fève en particulier).

- **De l'observation de terrain :** Les vérités terrain ne sont pas absolues et se heurtent à quelques difficultés. Ne sont pas identifiées les mêmes choses à quelques jours ou semaines près. Il est parfois possible de déterminer la nature d'une culture qui vient d'être plantée (par expertise ou renseignement auprès de l'agriculteur) sans que sa signature ne soit encore perceptible. Concernant le fourrage, il se distingue aisément du blé en été quand ce dernier est absent. Au contraire, la subjectivité de l'observateur rentre parfois en jeu au printemps quand le fourrage est vert et peut être confondu avec du blé.
- **De pratiques agricoles différentes :** Les méthodes de travail diffèrent selon les agriculteurs, aussi bien par rapport au matériel employé qu'au niveau des dates de plantation et de récolte des cultures. Si l'on s'intéresse aux deux profils de NDVI de piments de la figure 19, hormis un décalage de plusieurs mois quant à la date de semis, on remarque que la croissance du premier est lente et perturbée. L'hypothèse la plus probable est le pâturage de vaches pratiqué régulièrement dans ce type de cultures. Enfin, les motivations économiques influent les stratégies de récoltes (dates différentes de récoltes/fauchage, arrêt brutal d'une culture) et d'irrigation (Figure 19, haut gauche : céréale hiver non irriguée, haut droite : céréale hiver irriguée).
- **Des conditions météorologiques de l'année en cours :** Un paramètre supplémentaire à prendre en compte reste la variabilité des conditions météorologiques d'une année sur l'autre. Le problème ne s'observe pas pour nos données puisque toutes issues d'une même série temporelle mais il est important d'en être conscient pour le développement de la méthode qui devra y être

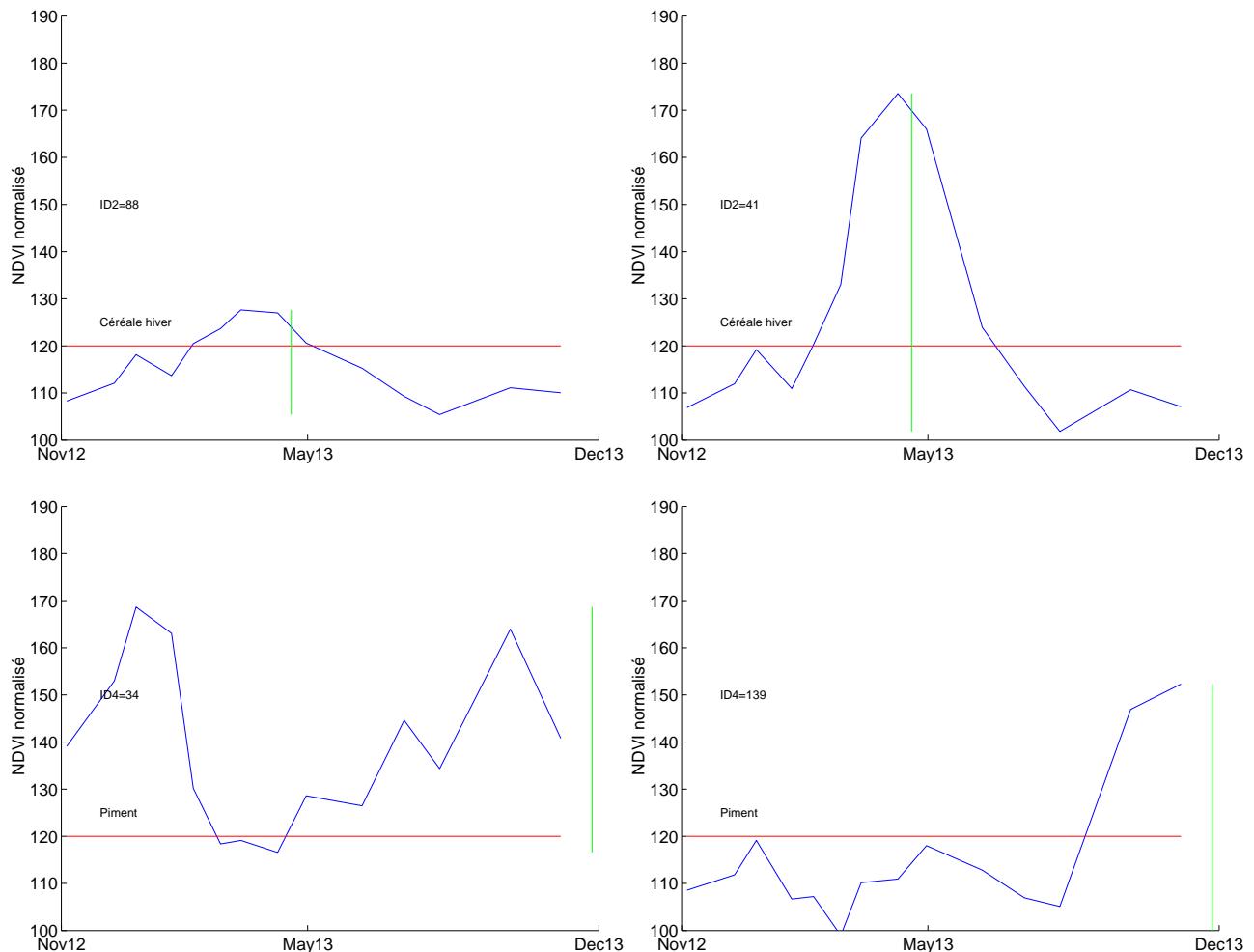


FIGURE 19 – Exemple de profils de cultures d'une même classe, cultivés avec des pratiques différentes. En haut à gauche : céréale d'hiver non irriguée, en haut à droite : céréale d'hiver irriguée, en bas : profils de parcelles de piments conduites différemment.

robuste. Ce phénomène apparaît particulièrement marqué en zone semi-aride où les cultures sont très sensibles au manque d'eau et aux fortes pluies.

L'analyse visuelle des données nous a renseigné sur l'allure générale de nos profils. Il en ressort une certaine variabilité intra-classes à ne pas négliger, mais nous conforte globalement quant à la possibilité de distinction des différentes classes de cultures. Commencent à émerger instinctivement plusieurs critères de distinction qu'il s'agit de formaliser.

2.2 Mise en place de la méthode

Nous présentons ici la réflexion menée sur les différentes options envisagées avant d'aboutir à la solution retenue. La robustesse visée de la méthode au bruit des données en vue de son automatisation guide le choix des critères retenus.

2.2.1 Extraction des profils de référence

L'intégralité d'un profil n'est pas toujours exploitable en tant que donnée de référence : pour les cultures annuelles, n'est significative que la section correspondant à la date de passage sur le terrain.

On développe un outil permettant d'extraire puis d'enregistrer les sections de profils que l'on souhaite conserver en référence (stocker dans la base de données). Sur la base d'un affichage similaire à celui utilisé pour la visualisation des profils (Figure 20), l'utilisateur peut définir successivement pour chaque profil :

- si il souhaite l'exploiter ou non
- préciser interactivement à l'aide d'un clic de souris les dates de début et de fin de la section de profil extraite
- un marqueur (0, 1 ou 2) visant à préciser si le bout de profil extrait est entier (0) ou interrompu par l'acquisition des images (1 : le profil est interrompu par le début de la série temporelle et seule partie finale du profil conservée, 2 : inversement, la partie initiale du profil est conservée).

Les sections de profils ainsi extraites sont enregistrées dans un fichier .txt de données de références. Elles sont triées par classe, et il figure pour chaque section : son identifiant, les valeurs du profil interpolées au pas de temps journalier, les dates de début et de fin des valeurs.

Le découpage manuel des données doit répondre à une logique cohérente et être en accord avec ce que l'on souhaite détecter par la suite. On choisit de conserver le cycle végétatif dans sa totalité, à savoir depuis la date de semence jusqu'à la fin de la senescence. Cela constitue l'information la plus pertinente à conserver dans la mesure où elle représente une réalité bio-physique significative. Concrètement, ces dates sont estimées à partir des minimums atteints en début et fin de cycle mais également grâce la période temporelle qui s'écoule entre la date de semis et la date à laquelle le NDVI devient visible.

Une opération fastidieuse s'en suit : il s'agit d'extraire pour chacun des quelques 700 profils leurs sections exploitables. Le choix a été fait de les identifier manuellement dans la mesure où cette opération ne sera effectuée qu'une seule fois. On aurait pu envisager leur sélection automatique, avec le risque d'introduire des erreurs. Étant donnée l'importance des données de références, sur lesquelles la méthode est basée, on priviliege d'apporter l'expertise de l'utilisateur pour leur validation.

2.2.2 Extraction des cycles de végétation

Une des difficultés préalables se révèle être l'extraction de ce cycle de croissance, puis de ses caractéristiques, à partir du profil brut extrait de l'image. Le problème a déjà été largement traité par la communauté scientifique [11,19] et l'on trouve dans la littérature plusieurs méthodes abordant cette problématique. On distingue les méthodes simples dites de seuillage, de celles basées sur une modélisation mathématique des profils pour en extraire les paramètres clefs.

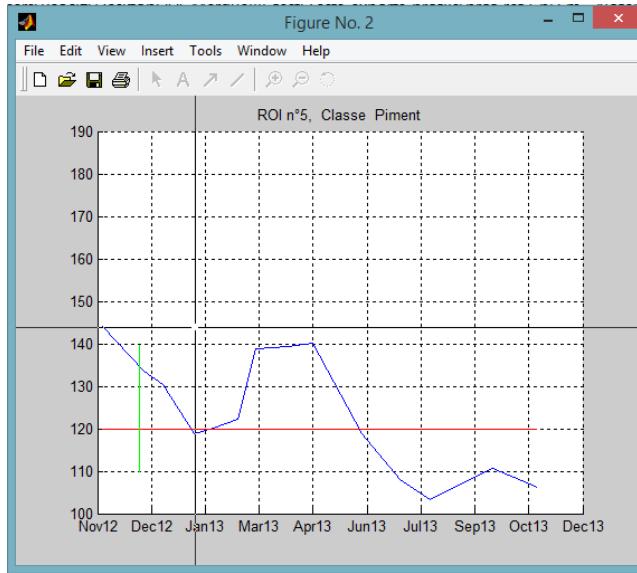


FIGURE 20 – Visualisation de l’outil de sélection des profils de référence

Méthodes d’interpolation

La méthode d’interpolation choisie conditionne la forme du profil et peut faciliter ou non l’analyse du profil et l’extraction d’information. Parmi les plus usuelles, on peut citer :

- l’interpolation linéaire
- l’interpolation cubique
- l’interpolation par spline
- l’interpolation par moindres carrés polynomiaux

Ces différentes méthodes d’interpolation sont appliquées sur différents profils types afin d’étudier leurs différents comportements (Figure 21)

L’interpolation par spline semble peu adaptée à ce que l’on recherche : elle semble avoir tendance à introduire des variations locales de valeurs au lieu de permettre un lissage plus global que l’on recherche. Au contraire, l’interpolation cubique introduit peu de variations par rapport à l’interpolation linéaire, mais l’intérêt de son utilisation est donc moindre.

Certains essais ont également été menés en utilisant des interpolations par moindres carrés polynomiaux. Si le résultat peut être satisfaisant sur certains profils avec un ordre de polynôme bien particulier, il est difficile d’imposer un unique ordre de polynôme qui puisse permettre une interpolation convenable pour tous les différents types de profils.

Finalement, l’interpolation linéaire présente l’avantage d’être simple à mettre en oeuvre et de travailler sur des valeurs les plus brutes possibles, sans risque d’introduction d’erreurs par des interpolations dont on ne maîtrisera pas toutes les subtilités.

Méthodes de filtrages, lissages et de modélisation mathématique

Une rapide recherche a également été menée sur les autres méthodes permettant de lisser et/ou de modéliser les séries temporelles dont nous disposons. Plusieurs méthodes de filtrage et de lissage existent (filtre BISE, filtre polynomial de Savitzky-Golay, ...), ainsi que des méthodes de modélisation mathématique (modélisation par double sigmoïde, ...). Cependant, ces méthodes requièrent pour la plupart de définir certains paramètres de lissages (ex : fenêtre temporelle du filtre utilisé, nombre de valeurs utilisées pour le lissage) et il est une fois de plus difficile d’imposer un unique paramètre sur une telle variété de forme de profils. De plus, des incertitudes persistent sur l’origine de certaines variations de NDVI, notamment quant à l’influence des erreurs de corrections d’images. Ceci nous empêchent de définir raisonnablement des seuils visant à lisser les erreurs induites par les corrections d’images sans

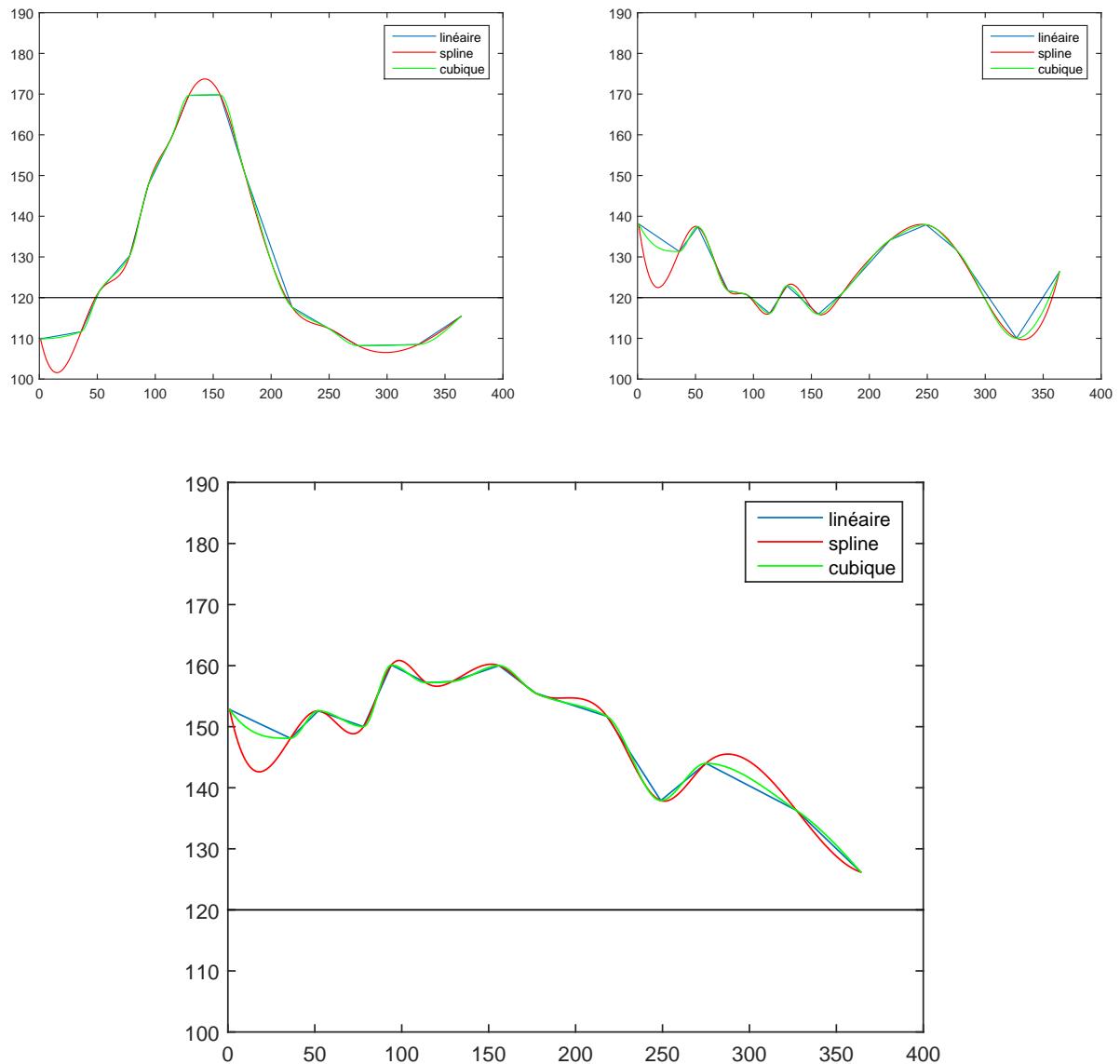


FIGURE 21 – Profils temporels de NDVI interpolés selon différentes méthodes (Gauche : profil de céréale d'hiver identifié à la campagne 2, milieu : profil de piment identifié en automne, droite : profil de luzerne).

risquer de gommer certaines variations de NDVI importantes d'un point de vue phénologique.

A partir de ce constat, on choisit finalement pour cette étude d'extraire les cycles des cultures du profil du pixel en minimisant les traitements et paramétrages. On imagine une méthode de seuillage basée sur des critères robustes.

Les valeurs de NDVI observées en-dessous du seuil 0.2 correspondent à des sols nus et leur variabilité s'explique par des travaux du sol ou des changements d'humidité. Il s'avère délicat d'exploiter ces valeurs et une première opération se dessine naturellement : chaque profil est seuillé à 0.2 (les valeurs inférieures deviennent classées comme du sol nu). En résulte des 'tronçons' de profils aux formes diverses, plus ou moins exploitables. Certains correspondent directement à des cycles phénologiques tandis que d'autres rendent compte du développement de végétation parasite (adventice) ou bien résultent du bruit lié aux mesures. Le parti a été pris de séparer les cycles phénologiques des autres 'tronçons' par un seuillage basé sur la valeur de leur intégrale. Ce choix permet de se baser simultanément sur la durée et l'amplitude du signal, ce qui correspond à une réalité biophysique et constitue un critère de distinction robuste. On prendra soin de distinguer les traitements appliqués aux 'tronçons' de profil entiers et à ceux interrompus par l'acquisition des images en leur appliquant des seuils de valeurs différentes.

Le principal problème qui persiste réside dans le fait que cette méthode ne permet pas de détecter un changement de culture dans le cas où le NDVI ne franchit pas le seuil 0.2. La situation s'observe lorsque plusieurs mois séparent les images d'une série temporelle : aucune des acquisitions n'enregistre le passage en sol nu de la parcelle et l'interprétation du profil s'en voit affectée. Etant donnée la rareté des nuages dans notre zone, on dispose facilement d'images acquises tous les mois. Un temps de revisite plus important ne serait pas pertinent au vue des échelles temporelles des variables biophysiques. On décide dans un premier temps de ne pas traiter ce problème même s'il présente une piste d'amélioration pour la suite.

Le traitement proposé précédemment ne prend sens que si appliqué aux cultures annuelles. On doit conserver la totalité des profils cultures permanentes pour leur analyse même si il y a périodiquement des passages en dessous de 0.2. L'idéal serait de parvenir à les détecter préalablement à cette phase en vue de leur appliquer un traitement alternatif. D'après les caractéristiques de ces profils soulignées dans la partie 2.1, plusieurs pistes sont étudiées :

- l'utilisation de critères de Fourier. En effet, comme on a pu le remarquer précédemment, on observe des variations de NDVI pseudo-périodiques concernant les cultures permanentes. De là, il serait envisageable de pouvoir les détecter grâce à une décomposition en fréquences principales. Les profils dont il ressortirait des motifs périodiques seraient considérés comme appartenant à des cultures permanentes, et dans le cas contraire comme relevant de cultures annuelles. Cependant, l'étude des fréquences principales contenues dans les profils ne prendrait sens que si appliquée sur un intervalle de temps minimum. Or on souhaite que la méthode puisse être appliquée pour classer des profils issus de toutes séries temporelles, y compris celles contenant peu d'images et s'étalant sur une petite période de temps. Cette piste de réflexion n'est donc pas creusée jusqu'au bout.

- le recours à des critères dits « experts », élaborés à partir d'une certaine expertise (seuils, différences min/max, ...). Expérimentalement, on arrive facilement à détecter plus de 90% des arbres à feuilles caduques et des autres cultures permanentes en se basant sur des critères propres à notre jeu d'images. Toutefois, la définition de critères robustes basés sur les caractéristiques phénologiques des cultures (Table 3) n'aboutissent pas à des résultats satisfaisants.

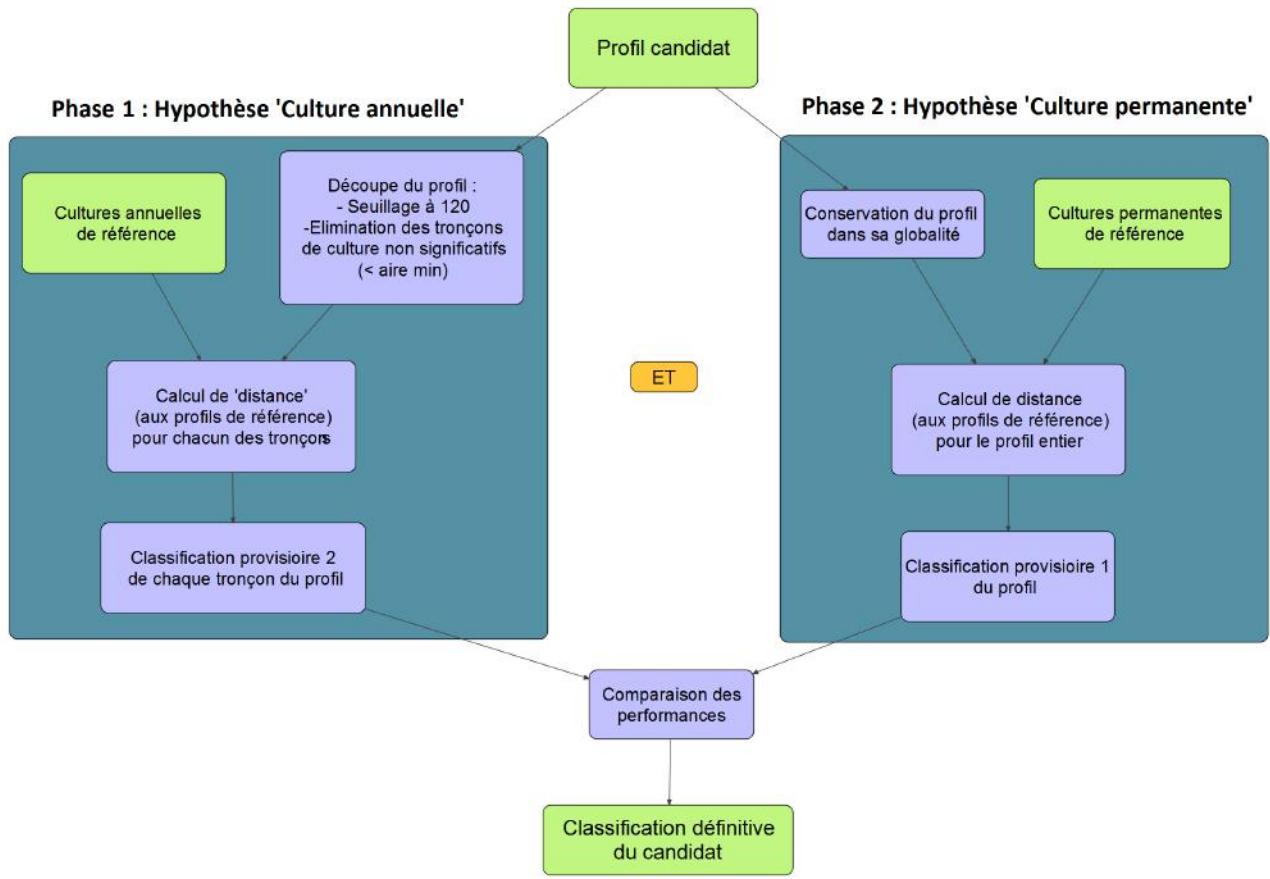


FIGURE 22 – Schéma de fonctionnement des deux phases de comparaison

Type de culture à détecter	Critère(s) appliqué(s)
Arbres à feuilles caduques	- Minimum du profil ∈ [1er février : 31 mars] - Maximum du profil ∈ [15 avril : 30 juin]
Autres cultures permanentes	- $ \text{Min-Max} \leq 30$ - Pourcentage de valeurs de sol nu < 80 %

TABLE 3 – Critères dits “experts” proposés pour détecter les cultures permanentes

Une autre approche est envisagée pour palier à cette difficulté. Plutôt que de chercher à déterminer préalablement à la découpe du profil temporel le caractère annuel ou non de la culture, on imagine deux phases de traitements distinctes que subirait simultanément le profil candidat (Figure 22). L'une correspond à l'hypothèse (1) : 'culture annuelle' et le profil est découpé selon la procédure décrite précédemment. Chaque 'tronçon' obtenu est comparé aux profils de références (de cultures annuelles seulement) afin d'identifier leur nature. Au contraire, dans la seconde phase (effectuée en parallèle), on considère que la culture est permanente (Hypothèse (2)) et la totalité du profil est conservée pour être identifiée. A l'issue de ces deux phases, la pertinence de chacune des deux hypothèses est estimée grâce à la performance des comparaisons effectuées. En effet, lorsqu'une culture permanente est découpée lors de la phase 1, il en résulte des 'tronçons' de profils difficilement identifiables puisque leurs formes diffèrent de celles enregistrées dans la base de données. La performance de la comparaison s'en ressent et le résultat obtenu à l'issue de la phase 2 est privilégié.

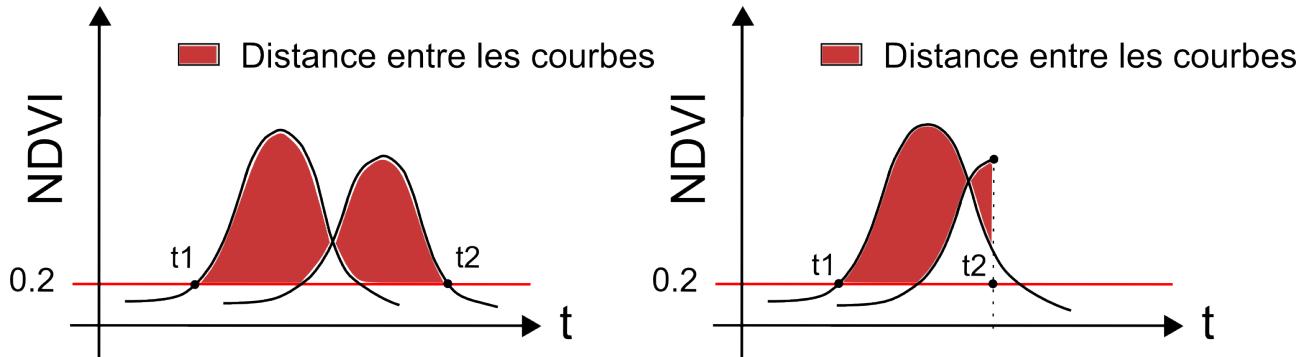


FIGURE 23 – Représentation de la distance D_n entre deux profils (Gauche : cas de deux profils entiers, Droite : cas d'un profil interrompu)

2.2.3 Critères de comparaison

Il s'agit de comparer les profils candidats à ceux de référence.

Plusieurs méthodes existent pour estimer la similitude entre deux courbes. Une méthode simple à mettre en œuvre est le calcul de l'aire située entre les deux courbes :

$$D_n = \int_{t_1}^{t_2} |f - g|, \text{ avec } f \text{ et } g \text{ deux fonctions quelconques}$$

Elle permet de rendre compte de la différence d'amplitude des courbes mais également de leur décalage temporel. On choisit d'appliquer ce calcul de 'distance' aux portions de courbes supérieures à 0.2 pour la raison évoquée précédemment (*). On parlera par la suite de 'distance' entre les courbes.

Les bornes de l'intégrale sont à définir différemment selon que les profils comparés soient entiers ou non.

- Si les deux profils sont entiers, on travaille naturellement sur l'union de leurs durées (Figure 23).
- Si au moins l'un des deux profils est interrompu (par l'acquisition des données), la ou les dates de coupures du ou des profils deviennent bornes d'intégrale.

Enfin, un paramètre supplémentaire à considérer pour le calcul de similitude des profils reste 'l'intervalle temporel' de comparaison, à savoir $t_2 - t_1$ (différence entre les deux bornes d'intégrale). Il est nécessaire d'intégrer cet aspect temporel en normalisant les distances afin de pouvoir les comparer entre elles.

Cette distance normalisée s'écrit :

$$D_{n_{norm}} = \frac{D_n}{t_2 - t_1} = \frac{\int |f - g|}{t_2 - t_1}$$

2.2.4 Traitement post-comparaison

S'il n'est pas fiable d'exploiter les valeurs NDVI inférieures à 0.2 pour le calcul de distance entre profils, on peut les utiliser après comparaison dans l'idée de mieux rendre compte de la réalité phénologique des cultures. En effet, un certain temps s'écoule entre la date de semis et la date à laquelle le NDVI devient visible (franchit le seuil 0.2). A fortiori, le même phénomène s'observe concernant la sénescence. Il est plus juste d'un point de vue biologique de cartographier la nature de l'occupation du sol à partir de la date de semis. On décide d'identifier, postérieurement à la comparaison des profils, ces débuts et fins de cycles en fixant arbitrairement 1 mois entre la date de semis et de NDVI visible puis 15 jours² entre la fin de la sénescence et la fin du cycle complet. Si, au cours de ces intervalles

2. Les durées ainsi fixées pourront être affinées par la suite grâce à des observations de terrain

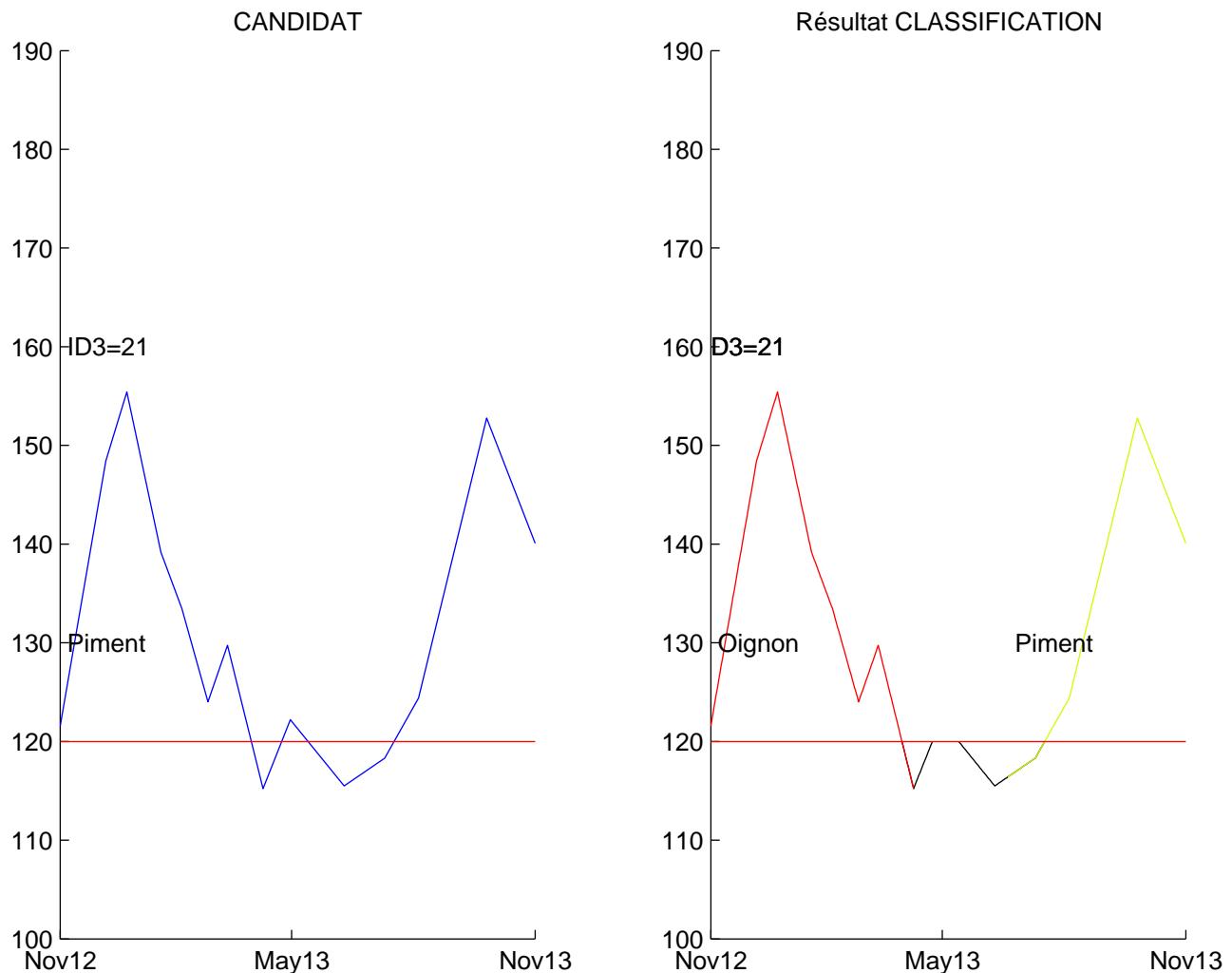


FIGURE 24 – Exemple de résultats de la phase obtenu après 'post-traitement'. A gauche : en bleu, le profil dit 'candidat'. A droite : résultat de la classification du profil. Un premier cycle de développement est identifié comme oignon (rouge), suivi d'une période de sol nu (noir), puis d'un nouveau cycle de piment (jaune).

temporels pré ou post végétation, un minimum est observé sur le profil de NDVI, il est choisi comme date de début ou de fin de cultures (Figure 24).

2.2.5 Méthode retenue

Finalement, sur la base de la méthode présentée à la partie précédente (Figure 22), la méthode est entièrement développée sous Matlab selon les algorithmes finaux suivants :

Notation	Définition
ST	Série temporelle à classer
C	Ensemble des profils candidats
C_{ij}	Candidat correspondant au pixel de coordonnées (i,j)
C_{ijk}	Bout de profil numéro k du candidat C_{ij}
$C_{ijk}(t)$	Valeur du bout de profil à la date t
$Classif_{ijk}$	Résultat de la classification pour chaque C_{ijk}
$R_{ijannuels}$	Profil annuel de référence numéro j de la classe i
$R_{ijpermanents}$	Profil permanent de référence numéro j de la classe i
$aire_{min}$	Valeur d'aire minimale fixée arbitrairement pour le moment. Sa valeur sera affinée à la partie suivante.

TABLE 4 – Notations

Algorithme 1 Algorithme de découpage des profils

Entrée : $C = \{C_{ij} | (i,j) \in ST\}$

Pour chaque candidat C_{ij}

Etape 1 : Interpolation au pas de temps journalier

Etape 2 : Calcul des points d'intersection avec le seuil 120 (qui correspond à une valeur de NDVI=0.2, et donc au seuil sol nu)

Etape 3 : Découpe du profil en bouts de profils C_{ijk}

Pour chaque bout de profil C_{ijk}

Si $\exists C_{ijk}(t) > 120$ et $aire(C_{ijk}) > aire_{min}$ **alors** nature(C_{ijk}) = 'culture'

Sinon nature(C_{ijk}) = 'sol nu'

Si C_{ijk} n'est pas entier **alors** marqueur(C_{ijk}) = 1 ou 2

Sinon marqueur(C_{ijk}) = 0

Fin Pour

Fin Pour

Sortie : $\{C_{ijk} | (i,j) \in ST, k \in \{1, \dots, \text{nombre de bouts de culture identifiés pour } C_{ij}\}\}$

Algorithme 2 Algorithme de classification utilisé dans la phase 1

Entrée : $C_{ijk}, R_{ijannuels}$

Pour chaque C_{ij}

Pour chaque C_{ijk} de nature='culture'

$aire_{min}=10^{10}$

Pour chaque $R_{ijannuels}$

$aire = \frac{\int |R_{ijannuels} - C_{ijk}|}{t_2 - t_1}$

Si $aire < aire_{min}$ **alors** $aire_{min} = aire$

Fin Pour

$\text{nature}(C_{ijk}) = \text{nature}(R_{ijannuels})$ qui minimise la distance

Fin Pour

Fin Pour

Sortie : $Classif_{ijk}1$

Algorithme 3 Algorithme de classification utilisé dans la phase 2

Entrée : C_{ijk} , $R_{ijpermanents}$

Pour chaque C_{ij}

$aire_{min}=10^{10}$

Pour chaque $R_{ijpermanents}$

$aire = \frac{\int |R_{ijpermanents} - C_{ijk}|}{t2-t1}$

Si $aire < aire_{min}$ **alors** $aire_{min} = aire$

Fin Pour

$nature(C_{ijk})=nature(R_{ijpermanents})$ qui minimise la distance

Fin Pour

Sortie : $Classif_{ijk}2$

Algorithme 4 Algorithme de classification de comparaison des phases 1 et 2

Entrée : $Classif_{ijk}1$, $Classif_{ijk}2$

Pour chaque C_{ijk} de nature='culture'

Si moyenne (distances des cycles de développement du profil aux références les plus proches) < aire(distance du profil entier à la référence la plus proche)

Alors $Classif_{ijk} = Classif_{ijk}1$

Sinon $Classif_{ijk} = Classif_{ijk}2$

Fin Pour

Sortie $Classif_{ijk}$

Quatrième partie

Résultats

Une méthode permettant de reconstituer les successions de cultures a pu être définie dans la partie précédente. Nous allons maintenant exposer et analyser les résultats obtenus pour étudier la pertinence de cette méthode. Tout d'abord, nous présenterons l'affichage retenu pour la visualisation des résultats ainsi que son intérêt pour leur interprétation. Ensuite, la mise en place de typologies adaptées à la classification et à la validation des résultats permettra de mettre en place plusieurs tests de performance de la méthode. Certains seront notamment utilisés pour affiner les paramètres employés dans les algorithmes, tandis que d'autres permettront de rendre finalement compte de la performance globale de la méthode retenue. Enfin, une réflexion sera menée sur les limites de la méthode et sur ses perspectives d'évolution.

1 Exemples de visualisation de résultats

Afin de faciliter l'interprétation des résultats obtenus au fur et à mesure du l'évolution de la méthode, est programmé sous Matlab un affichage de la classification temporelle du profil (Figure 25).

Y figurent, pour chaque profil de référence testé, les résultats des deux classifications issues des phases 1 et 2. Ce résultat comporte l'identifiant, la nature et l'allure des profils de référence issus de la comparaison du profil, superposés au profil candidat. Ceci permet d'étudier visuellement la similitude des profils tests et de référence, et donc d'estimer la pertinence de la classification. L'outil s'avère également utile pour analyser les différentes sources d'erreurs possibles pour y apporter des solutions. Par exemple, concernant le profil candidat ID3=88, on remarque que si la culture fut identifiée lors de la campagne de terrain 3 comme de la tomate, le profil est classé à cette date par l'algorithme comme du piment. On constate cependant que le profil testé est très proche du profil de référence sélectionné après comparaison. L'erreur de classification provient donc ici du principe même de la méthode, et de la confusion possible de certaines classes en fonction de leurs signatures spectrales. Au contraire, la classification s'avère juste pour le candidat ID3=76.

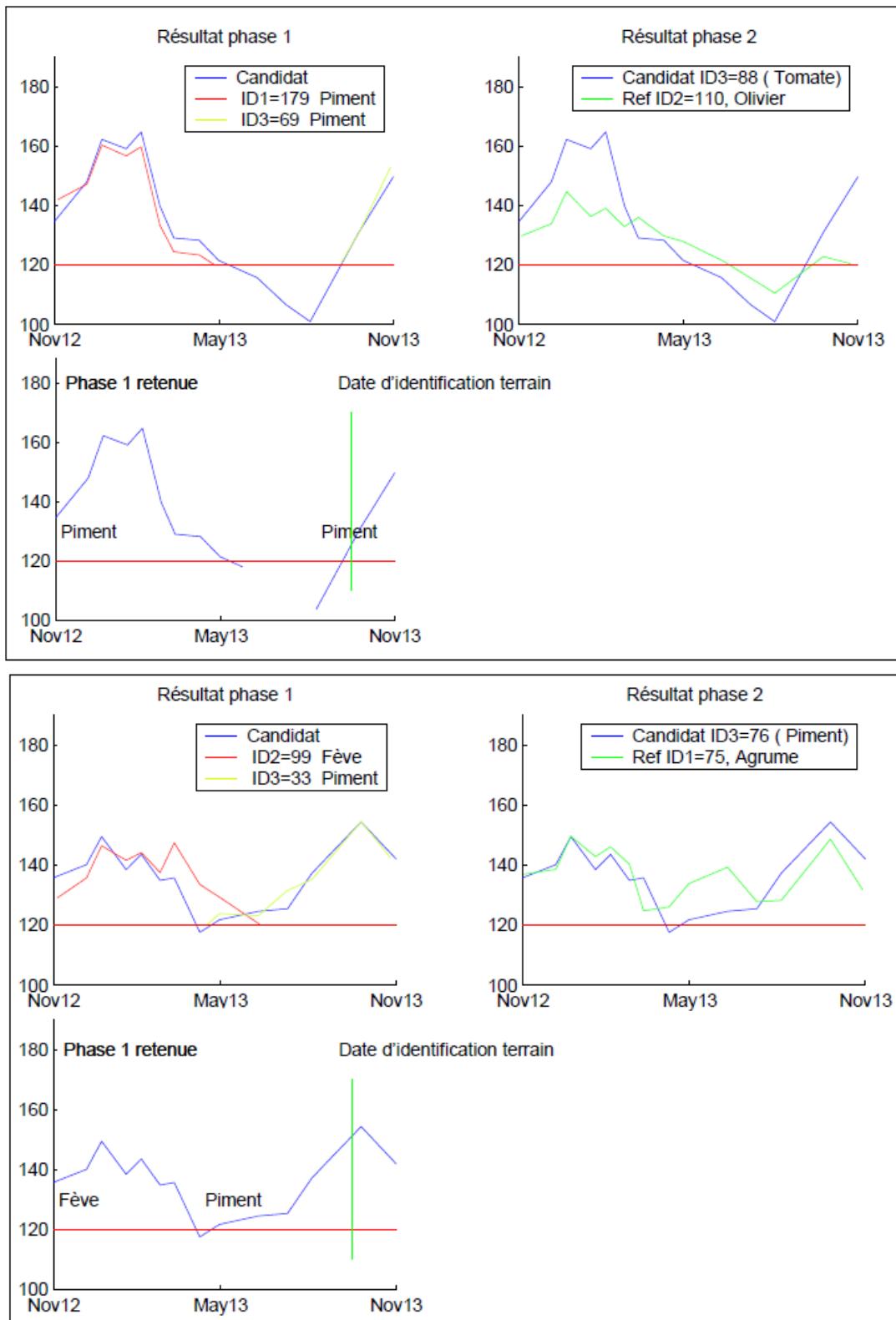


FIGURE 25 – Exemple d'affichage du résultat de la classification d'un profil

2 Analyse et typologies des données terrain

2.1 Typologie relative aux données de références

Les calculs de performance qui seront menés au cours des différents tests qui vont suivre doivent refléter la qualité de l'algorithme et de la méthode de classification proposée. Pour cela, il est indispensable de trier le jeu de données pour disposer de données propres et représentatives. Ce tri permet de distinguer plusieurs groupes de données et facilite l'interprétation des résultats. On distingue quatre types parmi nos 700 données de références :

- **Les données de type 1, « Typiques »** : sont regroupées ici les données les plus propres. On espère pouvoir en identifier une grande majorité grâce à notre méthode.
- **Les données de type 2, « Atypiques Significatives »** : leur profil NDVI est atypique par rapport à la majorité des cultures de leur classe mais il reste tout même caractéristique de la nature de la culture. L'algorithme peut potentiellement les détecter.
- **Les données de type 3, « Mal identifiées »** : elles présentent a priori un problème d'identification et ne sont absolument pas caractéristiques de leur classe. L'algorithme ne peut pertinemment les détecter et ces données ne sont pas exploitables pour la validation.
- **Les données de type 4, « Mélanges ou minoritaires »** : y figurent celles dont le nom correspond à un mélange de cultures ou à une classe très minoritaire (Sorgho, maïs, ..). Les signaux de mélanges de cultures difficilement exploitables, et nous ne chercherons pas pour l'instant à les identifier. Pour les cultures minoritaires, on ne possède pas suffisamment de données de références pour prétendre pouvoir les distinguer, d'autant plus que l'enjeu est moindre.

2.2 Typologie relative à la classification

A partir de ces considérations, de la précision recherchée mais aussi suite à l'analyse visuelle des profils (cf partie 2.1.1), on définit une typologie relative à la classification :

Cultures annuelles	Cultures permanentes
Piment	Olivier
Céréale hiver	Rosaceae (Pêcher, Amandier, Prunier, Abricotier, Pommier)
Fourrage annuel	Agrume
Oignon	Grenadier
Tomate	Fourrage permanent
Fève	Luzerne
Pastèque	Jachère
Melon	

TABLE 5 – Typologie retenue pour la classification

3 Tests d'optimisation de la méthode

Une série de tests est menée sur les données de type 1 et 2 afin d'affiner les paramètres utilisés dans la méthode de classification. La validation des résultats s'effectue par leave-one-out afin d'exploiter la configuration optimale de la base de données, et a fortiori de la méthode.

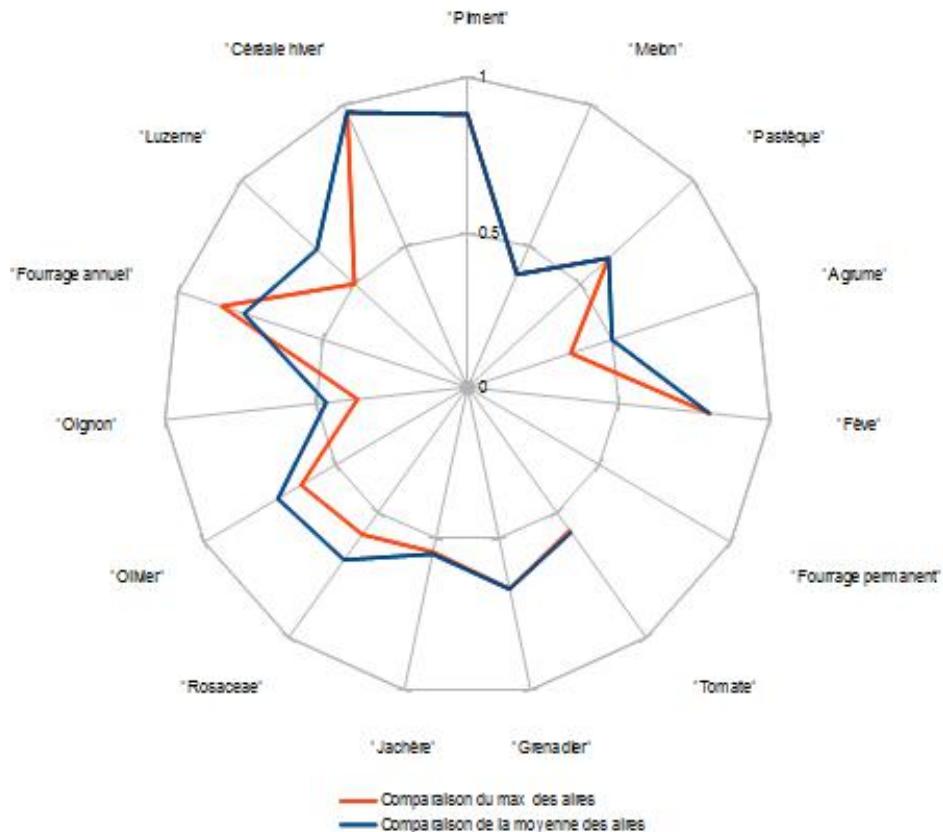


FIGURE 26 – Comparaison du mode de sélection entre résultats issus des phases 1 et 2

3.1 Mode de comparaison des résultats issus des phases 1 et 2

On s'intéresse ici à la manière d'estimer la performance des classifications issues des deux phases de traitement.

La performance de la comparaison d'un tronçon de profil avec ceux de références s'estime par la valeur de la distance D_n calculée entre les deux bouts de profils.

Cependant, estimer la performance globale de la classification d'un profil entier revient à étudier les différentes valeurs de similitudes calculées pour chacun des bouts de profils. Plusieurs indicateurs de performance relative entre les résultats des phases 1 et 2 méritent d'être comparés. Le premier consiste simplement à comparer la moyenne des distances calculées pour chacun des 'bouts' de profils issus de la phase 1, à la valeur de la distance issue de la phase 2. La phase qui présente la moyenne la plus basse est retenue. Un point de vue différent amène à comparer non plus la moyenne, mais la valeur maximale des distances issues de la première phase, à la distance obtenue en phase 2. Ce concept illustre l'idée que le découpage n'avait pas lieu d'être si l'une des distances calculées pour un tronçon est trop grande, auquel cas on conserve le résultat issu de l'hypothèse 'culture permanente'.

Ces deux méthodes reflètent des sensibilités différentes aux valeurs et les tests suivants nous aident à mieux cerner laquelle privilégié.

Le graphique de la figure 26 rend compte de la précision obtenue pour chaque classe selon les deux modes de comparaison testés. Excepté pour le fourrage annuel, la précision obtenue par comparaison des moyennes est toujours supérieure ou égale à celle obtenue en se basant sur les valeurs des maximums. Au vu de ces résultats, on priviliege par la suite l'utilisation d'une comparaison basée sur la moyenne pour estimer le meilleur résultat issus des deux phases.

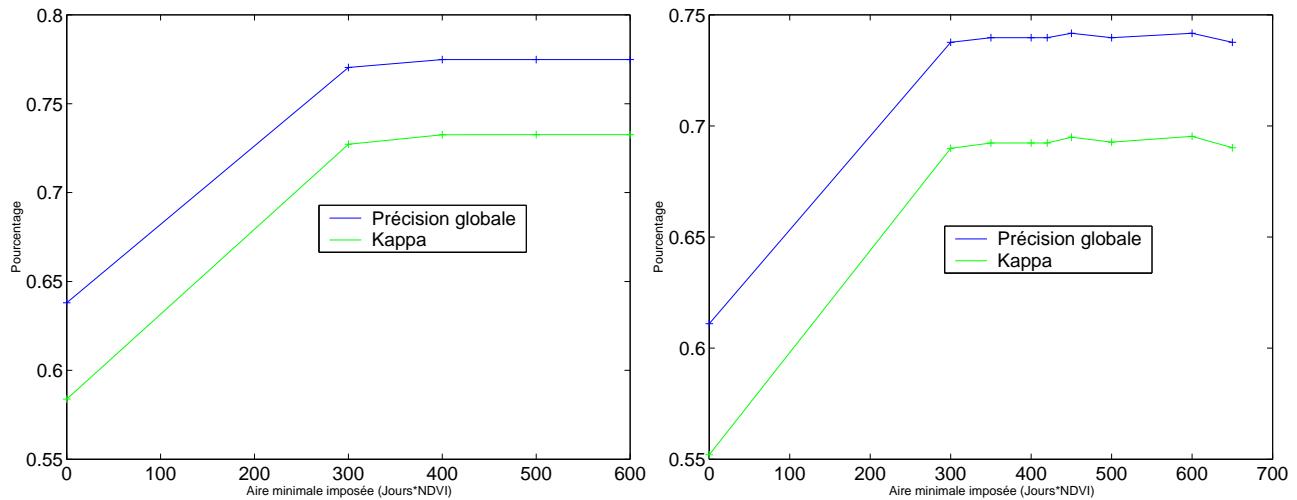


FIGURE 27 – Evolution de la performance de la classification en fonction de l'aire minimale imposée pour les cycles entiers identifiés après le seuillage à $NDVI = 0.2$ (Gauche : données type 1, Droite : données type 1+2)

3.2 Valeurs des aires minimales

Lors de la phase d'extraction des 'tronçons' de profils, le choix a été fait de distinguer les cycles réels de végétation du bruit parasite par seuillage d'intégrale (cf partie 2.2.2). La valeur du seuil s'estime approximativement à partir de variables biologiques, notamment par la durée du cycle ainsi que la valeur NDVI minimale atteinte. Aussi, en considérant un cycle entier, une valeur de seuil cohérente s'approche à 400 jours*(NDVI normalisé). Il est possible de l'affiner grâce à des tests menés sur les données de type 1 et 2 qui, fiables, intègrent toutefois la variabilité des données rencontrées.

Les aires minimales optimales à imposer se révèlent clairement dans les deux cas : la valeur 400 maximise la performance de la classification des données les plus propres, tandis que l'on doit imposer une aire minimale de 450 pour obtenir les meilleurs résultats dans le cas des données type 1+2 (Figure 27). Est retenue la valeur qui offre la meilleure classification possible pour les deux situations, soit 450.

Des expériences similaires permettent d'accéder aux seuils d'intégrales optimums dans le cas où les profils sont interrompus ie leur cycle de végétation incomplet.

Deux asymptotes se dessinent nettement dans chacun des cas pour les valeurs de kappa et de précision globale. Les valeurs d'optimisation sont respectivement 100 et 150 jours*(NDVI normalisé) pour le cas des données type 1 et type 1+2 (Figure 28). On retient pour la suite celle permettant d'optimiser les deux groupes de données à savoir ici 150.

3.3 Influence de la base choisie pour le calcul d'aire

La base retenue pour le calcul des aires n'est pas neutre et influe sur les résultats obtenus aussi bien à l'issue de chacune des phases que lors de la comparaison de leur performance. On entend par 'base' la valeur seuil de NDVI à partir de laquelle l'intégrale est calculée. Étant donnée la renormalisation de nos données, une base de 0 amène à des aberrations lors du calcul de la distance normalisée. La distance brute D_n calculée apparaît exagérément grande par rapport à la durée temporelle $t_2 - t_1$ et la distance normalisée D_{n_n} qui en résulte s'en écarte très peu. Si ceci ne pose pas de problème majeur pour identifier relativement la nature de chaque bout de profil lors de la phase 1, on aboutit à des aberrations lorsque l'on est amené à comparer le résultat des phases entre elles. En effet, la distance D_2 entre les profils utilisée lors de la phase 2 ne tient pas compte d'une quelconque base puisqu'elle consiste simplement en une différence d'aires, les deux profils étant comparés sur leur intervalle temporel commun. Une fois

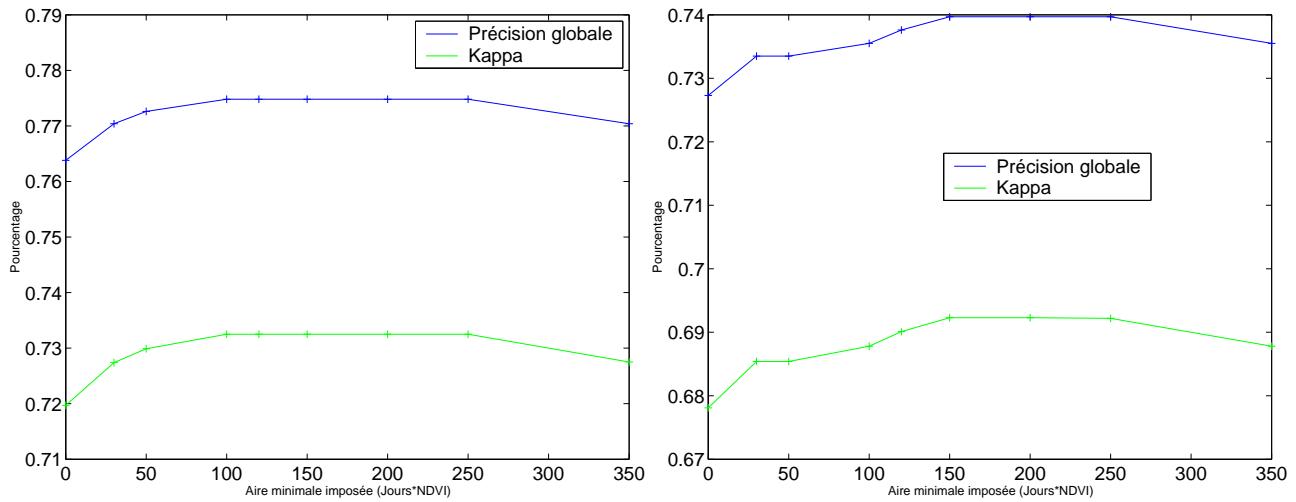


FIGURE 28 – Evolution de la performance de la classification en fonction de l'aire minimale imposée pour les bouts de profils interrompus (Gauche : données de type 1, Droite : données de type 1+2)

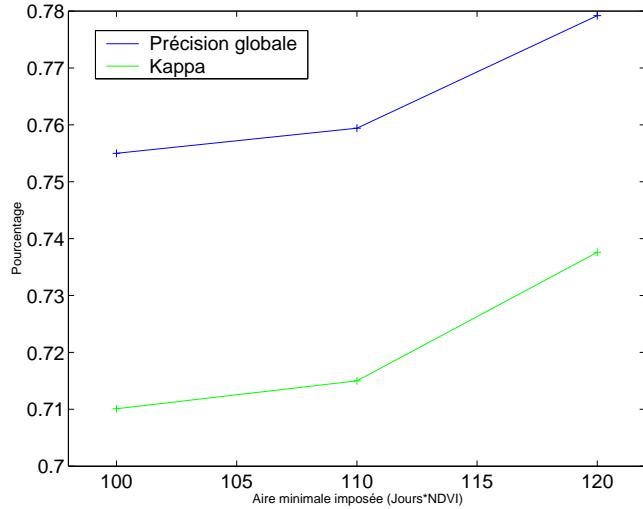


FIGURE 29 – Evolution de la performance de la classification en fonction de la base utilisée pour le calcul des aires

normalisée par le temps, cette distance D_2 se révèle anormalement petite face à D_n , et la comparaison s'avère délicate.

Instinctivement, on devine qu'il est plus approprié d'utiliser une base de 120. Différentes valeurs sont comparées pour estimer l'influence de ce choix.

Les valeurs obtenues confirment l'hypothèse de départ. La performance de la classification se dégrade lorsque l'on utilise pour base de calcul d'aire une valeur inférieure à 120 (Figure 29). Le poids relatif affecté aux valeurs NDVI inférieures à 120 augmente par rapport au poids du reste du profil dans la configuration d'une base inférieure à 120 (Figure 30). Les erreurs que cela engendre peuvent s'expliquer par la faible représentativité et fiabilité de ces valeurs auxquelles est accordée une plus grande importance.

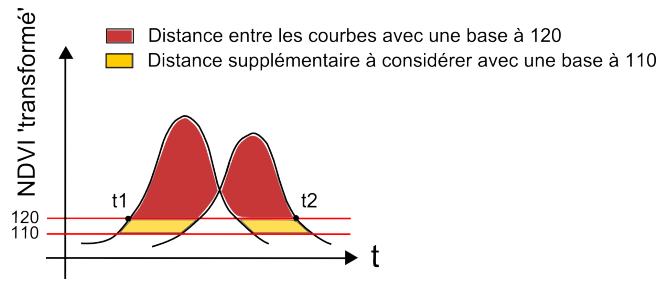


FIGURE 30 – Illustration de l'influence de la valeur choisie pour base du calcul de l'aire située entre les courbes

3.4 Regroupement de classes

Au vu des premiers résultats issus de la matrice de confusion, il est légitime de reconsidérer la typologie préalablement établie.

Sans regroupement														
	"Piment"	"Cézale In"	"Luzerne"	"Fouage a"	"Oignon"	"Oliver"	"Rosaceae"	"Jacinthe"	"Tomate"	"Fouage d"	"Fée"	"Agathe"	"Patisse"	"Mélanger"
"Piment"	1,20	0	0	3	0	0	2	0	0	17	0	0	0	144
"Cézale In"	0	59	0	1	0	1	0	0	0	0	0	0	0	53
"Luzerne"	0	2	1	1	2	2	3	0	0	0	1	0	0	2
"Fouage annuel"	0	0	0	0	0	0	0	0	0	0	0	0	0	31
"Oignon"	0	0	0	0	0	0	0	0	0	1	0	0	0	14
"Oliver"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Rosaceae"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Jacinthe"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Tomate"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Fouage permanent"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Fée"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Agathe"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Patisse"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Mélanger"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0,9219	0,9695	0,9695	0,7826	0,4615	0,7347	0,7407	0,913	0	0,5284	NaN	0	0,4624	0
Précision globale	0,7638	0,7197												
Kappa	0,7792	0,7375												

Avec regroupement des classes														
	"Piment"	"Cézale In"	"Luzerne"	"Fouage a"	"Oignon"	"Oliver"	"Autre Buil"	"Jacinthe"	"Tomate"	"Fouage d"	"Fée"	"Agathe"	"Patisse"	"Mélanger"
"Piment"	1,22	0	0	1	0	0	0	0	0	17	0	0	0	144
"Cézale In"	0	51	0	2	1	0	0	0	0	0	0	0	0	53
"Luzerne"	0	0	0	0	20	3	0	0	0	0	0	0	0	0
"Fouage annuel"	0	0	0	0	0	0	0	0	0	1	0	0	0	0
"Oignon"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Oliver"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Autre feuilles caduques"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Jacinthe"	0	1	0	0	0	0	0	0	0	0	0	0	0	0
"Tomate"	0	15	0	0	0	0	0	0	0	0	0	0	0	0
"Fouage permanent"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Fée"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Agathe"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Patisse"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
"Mélanger"	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0,9243	0,9623	0,9697	0,3	0,4236	0,3	0,833	0,913	0	0,5294	NaN	0,4624	0	0
Précision globale	0,7792	0,7375												
Kappa	0,7792	0,7375												

TABLE 6 – Matrices de confusion obtenues avant (haut) puis après (bas) regroupement des Grenadiers et Rosaceae. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.

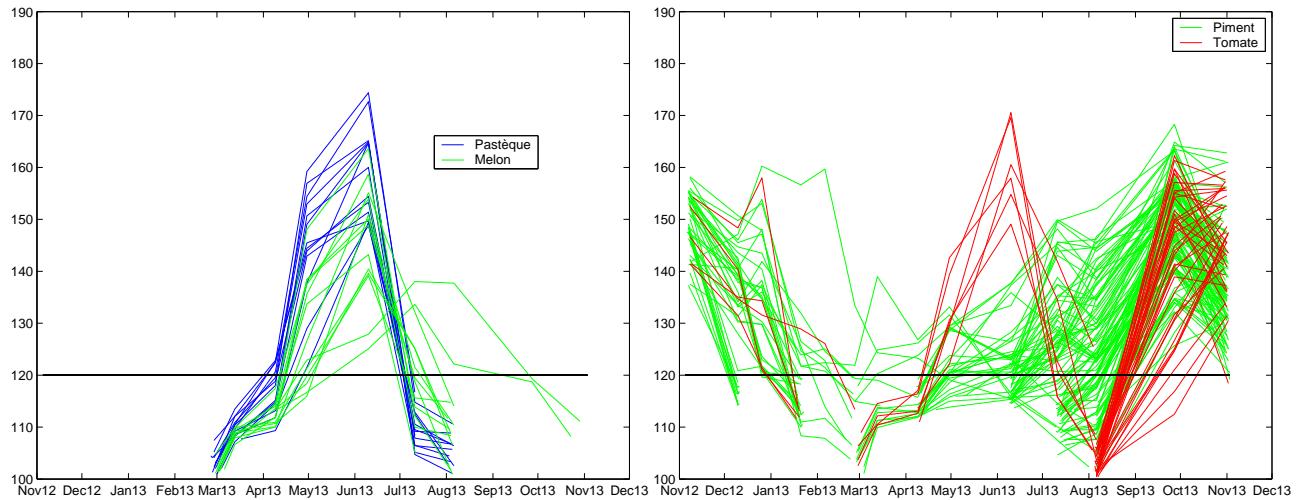


FIGURE 31 – Superposition des profils stockés dans la base de données pour les couples de classe Pastèque/Melon (Gauche) et Piment/Tomate (Droite)

En considérant les données de type 1, aucun des 5 grenadiers testés n'est bien classé. Deux sont identifiés comme des Rosaceae, deux autres comme des piments et un dernier comme un olivier. Ceci remet en cause le bien fondé des algorithmes mais aussi la pertinence des classes retenues pour la classification. Il avait été choisi pour notre typologie de différencier les grenadiers des rosaceae dans la mesure où leurs profils NDVI semblaient visuellement distinguables bien que très proches.

Visiblement, la distinction s'opère difficilement à partir des données dont on dispose qui restent peu nombreuses pour la classe Grenadier. Les classes Grenadier et Rosaceae sont regroupées sous le nom de Arbre à feuilles caduques.

On retrouve classés parmi les arbres à feuilles caduques les deux grenadiers précédemment mal identifiés et la précision de la classe s'en voit améliorée (Figure 6). La fusion de ces classes s'avère donc pertinente.

3.5 Ajout de critères « experts »

Les résultats précédents révèlent des confusions notables entre les classes piment/tomate et pastèque/melon. Si l'on superpose pour ces deux couples de classes les profils de références dont l'on dispose, on s'aperçoit que leur signaux se mélangent (Figure 31).

On imagine l'introduction de critères qualifiés d' « experts » visant à moduler dans certains cas le classement issu de la classification des profils. En moyenne, les intégrales des profils de la classe pastèque sont supérieures à celles de la classe melon. De même, les pentes associées aux profils tomates durant la période mi-août/mi-septembre semblent plus importantes que celles présentées par les piments. Aussi, il est possible d'envisager de nouveaux critères concernant ces classes.

- Pour chacune des classes melon et pastèque, la moyenne des intégrales de leur profils stockés en référence est calculée. Ceci permet d'affiner la classification retenue pour un profil classé en tant que pastèque ou melon : on lui affecte finalement la nature de la classe dont la moyenne des intégrales est la plus proche de l'intégrale de son profil.

- Le même processus basé sur la pente des profils entre mi-août et mi-septembre permettrait de mieux distinguer les piments des tomates.

	'Piment'	'Tomate'	'Pastèque'	'Melon'	'Oignon'	'Olivier'	'Arbre feuilles Caduques'	'Fourrage Annuel'	'Luzeme'	'Fève'	Totaux	Rappels
'Piment'	122	17	0	0	3	0	0	0	1	1	144	0.8472
'Tomate'	15	27	3	1	1	0	0	3	0	0	50	0.54
'Pastèque'	0	1	10	2	0	0	0	0	0	0	13	0.7692
'Melon'	0	2	3	6	0	1	1	0	0	0	13	0.4615
'Oignon'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Olivier'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Arbre feuilles caduques'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Fourrage annuel'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Luzeme'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Fève'	0	0	0	0	0	0	0	0	0	0	0	NaN
Totaux	137	47	16	9	4	1	1	3	1	1	0	0
Précisions	0.8905	0.5745	0.625	0.6667	0	0	0	0	0	0	0	0
Précision globale	0.75											
Kappa	0.5346											

	'Piment'	'Tomate'	'Pastèque'	'Melon'	'Oignon'	'Olivier'	'Arbre feuilles Caduques'	'Fourrage Annuel'	'Luzeme'	'Fève'	Totaux	Rappels
'Piment'	130	3	0	0	3	0	0	0	1	1	144	0.9028
'Tomate'	19	23	3	1	1	0	0	3	0	0	50	0.46
'Pastèque'	1	0	10	2	0	0	0	0	0	0	13	0.7692
'Melon'	2	0	1	8	0	1	1	0	0	0	13	0.6154
'Oignon'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Olivier'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Arbre feuilles caduques'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Fourrage annuel'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Luzeme'	0	0	0	0	0	0	0	0	0	0	0	NaN
'Fève'	0	0	0	0	0	0	0	0	0	0	0	NaN
Totaux	152	32	14	11	4	1	1	3	1	1	0	0
Précisions	0.8553	0.7188	0.7143	0.7273	0	0	0	0	0	0	0	0
Précision globale	0.7773											
Kappa	0.5616											

TABLE 7 – Matrices de confusion obtenues avant (haut) puis après (bas) application des critères experts. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.

Le caractère statistique de cette démarche n'affecte pas la robustesse de la méthode, contrairement à l'utilisation de valeurs de seuils fixes. S'il est évident que ces critères supplémentaires ne peuvent pallier à la mauvaise classification des profils trop dissemblables de leur classe, ils peuvent améliorer les résultats concernant les plus tangents.

Concernant les classes pastèques/melons, l'emploi de ces critères de pente permet de rectifier la classification pour deux profils 'melon' (Figure 7). Le rappel et la précision des classes augmentent. Pour les classes piments/tomates, l'introduction du critère permet de limiter les confusions concernant la classification de piments en tomates : la classification est rectifiée pour 8 profils de piments, et le rappel associé à la classe augmente. Cependant, de nouvelles erreurs apparaissent concernant la classification de profils de tomates. Si la précision de cette classe passe de 57% à 72% avec l'utilisation de ces critères, son rappel diminue quelque peu puisque 5 profils de tomates auparavant bien classés aboutissent maintenant dans la classe piment. Globalement, la légère diminution de la précision de la classe piment est contre-balancée par l'augmentation de celle de la classe tomate, et le critère augmente la précision globale de la classification.

Finalement, ces deux critères experts sont introduits dans la classification (lors de la phase 1).

4 Performance finale de la méthode retenue

De ces analyses préliminaires, des choix ont pu être faits en vue d'optimiser la performance de la méthode. On s'intéresse dans la partie qui suit à estimer la précision et la robustesse de la méthode retenue.

4.1 Statistiques finales

L'analyse des matrices de confusion et de leurs indices de performances associés renseignent sur la qualité de la méthode proposée.

	'Piment'	'Céréale Hiver'	'Luzerne'	'Fourrage Annuel'	'Oignon'	'Olivier'	'Arbre feuilles Caduques'	'Jachère'	'Tomate'	'Fourrage Permanent'	'Fève'	'Agrume'	'Pastèque'	'Melon'	Totaux	Rappels
'Piment'	130	0	1	0	3	0	0	0	9	0	1	0	0	0	144	0.9028
'Céréale hiver'	0	47	0	3	0	0	0	1	0	0	0	0	0	0	48	0.9792
'Luzerne'	0	0	2	1	0	2	0	0	0	0	1	0	0	0	6	0.3333
'Fourrage annuel'	1	1	0	1	1	0	0	1	1	0	0	0	0	0	14	0.6426
'Oignon'	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0.2222
'Arbre feuilles caduques'	0	0	0	0	0	0	2	1	0	0	0	0	0	0	35	0.7453
'Jachère'	0	0	0	0	0	0	0	27	1	0	0	0	0	0	23	0.9890
'Tomate'	7	0	3	1	0	0	0	0	28	0	0	0	0	1	50	0.5
'Fourrage permanent'	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0
'Fève'	1	1	0	0	0	0	0	0	0	4	0	0	0	0	6	0.6667
'Agrume'	0	0	0	0	0	3	0	0	0	0	5	0	0	0	8	0.825
'Pastèque'	1	0	0	0	0	0	0	0	0	0	0	10	2	13	0.7892	
'Melon'	2	0	0	0	0	1	1	0	0	0	0	1	6	13	0.6154	
Totaux	158	49	3	13	7	41	22	30	35	0	6	10	14	11	0	0
Précisions	0.8228	0.9592	0.6667	0.6623	0.2867	0.7805	0.9091	0.9	0.7143	NaN	0.6667	0.5	0.7143	0.7273	0	0
Précision globale	0.8045															
Kappa	0.7585															

	'Piment'	'Céréale Hiver'	'Luzerne'	'Fourrage Annuel'	'Oignon'	'Olivier'	'Arbre feuilles Caduques'	'Jachère'	'Tomate'	'Fourrage Permanent'	'Fève'	'Agrume'	'Pastèque'	'Melon'	Totaux	Rappels
'Piment'	131	0	1	1	3	5	0	0	9	0	2	0	0	0	152	0.8618
'Céréale hiver'	0	51	0	2	0	0	0	6	0	0	0	0	0	0	56	0.8944
'Luzerne'	0	0	2	1	0	2	0	0	0	0	1	0	0	0	6	0.3333
'Fourrage annuel'	3	2	0	22	2	0	0	1	2	0	2	0	0	0	34	0.6471
'Oignon'	0	0	0	0	6	0	1	0	0	0	1	0	0	0	14	0.4286
'Olivier'	2	0	0	0	0	36	2	1	0	0	2	2	0	0	48	0.8125
'Arbre feuilles caduques'	1	0	0	0	0	4	24	2	0	0	0	0	0	0	31	0.7742
'Jachère'	1	0	0	0	0	2	1	42	0	0	0	0	0	0	46	0.913
'Tomate'	12	0	0	3	1	0	0	0	28	0	0	3	1	50	0.49	
'Fourrage permanent'	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0
'Fève'	0	1	0	0	1	1	0	0	0	4	0	0	0	0	7	0.5714
'Agrume'	0	0	0	0	0	3	0	0	0	0	5	0	1	9	0.5556	
'Pastèque'	1	0	0	0	0	0	0	0	0	0	0	10	2	13	0.7892	
'Melon'	2	0	0	0	0	0	1	1	0	0	0	0	1	13	0.6154	
Totaux	166	54	3	29	13	56	26	53	34	0	11	10	14	12	0	0
Précisions	0.7952	0.9444	0.6667	0.7588	0.4615	0.6964	0.8276	0.7925	0.6766	NaN	0.3636	0.5	0.7143	0.6667	0	0
Précision globale	0.7583															
Kappa	0.7117															

TABLE 8 – Matrices de confusion finales obtenues sur les données de type 1 (haut) et sur les données type 1+2 (bas). Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.

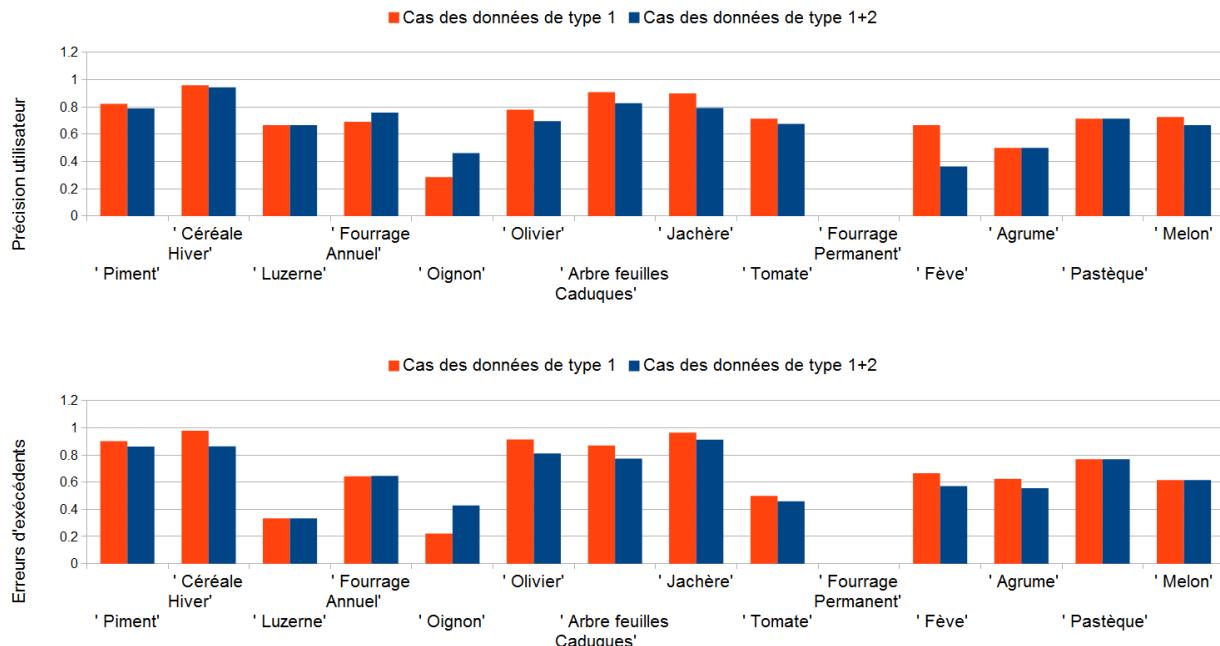


FIGURE 32 – Histogramme des précisions utilisateurs (Haut) et des erreurs d'excédents (Bas) obtenues par classe

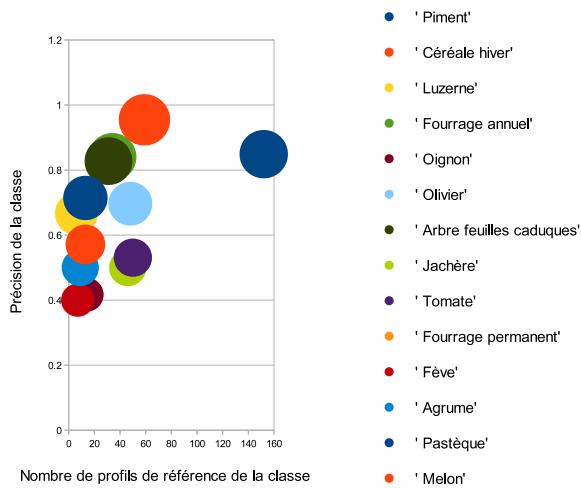


FIGURE 33 – Réprésentation de la précision obtenue par classe en fonction du nombre de profils de la classe présents dans la base de données.

La précision globale de 76 % sur les données de type 1+2 s'avère convenable dans le cadre de notre utilisation. Les précisions globales et relatives à chaque classe concernant les données de type 1 sont légèrement meilleures. Toutefois, le faible écart de performance obtenu entre les deux jeux de données atteste de la robustesse de la méthode.

La précision utilisateur s'élève respectivement à 79 % et 94 %, sur les données de type 1+2, pour les classes piment et céréale d'hiver qui apparaissent être les plus courantes dans notre zone d'étude d'après nos échantillons de terrain. De plus, ces classes représentent un fort pourcentage de la surface agricole totale et a fortiori de la consommation d'eau.

Les performances obtenues relativement pour chaque classe s'expliquent différemment selon diverses sources d'erreurs. Les confusions induites par la forte similitude entre certaines classes furent évoquées lors de la construction de critères experts (partie 3.5) qui ne résolvent pas toutes les erreurs : concernant les données de type 1+2, 3 agrumes sur 9 sont considérés comme des oliviers et 6 oignons sur 14 comme des piments. Outre les confusions inter-classes, une forte variabilité intra-classe limite la performance du classifieur et peut contribuer à expliquer les faibles résultats obtenus concernant les oignons (Figure 32). Le cas s'observe également par rapport à la classe 'fourrage permanent' pour laquelle aucun des profils n'est bien classé (Figure 32). Enfin, ce problème soulève celui du nombre de données de référence disponibles pour chaque classe, dont un plus grand nombre peut aider à limiter l'erreur induite par la forte de la variabilité des données.

L'étude de l'évolution de la précision utilisateur des classes en fonction du nombre de leurs profils de référence dont l'on dispose (Figure 33) laisse entrevoir une certaine corrélation entre les deux grandeurs. Les classes piment et céréale hiver en témoignent, notamment pour la classe piment dont la confusion avec la classe tomate est possible. Le phénomène s'observe également concernant les classes 'fève' et 'oignon' pour lesquelles on dispose respectivement de 7 et 14 profils exploitables.

Relativement peu d'oliviers sont mal classés (rappel de la classe de 0.81) contrairement à ce que dont on pourrait s'attendre au regard des faibles réponses NDVI qu'ils présentent. Cependant, certaines cultures y compris annuelles (piments) sont considérées par l'algorithme comme des oliviers.

	' Fourrage Annuel'	' Jachère'	' Fève'	' Olivier'	' Arbre feuilles Caduques'	' Luzeme'	' Agrume'	' Oignon'	' Piment'	' Tomate'	' Pastèque'	' Melon'	Total
' Olivier + Abricotier'	0	0	0	2	2	0	0	0	0	0	0	0	4
' Olivier + Pommier'	0	0	0	1	2	0	0	0	0	0	0	0	3
' Olivier + Céréale + Luzeme'	0	0	0	0	0	1	0	0	0	0	0	0	1
' Olivier + Pêcher + Agrume'	0	0	0	0	0	0	1	0	0	0	0	0	1
' Olivier + Pêcher'	0	0	0	2	7	0	0	0	0	0	0	0	9
' Olivier + Pomme de terre'	0	1	0	1	0	0	0	0	0	1	0	0	3
' Olivier + Fève'	1	1	1	2	0	0	0	0	0	1	0	0	6
' Olivier + Tomate'	1	0	0	1	1	0	0	0	2	4	0	0	9
' Olivier + Piment'	0	0	1	1	1	0	0	1	14	0	0	0	18
' Olivier + Fourrage'	1	0	0	0	0	0	0	0	0	0	0	0	1
' Olivier + Agrume'	0	0	0	0	0	0	2	0	0	0	0	0	2
' Olivier + Grenadier'	0	0	0	1	0	0	0	0	0	0	0	0	1
' Pêcher + Fève'	0	0	0	1	0	0	0	0	0	0	0	0	1
' Petit-pois + Fève + Oignon'	0	0	1	0	0	0	0	0	0	0	0	0	1
' Piment + Fève'	1	0	1	0	0	0	0	0	0	0	0	0	2
' Tomate + Fève'	0	0	2	0	0	0	0	0	0	0	0	0	2
' Fève + Oignon'	0	1	2	0	0	0	0	0	1	0	0	0	4
' Melon + Pastèque'	0	0	0	0	0	0	0	0	0	0	1	0	1
' Piment + Melon'	0	0	0	0	1	0	0	0	3	0	0	0	4

TABLE 9 – Matrice de confusion correspondant au test de profils de mélanges de cultures. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.



FIGURE 34 – Pourcentage d'identifications pertinentes pour les mélanges de cultures

4.2 Réponse aux mélanges de cultures

Pour un champ sur lequel sont cultivés plusieurs types de cultures, le signal enregistré pour le pixel correspondant est difficilement exploitable puisque résultat d'un mélange de signatures. Il apparaît légitime de ne pas en tenir compte lors du calcul de la précision de la méthode développée au moins dans un premier temps. Ces imprécisions relèvent des caractéristiques des capteurs de télédétection et non de la méthode de classification employée. La réponse de notre algorithme à ces données mérite pourtant d'être étudiée puisque ces pratiques se rencontrent fréquemment dans notre zone d'étude. On conserve la typologie précédemment établie et tente de classer les données de mélanges de cultures.

Dans 77 % des cas (Figure 34), l'algorithme identifie un des types de cultures du mélange et la classification est cohérente. Cependant, la nature de la culture ainsi identifiée semble difficilement prévisible, mis à part pour la configuration olivier+piment dans laquelle la réponse en NDVI du piment l'emporte sur celle de l'olivier (Figure 9).

La réponse des profils des cultures minoritaires non présentes dans la base de données est également étudiée.

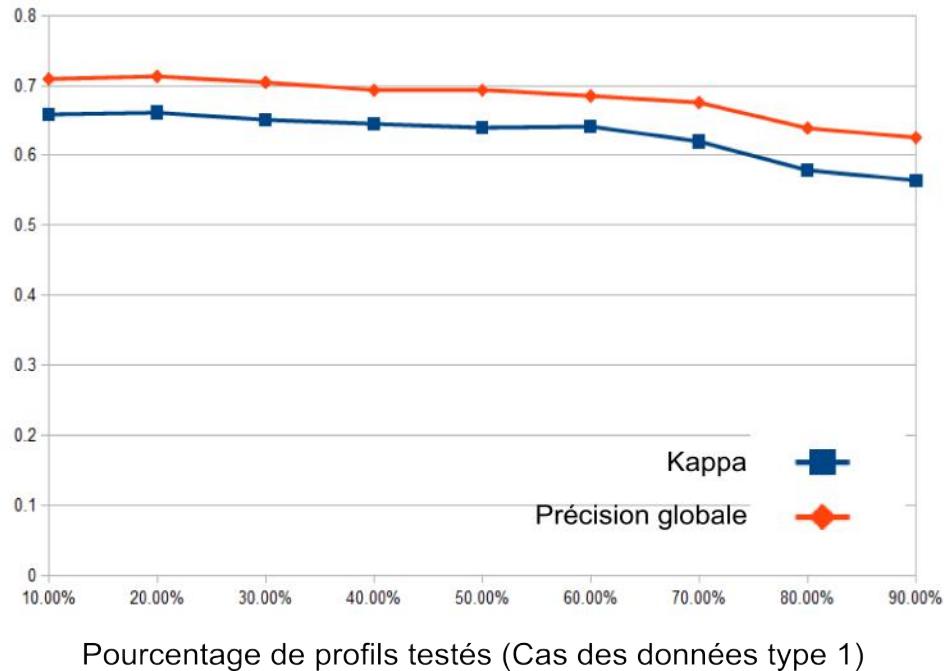


FIGURE 35 – Evolution de la performance de la classification en fonction du nombre de données présentes dans la base de données (Cas des données de type 1)

	* Fourrage Annuel*	* Jachère*	* Fève*	* Olivier*	* Arbre feuilles Caduques*	* Luzerne*	* Agrume*	* Oignon*	* Piment*	* Tomate*	* Pastèque*	* Melon*	Total
* Mais + Courgette*	0	0	0	0	1	0	0	0	0	0	0	0	1
* Artichaut*	1	0	0	0	0	0	0	0	0	0	0	0	1
* Sorgho*	0	1	0	0	0	0	0	3	0	0	0	1	5
* Pépinière*	0	0	0	1	0	0	1	0	0	0	0	0	2
* Pomme de terre*	0	5	0	2	0	0	0	1	0	1	1	0	10
* Carotte*	1	0	0	0	0	0	0	0	0	0	0	0	1
* Eucalyptus*	0	0	0	1	0	0	0	0	0	0	0	0	1
* Piment ?	0	1	0	1	0	0	0	0	0	0	0	0	2
* Sere*	0	0	0	0	1	0	0	0	0	0	0	0	1
* Mais*	0	0	0	0	0	0	0	0	1	0	0	0	1
* Vigne*	0	0	0	1	0	0	0	0	0	0	0	0	1

TABLE 10 – Matrice de confusion correspondant au test de cultures minoritaires. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.

Aucune cohérence particulière ne ressort de ces résultats et leur classification s'avère totalement imprévisible (Figure 10). Cependant, seules quelques parcelles isolées abritent de telles cultures et leur surface totale reste minime.

4.3 Influence du nombre de profils de référence stockés dans la base de données

La performance de notre méthode repose sur la qualité de la base de données utilisée. L'utilisation d'échantillons bootstrap obtenus par tirages aléatoires permet de tester les résultats obtenus par notre méthode pour différentes configurations de base de données. Un tirage consiste à sélectionner aléatoirement pour chaque classe $p\%$ des données qui sont alors utilisées comme échantillons tests tandis que les autres sont stockées dans la base de données. En considérant dans un premier temps l'ensemble des données de type 1, puis la combinaison des données de type 1 + 2, on effectue grâce à un script Matlab que l'on développe 10 tirages pour chacun des pourcentages $p \in \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$. La classification est testée sur chacun des échantillons qui en résultent et les résultats moyennés pour chaque pourcentage.

Dans le cas des données de type 1, les indices kappa et de précision globale diminuent d'environ 3%

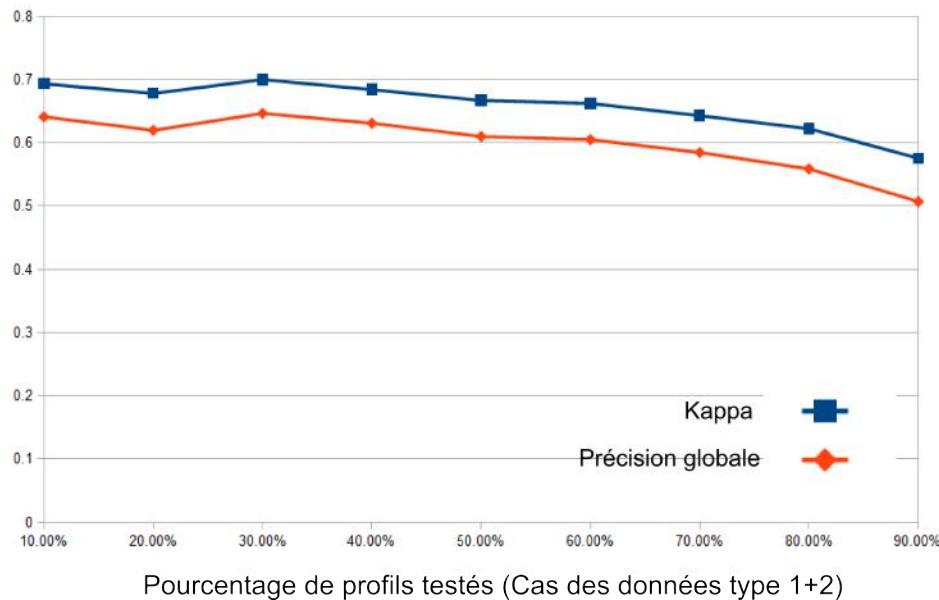


FIGURE 36 – Evolution de la performance de la classification en fonction du nombre de données présentes dans la base de données (Cas des données de type 1+2)

entre les configurations où 90 % puis 60 % des données sont utilisées en référence. Ces faibles variations rendent compte de la robustesse de la méthode au nombre de données présentes en référence, et par là aussi indirectement à leur variabilité. La précision se dégrade en-dessous du seuil de 60 % de données testées c'est-à-dire de 40 % de données utilisées en référence : le bon fonctionnement de la méthode est conditionné par une base de données suffisamment étouffée bien que la variation totale de performance n'excède les 10 %.

Concernant les données de type 1+2, la valeur relative au 20% de profils testés est anormale et peut s'expliquer par un nombre de tirages trop faible, laissant une influence relativement importante aux tirages exceptionnels. Si l'on ne considère pas cette valeur, la précision globale et le kappa diminuent de respectivement de 4 et 3 % entre les configurations où 90 % puis 60 % des données sont utilisées en référence. Ensuite, leurs valeurs se dégradent d'environ 10% entre les configurations où 60% et 90% des données sont testées.

Finalement, on peut conclure de ces chiffres que la configuration '40 % de données en référence' offre un bon compromis entre précision et nombre de données disponibles.

5 Perspective d'améliorations dans le cadre d'études ultérieures

Plusieurs pistes d'améliorations possibles ressortent de nos résultats :

- D'un point de vue technique, il sera nécessaire pour une utilisation régulière de la méthode **d'optimiser le temps de calcul**. Par construction, la méthode est coûteuse en temps puisqu'elle nécessite la comparaison de chaque section de chaque profil à chaque donnée de référence. Une seconde phase d'extraction de la classification obtenue pour plusieurs dates choisies participe à alourdir les calculs. Plusieurs solutions peuvent être envisagées pour y remédier. Provisoirement, l'utilisation de données cadastrales permettrait une classification orientée objet et non plus pixel, avec toutefois le risque de perdre de l'information brute. Ensuite, d'après les résultats présentés dans la partie 4.3, l'utilisation d'un nombre démesuré de données de références utilisées n'est pas nécessaire. On pourra optimiser le coût en temps de calcul de la

-
- méthode en optimisant leur nombre bien qu'il soit voué à augmenter au fil des campagnes de terrain. Enfin, on pourrait envisager une utilisation mixte de différentes méthodes. Combiner une approche arbre de décision pourrait permettre d'appliquer certains critères très grossiers au préalable de notre méthode afin de limiter, pour un candidat donné, les possibilités de nature de culture et limiter ainsi le nombre de comparaisons qui s'ensuit. D'autres regroupements très globaux et basés sur des méthodes statistiques pourraient être imaginés.
- Il sera indispensable d'**étoffer la base de données** grâce à des séries temporelles d'années différentes. Si cela vise à améliorer les résultats pour la plupart des classes, du bruit supplémentaire lié aux conditions météorologiques variables apparaîtra sûrement. Si la méthode proposée ici tente d'y être robuste, les précisions liées aux classes confuses (melon/pastèque, piment/tomate/oignon) risquent d'en être affectées et de faire émerger de nouveaux problèmes à traiter.
 - **Limiter les incertitudes liées aux corrections radiométriques** permettrait non seulement d'homogénéiser les données extraites de séries temporelles différentes, mais aussi d'envisager l'introduction de nouveaux critères 'experts' élaborés à partir de données plus fiables.
 - Il serait intéressant de parvenir à **distinguer le 'blé pluvial' du 'blé'**, par l'introduction d'une nouvelle classe. L'enjeu en terme de surfaces agricoles concernées n'est pas négligeable.
 - Un travail plus poussé sur le **lissage des valeurs et leur paramétrisation** pourrait aboutir à limiter le bruit des données et certaines confusions qui y sont associées. On imagine qu'une modélisation des profils faciliterait l'exploitation de séries temporelles comportant un nombre plus restreint d'images.
 - Il serait envisageable d'**appliquer le principe de notre méthode à une seule image acquise à une date t**. Une classification basée sur une comparaison 1D pourrait permettre d'accéder à des hypothèses de natures de cultures instantanément après l'acquisition de l'image.
 - **L'utilisation de masques visant à détecter les oliviers** et obtenus par des méthodes de traitement d'images très haute résolution spatiale permet d'envisager de meilleurs résultats quant à leur identification, bien que la précision actuelle soit déjà suffisante (mais peut être dégradée par l'utilisation de nouvelles séries temporelles aux erreurs de corrections d'images différentes). Ceci limiterait également la confusion des autres classes en 'olivier' lors de la classification.

6 Rendus finaux

A partir de la classification temporelle de chaque profil NDVI issu de chacun des pixels de l'image, on programme sous Matlab l'affichage de cartes d'occupation du sol à pas de temps variable et précisé par l'utilisateur. On utilise IDL afin de les géo-référencer et de les enregistrer dans un format standard de ENVI pour les cartes d'occupation du sol. La projection utilisée est l'UTM zone 32N, associée au système de référence géodésique Carthage. Compte tenu du temps de calcul nécessaire à la classification de l'image dans sa totalité, on présente ici des extraits de la zone (Figure 37).

L'affichage ci-dessus rend compte de la classification 'brute' de chacun des pixels. Les motifs correspondants aux champs ressortent clairement malgré un certain bruit que l'on pourrait limiter par l'utilisation d'un filtre majoritaire. Si la classification obtenue apparaît cohérente au regard du calendrier culturel, elle permet de souligner la complexité de l'organisation spatiale de la zone et la diversité des pratiques agricoles que l'on rencontre : les dates de semis d'une même culture se révèlent espacées de plusieurs mois pour certaines parcelles. Par construction, la nature des parcelles abritant des cultures permanentes n'évolue pas dans le temps.

Les routes et bâtiments sont classés en jachère compte tenu de leurs faibles valeurs de NDVI, avoisinant les 0. Il en est de même pour les zones d'eau dont les valeurs de NDVI sont négatives.

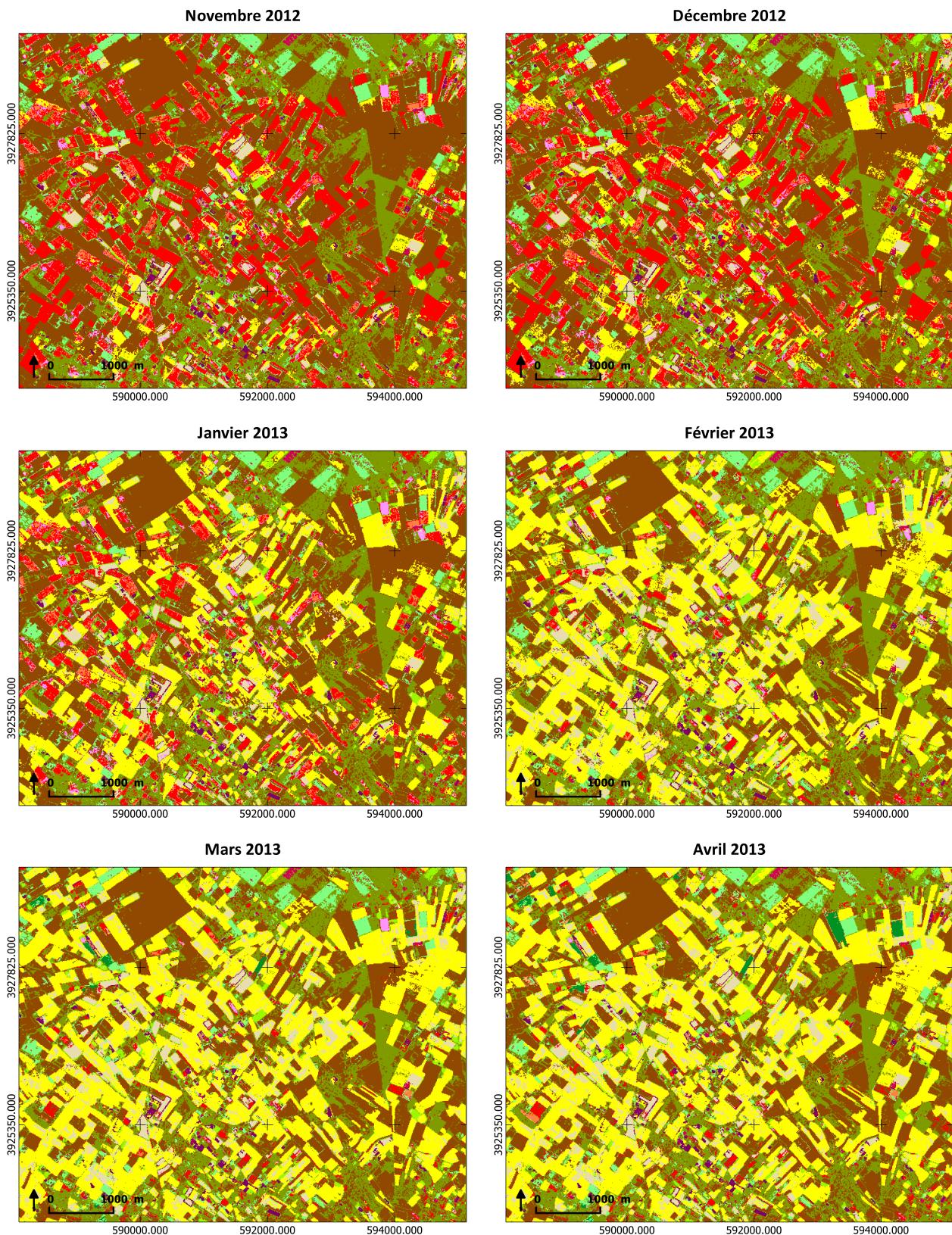
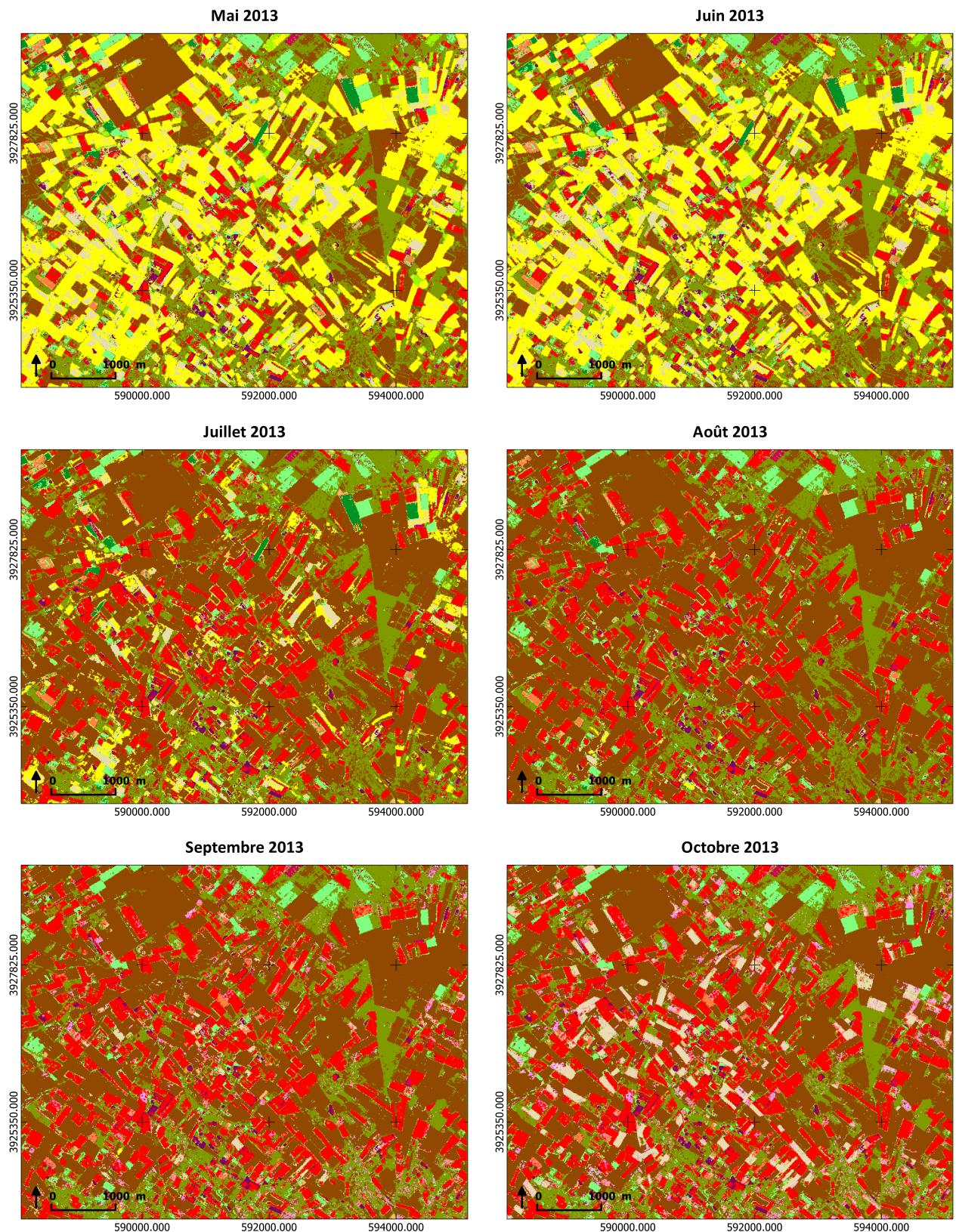


FIGURE 38 – Exemple de rendus de carte d'occupation du sol tous les mois



Cinquième partie

Conclusion

L'objectif de ce stage était de proposer puis de tester la possibilité d'une méthode basée sur la détection de plusieurs cycles de développement en exploitant le potentiel offert par les séries temporelles hautes résolutions spatiales et temporelles (Sentinel-1 et 2), dont l'apparition va révolutionner le monde de l'observation satellitaire.

Plusieurs étapes furent essentielles pour répondre à ces attentes. Dans un premier temps, un gros travail de visualisation des données de référence permit de mieux cerner leurs caractéristiques et de faciliter leur analyse. Ceci aboutit à proposer une méthode de classification reposant sur deux phases d'identification parallèles : l'une correspond à l'hypothèse « culture permanente » et consiste à considérer le profil dans son ensemble, tandis que la seconde suppose la présence de cultures annuelles et nécessite au préalable une phase de 'découpe' du profil visant à en extraire les différents cycles végétatifs. Dans les deux cas, l'identification repose sur une approche base de données : le cycle de développement à l'étude est comparé par un calcul de distance à tous ceux présents dans la base de données, et la nature de celui qui maximise la ressemblance est conservée. Finalement, on choisit pour classification définitive le résultat le plus pertinent de celui issu des deux phases.

Dans un second temps, la méthode ainsi proposée fait l'objet de plusieurs tests visant à affiner la valeurs des paramètres fixés dans l'algorithme, ainsi que les différentes typologies utilisées. Enfin, on cherche à estimer la performance et la robustesse de la méthode ainsi retenue. On obtient une précision globale et une valeur de kappa de respectivement 76 % et 71% ce qui, en pouvant paraître loin des 90 % fréquemment obtenus en zones occidentales, reste très satisfaisant étant donnés les fortes variabilités observées dans notre zone d'étude ainsi que le nombre de classes retenues dans la typologie (14 classes) .

Ainsi, cette méthode permet d'obtenir des hypothèses satisfaisantes quant à la nature des cultures de la zone et permettra d'affiner les bilans hydriques obtenus par le modèle hydrologique SAMIR. Cependant, quelques incertitudes persistent concernant la méthode, bien que de nombreuses pistes d'améliorations soient envisageables. On se questionne notamment quant à l'influence du jeu d'images utilisé et les corrections radiométriques peuvent s'avérer être l'une des difficultés majeures à surmonter par la suite. Toutefois, on espère pouvoir continuer à étoffer suffisamment la base de données pour rendre par la suite la méthode complètement automatique dans tous types de situations.

Ce travail fut l'occasion de se familiariser avec certains outils de développement et de découvrir la satisfaction que la mise en place d'une méthode et d'outils associés peut procurer, ainsi que le travail de recherche en général. Ensuite, il permit de pointer l'importance de la qualité des données utilisées, aussi bien satellitaires que des observations de terrain, qui conditionnent directement les résultats obtenus.

Table des figures

1	Carte de localisation de notre zone d'étude	6
2	Illustration d'une partition de l'espace à 2 dimensions obtenu par maximum de vraisemblance [9]	9
3	Illustrations de la recherche de l'hyperplan dans le cas de la méthode SVM [15]	10
4	Visualisation des signatures spectrales obtenues pour deux pixels à partir d'une série temporelle NDVI visible [9]	11
5	Dates de transition remarquables d'un cycle de développement d'une culture annuelle, observées à l'aide d'un profil NDVI [10]. Le trait plein correspond à une série temporelle idéale d'un indice de végétation (pour nous le NDVI), et le trait pointillé sa dérivée seconde. Les points indiquent les différentes dates de transition remarquables.	12
6	Répartition temporelle des images de la série	16
7	Répartition spatiale des différentes campagnes de terrain	18
8	Schéma du fonctionnement général de la méthode	20
9	Calendrier cultural ('lab' : labouré, 'plan' : planté, 'rec' : récolté, 'taille' : taillé, 'sem' : semé).	21
10	Exemple de superposition de profils par classe et par campagne de terrain (1)	22
11	Exemple de superposition de profils par classe et campagne de terrain (2)	23
12	Exemple de superposition de profils par classe et campagne de terrain (3)	24
13	Exemple de superposition de profils par classe et campagne de terrain (4)	25
14	Exemple de superposition de profils par classe et campagne de terrain (5)	26
15	Biais observés dans les valeurs de NDVI	28
16	Exemple de profil de piment dont les variations de NDVI relèvent potentiellement de sa rugosité	29
17	Exemple de parcelles d'oliviers (haut) avec différentes distances inter-rangs, et profil NDVI associés (bas), visualisés sous ENVI	29
18	Exemple de profil illustrant un mélange de signaux. ID2=110 signifie que l'on affiche le profil correspondant à la 110ème ROI de la campagne de terrain 2, identifiée à cette date comme parcelle d'oliviers.	30
19	Exemple de profils de cultures d'une même classe, cultivés avec des pratiques différentes. En haut à gauche : céréale d'hiver non irriguée, en haut à droite : céréale d'hiver irriguée, en bas : profils de parcelles de piments conduites différemment.	31
20	Visualisation de l'outil de sélection des profils de référence	33
21	Profils temporels de NDVI interpolés selon différentes méthodes (Gauche : profil de céréale d'hiver identifié à la campagne 2, milieu : profil de piment identifié en automne, droite : profil de luzerne).	34
22	Schéma de fonctionnement des deux phases de comparaison	36
23	Représentation de la distance D_n entre deux profils (Gauche : cas de deux profils entiers, Droite : cas d'un profil interrompu)	37
24	Exemple de résultats de la phase obtenu après 'post-traitement'. A gauche : en bleu, le profil dit 'candidat'. A droite : résultat de la classification du profil. Un premier cycle de développement est identifié comme oignon (rouge), suivi d'une période de sol nu (noir), puis d'un nouveau cycle de piment (jaune).	38
25	Exemple d'affichage du résultat de la classification d'un profil	42
26	Comparaison du mode de sélection entre résultats issus des phases 1 et 2	44
27	Evolution de la performance de la classification en fonction de l'aire minimale imposée pour les cycles entiers identifiés après le seuillage à NDVI = 0.2 (Gauche : données type 1, Droite : données type 1+2)	45

28	Evolution de la performance de la classification en fonction de l'aire minimale imposée pour les bouts de profils interrompus (Gauche : données de type 1, Droite : données de type 1+2)	46
29	Evolution de la performance de la classification en fonction de la base utilisée pour le calcul des aires	46
30	Illustration de l'influence de la valeur choisie pour base du calcul de l'aire située entre les courbes	47
31	Superposition des profils stockés dans la base de données pour les couples de classe Pastèque/Melon (Gauche) et Piment/Tomate (Droite)	49
32	Histogramme des précisions utilisateurs (Haut) et des erreurs d'excédents (Bas) obtenues par classe	51
33	Réprésentation de la précision obtenue par classe en fonction du nombre de profils de la classe présents dans la base de données.	52
34	Pourcentage d'identifications pertinentes pour les mélanges de cultures	53
35	Evolution de la performance de la classification en fonction du nombre de données présentes dans la base de données (Cas des données de type 1)	54
36	Evolution de la performance de la classification en fonction du nombre de données présentes dans la base de données (Cas des données de type 1+2)	55
37	Localisation de la zone culturelle classifiée et légende associée	57
38	Exemple de rendus de carte d'occupation du sol tous les mois	58

Liste des tableaux

1	Critères utilisés pour la classification des profils NDVI par la méthode développée au CESBIO [7], avec Sn=0.18 (seuil sol nu), Sa=0.45 (valeur de NDVI considérée comme maximale pour les arbres sur sol nu) et Sr=0.15 (amplitude minimale de NDVI pour les arbres sur cultures annuelles).	13
2	Exemple théorique de matrice de confusion	15
3	Critères dits “experts” proposés pour détecter les cultures permanentes	36
4	Notations	39
5	Typologie retenue pour la classification	43
6	Matrices de confusion obtenues avant (haut) puis après (bas) regroupement des Grenadiers et Rosaceae. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.	48
7	Matrices de confusion obtenues avant (haut) puis après (bas) application des critères experts. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.	50
8	Matrices de confusion finales obtenues sur les données de type 1 (haut) et sur les données type 1+2 (bas). Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.	51
9	Matrice de confusion correspondant au test de profils de mélanges de cultures. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.	53
10	Matrice de confusion correspondant au test de cultures minoritaires. Sont surlignées en orange les valeurs de la diagonale (qui correspondent au nombre de profils bien classés par classe), en vert clair les valeurs non nulles de la matrice, et en vert foncé les valeurs supérieures à 5.	54

Liste des algorithmes

1	Algorithme de découpage des profils	39
2	Algorithme de classification utilisé dans la phase 1	39
3	Algorithme de classification utilisé dans la phase 2	40
4	Algorithme de classification de comparaison des phases 1 et 2	40

Bibliographie

- [1] Remote sensing and digital image processing. Custom edition for University of Calgary. Pearson, 2007.
- [2] Envcal-suivi de l'environnement par teledetection. Les indices simples, [en ligne]. Disponible sur : <http://e-cours.univ-paris1.fr/> (consulté le : 10/06/2015)
- [3] **Kastens Jude H., Wardlow Brian D., Egbert Stephen L.** Analysis of time-series modis 250 m vegetation index data for crop classification in the u.s. central great plains. *Remote Sensing of Environment*, 108, 290-310, 2007.
- [4] **Furlan D., Rodrigues A., Cunha M., Marcel. A., and Ballester M-V.** Land cover classification in rondonia (amazonia) using ndvi time series data from spot-vegetation . 33rd Canadian Symposium on Remote Sensing, 2012.
- [5] **Lacombe G.** Evolution et usages de la ressource en eau dans un bassin versant aménagé semi-aride. Le cas du Merguellil en Tunisie Centrale. Mémoire de thèse. Université de Montpellier II, 2007.
- [6] **Masse A.** Développement et automatisation de méthodes de classification à partir de séries temporelles d'images de télédetection : Application aux changements d'occupation des sols et à l'estimation du bilan carbone . Mémoire de thèse. Université Paul Sabatier de Toulouse, 2013.
- [7] **Helson D. Er-Raki S. Olioso A. Chehbouni A.G. Simonneaux V., Duchemin B.** The use of highresolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central morocco. *International Journal of Remote Sensing*, 29 :1, 95 - 116, 2008.
- [8] **Ingla J. Tupin F., Nicolas J.-M.** Imagerie de télédétection . Lavoisier, 2014.
- [9] **Simonneaux V.** Caractérisation de l'occupation du sol par télédétection. Support power point de présentation au CESBIO, 2003.
- [10] **Schaaf C. Strahler A. Hodges J. Gao F. Reed B. Huete A. Zhang X., Friedl M.** Monitoring vegetation phenology using modis. *Remote Sensing of Environment* 84, 471-475, 2003.
- [11] **Erwann Joachim F.** Application de la télédétection à l'analyse de la variabilité climatique des régions boréales et subarticques du Canada et à la validation du modèle régional canadien du climat. Mémoire de thèse. Université de Sherbrooke, 2003.
- [12] **Tardy B.** Comparaison d'algorithmes de classification pour la production de cartes d'occupation des sols à partir de séries temporelles d'images optiques à haute résolution. Mémoire de stage, Université Paul Sabatier, 2014.
- [13] **Lamachere J.-M., Puech C.** Télédétection, états de surface et observations au sol.
- [14] Statistiques mondiales. Tunisie, [en ligne]. Disponible sur : www.statistiques-mondiales.com (consulté le 24/05/2015)
- [15] Wikipedia. Machine à vecteurs de support, [en ligne]. Disponible sur : <https://fr.m.wikipedia.org/wiki/> (consulté le : 25/05/2015)

-
- [16] Wikipedia. Validation croisée, [en ligne]. Disponible sur : https://fr.m.wikipedia.org/wiki/Validation_crois%C3%A9e (consulté le : 29/05/2015)
- [17] EOEd - Petit guide de la télédétection. Corrections radiométriques, [en ligne]. Disponible sur : eoedu.belspo.be (consulté le : 06/06/215)
- [18] Introduction à la télédétection. La télédétection optique, [en ligne]. Support de cours de l'ENSG disponible sur : fad.ensg.eu/moodle/ (consulté le 06/06/2015)
- [19] **Hird Jennifer N., McDermid Gregory J.** Noise reduction of NDVI time series : An empirical comparison of selected techniques. *Remote Sensing of Environment*, 113, 248-258, j2008.
- [20] **Ducrot, D.** Methodes d'analyses et d'interprétation d'images de télédétection multi-sources. Extraction des caractéristiques du paysage. Mémoire HDR. INP Toulouse, 2005.
- [21] Envcal-suivi de l'environnement par télédétection. Réflexion, [en ligne]. Disponible sur : <http://e-cours.univ-paris1.fr/>

Annexes

ANNEXE 1 : Photographies de la zone

ANNEXE 2 : Résumé

ANNEXE 3 : Présentation du modèle SAMIR

ANNEXE 1 :

Photographies de la zone



Cultures de piments au milieu d'oliviers



Cultures de tomates au milieu d'oliviers



Fruitiers caduques



Fauchage de blé au Maroc . Les mêmes conditions sont observées en Tunisie.



Champs de blé

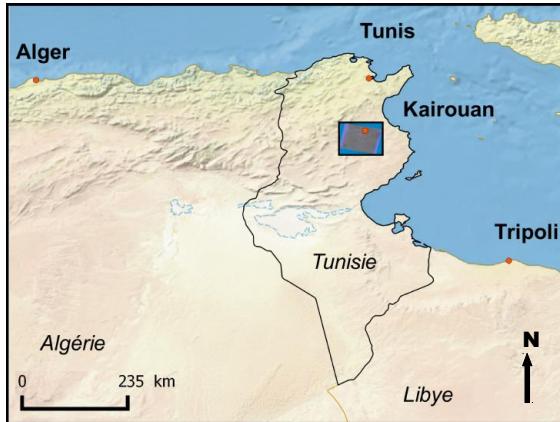
ANNEXE 2

Résumé :

Cartographie d'occupation du sol à partir de séries temporelles d'images haute résolution visible NDVI

Introduction

Au Maghreb, la forte variabilité des cycles de développement des cultures, liée aux différentes pratiques agricoles ainsi qu'aux conditions météorologiques instables, rend difficile l'utilisation de méthodes classiques de cartographie d'occupation du sol. Nous proposons de développer puis de tester de nouvelles méthodes plus automatiques et robustes, permettant d'exploiter le potentiel offert par les séries temporelles haute résolution visible NDVI dans le bassin du Merguellil, en Tunisie. Nous disposons pour cela d'une série temporelle SPOT5 de 14 images corrigées en réflectance, acquises entre novembre 2012 et novembre 2013, avec une résolution spatiale de 10 m en multispectral. Elle permet de simuler convenablement les futures images Sentinel-2 prochainement disponibles.



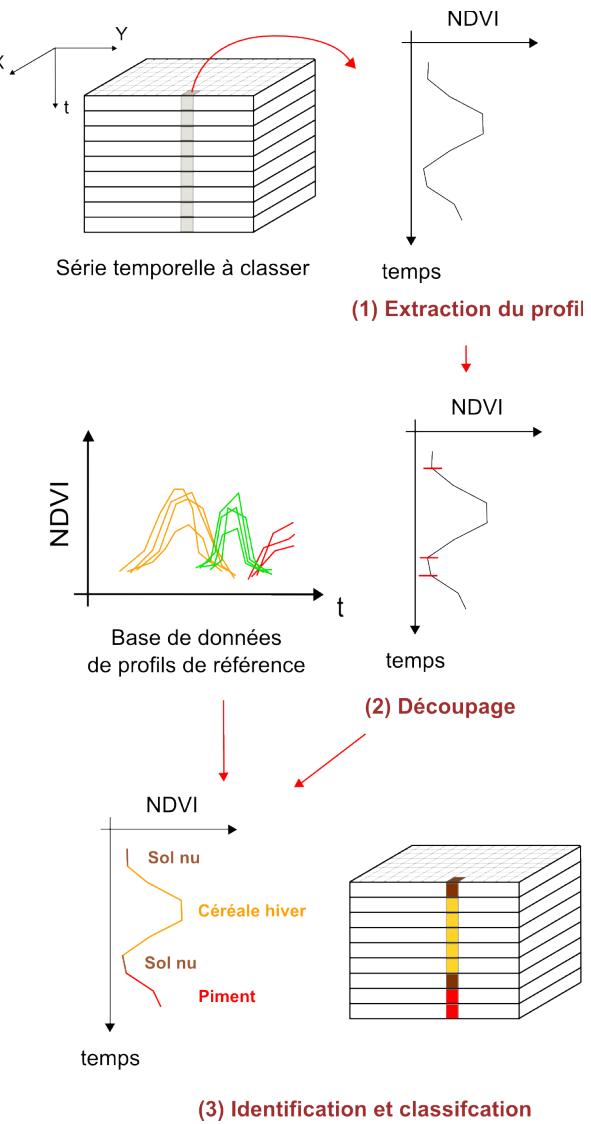
Localisation de la zone d'étude

Méthode

1 - Principe général

Pour une série temporelle à classer, on cherche à obtenir pour chaque pixel les différentes occupations du sol qui se sont succédées. Le profil temporel NDVI de chaque pixel est reconstruit à partir de la série temporelle (1). Il présente l'avantage de rendre compte de réels paramètres biophysiques et va servir de base à la classification.

Il convient d'extraire pour chacun de ces profils les paramètres en question, à savoir les dates de début de croissance et de fin de sénescence. On obtient une première segmentation temporelle du profil (2), préalable à l'identification de la nature des cycles de cultures ainsi obtenus. A ce stade, une approche base de données est utilisée (3). Chaque cycle de développement de chaque profil est comparé à des données de référence en vue d'identifier sa nature. La comparaison repose aussi bien sur des critères de similitude de forme que de synchronisation temporelle.

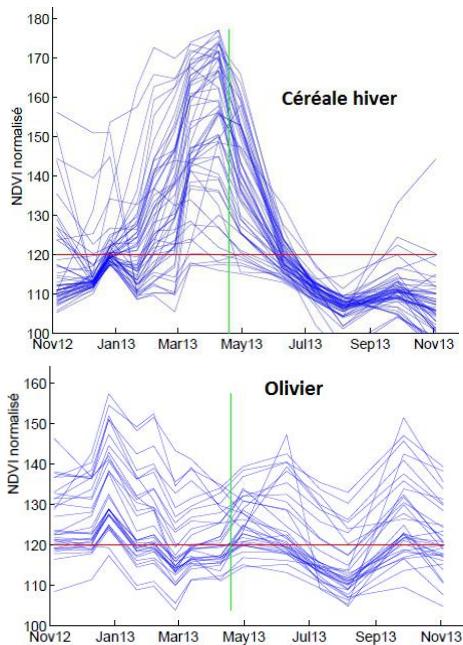


2 – Démarche de développement

Dans ce cadre général, la définition précise des algorithmes à développer nécessite une certaine connaissance des caractéristiques phénologiques des cultures observées. A partir des données de terrain dont l'on dispose, on s'attache à analyser les profils temporels NDVI de chaque classe de cultures, afin de tenter d'en dégager des caractéristiques propres à chacune d'entre elles. Un affichage des profils par classe nous permet dégager certaines tendances générales.

Tout d'abord, on distingue facilement les signatures spectrales des cultures annuelles de celles des permanentes. Les cultures annuelles présentent une variation de NDVI rapide et de forte amplitude. Le cycle de développement, d'une période maximale de 5 mois, est précédé puis suivi d'une période de sol nu. Au contraire, les cultures permanentes affichent en général des motifs de variations pseudo-périodiques d'amplitudes plus faibles. Se distinguent également parmi ces dernières les arbres à feuilles persistantes des arbres à feuilles caduques, pour lesquels la chute des feuilles observée en automne se traduit par un passage du NDVI en-dessous du seuil sol nu (fixé à NDVI=0.2).

On remarque également que s'il est possible de dégager des caractéristiques propres à chaque classe, une analyse plus détaillée des profils révèle une variabilité intra-classe à ne pas négliger. Celle-ci peut provenir d'espacements différents des cultures au sol, de l'impact de la rugosité des surfaces sur la réflexion du signal enregistré, de la subjectivité de l'observation de terrain, de pratiques agricoles différentes ou encore des conditions météorologiques de l'année en cours.



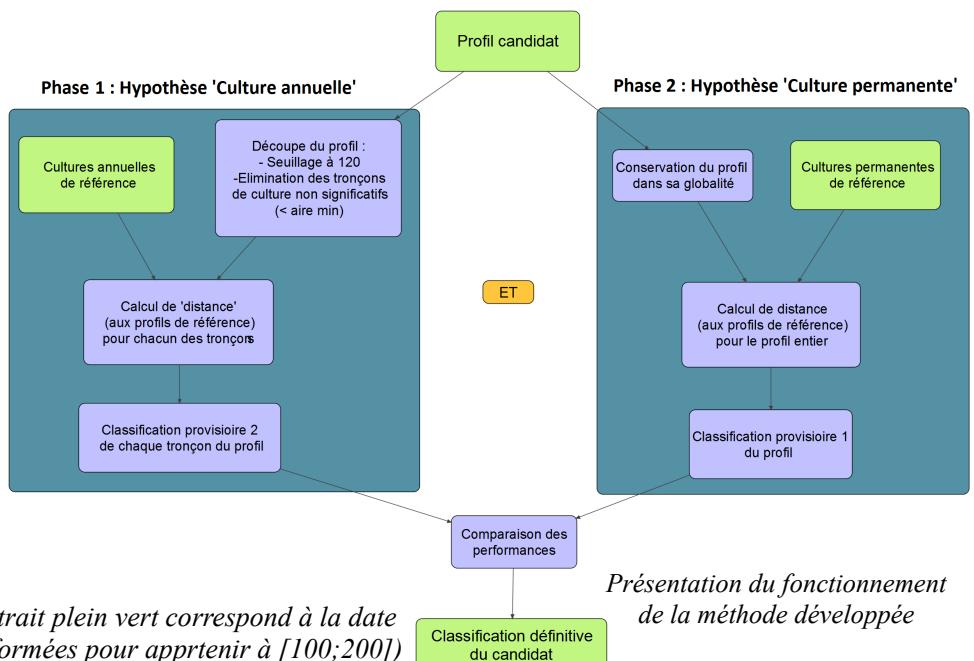
Exemple de superposition de profils par classe (le trait plein vert correspond à la date d'identification et les valeurs NDVI ont été retransformées pour appartenir à [100;200])

Cette analyse préalable des données nous permet d'aborder ensuite la définition de la méthode avec une meilleure expertise. Elle permet également de nous aider à extraire des profils de référence leurs sections exploitables. Ceci est effectué sur les quelques 700 données de références dont l'on dispose grâce à un outil que nous avons développé. Celui-ci permet à l'utilisateur de sélectionner de façon interactive les cycles de développement à enregistrer dans la base de données.

3 – Méthode retenue

Après plusieurs expérimentations plus ou moins concluantes, on parvient à définir une méthode qui s'appuie sur des critères robustes. Elle repose sur deux phases d'identification parallèles, pour un même profil 'candidat' : l'une correspond à l'hypothèse 'culture permanente' et consiste à considérer le profil dans son ensemble, tandis que la seconde suppose la présence de cultures annuelles et nécessite au préalable une phase de découpe du profil visant à en extraire les différents cycles végétatifs. Dans les deux cas, l'identification repose sur la comparaison, par un calcul de distance, du cycle de développement à l'étude avec tous ceux présents dans la base de données. La nature de celui qui maximise la ressemblance est conservée. Finalement, on choisit pour classification définitive le résultat le plus pertinent de celui issu des deux phases.

Détaillons plus en détails les algorithmes utilisés. Lors de la phase 1, les différents cycles végétatifs présents au sein d'un même profil temporel sont extraits par seuillage des valeurs à 0.2 (seuil sol nu). Une aire minimale concernant leur intégrale est fixée expérimentalement pour ne pas conserver les cycles résultant du bruit des mesures ou encore du développement de végétation parasite (adventice).



L'identification s'effectue ensuite, dans le cas des deux phases, par le calcul de l'aire située entre les deux profils : le candidat et celui de référence. Ce calcul d'intégrale entre deux dates t_1 et t_2 permet de rendre

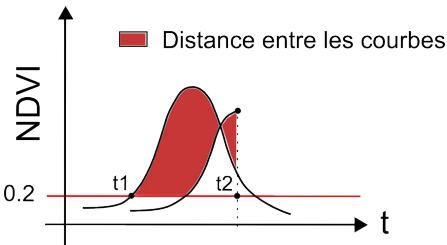


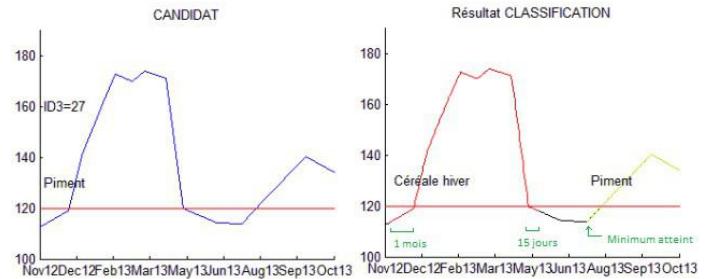
Schéma de la distance entre les courbes, dans le cas d'un profil interrompu par l'acquisition d'images

compte de la différence d'amplitude des courbes, mais également de leur décalage temporel. On choisit d'appliquer ce calcul de 'distance' aux portions de courbes supérieures à $NDVI=0.2$, étant donné la faible significativité des valeurs situées en-dessous. De plus, on fera attention à distinguer les cycles végétatifs entiers, de ceux interrompus par l'acquisition des images de la série temporelle. Dans ce dernier cas, les bornes de l'intégrale correspondront aux éventuelles dates de coupures. Dans tous les cas, la valeur de cette intégrale est normalisée : on la divise par l'intervalle temporel t_2-t_1 . Ceci permet de comparer entre elles les différentes distances obtenues. Enfin, le choix de la phase retenue pour la classification repose sur la comparaison de la moyenne des distances calculées pour chacun des 'tronçons' de profils issus de la phase 1, à la valeur de la distance issue de la phase 2. La phase qui présente la moyenne des distances la plus basse (et donc une meilleure similitude globale entre le candidat et les références) est retenue.

Un traitement post-comparaison est ensuite appliqué pour affiner la classification temporelle du profil et mieux rendre compte de la réalité phénologique des cultures. En effet, un certain temps s'écoule entre la date de semis et la date à laquelle le NDVI devient visible (franchit le seuil 0.2). A fortiori, le même phénomène s'observe concernant la sénescence. Pour rendre compte de cette réalité dans notre classification, on décide d'identifier ces débuts et fins de cycles en fixant arbitrairement 1 mois entre la date de semis et de NDVI visible puis 15 jours entre la fin de la sénescence et la fin du cycle complet. Si, au cours de ces intervalles temporels pré ou post végétation, un minimum est observé sur le profil de NDVI, il est choisi comme date de début ou de fin de cultures.

Résultats

La visualisation de la classification temporelle de chaque pixel est programmée sous Matlab, et aide à l'interprétation des résultats.



Exemple d'affichage de la classification temporelle d'un profil

Afin de calculer la performance globale de la classification, il est nécessaire de disposer de données propres et représentatives. Le jeu de données est trié en quatre groupes :

- les données de type 1, « typiques ». Ce sont les plus propres et on espère pouvoir les identifier en majorité.
- les données de type 2, « atypiques significatives », que l'algorithme peut potentiellement détecter même si leur allure est atypique.
- les données de type 3, « mal identifiées », qui présentent un problème d'identification.
- les données de type 4, « mélanges ou minoritaires », qui correspondent à des mélanges de cultures ou à des espèces peu fréquentes.

Ainsi, afin de rendre compte de la qualité de l'algorithme et de la méthode proposée, les tests de performance seront principalement menés sur les données de type 1 et/ou 2.

De plus, l'analyse des profils effectuée en amont permet de définir une typologie relative à la classification. Dans un premier temps, 15 classes sont conservées. On distingue les cultures annuelles (piment, céréale d'hiver, fourrage, oignon, tomate, fève, pastèque, melon, fourrage annuel) des cultures permanentes (olivier, *rosaceae*, agrume, grenadier, fourrage permanent, luzerne, jachère). Une fois ces typologies établies, une série de tests est menée en leave-one-out afin d'affiner les paramètres utilisés dans la méthode de classification. Le calcul de l'indice kappa et de l'Overall Accuracy associés à la classification pour plusieurs valeurs de seuils d'intégrales aboutit à retenir la valeur de 450 unités dans le cas des cycles entiers, et de 150 dans le cas des cycles interrompus par l'acquisition des images. De même, des tests similaires permettent d'accéder à la valeur optimale de la base choisie pour le calcul des aires de comparaison des profils. En effet, celle-ci influe sur le poids relatif affecté aux valeurs NDVI inférieures à 120 par rapport au

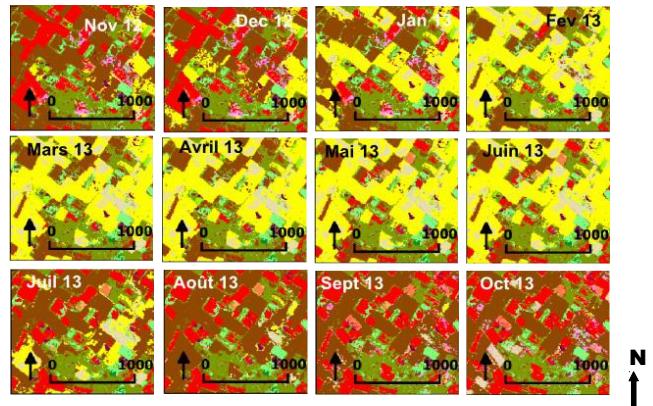
poids du reste du profil. Au final, la valeur de 120 s'avère être celle offrant la meilleure performance pour la classification.

A la vue des premiers résultats issus des matrices de confusion, certains regroupements de classes sont étudiés. La performance de la classification apparaît meilleure lorsque les classes 'grenadier' et 'rosaceae' sont regroupées. La typologie est modifiée et ces deux classes apparaissent sous le nom 'd'arbres à feuilles caduques'.

Les confusions observées entre certaines classes, notamment entre les classes piment/tomate et melon/pastèque, mènent à réfléchir sur l'introduction de critères « experts » visant à moduler dans certains cas le résultat issu de la classification des profils. Des critères statistiques, et donc robustes, sont imaginés. En effet, on remarque globalement que la pente des profils de tomate est supérieure à celle des profils de piments entre mi-août et mi-septembre. De même, l'intégrale des profils de pastèque est en moyenne supérieure à celle des melons. Ainsi, dans le cas du couple de classes piments/tomates, la pente du profil candidat est comparée aux pentes moyennes des profils des classes piments et tomates stockés dans la base de données. La classe minimisant l'écart de pente est retenue. Le même processus basé sur l'intégrale des profils est utilisé dans le cas melon/pastèque. Leur emploi permet d'augmenter globalement la précision des classes concernées.

Finalement, la précision globale de notre méthode atteint les 71 % sur les données de type 1+2, ce qui s'avère prometteur dans le cadre de notre utilisation. La précision utilisateur s'élève respectivement à 79 % et 94 % pour les classes piment et céréale d'hiver qui sont les plus courantes dans notre zone d'étude. De plus, un nouveau test permet de calculer que dans le cas de mélanges de cultures, 77 % des profils sont identifiés par l'algorithme comme l'un ou l'autre des types de cultures présents sur la parcelle. Enfin, l'utilisation d'échantillons 'bootstraps' pour la classification permet d'évaluer l'influence du nombre de données stockées en référence. Si la performance ne cesse de croître avec l'augmentation du nombre de ces données, on observe une diminution nette de l'influence de leur nombre à partir de 40 % des données utilisées comme référence pour chaque classe.

Concernant les rendus, on programme un affichage de la classification obtenue pour un pas temporel choisi par l'utilisateur.



Exemple de classification mensuelle d'une zone

Perspectives d'améliorations

Il sera nécessaire de tester la classification d'une autre série temporelle acquise sur une année différente. On s'attend alors à une diminution de la performance, liée à des confusions dues aux erreurs de corrections radiométriques et/ou à des conditions météorologiques différentes. Pour y palier, pourrait être envisagé une amélioration des corrections radiométriques des images, un étoffement de la base de données, l'utilisation de techniques de traitement d'images très haute résolution pour la détection des oliviers ainsi qu'un travail approfondi sur le lissage et la paramétrisation des profils. De plus, on pourrait imaginer appliquer cette méthode à une unique image, grâce à une comparaison en une dimension des valeurs de NDVI. Ceci permettrait d'obtenir instantanément une première hypothèse quant à la nature de la culture dès l'acquisition d'une image.

Conclusion

Cette méthode innovante permet de reconstruire la succession temporelle des différents types de cultures observés sur une parcelle, en exploitant le potentiel offert par les séries temporelles haute résolution visible (NDVI). Elle repose sur une approche base de données, qui permet de palier à la variabilité des cycles de développements observée dans la zone, où les méthodes classiques de cartographie d'occupation s'avèrent peu efficaces. Ceci permet également d'envisager de s'affranchir à long terme de données de terrain, une fois que la base de donnée sera suffisamment étoffée. Les résultats obtenus démontrent le potentiel de la méthode : elle présente pour ce jeu d'images une précision globale de 71%, associée à un kappa de 76%. Toutefois, des incertitudes persistent quant à l'influence du jeu d'images utilisé, bien que de nombreuses pistes d'améliorations soient envisageables.

Annexe 3 : Présentation du modèle SAMIR

ET : Evapotranspiration : Quantité d'eau transférée vers l'atmosphère, par l'évaporation au niveau du sol et par la transpiration des plantes.

SAMIR (SAtellite Monitoring of IRrigation) est un logiciel *de spatialisation de l'évapotranspiration*, qui permet le suivi du bilan hydrique des surfaces irriguées et donc de faciliter la gestion des ressources en eau à l'échelle du bassin-versant. Il repose aussi bien sur des informations de terrain, que sur des mesures satellitaires et des modèles physiques.

I . Principe du pilotage de l'irrigation

Le principe du pilotage de l'irrigation dépend de l'échelle considérée. Au niveau de quelques parcelles, un agriculteur est surtout intéressé par la consommation des cultures au jour le jour. Il peut ainsi ajuster les apports en eau, à condition d'en disposer à volonté. Dans ce cas, les volumes d'irrigation peuvent facilement être obtenus par la méthode FAO (Food and Agriculture Organization of the United Nation) sous réserve d'avoir accès à des données climatiques et de pouvoir quantifier la végétation présente dans les champs. Des mesures d'humidité du sol, sans être nécessaires, peuvent améliorer les estimations. Dans ce contexte, les données nécessaires sont facilement accessibles, et il s'agit surtout de bien maîtriser la technique d'estimation des besoins.

Au contraire, l'information nécessaire pour une gestion régionale de l'irrigation est beaucoup plus importante et moins facilement accessible. Il s'agit notamment de savoir ce qui a été consommé depuis le début d'une saison à l'échelle de secteurs d'irrigations, mais encore de connaître la consommation à un instant t , ou d'estimer ce qui sera nécessaire pour terminer la saison. C'est dans ce cadre que s'inscrit le logiciel SAMIR, outil d'aide à la gestion de l'irrigation, à vocation opérationnelle et basé sur l'utilisation de séries d'images satellitaires.

II . Fonctionnement de SAMIR

Principe général

SAMIR permet de réaliser une estimation spatialisée de l'ET et du bilan hydrique des cultures à l'échelle des paramètres irrigués. En sont déduits les apports en irrigation.

Calcul du bilan hydrique

L'ET d'un couvert est la somme de la transpiration de la végétation et de l'évaporation du sol. La méthode FAO modélise l'ET de n'importe quel type de végétation en la comparant, via des coefficients culturaux empiriques (calculés dans notre cas sur la base de valeurs NDVI issues des images satellites, en fonction de l'occupation du sol), à celle d'un gazon « standard » bien arrosé, placé dans les mêmes conditions de température, d'humidité de l'air, de vitesse du vent et de rayonnement solaire. Cette ET de référence (ET0) représente la « demande climatique » à un moment donné, elle peut être calculée au moyen de l'équation de Penman-Monteith. À partir de là, l'ET d'un couvert quelconque, notée ETR (ET réelle), est obtenue par l'équation suivante :

$$ETR = ET0 \times (Kcb \times Ks + Ke)$$

avec :

- Kcb : coefficient basal crop déterminant la transpiration de la fraction de sol couverte par la végétation (fc) ;
- Ks : coefficient de stress hydrique ;

– K_e : coefficient déterminant l'évaporation de la fraction de sol nu ($1 - f_c$).

Le calcul de ETR nécessite donc trois types de données : des variables climatiques pour le calcul de l'ET₀, l'occupation du sol et le développement de la végétation permettant d'estimer les coefficients culturaux K_{cb} et K_e . Le coefficient K_s est calculé à partir de l'état hydrique du sol et permet de réduire l'évaporation maximale du couvert en conditions bien arrosées, correspondant à K_{cb} .

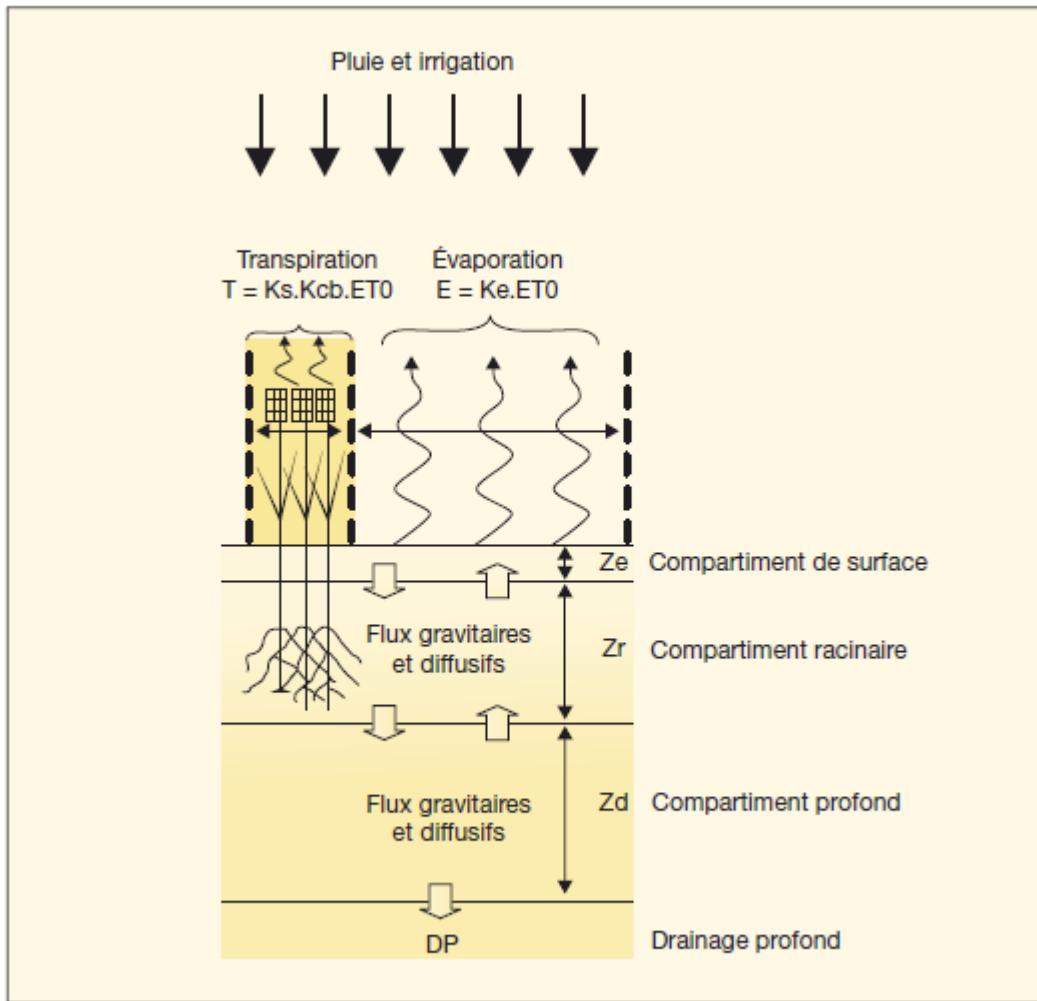
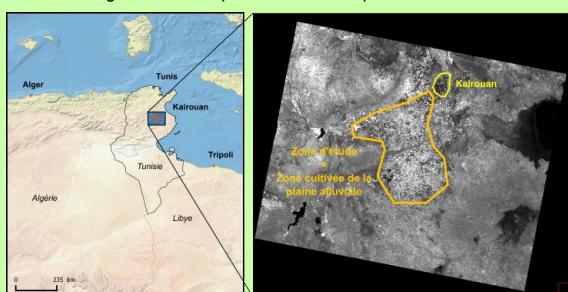


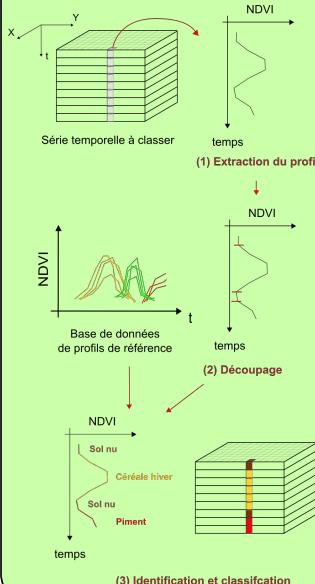
Figure 1. Modélisation des transferts d'eau entre sol, végétation et atmosphère dans SAMIR.

Introduction

Au Maghreb, la forte variabilité des cycles de développement des cultures, liée aux différentes pratiques agricoles ainsi qu'aux conditions météorologiques instables, rend difficile l'utilisation de méthodes classiques de cartographie d'occupation du sol. Nous proposons de développer puis de tester de nouvelles méthodes plus automatiques et robustes, permettant d'exploiter le potentiel offert par les séries temporelles haute résolution visible NDVI dans le bassin du Merguellil, en Tunisie. Nous disposons pour cela d'une série temporelle SPOT5 de 14 images corrigées en réflectance, acquises entre novembre 2012 et novembre 2013, avec une résolution spatiale de 10 m en multispectral. Elle permet de simuler convenablement les futures images Sentinel-2 prochainement disponibles.



Méthode générale



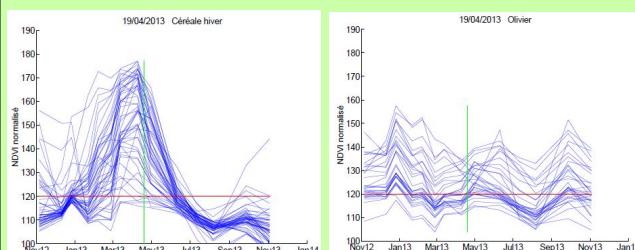
Pour une série temporelle à classer, on cherche à obtenir pour chaque pixel les différentes occupations du sol qui se sont succédées. Le profil temporel NDVI de chaque pixel est reconstruit à partir de la série temporelle (1). Il présente l'avantage de rendre compte de réels paramètres biophysiques et va servir de base à la classification.

On identifie les dates de début de croissance et de fin de sénescence d'un profil : on obtient une première segmentation temporelle (2), préalable à l'identification de la nature des cycles de cultures ainsi obtenus.

Une approche base de données est utilisée (3) : chaque cycle de développement de chaque profil est comparé à des données de référence en vue d'identifier sa nature. La comparaison repose aussi bien sur des critères de similitude de forme que de synchronisation temporelle.

Démarche de développement

Nécessité d'appropriation des données terrain et des caractéristiques phénologiques des cultures -> visualisation des profils par classe :

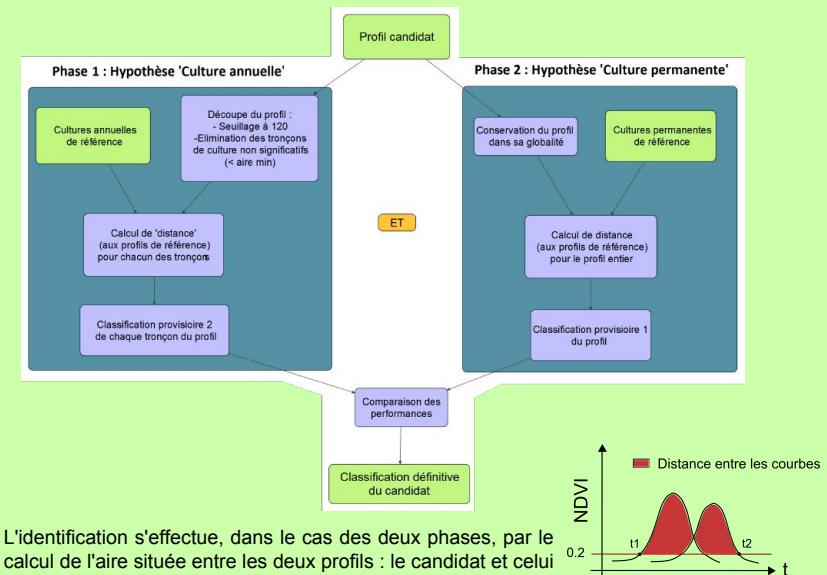


Il en ressort :

- Distinction des signatures spectrales des cultures annuelles (variation de NDVI rapide et de forte amplitude) de celles des permanentes (en général, motifs de variations pseudo-périodiques d'amplitudes plus faibles).
- Des caractéristiques propres à chaque classe de cultures.
- Une variabilité intra-classe, différentes ou encore des conditions météorologiques de l'année en cours.

Méthode détaillée retenue

Elle repose sur deux phases d'identification parallèles, pour un même profil 'candidat' :



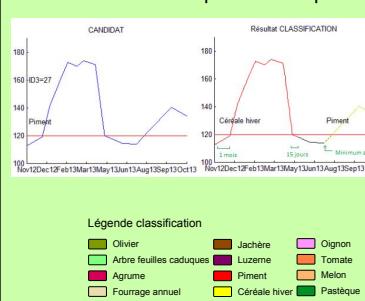
L'identification s'effectue, dans le cas des deux phases, par le calcul de l'aire située entre les deux profils : le candidat et celui de référence.

- > Tests d'affinage des paramètres de la classification (seuils utilisés, ...)
- > Rajouts de critères "experts" pour les classes confuses (couples piments/tomates et melons/pastèque)
- > Test du nombre optimal de données à stocker en référence

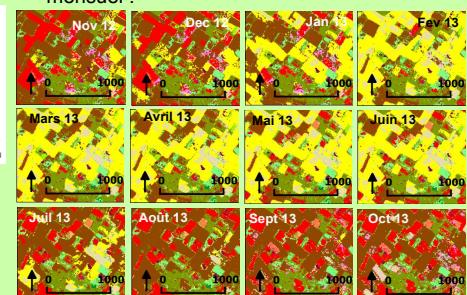
Résultats

Affichages :

-> Classification temporelle d'un profil :



-> Classification d'un extrait de la zone au pas mensuel :



Performance :

- > Précision globale de 71%
- > Précisions utilisateurs de respectivement 79% et 94% pour les classes piment et céréale d'hiver, qui sont les plus répandues dans la zone.
- > 77% des profils de mélanges de cultures sont identifiés comme étant l'un ou l'autre des types de cultures présents sur la parcelle.

Perspectives d'amélioration

- > Tester la classification sur un autre jeu d'images
- > Améliorer les corrections radiométriques des images
- > Etoffer la base de données dans les prochaines années
- > Tester l'utilisation de paramétrages et de modélisations des profils
- > Détecter les oliviers par des méthodes de traitement d'images très haute résolution
- > Appliquer la méthode avec moins d'images, voir avec une seule

Conclusion

Potentiel de la méthode démontré :

- > Reconstruction de la succession temporelle des différents types de cultures
- > Approche base de données
- > Performance et robustesse satisfaisante
- > Pistes d'améliorations à creuser

RESUME

Au Maghreb, la forte variabilité des cycles de développement des cultures, liée aux différentes pratiques agricoles ainsi qu'aux conditions météorologiques instables, rend difficile l'utilisation de méthodes classiques de cartographie d'occupation du sol. Nous proposons de développer puis de tester de nouvelles méthodes plus automatiques et robustes, permettant d'exploiter le potentiel offert par les séries temporelles haute résolution visible NDVI dans le bassin du Merguellil, en Tunisie.

Une méthode permettant de reconstituer les successions de cultures a pu être mise en œuvre et testée à l'aide d'analyse de données satellitaires SPOT (simulant les futures images Sentinel-2), couplée à des observations de terrain. La reconstitution des successions de cultures est alors obtenue avec une précision globale de 72 %. Une approche base de données est utilisée pour identifier les différents cycles de développement cultureaux à partir du profil temporel NDVI de chaque pixel. Les résultats obtenus démontrent la pertinence de l'approche innovante ici testée.

Mots clés : irrigation, agriculture, classification, télédétection, Tunisie, occupation du sol, série temporelle, NDVI, algorithme

ABSTRACT

Agricultural practices and weather conditions leads to a high variability of crop development cycles in the Maghreb, which makes land cover mapping using conventional methods difficult. This study aims to develop and test new automatic and robust methods that would exploit the potential the high-resolution imagery time series offer. The study is conducted in the Merguellil watershed, in Tunisia.

From SPOT satellite data (with which futures Sentinel-2 images can be simulated) and ground data, we managed to develop a method that enables us to retrace crop successions with a global accuracy that is up to 72%. The identification of vegetation development cycles relies on a database approach based on the NDVI temporal profile of each pixel. Preliminary results confirm the soundness of the innovative approach we have tested.

Keywords : irrigation, agriculture, classification, remote sensing, Tunisia, land cover map, time series, NDVI, algorithm.