

TABLE DES MATIÈRES

REMERCIEMENTS	4
RÉSUMÉ	7
ABSTRACT	8
TABLE DES MATIÈRES	9
ACRONYMES	11
INTRODUCTION	13
CHAPITRE 1. CONTEXTE ET ÉTAT DE L'ART :	18
SÉCURITÉ ALIMENTAIRE ET TÉLÉDÉTECTION POUR LE SUIVI DES CULTURES	18
1. CONTEXTE : SECURITE ALIMENTAIRE	18
1.1. Définitions et situation en 2010 dans le monde	18
1.2. La sécurité alimentaire en Afrique subsaharienne	20
1.2.1. Pression démographique	20
1.2.2. Pression foncière	20
1.2.3. Déclin de la fertilité du sol	21
1.2.4. Autres obstacles	21
1.3. L'exemple du Mali	22
1.3.1. Présentation générale	22
1.3.2. Les systèmes agricoles	23
1.3.3. La sécurité alimentaire au Mali	25
2. TELEDETECTION ET SUIVI DES CULTURES	26
2.1. Cartographie du domaine cultivé	27
2.2. Cartographie des systèmes agricoles	30
2.3. Suivi temporel et estimation du rendement	31
3. LES SYSTEMES D'ALERTE PRECOCE POUR LA SECURITE ALIMENTAIRE EN AFRIQUE DE L'OUEST	33
3.1. Définitions et présentation des systèmes existants	33
3.2. Les données d'entrée des systèmes d'alerte précoce pour le suivi agricole	33
3.2.1. Les données satellitaires	34
3.2.2. Les données de terrain	37
CHAPITRE 2. DONNÉES ET OUTILS UTILISÉS	39
1. DONNEES SATELLITAIRES	39
1.1. Les produits satellitaires MODIS	39
1.2. Les images Landsat	42
1.3. Les images SPOT	42
1.4. Indices utilisés	43
1.4.1. Indices spectraux	43
1.4.2. Indices texturaux	44
1.4.3. Indices phénologiques	45
2. LES PRODUITS GLOBAUX EXISTANTS	45
3. DONNEES TERRAIN	45
3.1. Points GPS	45
3.2. Base de données CMDT / IER	48
3.3. Autres données	50
4. LES OUTILS	51
4.1. Précision de classification	51
4.2. Fouille de données	52
4.3. Random Forest	53
4.3.1. Définitions	53
4.3.2. Le paquet R randomForest	54
4.3.3. Evaluation du modèle de classification	54

CHAPITRE 3. CARTOGRAPHIE DU DOMAINE CULTIVÉ :	55
MÉTHODES ET PRINCIPAUX RÉSULTATS	55
1. METHODE GEOMATIQUE	55
2. RESULTATS	57
2.1. A l'échelle locale	57
2.1.1. Validation quantitative	57
2.1.2. Comparaison de la distribution spatiale des cultures	57
2.2. A l'échelle nationale	58
2.2.1. Validation quantitative	58
2.2.2. Comparaison de la distribution spatiale des cultures	58
3. COMPARAISON A LA FOUILLE DE DONNEES	59
3.1. Extraction de connaissances	59
3.2. Comparaison aux points terrain	60
3.3. Comparaison avec la référence SPOT	60
4. DISCUSSIONS	61
CHAPITRE 4. CARACTÉRISATION DES SYSTÈMES AGRICOLES	64
1. METHODE	64
1.1. Préparation du jeu de données	64
1.2. Modèle Random Forest	65
2. RESULTATS	66
2.1. Précision du modèle	66
2.2. Importance des variables	67
2.3. Importance des jeux de variables	67
2.4. Cartographie des systèmes agricoles du Sud Mali à l'échelle villageoise	68
3. DISCUSSIONS	69
CHAPITRE 5. SUIVI DE LA CROISSANCE DES CULTURES	72
1. METHODE	73
1.1. Evaluation du produit MCD12Q2	73
1.2. Extraction des indicateurs phénologiques observés (MODIS MCD12Q2)	73
1.3. Extraction des indicateurs de la phénologie simulés (LAI)	74
1.4. Comparaison des deux types d'indicateurs phénologiques	74
2. RESULTATS	75
2.1. Evaluation du produit MODIS MCD12Q2	75
2.2. Les valeurs d'indicateurs phénologiques sur les stations synoptiques	78
2.3. Les indicateurs phénologiques simulés (LAI)	78
2.4. Comparaison en 2007 sur 8 stations synoptiques	79
2.5. Comparaison de 2001 à 2008 sur 2 stations	79
3. DISCUSSIONS	81
CONCLUSIONS ET PERSPECTIVES	84
1. CONTRIBUTIONS DE LA THESE	84
1.1. Apports méthodologiques	84
1.2. Apports thématiques	85
2. PERSPECTIVES	86
2.1. Perspectives de recherche	87
2.2. Perspectives opérationnelles	88
2.2.1. Transférabilité	88
2.2.2. Intégration des résultats dans les systèmes d'alerte précoce	89
LISTE DES FIGURES ET DES TABLEAUX	90
BIBLIOGRAPHIE	92
ANNEXES	99

ACRONYMES

Organismes, institutions, programmes

AGRHYMET	Centre Régional Agro-Hydro-Météorologique
CILSS	Comité permanent Inter-états de Lutte contre la Sécheresse au Sahel
CMDT	Compagnie Malienne Du Textile
CNRM	Centre National de Recherches Météorologiques
DIAPER	DIagnostic PERmanent pour la sécurité alimentaire régionale (Programme conduit par le centre Agrhymet)
ESA	European Spatial Agency
FAO	Food and Agriculture Organization of the United Nations
FEWS NET	Famine Early Warning Systems Network
GIEWS	Global Information and Early Warning System (SMIAR en français)
GMES	Global Monitoring for Environment and Security
GMFS	Global Monitoring for Food Security
GSE	Surveillance Globale de l'Environnement et de la Sécurité
IER	Institut d'Economie Rurale (Mali)
IGPB	International Geosphere-Biosphere Programme
JRC	Joint Research Centre
NASA	National Aeronautics and Space Administration
PAM	Programme Alimentaire Mondial
SAP	Système d'Alerte Précoce
SAP	Système d'Alerte Précoce (au Mali)
SMIAR	Système Mondial d'Information et d'Alerte Rapide sur l'alimentation et l'agriculture (GIEWS en anglais)
UEMOA	Union Economique et Monétaire Ouest-Africaine
USAID	United States Agency for International Development

Satellites, indicateurs

AVHRR	Advanced Very High Resolution Radiometer
EVI	Enhanced Vegetation Index
HRS	Haute Résolution Spatiale
LAI	Leaf Area Index
MNE	Modèle Numérique d'Elevation

MNT	Modèle Numérique de Terrain
MODIS	MODerate resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
RFE	Rainfall Estimates
SPOT	Satellite pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
THRS	Très Haute Résolution Spatiale
TM	Thematic Mapper
VPI	Vegetation Productivity Index

Autres

GPS	Global Positioning System
LCCS	Land Cover Classification System
RGA	Recensement Général de l'Agriculture (au Mali)

INTRODUCTION

L'insécurité alimentaire s'aggrave tout particulièrement en Afrique, où la proportion de la population souffrant de faim chronique est la plus élevée au monde. Face à l'augmentation du nombre de catastrophes naturelles, à la démographie grandissante, à l'apparition des cultures dédiées à la production de biocarburants, à l'accaparement des terres par des investisseurs étrangers etc., il existe un besoin de suivi systématique et précis des systèmes agricoles et de leur adaptation à un environnement en évolution, pour pouvoir évaluer les impacts sur la sécurité alimentaire.

Les systèmes d'alerte précoce visent à fournir une information fiable et préventive sur les risques potentiels de crise alimentaire ou d'insécurité alimentaire localisée (Brown *et al.*, 2007). Dans ces systèmes, estimer les productions est capital pour pouvoir suppléer au manque de nourriture par l'aide alimentaire ou les importations dans les pays en développement. Les estimations se font à l'échelle régionale et globale, et sont fournies le plus tôt possible durant la saison de croissance des plantes, et mises à jour périodiquement jusqu'à la récolte (Herman *et al.*, 1997; Los, 1998). Pour prévoir la production, et éventuellement détecter des anomalies par rapport à d'autres années, l'objectif du système d'alerte précoce est d'estimer les superficies cultivées, puis un rendement moyen pour les différentes cultures. La multiplication de ces deux facteurs donne la production de la culture, et participe donc à l'évaluation de l'offre dans les systèmes de sécurité alimentaire. Pour ce faire, les systèmes d'alerte précoce tels que le FEWS NET (Famine Early Warning Systems NETwork) de l'USAID (United States Agency for International Development) ou le GIEWS (Global Information and Early Warning System) de l'ESA (Agence Spatiale Européenne) utilisent des méthodes différentes et complémentaires pour estimer surfaces et rendements. Outre les enquêtes agricoles, les données climatiques et les sorties de modèles statistiques ou agro-météorologiques, tous les systèmes utilisent peu ou prou un suivi par satellite dans leur méthodologie d'estimation de la production.

En effet, à partir des années 80, avec le développement d'un indice de végétation permettant de mesurer l'état des couverts végétaux quotidiennement en tout point du globe à partir du satellite météorologique NOAA (National Oceanic and Atmospheric Administration), la télédétection est devenue la pierre angulaire des programmes d'alerte précoce (Tucker *et al.*, 1985; Tucker, 1979; Tucker and Choudhury, 1987). Les systèmes d'alerte précoce utilisent des données à moyenne ou basse résolution spatiale et à haute fréquence temporelle pour le suivi de l'agriculture à l'échelle globale (e.g. Becker-Reshef *et al.*, 2010). L'analyse de séries temporelles d'images satellitaires a permis le développement de nombreuses applications en agriculture, couvrant les grands domaines de la cartographie de l'occupation et de l'utilisation du sol (cartographie du domaine cultivé et caractérisation des pratiques agricoles) et du suivi de la production végétale (détection des anomalies de croissance et estimation du rendement), qui participent à l'estimation de la production agricole.

- Concernant la cartographie de l'occupation des sols à l'échelle continentale et/ou mondiale, il existe des produits satellitaires dits « globaux ». L'espace agricole y est décrit par une, voire deux classes, et les espèces cultivées ne sont jamais spécifiées. L'inter-comparaison des produits globaux (Fritz and See, 2008; Giri *et al.*, 2005; Hansen and Reed, 2000b; Herold *et al.*, 2008; McCallum *et al.*, 2006) met en évidence de fortes divergences spatiales et

thématiques, en particulier sur la localisation des cultures dans les zones de transition (Kaptué Tchuenté *et al.*, 2011), telle que la zone sub-saharienne. La plupart des systèmes d'alerte précoce cartographient également les surfaces cultivées de certains pays à partir d'images à moyenne résolution spatiale ainsi qu'à partir de quelques images à haute résolution sur des zones tests.

- Concernant la caractérisation des pratiques agricoles, la littérature fait part de travaux utilisant des indices de végétation de séries d'images temporelles pour distinguer les espèces cultivées, et des pratiques agricoles telles que la date de semis, de récolte, les rotations etc. A l'échelle régionale, les exemples de discrimination des espèces cultivées (Wardlow *et al.*, 2007), de la caractérisation des systèmes à double culture (Arvor *et al.*, 2012) ou des cultures irriguées à partir d'imagerie MODIS (Thenkabail *et al.*, 2005) prennent place essentiellement dans des zones où l'agriculture est homogène, où les espaces cultivés sont grands et contigus, tels que dans l'état du Kansas (USA), du Mato-Grosso (Brésil) ou de la plaine Indo-Gangétique (Inde, Pakistan, Bangladesh) respectivement. A notre connaissance, la télédétection n'est jamais utilisée pour la caractérisation des pratiques agricoles dans les systèmes d'alerte précoce.
- Concernant le suivi de la croissance de la végétation, l'étude de séries temporelles d'images satellitaires permet le suivi de la phénologie des cultures et de formations végétales naturelles (Jakubauskas *et al.*, 2002; Sakamoto *et al.*, 2005; Wardlow and Egbert, 2008; Zhang *et al.*, 2003). Dans le cas des systèmes d'alerte précoce, les produits « anomalies de croissance » montrent combien la période présente est différente de la moyenne de toutes les images acquises précédemment pour la même période. Le GMFS (Global Monitoring for Food Security) produit par exemple des cartes d'indicateurs de productivité de la végétation (VPI), à partir de données SPOT VEGETATION à 1 km de résolution tous les 10 jours pour prévenir des événements extrêmes.
- Enfin, pour la prévision du rendement, de nombreuses études s'intéressent aux relations statistiques entre le rendement et le NDVI ou le climat (Lobell and Field, 2007; Maselli *et al.*, 2000; Rasmussen, 1992). Il est également possible d'estimer des variables biophysiques de surface telles que le LAI (Leaf Area Index), à partir de données de télédétection, et de les utiliser comme variable de forçage dans des modèles de rendement à l'échelle globale ou régionale (Doraiswamy *et al.*, 2004). En dehors des cartes de VPI et d'anomalies de croissance, la télédétection n'est que rarement utilisée dans les systèmes d'alerte précoce pour la quantification du rendement en Afrique de l'Ouest. Ce sont plutôt des données climatiques dans les modèles de plante et des statistiques agricoles qui sont utilisées.

En conclusion, il n'existe que peu d'études précises sur la cartographie, la caractérisation et le suivi des systèmes agricoles en Afrique de l'Ouest à l'échelle nationale, échelle pertinente pour les suivis d'alerte précoce. Ceci est en partie dû au fait que dans cette région, l'agriculture est très fragmentée, peu intensifiée, avec une variabilité spatiale très importante, et des pratiques culturales très différentes en fonction des régions (intensification, cultures associées, irrigation, arbres

disséminés sur les parcelles, etc.) (Malingreau *et al.*, 1987). D'autres difficultés surviennent du fait que la phénologie des cultures est synchronisée avec celle de la végétation naturelle en raison du régime des précipitations et qu'il est donc difficile de séparer un agrosystème d'un écosystème sur la base d'une signature temporelle (Fritz *et al.*, 2008). L'Afrique de l'Ouest est pourtant une zone très touchée par les problématiques de sécurité alimentaire, et qui nécessite la mise en place d'un suivi des cultures précis et opérationnel à l'échelle nationale.

Ainsi, l'objectif général de ce travail de thèse est d'explorer de nouvelles pistes méthodologiques en télédétection permettant une meilleure estimation de la production agricole en Afrique de l'Ouest, afin de mieux répondre aux besoins des systèmes d'alerte pour la sécurité alimentaire. Pour cela, cette thèse s'attache à répondre à trois sous-objectifs (Figure 1):

- cartographier le domaine cultivé à l'échelle nationale et qualifier le résultat vis-à-vis des produits globaux existants,
- caractériser les principaux systèmes agricoles du pays,
- suivre la phénologie des cultures à partir de produits MODIS et de simulations agro-météorologiques.

Nos recherches sont développées sur le Mali, pays représentatif de la zone soudano-sahélienne.

La thèse comprend 5 chapitres. Le premier chapitre décrit le contexte général (sécurité alimentaire et systèmes d'alerte précoce en Afrique) et dresse un état de l'art des thématiques concernées, à savoir l'utilisation de la télédétection pour le suivi global de l'agriculture. Le deuxième chapitre décrit les données et outils utilisés dans cette thèse. Les trois chapitres suivants s'attachent à rendre compte des méthodes développées et des résultats obtenus dans ce travail de thèse au travers des cinq articles publiés et/ou soumis (Figure 1). Ainsi, pour la cartographie du domaine cultivé, nous ferons référence à l'article 1 (ainsi qu'à deux articles annexes, 4 et 5). Pour la caractérisation des systèmes agricoles, à l'article 2, et enfin, pour le suivi de la croissance des cultures, à l'article 3.

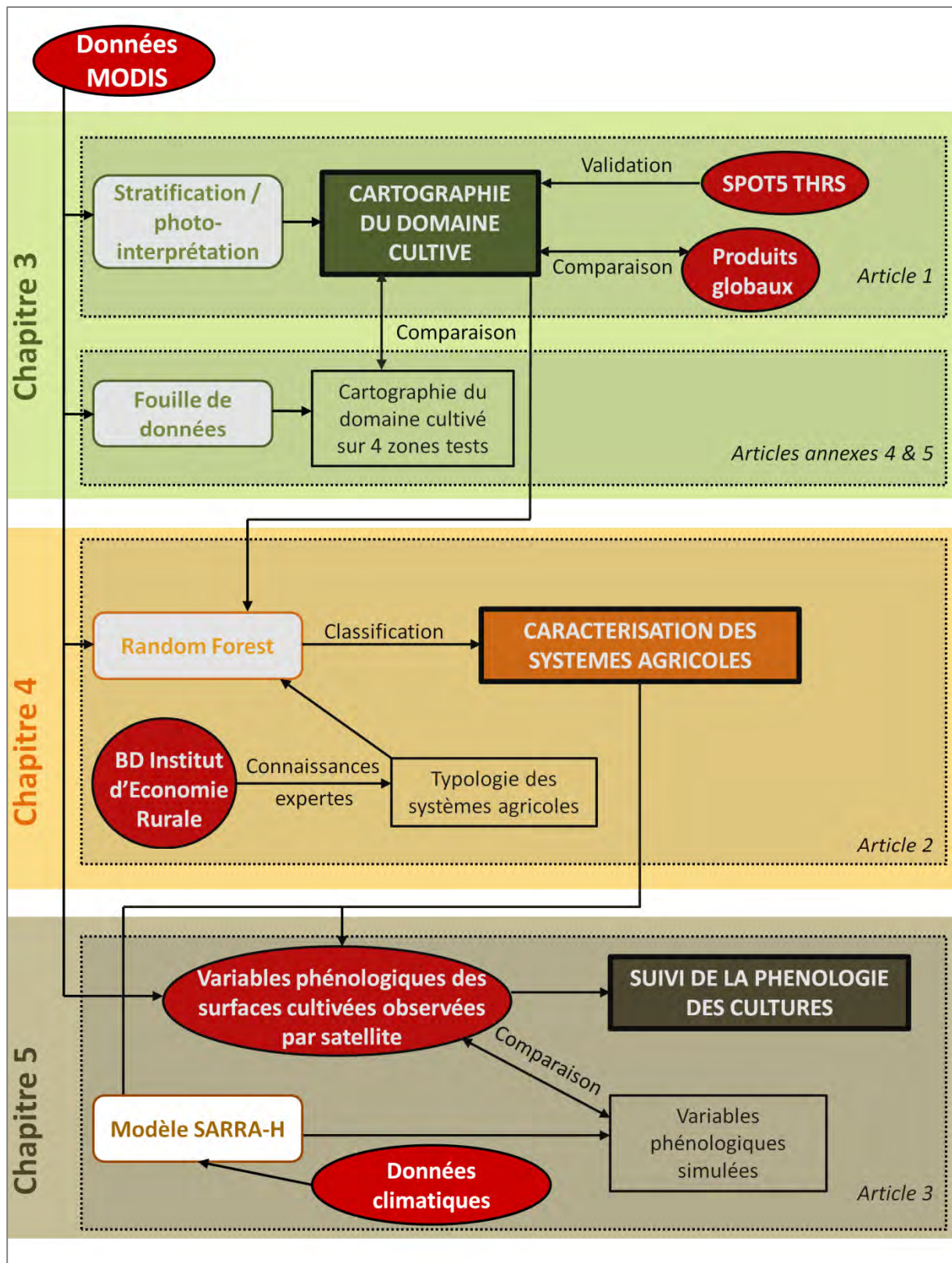


Figure 1 : Organisation générale de la thèse.

Ci-dessous, les références complètes des articles pré-cités, présentés par objectif.

- La cartographie du domaine cultivé:

Article 1 : Vintrou, E., Desbrosse, A., Bégué, A., Traoré, P. C. S., Baron, C., Lo Seen, D. (2012). Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation* 14(1): 83-93.

Ainsi que deux articles en annexes :

Article 4 : Pitarch Y., Vintrou E., Badra F., Bégué A., Teisseire M. (2011). Mining Sequential Patterns from MODIS Time Series for Cultivated Area Mapping, in: S. Geertman, et al. (Eds.), *Advancing Geoinformation Science for a Changing World*, Springer Heidelberg Dordrecht London New-York. pp. 45-62.

Article 5 : Vintrou, E., Pitarch, Y., Bégué, A., Teisseire, M. (2012). Fouille de données et cartographie des surfaces agricoles. Des relevés terrain aux données satellitaires : application au Mali. *Revue Internationale de Géomatique*, 21(4): 469-488.

- La caractérisation des systèmes agricoles :

Article 2 : Vintrou, E., Soumaré, M., Bernard, S., Bégué, A., Baron, C., Lo Seen, D. (2012). Mapping Fragmented Agricultural Systems in the Sudano-Sahelian region using Random Forest and Ensemble Metrics of Coarse Resolution MODIS Imagery. (Special Issue) *Photogrammetric Engineering and Remote Sensing* (Accepted in January 2012).

- Le suivi de la phénologie des surfaces cultivées :

Article 3 : Vintrou, E., Bégué, A., Baron, C., Lo Seen, D. (2012). Analysing MODIS phenometrics information content in heterogeneous agricultural landscapes of West Africa (In preparation).

Enfin, nous concluons la thèse par un bilan des travaux réalisés et par des perspectives identifiées en recherche et sur le plan opérationnel.

Chapitre 1. CONTEXTE ET ÉTAT DE L'ART : SÉCURITÉ ALIMENTAIRE ET TÉLÉDÉTECTION POUR LE SUIVI DES CULTURES

1. Contexte : Sécurité alimentaire

1.1. Définitions et situation en 2010 dans le monde

La perspective d'une planète comptant 9 milliards d'habitants d'ici à 2050 rend la question de l'accès à l'alimentation toujours plus brûlante. Selon l'Organisation des Nations Unies pour l'alimentation et l'agriculture, la FAO, on peut parler de sécurité alimentaire lorsque des aliments sont disponibles à tout moment, que tous y ont accès, que ces aliments sont appropriés du point de vue nutritionnel, tant en quantité qu'en qualité et en variété, et qu'ils sont acceptables sur le plan culturel. Elle comporte quatre dimensions : i) la disponibilité (importations, stockage ou aide alimentaire), ii) l'accès (pouvoir d'achat et infrastructures), iii) la salubrité et la qualité (hygiène, accès à l'eau) et iv) la stabilité des trois premières dimensions. Il faut donc bien noter que la sécurité alimentaire n'est pas nécessairement acquise lorsque l'offre alimentaire est suffisante : il faut également savoir qui produit la nourriture, qui a accès aux informations nécessaires sur la production agricole, qui a un pouvoir d'achat suffisant pour acquérir la nourriture etc. D'après cette définition, il y a aujourd'hui environ deux milliards de personnes — une personne sur trois — en situation d'insécurité alimentaire. Elles ne peuvent cultiver elles-mêmes suffisamment de produits alimentaires ou n'ont pas les moyens d'en acheter en quantités suffisantes sur les marchés locaux.

On parle de sous-alimentation lorsque l'apport calorique est inférieur aux besoins énergétiques alimentaires minimaux. Les besoins énergétiques minimaux correspondent à l'énergie nécessaire pour avoir une activité physique légère et maintenir un poids minimal acceptable, selon la taille. Ils varient d'un pays et d'une année à l'autre, et sont fonction du sexe et de l'âge de la population. La plupart des personnes sous-alimentées vivent dans les pays en développement. Deux tiers d'entre elles sont concentrées dans sept pays seulement (Bangladesh, Chine, République démocratique du Congo, Éthiopie, Inde, Indonésie et Pakistan) et plus de 40 pour cent d'entre elles vivent en Chine et en Inde (Figure 2). En pourcentage de la population nationale, les pays les plus touchés en 2007 sont situés sur le continent africain (Figure 3). Selon les estimations de la FAO, le nombre de personnes sous-alimentées a diminué dans le monde entre 2009 (1,023 milliard) et 2010 (925 millions) (Figure 4). C'est l'Asie qui a contribué le plus fortement à cette diminution: la faim touche aujourd'hui 80 millions de personnes en moins sur ce continent. L'Afrique sub-saharienne a également progressé puisque le nombre de personnes souffrant de la faim y a baissé de 12 millions dans cet intervalle de temps. Cependant, le nombre de personnes en proie à la faim est plus élevé en 2010 qu'avant les crises alimentaire et économique de 2008-2009 (Figure 4).

Face à ces chiffres alarmants, l'Objectif numéro 1 du Millénaire pour le Développement, confirmé au Sommet mondial de la sécurité alimentaire en Novembre 2009, est de réduire de moitié le nombre de personnes sous-alimentées à l'horizon 2015.

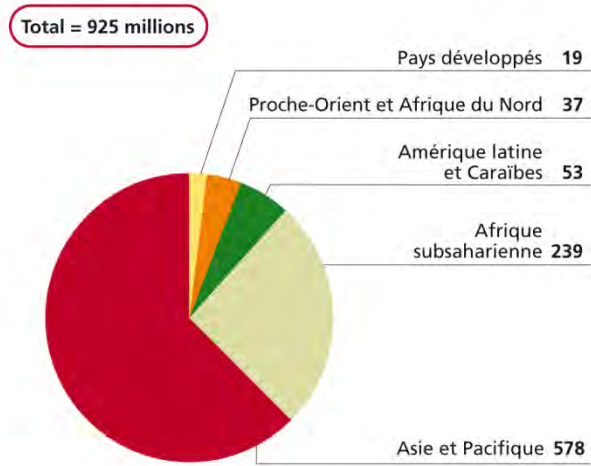


Figure 2 : Répartition par régions du monde du nombre de personnes (en millions) sous-alimentées (Source FAO, 2010)

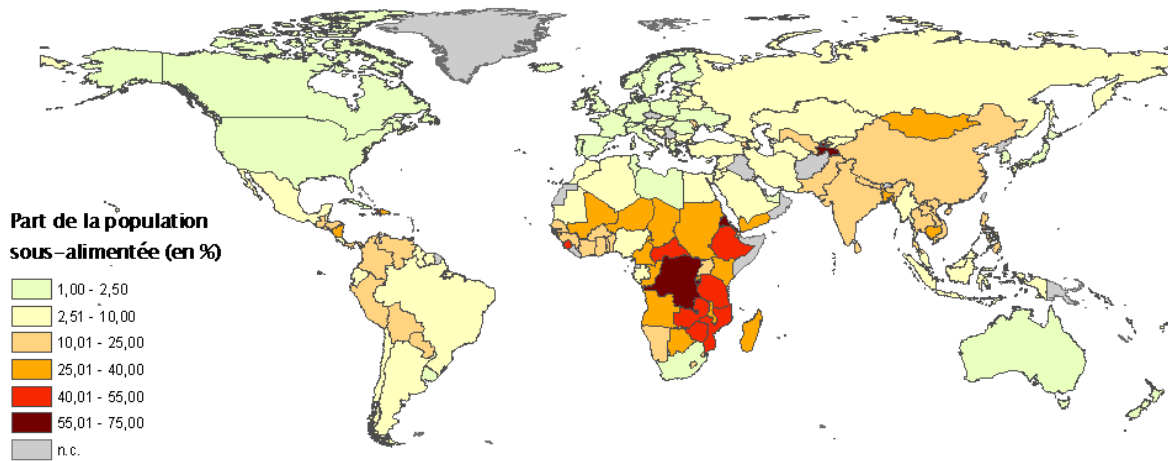


Figure 3 : Concentration par pays des personnes souffrant de sous-alimentation en 2007 (selon les données 2007-2008 de l'indice de développement humain) (Source PNUD 2007/2008)

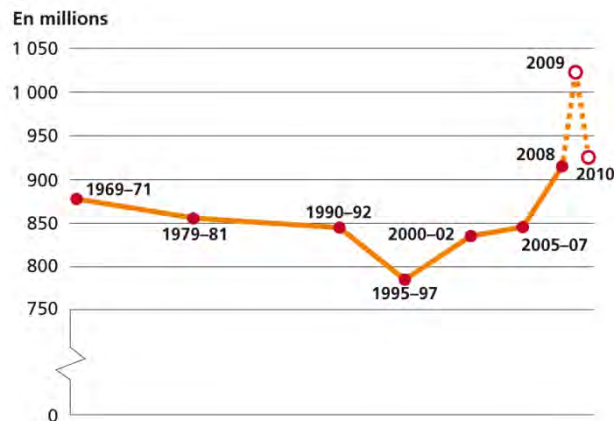


Figure 4 : Nombre de personnes sous-alimentées dans le monde (Source : FAO, 2010)

1.2. La sécurité alimentaire en Afrique subsaharienne

Aujourd'hui, près de 33 pour cent de la population d'Afrique subsaharienne, soit environ 200 millions de personnes, sont sous-alimentées. Les famines frappent généralement les pays les plus vulnérables, dont certaines populations fragiles sont les moins à même de résister à des chocs économiques ou climatiques ou à des conflits. La prévention joue désormais un rôle essentiel dans le ciblage des populations les plus nécessiteuses et prépare les réponses d'urgence dont l'aide alimentaire. Les organismes travaillant avec les communautés les plus vulnérables tentent donc de répondre aux besoins immédiats des personnes qui se trouvent dans des situations de crise alimentaire, mais aussi de réduire la vulnérabilité de ces communautés en leur fournissant une aide à plus long terme.

Pour prévenir de futures crises, la première considération semble naturellement être la question des disponibilités alimentaires. En premier lieu parce qu'elle garantit un accès local, donc rapide, à des ressources agricoles. En second lieu parce que les populations les plus pauvres sont majoritairement en zone rurale et que toute amélioration de la production agricole devrait leur profiter. Ces disponibilités ont augmenté au niveau mondial au cours des vingt dernières années, mais dans les pays en développement comme ceux d'Afrique (et plus particulièrement ceux qui consomment du manioc ou de l'igname), les disponibilités ont régressé durant la même période. Le rendement moyen des céréales stagne et, pour atteindre un niveau d'approvisionnement alimentaire correspondant au nombre de bouches à nourrir, des efforts importants doivent être réalisés pour la transformation des systèmes de production vers des systèmes plus productifs et plus durables.

Les obstacles à l'augmentation adéquate de la production agricole en Afrique sub-saharienne sont nombreux.

1.2.1. Pression démographique

Dans beaucoup de pays en développement, la croissance démographique évolue si rapidement que la production agricole peine à suivre le rythme de l'accroissement de la demande alimentaire. De par le dynamisme de la population (proportion élevée de la population qui est en âge de procréer), l'augmentation de l'espérance de vie, et des taux de fécondité qui ne diminuent que très lentement, on ne peut compter dans l'immédiat sur une diminution significative du taux de croissance de la population africaine. En Afrique subsaharienne, la population est considérée comme une ressource importante pour le développement et la création de richesses. Sur cette base, on s'attend à ce que la population en Afrique subsaharienne soit au moins doublée dans les 25 prochaines années, ce qui constitue un défi considérable pour le développement de l'agriculture.

1.2.2. Pression foncière

Le niveau d'urbanisation est encore très modéré en Afrique : il atteint aujourd'hui les 37%, contre 50% au niveau mondial (Devèze, 2008). C'est pourquoi malgré l'exode rural, la population rurale continue de s'accroître chaque année, contrairement à la plupart des autres régions du monde où l'urbanisation « vide les campagnes ». Aujourd'hui, avec l'augmentation de la pression foncière et

la saturation de l'espace, les nouveaux arrivants se voient prêter des terres avec une durée limitée, avec souvent des contreparties en nature. Les modifications de gestion du foncier proviennent aussi du développement des cultures commerciales, qui permettent une valeur ajoutée au revenu familial. Il en résulte un développement progressif de la vente de terre, puisque la pratique des cultures commerciales procure des revenus monétaires permettant l'achat de terre.

1.2.3. Déclin de la fertilité du sol

Dans la majeure partie de l'Afrique subsaharienne, les sols ont une fertilité extrêmement faible et les éléments nutritifs exportés ne sont pas remplacés de manière adéquate. L'Afrique subsaharienne a la plus basse consommation d'engrais minéraux, environ 10 kg d'éléments nutritifs (N, P₂O₅, K₂O) par hectare et par an, par rapport à une moyenne de 90 kg au niveau mondial (FAO, 2003). De plus, la baisse de la fertilité des sols dans cette région est également due à la généralisation de la culture continue et à l'abandon de la jachère. Un autre problème récurrent vient du fait que lorsque les agriculteurs itinérants n'ont aucun droit sur la terre où ils s'installent, ils ont peu d'incitation à améliorer la productivité future de ces terres. En outre, le brûlis annuel de la végétation réduit de manière drastique le retour de la matière organique au sol, en perdant ainsi les bénéfices (fertilité, meilleure structure, conservation de l'eau, biodiversité, etc.) et la terre s'appauvrit.

1.2.4. Autres obstacles

Dans leur rapport sur la sécurité alimentaire et le développement agricole en Afrique subsaharienne, Kidane *et al.* (2006) évoquent de nombreux autres obstacles majeurs, à commencer par les troubles politiques et armés dans la région, qui sont pour eux les principaux freins au développement agricole. Ils mentionnent également la faiblesse des capacités institutionnelles qui contribue à l'inefficacité de politiques capables de relever les défis du développement agricole et rural, et qui explique de ce fait, l'absence d'infrastructures permettant la commercialisation efficace des produits agricoles. Le rapport évoque également la fuite des cerveaux, la mise en œuvre hâtive de réformes mal conçues et la tendance à privilégier les villes. Le dernier point important réside dans la nécessité de renforcement de l'appui public à l'agriculture, car il n'existe pas de mécanismes pouvant aider à minimiser ou partager les risques supportés par les producteurs.

Les facteurs de l'insécurité alimentaire – et donc les obstacles à l'augmentation adéquate de la production agricole - en Afrique sub-saharienne sont la pression démographique, la pression foncière, la faible fertilité des sols, ainsi que les troubles politiques ou les dysfonctionnements des institutions locales. Pourtant, la production agricole doit augmenter, que ce soit par un accroissement des superficies là où c'est encore possible, par des gains de productivité du travail, ou par l'augmentation des rendements, comme le suggère la FAO.

1.3. L'exemple du Mali

1.3.1. Présentation générale

Le Mali est un pays enclavé d'une superficie de 1 241 138 km²; ses accès à la mer sont situés à plus de 1000 km (Dakar, Abidjan, Lomé notamment) (Figure 5). Les zones saharienne et sahélienne ayant une pluviométrie moyenne inférieure à 400 mm occupent près de 75% de la surface totale. Ces régions accueillent moins de 10% de la population. En dehors du massif des Iforas dans le nord-est et les Monts mandingues au Sud-Ouest, le pays est une juxtaposition de plaines et de plateaux. Les deux plus grands fleuves de l'Afrique de l'Ouest, le Niger et le Sénégal, le traversent et lui confèrent ainsi d'importantes potentialités en terres irrigables et des possibilités de navigation contribuant au désenclavement de certaines zones. La population est estimée à 13 millions d'habitants, avec un taux de croissance annuelle de l'ordre de 2,5%. Elle est largement concentrée dans les régions du centre (Bamako, Ségou, Mopti) et du sud (Sikasso); elle est urbanisée pour environ 30 % du total.

La saison des pluies se situe globalement entre le mois de mai et le mois de novembre. Les précipitations totales et la durée de la saison des pluies augmentent du Nord au Sud, déterminant plusieurs zones climatiques. Les zones climatiques définies traditionnellement, au nombre de sept, sont les suivantes : zone soudano-guinéenne (pluviométrie supérieure à 1200 mm) ; zone sud soudanienne (1000–1200 mm) ; zone nord soudanienne (800–1000 mm) ; zone soudano-sahélienne (400–800 mm) ; zone sahélienne (300–400 mm) ; zone sahélo-saharienne (200-300 mm) ; zone Saharienne (< 200 mm).

L'agriculture et l'élevage constituent les principales activités en milieu rural. La production agricole très diversifiée comporte des cultures vivrières, de rente, fruitières et maraîchères, les produits d'élevage, de forêt et de pêche. La production vivrière essentiellement pluviale est dominée par les céréales : mil, sorgho, maïs, riz, fonio. Les légumineuses alimentaires comprennent le niébé, l'arachide et le voandzou. Les cultures de rente sont dominées par le coton suivi de la canne à sucre. Viennent ensuite le tabac et le thé.



Figure 5 : Carte du Mali (www1.alliancefr.com).

1.3.2. Les systèmes agricoles

A l'échelle de l'exploitation, le système de culture (ou *cropping system* en anglais) peut être identifié à l'aide d'un état de parcelle recensant les successions culturales sur les différentes parcelles (Dufumier, 1996). Lorsque l'on s'intéresse à la sécurité alimentaire à l'échelle d'un pays, l'échelle considérée est souvent celle d'une région administrative. C'est probablement une des raisons qui ont conduit certains agronomes et géographes à proposer de nouveaux concepts pour identifier au sein de l'espace cultivé des sous-ensembles présentant des caractéristiques agro-écologiques homogènes. Ce sont par exemple « les unités paysagères » proposées par Deffontaines and Thion (2001), ou encore les notions de « terroirs » que l'on retrouve chez Sautter and Pelissier (1964), qui représentent l'espace exploité par une communauté villageoise. On s'intéressera donc à des entités agricoles, présentant des caractéristiques agro-écologiques homogènes, que nous nommerons « systèmes agricoles » (pour éviter toute confusion avec les définitions précitées) et qui pourront se définir par une ou deux cultures dominantes, reliées à des pratiques culturales au sens large (intensification, irrigation, association...).

Au Mali, l'espace agricole (là où la culture pluviale est possible) peut être découpé suivant la quantité de pluies qui tombe par an, et suivant la durée des pluies (Figure 6).

- Les systèmes agricoles nord-sahéliens (<300 mm) sont à pastoralisme dominant, alors que les systèmes sahéliens (300-400 mm) sont des systèmes agro-pastoraux à base de mil-sorgho essentiellement. Ce sont des céréales résistantes à la sécheresse, à cycle court, qui s'adaptent très bien à la zone. Les rendements sont faibles en raison de la faible utilisation d'intrants et du bas taux d'équipement des exploitations. L'élevage (surtout des petits ruminants) est important.

- La zone soudano-sahélienne (400-800 mm) peut se découper en deux parties. Au nord (400-600 mm), on pratique surtout des cultures ayant un cycle végétatif bref de 90 jours, c'est-à-dire principalement du sorgho et du mil. Dans le vieux bassin cotonnier (autour de Koutiala – 600-800 mm), le système agricole est du type agro-sylvo-pastoral à dominance cultures pluviales (coton et céréales). Il est centré autour de la production cotonnière, principale activité de rente. Cette culture est pratiquée dans une rotation de type triennal coton-sorgho-mil ou arachide ou biennal coton-céréales. Le maïs est cultivé seul ou en association avec le mil.

- La zone nord-soudanienne (autour de Bougouni – 800-1000 mm) présente des systèmes de production organisés autour du coton et du maïs, cultivés en alternance avec de l'arachide et du niébé. Dans la zone d'influence du barrage de Sélingué, les systèmes agricoles tournent autour du maïs, de l'arachide et du riz. Le fonio, l'igname, la patate douce et le manioc sont les principales cultures de diversification. Les cultures maraîchères se développent dans les zones à disponibilité en eau pendant la saison sèche.

- La zone sud-soudanienne (1000-1200 mm) présente des conditions climatiques adaptées à la culture de riz pluvial dans les bas-fonds, mais aussi du sorgho, du coton et de l'arboriculture fruitière.

Cette typologie des systèmes agricoles est complétée par deux autres systèmes caractérisés

par la disponibilité de l'eau à tout moment de l'année : les systèmes irrigués (aménagement des grandes vallées alluviales) et les systèmes périurbains (eaux de surfaces et souterraines peu profondes, proximité des villes pour leur approvisionnement). La majeure partie des terres irriguées se trouve dans la région du Delta Intérieur du fleuve Niger, en zone soudano-sahélienne. Les terres sont essentiellement utilisées pour l'élevage et la riziculture.

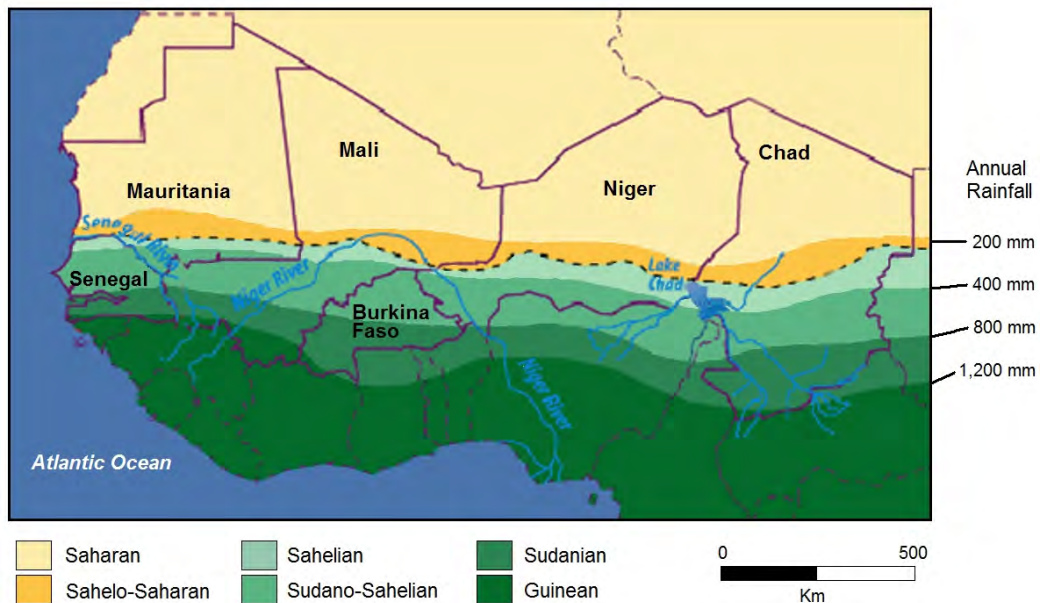


Figure 6 : Précipitations annuelles et régions agro-climatiques en Afrique de l'Ouest (Source : FEWS 2007).

Les activités de l'agriculture et de l'élevage s'organisent suivant un calendrier bien précis (Figure 7). L'essentiel des activités agricoles ont lieu entre mai et novembre, avec un étalement plus ou moins grand suivant si l'on se trouve au Sud ou au Nord du Mali. A mesure que l'on se déplace vers le Nord, la saison des pluies démarre plus tard et se termine plus tôt.

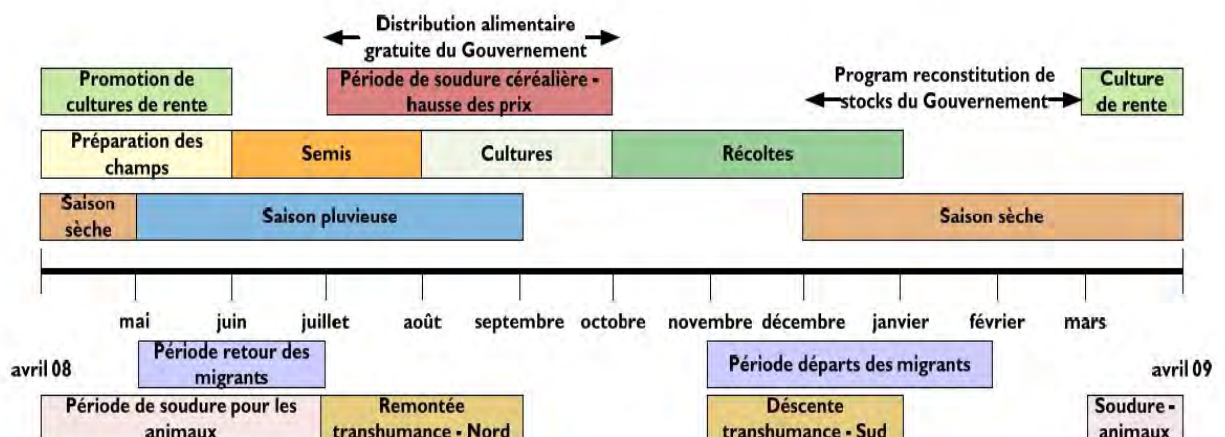


Figure 7 : Calendrier saisonnier et événements significatifs pour l'agriculture au Mali. (Source : FEWS NET Mali).

1.3.3. La sécurité alimentaire au Mali

Le Mali est à l'image des autres pays d'Afrique sub-saharienne. La prévalence de la sous-alimentation dans la population totale est de 12% d'après la carte de la Faim FAO (2005-2007). Le nombre de personnes sous-alimentées est de 1.5 millions sur une population totale de 12.1 millions. Malgré son potentiel considérable en ressources agricoles, le Mali fait donc face à des problèmes sérieux d'insécurité alimentaire. Les céréales constituent la base de l'alimentation; mais en dehors du riz irrigué, la production des céréales traditionnelles (mil, sorgho et maïs) reste tributaire des aléas climatiques et subit de fortes variations.

Le problème majeur de la sécurité alimentaire au Mali est celui de la pauvreté, une partie importante de la population n'ayant pas un accès garanti à une alimentation de base. Malgré une diversification assez avancée, la faiblesse des revenus interdit cette accessibilité au marché pour compléter la production familiale lorsqu'elle existe.

Pourtant, le Mali est théoriquement doté de la capacité d'être un « grenier alimentaire » pour tout l'Ouest africain, avec une superficie irrigable de près de 2 millions d'hectares (dont 130 000 seulement sont actuellement aménagés) dans la seule zone de l'Office du Niger. Mais il se retrouve chaque année dans une situation critique en termes de production céréalière. Les conséquences de cette fragilité sont des hausses de prix sur les marchés, des importations de céréales et des recours à l'aide alimentaire internationale.

La difficulté au Mali réside dans le fait que même lorsque la campagne agricole est bonne (comme en 2008), la sécurité alimentaire des Maliens n'est pas assurée, car les prix de vente des céréales les plus consommées ne diminuent plus, même lorsque l'offre est abondante. Les raisons de ce paradoxe sont multiples (liées au marché international, aux changements de comportements alimentaires etc.), mais l'une des causes principales avancée par Oumarou Mohamed Lamine¹ est la forte syndicalisation des producteurs de céréales. Ces derniers s'organisent pour ne pas vendre les productions au dessous d'un certain seuil, en prenant en compte la hausse du prix des intrants et du coût de la vie en général. Ils mettent également en place des stocks céréaliers, pour ne pas vendre immédiatement après la récolte, et attendre des prix plus intéressants, notamment lors de la période dite de « soudure ». Tout ceci engendre un sérieux problème pour les acheteurs, qui du coup ne bénéficient jamais de prix faibles.

Pour sortir de ce cercle vicieux, le gouvernement malien propose d'aménager plus de terres irrigables, qui offrent le plus grand potentiel. En plus de l'irrigation, le Programme Régional de Sécurité Alimentaire « Appui à la mise en œuvre de la Politique agricole de l'Union en matière de sécurité alimentaire » (UEMOA-Mali) préconise de :

- faciliter les transports (régulation spatiale) et améliorer le stockage (réduction des pertes, régulation temporelle) ;
- améliorer les dispositifs de prévention et de gestion des crises conjoncturelles ;
- mettre en place des politiques spécifiques visant les groupes vulnérables ;
- développer une politique régionale en matière de sécurité alimentaire.

¹ Oumarou Mohamed Lamine est un agronome malien. Ces informations sont tirées du site : <http://maliagroinfos.unblog.fr/2011/05/25/la-securite-alimentaire-au-mali/>

Comme dans les autres pays d'Afrique subsaharienne, les approvisionnements alimentaires proviennent de trois sources: la production alimentaire nationale, les importations, et l'aide alimentaire. Ces 30 dernières années, la croissance de la production céréalière a en effet été irrégulière : la croissance relativement rapide des années 1980-1990 s'est estompée après 1997, ce recul s'accompagnant d'un accroissement des importations. D'après le rapport de la FAO sur l'accroissement des soutiens publics (Kidane *et al.*, 2006), la part des importations dans la consommation totale de céréales, évaluée à environ 5% juste après l'indépendance, se situe ces dernières années aux alentours de 25%. L'aide alimentaire a augmenté rapidement dans les années 90, atteignant 10% de la consommation totale de céréales, alors qu'elle est aujourd'hui presque négligeable. Pour le Mali en particulier, ce même rapport le situe dans le groupe de pays dont la sécurité alimentaire « stagne » (avec le Burkina, le Nigéria, le Soudan et le Zimbabwe).

Les systèmes agricoles maliens se différencient suivant un gradient climatique, tant au niveau des espèces cultivées et du calendrier agricole que des logiques de production (cultures vivrières, de rente, irrigation etc). Ils sont fragiles, peu intensifiés, sensibles aux variations climatiques, et n'assurent pas la sécurité alimentaire du Mali, puisque le pays importe 25% des céréales consommées, alors même que le potentiel de production agricole de ce pays (notamment au niveau des surfaces irriguées) est très grand.

La situation alimentaire au Mali reste donc très précaire et surtout variable d'année en année. Les réseaux d'observation au sol ne sont généralement pas de densité et de qualité suffisantes pour être utilisés seuls. La télédétection apparaît ainsi être un outil indispensable de suivi agricole à l'échelle régionale, en complément à d'autres sources d'informations.

2. Télédétection et suivi des cultures

La télédétection par satellite permet des observations régulières, répétitives et objectives de la quasi-totalité de la planète, à divers pas d'espace et de temps, dans plusieurs domaines de longueur d'onde.

L'imagerie satellite à moyenne et basse résolutions spatiales (> 100 m de résolution) est utilisée de manière régulière depuis bientôt une trentaine d'années pour étudier la biosphère terrestre sur de vastes étendues (Figure 8). A l'échelle de la planète, les satellites « grand champ » permettent de suivre les variations saisonnières et annuelles des couverts végétaux. L'arrivée des nouvelles images de télédétection à haute résolution (5 - 30 m) et très haute résolution spatiale (moins de 5 m) permettent la perception d'une nouvelle gamme d'objets au sol non observables jusqu'à maintenant. Dans le même temps, les nécessités de la surveillance des territoires et en particulier des cultures imposent un suivi temporel fin, ne pouvant être obtenu que par des systèmes d'observation à grand

champ ou géostationnaires (Tableau 1).

ECHELLE	Résolution spatiale	Résolution temporelle
Globale	5 km - 1 km	Toutes les heures
	1 km - 250 m	Quotidien
Régionale / Nationale	250 m - 60 m	1 couverture journalière à décadaire
Sous-nationale	60 m - 10 m	1 couverture décadaire à mensuelle
Locale	< 10 m	1 couverture annuelle

Tableau 1 : Résolutions spatiale et temporelle de données satellitaires pour les applications en agriculture.

2.1. Cartographie du domaine cultivé

A l'échelle locale, les images à haute ou très haute résolution spatiale comme SPOT5 ou LANDSAT (Figure 8) sont couramment utilisées pour la cartographie de l'occupation du sol et en particulier des surfaces cultivées (Murakami *et al.*, 2001; Thenkabail, 2003).

A l'échelle régionale, l'imagerie satellitaire à moyenne ou basse résolution spatiale ne permet pas toujours de distinguer facilement le domaine cultivé des autres types d'occupation du sol. De nombreux efforts ont été faits ces vingt dernières années et ont fait émerger deux méthodes clés : i) l'utilisation de statistiques et de données à haute résolution en appui, et ii) le « démixage » (ou unmixing) des pixels. Par exemple, Husak *et al.* (2008) proposent une cartographie des surfaces cultivées en Ethiopie, basée sur des images Landsat et des statistiques du gouvernement éthiopien, et s'appuient sur des images IKONOS comme vérité terrain. Lobell and Asner (2004) utilisent des images MODIS et Landsat pour estimer l'aire cultivée sur deux sites (Mexique et Etats-Unis) : la première étape consiste à séparer les signaux temporels de différentes occupations du sol à partir d'imagerie Landsat, méthode plus connue sous le nom de « linear unmixing » (Adams *et al.*, 1995; Elmore *et al.*, 2000) et la seconde à quantifier le pourcentage des différentes occupations du sol au sein d'un pixel MODIS. Cependant, dans des pays comme ceux d'Afrique de l'Ouest, où l'agriculture est généralement fragmentée, ces approches restent complexes à cause de la grande diversité d'occupations du sol contenue au sein d'un même pixel (Lobell *et al.*, 2004). Des difficultés de détection résultent également de l'extrême hétérogénéité des parcellaires, des calendriers agricoles et des pratiques culturelles en Afrique de l'Ouest (Malingreau *et al.*, 1987).

Enfin, il existe également des produits dits « globaux » qui cartographient l'occupation du sol à l'échelle continentale et/ou mondiale. Parmi les produits globaux existants, nous présentons en Tableau 2 le Global Land Cover 2000 (GLC2000) for Africa, le produit GLOBCOVER, le produit global MODIS V05 (MCD12Q1) et enfin, le produit ECOCLIMAP-II, qui sont représentatifs des produits globaux actuellement disponibles.

- Le GLC2000 (Global Land Cover 2000) est un produit global développé au Joint Research Centre (JRC) de l'Union Européenne. Il est fait à partir de données SPOT VEGETATION journalières à la résolution de 1 km, et donne l'occupation du sol pour l'année 2000. Dans la carte GLC2000, la légende comporte 27 classes d'occupation du sol, dont trois concernent l'agriculture : les zones agricoles sont définies comme les terres qui contiennent plus de 50% de cultures ou/et de pâturages.
- Le Globcover est une initiative de l'Agence Spatiale Européenne (ESA), qui a voulu fournir un produit à meilleure résolution spatiale (300 m). Cette carte à 22 classes se base donc sur des images du capteur MERIS, à bord du satellite ENVISAT, des années 2005 et 2006. Elle contient 4 classes de cultures : une classe de cultures irriguées, une classe de cultures pluviales, et deux classes mixtes (culture et végétation naturelle, avec des pourcentages différents : entre 50 et 70% de cultures dans la première, et entre 20 et 50% de cultures dans la seconde).
- Le produit global MODIS (MODIS V05) est distribué par l'Agence Spatiale Américaine (NASA LP DAAC) à une résolution de 500 m. C'est un produit renouvelé chaque année, basé sur des images MODIS-Terra et MODIS-Aqua. Il contient 17 classes, dont une classe de culture et une classe mixte « culture – végétation naturelle » (le pourcentage de cultures dans cette classe est indiqué comme inférieur à 60%).
- Enfin, le produit ECOCLIMAP-II est une carte d'occupation du sol développé par le Centre National de Recherches Météorologiques (CNRM), à une résolution spatiale de 1 km, en fusionnant un produit déjà existant (ECOLIMAP-I) et le GLC2000. Ce produit propose pour la partie Afrique de l'Ouest 37 classes d'occupation du sol, dont 3 concernent les cultures (deux classes pures et une classe mixte dont le pourcentage de culture n'est pas indiqué). C'est un produit qui doit être renouvelé tous les 5 ans.

Ces produits globaux ont été testés, comparés et validés dans de nombreuses études. Pour ce faire, les légendes de ces différents produits doivent être harmonisées, car certains utilisent la nomenclature Land Cover Classification System (LCCS), et d'autres celle de l'International Géosphère-Biosphère Programme (IGBS). Les méthodes d'harmonisation sont nombreuses, et nécessitent souvent d'agréger certaines classes. Après ces « corrections », Giri *et al.* (2005) ont proposé une comparaison des produits MODIS V05 et GLC2000 et ont montré une relative cohérence en terme de surfaces occupées par chaque classe, excepté pour les classes de savane et de zones humides. Entre les deux produits, les surfaces occupées par les cultures variaient de 20% à l'échelle mondiale et seuls 52,1% des surfaces cultivées étaient au même endroit. En comparant ces deux mêmes produits à l'échelle du Soudan, Fritz and See (2008) ont montré que le produit MODIS V05 avait tendance à sous-estimer les surfaces cultivées, alors que le produit GLC2000 les surestimait. Kaptué Tchuenté *et al.* (2011) ont comparé les quatre produits présentés en Tableau 2 à l'échelle de l'Afrique (avec la version V04 du produit MODIS cependant) et concluent sur une bonne corrélation des classes d'occupation du sol sur le continent africain. Cependant, lorsque seule la classe cultivée

est considérée, le nombre de pixels peut aller du simple au double entre les différents produits. En ce qui concerne la cartographie des surfaces cultivées, les plus grandes différences se retrouvent dans les zones où l'agriculture est fragmentée, comme cela est le cas en Afrique de l'Ouest. En travaillant à des résolutions de 500 m ou 1 km, les pixels sont généralement mixtes, regroupant différents types d'occupation du sol. Il est donc difficile de savoir sur quels produits s'appuyer.

De plus, les produits globaux sont focalisés sur de larges écosystèmes plutôt que sur les systèmes agricoles et possèdent tout au plus 2 à 3 classes représentant la culture, souvent mélangée à la végétation naturelle, et donc ne permettent pas la caractérisation des systèmes agricoles, condition indispensable au suivi des cultures et estimations de rendement pour un système d'alerte précoce.

Produit	GLC2000 for Africa	GLOBCOVER	MODIS V05 (MCD12Q1)	ECOCLIMAP-II
Auteur	Mayaux <i>et al.</i> , 2004	Arino <i>et al.</i> , 2008	Friedl <i>et al.</i> , 2010	Kaptué Tchuenté <i>et al.</i> , 2010
Plateforme - Capteur	SPOT-VGT-1	ENVISAT-MERIS	MODIS-Terra MODIS-Aqua	MODIS-Terra MODIS-Aqua
Données d'entrée	4 bandes spectrales et NDVI	13 bandes spectrales	3 à 7 bandes	Leaf Area Index
Date d'acquisition	01/11/1999 – 31/12/2000	01/12/2004 – 31/06/2006	2007	01/01/2000 - 31/12/2007
Résolution spatiale	1 km	300 m	500 m	1 km
Nombre de classes	27	22	17	37
Légende	Land Cover Classification System	Land Cover Classification System	International Geosphere-Biosphere Programme	Land Cover Classification System
Méthode de classification	Classification avec expertise locale/régionale	Classification avec expertise locale/régionale	Classification supervisée avec arbre de décision	Classification supervisée fusionnant ECOCLIMAP-I et GLC2000
Validation	Intercomparaison Echantillonnage statistique Données à haute résolution	Echantillonnage statistique	“sites d'entraînement” et “valeurs de confiance agrégées par classe et par région”	Matrice de confusion Intercomparaison Données à haute résolution
Lien internet	(1)	(2)	(3)	(4)

(1) http://bioval.jrc.ec.europa.eu/products/glc2000/data_access.php

(2) <http://postel.mediasfrance.org/fr>

(3) <http://wist.echo.nasa.gov/api>

(4) http://www.cnrm.meteo.fr/gmme/PROJETS/ECOCLIMAP/page_ecoclimap.htm

Tableau 2: Caractéristiques des produits globaux d'occupation du sol.

Les produits globaux sont sans cesse améliorés, et la communauté scientifique s'attache aujourd'hui à les valider. Fritz *et al.* (2009) proposent un outil intéressant, appelé geo-wiki, dans lequel des volontaires sont mobilisés pour déterminer, d'après leurs connaissances ou à l'aide de Google Earth, les zones où les produits globaux sont corrects et celles où ils ne le sont pas. Chaque utilisateur peut également intégrer dans l'outil ses propres photographies ou points GPS de terrain. A terme, une nouvelle carte hybride focalisée sur les zones cultivées va être produite (Fritz *et al.*, 2011).

2.2. Cartographie des systèmes agricoles

Différentes méthodes ont été mises au point pour discriminer les différentes espèces cultivées ou pratiques (irrigation, nombre de cycles culturaux, dates de semis ou de récolte...), à partir d'indices de végétation de séries temporelles MODIS ou AVHRR. Les approches les plus courantes sont par exemple l'analyse du profil temporel de NDVI (Defries and Townshend, 1994; Reed *et al.*, 1994), la classification multi-dates (Ruelland *et al.*, 2008) et les ACP (Analyses en Composantes Principales) (Townshend *et al.*, 1987; Tucker *et al.*, 1985). En voici quelques exemples :

- Wardlow *et al.* (2007) et Arvor *et al.* (2012) ont utilisé des indices de végétation MODIS pour discriminer différentes signatures temporelles pour chacune des grandes cultures de l'Etat du Kansas et du Brésil respectivement. Les différentes cultures du Kansas et du Brésil ont été facilement discriminées pendant la période de croissance puisque leur calendrier est très différent.
- Potgieter *et al.* (2007) ont réalisé une étude similaire, en comparant des signatures temporelles d'EVI en Australie, par la méthode de transformation de Fourier, pour séparer l'orge, le blé et le pois-chiche de la non culture.
- Sur une petite région du Sud du Brésil, Soares *et al.* (2008) ont essayé de déterminer les relations entre différentes métriques spatiales (aire, densité, forme, connectivité, contagion etc) et la culture au sol (maïs, canne à sucre et soja). Ils ont travaillé à partir d'images Landsat et MODIS, et mis en évidence que l'imagerie MODIS pouvait être utilisée pour cartographier les types de surfaces cultivées avec une précision similaire à celle de Landsat seulement pour les cultures de faible fragmentation, comme le soja.
- Concernant les pratiques, Thenkabail *et al.* (2005) ont testé différents indices MODIS pour cartographier les surfaces irriguées en Inde, alors que Wardlow *et al.* (2007) ont réussi à distinguer le maïs irrigué du non irrigué grâce à un seuil de NDVI à atteindre au cours de la saison de croissance.
- La gestion des terres cultivées a également été étudiée par Qiu *et al.* (2003), qui ont détecté les simple, double et triple cultures en Chine, à partir d'images Landsat, ou par Panigraphy and Sharma (1997), qui ont cartographié les rotations de cultures en Inde à partir de séries temporelles d'images indiennes (IRS).

Tous ces exemples s'intéressent à des sites d'études où l'agriculture est homogène, où les espaces cultivés sont grands et contigus. Or, tout comme pour la cartographie, les méthodes présentées ne sont pas pertinentes pour des pays où l'agriculture est fragmentée, comme c'est le cas

en Afrique de l'Ouest. Ozdogan and Woodcock (2006) ont montré la relation entre la résolution de l'image et les erreurs dans la classification des surfaces cultivées. En effet, les erreurs proviennent souvent des pixels « mixtes », i.e. au Mali, des pixels qui contiennent des parcelles de cultures différentes, ou bien de la végétation naturelle. Les champs sont en général plus petits que la résolution spatiale considérée.

2.3. Suivi temporel et estimation du rendement

A l'échelle globale, la littérature montre que le suivi de la croissance végétale se fait à partir de capteurs tels que MODIS (Terra) ou AVHRR (NOAA), grâce à la haute fréquence temporelle de ces images à basse ou moyenne résolution spatiale (Carrão *et al.*, 2008; Defries *et al.*, 1994; Doraiswamy *et al.*, 2004; Hountondji *et al.*, 2006; Justice *et al.*, 1985). A l'échelle régionale, avec des images à moyenne résolution spatiale comme LANDSAT (30 m), on peut également suivre des zones de culture en utilisant des séries temporelles durant la saison de croissance des plantes (Ippoliti-Ramilo *et al.*, 2003; Lowry *et al.*, 2007; Pax-Lenney and Woodcock, 1997). Enfin, avec des images à haute résolution spatiale (décamétrique), acquises par des satellites comme SPOT5 (Figure 8), il est possible d'élaborer un suivi très fin, à l'échelle de la parcelle (El Hajj *et al.*, 2009) avec une répétitivité décadaire à mensuelle (Tableau 1), mais les coûts engendrés sont rarement compatibles avec les budgets civils.

D'autres applications du suivi temporel concernent :

- la détection d'anomalies, qui peut se faire grâce à l'observation des changements de réflectance au cours d'une saison de végétation (Jakubauskas *et al.*, 2002; Sakamoto *et al.*, 2005; Wardlow *et al.*, 2008; Zhang *et al.*, 2003).
- l'étude des relations statistiques entre le rendement et le NDVI ou le climat, pour estimer le rendement. Ce sont Tucker *et al.* (1980) et Wiegand *et al.* (1979) qui ont identifié les premiers la relation entre le NDVI et le rendement d'une culture. Rasmussen (1992) et Maselli *et al.* (1993) ont utilisé des images AVHRR au Burkina et au Niger pour prévoir les rendements de mil et sorgho respectivement. Groten (1993) a estimé le rendement des cultures avec une précision de 15%, 60 jours avant la récolte au Burkina Faso, grâce à 5 années d'images AVHRR. Quant à Lobell and Field (2007), ils ont montré que plus de 30% de la variabilité interannuelle de rendements des six cultures les plus répandues dans le monde pouvaient s'expliquer par la température et les précipitations.
- le couplage de variables biophysiques de surface tels que le LAI (Leaf Area Index) à partir de données de télédétection, avec des modèles de plante permettant d'estimer le rendement à l'échelle globale ou régionale (Doraiswamy *et al.*, 2004; Doraiswamy *et al.*, 2005; Kastens *et al.*, 2005; Ma *et al.*, 2011; Xu *et al.*, 2011).
- l'évaluation par télédétection du développement phénologique de cultures, à partir d'indicateurs de séries temporelles définis comme des « métriques phénologiques » (Sakamoto *et al.*, 2005; Vina *et al.*, 2004).

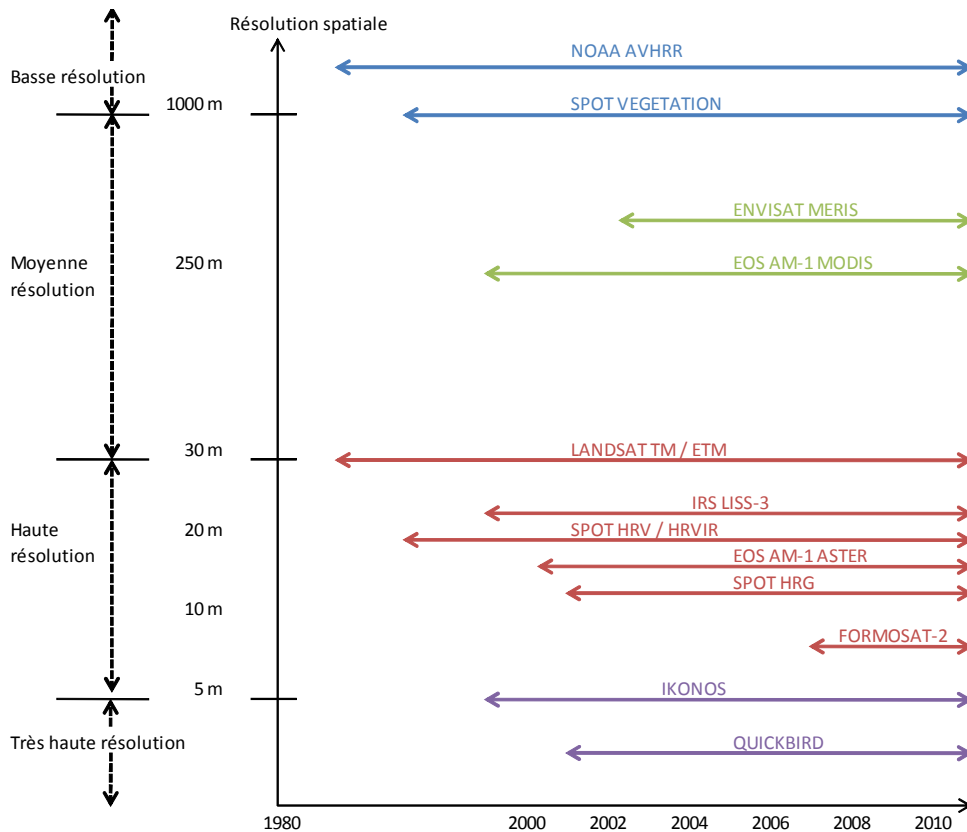


Figure 8 : Principaux capteurs fournissant des images de basse à très haute résolution spatiale dans le domaine du visible et de l'infrarouge.

De nombreuses études utilisent la télédétection pour la cartographie et le suivi des systèmes agricoles à l'échelle régionale ou globale. Dans ce cas, les images sont de résolution spatiale moyenne à basse, offrent une haute répétitivité temporelle, et donc la possibilité d'un suivi précis dans le temps. Les produits globaux constituent un apport certain dans la cartographie du domaine cultivé dans le monde, et en particulier en Afrique où la sécurité alimentaire reste précaire. Ils divergent cependant dans les régions où l'agriculture est fragmentée. Il en est de même pour la cartographie des systèmes agricoles : la plupart des travaux recensés développent des méthodes dans des zones où l'agriculture est homogène, où les espaces cultivés sont grands et contigus.

Il n'existe donc que peu d'études précises sur la cartographie, la caractérisation et le suivi des systèmes agricoles en Afrique de l'Ouest à l'échelle nationale, échelle pertinente pour les suivis d'alerte précoce. Ceci est dû en partie au fait que l'agriculture est très fragmentée dans cette région, avec une variabilité spatiale très importante et des pratiques culturelles très différentes en fonction des régions. D'autres difficultés surviennent du fait que la phénologie des cultures est synchronisée avec celle de la végétation naturelle en raison du régime des précipitations, ou encore de l'ennuage fort pendant la saison de croissance qui ne permet pas toujours un suivi temporel fin par satellite.

3. Les systèmes d'alerte précoce pour la sécurité alimentaire en Afrique de l'Ouest

3.1. Définitions et présentation des systèmes existants

Les systèmes d'alerte précoce visent à fournir une information fiable et préventive sur les risques potentiels de crise alimentaire ou d'insécurité alimentaire localisée (Brown *et al.*, 2007). Ce sont des systèmes d'information, de suivi et d'outils d'aide à la décision pour :

- estimer la production des principales cultures ;
- prévoir la production et les crises alimentaires ;
- apprécier l'évolution de la situation dans les zones les plus vulnérables ;
- suivre les populations vulnérables ;
- définir le contexte alimentaire.

Ces systèmes ont à la fois un rôle d'alerte en cas de crise (sécheresses, conflits) mais également de surveillance, sur le plus long terme, des besoins liés à des facteurs environnementaux (gestion de la fertilité des sols, démographie en hausse etc), et réfléchir sur la viabilité des systèmes agricoles actuels.

Les systèmes d'alerte précoce utilisent des méthodes différentes, sont associés dans des réseaux d'échanges des informations, et essaient de divulguer aussi rapidement que possible les informations alarmantes. Tous ces systèmes n'abordent pas tous les aspects de la sécurité alimentaire. Certains s'intéressent à l'accès aux marchés, d'autres à la disponibilité alimentaire, d'autres encore à la stabilité des différents piliers de la sécurité alimentaire (disponibilité, accessibilité...). Les activités du Global Monitoring for Food Security (GMFS) se limitent par exemple aux facteurs influençant la disponibilité alimentaire et pouvant être surveillés par télédétection alors que Global Information and Early Warning System (GIEWS) s'intéresse davantage aux marchés, à l'offre et la demande de céréales.

Nous décrivons en Annexe 6 le fonctionnement global du Système d'Alerte Précoce (SAP) du Mali, du Famine Early Warning Systems Network (FEWS NET), du Global Information and Early Warning System (GIEWS), du système Global Monitoring for Food Security (GMFS) et de l'Agrhymet, qui sont les cinq structures d'envergure analysant les besoins des populations vulnérables avec un suivi systématique d'un ou plusieurs pays d'Afrique de l'Ouest, et différents niveaux d'analyses et informations divulguées.

3.2. Les données d'entrée des systèmes d'alerte précoce pour le suivi agricole

Les données d'entrées indispensables au fonctionnement des systèmes d'alerte pour le suivi agricole (disponibilité alimentaire) sont les images satellitaires, des observations in-situ et des modèles de rendements. La disponibilité de ces données dans le temps constitue la contrainte majeure de l'estimation des productions agricoles. Les systèmes de suivi et d'alerte traduisent certaines ces données entrantes en informations pertinentes pour les professionnels et les filières (cartes d'occupation du sol, suivi des récoltes, détection des anomalies de croissance...).

3.2.1. Les données satellitaires

D'après les différents types de capteurs et d'images décrits précédemment, nous avons pu voir que pour le suivi de l'agriculture à l'échelle globale, ce sont les images à moyenne ou basse résolution spatiale et à haute fréquence temporelle qui sont utilisées. Ces images servent à localiser les cultures et à suivre leur état. D'autres images à plus haute résolution peuvent ensuite être utilisées pour la caractérisation de ce que nous avons appelé les « systèmes agricoles ».

L'estimation de la production agricole se déroule en 4 étapes : i) la cartographie des superficies emblavées par culture chaque année, ii) la détection d'anomalies par rapport à une année précédente, iii) le couplage de ces résultats à un modèle de plante, des données météorologiques et/ou des statistiques agricoles pour prévoir le rendement, et iv) la prévision de production en tant que telle. Dans chacune des trois premières étapes, des données satellitaires sont utilisées. Quelles sont-elles ?

Nous donnerons en exemple les deux systèmes d'alerte existants les mieux renseignés, à savoir le GMFS et le FEWS NET, et leur utilisation des données de télédétection. Le GMFS est déployé en Ethiopie, au Soudan, au Sénégal, au Zimbabwe, au Mozambique et au Malawi, et ne fournit donc aucune donnée sur le Mali ; cependant, ce système d'alerte est très bien renseigné et sa méthode mérite d'être analysée.

		Surfaces	Détection d'anomalies	Suivi agro-climatique	Pays d'Afrique concernés
GMFS	Données	ENVISAT-MERIS	SPOT VEGETATION et ENVISAT-MERIS	METEOSAT	Ethiopie, Soudan, Sénégal, Zimbabwe, Mozambique et Malawi + sites test
	Méthodes	Classification sur images à haute résolution + réseau de neurones pour le calcul de pourcentage de chaque occupation du sol dans un pixel MERIS	Cartes d'indicateurs de productivité de la végétation + carte de fAPAR tous les 10 jours, à comparer aux saisons précédentes pour prévenir les événements extrêmes	Cartes d'estimations des pluies, et du "Dry matter productivity" qui mesure le taux de croissance de la plante	
FEWSNET	Données	Landsat, Ikonos et Quickbird	NOAA-AVHRR et MODIS-TERRA via le NDVI.	Croisement des données NOAA-AVHRR et MODIS-TERRA avec le RFE	Burkina Faso, Tchad, Liberia, Mali, Mauritanie, Niger, Nigeria, Senegal, Sierra Leone + Afrique de l'Est et du Sud
	Méthodes	Classification sur images à haute résolution sur des sites test. Le Zimbabwe a été cartographié entièrement récemment.	Comparaison de l'année en cours avec la moyenne du NDVI des années précédentes.	Incorporation dans un modèle de plante. Performance de la culture en fonction de la disponibilité en eau par des cartes de Water Requirement Satisfaction Index (WRSI)	

Tableau 3 : Données et méthodes utilisées par les systèmes d'alerte FEWS NET et GMFS pour la cartographie et le suivi des surfaces cultivées.

Cartographie des cultures

Le GMFS propose des cartes du domaine cultivé à l'échelle nationale, générées chaque année, à partir d'images multi-temporelles ENVISAT-MERIS-FR (Tableau 3). Pour caractériser le

domaine cultivé et associer un type de culture à chaque pixel, le programme utilise une méthode « up-scaling ». Des classifications à haute résolution spatiale sont réalisées sur de petites zones tests, et utilisées pour entraîner un réseau de neurones qui permettra ensuite de calculer le pourcentage de chaque occupation du sol dans un pixel MERIS (300 m). Le réseau de neurones est ensuite appliqué à l'ensemble des images, à l'échelle nationale, en extrapolant les informations générées sur les zones tests.

Le FEWS NET se concentrait jusqu'à récemment uniquement à la détection d'anomalies de croissance, avec peu d'attention portée aux variations de surfaces cultivées pour une culture en particulier. Aujourd'hui le FEWS NET s'intéresse également à la cartographie précise des surfaces cultivées. C'est le cas par exemple au Zimbabwe, où une carte des cultures a été produite en 2005 grâce à 24 images Landsat. Ils utilisent également ponctuellement des images Ikonos et Quickbird (Tableau 3).

Détection d'anomalies

Le FEWS NET utilise des données NOAA-AVHRR et MODIS-TERRA, pour un suivi de la végétation, via l'indice NDVI (Tableau 3). Les anomalies sont étudiées : ce sont généralement des produits sous forme d'images, qui montrent combien la période présente est différente de la moyenne de toutes les images précédentes pour la même période.

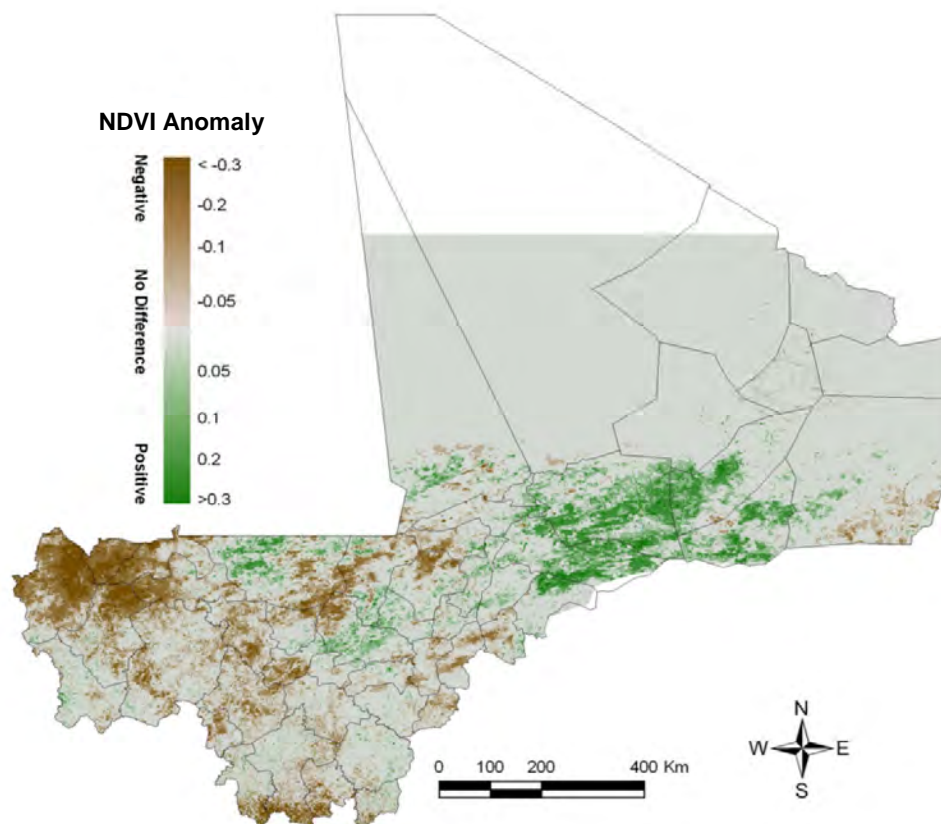


Figure 9 : Anomalies du NDVI au 31 Août 2011 par rapport à la moyenne 2001 – 2010 (Source : bulletin FEWS NET Septembre 2011).

Le GMFS produit des cartes d'indicateurs de productivité de la végétation (VPI), indice proposé par Sannier *et al.* (1998), basé sur des données SPOT VEGETATION à 1 km de résolution (Tableau 3). Cette méthode donne la distribution statistique du NDVI tous les 10 jours, pour prévenir des événements extrêmes. Concrètement, pour chaque pixel, la valeur du NDVI est comparée aux données historiques, et le NDVI est classé dans un « groupe de productivité », en fonction du pourcentage de différence avec les années précédentes (0%, 20%, 40%, 60%, 80% ou plus de 100%). Le GMFS produit aussi des cartes d'indice de la fraction de rayonnement photosynthétiquement actif absorbé par la végétation pour sa photosynthèse chaque saison (fAPAR), à partir d'imagerie ENVISAT-MERIS-FR à 250 m de résolution à 10 jours qui peuvent être comparés aux saisons précédentes.

Suivi agro-climatique

Pour le suivi climatique de FEWS NET, les données de NDVI sont croisées avec de l'imagerie RainFall Estimation (RFE) qui est un produit automatisé basé sur des images METEOSAT en infrarouge à 10 km de résolution (Figure 10) (Tableau 3). Ces données sont utilisées en input de modèles hydrologiques ou agro-météorologiques et fournissent également des informations sur le climat (i.e. permettent de comparer des quantités de pluies vis-à-vis de périodes antérieures). Le démarrage de la saison des pluies est un produit calculé d'après les données RFE. Le produit montre la date du démarrage de la saison, comparé à la moyenne climatologique (Figure 10). La saison est considérée comme « démarrée » lorsque l'on assiste à 3 périodes de 10 jours consécutifs avec plus de 20 mm de précipitations. La date du démarrage est ensuite comparée à la moyenne des dates de démarrage sur les 10 années précédentes (lorsque les données sont disponibles). Ainsi, les régions présentant un retard dans le démarrage peuvent être identifiées.

Les données de pluies sont cependant hétérogènes, dans le sens où le RFE ne couvre pas toutes les régions du monde. Le NDVI est donc la source d'information la plus importante pour le programme FEWS NET car il permet d'évaluer l'impact de la pluie sur la végétation. Ces données de pluie et de NDVI sont croisées avec des modèles de plante pour suivre l'évolution de la croissance des plantes. Cependant, deux saisons de croissance avec la même quantité de pluie peuvent amener à des rendements de culture bien différents si la distribution dans le temps est différente. Le programme FEWS NET a donc implémenté des mesures de l'impact des pluies sur chaque culture grâce au « Water Requirement Satisfaction Index » (WRSI) (Senay and Verdin, 2003; Verdin and Klaver, 2002). Cet indice donne la performance de la culture en fonction de la disponibilité en eau durant la période de croissance.

Le GMFS produit également des cartes d'estimations des pluies à partir de données METEOSAT à 3 km de résolution (Tableau 3). Il travaille également avec un indicateur appelé « Dry matter productivity » (DMP) qui mesure le taux de croissance de la plante. Quand la végétation est en bonne santé et que les nutriments ne sont pas limitants, le DMP est proportionnel à la quantité de lumière reçue par la plante. Les estimations de productivité de la végétation peuvent être obtenues en croisant les images de télédétection et des données météorologiques (rayonnement solaire et température). Leur calcul est basé sur la méthode Monteith (1972).

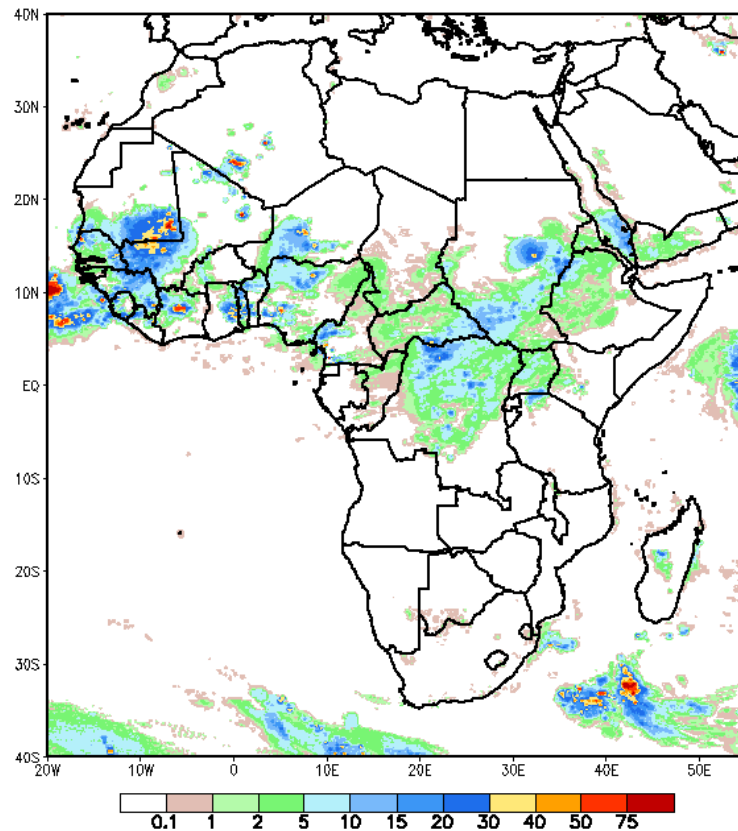


Figure 10 : NOAA Climate Prediction Center FEWS NET Rainfall Estimate (mm) en Afrique (7 Aout 2011).

3.2.2. Les données de terrain

Les enquêtes agricoles sont évidemment l'instrument de base de la détermination des disponibilités alimentaires. En principe, tous les pays du CILSS disposent d'une enquête agricole, mais il arrive qu'elle soit mise en œuvre trop tardivement ou annulée pour des raisons financières. Les enquêtes sont généralement faites sous forme de sondage (sur les superficies, l'estimation de la production à partir des carrés de rendements...) et reposent sur les déclarations des paysans. Elles font souvent l'objet de polémiques lorsque les résultats définitifs de production ne correspondent pas aux prévisions de récolte. Ce sont parfois les services nationaux qui sont suspectés de manipuler les chiffres pour influencer les résultats du bilan céréalier. De plus, ces enquêtes sont fastidieuses, et l'harmonisation des données produites par des dispositifs différents par région est compliquée. Des efforts notables ont été faits grâce au programme « Amélioration des instruments du DIAGNOSTIC PERMANENT pour la sécurité alimentaire régionale » (DIAPER) conduit par le centre Agrhymet sur la période 1984-2000, qui était un projet d'appui technique aux services nationaux des pays membres du CILSS contribuant à la collecte de données statistiques sur l'agriculture et l'élevage.

Au-delà des données statistiques ou satellitaires sur lesquelles s'appuient les systèmes d'alerte pour prévoir la production, des campagnes de terrain sont organisées chaque année dans

différents pays africains pour valider les produits cartographiques. Par exemple, en parallèle des données de pluies acquises par imagerie satellitaire, des données pluviométriques in-situ sont également relevées. La plupart des systèmes d'alerte utilisent également des évaluations à dire d'experts.

Les systèmes de surveillance de la sécurité alimentaire utilisent différentes plateformes satellitaires, ainsi que des données de terrain pour fonctionner. La démarche est la suivante : i) localiser les cultures, ii) caractériser le type de culture, iii) prévoir un rendement grâce à des données de pluies in situ complétées par des estimations par satellite et/ou des statistiques agricoles.

Le Mali n'a jamais été cartographié avec précision par aucun des systèmes d'alerte existants. Il se trouve dans une zone particulière en termes de régime pluviométrique et d'agriculture. Les systèmes agricoles maliens sont complexes, et gagneraient à être identifiés et caractérisés par télédétection pour injecter de l'information utile, continue dans le temps et dans l'espace, dans les systèmes d'alerte précoce.

Chapitre 2. DONNÉES ET OUTILS UTILISÉS

1. Données satellitaires

Trois types de données satellitaires ont été utilisés pour répondre à nos besoins (Tableau 5):

- des images MODIS et Landsat pour les développements méthodologiques,
- des images SPOT pour la validation de nos résultats.

Le système de coordonnées choisi pour ces images est l'UTM WGS 84 (Zone 30 Nord).

1.1. Les produits satellitaires MODIS

Les séries temporelles d'images satellitaires permettent d'exploiter la double dimension spatiale et temporelle des données de télédétection, offerte par le retour régulier des satellites au dessus des mêmes zones de la Terre. Les données et produits satellitaires utilisés dans cette thèse proviennent de la base de données de la NASA Land Process Distributed Archive Center (LP DACC).

Le choix du capteur MODIS a été motivé par quatre raisons : 1) la disponibilité et la facilité d'accès à la donnée (les données du capteur MERIS présentent un accès restreint à l'heure actuelle), 2) la capacité de couverture journalière de grandes étendues, 3) la résolution spatiale de 250 m la plus fine parmi l'offre en imagerie pour cette gamme de résolution spatiale (en réalité, les pixels ont une taille de 231.65 m x 231.65 m au nadir), et 4) la profondeur temporelle des archives (disponibles depuis février 2000).

Parmi les produits MODIS, nous avons utilisé (Tableau 5) :

- les synthèses temporelles de NDVI à 16 jours, à 250 m (MOD13Q1 ou « *Vegetation Indices 16 days L3 Global 250 m* ») (Figure 11) qui couvrent le Mali en deux tuiles. Nous avons récupéré la série complète (24 images) de 2007 pour être en cohérence avec les dates d'acquisition des images SPOT sur les sites d'études (voir ci-après). Sur ces 24 images, nous avons sélectionné celles qui présentaient une couverture nuageuse inférieure à 10%, en prenant soin d'avoir au moins une image par mois. Les dates des synthèses temporelles sélectionnées ont été les suivantes : 01/01, 03/02, 06/03, 23/04, 25/05, 26/06, 12/07, 29/08, 30/09, 16/10, 17/11, de l'année 2007. Pour une série temporelle MODIS, la date annoncée constitue le premier jour des 16 jours « synthétisés ». Toutes les images sont acquises en projection sinusoidale et ont été reprojctées en UTM WGS84 zone 30 Nord.

- les synthèses temporelles de EVI (Enhanced Vegetation Index) de 8 jours à 250 m (MOD13Q1 ou « *Vegetation Indices 16 days L3 Global 250 m* »), des années 2000 à 2010. Cet indice est calculé à partir du produit de réflectance terrestre de base (MODIS Terra surface reflectance) qui a été corrigé des effets atmosphériques (dispersion moléculaire, absorption de l'ozone et des aérosols).

- le produit MODIS Land Cover Type V005 (MCD12Q1 ou « *Land Cover Type Yearly L3 Global 500 m* ») de l'année 2007, qui est un produit annuel de cartographie de l'occupation du sol à l'échelle mondiale.

- le produit MODIS Land Cover Dynamics V005 (MCD12Q2 ou « *Land Cover Dynamics Yearly L3 Global 500 m* », qui synthétise en 7 variables la phénologie d'un couvert végétal, des années 2001 à 2009. Il renseigne sur la date de début de croissance végétale, de maximum, de

sénescence, de minimum de croissance végétale ainsi que sur les valeurs d'EVI pour les dates du minimum et du maximum. La dernière bande est l'intégrale d'EVI pendant la saison des pluies (Tableau 4).

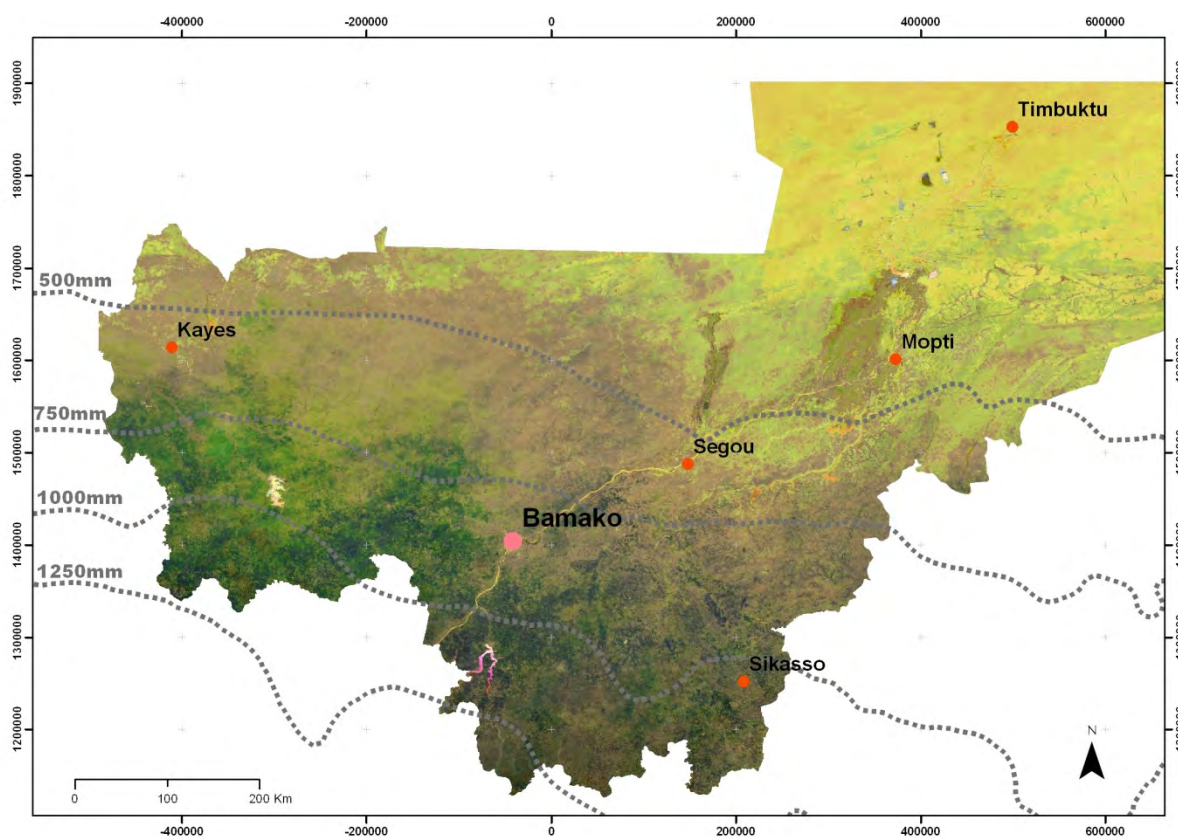


Figure 11 : Composition colorée RVB à partir de séries temporelles de NDVI au Sud Mali en 2007 (NDVI Mars / NDVI Juin / NDVI Septembre – MOD13Q1).

Band	Phenological variable name	Units	Valid range	Scale factor
1	Onset_Greenness_Increase cycle 1	Days since January 1, 2000	0–32766	1.0
2	Onset_Greenness_Increase cycle 2			
3	Onset_Greenness_Maximum cycle 1	Days since January 1, 2000	0–32766	1.0
4	Onset_Greenness_Maximum cycle 2			
5	Onset_Greenness_Decrease cycle 1	Days since January 1, 2000	0–32766	1.0
6	Onset_Greenness_Decrease cycle 2			
7	Onset_Greenness_Minimum cycle 1	Days since January 1, 2000	0–32766	1.0
8	Onset_Greenness_Minimum cycle 2			
9	NBAR_EVI_Onset_Greenness_Min cycle 1	NBAR EVI units	[-10 000 , +10 000]	0.0001
10	NBAR_EVI_Onset_Greenness_Min cycle 2			
11	NBAR_EVI_Onset_Greenness_Max cycle 1	NBAR EVI units	[-10 000 , +10 000]	0.0001
12	NBAR_EVI_Onset_Greenness_Max cycle 2			
13	NBAR_EVI_Area cycle 1	NBAR EVI units	[-3 660 , +3 660]	0.01
14	NBAR_EVI_Area cycle 2			
15	Phenology_Quality	Concatenated flags	N/A	N/A

Tableau 4 : Composition du produit annuel MCD12Q2 V005.

	Nom du produit	Synthèses temporelles	Résolution spatiale	Canaux téléchargés							Téléchargement	Format de téléchargement	Projection de téléchargement	Périodes téléchargées		
				B	V	R	PIR	MIR	NDVI	EVI						
MODIS	MOD13Q1	16j	250 m							x		HDF-EOS	Sinusoïdale	2007		
	MOD13Q1	8j	250 m								x			https://lpdaac.usgs.gov/lpdaac/get_data	2000-2009	
	MCD12Q1 (occ. sol)	Produit annuel	500 m													2007
	MCD12Q2 (pheno)	Produit annuel	500 m													
SPOT	SPOT 5 - HRG	Image journalière	10 m		x	x	x	x				HDF-EOS	Geog. Lat Long		12/12/2007 14/11/2007 20/11/2007	
LANDSAT	Landsat TM Multi spectral	Image journalière	30 m	x	x	x	x	x			http://glcf.umd.edu/data/			HDF-EOS	Transverse Mercator	Octobre et Novembre 2007

Tableau 5 : Principales caractéristiques des données satellitaires utilisées.

1.2. Les images Landsat

La résolution spatiale des images Landsat est de 30 m (ou 15 m en mode panchromatique) et les domaines spectraux explorés concernent le visible, l'infrarouge proche et moyen ainsi que l'infrarouge thermique (Tableau 5). Le radiomètre TM permet d'offrir des images couvrant un champ d'observation de 185 x 185 km².

Nous avons choisi 14 images Landsat 4-5 TM, pour couvrir tout le Mali. Ce sont des images journalières d'octobre et novembre 2007.

1.3. Les images SPOT

Nous avons acquis des images SPOT sur 4 sites en 2007, le long du gradient climatique malien (Tableau 6 et Figure 12). Il s'agissait d'acquérir les images les plus récentes possibles, présentant un bon échantillonnage des zones climatiques, et donc des systèmes agricoles présentés en Chapitre 1. Le site de Boro, au Nord, représente la zone sahélienne avec une agriculture céréalière clairsemée. Le site de Cinzana, en zone soudano-sahélienne présente également une agriculture en mil/sorgho, mais sur de larges étendues. Plus au Sud, le site de Koutiala est situé dans la zone cotonnière, où la pluviométrie est d'environ 800 mm par an. Le dernier site, Sikasso, est situé dans la zone la plus humide, avec une végétation naturelle dense, où les agriculteurs cultivent majoritairement du coton, du maïs et des arbres fruitiers.

Site d'étude (zone éco-climatique)	Précipitations moyennes annuelles	Cultures principales	Végétation naturelle (en majorité)
Boro (sahélienne)	500 mm	Mil et sorgho	Végétation dégradée et sol nu
Cinzana (soudano- sahélienne)	650 mm	Mil et sorgho	Végétation dégradée et sol nu
Koutiala (soudano- sahélienne)	800 mm	Coton, mil et sorgho	Végétation ouverte et fermée
Sikasso (soudanienne)	1000 mm	Maïs, coton et fruits	Végétation naturelle dense

Tableau 6 : Principales caractéristiques des 4 zones d'étude retenues pour l'acquisition des images SPOT.

Des images du satellite SPOT-5 ont été acquises sur ces sites (60 x 60 km²) grâce au montage d'un dossier ISIS auprès du CNES (Tableau 5). Les instruments HRG 1&2 du satellite SPOT-5 permettent d'acquérir le rayonnement dans quatre bandes spectrales : Vert, Rouge, Proche Infra-Rouge (PIR) et Moyen Infra-Rouge (MIR). Ces quatre bandes correspondent respectivement aux gammes de longueurs d'ondes allant de 0,50 à 0,59 μm , de 0,61 à 0,68 μm , de 0,78 à 0,89 μm et 1,58 à 1,75 μm . Les images acquises par ces instruments sont caractérisées par une haute résolution spatiale : un pixel dans une image SPOT-5 représente un carré de 10 x 10 m² dans le Vert, Rouge et PIR, et 20 x 20 m² dans le MIR. Un traitement informatique spécifique des images HRG permet

d'accéder à une résolution de 5 m dans le mode multispectral et à 2,5 m dans le mode panchromatique.

Les images SPOT ont été livrées au niveau 3 de prétraitement (images orthorectifiées, directement superposables à d'autres données cartographiques géoréférencées). En l'absence d'images acquises en 2009 (année de terrain – voir ci-après) et 2008 sur les sites d'étude, nous avons acquis des images à 2.5 m de résolution spatiale en 2007 : 12 décembre 2007 pour Boro, 14 novembre 2007 pour Cinzana et 20 novembre 2007 pour Koutiala et Sikasso.

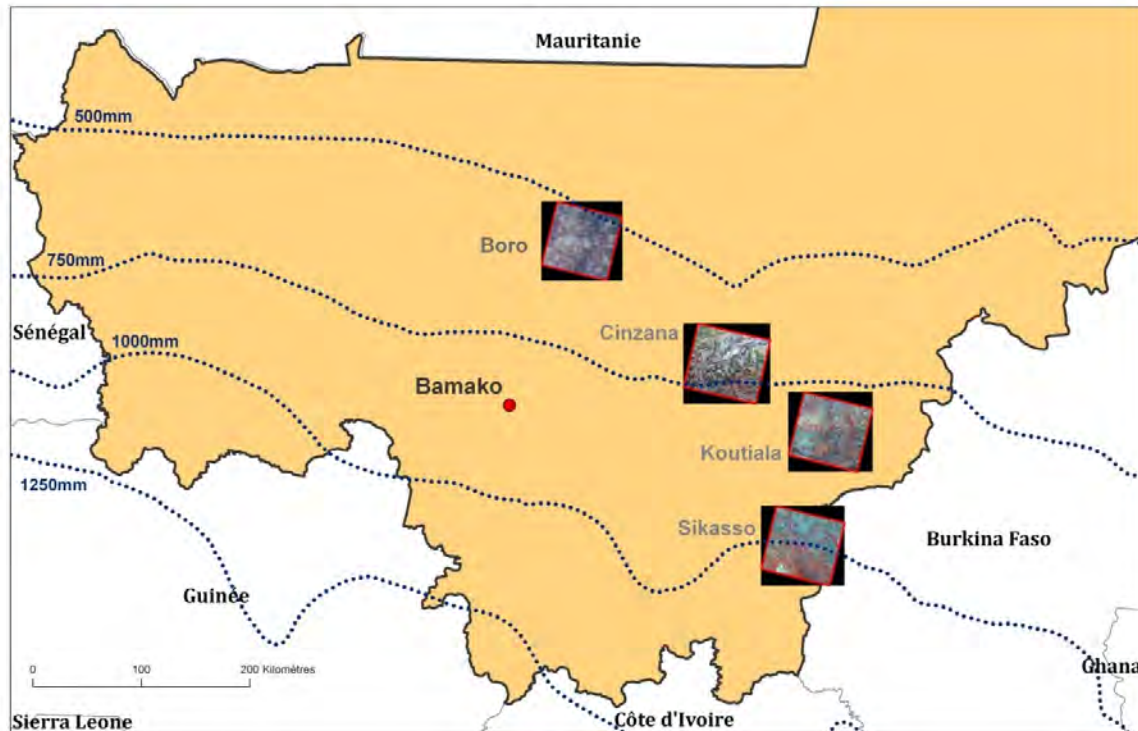


Figure 12 : Localisation des 4 images SPOT acquises en 2007 au Mali : Boro, Cinzana, Koutiala et Sikasso et Quicklook des images.

1.4. Indices utilisés

Après traitements des images, trois types d'indices ont été utilisés dans nos travaux.

1.4.1. Indices spectraux

Les indices de végétation sont des combinaisons, linéaires ou non, de réflectances dans les bandes spectrales R (rouge), PIR (proche infrarouge) et MIR (moyen infrarouge). Ils permettent de caractériser le couvert végétal en termes de vigueur de la végétation. Parmi ces indices, le « Normalized Difference Vegetation Index » (Rouse, 1974) est de loin le plus utilisé. Il permet de caractériser l'activité photosynthétique de la surface, et donc de discriminer facilement des surfaces végétalisées (0,9 pour de la végétation verte et dense) des surfaces de sols nus (environ 0,1 pour sol nu). Ces propriétés font que le NDVI est fréquemment utilisé comme une mesure indirecte de la biomasse. La formule du NDVI est la suivante :

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

Il existe également le « Enhanced Vegetation Index » (EVI) qui comprend une correction atmosphérique et ne sature pas sur les couverts à forte densité végétale (Huete *et al.*, 2002). La formule de l'EVI pour l'algorithme MODIS est la suivante :

$$EVI = 2.5 * \frac{PIR - R}{PIR + 6 * R - 7.5 * B + 1}$$

1.4.2. Indices texturaux

L'hétérogénéité spatiale de la couverture végétale est étudiée au travers de la distribution spatiale des variables radiométriques. La variabilité spatiale d'une image est représentée par le concept de texture.

Deux jeux d'indices ont été calculés à partir des synthèses temporelles MODIS, en choisissant uniquement 3 dates au sein de la période de croissance des cultures (mars, juin, septembre) pour alléger les calculs. Les deux indices proposés par le logiciel ERDAS Imagine sont la variance et la dissymétrie (« *skewness* »). L'indice de variance donne au pixel central la valeur moyenne de variance calculée sur tous les pixels du motif (Figure 13). L'indice de dissymétrie caractérise le degré d'asymétrie de la distribution des valeurs d'une population (distribution non gaussienne avec asymétrie droite ou gauche).

La génération de ces indices est basée sur l'orientation de groupes de pixels, dans une direction spécifique (horizontal, vertical ou diagonal) et à une distance spécifique. Cet agencement est appelé « motif » ou « pattern » en anglais. Dans notre approche, nous avons déterminé empiriquement :

- une taille de motif de 45 pixels MODIS pour le travail de stratification à l'échelle nationale (voir Chapitre 3), ce qui correspond à une fenêtre d'environ 10 km x 10 km.
- une taille de motif de 7 pixels MODIS pour le travail de cartographie des systèmes agricoles à l'échelle du village (voir Chapitre 4), ce qui correspond à une fenêtre d'environ 1.6 km x 1.6 km.
- une direction diagonale d'un pixel pour le déplacement de la fenêtre de texture.

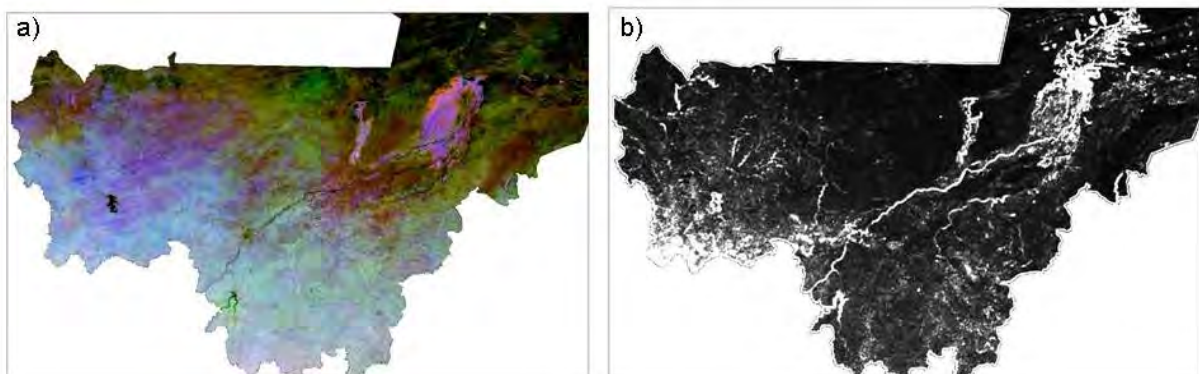


Figure 13 : Séries temporelles MODIS 2007 (mars /juin/septembre) a) NDVI et b) indice de variance.

1.4.3. Indices phénologiques

Ces indices sont issus du produit MCD12Q2 décrit précédemment (Tableau 4). Ils retracent la phénologie d'une culture au moyen de 7 métriques (Figure 14). Le premier indice décrit la date de démarrage de la croissance de la plante, le second correspond à la date de maturité, le troisième à la date de début de la sénescence, le quatrième à la date d'entrée en dormance, c'est-à-dire de fin de cycle. Le cinquième et le sixième indices correspondent aux valeurs de l'indice de végétation à la date de démarrage et de maturité respectivement. Le septième indice est l'intégrale de l'indice de végétation durant la saison de croissance.

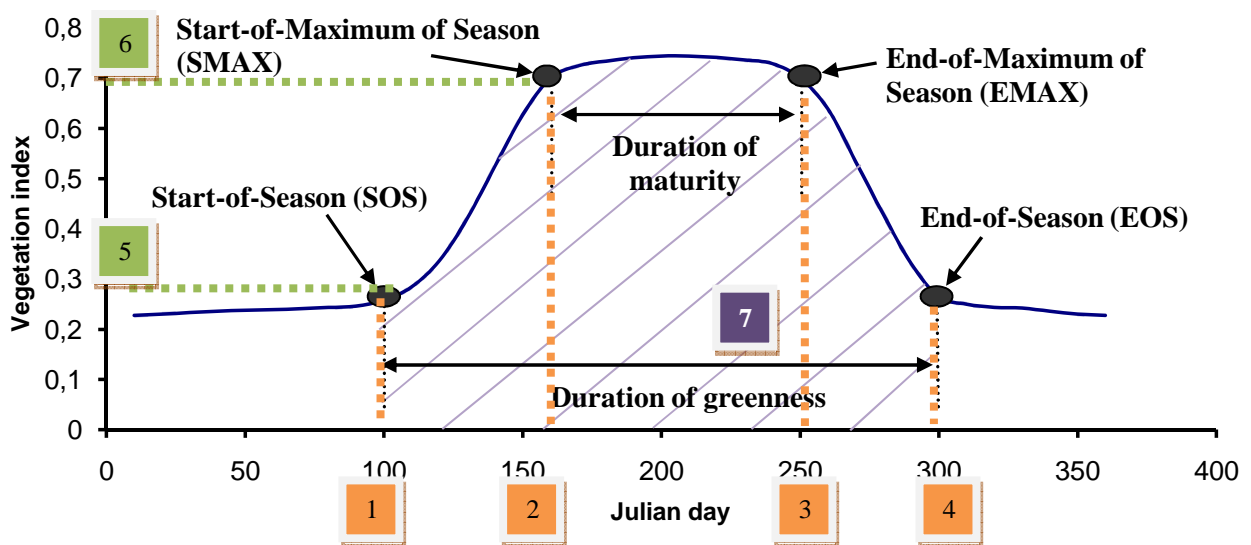


Figure 14 : Indicateurs phénologiques issus du produit MOD12Q2. 1. Date de démarrage de la saison (Start-of-Season) ; 2. Date de maturité (Start-of-Maximum of Season) ; 3. Date de sénescence (End-of-Maximum of Season) ; 4. Date du minimum de la saison (End-of-Season) ; 5. et 6. Valeur de l'indice de végétation pour les dates du minimum et du maximum ; 7. Intégrale de l'indice de végétation durant la saison de croissance. (Source : Zhang *et al.*, 2003).

2. Les produits globaux existants

Les produits globaux utilisés, à savoir, le GLC2000, le Globcover, MODIS V05 ainsi qu'Ecoclimap-II ont été décrits au chapitre 1, section 2.1.1, leur présentation, comparaison et validation nécessitant de fait une revue bibliographique.

3. Données terrain

3.1. Points GPS

Dans le but de mieux connaître le terrain d'étude et de relever des points GPS qui constitueront la vérité terrain des classifications, des missions de terrain ont eu lieu au Mali lors d'une expatriation de 6 mois en 2009, à l'Institut d'Economie Rurale (IER) à Bamako, au laboratoire de

télétection. Quatre missions de 10 jours chacune ont été effectuées sur les 4 sites d'études (Cinzana, Koutiala, Sikasso, Boro). Au total, 980 points GPS ont été enregistrés en effectuant des transects de village en village, sur chacun des sites étudiés (extrait en Figure 15), avec une photo associée pour chaque point. Des paysans de chacune des régions étudiées ont également été interrogés. Les questions concernaient leur production, leur façon de cultiver, les rendements moyens, et ensuite à l'aide de dessins, sur les différences qu'ils percevaient entre leur parcelle et celles du voisin. Ce travail nous a permis de mieux appréhender les différents types de paysages agricoles au Mali : nous avons appris à reconnaître sur les images SPOT ce qu'il y a au sol grâce aux photographies prises avec les points GPS, mais aussi à reconnaître les différentes cultures, à comprendre les pratiques des agriculteurs, qui diffèrent dans chaque région. Il nous est apparu impossible de dégager des particularités claires pour chaque site étudié. En effet, les différentes pratiques culturelles dépendent essentiellement des pluies, mais également des moyens financiers de chaque village (qui eux conditionnent l'équipement et la fertilisation), de la main d'œuvre, de la distribution des engrais par la Compagnie Malienne Du Textile (CMDT) et des subventions de l'Etat. La différence la plus importante pour notre étude (qui implique la télétection) était sur la taille des champs. Nous avons remarqué des tailles de champs supérieures sur le site de Cinzana par rapport aux autres sites.

Une illustration non exhaustive de différents paysages rencontrés lors des missions est donnée en Figure 16.

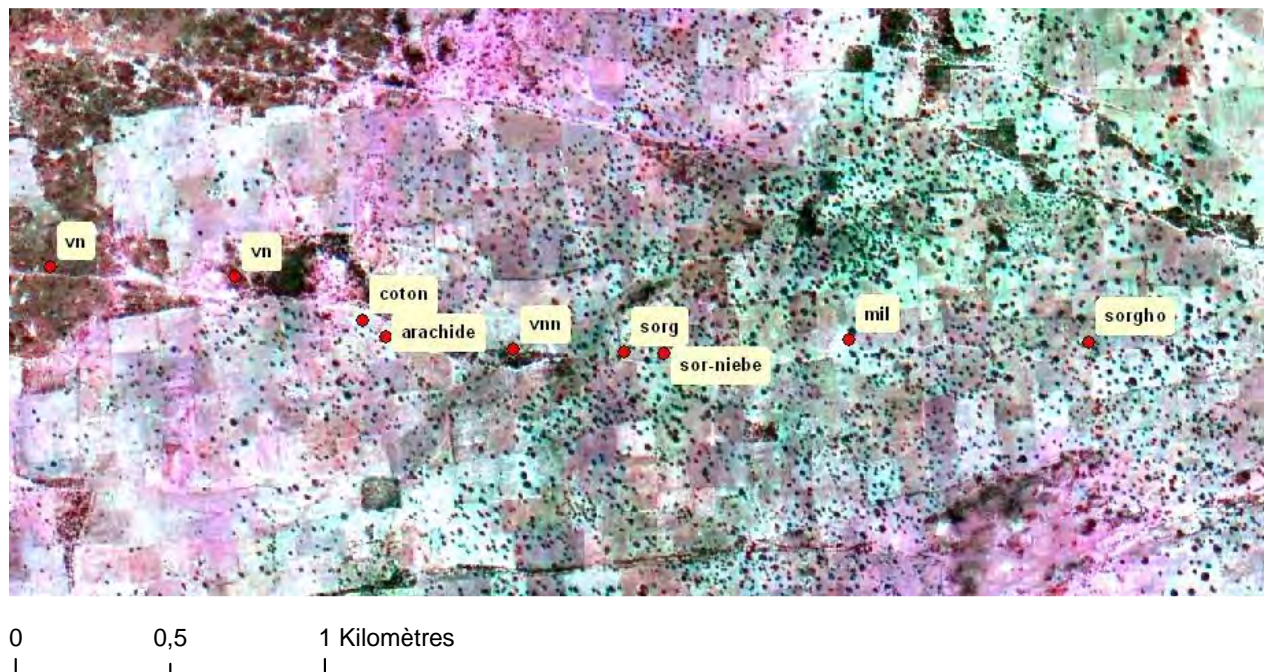


Figure 15 : Extrait des points GPS enregistrés dans la région de Koutiala, sur fond d'image SPOT (14/11/2007 ©CNES, distribution Spot Image). (Vn = végétation naturelle).



Zones cuirassées, impropres aux cultures



Tas de fumure avant labour des champs



Labour à la charrue



Coton cultivé avec intrants



Cultures mil/sorgho avec plus ou moins d'arbres et différents sols



Bas fond à Koutiala



Végétation naturelle et zone de pâturage



Sorgho à maturité



Culture du riz aux abords du fleuve

Figure 16 : Illustration de quelques paysages rencontrés lors de missions sur les sites d'études (E.Vintrou).

3.2. Base de données CMDT / IER

Une base de données de l'Institut d'Economie Rurale (IER) de Bamako et de la Compagnie Malienne Du Textile (CMDT) nous a été fournie par l'intermédiaire de Mamy Soumaré, qui a effectué ses travaux de thèse sur la dynamique des systèmes agraires dans la zone cotonnière (Soumare, 2008). M. Soumaré a proposé un zonage par régions agricoles et systèmes agraires locaux dans les zones cotonnières du Mali (Figure 17). Pour ce faire, il a proposé deux démarches complémentaires : une quantitative de type *Bottom-up* et une autre qualitative de type *Top-down* essentiellement basée sur l'analyse du paysage et des diagnostics agraires, pour comprendre la dynamique agricole d'une région sans perdre de vue les comportements des acteurs locaux. Au total, ses travaux de thèse recensent 6200 exploitations décrites à partir de 50 variables principales portant sur leur structure : la population, les équipements agricoles, l'élevage et les cultures. Le recueil des données a été étalé sur 3 ans (2001-2004) avec une méthodologie inchangée dans le temps et dans l'espace. L'ensemble des données – 50 variables * 6200 exploitations – a été traité avec le logiciel de statistique R[®] pour simplifier le jeu de données. Une Analyse en Composantes Principales (ACP) a été réalisée sur les 31 variables les plus discriminantes. Les deux premiers axes font ressortir d'une part, la taille des exploitations et d'autre part les options d'intensification. Ce sont la population totale, la superficie totale, le nombre de bovins hors bœufs de labour, la surface cultivée par actif, le nombre de bœufs de labour par actif, les surfaces en maïs, mil, sorgho et coton par actif qui participent le plus à la formation de ces axes. Les résultats des enquêtes à l'échelle de l'exploitation ont été ramenés à l'échelle du village par extrapolation.

En définitive l'ensemble des démarches suivies dans ces travaux de thèse (Soumare, 2008) - approche ascendante et quantitative et démarche descendante et qualitative- a permis d'identifier 15 systèmes agraires locaux. Ces 15 systèmes sont regroupés en grandes zones agricoles : la zone nord soudano-sahélienne, le vieux bassin cotonnier, la zone hétérogène en équilibre, la zone sud d'extension, la zone de la Haute Vallée du Niger et la nouvelle zone cotonnière (Figure 17).

Dans chacune de ces zones, un minimum de 10 villages a été enquêté, mais certaines présentaient des données manquantes. Nous avons donc retenu 100 villages dont les données étaient complètes, répartis dans quatre grandes zones climatiques (Figure 18). Nous retenons pour notre étude différents indicateurs ramenés à l'échelle du village.

Outre ces enquêtes, l'IER nous a également fourni une base de données géo-référençant 4000 villages du Sud du Mali.

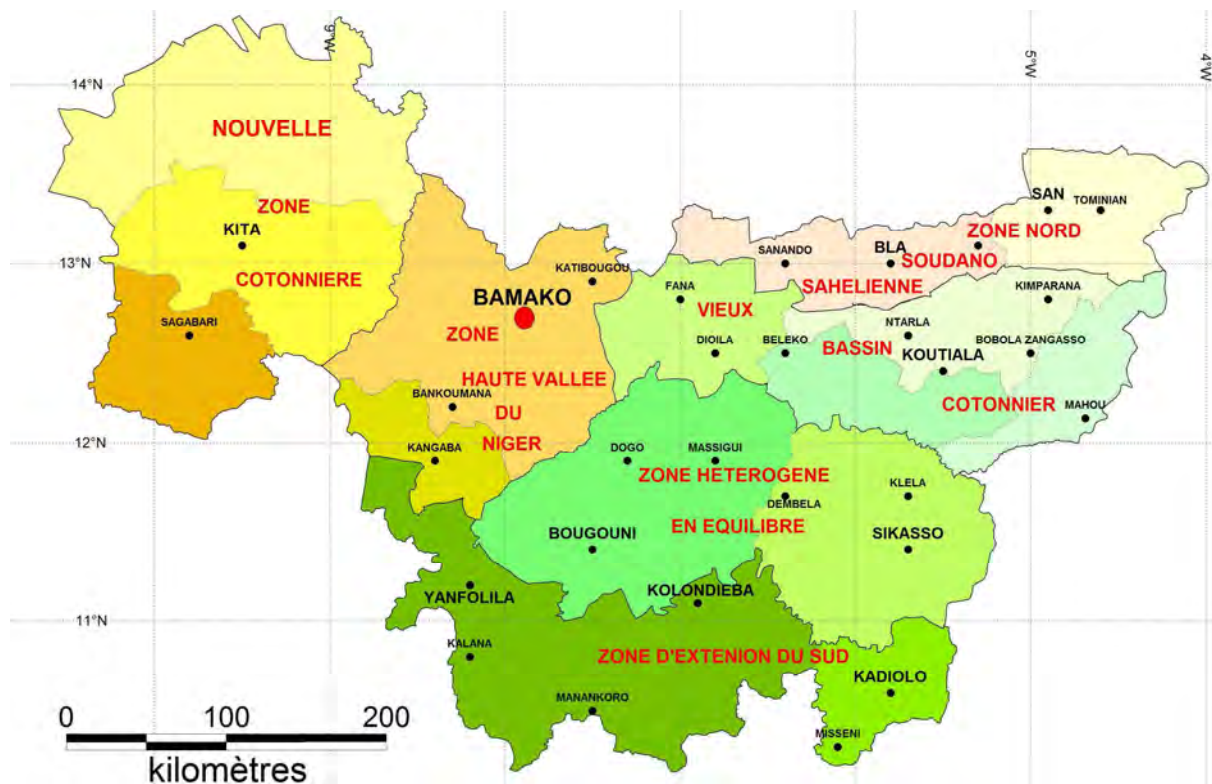


Figure 17 : Régions agricoles et systèmes agraires locaux dans les zones cotonnières du Mali (Soumare, 2008).

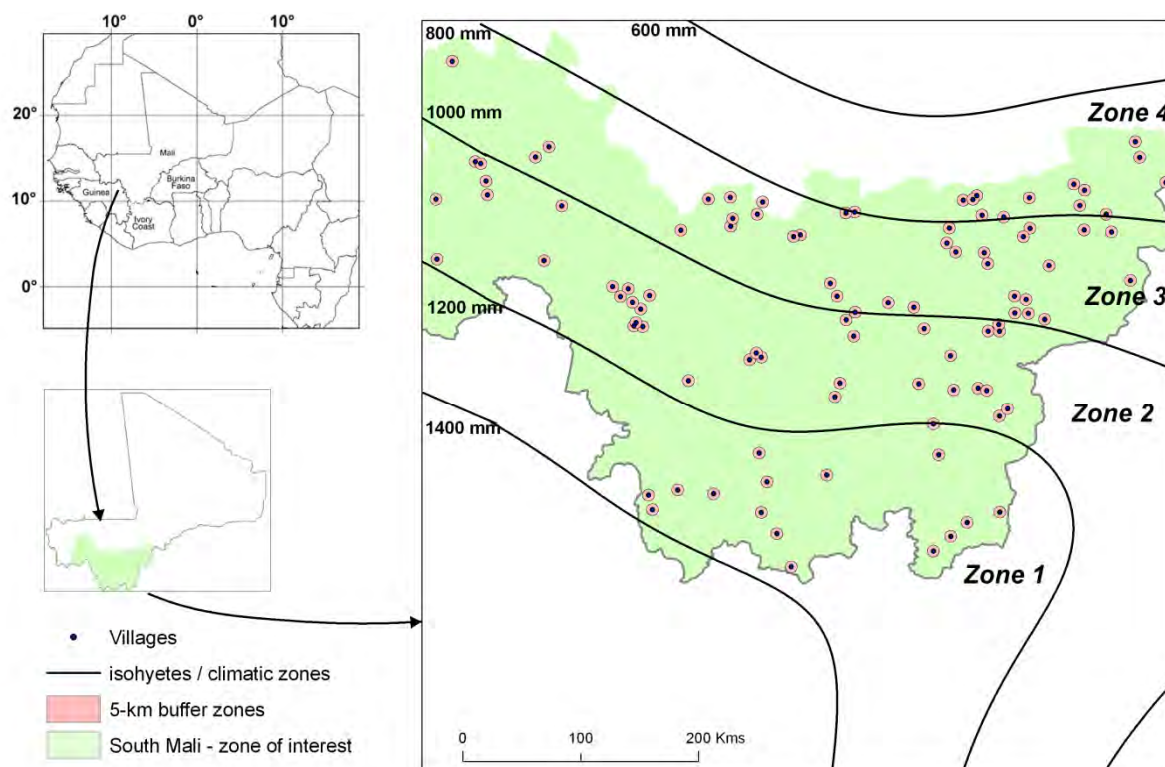


Figure 18 : Localisation des 100 villages enquêtés de la base de données de l'IER et de la CMDT.

3.3. Autres données

Données climatologiques et pluviométriques

Des données pluviométriques et climatiques journalières enregistrées dans 20 stations synoptiques (Figure 19) ont été recueillies par l'Agrhymet pour la période d'acquisition satellite (2000-2010). Sur ces 20 stations, 12 se sont révélées inexploitable en raison d'un grand nombre de données manquantes ou de localisation en dehors des zones de cultures. Au final, 8 stations ont été utilisées, du Nord au Sud : Nara, Segou, San, Kita, Bamako, Koutiala, Bougouni et Sikasso. Ces données sont les précipitations journalières (mm), le potentiel d'évapotranspiration (mm), le rayonnement global (J/m^2), et les valeurs minimale, maximale et moyenne de température ($^{\circ}C$).

Elévation

Les données du Modèle Numérique de Terrain (MNT) mondial Shuttle Radar Topography Mission (SRTM) ont été importées depuis <http://srtm.csi.cgiar.org/>. Ce sont des données d'élévation à 90 m de précision au nadir.

Population

Les données de population par village au Mali proviennent de la Mission de Décentralisation et des Réformes Institutionnelles (MDRI) de 1998.

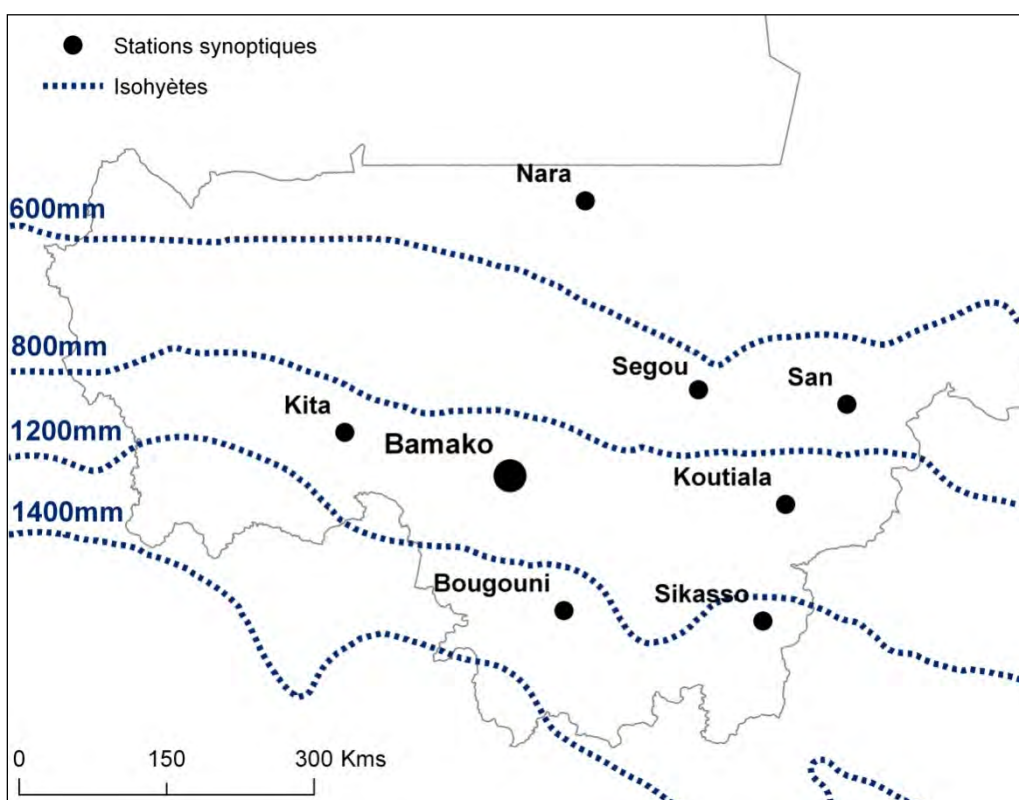


Figure 19 : Localisation des 8 stations synoptiques utilisées et des isohyètes (Source : IER Bamako).

4. Les outils

Au cours de ces travaux, différents outils et modèles ont été utilisés pour le traitement des données. Il nous semble important de décrire ici leur fonctionnement :

- Un outil classique de validation de classification : la matrice de confusion
- Un outil original d'extraction de connaissances : la fouille de données
- Un algorithme de classification par forêts aléatoires d'arbres : le Random Forest.

4.1. Précision de classification

Une matrice de confusion est un outil servant à évaluer la qualité d'une classification. Cette matrice affiche les statistiques de la précision de classification d'une image, notamment le degré de classification erronée parmi les diverses classes. Elle est obtenue en comparant les données classées (en ligne) avec des données de référence (en colonne), les valeurs étant des nombres de pixels ou d'objets. Ainsi, la diagonale de la matrice correspond aux individus (pixels ou objets) bien classés (en vert sur le Tableau 7). Si l'on considère la colonne i (individus qui dans la réalité appartiennent à la classe i), on trouvera dans la ligne j (case j,i) le nombre d'individus qui sur le terrain sont de la classe i mais ont été attribués à la classe j par la procédure de classification.

		Référence			Total
		Classe 1	Classe i	Classe n	
Classification	Classe 1	x11	x1i	x1n	M1
	Classe i	x $i+$	xii	x+i	Mi
	Classe n	xn1	xni	xnn	Mn
Total		N1	Ni	Nn	N

Tableau 7 : Notation sur une matrice de confusion.

De cette matrice, on peut calculer des estimateurs globaux qui donnent une appréciation globale de la qualité des résultats de classification :

- la précision globale (PG) désigne le nombre d'individus (objets ou pixels) bien classés, rapporté au nombre total d'individus. Elle correspond donc à la proportion d'individus bien classés mais ne donne pas d'information sur les classes possédant une meilleure fiabilité.

$$PG = \frac{\sum_{i=1}^n x_{ii}}{N}$$

On distingue d'autres mesures de fiabilité, comme les indicateurs par classe exprimant la

qualité de la classification pour chacune des classes :

- La précision pour le réalisateur (PR ou « Producer accuracy » en anglais, ou encore « Rappel » en statistiques) représente le pourcentage d'individus d'une classe de référence affectés à la même classe par la classification.

$$PR_i = \frac{x_{ii}}{N_i}$$

- La précision pour l'utilisateur (PU ou « User accuracy » en anglais, ou encore « Précision » en statistiques) représente le pourcentage d'individus (pixels ou objets) d'un groupe issu de la classification qui sont correctement classés vis-à-vis de la référence.

$$PU_i = \frac{x_{ii}}{M_i}$$

- L'erreur de commission est égale à (100% - Précision pour l'utilisateur).
- L'erreur d'Omission est égale à (100% - Précision pour le réalisateur).
- Le F-mesure (« F-score » en anglais) combine la précision pour l'utilisateur et la précision pour le réalisateur, tout en les pondérant de façon égale.

$$F_{\text{mesure}} = 2 \times \frac{PU \times PR}{PU + PR}$$

4.2. Fouille de données

La fouille de données (ou Data Mining en anglais) peut être définie par « *Processus non trivial permettant l'extraction automatique de connaissances d'une base de données pour obtenir de nouvelles données, valides, potentiellement utiles et compréhensibles* » (Fayyad et al., 1996). Plus largement, le concept d'extraction de connaissances consiste à sélectionner uniquement les données potentiellement utiles de la base (opération de filtrage), sur lesquelles on effectue une phase de prétraitement (gestion des données manquantes ou invalides) (Figure 20). Ensuite, les données obtenues passent par une phase de formatage, afin de les préparer au processus de Data Mining. La connaissance extraite est fréquemment exprimée sous forme de règles, construites à partir de motifs fréquents contenus dans la base de données. Finalement, la dernière étape est une étape d'analyse et d'interprétation de la connaissance extraite par la fouille de données, pour la rendre lisible et compréhensible par l'utilisateur.

Les travaux couplant la télédétection et la fouille de données sont récents et correspondent essentiellement à des approches utilisant les arbres de décision (Aksoy et al., 2004) ou la détection de changement au sein de séries temporelles (Romani, 2010; Silva et al., 2008). Il existe également des propositions de recherche de motifs comme outil d'exploration des données satellitaires, nous pouvons citer Julea et al. (2011) et Petitjean et al. (2011) qui extraient des patrons de comportement

ou identifient des profils de changement. Un des enjeux de l'application de ces techniques en télédétection réside dans le fait que les données relatives à une même réalité terrain sont par nature multi-sources, puisqu'elles résultent du croisement de données issues de différents capteurs, de relevés terrain et bases de données externes. Leur description se fait par ailleurs dans plusieurs dimensions et combine des informations spectrales, spatiales et temporelles. Pour cette raison, nous avons utilisé l'algorithme d'extraction d'itemsets multidimensionnels M^3SP (Plantevit *et al.*, 2010), qui peut prendre en compte le caractère multidimensionnel des données.

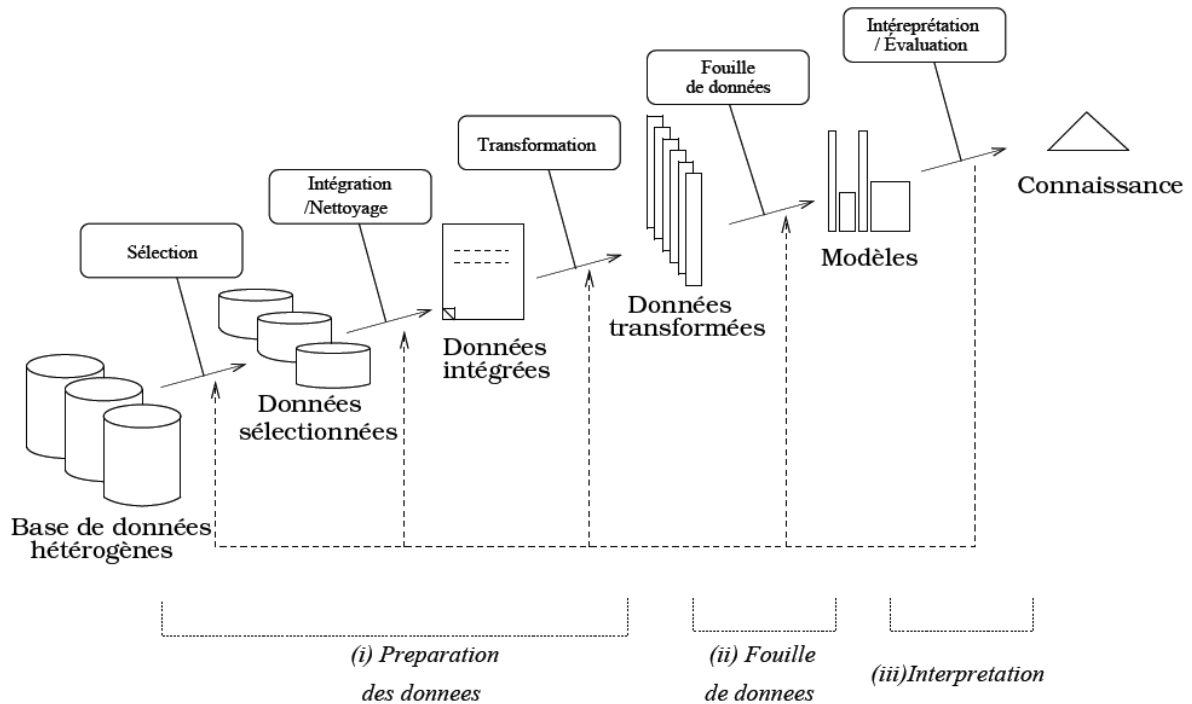


Figure 20 : Représentation classique du processus d'extractions de connaissances (d'après Fayyad *et al.*, 1996).

4.3. Random Forest

4.3.1. Définitions

Random Forest est une technique d'apprentissage supervisée, introduite par Breiman (2001), qui combine une technique d'agrégation, le « *bagging* », et une technique particulière d'induction d'arbres de décision. Les forêts aléatoires d'arbres (Random Forests ou RF) sont constituées de plusieurs arbres élémentaires travaillant en parallèle. Chaque arbre de classification (~ 500 arbres dans une forêt typique) est construit à partir d'un ensemble composé d'un échantillon aléatoire d'environ deux-tiers de la totalité des données. La classification d'un nouvel objet se fait en examinant le vote de chaque arbre (avec un poids égal), et en attribuant la classe proposée par la majorité d'entre eux. Le tiers-restant du jeu de données représente les observations « out-of-bag ». Elles sont utilisées pour tester en interne le modèle de classification, i.e. le modèle de classification est construit

sur 2/3 du jeu de données et testé sur le tiers-restant, cela étant répété pour chaque arbre de classification.

D'après l'état de l'art pour la cartographie de l'occupation du sol à partir d'images satellite, le classificateur RF a été jugé stable et relativement rapide, impliquant quelques paramètres définis par l'utilisateur et les précisions des rendements globaux qui sont soit comparables ou meilleurs que d'autres classificateurs, comme le maximum de vraisemblance et les arbres de décision classiques (Lawrence *et al.*, 2004; Na *et al.*, 2010), les arbres de décision AdaBoost et les réseaux neuronaux (Chan and Paelinckx, 2008), ou des machines à vecteurs de support (Pal, 2005).

4.3.2. Le paquet R *randomForest*

L'algorithme des Random Forests, codé par Breiman and Cutler (2005), a été importé dans le logiciel libre R par Liaw and Wiener (2002) via le paquet *randomForest*. Ce paquet est librement utilisable et est utilisé dans le traitement de nombreuses applications réelles.

L'algorithme possède uniquement deux paramètres à fixer par l'utilisateur. Il s'agit :

- du nombre de variables choisies aléatoirement à chacun des noeuds des arbres. Ce paramètre est nommé *mtry* dans le paquet. Il peut varier de 1 à p , mais Bernard (2009) a recommandé une valeur de \sqrt{p} en classification.
- du nombre d'arbres de la forêt. Ce paramètre est nommé *ntree* et sa valeur par défaut est 500. Il n'y a pas de pénalité pour l'utilisation d'un trop grand nombre d'arbres, comme cela a été prouvé mathématiquement par Breiman (2001) et confirmé expérimentalement dans deux autres études (Bernard *et al.*, 2009; Latinne *et al.*, 2001).

4.3.3. Evaluation du modèle de classification

Erreur Out-Of-Bag

Le classifieur Random Forest ne nécessite pas de validation externe. La précision du classifieur est calculée en interne. En effet, en plus de construire un prédicteur, l'algorithme de Random Forest calcule une estimation de son erreur de généralisation sur le tiers de données non utilisé pour la construction du modèle de classification : l'erreur Out-Of-Bag (OOB).

Importance des variables

Le paquet *randomForest* permet d'avoir des informations sur les variables des données que l'on étudie. Quelles sont les variables vraiment nécessaires pour expliquer la classification ? De quelles variables peut-on se passer ? Un indice d'importance des variables spécifiques aux forêts est introduit par Breiman (2001) ; il utilise la permutation aléatoire des variables. Plus les permutations aléatoires d'une variable engendrent une forte augmentation de l'erreur, plus la variable est importante. A l'inverse, si les permutations n'ont aucun effet sur l'erreur, la variable est considérée comme très peu importante.

Chapitre 3. CARTOGRAPHIE DU DOMAINE CULTIVÉ : MÉTHODES ET PRINCIPAUX RÉSULTATS

En Afrique de l'Ouest, l'estimation des surfaces cultivées reste problématique, et l'utilisation des systèmes d'observation satellitaire actuels peinent à aboutir en raison d'un domaine cultivé très fragmenté, d'une grande hétérogénéité spatiale due aux conditions environnementales et aux pratiques culturales, et de la synchronisation des phénologies des agro-systèmes et des écosystèmes liée au régime des précipitations. C'est pourquoi nous avons développé une méthode originale de détection du domaine cultivé (comprenant les surfaces en cultures et les jachères) au Mali par télédétection. Pour cela, nous avons considéré que les paysages perçus sur les images à moyenne résolution sont décrits par : i) une structure (liée à la morpho-pédologie et à l'utilisation du sol) et ii) une phénologie particulière de la végétation (en lien avec les zones éco-climatiques).

Ainsi, pour décrire et cartographier les paysages maliens, nous avons à la fois utilisé l'information spectrale, texturale et temporelle de séries temporelles d'images à moyenne résolution spatiale (MODIS, 250 m), ce qui constitue une approche « multi-indicateurs » originale de par l'utilisation de la totalité de l'information potentiellement contenue dans les images MODIS. La méthode utilisée - segmentation, photo-interprétation et classification – pour cartographier le domaine cultivé au Mali est une méthode éprouvée. Ce travail a été valorisé par un article publié dans une revue de rang A (Annexe 1):

Article 1 : Vintrou E., Desbrosse A., Bégué A., Traoré P.C.S., Baron C., Lo Seen D. (2012) Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*. **14**(1): 83-93.

Cette méthode, appelée par la suite « méthode géomatique » a été comparée à une deuxième méthode utilisant la fouille de données. La fouille de données, ou *Data Mining* en anglais, est un processus permettant l'extraction de connaissances automatiques sur une base de données hétérogènes, sans à priori, pour les rendre compréhensibles par l'utilisateur. Cette deuxième méthode est donc une alternative aux méthodes classiques pour définir un mécanisme d'apprentissage basé sur des données multi-sources. Elle se déroule en deux temps : (i) l'extraction de connaissances et (ii) le processus de classification à partir des règles extraites. L'extraction de connaissances (i) à partir de notre jeu de données est présentée dans deux articles de revues (Pitarch *et al.*, 2011; Vintrou *et al.*, 2012b) donnés en annexes 4 et 5. Sont présentés dans cette section uniquement les résultats du processus de classification (ii) et la comparaison entre les deux méthodes (fouille de données vs « méthode géomatique »).

1. Méthode géomatique

Cette méthode comprend deux étapes : la stratification du territoire et la classification culture / non culture dans chacune des zones (Figure 21a).

Les indicateurs spectraux (NDVI) et texturaux d'imagerie MODIS (décrits au chapitre 2, section 1.4) ont d'abord été utilisés pour stratifier le territoire en zones homogènes, afin de minimiser la variabilité spatiale des différentes classes d'occupation du sol et ainsi de limiter la confusion entre les classes. Cette stratification est faite avec le logiciel Ecognition-Definiens, en procédant à une segmentation, puis à une classification supervisée des segments. Le résultat de la stratification est présenté en Figure 21b.

Dans chacune des 13 zones, une classification pixellaire non supervisée à 20 classes a été réalisée. Chaque classe est ensuite assignée à une des deux classes prédéfinies (culture ou non culture) par photo-interprétation d'images Landsat 2007 (résolution 30 m). Le produit cartographique résultant nommé MODIS_{crop} est présenté en Figure 21b.

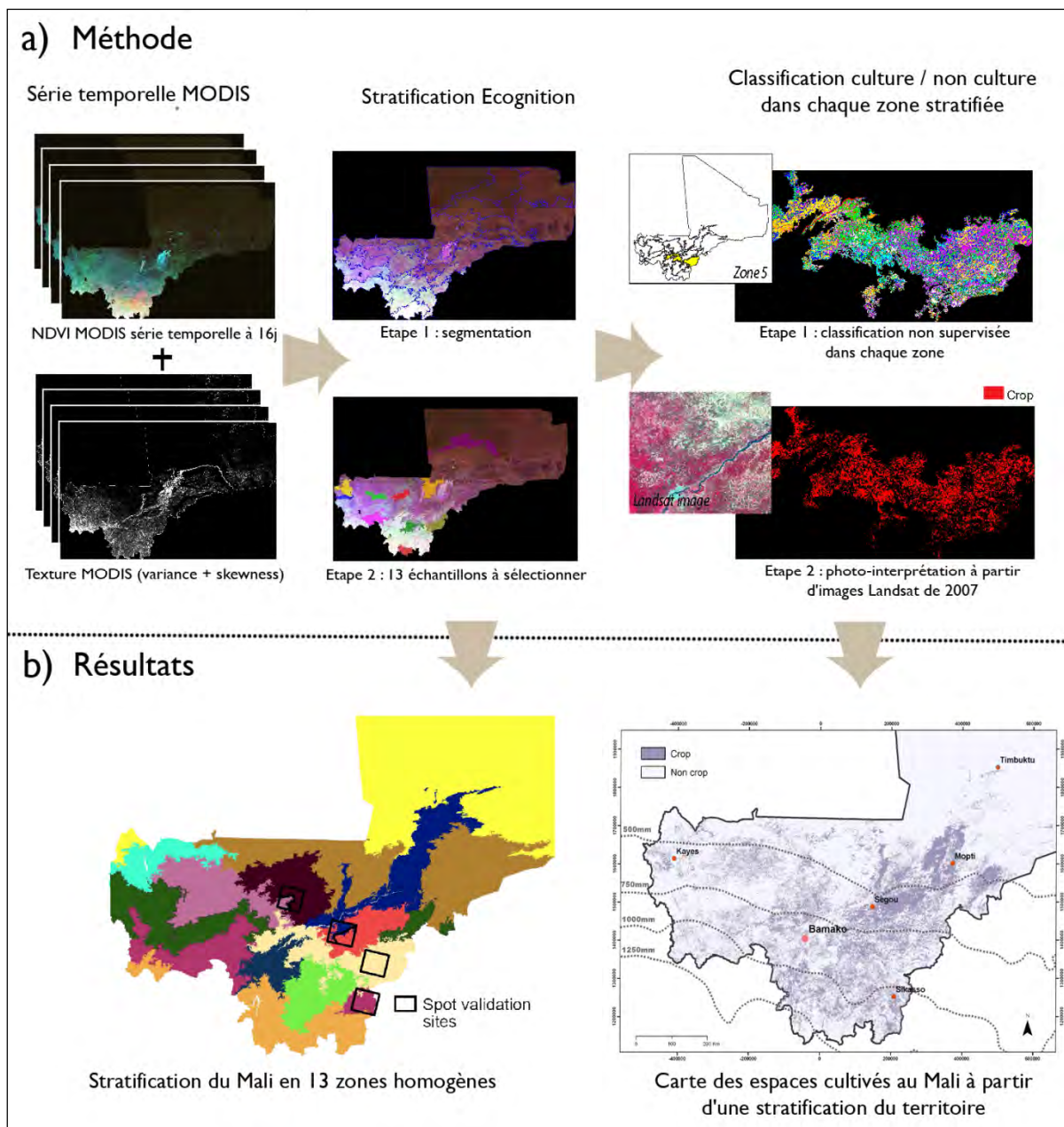


Figure 21: La méthode géomatique pour la détection du domaine cultivé au Mali : a) les différentes étapes de traitements et b) les principaux résultats.

2. Résultats

La carte du domaine cultivé a ensuite été : i) validée avec un jeu de données « référence », issu du traitement de 4 images SPOT (2.5 m de résolution), discriminant culture et non culture sur 6 sites (cf chapitre 2, Section 1.2 : les images SPOT de Boro et de Cinzana sont découpées en deux sites chacune selon un gradient éco-climatique), et ii) comparée avec 4 produits globaux existants (Globcover, GLC2000, MODIS V05 et Ecoclimap-II), à l'échelle locale et nationale. Les comparaisons se sont faites au moyen de matrices de confusion, comme expliqué en chapitre 2, section 4.1, après ré-échantillonnage de tous les produits à 2.5 m. La fonction « Matrix » du logiciel ERDAS permet de comparer deux classifications à partir d'une classification de référence (SPOT).

2.1. A l'échelle locale

2.1.1. Validation quantitative

Les résultats montrent que la superficie totale cultivée est bien évaluée avec le produit MODIS_{crop} pour 5 sites de validation sur 6 (< 6% de différence absolue) (Figure 22). Le produit global MODIS V05 présente des résultats similaires à ceux de MODIS_{crop} à l'échelle locale.

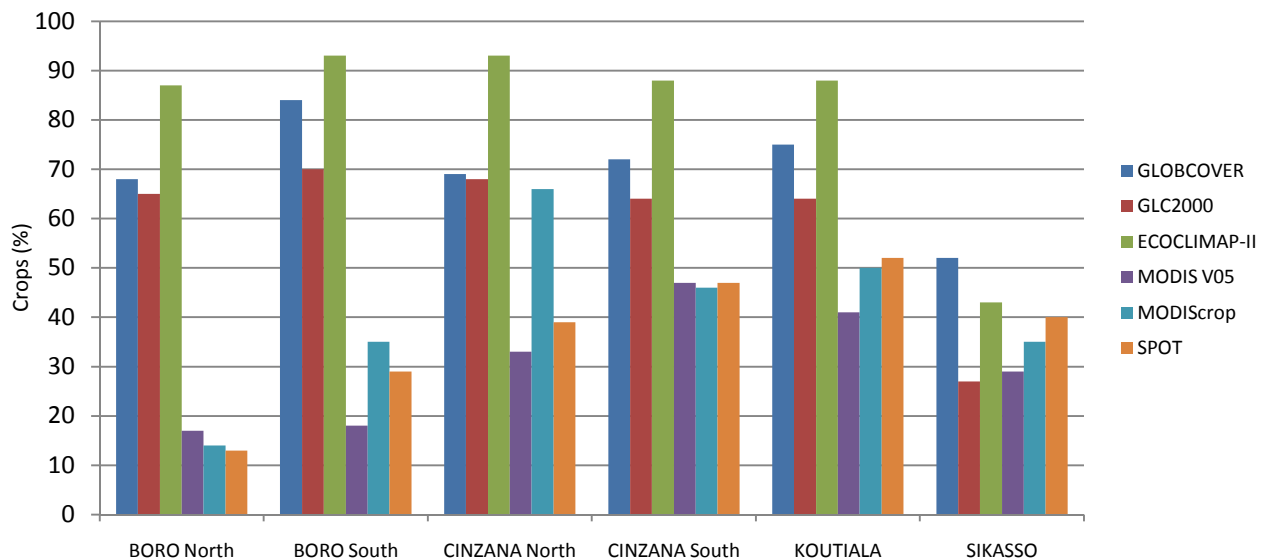


Figure 22 : Comparaison du domaine cultivé sur la référence (SPOT) du produit MODIS_{crop} et des produits globaux.

2.1.2. Comparaison de la distribution spatiale des cultures

Les deux produits qui quantifient le mieux les surfaces cultivées à l'échelle globale (MODIS V05 et MODIS_{crop}) ont été comparés à la référence SPOT via une matrice de confusion pour évaluer la précision de la distribution spatiale des cultures (Tableau 8). Le plus fort pourcentage de pixels correctement classés est obtenu avec le produit MODIS_{crop} (70%), qui présente une précision globale supérieure de 10% à celle de MODIS V05. De plus, les erreurs de commission et d'omission de MODIS_{crop} sont comparables pour chaque classe, ce qui implique que les surfaces cultivées ont été

estimées correctement, ce qui n'est pas le cas pour MODIS V05.

Nous avons par ailleurs mis en évidence que la précision de l'utilisateur est corrélée avec la taille des patches de culture ($R^2 = 0.8$). Comme attendu, les meilleures précisions sont obtenues dans les zones où l'agriculture est la moins fragmentée (Cinzana, Koutiala).

		MODIS _{crop}	MODIS V05*	MODIS V05
CROP	Erreur d'omission	37,1%	29,0%	60,3%
	Erreur de commission	40,3%	78,7%	42,6%
NON CROP	Erreur d'omission	25,6%	49,8%	27,0%
	Erreur de commission	23,5%	18,3%	38,1%
Pourcentage de pixels correctement classés		70,0%	58,3%	60,1%

MODIS V05*: avec la classe mixte considérée comme entièrement cultivée

MODIS V05: avec la classe mixte pondérée par 0,5.

Tableau 8: Matrice de confusion entre MODIS_{crop} et SPOT et entre MODIS V05 et SPOT pour les six sites de validation.

2.2. A l'échelle nationale

2.2.1. Validation quantitative

A l'échelle du pays, les superficies cultivées estimées par les produits GLC2000 et Globcover sont de l'ordre de 13 millions d'hectares. Les produits globaux MODIS V05 et ECOCLIMAP-II donnent des valeurs extrêmes, de 5 millions d'hectares pour le premier à 18 millions d'hectares pour le second. Le produit MODIS_{crop} donne une valeur moyenne à tous ces produits : 11,3 millions d'hectares cultivés, ce qui correspond toutefois au double de la valeur donnée par la FAO (Source : FAOSTAT, 2007).

2.2.2. Comparaison de la distribution spatiale des cultures

La comparaison des produits MODIS V05* et MODIS_{crop} révèle un fort désaccord sur la distribution spatiale des cultures à l'échelle nationale (77% si l'on considère uniquement la classe culture, 25% autrement) (Figure 23). Plusieurs raisons peuvent expliquer ces différences. Premièrement, les images ont des résolutions spatiales différentes. Deuxièmement, la nomenclature diffère entre les deux produits, avec une classe mixte pour MODIS V05*. Après vérifications sur Google Earth et par expertise locale, les zones encadrées en vert sur la Figure 23, détectées comme cultivées uniquement sur MODIS_{crop}, ont été confirmées comme cultivées.

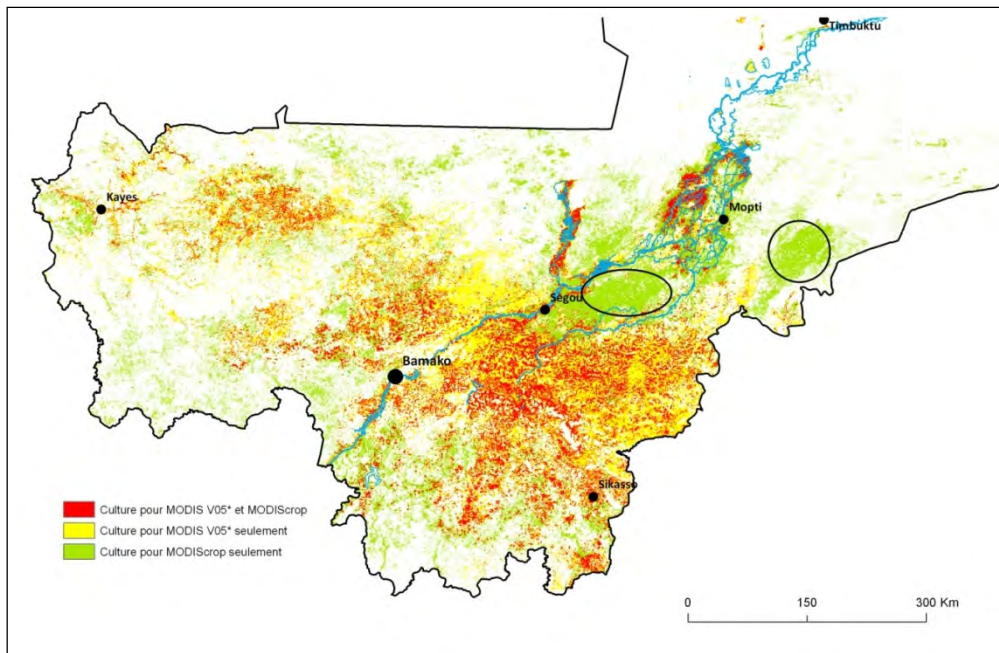


Figure 23 : Comparaison des produit MODIS_{crop} et MODIS V05.

3. Comparaison à la fouille de données

3.1. Extraction de connaissances

Un processus d'extraction de connaissances se fait en quatre étapes (Figure 24): (1) chargement des données, (2) préparation des données d'apprentissage (on associe à chaque point collecté une valeur de plusieurs indices de télédétection - NDVI et texture, calculée en prenant la moyenne de la valeur de l'indice sur les pixels contenus dans un polygone carré de 1 km de côté centré sur ce point), (3) fouille de données (l'algorithme M³SP appliqué sur les données d'apprentissage formatées recherche des motifs séquentiels fréquents) et (4) interprétation et validation par expertise des séquences obtenues sur les séries temporelles MODIS.

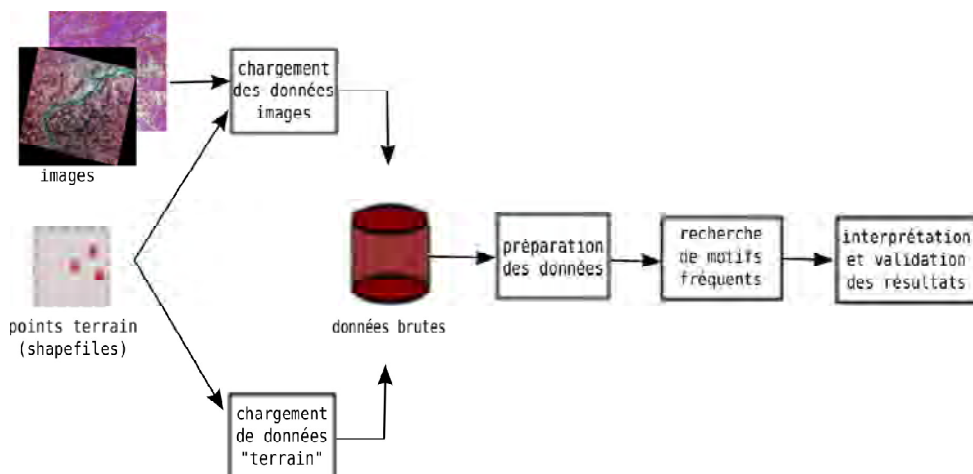


Figure 24 : Les différentes étapes du processus d'extraction de connaissances.

La méthode d'extraction de connaissance a été valorisée par deux articles (Annexes 4 et 5).

A partir des motifs extraits, on procède à la classification proprement dite. Les résultats de classification culture / non culture obtenus sont présentés ci-après et comparés aux classifications obtenues par la méthode dite « géomatique ».

3.2. Comparaison aux points terrain

Nous avons évalué les résultats de classification avec les mesures de Précision pour l'utilisateur et pour le réalisateur pour la classe culture, également appelés respectivement « Précision » et « Rappel » en statistiques (cf Chapitre 2, section 4.1), sur les résultats cartographiques des deux approches (géomatique et fouille de données). Nous avons utilisé pour l'approche par fouille de données les meilleures valeurs de Précision et de Rappel. En effet, celles-ci varient suivant le support (Pitarch *et al.*, 2011).

D'après le Tableau 9, les deux approches ont un F-mesure équivalent. Cependant, si l'on analyse les résultats de Précision et de Rappel, on peut noter que l'approche de télédétection offre des résultats plus stables, avec une Précision équivalente au Rappel, c'est-à-dire une Commission de pixels équivalente à l'Omission. Dans notre étude, et dans une thématique de sécurité alimentaire, il nous semble plus important de ne pas surestimer la classe culture. Nous pensons donc que la Précision est plus importante que le Rappel, et nous pouvons donc en conclure que l'approche dite « géomatique » a donné de meilleurs résultats. Cependant, le Tableau 9 montre également d'excellents résultats pour l'approche par Fouille de données, notamment pour le Rappel. Il faut également considérer que cette approche par fouille tient compte des points terrain pour bâtir son algorithme de classification, et que le calcul de F-mesure est effectué sur 10% des points restants.

Approche	Précision (Précision pour l'Utilisateur)	Rappel (Précision pour le réalisateur)	F-Mesure
Télédétection	0.7061	0.7422	0.7237
Fouille de données	0.6004	0.9155	0.7252

Tableau 9 : Comparaison des deux approches pour la détection du domaine cultivé.

3.3. Comparaison avec la référence SPOT

Nous avons comparé les deux résultats de classification avec les images référence SPOT sur les 6 sites de validation (Tableau 10). La matrice de confusion montre que pour la classe culture, c'est le produit MODIS_{crop} qui minimise les erreurs d'omission, alors que le produit MODIS_{fouille} minimise celles de commission. La fraction de pixels correctement classés est supérieure pour le produit MODIS_{crop}, obtenu par la méthode dite « géomatique ».

		MODIS _{crop}	MODIS_fouille
CROP	Omission error	37,1%	68,2%
	Commission error	39,0%	25,5%
NON CROP	Omission error	25,6%	20,93%
	Commission error	24,0%	41,44%
Fraction of correctly classified pixels		70,0%	57,8%
F-mesure		0,615	0,446

Tableau 10: Matrice de confusion entre MODIS_{crop} et SPOT et entre MODIS_fouille et SPOT pour les six sites de validation.

4. Discussions

Nous avons développé deux méthodes pour la cartographie du domaine cultivé au Mali. La première méthode comprend deux étapes : la stratification du territoire et la classification culture / non culture dans chacune des zones pré-définies, à partir d'indicateurs MODIS uniquement (et d'imagerie Landsat pour la photo-interprétation). La seconde méthode est une approche par fouille de données, basée sur les indicateurs MODIS et les points de terrain. Ces deux méthodes se sont révélées satisfaisantes. La première a montré une fraction de pixels correctement classés à 70% à partir de la référence SPOT sur les six sites de validation, contre 57,8% pour l'approche par fouille. La validation à partir des points de terrain a montré des résultats équivalents pour les deux méthodes, avec un F-mesure de 0,72.

Les erreurs de classification par classe (cultivée et non cultivée) se sont situées entre 24 et 39% pour l'approche « géomatique » et entre 20 et 68% pour l'approche par fouille, que nous pouvons expliquer par deux facteurs principaux : i) des tâches de cultures trop petites pour être détectées par MODIS à 250 m de résolution ou séparées de la végétation naturelle environnante, comme cela a été confirmé par Wardlow *et al.* (2007), ii) des erreurs dans la photo-interprétation à partir des images Landsat pour la première méthode ou des erreurs de classification dans l'algorithme pour la seconde méthode.

La limite de l'approche par fouille réside dans le fait qu'elle n'a été utilisée que sur les sites où existaient des relevés terrain. Les motifs de classification construits sur ces sites de validation sont en effet spécifiques à ces zones géographiques et l'application de ceux-ci sur d'autres zones a montré des résultats peu satisfaisants. Néanmoins, l'avantage de cette approche est de pouvoir fonctionner plus rapidement, l'extraction de connaissances étant « automatique », et la classification ne nécessitant que quelques secondes, alors que la photo-interprétation par l'approche « géomatique » est fastidieuse. De plus, l'approche par fouille de données a permis la mise en évidence de corrélations entre occupation du sol et indicateurs de télédétection qui n'avaient pas pu être mis en évidence par la première approche. En effet, dans les travaux de Pitarch *et al.* (2011) et Vintrou *et al.* (2012b), la texture s'est révélée être une dimension plus importante que le NDVI pour la distinction culture – non

culture. La prochaine étape consisterait à extraire des motifs sur d'autres zones pour pouvoir produire une cartographie des cultures à l'échelle nationale par fouille de données.

L'approche « géomatique » est avantageuse par le peu de relevés terrain nécessaires à sa mise en œuvre et l'applicabilité à l'ensemble du territoire. Grâce à cette première approche, la cartographie du domaine cultivé à l'échelle nationale au Mali (MODIS_{crop}) est un résultat prometteur. Les produits globaux existants donnent en effet à l'échelle locale de moins bons résultats sur la distribution spatiale des cultures, ainsi que de grandes différences d'un point de vue quantitatif. Ces résultats vont dans le sens de nombreuses études, qui mettent en avant les désaccords importants entre les produits globaux, spécialement dans la région ouest-africaine (Fritz *et al.*, 2008; Hansen and Reed, 2000a; Herold *et al.*, 2008; McCallum *et al.*, 2006) et en particulier pour la classe agriculture et sa distribution spatiale (Fritz *et al.*, 2010; Kaptué Tchuenté *et al.*, 2011). Giri *et al.* (2005) et Herold *et al.* (2006) ont expliqué ces différences en grande partie par les problèmes d'harmonisation des nomenclatures entre les différents produits. Nous pensons également que la fragmentation de l'agriculture ouest-africaine et la résolution spatiale utilisée jouent un rôle majeur dans la précision des produits globaux, ce qui corrobore avec les travaux de Achard *et al.* (2001) et de Pittman *et al.* (2010). Nous avons en effet montré dans cette étude que plus le site de validation avait une agriculture fragmentée, plus la précision de classification était faible. Nous avons également quantifié le seuil de détection d'un patch de culture MODIS à 200 hectares, c'est-à-dire à 37 pixels MODIS. Si l'on considère ce seuil de 200 hectares pour les autres produits globaux, cela correspond à 2 pixels (AVHRR et VEGETATION) et 8 pixels (MODIS V05), ce qui implique évidemment plus de difficultés à détecter le patch cultivé au sein de la végétation naturelle. Il est intéressant de noter que Doraiswamy *et al.* (2004) évoquent la possibilité de suivre des champs cultivés de plus de 25 hectares grâce à l'imagerie MODIS aux Etats-Unis. En Afrique de l'Ouest, en raison du « morcellement » du domaine cultivé, le patch de culture doit donc être 8 fois plus grand pour être détecté par les séries temporelles de MODIS NDVI.

Par ailleurs, la méthode Pareto (Boschetti *et al.*, 2004; Vintrou *et al.*, 2012a) a permis de séparer les sources d'erreurs provenant de la performance de l'algorithme de classification et celles provenant de la résolution des images. A partir d'un jeu de données à haute résolution (SPOT), il est possible de déterminer le minimum d'erreurs de commission et d'omission pouvant être atteint par la classification d'image à moyenne résolution. Sur nos sites d'études, 20 à 40% des erreurs sont incompressibles, i.e. dues à la résolution du capteur MODIS et à la fragmentation de l'agriculture au Mali.

Ces raisons peuvent expliquer les différences entre notre produit MODIS_{crop} et le produit MODIS V05 à l'échelle nationale. Par exemple, dans le Sud du Mali, notre produit a pu détecter des patches de cultures que le produit MODIS V05 n'a pas détecté, probablement parce que les patches de cette zone sont de trop petite taille. D'autre part, Ruelland *et al.* (2010) ont montré que dans l'aire soudanienne, les cultures ne sont pas stables au cours du temps, avec des périodes successives d'abandon et de recolonisation des terres. C'est pourquoi les divergences pourraient également être dues aux différentes dates d'acquisition d'images des différents produits.

POINTS-CLÉS : DETECTION DU DOMAINE CULTIVÉ

- **Nous avons examiné la possibilité de cartographier les espaces cultivées au Mali à partir d'imagerie MODIS à 250 m.**
- **Les désaccords entre produits globaux existants d'occupation du sol et le besoin urgent de cartographier ces espaces cultivés pour les systèmes d'alerte précoce sont les raisons qui ont motivé cette étude.**
- **Une méthode de stratification du territoire à partir d'indicateurs spectraux, spatiaux, texturaux et temporels d'imagerie MODIS et d'une classification objet a été développée.**
- **Nous avons démontré par deux méthodes différentes que les séries temporelles de NDVI MODIS sont capables de distinguer les surfaces cultivées de la végétation naturelle à l'échelle nationale, avec une précision de 70% pour la méthode « géomatique ».**
- **La méthode développée dans cette étude peut être reproduite dans d'autres pays d'Afrique de l'Ouest.**
- **L'étape suivante consistera à caractériser ces espaces cultivés, pour produire des cartes de types de systèmes agricoles qui pourront alimenter les modèles de culture et donc améliorer les prévisions de rendements.**

Chapitre 4. CARACTÉRISATION DES SYSTÈMES AGRICOLES

Afin d'améliorer l'estimation de la production des cultures en Afrique de l'Ouest, nous avons posé comme second objectif, après la cartographie des cultures, le développement d'une méthode permettant la cartographie des principaux systèmes agricoles du pays. Il s'agit de discriminer et d'identifier les systèmes agricoles à des fins de suivi et de modélisation des rendements de chaque grand type de culture. Nous avons émis l'hypothèse qu'un système agricole, au travers des pratiques culturales (espèces cultivées, intensification...), façonne le paysage, et que celui-ci peut être caractérisé par un ensemble de variables obtenues par satellite à moyenne résolution spatiale : i) une signature spectrale spécifique, indiquant une quantité de biomasse présente, ii) une signature spatiale et texturale spécifique, avec une organisation des éléments du paysage, et iii) une signature temporelle spécifique, représentant les variations saisonnières. Nous avons donc testé la possibilité d'utiliser des images à moyenne résolution telles que MODIS pour effectuer cette caractérisation, au travers d'indicateurs spectraux, spatiaux, texturaux et temporels, et de données de terrain. Grâce à un algorithme de classification (Random Forest), nous avons déterminé les liens entre le type de système agricole et les indicateurs de télédétection à l'échelle du village, et ensuite construit un modèle capable de prédire le système agricole majoritaire dans un village à partir des indicateurs de télédétection.

Cette étude a été valorisée par un article soumis dans une revue de rang A (Annexe 2):

Vintrou, E., Soumaré, M., Bernard, S., Bégué, A., Baron, C., Lo Seen, D. (2011). Mapping Fragmented Agricultural Systems in the Sudano-Sahelian region using Random Forest and Metrics of Coarse Resolution MODIS Imagery. (Special Issue) *Photogrammetric Engineering and Remote Sensing*. (Version acceptée en janvier 2012).

1. Méthode

La méthode est résumée dans la Figure 25.

1.1. Préparation du jeu de données

a. Constitution de la BD « Typologie des systèmes agricoles du Sud Mali » par village

Nous avons utilisé la base de données IER décrite au chapitre 2, section 3.2. Nous avons d'abord défini par expertise trois types de systèmes agricoles. Chacun des 100 villages de la base (Figure 18) a été assigné à une classe A, B, ou C, en considérant les cultures majoritairement cultivées sur ses terres :

- La classe A représente 29% du jeu de données. Elle regroupe les villages dont le domaine cultivé est composé à plus de 50% de mil et de sorgho. Le pourcentage de terre cultivée en coton dans ces villages est inférieur à 10%. La classe A représente la classe de cultures vivrières.
- La classe B représente 37% du jeu de données et regroupe des villages dont le

domaine cultivé est composé à plus de 40% de maïs et de coton. La classe B se réfère aux cultures intensives.

- La classe C représente 34% du jeu de données, et comporte des villages dont l'agriculture est moins intensive. Le domaine cultivé est à la fois composé de cultures vivrières (sorgho) et de rente (coton). Cette classe représente la classe « mixte ».

b. Echelle de travail

Pour pouvoir étudier le lien entre le type de systèmes agricoles et les indicateurs de télédétection, nous avons choisi une échelle de travail commune : l'échelle villageoise. L'agriculture malienne étant souvent sous forme de cercles autour d'un village, nous avons représenté ce village par une zone tampon de 5 km de rayon.

c. Calcul des indicateurs de télédétection

D'après notre hypothèse de départ, nous avons extrait différentes moyennes d'indicateurs par buffer, sur des images MODIS. Les 4 grands types d'indicateurs sont : le NDVI (4 indicateurs), la texture (12 indicateurs), la phénologie (7 indicateurs) et la fragmentation (3 indicateurs). Ils ont été extraits sur le masque de culture préalablement bâti (Vintrou *et al.*, 2012a) et présenté dans le chapitre précédent. A ces indicateurs de télédétection, nous avons ajouté quelques indicateurs socio-environnementaux tels que la population, l'élévation et la zone climatique (Figure 18).

1.2. Modèle Random Forest

Pour établir le lien entre le type de systèmes agricoles (A, B ou C) et les indicateurs de télédétection extraits sur le masque de culture, nous avons utilisé le classifieur Random Forest.

Le modèle a été entraîné sur les 100 villages, en considérant 30 indicateurs de télédétection (Figure 25) et l'appartenance à une classe qui donne le type de systèmes agricoles majoritaire dans le village.

Nous avons ensuite estimé l'importance de chaque variable, puis analysé la contribution de chaque grand groupe de variables, à savoir : le NDVI, la texture, les indicateurs phénologiques et les indices de fragmentation.

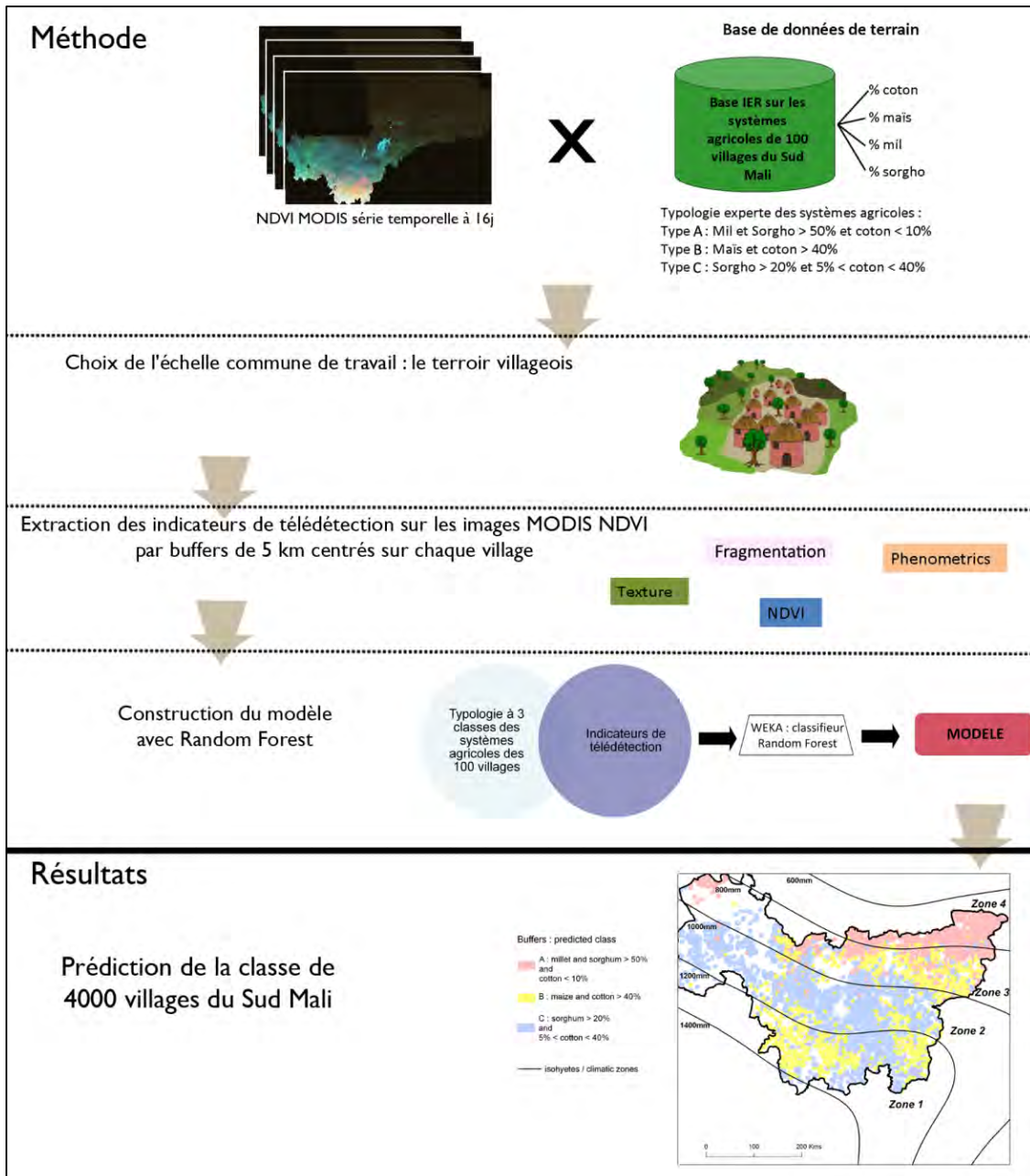


Figure 25 : Méthodologie et résultat pour la caractérisation des systèmes agricoles au Mali.

2. Résultats

2.1. Précision du modèle

L'erreur OOB a été utilisée pour l'évaluation de la précision. Le modèle Random Forest a fourni une bonne prédiction de la classe de systèmes agricoles à partir de la base de données d'apprentissage de 100 villages, avec une précision globale estimée de 60%. Les précisions par classe varient entre 50 et 70% (Tableau 11). Le modèle Random Forest semble être plus performant pour les villages de la classe A et B (F-score de 0.66 et 0.63 respectivement). On peut également noter d'importantes confusions entre la classe B et C. Enfin, les erreurs d'omission et de commission sont équivalentes pour chaque classe (<10% d'écart) (Tableau 11).

Pour évaluer la stabilité de l'algorithme, nous avons répété le procédé 50 fois et calculé les écarts-types de la précision globale et de la précision par classe. Ils se sont avérés acceptables (<0.04 pour la précision par classe, et <0.02 pour l'erreur OOB).

	Reference data				User Accuracy UA (commission error)	Producer Accuracy PA (omission error)	F-score ⁽¹⁾	
	A	B	C	Total				
Classified data	A	18	1	10	29	62% (38%)	70% (30%)	0.66
	B	5	22	10	37	59% (41%)	67% (33%)	0.63
	C	4	10	20	34	59% (41%)	50% (50%)	0.54
Total	27	33	40	100		Overall accuracy:	60%	

⁽¹⁾ F-score = $(2 * UA * PA) / (UA + PA)$

Tableau 11 : Matrice de confusion pour les 100 villages classés.

2.2. Importance des variables

La Figure 26a montre le classement de l'importance des variables et la distribution des principales variables par classe. Parmi les 8 premières variables, nous trouvons, par ordre d'importance, des métriques spectrales, phénologiques et texturales. La moyenne annuelle de NDVI sur le masque de culture constitue la variable la plus importante pour distinguer les types de systèmes agricoles. Viennent ensuite deux indicateurs phénologiques, les dates de début du maximum de la saison et la date de fin du maximum de la saison. Le maximum annuel de NDVI, ainsi que le moyenne au cours de la saison des pluies (mai-novembre) par village semblent également prendre part à la caractérisation des 3 types de systèmes agricoles. Pour la texture, le maximum de variance en mars (maximum variance 1) et le maximum de skewness en juin (maximum skewness 2) ont également été classés avec une importance particulièrement élevée. Enfin, le maximum de l'élévation (maximum altitude) a contribué sensiblement (rang 8) à la classification.

2.3. Importance des jeux de variables

Nous avons ensuite étudié la contribution des grands groupes de variables, à savoir les indicateurs de NDVI, de texture, de phénologie, la fragmentation, puis en marge, les données socio-économiques (élévation, population, climat). Les principaux résultats sont:

- le groupe de variables de texture obtient de meilleurs résultats de classification que le groupe de variables de NDVI (53% contre 45%).

- l'ajout du groupe des variables socio-environnementales au groupe de variables de NDVI n'améliore pas les résultats.

- l'ajout du groupe des variables de fragmentation au groupe « NDVI » améliore le résultat de classification pour les classes A et C (+7 et +14% respectivement).
- l'ajout du groupe des variables de texture au groupe « NDVI » et au groupe des variables socio-environnementales augmente les précisions de classification des classes A et C (7 et 14% respectivement).
- l'ajout des variables phénologiques au groupe « NDVI » n'est bénéfique qu'à la classe A (+13% de précision).
- les meilleurs résultats sont obtenus en utilisant toutes les variables (60%).

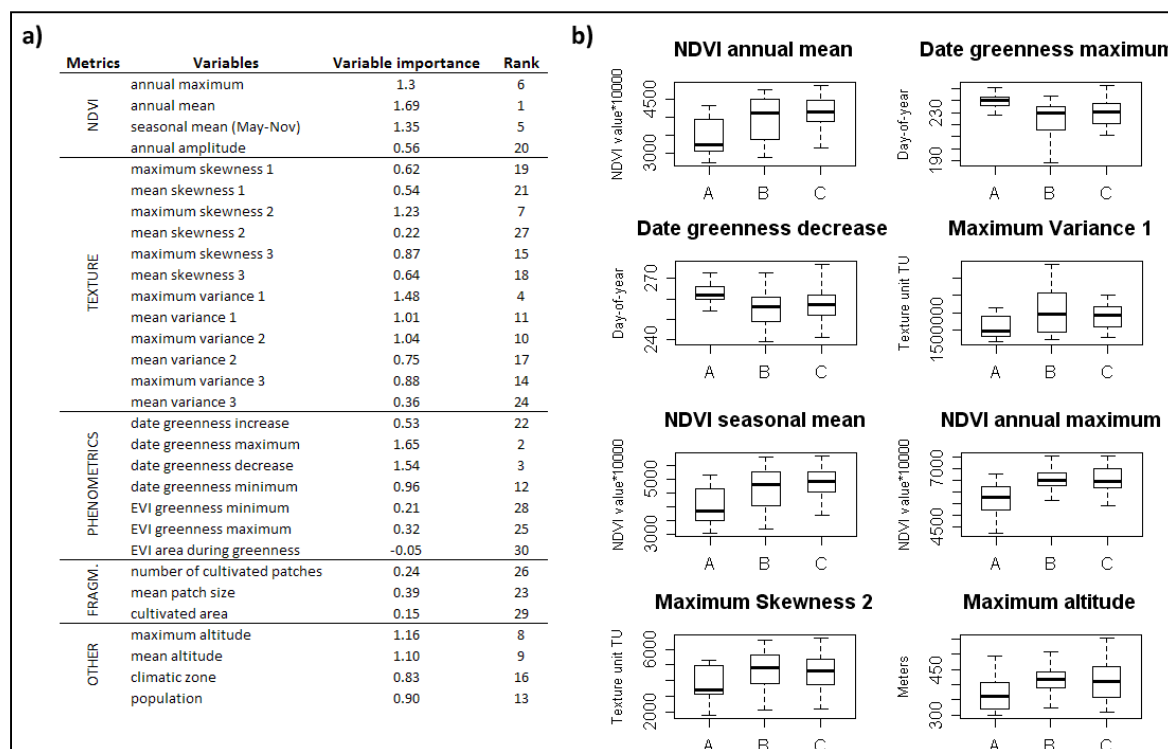


Figure 26 : Importance des variables (a) et distribution des principales variables par classe (b).

2.4. Cartographie des systèmes agricoles du Sud Mali à l'échelle villageoise

En appliquant le modèle sur la base de données des 4000 villages du Sud Mali fournie par l'IER, la cartographie des systèmes agricoles résultante distingue les principales zones d'agriculture (Figure 27):

- Le Sud de la zone soudano-sahélienne (zone 4), autour de Tominian et Ségou, où la classe A prédomine. En effet, le sorgho et le mil sont des variétés résistantes à la sécheresse, avec un cycle court de 90 jours (Kouressy *et al.*, 2008), adaptées à cette zone climatique.
- Dans le vieux bassin cotonnier, autour de Koutiala (zone 3), les systèmes agricoles sont centrés sur la production de coton (classe B). Cette zone évolue et présente aussi des villages de classe C, suite à la crise du coton.
- Les zones soudaniennes et soudano-guinéennes (zones 1 et 2) sont aussi basées sur le système cotonnier, mais considérées comme plus diversifiées, en particulier autour de Sikasso

(Soumare *et al.*, 2008), avec la présence de villages de classe B et C. Ceci peut s'expliquer par une saison des pluies plus longues, qui offre un plus large choix d'espèces et de variétés.

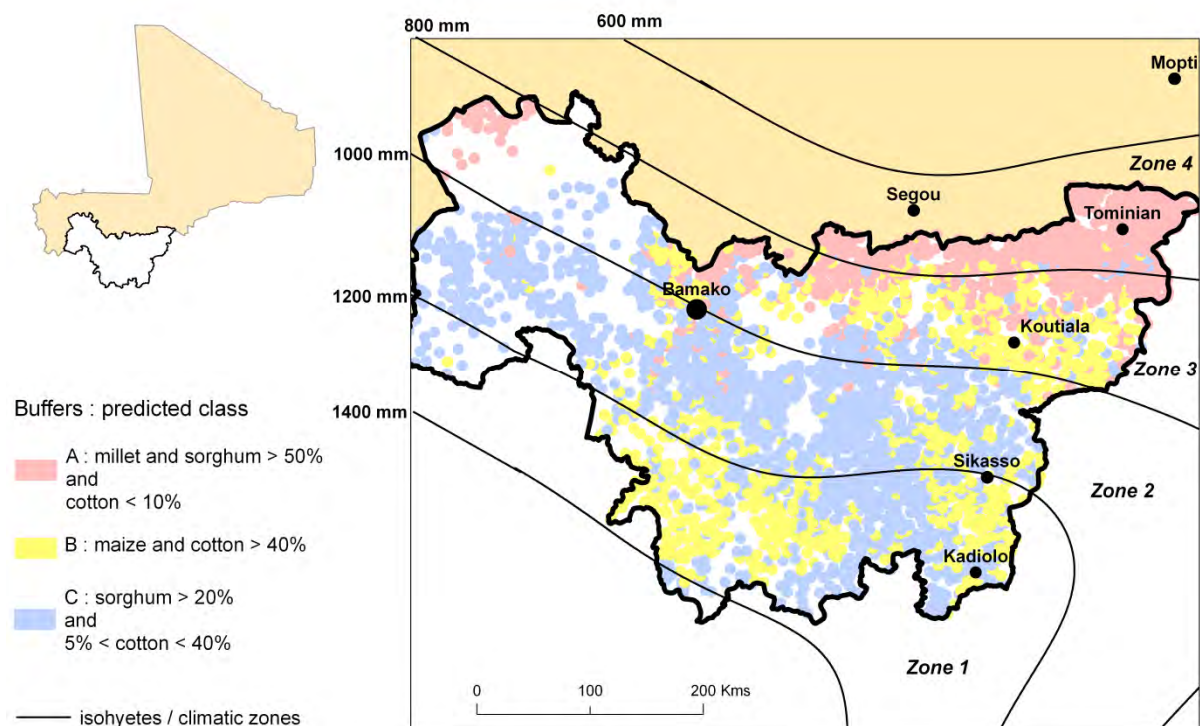


Figure 27 : Cartographie des systèmes agricoles du Sud Mali à l'échelle villageoise (buffers de 5 km autour de chaque village, BD villageoise IER).

3. Discussions

Nous avons développé une méthode permettant la cartographie des principaux systèmes agricoles du Mali à partir d'imagerie MODIS à 250 m de résolution spatiale et de données de terrain. La cartographie des types de systèmes agricoles a nécessité un premier travail de typologie effectué à partir de l'analyse d'une BD d'enquêtes de terrain de l'IER sur 100 villages. Trois types de systèmes agricoles ont pu être déterminés à l'échelle du village : à dominante de céréales (mil, sorgho), à dominante de cultures intensives telles que le maïs et le coton, et mélange de sorgho et de coton. La méthode de caractérisation de ces surfaces cultivées a ensuite déterminé des liens entre le type de système agricole et les indicateurs spectraux, spatiaux et temporels de télédétection, à l'échelle du village. Le modèle Random Forest, à partir des données d'apprentissage sur les 100 villages, a fourni une bonne prédiction du type de systèmes agricoles avec une précision globale estimée de 60% sur 4000 villages du Sud Mali. Cette relativement bonne précision confirme notre hypothèse de départ, à savoir qu'un système agricole se caractérise à la fois par des variables spectrales, spatiales et temporelles spécifiques, et qu'il est possible de les distinguer à partir d'imagerie MODIS à moyenne résolution spatiale. Nous pensons que les 40% d'erreurs sont dus : i) au faible jeu de données de 100 villages seulement, ii) aux enquêtes IER/CMDT qui ont été menées dans d'autres objectifs que le

notre, iii) aux données des enquêtes qui datent de plus de 5 ans et iv) à la typologie de « système agricole majoritaire » par village.

Nous avons également noté d'importantes confusions entre les classes B et C. Ceci peut s'expliquer par les ressemblances de ces deux classes : elles contiennent toutes deux du maïs et du coton, dans différentes proportions mais avec des pratiques agricoles similaires, de par la fertilisation favorisée par la culture du coton. De plus, les agriculteurs maliens faisant face actuellement à une crise du coton, les pratiques agricoles deviennent moins intensives et les villages évoluent souvent de la classe B vers la classe C (Belieres *et al.*, 2009).

Ces résultats confirment la pertinence de l'utilisation des Random Forest, car, en plus de sa facilité d'utilisation (deux paramètres à fixer), il accepte des échantillons de petite taille et autorise les variables corrélées (la plupart des métriques utilisées dans cette étude étaient corrélées).

Dans un second temps, nous avons étudié la contribution des grands groupes de variables utilisés pour la classification des systèmes agricoles, à savoir les indicateurs de NDVI, de texture, de fragmentation et de la phénologie. La contribution des indicateurs dans le classement des systèmes agricoles peut s'expliquer par la nature des classes définies. La classe A représente la classe de cultures vivrières, avec une saison des cultures courte et un choix limité d'espèces (mil ou sorgho). Cette classe a une phénologie plus distincte que les classes B et C réunies. C'est pour cela que les variables phénologiques ont aidé à la distinction des villages de classe A. Prenons également l'exemple de la texture, qui semble aider à la distinction des villages de classes A et C. L'hypothèse avancée est que la classe A et la classe C présentent une couverture du sol hétérogène, avec du sol nu entre les champs pour la classe A (zone sèche) et de la végétation naturelle pour la classe C, et donc des ruptures abruptes dans la texture des paysages agricoles, contrairement à la classe B, qui contient des cultures intensifiées et donc une couverture supposée plus homogène. Ces résultats sur la texture à moyenne résolution spatiale sont à valider.

Pour conclure, la carte des principaux systèmes agricoles (cultures vivrières, cultures intensives et agriculture mixte) produite dans cette étude est importante car elle est une source originale et indépendante d'information pour le Sud du Mali. Parce que les facteurs locaux comme le climat, le sol, la disponibilité en eau, l'accès aux marchés et aux engrais, influence les systèmes agricoles, la cartographie de ces systèmes peut aider à déterminer quelle région et quelle population peuvent être vulnérables à différents risques. Il est alors essentiel de localiser ces systèmes afin de procéder à une analyse précise de l'état de la sécurité alimentaire. En termes de prévisions de rendement, des cartes de systèmes agricoles peuvent être utiles dans la modélisation agro-météorologique, en réduisant le nombre de conditions initiales (les scénarios d'espèces cultivées et des pratiques d'intensification sont fixés). L'identification des principaux systèmes agricoles peut ainsi permettre d'assurer le forçage précis des paramètres de cultures concernées dans un modèle de rendement (Doraiswamy *et al.*, 2005).

POINTS-CLÉS : CARACTERISATION DES SYSTÈMES AGRICOLES

- Un système agricole est caractérisé par i) une signature spectrale spécifique, ii) une signature spatiale spécifique, avec une organisation des éléments du paysage, et iii) une signature temporelle spécifique, représentant les variations saisonnières.
- Une typologie des systèmes agricoles a été construite à partir d'une base de données de l'IER. Les trois classes sont : la classe de cultures vivrières (mil-sorgho), la classe de cultures intensives (maïs-coton) et la classe mixte (sorgho, coton).
- Un modèle de classification Random Forest a été développé à l'échelle du village, à partir de données sur le principal système agricole villageois (typologie à 3 classes) ainsi que 26 variables de télédétection (spectrales, texturales et phénologiques) et 4 variables socio-économiques calculées sur le masque de culture.
- Pour la prédiction de la classe de systèmes agricoles prédominant par village au Mali, la précision du modèle est de 60%.
- L'importance des variables est différente pour chaque classe. En considérant les grands groupes de variables séparément (NDVI, texture, phénologie), la texture est apparue être la variable la plus discriminante.
- Le couplage des cartes de systèmes agricoles, qui fournissent une information sur les pratiques, avec un modèle de plante et des données climatiques pourraient assurer un suivi annuel des cultures de meilleure précision.

Chapitre 5. SUIVI DE LA CROISSANCE DES CULTURES

Nous avons proposé une typologie des systèmes agricoles au Mali. L'étape suivante conduisant à une meilleure estimation du rendement consiste à suivre la phénologie de la culture au cours de la saison de croissance pour détecter d'éventuelles anomalies ou forcer des modèles de rendement par de l'information fiable et continue dans l'espace. L'utilisation d'indicateurs phénologiques issus d'imagerie satellitaire pour caractériser la végétation constitue une approche déjà bien répandue depuis les années 90 (Reed *et al.*, 1994; White *et al.*, 1997). Ces indicateurs phénologiques renseignent quant aux dates de début de période de croissance, de maximum, de sénescence et de minimum végétal. Ainsi, avec la possibilité de détecter de façon objective, homogène dans le temps et à l'échelle globale la phénologie de la végétation, les séries d'images satellite à moyenne résolution permettent aujourd'hui d'étudier les motifs phénologiques en relation avec la variabilité climatique et l'action de l'homme (De Beurs and Henebry, 2005; Zhang *et al.*, 2009). Une revue des différentes méthodes pour le calcul d'indicateurs phénologiques sur les séries temporelles d'images peut être trouvée dans White *et al.* (2009) et Schwartz and Hanes (2010). La plupart des auteurs calculent des métriques phénologiques à partir de l'indice de végétation NDVI (Funk and Budde, 2009; Rojas *et al.*, 2011), qui sont également souvent couplées avec des données de précipitations (Balme *et al.*, 2005; Brown and de Beurs, 2008; Heumann *et al.*, 2007; Wessels *et al.*, 2011). Cependant, ces données climatiques sont soit interpolées à partir des stations synoptiques (peu nombreuses dans la plupart des pays du Sud et dont les données sont difficilement accessibles), soit estimées à partir de données satellitaires, mais avec une résolution spatiale faible et une précision qui ne répond pas toujours aux exigences des modèles de cultures. Ainsi, estimer directement les indicateurs phénologiques à partir des séries temporelles d'indices de végétation de résolution spatiale moyenne améliorerait d'une part la résolution spatiale des indicateurs phénologiques et d'autre part serait plus proche de la réalité terrain en mesurant directement la phénologie de la végétation qui dépend non seulement de la pluie mais aussi de l'occupation du sol.

Le seul produit global actuellement disponible qui inclut des variables phénologiques est le produit MODIS *Land Cover Dynamics* (MCD12Q2) et il n'est actuellement pas utilisé dans les systèmes d'alerte précoce.

En considérant que les données de terrain sont difficilement accessibles au Mali, nous avons comparé ces variables phénologiques observées par télédétection à des variables simulées par le modèle de plante SARRA-H (Dingkuhn *et al.*, 2003), à partir de profils de LAI. Ce modèle a déjà été entraîné sur différentes espèces et variétés d'Afrique de l'Ouest, y compris les variétés de mil, de sorgho, et de maïs que l'on retrouve au Mali. Il est actuellement utilisé par Agrhymet pour l'estimation des rendements des cultures.

Le but de cette étude était donc d'analyser le contenu en information du nouveau produit MODIS MCD12Q2 sur les paysages agricoles fragmentés d'Afrique de l'Ouest. Pour cela, nous avons procédé en deux temps :

- nous avons d'abord évalué le produit MCD12Q2 (précision, validité) en rapport avec nos connaissances sur les variabilités climatiques et sur la localisation des systèmes agricoles au Mali.
- nous avons ensuite utilisé la modélisation qui permet de formaliser ces connaissances pour simuler la croissance des cultures dans les différentes situations rencontrées, et comparé les variables phénologiques issues du produit satellitaire MCD12Q2 aux dates de transition phénologique simulées par le modèle de plante SARRA-H.

Cette étude sera prochainement soumise dans une revue de rang A (Annexe 3) :

Vintrou, E., Bégué, A., Baron, C., Lo Seen, D. (2012). Analysing MODIS phenometrics information content in heterogeneous agricultural landscapes of West Africa. En préparation.

1. Méthode

La méthode utilisée pour évaluer l'information contenue dans le produit MCD12Q2 est résumée dans la Figure 28.

1.1. Evaluation du produit MCD12Q2

Ce produit phénologique annuel est détaillé en Chapitre 2, Section 1.4.3. Rappelons que l'algorithme de Zhang (2003) permet de calculer sept variables phénologiques à partir d'indicateurs EVI issus des séries temporelles MODIS. Nous avons utilisé les 4 premières, à savoir : (1) le démarrage de la saison de croissance de la plante ou *Start-of-Season* (SOS), (2) la date de début du maximum de la croissance, ou *Start-of-Maximum* (SMAX), (3) la date de fin de maximum de la croissance, ou *End-of-Maximum* (EMAX), et (4) la date de fin de la saison, ou *End-of-Season* (EOS).

Pour analyser la phénologie des cultures et conserver uniquement les pixels cultivés « purs », nous avons tout d'abord utilisé un filtre sur le masque des cultures, pour ne conserver que les pixels entourés de plus de 7 pixels cultivés sur le produit MCD12Q2, puis appliqué ce masque des cultures filtré sur le produit MCD12Q2. Nous avons ensuite exploré le produit MCD12Q2 en enlevant les pixels qui ne possédaient pas un cycle complet (absence d'un ou plusieurs indicateurs de la phénologie en 2007). Nous avons étudié la validité de ce produit sur l'année 2007.

1.2. Extraction des indicateurs phénologiques observés (MODIS MCD12Q2)

Les quatre indicateurs de la phénologie ont été extraits sur le masque de culture filtré et sur chaque station synoptique représentée par un polygone de taille 10 x10 km². Nous avons extrait la moyenne de chacune de ces variables, chaque année entre 2001 et 2008, et comparé les dates de transition phénologique obtenues avec les profils de synthèses temporelles d'EVI (produit MODIS) à 8 jours sur chaque station (Chapitre 2, section 1.1).

1.3. Extraction des indicateurs de la phénologie simulés (LAI)

Les profils LAI ont été simulés à l'aide du modèle de plante SARRA-H. Le modèle de culture SARRA-HV32 développé par le CIRAD permet de simuler les dynamiques des biomasses des céréales et plus particulièrement de plusieurs variétés de mil, maïs et sorgho cultivées en milieu paysan. SARRA-H est un modèle de culture « robuste » qui fonctionne à l'échelle parcellaire avec un pas de temps journalier (Dingkuhn *et al.*, 2003). Les pratiques agricoles les plus courantes, choix des variétés et notamment stratégies de semis, peuvent être simulées dans le modèle, ce qui permet de calculer les dates de semis réussis en fonction de la mise en place de la saison des pluies. Ce modèle est notamment utilisé par le Centre Régional Agrhymet dans la mise en oeuvre d'un système d'alerte précoce pour les pays du CILSS.

Les simulations de LAI ont été faites sur les stations synoptiques du Mali pour les principales cultures céréalières. Les scénarios de simulation ont été faits par expertise selon deux critères importants de gestion des cultures au regard des pratiques paysannes : l'adaptation des espèces et variétés, et des pratiques de semis en fonction des zones climatiques (Vintrou *et al.*, 2011). Les variétés utilisées par station sont résumées dans le Tableau 12. Les données climatiques d'entrée du modèle sont les données de pluie et de température mesurées sur les stations (en partenariat avec Agrhymet).

STATIONS	CROPS
Bamako	Sorghum guinea, Millet kenikeba, Millet choho, Maize
Bougouni	Sorghum guinea, Maize
Kayes	Sorghum caudatum, Millet kenikeba, Maize
Kita	Sorghum guinea, Millet kenikeba, Millet choho, Maize
Koutiala	Sorghum guinea, Millet kenikeba, Maize
Nara	Sorghum caudatum, Millet kenikeba, Millet souna
San	Sorghum caudatum, Millet souna
Segou	Sorghum guinea, Millet kenikeba, Millet choho
Sikasso	Sorghum guinea, Millet kenikeba, Millet choho, Maize

Tableau 12 : Choix des variétés à simuler par le modèle SARRA-H par station synoptique.

Les variables phénologiques des profils de LAI ont ensuite été calculées selon la méthode de Zhang *et al.* (2003) pour être comparables au produit phénologique MODIS, et moyennées par station synoptique.

1.4. Comparaison des deux types d'indicateurs phénologiques

Les indicateurs phénologiques issus de MODIS et du LAI ont été comparés i) spatialement, sur 8 stations en 2007, et ii) temporellement, sur 2 stations (Segou et Sikasso), de 2001 à 2008, correspondant aux stations pour lesquelles nous avons des données météorologiques et climatiques sur 10 ans (Chapitre 2, section 3.3).

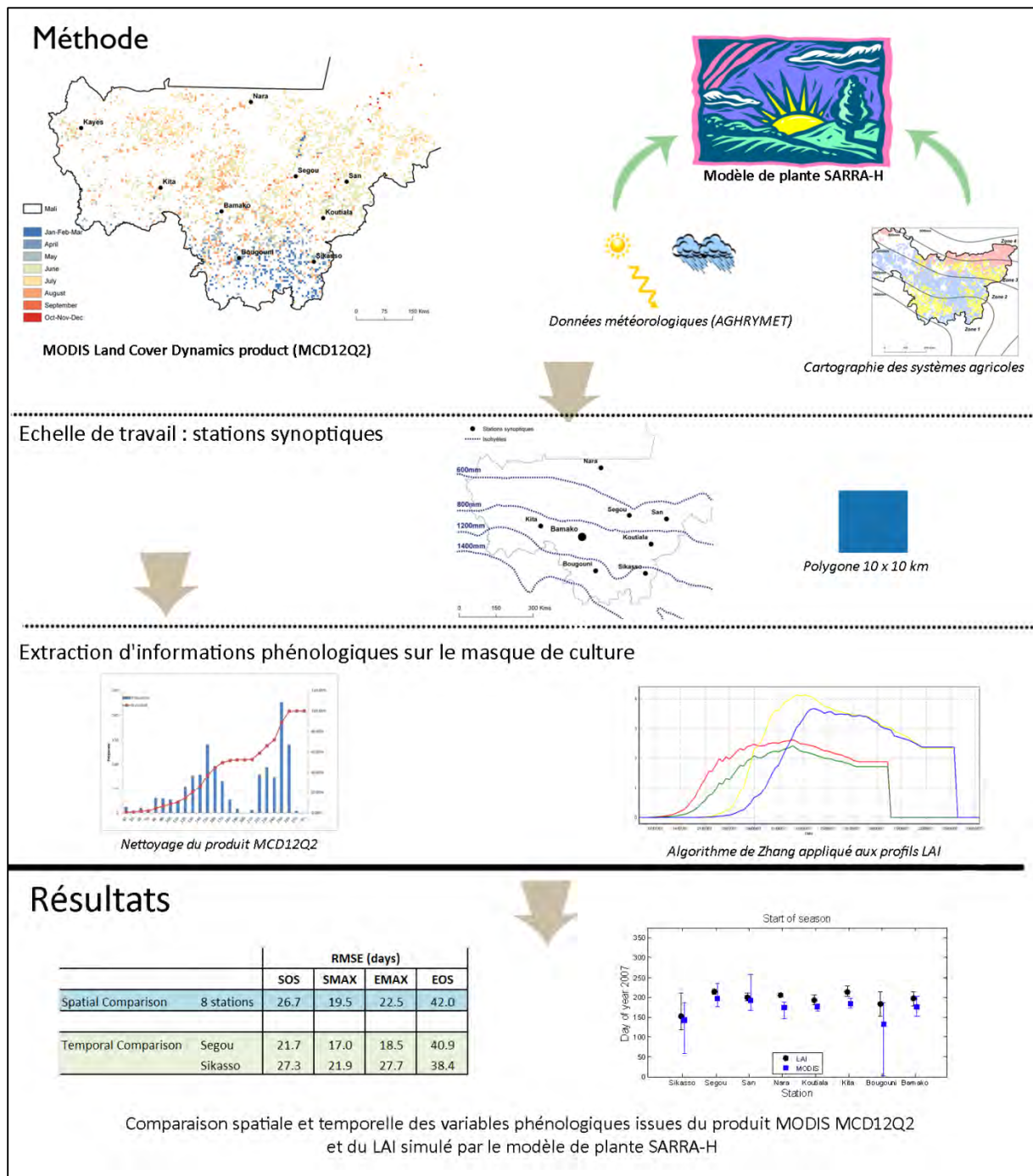


Figure 28 : Méthodologie et résultats pour le suivi de la croissance des cultures.

2. Résultats

2.1. Evaluation du produit MODIS MCD12Q2

A l'échelle nationale, 69% de pixels cultivés ont un cycle complet en 2007. Seuls 5% des pixels cultivés présentent 2 cycles complets sur la même année, ce qui concorde avec les systèmes de culture dans leur majorité pluviaux au Mali, qui présentent donc 2 cycles au cours d'une même saison uniquement en condition d'irrigation.

La Figure 29 présente le démarrage de la saison (SOS) dans chaque station synoptique en 2007. Nous pouvons noter un gradient Nord-Sud du démarrage de la saison de croissance des cultures. En effet, la saison commence plus tard dans le Nord (San, Segou, Nara), en relation avec le régime pluviométrique. Nous observons également de grandes différences de démarrage de croissance des cultures au sein même d'une station.

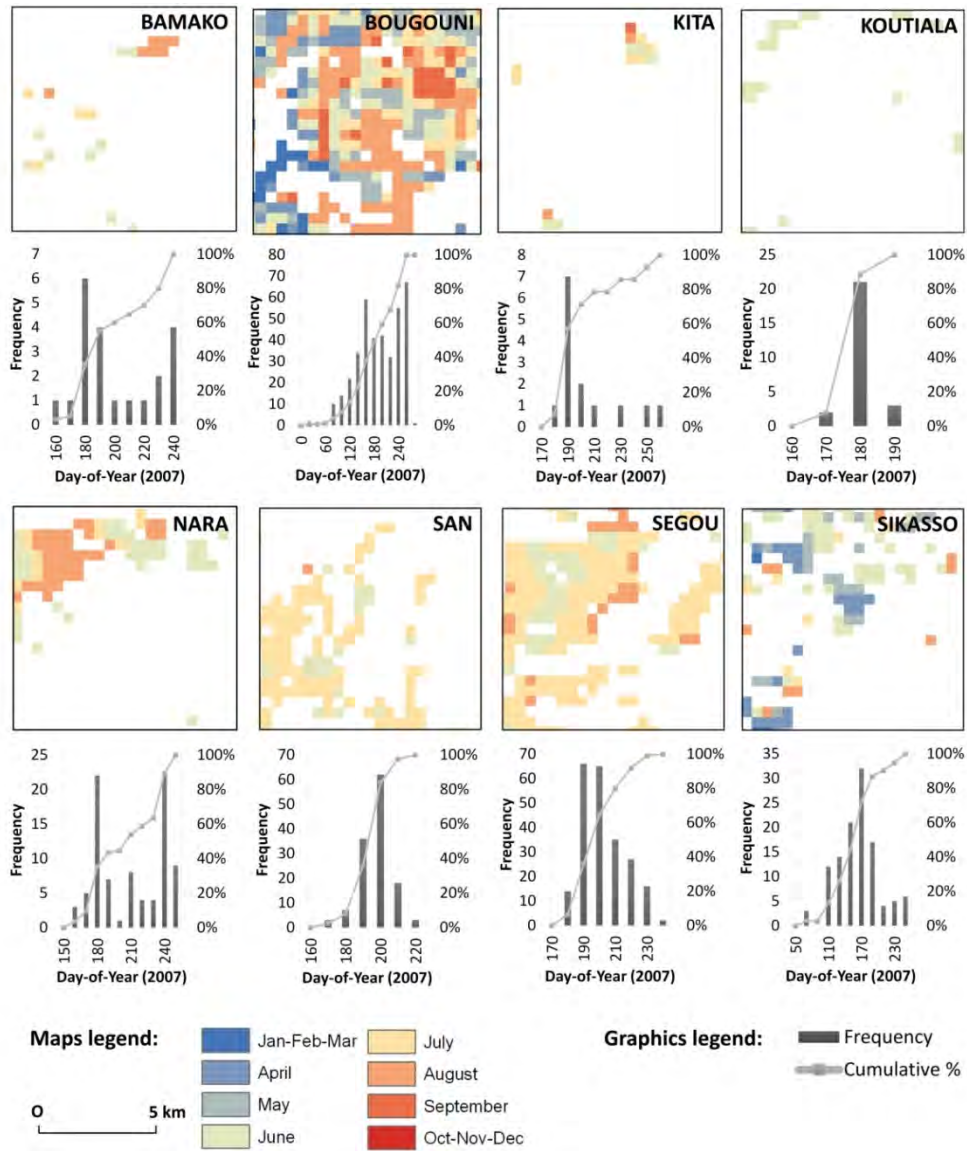


Figure 29 : Cartographie de l'indicateur « Start-of-Season » dans les 8 stations synoptiques en 2007 et histogrammes des valeurs.

Pour comprendre ce phénomène, nous avons analysé les histogrammes des valeurs des indicateurs phénologiques dans chaque station et mis en évidence quelques incohérences : pour les stations de Bougouni, Sikasso, Nara, Bamako et Kita, nous avons deux pics à haute fréquence pour le démarrage de la saison (Figure 29). Nous avons émis l'hypothèse le deuxième pic, très tardif par rapport à la saison des pluies, était dû à un problème d'ajustement de l'algorithme de Zhang. Pour

vérifier cela, nous avons pris l'exemple de la station de Bougouni et de 3 pixels aux valeurs de SOS très différentes (Figure 30). Ces pixels ont tous des profils d'EVI avec des valeurs manquantes aux synthèses 27 et 28. Cependant :

- le profil du premier pixel (en bleu) possède également deux valeurs manquantes d'EVI (synthèses 29 et 30). Malgré cela, l'algorithme de Zhang fonctionne.
- le profil du second pixel (en rouge) possède trois valeurs manquantes d'EVI (synthèses 29, 30 et 31). Dans ce cas, l'algorithme de Zhang détecte le démarrage de la saison des cultures à la synthèse 31, qui est en fait son maximum.
- enfin le profil du troisième pixel (en vert) possède une pente plus importante, dès la synthèse 10. Dans ce cas, l'algorithme détecte un démarrage beaucoup plus tôt, malgré les trois valeurs d'EVI manquantes (synthèses 29, 30 et 31).

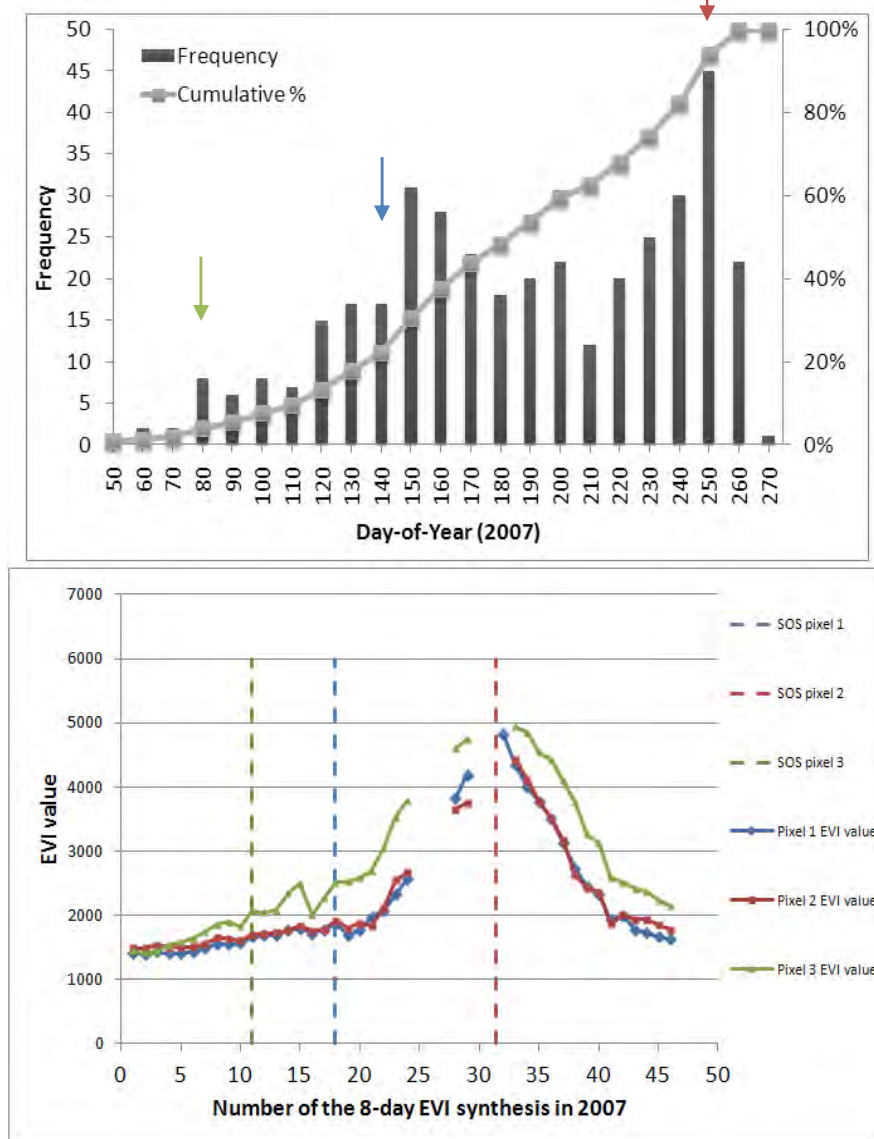


Figure 30 : Démarrage de la saison des cultures pour la station de Bougouni : histogrammes des valeurs et profils d'EVI sur 3 pixels.

Nous avons supprimé les pixels ayant un démarrage de végétation dans le « deuxième pic » pour les stations concernées (Bamako, Bougouni, Kita, Nara, Sikasso), en seuillant les dates de SOS (valeur seuil = valeur avec fréquence minimale entre les deux pics).

2.2. Les valeurs d'indicateurs phénologiques sur les stations synoptiques

Le Tableau 13 présente les valeurs des indicateurs phénologiques sur les 8 stations synoptiques en 2007, avec et sans corrections des histogrammes. Les corrections des histogrammes entraînent un démarrage de la saison 15 jours à un mois plus tôt que sur le produit sans correction. Le gradient nord-sud de l'indicateur de début de saison de croissance est confirmé, avec un démarrage en mai à Bougouni et Sikasso, en juin à Koutiala et Bamako, et en juillet à San, Ségou et Kita. Le début du maximum de la saison est concentré dans la deuxième moitié du mois d'août alors que la fin de maximum de croissance est concentrée en Septembre et la fin de la saison en Décembre. Concernant les écarts-types des valeurs des indicateurs de la phénologie à l'intérieur des stations, nous remarquons une large gamme de valeurs dans le Sud (Bougouni et Sikasso), avec environ 30 jours d'écart, tandis que les valeurs de début de saison sont plus concentrées autour de la valeur moyenne dans le Nord (15 jours pour San, 8 jours pour Nara par exemple).

Station	Histograms corrections	SOS (SD in days)	SMAX (SD in days)	EMAX (SD in days)	EOS (SD in days)
Bamako	Before	July-17 (28.7)	Aug.-28 (8.9)	Sept.16 (7.7)	Nov.25 (11.9)
	After	June-26 (12.9)	Aug.25 (8.7)	Sept.15 (7.9)	Nov.29 (12.8)
Bougouni	Before	July-1 (58.7)	Aug.26 (25.7)	Sept.19 (19.5)	Nov.26 (32.4)
	After	May-12 (31.4)	Aug.14 (30.6)	Sept.11 (23.8)	Nov.29 (43.6)
Kita	Before	July-18 (29.4)	Sept.3 (15.3)	Sept.26 (9.8)	Dec.12 (9.0)
	After	July-2 (6.8)	Aug.28 (13.1)	Sept.23 (9.8)	Dec.13 (8.9)
Koutiala	None	June-26 (4.7)	Aug.11 (7.6)	Sept.10 (4.1)	Nov.22 (8.9)
Nara	Before	July-23 (32.7)	Aug.24 (18.8)	Sept.14 (15.6)	Nov.13 (29.2)
	After	June-23 (8.3)	Aug.12 (18.8)	Sept.7 (18.5)	Nov.7 (38.0)
San	None	July-12 (10.2)	Sept.7 (8.6)	Sept.26 (8.5)	Dec.7 (17.3)
Segou	None	July-16 (15.7)	Sept.5 (8.0)	Sept.28 (6.8)	Dec.12 (12.4)
Sikasso	Before	June-5 (43.4)	Aug.19 (22.4)	Sept.22 (19.4)	Dec.11 (31.8)
	After	May-23 (31.8)	Aug.15 (22.8)	Sept.21 (19.7)	Dec.10 (34.0)

Tableau 13 : Valeurs des quatre indicateurs phénologiques sur les 8 stations synoptiques en 2007: moyenne et écart type avant et après corrections des histogrammes.

2.3. Les indicateurs phénologiques simulés (LAI)

Deux exemples de profils LAI sont donnés en Figure 31. A partir de ces profils et grâce à l'algorithme de Zhang, nous avons calculé les quatre indicateurs de la phénologie (SOS, SMAX,

EMAX et EOS) pour chaque profil, que nous avons moyennés par station.

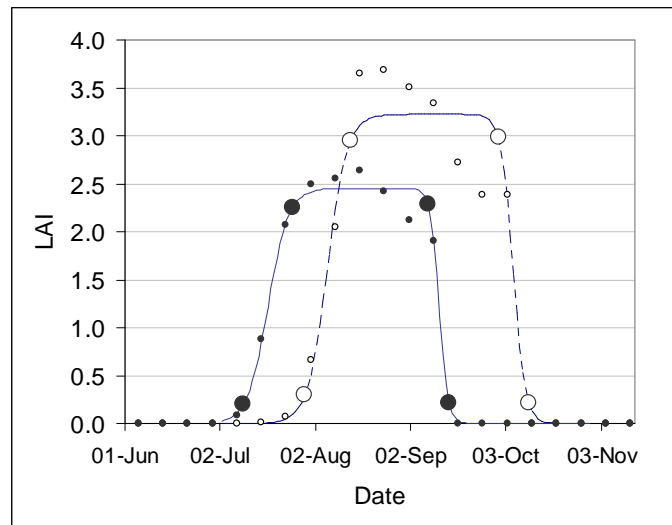


Figure 31 : Calcul du démarrage de saison des cultures (SOS), du début de maximum (SMAX), de la fin du maximum (EOS) et de la fin de saison (EOS) sur les profils LAI à partir de l'algorithme de Zhang. a) profil de sorgho caudatum fertilisé ; b) profil de mil souana non fertilisé en 2007.

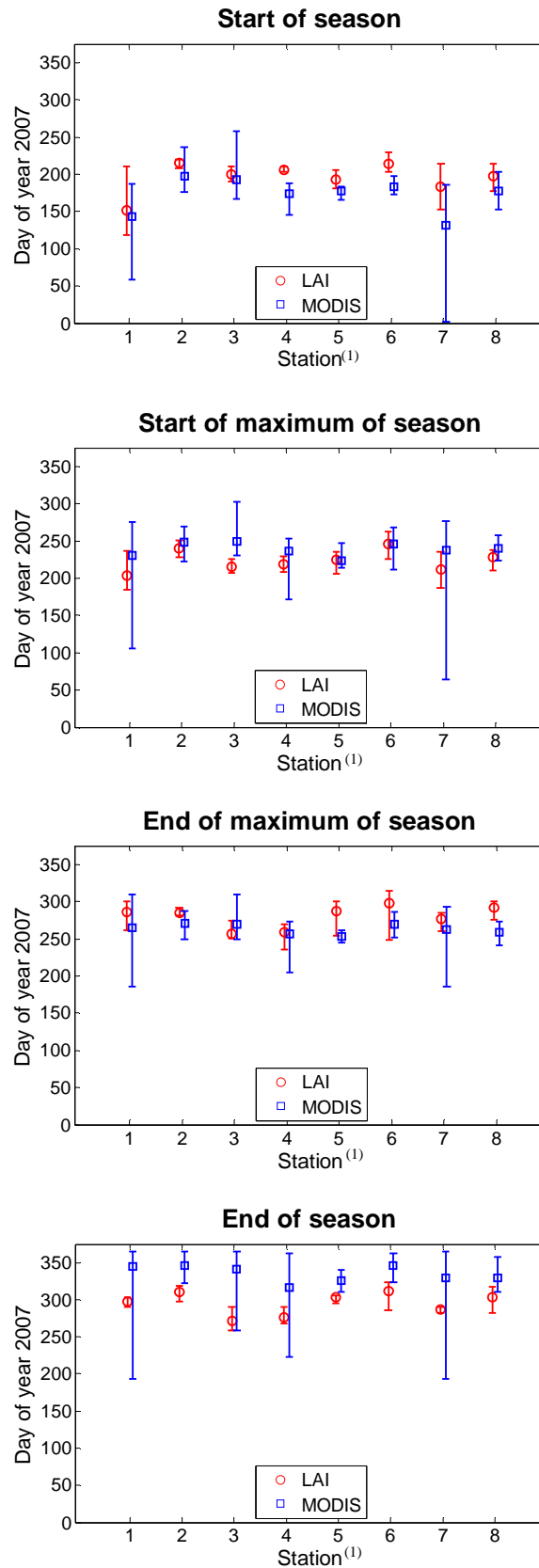
2.4. Comparaison en 2007 sur 8 stations synoptiques

En comparant les indicateurs phénologiques moyens issus du produit MCD12Q2 et ceux issus des profils LAI simulés par le modèle de plante SARRA-H, par station synoptique, en 2007, nous avons mis en évidence un écart de 20 jours en moyenne pour le démarrage, le début de maximum et la fin de maximum, et de 40 jours pour la fin de saison. La Figure 32 montre que le démarrage de la saison (SOS) et la fin du maximum de la saison (EMAX) sont toujours détectés plus tôt par MODIS que par le LAI. Au contraire, le début du maximum de la saison (SMAX) et la fin de la saison de croissance (EOS) sont plus tôt dans l'année d'après le LAI que d'après MODIS.

Nous notons également que la fin de saison présente de grandes disparités entre MODIS et le LAI, mais que pour les trois autres indicateurs phénologiques, les moyennes sont globalement proches, avec des fourchettes de valeurs plus larges pour MODIS que pour le LAI.

2.5. Comparaison de 2001 à 2008 sur 2 stations

La comparaison temporelle entre les indicateurs phénologiques issus de MODIS MCD12Q2 et ceux issus du modèle de plante a donné les mêmes résultats que sur l'année 2007 : un démarrage de saison et une fin de maximum de saison détectés 20 jours plus tôt par MODIS, et un début de maximum et une fin de saison de croissance détectés plutôt tôt par le LAI (20 et 40 jours respectivement).



⁽¹⁾ Station : 1. Bamako, 2. Bougouni, 3. Kita, 4. Koutiala, 5. Nara, 6. San, 7. Segou, 8. Sikasso

Figure 32 : Minimum, maximum et moyenne des 4 indicateurs phénologiques par station en 2007. Les lignes bleues représentent les indicateurs phénologiques issus de MODIS, les lignes rouges ceux issus du LAI.

3. Discussions

Pour le suivi de la phénologie des cultures au cours de la saison de croissance, nous avons proposé d'explorer les potentialités des variables phénologiques issues du produit MODIS MCD12Q2. Au Mali, 69% des pixels cultivés (à partir du masque de culture préalablement bâti) ont un cycle complet, c'est-à-dire qu'ils possèdent à la fois un démarrage, un début de maximum, une fin de maximum et une fin de saison de croissance en 2007. A l'échelle des stations synoptiques, nous avons vérifié l'adéquation des indicateurs phénologiques aux profils d'EVI en 2007, après avoir supprimé les pixels aux cycles incomplets. Une incohérence a été mise en évidence : certaines stations présentaient deux pics de valeurs distincts pour le démarrage de la saison. Il a été montré dans cette étude que cette incohérence est due à la méthode utilisée par l'algorithme pour détecter les taux de changement de courbure (Zhang *et al.*, 2003). En effet, les séries temporelles d'EVI sont lissées avec une valeur médiane calculée localement pour chaque point, en utilisant une fenêtre coulissante de 3 dates. C'est pourquoi cet algorithme ne fonctionne pas lorsqu'il y a plus de deux valeurs manquantes consécutives, ce qui est souvent le cas dans les pays tropicaux d'Afrique de l'Ouest. Zhang *et al.* (2009) ont estimé que la probabilité que l'erreur absolue dans l'estimation des indicateurs phénologiques soit supérieure à 5 jours est de 20% lorsqu'il y a deux valeurs consécutives manquantes dans le profil temporel d'EVI. Nos résultats confortent également ceux de Ganguly *et al.* (2010) qui ont montré que le Sahel était la zone qui comportait le plus de données manquantes sur le produit MODIS MCD12Q2, avec une plus grande incertitude pour l'indicateur phénologique de fin de saison que celui de début de saison.

Nous avons ensuite comparé ces variables phénologiques issues d'imagerie satellitaire avec les dates de transition phénologique simulées par le modèle de plante SARRA-H, qui sont des sorties de variables « idéalisées » mais qui rendent bien compte des connaissances que nous avons du terrain. A partir de la cartographie des systèmes agricoles (Vintrou *et al.*, 2011), nous avons bâti différents scénarios et des profils de LAI ont été simulés par le modèle de plante SARRA-H. L'observation la plus importante sur ces profils de LAI réside dans la fin de saison, qui finit de manière très abrupte, correspondant à la date de maturité physiologique de la culture. Cependant, les agriculteurs ont différentes pratiques et stratégies de récoltes : i) les récoltes avant maturité complète, soit en période de soudure, soit si les conditions sont trop pluvieuses, pour éviter les verses et moisissures de fin de saison et ii) les récoltes après maturité : certaines cultures restent sur pied alors que l'on ramasse leurs épis, ou sont récoltées plus tard pour améliorer le séchage des grains sur pied avant le stockage, ou encore par manque de main d'œuvre. C'est pourquoi cette date de fin de saison est difficilement comparable avec le taux de couverture du sol.

La comparaison en 2007 a montré un écart de 20 jours entre le démarrage de saison MODIS et celui du LAI, avec un démarrage plus précoce d'après l'imagerie satellitaire. L'hypothèse avancée est que les résolutions spatiales du produit MODIS MCD12Q2 et de notre masque de culture (500 m et 250 m) ne garantissent pas la surveillance exclusive des cultures. En effet, la couverture au sol est rarement uniforme sur 25 hectares au Mali, avec un mélange très fréquent de végétation naturelle et de cultures. Or la végétation naturelle verdit dès la première pluie, alors que l'agriculteur attend l'installation des pluies avant de semer. C'est pourquoi, en tenant compte de ce mélange (ne serait-ce

que par la présence d'arbres dans les champs), le démarrage est détecté plus tôt chez MODIS puisqu'il tient compte en partie de la végétation naturelle environnante. En ce qui concerne le démarrage du maximum de la saison, c'est le LAI qui présentait environ 20 jours d'avance sur le même indicateur phénologique MODIS. Nous supposons que les simulations du modèle SARRA-H suggèrent que le patch de cultures est homogène, avec une croissance simultanée de toutes les plantes, alors que notre masque de culture MODIS englobe certainement de la végétation naturelle et plusieurs types de culture couvrant le sol progressivement au sein d'un même pixel, et donc un maximum atteint plus tardivement. La comparaison de la fin du maximum n'a pas permis d'établir de liens clairs entre les deux types d'indicateurs phénologiques, même si les moyennes sont très proches. La fin de saison présentait par contre de grandes disparités.

Cet intervalle de 20 jours entre les indicateurs de la phénologie MODIS et SARRA-H peut être considéré sous forme d'un biais systématique puisqu'il se reproduit sur l'ensemble des stations étudiées et sur les 8 années considérées, à l'exception notable de la fin de saison des cultures. Nous pensons donc que le produit MCD12Q2 traduit bien les connaissances que nous avons des différences spatiales et temporelles des phénologies de cultures au Mali.

Ce produit nous semble apporter une information supplémentaire sur l'évolution de la croissance des cultures, mais il n'est pourtant pas utilisable directement par les systèmes d'alerte précoce, car délivré généralement l'année suivante. Cependant, l'algorithme pourrait être utilisé sur les séries temporelles d'EVI à 8 jours, et nous pensons qu'en milieu de saison de croissance, il serait possible de détecter les deux premiers changements de courbure, c'est-à-dire, le début de la saison et le début du maximum de la saison. Ces deux indicateurs de la phénologie sont en effet les plus importants pour juger d'un démarrage trop tardif et de possibles réductions de rendements.

Ces indicateurs phénologiques peuvent représenter un enrichissement possible (valeurs de forçage) quant aux simulations du modèle SARRA-H, utilisé notamment par l'Agrhymet pour les calculs de rendements potentiels. Cette information sur la phénologie, continue dans l'espace, pourrait être utilisée en complément des données climatiques agrégées, actuellement utilisées par les modèles.

POINTS CLÉS : SUIVI DE LA CROISSANCE DES CULTURES

- Nous avons analysé le contenu en information du nouveau produit MODIS MCD12Q2 sur les paysages agricoles fragmentés d'Afrique de l'Ouest. Ce produit donne une information sur le démarrage, le début du maximum, la fin du maximum et la fin de la saison de croissance, à partir de séries temporelles d'EVI MODIS à 500 m.
- Une incohérence a été mise en évidence au niveau du démarrage de la saison, lorsque plus de deux valeurs manquantes d'EVI sont observées, ce qui est fréquent dans les pays d'Afrique de l'Ouest.
- Nous avons formalisé nos connaissances des systèmes agricoles maliens par le modèle SARRA-H, et comparé les variables phénologiques issues d'imagerie satellitaire avec les dates de transition phénologique simulées par le modèle de plante. Les variables phénologiques du produit MODIS MCD12Q2 correspondent à ce que l'on connaît des cultures présentes dans la zone.
- L'élaboration de ces indicateurs de la phénologie à l'aide de l'outil satellitaire nécessite des séries temporelles régulières sur plusieurs années, et l'algorithme de Zhang (2003) pourrait être utilisé au cours de la saison de croissance pour détecter d'éventuels démarrages tardifs de la végétation ou un maximum de végétation faible conduisant à des rendements réduits.
- Le couplage des modèles d'estimation du rendement avec nos cartes de systèmes agricoles, qui fournissent une information sur les pratiques, et une intégration de l'information satellitaire sur la phénologie pourraient assurer un suivi annuel des cultures de meilleure précision qu'avec l'utilisation de données climatiques seules, comme c'est le cas actuellement dans les systèmes d'alerte précoce.

CONCLUSIONS ET PERSPECTIVES

Les systèmes d'alerte précoce régissent la surveillance de la sécurité alimentaire. L'Afrique de l'Ouest est une région particulièrement vulnérable, qui nécessite la mise en place d'un suivi des cultures précis et opérationnel à l'échelle nationale. Pour cela, les méthodes actuelles déployées dans cette région peinent à aboutir en raison d'une agriculture et d'un régime pluviométrique particuliers, ainsi que d'un réseau d'observations au sol de densité et de qualité insuffisantes.

L'objectif général de ces travaux de thèse est d'améliorer l'utilisation de la télédétection pour l'estimation de la production des cultures en Afrique de l'Ouest afin de mieux répondre aux besoins des systèmes d'alerte en sécurité alimentaire.

Pour cela, cette thèse s'est attachée à développer des méthodologies issues du traitement de séries temporelles MODIS de manière à répondre à trois sous-objectifs:

- cartographier le domaine cultivé à l'échelle nationale et qualifier le résultat vis-à-vis des produits globaux existants,
- caractériser les principaux systèmes agricoles du pays,
- suivre la phénologie des cultures à partir du produit MCD12Q2 et de simulations agro-météorologiques.

Ces travaux ont été développés sur le Mali, pays représentatif de la zone ouest-africaine. Malgré un potentiel agro-pastoral important, les ressources productives de ce pays ne sont pas exploitées de manière optimale et l'agriculture reste basée sur un système extensif dont la production reste très dépendante des aléas climatiques.

1. Contributions de la thèse

Les apports de cette thèse sont d'ordres méthodologiques et thématiques. Nous donnerons ici les principales contributions en rapport avec les objectifs formulés.

1.1. Apports méthodologiques

Nos résultats contribuent à montrer le potentiel des séries temporelles à moyenne résolution spatiale pour la cartographie, la caractérisation et le suivi des systèmes agricoles en Afrique de l'Ouest. Les méthodes ont été développées à partir de séries temporelles MODIS (250 m à 500 m de résolution spatiale) acquises sur le Mali, et plus précisément à partir d'indices spectraux, spatiaux, texturaux et temporels dérivés des images, ainsi que de données socio-économiques de terrain. La totalité de l'information potentiellement contenue dans les images MODIS à moyenne résolution spatiale a été utilisée, et contre toute attente, les indicateurs texturaux ont eu toute leur place dans les méthodes développées. En effet, la texture s'est révélée être un facteur explicatif plus important que la

simple information spectrale donnée par le NDVI pour la distinction culture / non culture dans l'approche de fouille de données pour la cartographie des cultures, ainsi que pour la caractérisation des systèmes agricoles.

Ces méthodes « multi-indicateurs » ont été développées à partir d'outils innovants.

Tout d'abord, la fouille de données, qui est une méthode originale d'extraction automatique de motifs séquentiels à partir de données complexes, a montré de bons résultats dans son approche préliminaire. Elle a constitué une alternative pertinente pour définir un mécanisme d'apprentissage sans à priori, basé sur des données multi-sources, et a également mis en évidence des critères d'implication qui auraient difficilement pu être identifiés autrement (i.e. la pertinence de l'utilisation de la texture à moyenne résolution spatiale pour la distinction culture / non culture).

Ensuite, l'utilisation de l'algorithme de classification Random Forest est novatrice, dans le sens où aucune étude, à notre connaissance, n'utilise cet outil pour la caractérisation des systèmes agricoles à l'échelle globale. Cet algorithme offre pourtant de nombreux avantages qui rendent son utilisation pertinente : la capacité à sélectionner automatiquement les variables discriminantes dans un fichier de données contenant un très grand nombre de variables potentiellement intéressantes, l'agrégation des performances (bagging) qui utilise l'aléatoire pour générer de la diversité dans les ensembles d'arbres, et enfin le fait de ne pas nécessiter de jeu de données de validation, ce qui permet d'exploiter la totalité du jeu de données comme apprentissage, avantage non négligeable pour des pays où l'on possède peu de données de terrain. Grâce à cet outil, nous avons pu cartographier des données socio-économiques (BD IER – typologie des systèmes agricoles) en les mettant en lien avec des indicateurs de télédétection.

Enfin, la stratification du territoire à l'échelle nationale mise en œuvre est également un des apports méthodologiques de cette thèse. Le fait de stratifier une zone d'étude afin de minimiser la variabilité spatiale des différentes classes d'occupation du sol - et ainsi de limiter la confusion entre les classes - n'est pas une méthode nouvelle. Toutefois, l'originalité de notre approche réside dans la prise en compte de la phénologie particulière de la végétation (en lien avec les zones éco-climatiques), mais également de la structure des paysages (liée à la morpho-pédologie et à l'utilisation du sol) pour le développement d'une méthode de stratification à partir d'indicateurs spectraux, spatiaux, texturaux et temporels d'imagerie MODIS et d'une classification objet. L'approche orientée objet est souvent réservée aux images à haute résolution spatiale. Or tout comme pour la texture, ces travaux de thèse ont démontré qu'une organisation de pixels en un paysage agricole pouvait être captée à l'échelle MODIS. Cette méthode de stratification nous est apparue pertinente, dès lors que la classification culture / non culture par zone a donné de bons résultats, et elle n'est pas utilisée actuellement par les systèmes d'alerte précoce.

1.2. Apports thématiques

Les apports thématiques concernent les systèmes d'alerte précoce pour la sécurité alimentaire. Nous avons proposé :

(i) Une stratification du territoire en zones éco-climatiques homogènes. Cette stratification pourrait être complémentaire de la cartographie des modes de vie du système d'alerte

FEWS NET. Cette dernière est faite uniquement à partir d'enquêtes de terrain sur la production agricole et l'accès au marché, sur différents groupes de richesse dans des zones choisies et n'utilise pas la télédétection.

(ii) Un masque des cultures à l'échelle nationale. Le produit cartographique obtenu présente une meilleure précision que les produits globaux « occupation du sol » existants (70% vs 50% pour la moyenne). Ceci représente une réelle avancée dans la cartographie des surfaces cultivées au Mali, étant donné les grandes divergences entre les différents produits globaux et leurs résolutions spatiales inférieures. Cependant, nous avons démontré qu'une part importante des erreurs de ce masque des cultures (20% à 40%) reste incompressible en raison de la fragmentation du domaine cultivé et de la résolution spatiale du produit utilisé (MODIS MOD13Q1 à 250 m). Ce masque ne peut donc servir uniquement pour un suivi des cultures annuel qualitatif.

(iii) Une carte des systèmes agricoles à 3 classes (cultures vivrières, cultures intensifiées, et classe mixte). Il est essentiel de localiser les différents systèmes agricoles dans le but de mener des analyses significatives sur la sécurité alimentaire. Parce que les facteurs locaux comme le climat, le sol, la disponibilité en eau, l'accès aux marchés et aux engrais influence les systèmes agricoles, la cartographie de ces systèmes peut aider à déterminer quelle région et quelle population peuvent être vulnérables à des risques différents. Par ailleurs, en termes de prévisions de rendement, des cartes de systèmes agricoles peuvent être utilisées dans la modélisation agro-météorologique de l'estimation des rendements, en apportant des informations sur les conditions initiales des simulations (espèces cultivées, stratégies de semis, pratiques d'intensification).

(iv) Un suivi de la croissance des cultures à partir du produit MODIS MCD12Q2. Les variables phénologiques de ce produit sont cohérentes avec ce que l'on connaît des cultures au Mali. L'approche qui consiste à utiliser des variables phénologiques issues d'imagerie satellitaire peut donc constituer une démarche pertinente dans l'optique de suivre la phénologie des zones cultivées au Mali. Nous pensons que l'intégration de ces données dans la méthodologie de l'estimation du rendement pourrait assurer un suivi annuel des cultures de meilleure précision qu'avec l'utilisation de données climatiques seules, comme c'est le cas actuellement dans les systèmes d'alerte précoce. En effet, ces indicateurs de la phénologie estimés par télédétection représentent une information continue dans l'espace, plus proche de la réalité terrain que les données climatiques interpolées à partir desquelles sont faites les simulations par les modèles de plante. Ils permettent de mesurer directement la phénologie de la végétation qui dépend non seulement de la pluie mais aussi de l'occupation du sol, ainsi que d'autres contraintes biotiques (maladies, sautériaux) non prises en compte dans les modèles.

2. Perspectives

Plusieurs axes de recherche et développement se dégagent de ces travaux. Ces axes portent d'une part sur l'amélioration des méthodes proposées pour le suivi de l'agriculture en Afrique de l'Ouest, et d'autre part sur l'applicabilité de ces méthodes dans les systèmes d'alerte précoce.

2.1. Perspectives de recherche

En ce qui concerne la cartographie et la caractérisation des systèmes agricoles au Mali, les résultats sont encourageants et incitent à poursuivre les recherches pour consolider les méthodes développées.

Tout d'abord, nous pourrions vérifier la généralité de nos méthodes à l'échelle nationale i) en validant notre cartographie du domaine cultivé par d'autres images à haute résolution spatiale et ii) en utilisant d'autres données de terrain pour la classification des systèmes agricoles, pour élargir la gamme des types de systèmes agricoles cartographiés et rendre l'algorithme de classification plus robuste. Nous pensons par exemple que les pourcentages des différentes cultures provenant de la base de données IER/CMDT pourraient être utilisés (nous avons gardé les cultures majoritaires par village uniquement). Par ailleurs, d'autres villages doivent être échantillonnés pour pouvoir prendre en compte la culture du riz, qui est très importante au Mali, surtout dans les zones irriguées qui ne faisaient pas partie de notre zone d'étude, et que nous n'avons donc pas pu cartographier dans ces travaux faute de données de terrain. Ceci permettrait d'étendre la validation à l'échelle globale, qui n'a pour l'instant été réalisée que sur le Sud du Mali, via quatre images SPOT (THRS) et via les enquêtes agricoles sur 100 villages par l'IER. Nous avons en effet conscience que ces études à l'échelle globale nécessitent un travail de validation plus poussé. Mais ces travaux constituent une première approche qui a montré de bons résultats sur les sites de validation, et ouvrent d'ores et déjà différentes pistes de recherche.

Il serait ainsi intéressant d'étudier le poids de la texture à moyenne résolution spatiale de manière plus approfondie. En effet, les résultats de cartographie et de caractérisation des systèmes agricoles ont montré des performances satisfaisantes, dans lesquelles les indicateurs de texture ont une place très importante. Cette signature texturale propre à un système agricole serait à vérifier, par différents tests dans l'extraction de la texture, de même que par la multiplication d'indicateurs, qui sont largement utilisés avec des images à haute résolution spatiale (les 14 indices texturaux d'Haralick par exemple), mais très peu à moyenne résolution spatiale.

Ensuite, l'utilisation de la fouille de données a montré de bons résultats dans son approche préliminaire, qui nous a semblée indispensable pour assurer des bases communes aux deux communautés engagées dans ce travail, les agronomes et les spécialistes de fouille de données. Il s'agirait par la suite de développer ce travail d'extraction de connaissances sur une base de données plus large et également de valoriser le travail de classification par fouille présenté dans ce manuscrit.

D'autres problématiques de recherche qui n'ont pu être approfondies durant cette thèse et qui seraient pertinentes à aborder, concernent le suivi des cultures. Deux questions restent en effet en suspens :

- Premièrement, les indicateurs phénologiques extraits à l'échelle d'un pixel MODIS de 500 m nous font penser que le poids de la végétation naturelle dans le signal de la phénologie peut être important. Au Mali et plus généralement en Afrique de l'Ouest, le régime des précipitations entraîne une synchronisation de la croissance des différents types d'occupation du sol, regroupée sur quelques mois. Il faudrait donc étudier la phénologie de la végétation naturelle et celle des surfaces cultivées de

façon disjointe. Cela permettrait d'évaluer l'influence de la présence de végétation naturelle au sein ou autour des champs cultivés au Mali sur les indicateurs satellitaires de la phénologie, ou éventuellement de capter une phénologie différente, avec un démarrage plus précoce pour la végétation naturelle, comme nous en avons fait l'hypothèse. Nous préconisons des comparaisons dans différentes zones climatiques et au cours de différentes années, afin de mieux comprendre le lien entre phénologie, climat et utilisation du sol.

- Deuxièmement, il serait intéressant d'étudier les impacts de l'information phénologique satellitaire sur l'estimation des rendements par modélisation agro-météorologique. Pour cela, il faudrait i) essayer d'améliorer l'algorithme de Zhang, qui ne fonctionne que lorsqu'il y a moins de deux valeurs d'indices de végétation manquantes sur les séries temporelles MODIS (or les pays tropicaux présentent fréquemment plus de deux quinzaines ennuagées), ii) imaginer comment intégrer cette information phénologique dans la méthodologie d'estimation du rendement, iii) évaluer les améliorations éventuelles de l'estimation du rendement. Par ailleurs, ces dernières années, le développement d'indices phénologiques dérivés du NDVI a émergé comme un axe de recherche clé dans l'étude des interactions biosphère-atmosphère, les changements climatiques, biologiques et anthropiques. Il pourrait donc être pertinent de tester les nouvelles métriques présentées dans ces travaux de thèse dans d'autres domaines d'application.

Enfin, il serait intéressant d'appliquer ces méthodes sur les futures données des satellites Sentinel-2 de l'ESA. Développés dans le cadre du programme GMES, ils seront mis en orbite en 2013 et 2015, pour assurer la continuité des missions comme SPOT et Landsat. Ces satellites seront dédiés à l'observation des surfaces continentales avec une résolution décamétrique, un temps de revisite de 5 à 10 jours, et une fauchée de 290 km. On peut donc imaginer une couverture du Mali (environ 600 000 km² sans la partie désertique) en moins d'une dizaine d'images.

2.2. Perspectives opérationnelles

2.2.1. Transférabilité

Dans le cadre de cette thèse, nous avons travaillé sur le Mali, pays dont l'agriculture est particulièrement fragmentée, comme dans la plupart des pays d'Afrique de l'Ouest. La cartographie et la caractérisation des systèmes agricoles nécessitent de l'imagerie MODIS, gratuite, ainsi que des données de terrain sur les systèmes agricoles à l'échelle d'un village. A l'exception de la stratification qui a été effectuée à partir d'un logiciel à licence payante (eCognition), les méthodes utilisées sont génériques (classification non supervisée, photo-interprétation, algorithmes de classification), rendant leur mise en place opérationnelle simple et rapide. Il serait donc intéressant de tester la transférabilité de ces méthodes dans d'autres pays d'Afrique de l'Ouest, l'étape de stratification pouvant être effectuée à partir de logiciels libres dans lesquels des modules de segmentation sont aujourd'hui disponibles.

2.2.2. Intégration des résultats dans les systèmes d'alerte précoce

Les différentes méthodes développées pourraient apporter de l'information aux systèmes d'alerte précoce ouest-africains. Le couplage de la cartographie des systèmes agricoles avec un modèle de plante et des données de terrain pourrait améliorer le fonctionnement d'un système d'alerte opérationnel, différent de l'existant. Tout d'abord, nous pensons que le masque de culture développé dans ces travaux pourrait aider les systèmes d'alerte précoce de deux manières : 1) pour le suivi de la croissance des cultures par télédétection, cela permettrait d'analyser uniquement les surfaces cultivées (notamment pour les cartes des anomalies de croissance) et de limiter les vérifications sur le terrain aux zones à risques, et 2) pour l'échantillonnage spatial des enquêtes nationales du Ministère de l'Agriculture du Mali par exemple (le Recensement Général de l'Agriculture a lieu tous les 10 ans environ). Ensuite, il est difficile de coupler des variables biophysiques de surface tels que le LAI avec les modèles de plante car les pixels de l'imagerie satellitaire utilisée pour le suivi de l'agriculture à l'échelle régionale ou globale sont très souvent mixtes, et n'aident donc pas à l'amélioration de la précision des modèles (mélange de différents types d'occupation du sol). Il faudra attendre l'arrivée de nouveaux capteurs à très haute résolution spatiale. C'est pourquoi, jusqu'à présent, la plupart des systèmes d'alerte sont très fortement basés sur les statistiques agricoles pour l'estimation de la production. L'observation de variables issues d'imagerie satellitaire est pour l'instant « externe » au système d'alerte, dans le sens où il évalue les anomalies de production par un suivi de NDVI par exemple, sans quantification directe. Les enquêtes sur le terrain à l'origine de la prévision de la production sont par essence localisées, et les prévisions agrégées. Le Système d'Alerte Précoce (SAP) du Mali, fournit par exemple des prévisions de production à l'échelle du cercle (région administrative malienne). Notre approche pourrait permettre des prédictions à l'échelle du village. Au-delà de l'échelle des prévisions, nous pensons également qu'il est possible d'extraire de jeux de données à moyenne résolution spatiale des informations sur les pratiques agricoles. La présence de coton dans le système agricole implique par exemple l'utilisation de fertilisants. La détection de surfaces irriguées a également déjà été démontrée dans d'autres travaux. Ce sont autant de pratiques « télédétectables » qui pourraient améliorer les prévisions des systèmes d'alerte.

Plus largement, nous avons axé ces travaux sur l'insécurité alimentaire conjoncturelle. Or, les idées convergent actuellement sur la nécessité d'un plus grand engagement des pays dans la recherche de solutions durables aux causes structurelles de l'insécurité alimentaire. L'insécurité alimentaire structurelle est celle dont la cause principale est la très faible capacité de réponse des ménages en raison des conditions d'extrême vulnérabilité. Ainsi les méthodes et les résultats obtenus devraient pouvoir être une base importante des études prospectives sur le long terme. Les observations sur les évolutions des systèmes agricoles (extension des surfaces cultivées, recours à l'irrigation, intensification etc) doivent permettre une évaluation de la vulnérabilité des agro-écosystèmes et des propositions de changements, avec la mise en œuvre de politiques et stratégies alimentaires adaptées.

LISTE DES FIGURES et DES TABLEAUX

Figure 1 : Organisation générale de la thèse.	16
Figure 2 : Répartition par régions du monde du nombre de personnes (en millions) sous-alimentées	19
Figure 3 : Concentration par pays des personnes souffrant de sous-alimentation en 2007 (selon les données 2007-2008 de l'indice de développement humain) (Source PNUD 2007/2008).....	19
Figure 4: Nombre de personnes sous-alimentées dans le monde (Source : FAO, 2010)	19
Figure 5 : Carte du Mali (www1.alliancefr.com).	22
Figure 6 : Précipitations annuelles et régions agro-climatiques en Afrique de l'Ouest (Source : FEWS 2007).	24
Figure 7 : Calendrier saisonnier et événements significatifs pour l'agriculture au Mali. (Source : FEWS NET Mali).	24
Figure 8 : Principaux capteurs fournissant des images de basse à très haute résolution spatiale dans le domaine du visible et de l'infrarouge.	32
Figure 9 : Anomalies du NDVI au 31 Août 2011 par rapport à la moyenne 2001 – 2010 (Source : bulletin FEWS NET Septembre 2011).	35
Figure 10 : NOAA Climate Prediction Center FEWS NET Rainfall Estimate (mm) en Afrique (7 Aout 2011).	37
Figure 11 : Composition colorée RVB à partir de séries temporelles de NDVI au Sud Mali en 2007 ..	40
Figure 12 : Localisation des 4 images SPOT acquises en 2007 au Mali : Boro, Cinzana, Koutiala et Sikasso et Quicklook des images.	43
Figure 13 : Séries temporelles MODIS 2007 (mars /juin/septembre) a) NDVI et b) indice de variance.	44
Figure 14 : Indicateurs phénologiques issus du produit MOD12Q2. 1. Date de démarrage de la saison (Start-of-Season) ; 2. Date de maturité (Start-of-Maximum of Season) ; 3. Date de sénescence (End-of-Maximum of Season) ; 4. Date du minimum de la saison (End-of-Season) ; 5. et 6. Valeur de l'indice de végétation pour les dates du minimum et du maximum ; 7. Intégrale de l'indice de végétation durant la saison de croissance. (Source : Zhang <i>et al.</i> , 2003).	45
Figure 15 : Extrait des points GPS enregistrés dans la région de Koutiala, sur fond d'image SPOT...	46
Figure 16 : Illustration de quelques paysages rencontrés lors de missions sur les sites d'études (E.Vintrou).	47
Figure 17 : Régions agricoles et systèmes agraires locaux dans les zones cotonnières du Mali (Soumare, 2008).	49
Figure 18 : Localisation des 100 villages enquêtés de la base de données de l'IER et de la CMDT...	49
Figure 19 : Localisation des 8 stations synoptiques utilisées et des isohyètes (Source : IER Bamako).	50
Figure 20 : Représentation classique du processus d'extractions de connaissances.....	53
Figure 21: La méthode géomatique pour la détection du domaine cultivé au Mali :.....	56
Figure 22 : Comparaison du domaine cultivé sur la référence (SPOT) du produit MODIS _{crop} et des produits globaux.....	57
Figure 23 : Comparaison des produit MODIS _{crop} et MODIS V05.	59
Figure 24 : Les différentes étapes du processus d'extraction de connaissances.....	59

Figure 25 : Méthodologie et résultat pour la caractérisation des systèmes agricoles au Mali.....	66
Figure 26 : Importance des variables (a) et distribution des principales variables par classe (b).	68
Figure 27 : Cartographie des systèmes agricoles du Sud Mali à l'échelle villageoise.....	69
Figure 28 : Méthodologie et résultats pour le suivi de la croissance des cultures.....	75
Figure 29 : Cartographie de l'indicateur « Start-of-Season » dans les 8 stations synoptiques en 2007 et histogrammes des valeurs.....	76
Figure 30 : Démarrage de la saison des cultures pour la station de Bougouni : histogrammes des valeurs et profils d'EVI sur 3 pixels.....	77
Figure 31 : Calcul du démarrage de saison des cultures (SOS), du début de maximum (SMAX), de la fin du maximum (EOS) et de la fin de saison (EOS) sur les profils LAI à partir de l'algorithme de Zhang. a) profil de sorgho caudatum fertilisé ; b) profil de mil souba non fertilisé en 2007.....	79
Figure 32 : Minimum, maximum et moyenne des 4 indicateurs phénologiques par station en 2007. Les lignes bleues représentent les indicateurs phénologiques issus de MODIS, les lignes rouges ceux issus du LAI.....	80

Tableau 1 : Résolutions spatiale et temporelle de données satellitaires pour les applications en agriculture.....	27
Tableau 2: Caractéristiques des produits globaux d'occupation du sol.....	29
Tableau 3 : Données et méthodes utilisées par les systèmes d'alerte FEWS NET et GMFS pour la cartographie et le suivi des surfaces cultivées.....	34
Tableau 4 : Composition du produit annuel MCD12Q2 V005.....	40
Tableau 5 : Principales caractéristiques des données satellitaires utilisées.....	41
Tableau 6 : Principales caractéristiques des 4 zones d'étude retenues pour l'acquisition des images SPOT.....	42
Tableau 7 : Notation sur une matrice de confusion.....	51
Tableau 8: Matrice de confusion entre MODIS _{crop} et SPOT et entre MODIS V05 et SPOT pour les six sites de validation.....	58
Tableau 9 : Comparaison des deux approches pour la détection du domaine cultivé.....	60
Tableau 10: Matrice de confusion entre MODIS _{crop} et SPOT et entre MODIS _{fouille} et SPOT.....	61
Tableau 11 : Matrice de confusion pour les 100 villages classés.....	67
Tableau 12 : Choix des variétés à simuler par le modèle SARRA-H par station synoptique.....	74
Tableau 13 : Valeurs des quatre indicateurs phénologiques sur les 8 stations synoptiques en 2007: moyenne et écart type avant et après corrections des histogrammes.....	78

BIBLIOGRAPHIE

- Achard, F., H. Eva, and P. Mayaux, 2001. Tropical forest mapping from coarse spatial resolution satellite data: production and accuracy assessment issues. *International Journal of Remote Sensing*, 22(14):2741-2762.
- Adams, J.B., D.E. Sabol, V. Kapos, R. Almeida, D.A. Roberts, M.O. Smith, and A.R. Gillespie, 1995. Classification of Multispectral Images Based on Fractions of Endmembers - Application to Land-Cover Change in the Brazilian Amazon. *Remote Sensing of Environment*, 52(2):137-154.
- Aksoy, S., K. Koperski, C. Tusk, and G. Marchisio, 2004. Interactive training of advanced classifiers for mining remote sensing image archives. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 773-782.
- Arino, O., P. Bicheron, F. Achard, J. Latham, R. Witt, and J.L. Weber, 2008. GLOBCOVER The most detailed portrait of Earth. *Esa Bulletin-European Space Agency*(136):24-31.
- Arvor, D., M. Margareth, V. Dubreuil, A. Bégué, and Y.E. Shimabukuro, 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, 32(2):702-713.
- Balme, M., S. Galle, and T. Lebel, 2005. Démarrage de la saison des pluies au Sahel: variabilité aux échelles hydrologique et agronomique, analysée à partir des données EPSAT-Niger. *Sécheresse*, 16(1):15-22.
- Becker-Reshef, I. et al., 2010. Monitoring global croplands with coarse resolution Earth observations: The GLobal Agriculture Monitoring (GLAM) project. *Remote Sensing*, 2(6):1589-1609.
- Belieres, J.F., J.É. Bidou, and I. Droy, 2009. Le cas des systèmes de production cotonniers au Mali. In: B. Thibaud and A. François (Editors),. *Systèmes de production et durabilité dans les pays du Sud*, KARTHALA Editions, pp.247-274.
- Bernard, S., 2009. *Forêts Aléatoires: De l'Analyse des Mécanismes de Fonctionnement à la Construction Dynamique*. Ph.D. dissertation, University of Rouen, Rouen (FR), 192 p.
- Bernard, S., L. Heutte, and S. Adam, 2009. On the selection of decision trees in Random Forests, *Proceedings of the International Conference on Intelligence Computing (ICIC)*. IEEE, Atlanta (Georgia, USA), pp. 302-307.
- Boschetti, L., S.P. Flasse, and P.A. Brivio, 2004. Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: The Pareto Boundary. *Remote Sensing of Environment*, 91(3-4):280-292.
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1):5-32.
- Breiman, L., and A. Cutler, 2005. Random Forests. Berkeley, URL: www.stat.berkeley.edu/users/breiman/RandomForests/ (last date accessed September 2011).
- Brown, M.E., and K.M. de Beurs, 2008. Evaluation of multi-sensor semi-arid crop season parameters based on NDVI and rainfall. *Remote Sensing of Environment*, 112(5):2261-2271.
- Brown, M.E., C.C. Funk, G. Galu, and R. Choularton, 2007. Earlier famine warning possible using remote sensing and models. *Eos*, 88:39.
- Carrão, H., P. Gonçalves, and M. Caetano, 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sensing of Environment*, 112(3):986-997.
- Chan, J.C.W., and D. Paelinckx, 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2999-3011.
- De Beurs, K.M., and G.M. Henebry, 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high latitude transects. *Global Change Biology*, 11(5):779-790.

- Deffontaines, J.P., and P. Thinon, 2001. Des entités spatiales significatives pour l'activité agricole et pour les enjeux environnementaux et paysagers. Contribution à une agronomie du territoire. *Courrier de l'Environnement, Inra*, 44:13-28.
- Defries, R.S., and J.R.G. Townshend, 1994. NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, 15(17):3567-3586.
- Devèze, J.C., 2008. *Défis agricoles africains*. KARTHALA Editions, 414 p.
- Dingkuhn, M. et al., 2003. Decision support tools for rainfed crops in the Sahel at the plot and regional scales. In: S.B. Tjark and W. M.C.S (Editors), *Decision Support Tools for Smallholder Agriculture in Sub-Saharan Africa: A Practical Guide*. IFDC, pp. 127-139.
- Doraiswamy, P.C., J.L. Hatfield, T.J. Jackson, B. Akhmedov, J. Prueger, and A. Stern, 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment*, 92(4):548-559.
- Doraiswamy, P.C., T.R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, and J. Prueger, 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, 97(2):192-202.
- Dufumier, M., 1996. *Les projets de développement agricole: manuel d'expertise*. KARTHALA Editions, 354 p.
- El Hajj, M., A. Bégué, S. Guillaume, and J.F. Martiné, 2009. Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices--The case of sugarcane harvest on Reunion Island. *Remote Sensing of Environment*, 113(10):2052-2061.
- Elmore, A.J., J.F. Mustard, S.J. Manning, and D.B. Lobell, 2000. Quantifying vegetation change in semiarid environments: Precision and accuracy of spectral mixture analysis and the Normalized Difference Vegetation Index. *Remote Sensing of Environment*, 73(1):87-102.
- FAO, 2003, 2007, 2010. FAOSTAT. URL : <http://faostat.fao.org/> (last date accessed December 2011).
- Fayyad, U.M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, 1996. *Advances in knowledge discovery and data mining*. Menlo Park, Calif.: AAAI Press.
- Friedl, M.A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168-182.
- Fritz, S., I. McCallum, C. Schill, C. Perger, R. Grillmayer, F.d.r. Achard, F. Kraxner, and M. Obersteiner, 2009. Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, 1(3):345-354.
- Fritz, S., and L. See, 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology*, 14(5):1057-1075.
- Fritz, S., L. See, and F. Rembold, 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *International Journal of Remote Sensing*, 31(9):2237-2256.
- Fritz, S. et al., 2011. Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets. *Geophysical Research Letters*, 38(4):L04404.
- Funk, C., and M.E. Budde, 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, 113(1):115-125.
- Ganguly, S., M.A. Friedl, B. Tan, X. Zhang, and M. Verma, 2010. Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, 114(8):1805-1816.
- Giri, C., Z. Zhu, and B. Reed, 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment*, 94(1):123-132.
- Groten, S., 1993. NDVI—crop monitoring and early yield assessment of Burkina Faso. *Remote Sensing*, 14(8):1495-1515.

- Hansen, M.C., and B. Reed, 2000a. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing*, 21(6-7):1365-1373.
- Hansen, M.C., and B. Reed, 2000b. A comparison of the IGBP DISCover and University of Maryland 1km global land cover products. *International Journal of Remote Sensing*, 21(6-7):1365-1373.
- Herman, A., V.B. Kumar, P.A. Arkin, and J.V. Kousky, 1997. Objectively determined 10-day African rainfall estimates created for famine early warning systems. *International Journal of Remote Sensing*, 18(10):2147-2160.
- Herold, M., P. Mayaux, C.E. Woodcock, A. Baccini, and C. Schmullius, 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112(5):2538-2556.
- Herold, M., C.E. Woodcock, A. di Gregorio, P. Mayaux, A.S. Belward, J. Latham, and C.C. Schmullius, 2006. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 44(7):1719-1727.
- Heumann, B.W., J.W. Seaquist, L. Eklundh, and P. Jonsson, 2007. AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982-2005. *Remote Sensing of Environment*, 108(4):385-392.
- Hountondji, Y.C., N. Sokpon, and P. Ozer, 2006. Analysis of the vegetation trends using low resolution remote sensing data in Burkina Faso (1982-1999) for the monitoring of desertification. *International Journal of Remote Sensing*, 27(5):871-884.
- Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, and L.G. Ferreira, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2):195-213.
- Husak, G.J., M.T. Marshall, J. Michaelsen, D. Pedreros, C. Funk, and G. Galu, 2008. Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. *Journal of Geophysical Research-Atmospheres*, 113(D14).
- Ippoliti-Ramilo, G.A., J.C.N. Epiphonio, and Y.E. Shimabukuro, 2003. Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries. *International Journal of Remote Sensing*, 24(7):1521-1534.
- Jakubauskas, M.E., D.R. Legates, and J.H. Kastens, 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, 37(1-3):127-139.
- Julea, A., N. Meger, P. Bolon, C. Rigotti, M.P. Doin, C. Lasserre, E. Trouve, and V.N. Lazarescu, 2011. Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(4):1417-1430.
- Justice, C.O., J.R.G. Townshend, B.N. Holben, and C.J. Tucker, 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, 6(8):1271-1318.
- Kaptué Tchuenté, A.T., J.-L. Roujean, and S. Faroux, 2010. ECOCLIMAP-II: An ecosystem classification and land surface parameters database of Western Africa at 1 km resolution for the African Monsoon Multidisciplinary Analysis (AMMA) project. *Remote Sensing of Environment*, 114(5):961-976.
- Kaptué Tchuenté, A.T., J.L. Roujean, and S.M. De Jong, 2011. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation*, 13(2):207-219.
- Kastens, J.H., T.L. Kastens, D.L.A. Kastens, K.P. Price, E.A. Martinko, and R.Y. Lee, 2005. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment*, 99(3):341-356.
- Kidane, W., M. Maetz, and P. Dardel (Editors), 2006. *Sécurité alimentaire et développement agricole en Afrique subsaharienne*. Dossier pour l'accroissement des soutiens publics, Rapport Principal, Série sur l'Assistance aux Politiques 2, FAO, Rome., 127 p.

- Kouressy, M. et al., 2008. Adaptation of Malian sorghums to climate variability. *Adaptation des sorghos du Mali a la variabilite climatique*, 17(2):95-100.
- Latinne, P., O. Debeir, and C. Decaestecker, 2001. Limiting the number of trees in random forests. *Multiple Classifier Systems*:178-187.
- Lawrence, R., A. Bunn, S. Powell, and M. Zambon, 2004. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote Sensing of Environment*, 90(3):331-336.
- Liaw, A., and M. Wiener, 2002. Classification and Regression by randomForest. *R news*, 2(3):18–22.
- Lobell, D.B., and G.P. Asner, 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, 93(3):412-422.
- Lobell, D.B., and C.B. Field, 2007. Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, 2:014002.
- Los, S.O., 1998. Estimation of the ratio of sensor degradation between NOAA AVHRR channels 1 and 2 from monthly NDVI composites. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1):206-213.
- Lowry, J. et al., 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: A case study of the Southwest Regional Gap Analysis Project (SWReGAP). *Remote Sensing of Environment*, 108(1):59-73.
- Ma, G., J. Huang, W. Wu, J. Fan, J. Zou, and S. Wu, 2011. Assimilation of MODIS-LAI into WOFOST model for forecasting regional winter wheat yield. *Mathematical and Computer Modelling*(0).
- Malingreau, J., E. Bartholomé, and E. Barisano, 1987. Surveillance de la production agricole en Afrique de l'Ouest ; Nécessité d'une intégration de différentes plates-formes satellitaires. *Proceedings of the Symposium SPOT*, Paris (FR), pp. 353-369.
- Maselli, F., C. Conese, L. Petkov, and M. Gilabert, 1993. Environmental monitoring and crop forecasting in the Sahel through the use of NOAA NDVI data. A case study: Niger 1986–89. *International Journal of Remote Sensing*, 14(18):3471-3487.
- Maselli, F., S. Romanelli, L. Bottai, and G. Maracchi, 2000. Processing of GAC NDVI data for yield forecasting in the Sahelian region. *International Journal of Remote Sensing*, 21(18):3509-3523.
- Mayaux, P., E. Bartholomé, S. Fritz, and A. Belward, 2004. A new land cover map of Africa for the year 2000. *Journal of Biogeography*, 31(6):861-877.
- McCallum, I., M. Obersteiner, S. Nilsson, and A. Shvidenko, 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, 8(4):246-255.
- Monteith, J., 1972. Solar radiation and productivity in tropical ecosystems. *Journal of Applied Ecology*, 9(3):747-766.
- Murakami, T., S. Ogawa, N. Ishitsuka, K. Kumagai, and G. Saito, 2001. Crop discrimination with multitemporal SPOT/HRV data in the Saga Plains, Japan. *International Journal of Remote Sensing*, 22(7):1335-1348.
- Na, X., S. Zhang, X. Li, H. Yu, and C. Liu, 2010. Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic. *Photogrammetric Engineering and Remote Sensing*, 76(7): 833-840.
- Ozdogan, M., and C.E. Woodcock, 2006. Resolution dependent errors in remote sensing of cultivated areas. *Remote Sensing of Environment*, 103(2):203-217.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217-222.
- Panigrahy, S., and S. Sharma, 1997. Mapping of crop rotation using multirate Indian Remote Sensing Satellite digital data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(2):85-91.

- Pax-Lenney, M., and C.E. Woodcock, 1997. Monitoring agricultural lands in Egypt with multitemporal landsat TM imagery: How many images are needed? *Remote Sensing of Environment*, 59(3):522-529.
- Petitjean, F., A. Ketterlin, and P. Gançarski, 2011. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678-693.
- Pitarch, Y., E. Vintrou, F. Badra, A. Bégué, and M. Teisseire, 2011. Mining Sequential Patterns from MODIS Time Series for Cultivated Area Mapping. In: S. Geertman, W. Reinhardt and F. Toppen (Editors), *Advancing Geoinformation Science for a Changing World*. Springer Heidelberg Dordrecht London New-York, pp. 45-62.
- Pittman, K., M.C. Hansen, I. Becker-Reshef, P.V. Potapov, and C.O. Justice, 2010. Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, 2(7):1844-1863.
- Plantevit, M., A. Laurent, D. Laurent, M. Teisseire, and Y.W. Choong, 2010. Mining multidimensional and multilevel sequential patterns. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(1):4.
- Potgieter, A.B., A. Apan, P. Dunn, and G. Hammer, 2007. Estimating crop area using seasonal time series of Enhanced Vegetation Index from MODIS satellite imagery. *Australian Journal of Agricultural Research*, 58(4):316-325.
- Qiu, J., H. Tang, S. Froking, S. Boles, C. Li, X. Xiao, J. Liu, Y. Zhuang, and X. Qin, 2003. Mapping Single-, Double-, and Triple-crop Agriculture in China at 0.5°x 0.5° by Combining Country-scale Census Data with a Remote Sensing derived Land Cover Map. *Geocarto International*, 18(2):3-13.
- Rasmussen, M.S., 1992. Assessment of millet yields and production in northern Burkina Faso using integrated NDVI from the AVHRR. *International Journal of Remote Sensing*, 13(18):3431-3442.
- Reed, B.C., J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen, 1994. Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*:703-714.
- Rojas, O., F. Rembold, J. Delincé, and O. Léo, 2011. Using the NDVI as auxiliary data for rapid quality assessment of rainfall estimates in Africa. *International Journal of Remote Sensing*, 32(12):3249-3265.
- Romani, L.A.S., 2010. Mining relevant and extreme patterns on climate time series with clipsminer. *Journal of Information and Data Management*, 1(2):245.
- Rouse, I., 1974. The explanation of culture change. *Science*, 185(4148):343-344.
- Ruelland, D., A. Dezetter, C. Puech, and S. Ardoin-Bardin, 2008. Long-term monitoring of land cover changes based on Landsat imagery to improve hydrological modelling in West Africa. *International Journal of Remote Sensing*, 29(12):3533-3551.
- Ruelland, D., F. Levavasseur, and A. Tribotté, 2010. Patterns and dynamics of land-cover changes since the 1960s over three experimental areas in Mali. *International Journal of Applied Earth Observation and Geoinformation*, 12:S11-S17.
- Sakamoto, T., M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno, 2005. A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96(3-4):366-374.
- Sannier, C., J. Taylor, W.D. Plessis, and K. Campbell, 1998. Real-time vegetation monitoring with NOAA-AVHRR in Southern Africa for wildlife management and food security assessment. *International Journal of Remote Sensing*, 19(4):621-639.
- Sautter, G., and P. Péliissier, 1964. Pour un atlas des terroirs africains: Structure-type d'une étude de terroir. *L'homme*, 4(1):56-72.
- Schwartz, M.D., and J.M. Hanes, 2010. Intercomparing multiple measures of the onset of spring in eastern North America. *International Journal of Climatology*, 30(11):1614-1626.

- Senay, G.B., and J. Verdin, 2003. Characterization of yield reduction in Ethiopia using a GIS-based crop water balance model. *Canadian Journal of Remote Sensing*, 29(6):687-692.
- Silva, M., G. Camara, M. Escada, and R. Souza, 2008. Remote Sensing Image Mining: Detecting Agents of Land Use Change in Tropical Forest Areas. *International Journal of Remote Sensing*, 29(16):4803-4822.
- Soares, D.d.M., L.S. Galvao, and A.R. Formaggio, 2008. Crop area estimate from original and simulated spatial resolution data and landscape metrics. *Scientia Agricola*, 65(5):459-467.
- Soumare, M., 2008. Dynamique et durabilité des systemes agraires à base de coton au Mali. Thèse de Doctorat en Géographie Humaine, Economique et Regionale, Université de Paris X Nanterre PP 373.
- Soumare, M., D. Bazile, M. Vaksman, M. Kouressy, K. Diallo, and C.H. Diakite, 2008. Agroecosystemic diversity and future of traditional cereals in the Malian cotton belt. *Cahiers Agricultures*, 17(2):79-85.
- Thenkabail, P., 2003. Biophysical and yield information for precision farming from near-real-time and historical Landsat TM images. *International Journal of Remote Sensing*, 24(14):2879-2904.
- Thenkabail, P.S., M. Schull, and H. Turrall, 2005. Ganges and Indus river basin land use/land cover (LULC) and irrigated area mapping using continuous streams of MODIS data. *Remote Sensing of Environment*, 95(3):317-341.
- Townshend, J.R.G., C. Justice, and V. Kalb, 1987. Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing*, 8(8):1189-1207.
- Tucker, C., C.L. Vanpraet, M. Sharman, and G. Van Ittersum, 1985. Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980-1984. *Remote Sensing of Environment*, 17(3):233-249.
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127-150.
- Tucker, C.J., and B.J. Choudhury, 1987. Satellite remote sensing of drought conditions. *Remote Sensing of Environment*, 23(2):243-251.
- Tucker, C.J., B. Holben, and J. Elgin, 1980. Relationship of spectral data to grain yield variation. *Photogrammetric Engineering and Remote Sensing*, 46:657-666.
- Verdin, J., and R. Klaver, 2002. Grid cell based crop water accounting for the famine early warning system. *Hydrological Processes*, 16(8):1617-1630.
- Vina, A., A.A. Gitelson, D.C. Rundquist, G. Keydan, B. Leavitt, and J. Schepers, 2004. Monitoring maize (*Zea mays* L.) phenology with remote sensing. *Agronomy Journal*, 96:1139-1147.
- Vintrou, E., A. Begue, C. Baron, and D. Lo Seen, 2011. Mapping Fragmented Agricultural Systems in the Sudano-Sahelian region using Random Forest and Metrics of Coarse Resolution MODIS Imagery. *Photogrammetric Engineering and Remote Sensing*, In revision.
- Vintrou, E., A. Desbrosse, A. Bégué, S. Traoré, C. Baron, and D. Lo Seen, 2012a. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):83-93.
- Vintrou, E., Y. Pitarch, A. Bégué, and M. Teisseire, 2012b. Recherche de motifs et cartographie des surfaces agricoles: des relevés terrain aux données satellitaires (application au Mali). *Revue Internationale de Géomatique*. In press.
- Wardlow, B.D., and S.L. Egbert, 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the US Central Great Plains. *Remote Sensing of Environment*, 112(3):1096-1116.
- Wardlow, B.D., S.L. Egbert, and J.H. Kastens, 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment*, 108(3):290-310.

- Wessels, K., K. Steenkamp, G. Von Maltitz, and S. Archibald, 2011. Remotely sensed vegetation phenology for describing and predicting the biomes of South Africa. *Applied Vegetation Science*, 14(1):49-66.
- White, M.A. et al., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, 15(10):2335-2359.
- White, M.A., P.E. Thornton, and S.W. Running, 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, 11(2):217-234.
- Wiegand, C., A. Richardson, and E. Kanemasu, 1979. Leaf area index estimates for wheat from LANDSAT and their implications for evapotranspiration and crop modeling. *Agronomy Journal*, 71(2):336-342.
- Xu, W., H. Jiang, and J. Huang, 2011. Regional Crop Yield Assessment by Combination of a Crop Growth Model and Phenology Information Derived from MODIS. *Sensors Letters*, 9(3):981-989.
- Zhang, X., M. Friedl, and C. Schaaf, 2009. Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. *International Journal of Remote Sensing*, 30(8):2061-2074.
- Zhang, X., M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.C.F. Hodges, F. Gao, B.C. Reed, and A. Huete, 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3):471-475.

ANNEXES

Annexe 1 - page 100

Vintrou E., Desbrosse A., Bégué A., Traoré P.C.S., Baron C., Lo Seen D. (2012) Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation* 14(1): 83-93.

Annexe 2 - page 112

Vintrou, E., Soumaré, M., Bernard, S., Bégué, A., Baron, C., Lo Seen, D. (2011). Mapping Fragmented Agricultural Systems in the Sudano-Sahelian region using Random Forest and Metrics of Coarse Resolution MODIS Imagery. (Special Issue) *Photogrammetric Engineering and Remote Sensing*. Version acceptée Janvier 2012.

Annexe 3 - page 134

Vintrou, E., Bégué, A., Baron, C., Lo Seen, D. (2012). Analysing MODIS phenometrics information content in heterogeneous agricultural landscapes of West Africa. In preparation.

Annexe 4 - page 160

Pitarch Y., Vintrou E., Badra F., Bégué A., Teisseire M. (2011) Mining Sequential Patterns from MODIS Time Series for Cultivated Area Mapping. *Advancing Geoinformation Science for a Changing World* : 45-62.

Annexe 5 - page 179

Vintrou E., Pitarch Y., Bégué A., Teisseire M. (2012) Recherche de motifs et cartographie des surfaces agricoles: des relevés terrain aux données satellitaires (application au Mali). *Revue Internationale de Géomatique*, 21(4): 469-488.

Annexe 6 - page 199

Les systèmes d'alerte précoce en Afrique de l'Ouest.

ARTICLE 1:

**Crop area mapping in West Africa
using landscape stratification of MODIS time series
and comparison with existing global land products**

International Journal of Applied Earth Observation and Geoinformation 14(1): 83-93.



Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products

Elodie Vintrou^{a,*}, Annie Desbrosse^a, Agnès Bégué^a, Sibiry Traoré^b, Christian Baron^a, Danny Lo Seen^a

^a CIRAD UMR TETIS, Maison de la Télédétection, 500 rue J.F. Breton, Montpellier F-34093, France

^b IER, Rue Mohamed V, BP 258, Bamako, Mali

ARTICLE INFO

Article history:

Received 28 October 2010

Accepted 30 June 2011

Keywords:

Cultivated area
MODIS time series
Stratification
Global land products
Mali

ABSTRACT

In Africa, food security early warning systems use satellite-derived data concerning crop conditions and agricultural production. Such systems can be improved if they are provided with a more reliable estimation of the cultivated area at national scale. This paper evaluates the potential of using time series from the MODerate resolution Imaging Spectroradiometer MOD13Q1 (16-day composite of normalized difference vegetation index at 250 m resolution) to extract cultivated areas in the fragmented rural landscapes of Mali. To this end, we first stratified Southern Mali into 13 rural landscapes based on the spatio-temporal variability of NDVI and textural indices, using an object-oriented classification scheme.

The accuracy of the resulting map (MODIS_{crop}) and how it compares with existing coarse-resolution global land products (GLC2000 Africa, GLOBCOVER, MODIS V05 and ECOCLIMAP-II), was then assessed against six crop/non-crop maps derived from SPOT 2.5 m resolution images used as references. For crop areal coverage, the MODIS_{crop} cultivated map was successful in assessing the overall cultivated area at five out of the six validation sites (less than 6% of the absolute difference), while in terms of crop spatial distribution, the producer accuracy was between 33.1% and 80.8%. This accuracy was linearly correlated with the mean patch size index calculated on the SPOT crop maps ($r^2 = 0.8$). Using the Pareto boundary as an accuracy assessment method at the study sites, we showed that (i) 20–40% of the classification crop error was due to the spatial resolution of the MODIS sensor (250 m), and that (ii) compared to MODIS V05, which otherwise performed better than the other existing products, MODIS_{crop} generally minimized omission–commission errors. A spatial validation of the different products was carried out using SPOT image classifications as reference. In the corresponding error matrices, the fraction of correctly classified pixels for our product was 70%, compared to 58% for MODIS V05, while it ranged between 40% and 51% for the GLC2000, the ECOCLIMAP-II and the GLOBCOVER.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The northern fringe of sub-Saharan Africa is a region that is considered to be particularly vulnerable to climate variability and change, and where food security remains a major challenge. To address this issue, major international research efforts coordinated by the African Monsoon Multidisciplinary Analyses (AMMA) project are being deployed. The aim of the project is to better understand the West African Monsoon and its variability and to improve predictions of the impact of this variability on West African societies. One of the necessary preliminary stages in analyzing such impacts on agriculture and food security is reliable estimation of the cultivated domain at a national level and at a scale compati-

ble with climate change. For that purpose, it is essential to develop global data sets of the geographic distribution of agricultural land use.

Remote sensing has become a widely used and effective method for estimating the areal coverage of land cover classes, and is an important step in monitoring crop conditions and yields. In sub-Saharan Africa, operational land cover mapping systems are restricted by the cost of high-resolution images. Thus, the monitoring of vast ecosystems at national or continental scales typically resorts to using free low- and medium-resolution images (Hountondji et al., 2006; Justice et al., 1985) such as NOAA-AVHRR, SPOT-VEGETATION and TERRA-MODIS. For the last two decades, large area land use/land cover (LULC) mapping has advanced thanks to efforts at regional (Boles et al., 2004; Carrão et al., 2008; Lowry et al., 2007; Wardlow et al., 2007) and global scales (Bartholomé and Belward, 2005; Defries and Townshend, 1994; Hansen et al., 2000; Loveland and Belward, 1997). In particular, new continental to global scale land cover data sets have emerged, such as the

* Corresponding author. Tel.: +33 4 67 54 87 54; fax: +33 4 67 54 87 00.

E-mail addresses: elodie.vintrou@cirad.fr, elodie.vintrou@gmail.com (E. Vintrou).

GLC2000 (Fritz et al., 2003) and GLOBCOVER (Arino et al., 2008) products. However, these products are generally more focused on large ecosystems than on agricultural systems, and are not expected to be suitable or reliable for the analysis of fragmented and heterogeneous African rural landscapes, because of subpixel heterogeneity (Lobell and Asner, 2004).

Two studies that compared different global products (Herold et al., 2006; McCallum et al., 2006) showed general agreement in terms of total areas at the global scale, but that at the regional scale, agreement was limited due to the difficulty of distinguishing between mixed classes of heterogeneous landscapes. Furthermore, there have been large areas of disagreement in the literature between different products concerning the Sahel belt of Africa (Fritz and See, 2008; Giri et al., 2005; Hansen and Reed, 2000; Herold et al., 2008; McCallum et al., 2006) and in particular disagreements about the extent of the cultivated area and its spatial distribution (Fritz et al., 2010; Kaptué Tchuenté et al., 2011).

The main reasons put forward to explain this difficulty in mapping cultivated areas in the Sahel belt are first, that the Sahel is a transition zone between the hyper-arid Sahara in the north and the more humid savannas and woodlands in the south, with specific weather conditions resulting in high regional variability in terms of agricultural systems (and calendars); and second, West African landscapes are complex, with small-scale farms, numerous trees in the fields, and synchronized phenology of crops and natural vegetation due to the rainfall regime.

Several studies have shown that image classifications of very heterogeneous areas lead to considerable confusion between different types of land cover (Husak et al., 2008). To reduce this confusion, the territory can be stratified into homogeneous areas. Previous attempts suggest that stratification is possible using coarse resolution satellite images (i.e., AVHRR) (Bastin et al., 1995; Mora and Iverson, 1997; Nelson and Horning, 1993; Roller and Colwell, 1986). However, these stratifications are generally performed empirically and visually on a limited number of AVHRR images. MODIS imagery, with its improved spatial and spectral resolution, could be used to stratify a large territory, prior to mapping cultivated areas, even in heterogeneous areas.

The purpose of this paper is to examine the ability of 250-m MODIS NDVI data to map the extent of cropland in Mali, a country that is representative of the Sahel belt of Africa. To this end, we first stratified the area of interest into homogeneous zones based on the radiometry and texture of the MODIS time series images. Then, crop/non-crop classifications were performed in each zone separately and merged at the national scale. To evaluate the accuracy of the resulting map and to compare it with existing coarse-resolution global land products (GLC2000 for Africa, GLOBCOVER, MODIS V05, ECOCLIMAP-II for AMMA zone), six detailed maps of cultivated areas derived from high-resolution images were used as a reference. The evaluations include the accuracy of both areal estimations and locations.

2. Materials and methods

2.1. The study area

Mali is a land-locked country with an area of 1,240,192 km², situated in West Africa between latitudes 10°N and 24°N (Fig. 1). It has a strong latitudinal climatic gradient that ranges from sub-humid to semi-arid and, further north, extends to arid and desertic. This climate gradient is representative of the Sahel belt of Africa. Mali encompasses the Sudano-Sahelian zone, where strong dependence on rainfed agriculture implies exposure to climate variability in addition to the impact of population growth on food security. Farming is the main source of income for many people of the

Sudano-Sahelian region. Millet and sorghum are the main food crops and they are sown at the beginning of the rainy season (between May and July) and harvested at the end (from September to November). The vast majority of the population (80%) are subsistence farmers who grow just enough food for their own needs. A few larger farms produce crops for sale (cash crops), mainly cotton and peanuts. Concerning inputs, improved seeds are used on only a small percentage of the cultivated area, for from 1% to 10% of all crops cultivated. The cash crops account for most fertilizers and pesticides even though consumption levels are very low. In this study, we did not include the Saharan zone in the north Mali where the sparse rainfall amounts to less than 200 mm per year.

2.2. Data

2.2.1. MODIS time series

The NASA Land Process Distributed Active Archive Center (LP DAAC) is the repository for all MODIS data. Among the MODIS products, we selected the “Vegetation Indices 16-Day L3 Global 250 m SIN Grid” temporal syntheses (MOD13Q1). MODIS temporal syntheses were used because of the high probability of cloud cover during the monsoon season in West Africa. With the MODIS 16-days NDVI time series product, the availability of cloud-free data is increased, compared to with 8-days temporal syntheses. For Mali, a set of 12 MODIS 16-day composite normalized difference vegetation index (NDVI) images at a resolution of 231.6 m was acquired for 2007 (the best quality composite image out of two for each month was kept). The year 2007 was chosen to overlap with the more recently available high-resolution data that are presented in Section 2.2.4.

2.2.2. Landsat images

A set of 10 Landsat ETM+ Orthoregocover images at a resolution of 30 m was used to fully cover Southern Mali. They were obtained free of charge from the Global Land Cover Facility (GLCF) (<http://glcf.umd.edu>). The images were acquired during the dry season (between October and January) in 1999, 2000, 2001 and 2002. The images were used to visually interpret the land cover types (crop or non-crop) of the MODIS time series classification.

2.2.3. Global land products

Among global land products, we chose the regional GLC2000 for Africa along with the more recent GLOBCOVER land cover, the MODIS V05 (MCD12Q1) and the ECOCLIMAP-II products (Table 1) as being representative of the most consistent global products currently available.

2.2.4. The validation data set

Validation sites were selected to sample the main agro-climatic regions of Central and Southern Mali (Fig. 1 and Table 2). Four multispectral SPOT images were acquired in November 2007 at 2.5 m resolution. They were processed and used as reference for the cultivated area at six validation sites (covering Koutiala, Sikasso, Boro and Cinzana districts; the last two scenes were split into northern and southern sites as they straddle two different agro-climatic zones). The images were classified using an object-based supervised classification method with eCognition4[®]. A ground data set collected during the 2009 and 2010 cropping seasons (744 GPS waypoints at the validation sites) was used to confirm the visual interpretation of the SPOT images. The selected samples represented about 10% of the image objects. Post-classification mainly based on visual interpretation was performed to correct misclassifications. After merging objects according to their land cover class (crop/non-crop), the final classification was exported in raster format.

Table 1
Characteristics of the four global products.

Product	GLC2000 for Africa	GLOBCOVER	MODIS V05 (MCD12Q1)	ECOCLIMAP-II
Author	Mayaux et al. (2004)	Arino et al. (2008)	Friedl et al. (2010)	Kaptué Tchuenté et al. (2010)
Platform – Sensor	SPOT-VGT-1	ENVISAT-MERIS	MODIS-Terra MODIS-Aqua	MODIS-Terra MODIS-Aqua
Input data	4 spectral bands and NDVI	13 spectral bands	3–7 bands	Leaf Area Index
Time of data collection	01/11/1999–31/12/2000	01/12/2004–31/06/2006	2007	01/01/2000–31/12/2007
Spatial resolution	1 km	300 m	500 m	1 km
Number of classes	27	22	17	37
Classification scheme	Land Cover Classification System	Land Cover Classification System	International Geosphere-Biosphere Programme	Land Cover Classification System
Classification method	Flexible classification using local/regional expertise	Flexible classification using local/regional expertise	Supervised classification using decision tree	Supervised classification by merging ECOCLIMAP-I and GLC2000
Validation scheme	Intercomparison Statistical sampling High resolution data	Statistical sampling	“Unseen training sites” and “confidence values aggregated by land cover class and continental region”	Error matrix Intercomparison High resolution data

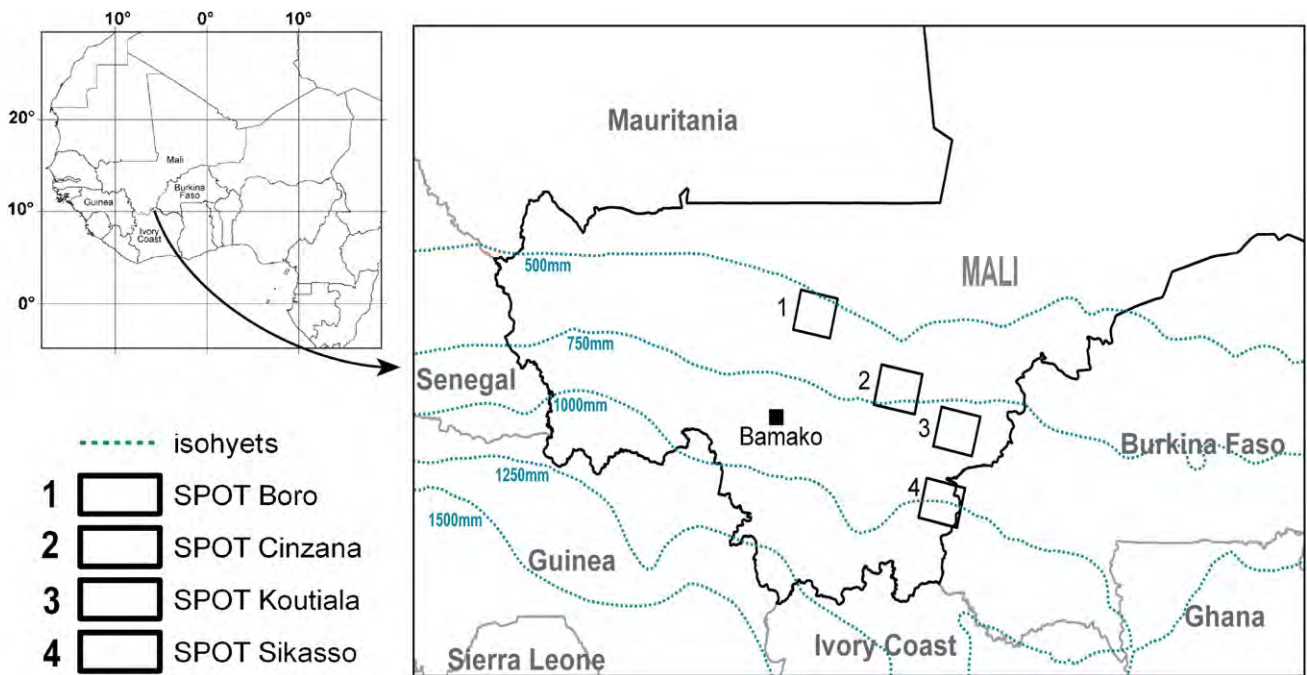


Fig. 1. Map of Mali and geographic footprints of the SPOT validation images along the rainfall gradient (yearly average between 1971 and 2000).

2.3. Methods

2.3.1. Stratification of Mali (MODIS_{strat})

It is generally accepted that an area broken down into patches can be analyzed more efficiently than an entire complex system. Landscape ecology focuses on the description of structures and patterns, and many authors have tried to reduce complexity using

different methods (Blaschke, 2006; Burnett and Blaschke, 2003; Farina, 2000; Forman, 1995; Saura, 2004).

We chose to stratify Mali landscapes only using the MODIS time series to reduce the spectro-temporal diversity of the image to be classified. Considering that prominent landscape characteristics include a temporal vegetation pattern (eco-climatic zones) and a landscape structure (morpho-pedologic zones and human activity),

Table 2
Main characteristics of the validation sites.

Site name (eco-climatic zone)	Mean annual rainfall	Main crops	Natural vegetation type
Boro (Sahelian)	500 mm/year	Millet, sorghum	High proportion of bare soils and sparse vegetation
Cinzana (Soudano-Sahelian)	600 mm/year	Millet, sorghum	High proportion of bare soils and sparse vegetation
Koutiala (Soudano-Sahelian)	750 mm/year	Millet, sorghum, cotton	Large areas of semi-open and closed natural vegetation
Sikasso (Soudanian)	1000 mm/year	Maize, cotton, fruit crops	Dense natural vegetation

Table 3
The global land product classes including crop components.

Product	Class number	Name	% crop in the class definition	Crop fraction
GLC2000 Africa	18	Croplands	>50	0.75
	19	Cropland with open woody vegetation	15–40	0.3
	20	Irrigated croplands	>50	0.75
GLOBCOVER	11	Post-flooding or irrigated croplands	No data	1
	14	Rain-fed croplands	No data	1
	20	Mosaic cropland/natural vegetation (grassland/shrubland/forest)	50–70	0.6
	30	Mosaic vegetation (grassland/shrubland/forest)/cropland	20–50	0.35
ECOCLIMAP-II	17	Dry tropical cropland	No data	1
	18	Cropland with open woody vegetation	No data	0.5
	19	Irrigated cropland	No data	1
MODIS V05	12	Croplands	No data	1
	14	Cropland/natural vegetation mosaics	≤60	0.5

we assumed that the spatio-temporal variability of NDVI and textural indices of MODIS images can be used to segment and classify the territory.

For the stratification, we used object-oriented image processing software. First, we conducted a segmentation of Mali based on radiometric and textural similarities in the MODIS data set. Radiometric data were composed of the 12-month NDVI time series, and textural data corresponded to the “contrast” and “dissimilarity” Haralick indices (Haralick, 1979), which are calculated on a 45×45 pixel window (about 13,000 ha) on three images from the MODIS NDVI time series (March, June and September 2007, corresponding to the dry season and the beginning and the end of the rainy season, respectively). After segmentation, a supervised classification enabled us to reduce the number of segments from 3000 to 13 zones that are homogeneous in terms of their NDVI time profiles and textures. The MODIS-derived landscape map of Mali produced was called MODIS_{strat}.

We compared our stratification to an existing one, which relies on a typology of agricultural systems based on cotton (Soumare, 2008). This stratification was developed by the Institute of Rural Economy in Bamako, and was called IER_{strat}.

2.3.2. Crop/non-crop classification of stratified MODIS time-series (MODIS_{crop})

For each stratified zone, 20 spectro-temporal clusters were obtained using a non-supervised classification algorithm (ISO-DATA). Each cluster was then assigned to one of the two predefined classes (crop/non-crop) after assessing the spatial distribution of the cluster pixels compared to a visual interpretation of the corresponding cover type(s) on Landsat ETM+ imagery. To support visual interpretation, a mean NDVI temporal profile was also produced for each cluster such that characteristic signatures corresponding to crop phenology could be recognized. Furthermore, the field surveys conducted in 2009 and 2010 were used to better identify cultivated areas during photo-interpretation. For example, crops are generally located in low-lying areas and never on the duricrust capped tabular hills (laterite cuirasses). The resulting MODIS map was called MODIS_{crop}.

2.3.3. Crop/non-crop maps derived from global products

The four global products (GLC2000, GLOBCOVER, MODIS V05 and ECOCLIMAP-II) and MODIS_{crop} were compared to the SPOT crop/non-crop maps of the validation sites. Prior to that, the crop classes in the legends of the four global products had to be harmonized. After deciding that cultivated areas include crops and fallows, only classes including a “crop” component were retained and aggregated into one crop class to obtain a simplified, two-class legend (crop/non-crop). For subsequent areal calculations, these classes were weighted before aggregation according to the

fraction of crop given in the definition of each “crop” class. When the definition provided no quantitative information, empirical values of 1 and 0.5 were applied for pure and mixed classes (e.g., cropland/shrubland mosaic), respectively (Table 3).

2.3.4. Accuracy analysis of crop/non-crop products

The accuracy of the five crop/non-crop products (MODIS_{crop} and the four global products) was checked using the six validation sites: Boro North, Boro South, Cinzana North, Cinzana South, Koutiala and Sikasso.

First, to assess the accuracy of crop area estimated by the five products, the cultivated areas were compared (1) with the six crop/non-crop reference maps at the local scale, and (2) with the statistics provided by FAO at the national scale.

Second, the accuracy of crop spatial distribution was assessed using (1) the error matrix for the five products, and (2) the Pareto boundary method (Boschetti et al., 2004) only for the MODIS products (MODIS_{crop} and MODIS V05). The Pareto boundary method was included because in our case, the assumption that the reference and classified data have the same spatial resolution, as required for the former method, does not hold. The Pareto method analyzes the influence of the low spatial resolution of the data set based on the accuracy of the end product. It is an original way to analyze the trade-off between omission and commission errors when assessing the accuracy of a low-resolution product compared to a high-resolution map used as a reference. A region of unreachable accuracy (delimited by the Pareto boundary) can be defined because, at a certain point, it is not possible to minimize both omission and commission errors at the same time. In our case, the Pareto boundary was generated using the SPOT reference data set for the six validation sites. The distance between the MODIS products and the boundary is an indicator of the performance of the classification algorithms. For MODIS V05, because no crop area fraction was given in the definition of the “mosaic” class (cropland and natural vegetation), we tested the Pareto method for two values of 50% (MODIS V05) and 100% (MODIS V05*).

The Pareto boundary is linked with landscape fragmentation. A large number of metrics and indices have been developed to characterize landscape composition and configuration based on categorical map patterns (McGarigal et al., 2009). In this study, we calculated the mean patch size (MPS) (Saura, 2004) of the crop class based on the SPOT 2.5 m reference images.

3. Results

3.1. Stratification of Mali

The stratification obtained with the MODIS time series consisted of 13 homogeneous zones (Fig. 2a). Generally speaking,

segmentation followed the north–south bio-climatic gradient from zones with 1250 mm of annual rainfall in the south (zone 1) to the desertic zone (zone 13) in the north. To evaluate the reliability of the stratification, we compared our result with the existing IER_{strat}

stratification (Fig. 2b). For example, our zones 4, 6 and 7 perfectly reproduced the gradient of the Kita cotton zone. In this region, cultivation started only recently along a south-to-north pioneer front. Likewise, the IER_{strat} Sikasso maize and cotton zone matched

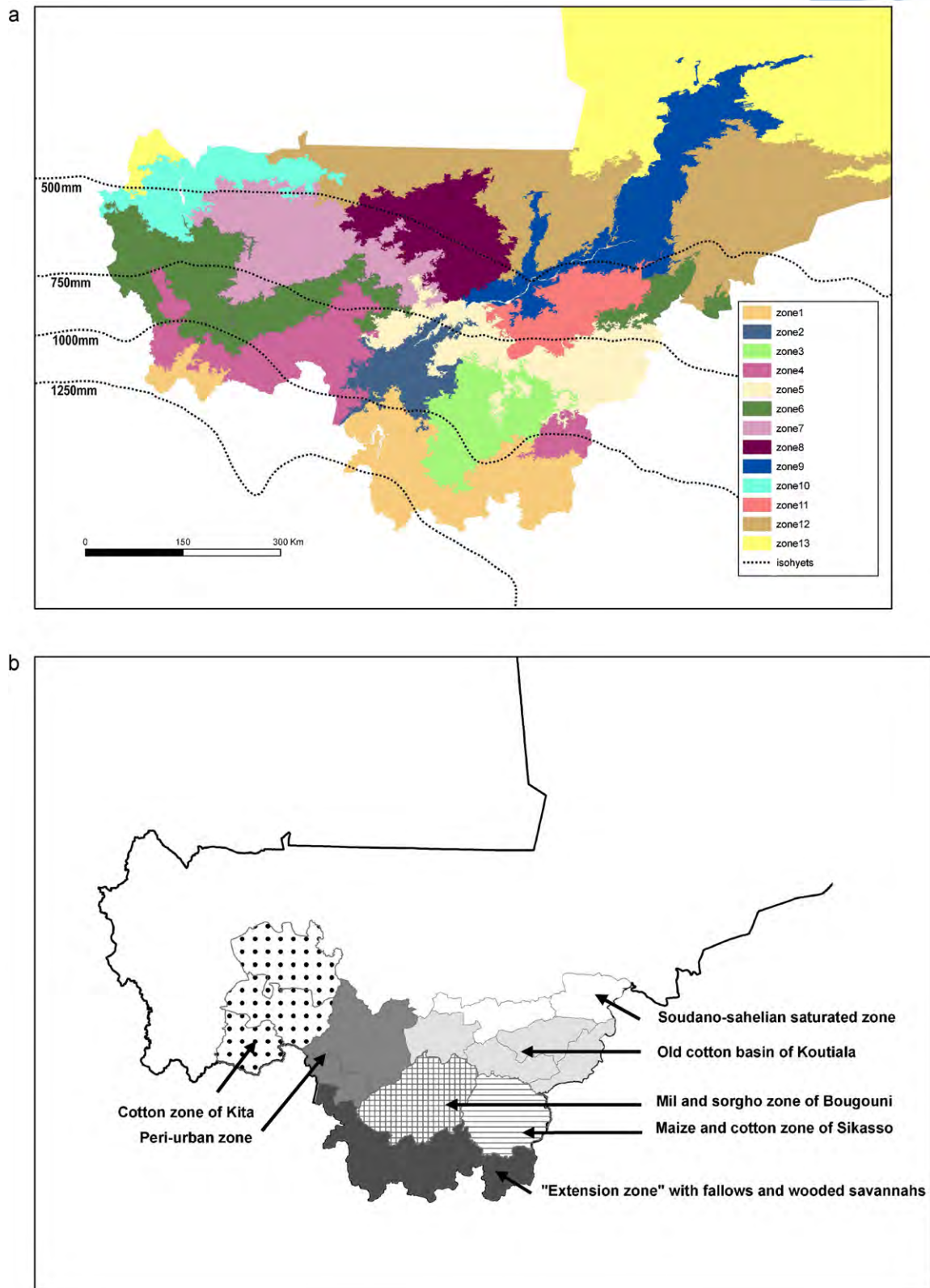


Fig. 2. Stratifications. (a) MODIS-derived landscape map of Mali (MODIS_{strat}) and (b) IER agricultural regions (IER_{strat}).

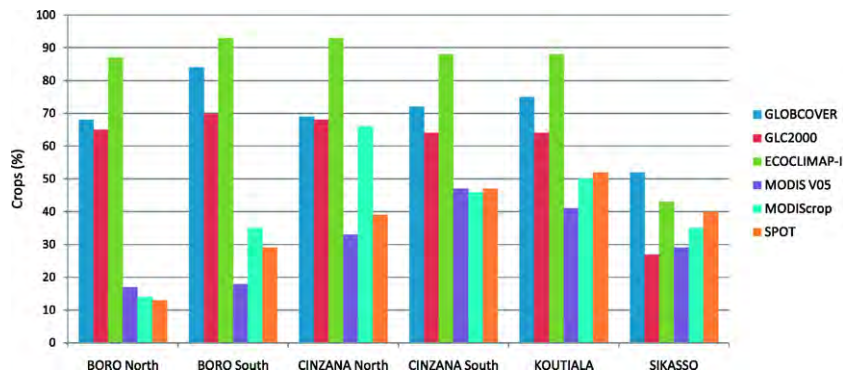


Fig. 3. Areal comparison of SPOT (reference data), MODIS_{crop} and crop maps of the validation sites produced by the global products.

our zone 4-east, and the IER_{strat} Bougouni cotton and mil/sorgho zone matched our zone 3, etc. These results show that the Malian landscapes in our stratification, which were characterized using radiometric and textural indices on MODIS images, are very similar to those in the IER stratification, which was mainly based on field measurements.

Furthermore, zones 1, 4 and 6 are not contiguous. This means that in Mali, some separate geographical areas may resemble each other from a radiometric and textural point of view. Actually, this is true of regions with the same amount of rainfall that are therefore likely to have similar landscapes because of similar temporal vegetation patterns, morpho-pedology and human activity.

3.2. Validation and comparison of products at the local scale

3.2.1. Areal validation and comparison

Compared to the SPOT derived reference data, the MODIS_{crop} map was quite successful in assessing the overall cultivated area of five out of the six validation sites (less than a 6% absolute difference) (Fig. 3). The Cinzana North site was the exception, here cultivated land was overestimated by MODIS_{crop} (65% versus 40% of cultivated area in the reference images). This zone is part of the Niger River zone (zone 9, Fig. 2a), and straddles different climatic regions, meaning it is complex and difficult to classify using coarse resolution images.

There was a marked discrepancy between the crop classes as assessed by the global products (Fig. 3). When using weighted crop classes (Table 3) and compared with the SPOT reference data, we found that the GLC2000, the ECOCLIMAP-II and the GLOBCOVER products overestimated the cultivated area (by about 2-fold). Only MODIS V05 showed a good level of accuracy, as previously mentioned (Giri et al., 2005; Herold et al., 2008; Ran et al., 2010) and was fairly close to the MODIS_{crop} cultivated map (less than 10% difference).

3.2.2. Validation and comparison of spatial distribution

For the six validation sites, error matrices were calculated between MODIS_{crop} and each of the four global products on one hand, and SPOT on the other, and shown in Table 4. The highest percentages of correctly classified pixels were obtained with MODIS_{crop} (70%) and MODIS V05 (58%). The percentage of correctly classified pixels for the GLC2000, the ECOCLIMAP-II and the GLOBCOVER ranged between 40% and 51%. In addition, MODIS_{crop} commission and omission errors were comparable for each class, implying that the cultivated area was estimated correctly, which is not the case for the other products.

MODIS_{crop} and MODIS V05 again showed a good level of spatial accuracy, whereas the GLC2000, the ECOCLIMAP-II and the GLOBCOVER products did not. Hence, for the rest of the analysis, comparison with global products focused mainly on MODIS V05.

We showed in Fig. 4 an interesting link between the detection of crops and the rural landscape fragmentation. There was a strong correlation between the producer accuracy of the MODIS_{crop} crop class and the crop mean patch size calculated on the 2.5-m resolution reference map ($R^2=0.8$, $N=6$, $P<0.025$ for One-Tailed Probabilities). The highest classification accuracies (>50%) occurred for sites with large cropped patches, such as the Cinzana and Koutiala (MPS > 200 ha). In contrast, sites with fragmented agricultural landscapes (MPS < 100 ha) like Boro were characterized by poor accuracy (< 40%).

The data were further analyzed using the Pareto boundary method (Fig. 5). The Pareto boundary of the most fragmented site (Boro) is above that of the other sites, and its unreachable region is considerably larger (Fig. 5). The Pareto boundary follows the 40% error isoline for Boro and the 20% isoline for the other sites. As a consequence, the accuracy levels that can be attained for any cultivated area map of the Boro region are considerably lower than those of the other sites.

Analysis of the Pareto regions enabled us to compare the MODIS_{crop} and MODIS V05 products (Fig. 5). MODIS_{crop} was found to be better than MODIS V05* for the Cinzana and Boro sites, and better than MODIS V05 for all sites. In other words, for all of our study sites, MODIS V05 and MODIS V05* were not better than MODIS_{crop}. Even if these products gave equivalent performances

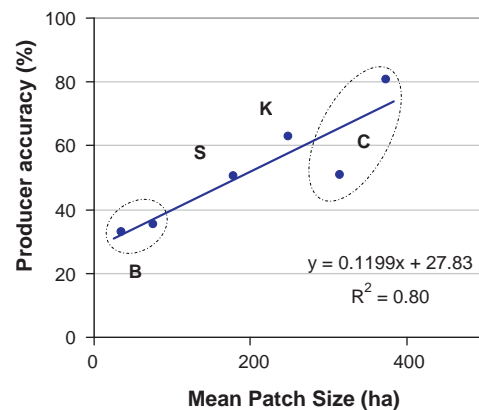


Fig. 4. Relations between the mean patch size index (MPS) of the crop class calculated on SPOT reference maps (C: Cinzana; B: Boro; K: Koutiala; S: Sikasso) and the corresponding producer accuracy calculated on MODIS_{crop}.

Table 4
Error matrices between MODIS_{crop} and SPOT, and between the four global products and SPOT for the six validation sites.

	MODIS _{crop}	MODIS V05*	MODIS V05	ECOCLIMAP-II*	ECOCLIMAP-II	GLC2000*	GLC2000	GLOBCOVER*	GLOBCOVER
Crop									
Omission error	37.1%	29.0%	60.3%	16.0%	17.5%	20.4%	22.6%	1.5%	24.5%
Commission error	39.0%	52.6%	51.8%	60.4%	58.9%	60.6%	57.9%	58.7%	56.9%
Non crop									
Omission error	25.6%	49.8%	27.0%	81.1%	74.6%	77.5%	67.3%	88.4%	63.0%
Commission error	24.0%	26.7%	34.3%	34.8%	30.3%	36.3%	30.4%	7.5%	29.5%
Fraction of correctly classified pixels	70.0%	58.3%	60.1%	41.8%	47.5%	44.6%	50.0%	45.2%	51.9%
F-score	0.615	0.568	0.436	0.538	0.548	0.527	0.545	0.580	0.548

MODIS V05*: with the mixed class considered as entirely under crops.

MODIS V05: with the mixed class weighted by 0.5.

ECOCLIMAP-II*: with the mixed class considered as entirely under crops.

ECOCLIMAP-II: with the mixed class weighted by 0.5.

GLC2000*: with the mixed class considered as entirely under crops.

GLC2000: with the mixed class weighted by 0.3.

GLOBCOVER*: with the mixed classes considered as entirely under crops.

GLOBCOVER: with the mixed classes weighted by 0.5.

F-score = $(2 \times UA \times PA) / (UA + PA)$, with user accuracy (UA) = 1 – omission error and producer accuracy (PA) = 1 – commission error.

in terms of areal accuracy (Fig. 3), MODIS_{crop} tended to minimize the omission–commission error in most of the sites and thus more accurate spatial distribution.

3.3. Comparison of the products at the national scale

3.3.1. Areal comparison

At national scale, the GLC2000 and GLOBCOVER global products gave a similar estimate of 13 Mha of cultivated land in Mali. MODIS V05 showed about 5 Mha under crops, whereas ECOCLIMAP-II product gave a very high estimate (18 Mha).

The MODIS_{crop} estimate of the total cropland area in Mali (11.3 Mha) was about twice that reported by FAO (4.9 Mha; FAOSTAT, 2007). The difference may be partly due to fallows: in the FAO definition of “arable land” temporary fallows are of less than five years while older fallows may have been included in our estimates. Furthermore, FAO statistics are computed from figures declared by a large number of different institutions. This complicates the use of a generic method, and suggests that because of harmonization problems, even official statistics may not be exempt of errors.

3.3.2. Comparison of spatial distribution

The MODIS_{crop} product in Fig. 6 shows the spatial distribution of cropland in Mali. As expected, the majority of the cultivated area is located in southeastern Mali.

Comparison of the MODIS V05* and MODIS_{crop} products (Fig. 6) revealed significant spatial disagreement at the national scale (77% when considering only the crop class, otherwise 25%). Several reasons may explain this lack of agreement. First, the images have different spatial resolutions. Second, the nomenclature differs between the products, with MODIS V05* including a mix crop/natural vegetation class. We tried to understand the difference between the two products by calling on external knowledge or data (visual verifications using Google Earth, Landsat images or local expertise). For example, the two large areas that appeared as crops only in the MODIS_{crop} product (see Fig. 7, circled green areas east of Segou and south-east of Bandiagara) were confirmed as being cultivated areas.

4. Discussion

Assessing the accuracy of coarse-resolution land cover products, such as our MODIS_{crop} map, is a critical and challenging task in West Africa because of landscape fragmentation and the medium

resolution of the sensor (Achard et al., 2001; Pittman et al., 2010). As an alternative approach to field surveys, six SPOT-derived cultivated area maps were used for the validation of the MODIS_{crop} map. We also compared our results to four recent global land products to determine their suitability in producing a crop map at national scale.

For crop areal estimations, the results of the MODIS_{crop} map were similar to those of the reference data on the validation sites, except for the Cinzana North site. This indicates that, even in regions where croplands are difficult to extract with reasonable accuracy using the MODIS 250-m satellite, a prior stratification of the territory and a large number of spectro-temporal clusters for the MODIS unsupervised classification can improve the classification.

Both MODIS_{crop} and MODIS V05 were found to perform well in terms of crop area estimation in the validation sites, whereas the GLC2000, GLOBCOVER and ECOCLIMAP-II estimations were considered out of range. The discrepancies between the global land cover products can be largely explained by differences in the resolution of the input data, as well as in the classification methods and above all by the discrepancies in the land cover nomenclature. Three of the four data sets use the FAO Land Cover Classification System (LCCS) (Di Gregorio, 2000), whereas the MODIS global land product uses the IGBP land cover classification. This problem of legend standardization has already been raised by other authors (e.g., Giri et al., 2005; Herold et al., 2006), and we consider this issue is more important than the interannual variation of cropland area, also mentioned by Fritz et al. (2010).

Considering the spatial distribution accuracy of the MODIS_{crop} product, our results indicate pixel-to-pixel crop accuracies of between 66% and 75%. These results can be explained by three main factors: (1) small cropland features that are spatially and/or spectrally inseparable from the surrounding non-crop areas in the MODIS 250-m NDVI data, as confirmed by Wardlow et al. (2007), (2) errors in photo-interpretation of the MODIS data set and hence misclassification of a few pixels, and (3) problems with the mixed classes in the nomenclature.

When the sites were compared, the crop class producer accuracy was inversely proportional to landscape fragmentation. Low accuracies were obtained for Boro where the cultivated area is highly fragmented (mean patch size < 80 ha), whereas good classification scores (>50%) were obtained for sites with large crop patches (>200 ha). The 200-ha patch size limit corresponds to a patch of 37 MODIS_{crop} pixels, which is a reasonable size for detection. When considering the other global products, this patch size

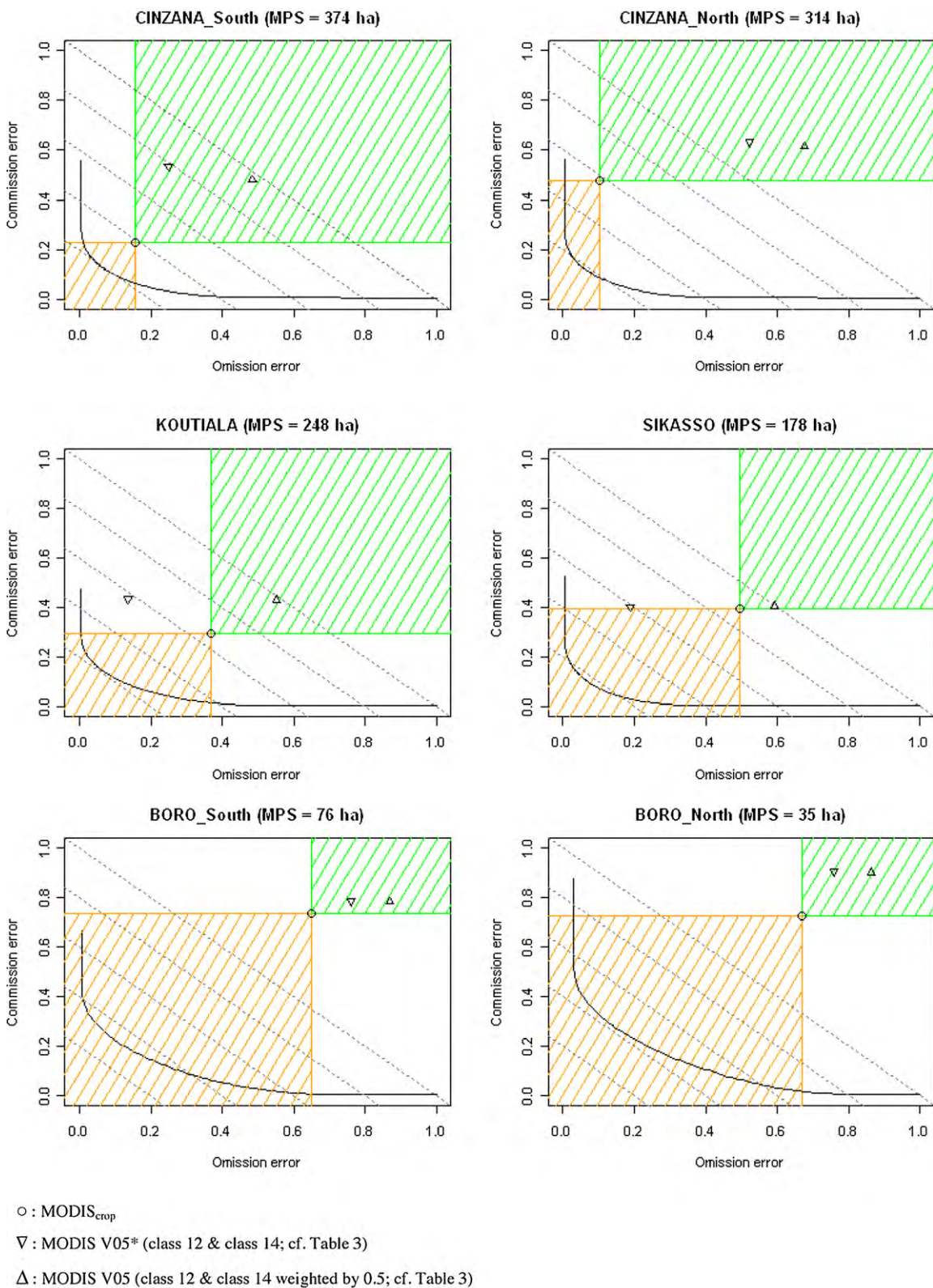


Fig. 5. Pareto figure calculated for a grid of 231.6 m × 231.6 m resolution with the reference cropped area derived from 2.5 m × 2.5 m SPOT images. The Pareto boundary (black line) represents the minimum level of omission and commission errors that could be theoretically obtained for a given landscape. Dotted lines correspond to isolines of omission/commission errors. The green area (upper right shaded area) is the region dominated by MODIS_{crop}; the orange region (lower left shaded area) is the region dominating MODIS_{crop}; the white area is the region of indifference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

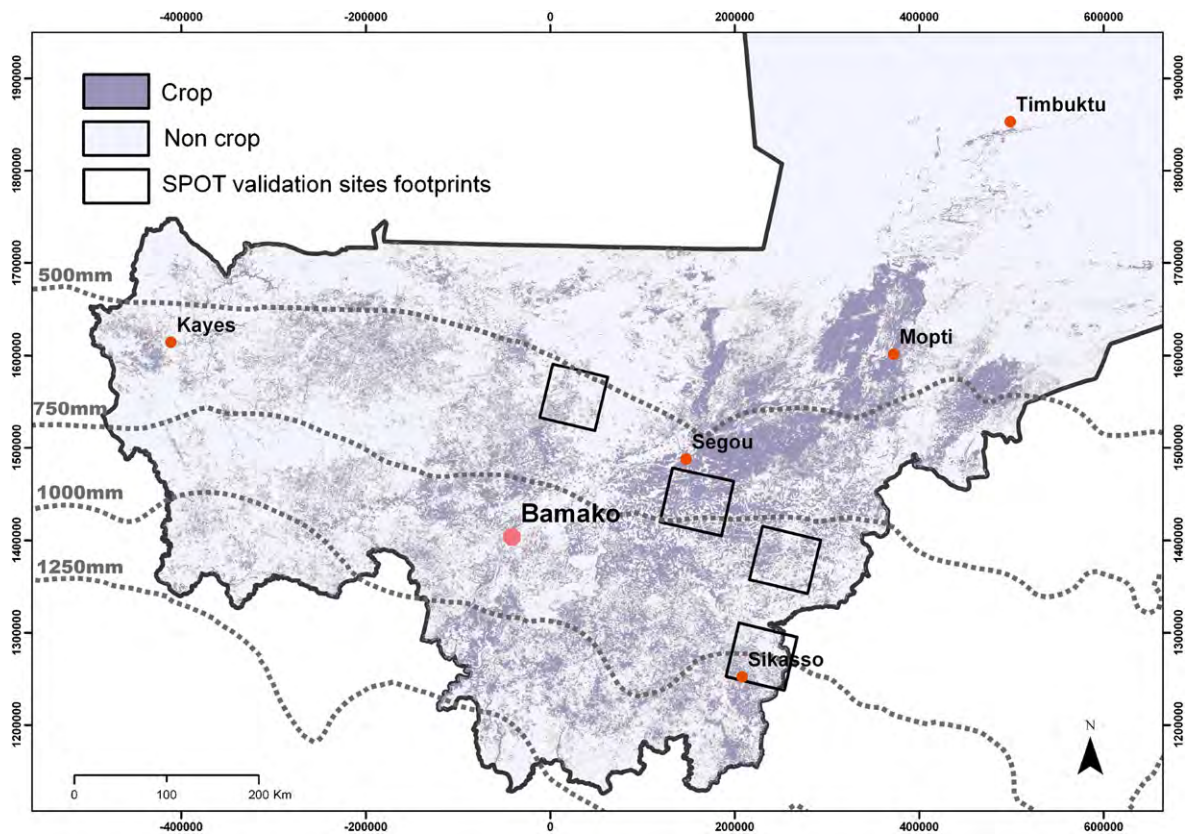


Fig. 6. Cultivated area in Mali estimated using stratified MODIS time series.

limit corresponds to two (AVHRR or VEGETATION) to eight pixels (MODIS V05), which consequently makes the patch much more difficult to detect and map. It is interesting that Doraiswamy et al. (2004) reported that MODIS 250 m resolution images are

adequate to monitor field sizes larger than 25 ha. In the fragmented landscapes of West Africa, the crop patch needs to be 8 times larger to be detected by MODIS NDVI time series. Our results corroborate earlier findings related to similar comparisons in a study conducted

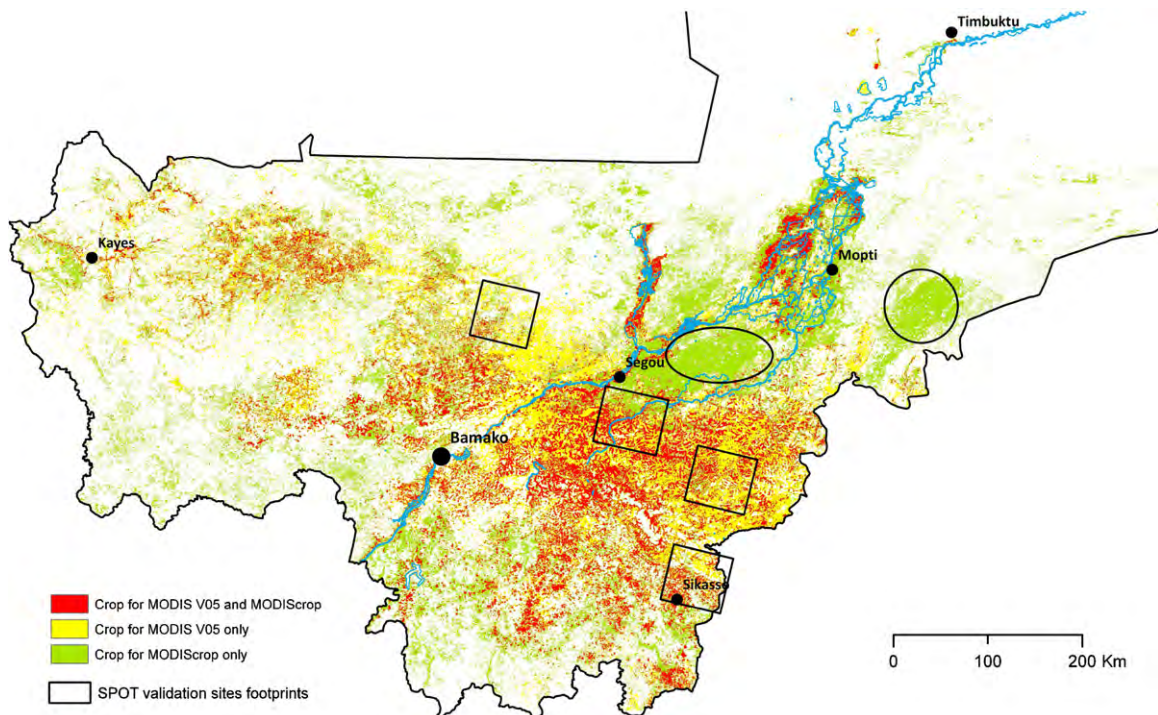


Fig. 7. Spatial comparison of cultivated area on MODIS V05* (class 12 and class 14, cf. Table 3) and MODIS_{crop} products. Green ellipsoids are discussed in the text. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

by [Mayaux and Lambin \(1995\)](#), who analyzed the effects of spatial aggregation on the estimation of tropical forest area derived from NOAA-AVHRR. Similarly, the study by [Smith et al. \(2003\)](#) on the impact of landscape characteristics on classification accuracy highlighted the impact of land-cover heterogeneity and patch size on classification accuracy, with the probability of correct classification decreasing with decreasing patch size and increasing heterogeneity.

These reasons may explain the discrepancies between MODIS_{crop} and MODIS V05. For example in Southern Mali, we detected crops that MODIS V05 did not, probably because the patches in this zone are too small to be detected. Furthermore, [Ruelland et al. \(2010\)](#) showed that in the Sudanian area, croplands are not stable over time. Over the 40-year period study by these authors, this zone underwent alternating periods of exploitation and recolonization by natural vegetation. Thus, it is difficult to determine if the differences between the two products are due to different image acquisition dates or not.

Apart from the rural landscape fragmentation aspect, differentiation between rainfed cropland and natural vegetation is expected to be difficult because of their almost synchronized phenology. However, depending on the region, they do display different levels of photosynthetic activity that also translate into differences in vegetation indices. It has been shown by [Bégué et al. \(2011\)](#) that in the Sahelian region, cropped areas have a higher annual NDVI than savanna areas (but lower values than forested areas, as expected). Conversely, the cropped areas in the Sudanian and Sudano-Guinean regions always have lower annual NDVI values than the natural vegetation (savannas and forests). Thus, differentiation between cropland and natural vegetation does not rely only on temporal differences in phenology, but more so on NDVI value levels. The good areal and spatial accuracies of MODIS_{crop} seem to indicate that our methodology using stratified MODIS NDVI time series does overcome this problem of synchronized phenology of crop and natural vegetation.

The use of the Pareto boundary helped to separate the sources of errors between the performance of the classification algorithm and the resolution of the remotely sensed data used in the classification procedure. Starting from a high-resolution reference data set, it is possible to determine the maximum user and producer accuracy values (i.e., minimum omission and commission errors) that could be jointly attained by a low-resolution map. Compared to the MODIS V05 product, which shows equivalent performances in terms of areal accuracy, MODIS_{crop} minimized the omission-commission error in most of the sites, which is particularly useful when dealing with agricultural statistics.

Further validation would be necessary, but that would require additional data. New initiatives based on collaborative networks are starting to emerge that indicates enormous potential for land cover validation. A first such initiative is the Geo-Wiki Project, where volunteers are asked to review hotspot maps of global land cover mismatch and determine, based on what they actually see in Google Earth and on their knowledge of local situations, whether the land cover maps are correct or not. Inputs are recorded in a database, along with uploaded photos, to be used in the future for the creation of a new and improved hybrid global land cover map ([Fritz et al., 2009, 2011](#)). Our work would in turn benefit from the additional validation data available there.

5. Conclusion

In this paper, we examined the capability of 250 m MODIS data for cropland mapping in Mali, a country that is representative of the Sahel belt of Africa. We first presented a method to map cultivated areas at a national scale using stratified MODIS NDVI

time series. We then compared the resulting crop map with the crop classes of existing global land products, in terms of areal estimation and spatial distribution accuracy. This was done using SPOT image classifications as the reference data set. It was found that the MODIS_{crop} product rated better than other recent global products.

As increasingly more environmental information is needed to develop sustainable policies, our study shows that the time series of moderate resolution imagery contain pertinent and accessible information that can be used in agricultural monitoring. We showed that MODIS time series can separate crop from non crop at national scale with acceptable accuracy in countries with fragmented agricultural landscapes like those of West Africa, using a prior stratification of the territory and a large number of spectro-temporal clusters for the MODIS unsupervised classification. The method could be applied to other West African countries such as Senegal, Niger or Burkina Faso.

The flexible approach described in this paper is particularly suitable for developing countries, where rapidly changing ecosystem processes require rapid updating of environmental information and where economic and technological constraints are high. The method presented requires only basic or widely used techniques (maximum likelihood was used as classification method and segmentation was used for the stratification, free softwares offer processing capabilities for both) and expertise for visual interpretation. Ground expertise is certainly an important and irreplaceable source of knowledge, especially in developing countries, and it would be a mistake not to include it during image data processing and analysis.

The next step will be to better characterize the cultivated area to produce maps of crop production systems from spatial, temporal, textural and morphological indicators. These maps could also be used as inputs for crop models to forecast yield.

Acknowledgments

This work was supported by the EEC (AMMA project) and CIRAD funds (E. Vintrou's fellowship). The four SPOT 5 images of 2007 were acquired with the support of the ISIS program (CNES). The authors would like to thank Michel Vaksman, Amadou Diarra and Maxime Claden for their help during field surveys and Mamy Soumaré for sharing his knowledge on local agriculture. We also thank Arnel Kaptué for his advice and suggestions on the use of ECOCLIMAP-II for AMMA zone.

References

- Achard, F., Eva, H., Mayaux, P., 2001. Tropical forest mapping from coarse spatial resolution satellite data: production and accuracy assessment issues. *International Journal of Remote Sensing* 22, 2741–2762.
- Arino, O., Bicheron, P., Achard, F., Latham, J., Witt, R., Weber, J.L., 2008. GLOBCOVER The most detailed portrait of Earth. *Esa Bulletin-European Space Agency* (136), 24–31.
- Bartholomé, E., Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing* 26, 1959–1977.
- Bastin, G.N., Pickup, G., Pearce, G., 1995. Utility of AVHRR data for land degradation assessment: a case study. *International Journal of Remote Sensing* 16, 651–672.
- Bégué, A., Vintrou, E., Ruelland, D., Claden, M., Dessay, N., 2011. Can a 25-year trend in Soudano-Sahelian vegetation dynamics be interpreted in terms of land use change? A remote sensing approach. *Global Environmental Change* 21 (2), 413–420.
- Blaschke, T., 2006. The role of the spatial dimension within the framework of sustainable landscapes and natural capital. *Landscape and Urban Planning* 75, 198–226.
- Boles, S.H., Xiao, X., Liu, J., Zhang, Q., Munkhtuya, S., Chen, S., Ojima, D., 2004. Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data. *Remote Sensing of Environment* 90, 477–489.
- Boschetti, L., Flasse, S.P., Brivio, P.A., 2004. Analysis of the conflict between omission and commission in low spatial resolution dichotomic thematic products: the Pareto boundary. *Remote Sensing of Environment* 91, 280–292.

- Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling* 168, 233–249.
- Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sensing of Environment* 112, 986–997.
- Defries, R.S., Townshend, J.R.G., 1994. NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing* 15, 3567–3586.
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS. *Remote Sensing of Environment* 92, 548–559.
- FAO, 2007. FAOSTAT. Available on line at <http://faostat.fao.org/>.
- Farina, A., 2000. The cultural landscape as a model for the integration of ecology and economics. *Bioscience* 50, 313–320.
- Forman, R.T.T., 1995. Some general-principles of landscape and regional ecology. *Landscape Ecology* 10, 133–142.
- Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* 114, 168–182.
- Fritz, S., Bartholome, E., Belward, A., Hartley, A., Stibig, H.-J., Eva, H., et al., 2003. Harmonization, Mosaicking, and Production of the Global Land Cover 2000 Database. Joint Research Center (JRC), Ispra, Italy.
- Fritz, S., See, L., 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology* 14, 1057–1075.
- Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F.d.r., Kraxner, F., Obersteiner, M., 2009. Geo-Wiki.Org: the use of Crowdsourcing to improve global land cover. *Remote Sensing* 1, 345–354.
- Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *International Journal of Remote Sensing* 31, 2237–2256.
- Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu, J., Hansen, M., Obersteiner, M., 2011. Cropland for sub-Saharan Africa: a synergistic approach using five land cover data sets. *Geophysical Research Letters* 38, L04404.
- Giri, C., Zhu, Z., Reed, B., 2005. A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets. *Remote Sensing of Environment* 94, 123–132.
- Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing* 21, 1331–1364.
- Hansen, M.C., Reed, B., 2000. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing* 21, 1365–1373.
- Haralick, R.M., 1979. Statistical and structural approaches to texture. *Proceedings of the IEEE* 67, 786–804.
- Herold, M., Woodcock, C.E., di Gregorio, A., Mayaux, P., Belward, A.S., Latham, J., Schmullius, C.C., 2006. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing* 44, 1719–1727.
- Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., Schmullius, C., 2008. Some challenges in global land cover mapping: an assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment* 112, 2538–2556.
- Hountondji, Y.C., Sokpon, N., Ozer, P., 2006. Analysis of the vegetation trends using low resolution remote sensing data in Burkina Faso (1982–1999) for the monitoring of desertification. *International Journal of Remote Sensing* 27, 871–884.
- Husak, G.J., Marshall, M.T., Michaelsen, J., Pedreros, D., Funk, C., Galu, G., 2008. Crop area estimation using high and medium resolution satellite imagery in areas with complex topography. *Journal of Geophysical Research-Atmospheres* 113, D14112, doi:10.1029/2007JD009175.
- Justice, C.O., Townshend, J.R.G., Holben, B.N., Tucker, C.J., 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* 6, 1271–1318.
- Kaptué Tchuenté, A.T., Roujean, J.-L., Faroux, S., 2010. ECOCLIMAP-II: an ecosystem classification and land surface parameters database of Western Africa at 1 km resolution for the African Monsoon Multidisciplinary Analysis (AMMA) project. *Remote Sensing of Environment* 114, 961–976.
- Kaptué Tchuenté, A.T., Roujean, J.L., De Jong, S.M., 2011. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation* 13, 207–219.
- Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment* 93, 412–422.
- Loveland, T.R., Belward, A.S., 1997. IGBP-DIS global 1 km land cover data set, DISCover: First results. *International Journal of Remote Sensing* 18, 3289–3295.
- Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, Z., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 21, 1303–1330.
- Lowry, J., Ramsey, R.D., Thomas, K., Schrupp, D., Sajwaj, T., Kirby, J., Waller, E., Schrader, S., Falzarano, S., Langs, L., Manis, G., Wallace, C., Schulz, K., Comer, P., Pohns, K., Rieth, W., Velasquez, C., Wolk, B., Kepner, W., Boykin, K., O'Brien, L., Bradford, D., Thompson, B., Prior-Magee, J., 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: a case study of the Southwest Regional Gap Analysis Project (SWReGAP). *Remote Sensing of Environment* 108, 59–73.
- Mayaux, P., Lambin, E.F., 1995. Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation. *Remote Sensing of Environment* 53, 1–15.
- Mayaux, P., Bartholomé, E., Fritz, S., Belward, A., 2004. A new land cover map of Africa for the year 2000. *Journal of Biogeography* 31, 861–877.
- McCallum, I., Obersteiner, M., Nilsson, S., Shvidenko, A., 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation* 8, 246–255.
- McGarigal, K., Tagil, S., Cushman, S.A., 2009. Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology* 24, 433–450.
- Mora, F., Iverson, L.R., 1997. Dynamic stratification of the landscape of Mexico: analysis of vegetation patterns observed with multitemporal remotely sensed images. *Geocarto International* 12, 73–87.
- Nelson, R., Horning, N., 1993. AVHRR-LAC estimates of forest area in Madagascar, 1990. *International Journal of Remote Sensing* 14, 1463–1475.
- Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. *Remote Sensing* 2, 1844–1863.
- Ran, Y.H., Li, X., Lu, L., 2010. Evaluation of four remote sensing based land cover products over China. *International Journal of Remote Sensing* 31, 391–401.
- Roller, N.E.G., Colwell, J.E., 1986. Coarse-resolution satellite data for ecological surveys. *Bioscience* 36, 468–475.
- Ruelland, D., Levvasseur, F., Tribotté, A., 2010. Patterns and dynamics of land-cover changes since the 1960s over three experimental areas in Mali. *International Journal of Applied Earth Observation and Geoinformation* 12, S11–S17.
- Saura, S., 2004. Effects of remote sensor spatial resolution and data aggregation on selected fragmentation indices. *Landscape Ecology* 19, 197–209.
- Smith, J.H., Stehman, S.V., Wickham, J.D., Yang, L., 2003. Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment* 84, 342–349.
- Soumare, M., 2008. Dynamique et durabilité des systèmes agricoles à base de coton au Mali. Thèse de Doctorat en Géographie Humaine, Economique et Régionale, Université de Paris X Nanterre, p. 373.
- Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment* 108, 290–310.

ARTICLE 2 :

**Mapping Fragmented Agricultural Systems
in the Sudano-Sahelian region using Random Forest and
Metrics of Coarse Resolution MODIS Imagery**

Photogrammetric Engineering and Remote Sensing, in press.

Mapping Fragmented Agricultural Systems in the Sudano-Sahelian environments of Africa using Random Forest and Ensemble Metrics of Coarse Resolution MODIS Imagery

Elodie Vintrou ^{a*}, Mamy Soumaré ^b, Simon Bernard ^c, Agnès Bégué ^a, Christian Baron ^a,
Danny Lo Seen ^a

^a CIRAD UMR TETIS, Maison de la Télédétection, 500 rue J.F. Breton, Montpellier, F-34093
France

^b IER, Rue Mohamed V, BP 258, Bamako, Mali

^c University of Rouen, LITIS EA 4108, BP 12 - 76801 Saint-Etienne du Rouvray, France

* Corresponding author. Tel.: +33 4 67 54 87 54; e-mail address: elodie.vintrou@cirad.fr

Abstract

We worked on the assumption that, agricultural systems shaped the landscape through human cropping practices, and that the resulting landscape can be described with a set of coarse resolution satellite-derived metrics (spectral, textural, temporal, and spatial metrics). A Random Forest classification model was developed at the village scale in South Mali, based on 100 samples, with data on the main type of agricultural system in each village (3-class typology), and 30 MODIS-derived and socio-environmental metrics calculated on agricultural areas. The model was found to perform well (overall accuracy of 60%) and was stable. Class A (food crops) and B (intensive agriculture) displayed good producer's accuracy (70% and 67% respectively), while class C (mixed agriculture) was less accurate (50%). The most important metrics were shown to be the annual mean of NDVI, followed by the phenology transition dates and texture metrics. However, when considering each set of metrics separately, texture emerged as the most discriminating factor (with 53% of global accuracy). This result, i.e., that even coarse resolution imagery contains textural information that can be used for crop mapping, is new. Such maps could be used in food security systems as an indicator of system vulnerability, or as spatial inputs for crop yield models.

Introduction

Sub-Saharan Africa is the only region in the world where per capita food production is declining, making continued investment in monitoring and aid program mandatory (Brown *et al.*, 2007). Different early warning systems, such the Famine Early Warning System Network (FEWS NET - USAID) or the Global Information and Early Warning System (GIEWS - FAO), use methods that can identify food security hazards early enough so that political and budgetary decisions can be made in a timely manner.

Key deliverables of food security systems in terms of crop monitoring consist of early estimates of cultivated area and crop-type distribution, cropping practices (cropping intensity, crop calendar, water and fertilizer inputs), detection of growth anomalies, and crop yield estimates. Remote sensing plays an important role in delivering information for crop monitoring in insecure countries (e.g. Hutchinson, 1991; Brown, 2008; Thenkabail *et al.*, 2009a; Thenkabail *et al.*, 2011; Thenkabail *et al.*, 2009c). Satellites help locate and estimate areas of crop types and crop potential yield at local scales (e.g. Bruzzone *et al.*, 1997). However, remote sensing techniques face numerous challenges for crop mapping of larger areas such as the continents and the globe (Thenkabail, 2010). Global monitoring typically resorts to using free low- and medium-resolution images (Hountondji *et al.*, 2006; Justice *et al.*, 1985), but these images do not enable precise localization of the crops. This is particularly true in regions with fragmented cropped areas, like West Africa, where 20% to 40% of crop classification error using MODIS is inherent to the structure of the landscape (Vintrou *et al.*, 2011). One way to get around this problem is to develop sub-pixel approaches (e.g. Ichokua and Karnieli, 1996) or to increase the scale of the observed objects.

Many papers deal with the sub-pixel approach, which gives correct results in the case of simple landscapes (e.g. De Fries *et al.*, 1997), but is not ideal for the complex landscapes of West Africa. In the Sudano-Sahelian region, peasant farming is the dominant production system with low cropland intensification, small fields, and marked between-field and within-field variability of vegetation. The presence of trees in the field and the rainfall regime make it even more difficult to distinguish between the natural vegetation and the cropped areas (Fritz and See, 2008).

In this context, we tested an approach that consisted in changing the observation scale, and consequently in changing the nature of the observed object. We chose to work at a local scale - a village -, and with one object - the agricultural system- compatible with the existing coarse resolution imagery. The agricultural system is described by cropping practices (cultivated species and level of intensification) and can be used as the basis for crop monitoring at large scales. There is a real need for better discrimination and identification of the main agricultural

systems in many regions of the world, because they indicate how the system will interact with the local environment and water conditions and thus influence yields and productions. These interactions can be estimated empirically or by using agro-meteorological models, so as to propose forecast scenarios of crop yield estimates.

Thus, in the context of fragmented landscapes, and due to the absence of adequate Earth Observing Systems, the objective of this study was to develop a new approach to classifying and mapping the main agricultural systems in West Africa using coarse resolution imagery. The method is based on the general assumption that human practices shape the landscape. We adapted this assumption to the cultivated domain by claiming that agricultural systems shape the cultivated part of the landscape, and as a result, that these systems could be captured through a set of landscape attributes, such as i) vegetation biomass, ii) organization of the landscape components, and iii) vegetation seasonality (phenology and crop calendar). These attributes are known to be accessible using time series of multispectral images. Among these characteristics, biomass estimation is certainly the most frequently investigated (e.g. Hutchinson, 1991; Prince and Goward, 1995). Concerning landscape organization, classical pixel-based classification methods do not make sufficient use of spatial concepts of neighborhood, proximity or homogeneity (Burnett and Blaschke, 2003). A number of authors used texture analysis to classify high and very high resolution images (e.g. Asner *et al.*, 2002; Kayitakire *et al.*, 2006; Peddle and Franklin, 1991), but very few studied possibility of using the texture of coarse resolution images in land classification (see Tsaneva *et al.*, 2010) for one example). Concerning vegetation seasonality, recent multitemporal studies demonstrated that MODIS data can be used to detect unique multi-temporal signatures for each of the major crop types (e.g. Wardlow *et al.*, 2007), to map double cropping (e.g. Arvor *et al.*, 2012), or to map irrigated areas (Thenkabail *et al.*, 2009b).

To investigate the applicability of moderate-resolution imagery for the classification of agricultural systems, we chose to use the example of Mali and, on the crop mask, calculated spectral, spatial, textural, and temporal metrics derived from MODIS time-series images. We used a training data set of observations collected in 100 villages in Mali and the remotely-sensed metrics associated with each village to build a Random Forest (RF) statistical model, capable of predicting the main type of agricultural system in every village in South Mali. Our paper is organized as follows: after a brief description of the study area, we present the training data set and the RF algorithm we built. We then investigate the importance of satellite-derived metrics for mapping agricultural systems, and present the resulting classification of the villages of South Mali. We conclude by discussing the accuracy and relevance of the method and the possible use of the resulting map in food security systems.

Study area and data sets

Study area

Mali is a land-locked country covering of 1,240,192 km² in West Africa between latitudes 10°N and 24°N (Figure 1). The topography consists mostly of low plateaus and basins with occasional rocky hills. The plateaus never exceed 300 or 400 m. Mali has a latitudinal climatic gradient ranging from sub-humid to semi-arid that, further north, extends to arid and desertic. Like other West African countries along the same latitudinal belt, food security relies on an adequate supply of rainfall during the monsoon season. Mali encompasses the Sudano-Sahelian zone, where strong dependence on rain-fed agriculture implies exposure to climate variability, in addition to the impacts population growth have on food security. Farming is the main source of income for many people in the Sudano-Sahelian region, where millet and sorghum are the main food crops. The vast majority of the population (80%) consists of subsistence farmers. A few larger farms produce crops for sale (cash crops), mainly cotton and peanuts. In this study, we did not include the Saharan zone in the north of the country where the sparse rainfall amounts to less than 200 mm per year.

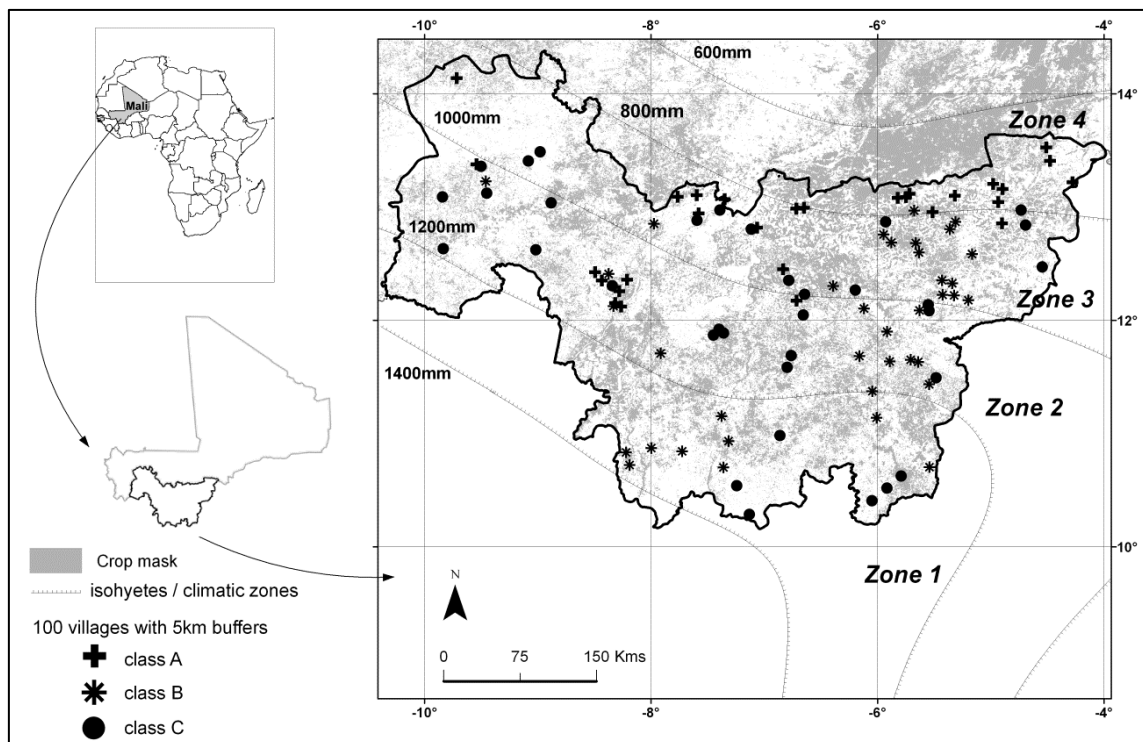


Figure 1. Map of Mali showing the geographic location of the 100 villages sampled along the rainfall gradient (1971-2000 year average).

Ground data

In this study, we used data gathered during a field survey on farmers' practices conducted in 2001-2004 by CMDT (*Compagnie Malienne Du Textile*) and IER (*Institut d'Economie Rurale*, Bamako) in South Mali (Figure 1). The survey was conducted on farms sampled in 100 villages. The sampling and survey protocols are detailed in Soumare, 2008). The villages to be surveyed were chosen to be spatially distributed in the Malian cotton belt and representative of the existing agricultural systems.

Table 1. Percentage area of each species cropped in a village, and the associated type of “agricultural system”. These examples are extracted from the CMDT/IER data set (Soumare, 2008).

Village name	Cotton (%)	Maize (%)	Millet (%)	Sorghum (%)	Type
Boukoro	0	0	64	13	A: Food crops millet + sorghum > 50% cotton < 10%
Dokolo	1	1	56	17	
Missango	2	2	19	66	
Neguena	9	6	58	16	
Ntieresso	0	3	49	26	
...	
Bamadougou	14	39	6	5	B: Intensive crops maize and cotton > 40%
Bladje	33	24	3	11	
Dougoumouso	49	17	14	16	
Doumanani	33	25	14	15	
Kenieroba	25	28	16	15	
...	
Doma	21	12	17	36	C: Mixed crops sorghum > 20% 5% < cotton < 40%
Fiena	10	4	51	21	
Koro-Koro	22	6	34	29	
Sinkolo	28	9	23	33	
Sinsin	26	9	14	34	
...	

We used the survey data to build a 3-class typology of the villages based on the agricultural system. The typology was created using expert knowledge and considering the main crop species cultivated in the village and the intensification of production (use of fertilizers, agricultural equipment, livestock, etc.). Each village was assigned to a class A, B or C (Table 1):

- Class A villages represent 29% of the data set. Class A include farms whose area under millet and sorghum represents more than 50% of their total cropped land, plus a few plots under cotton (<10%). This class groups the millet and sorghum-based agricultural systems, which are mostly located in the Sahelian zone, where the monsoon season is shorter. These

crops are considered to be of low value, i.e., in general, the little income these hard grains are able to generate does not allow farmers to invest in fertilizers. Class A corresponds to “food-producing agriculture”.

- Class B villages represent 37% of the data set, and include farms whose land under maize and cotton represents more than 40% of all their cropped land. Maize and cotton are common cash crops in Mali, and are usually found in the same agricultural system since they demand similar levels of rainfall and soil nutrients (Coulibaly, 1995). Cotton requires a crop rotation to maintain an adequate soil nutrient balance and to minimize pest pressure. Class B is referred hereafter as the “intensive agriculture” class.
- Class C villages represent 34% of the data set, and have less intensive practices, as they cultivate both coarse grain (sorghum) and a cash crop (cotton). Class C represents “mixed agriculture”, neither completely dedicated to grains, nor to cotton.

The crop mask

The crop mask used in this study is described in detail in Vintrou *et al.*, 2011) (Figure 1), who classified MODIS time series in two steps: 1) the Malian territory was stratified into homogeneous zones based on the radiometry and texture of the MODIS time series images, and 2) the cropped/non-cropped areas in each zone were classified separately, after which the areas were merged at the national scale. The accuracy of the resulting map ($MODIS_{crop}$), and how it compares with existing coarse-resolution global land products (GLC2000 Africa, GLOBCOVER, MODIS Land Cover V05 and ECOCLIMAP-II), was assessed along a climatic gradient against six crop/non-crop maps derived from SPOT 2.5 m resolution images used as references. The fraction of correctly classified pixels for $MODIS_{crop}$ was 70%, compared to 58% for the MODIS Land Cover product, while for the other global products, it ranged between 42% and 52%.

Satellite-derived metrics

The satellite-derived metrics were extracted from MODIS products acquired from the NASA Land Process Distributed Active Archive Centre (LP DAAC). Products concerning Mali were downloaded for the year 2007 (to be consistent with the crop mask year).

Spectral (NDVI) metrics

The spectral metrics were mainly based on the Normalized Difference Vegetation Index (NDVI), as this spectral index is one of the most successful indices to simply and quickly

identify vegetated areas and their “condition”. We used NDVI values from the *Vegetation Indices 16-Day L3 Global 250 m SIN Grid* product (MOD13Q1 V05) to calculate the spectral metrics on the crop mask. For Mali, a set of 12 MODIS NDVI 16-day composite images was acquired for 2007 (the best quality composite image out of two for each month was kept). Four NDVI-derived metrics were selected for the spectral metrics: the annual maximum, annual mean, the annual amplitude, and the seasonal mean (May-November).

Texture metrics

The texture metrics were computed to characterize variations in spatial data within an image. Two texture indices (variance and skewness) were calculated on the crop mask using the 2007 MOD13Q1 product for three different periods in the growing season (March, June, September). The generation of these indices was based on different orientations of pairs of pixels, with a specific angle (horizontal, diagonal, vertical, co-diagonal) and distance, called “patterns”. In our approach, we empirically determined a pattern size of 7 MODIS pixels (around 300 ha), which is close to the upper limit of the cropped patch size in South Mali (Vintrou *et al.*, 2011).

The “variance” index is the variance value of the pixels inside the 7 x 7 pixels window, while the “skewness” index measures to what extent the data within the window are skewed towards the highest or the lowest values.

Phenology metrics

We used the V005 MODIS *Global Land Cover Dynamics* (MCD12Q2) product (informally called the MODIS Global Vegetation Phenology product). For each pixel, it provides estimates of the timing of vegetation phenology at the global scale. It identifies the growth, maturity, and senescence of the vegetation that characterize seasonal cycles. It also provides pixel information related to the range and summation of the Enhanced Vegetation Index (EVI) computed from MODIS surface reflectance data. Using the crop mask, we calculated seven metrics from the MCD12Q2 product: the date of greenness increase, the date of greenness maximum, the date of greenness decrease, the date of greenness minimum, the EVI value at the date of the greenness minimum and maximum, and the EVI area during greenness.

Spatial (fragmentation) metrics

The fragmentation of the rural landscape was evaluated using three spatial metrics calculated for a given portion of the territory: the fraction of cropped area, the number of cultivated patches, and the mean cultivated patch size (MPS). These metrics were calculated using FRAGSTATS 2.0. software.

Socio-environmental data

We used a database from the *Mission de Décentralisation et des Réformes Institutionnelles* in Mali (Source: MDRI, 1998), from which we extracted the geographic coordinates and the number of inhabitants per village of the 4000 villages in South Mali.

In addition, each village was affiliated to one of the four climatic regions presented in Figure 1, and the maximum and the mean of elevation of the cultivated domain in each village was calculated. Elevation data are based upon the SRTM (Shuttle Radar Topography Mission) digital elevation model, with a resolution of 90 m at the equator.

Data processing

Choice of the spatial unit

To combine ground and satellite-derived data, we had to choose a common scale. The field surveys were conducted in villages in South Mali (Figure 1). Consequently, we decided to represent the village unit by calculating a buffer zone of 5 km around each village center. In fact, the organization of cultivated land in Mali often results in concentric rings around homesteads, with different agricultural systems in each ring. However, we decided to consider only the main agricultural system in each village, because the spatial resolution did not allow us to consider detailed cropped types within the 5 km buffer zone. Furthermore, no village level administrative map was available, and preliminary tests using Thiessen polygons as village spatial units revealed artifacts in areas where villages are located far away from one another.

Next, to characterize the agricultural systems at the village scale, the mean of the different metrics presented in Table 2 was calculated for the 5-km buffer zones inside the 2007 crop mask.

Table 2. MODIS-derived and socio-environmental metrics used to train the Random Forest model

NDVI metrics	annual maximum annual mean seasonal mean (May-November) annual amplitude
TEXTURE metrics ⁽¹⁾	maximum variance 1 mean variance 1 maximum variance 2 mean variance 2 maximum variance 3 mean variance 3 maximum skewness 1 mean skewness 1 maximum skewness 2 mean skewness 2 maximum skewness 3 mean skewness 3
PHENOLOGY metrics	date of greenness increase date of greenness maximum date of greenness decrease date of greenness minimum EVI value at the date of the greenness minimum EVI value at the date of the greenness maximum EVI area during greenness
FRAGMENTATION metrics	% of cultivated area number of cultivated patches mean patch size
OTHER metrics	maximum altitude mean altitude climatic zone population

⁽¹⁾ 1 = March; 2 = June; 3 = September.

Random Forest training and analysis

Random Forest (RF) is an ensemble learning technique that generates many classification trees that are then aggregated to compute a classification (Breiman, 2001). It uses a bagging-based approach (random sampling with replacement) to build a forest of classification trees. Each classification tree (~500 trees in a typical forest) is constructed from a randomly sampled set consisting of ~63% of the full data (Cutler *et al.*, 2007). Observations in the original data set that do not occur in a bootstrap sample are called out-of-bag observations (OOB). In land cover studies, the RF classifier has been found to be stable and relatively fast, involves few user-defined parameters and yields overall accuracies that are either comparable to or better than other classifiers, such as maximum likelihood and conventional decision trees (Lawrence *et al.*, 2006; Na *et al.*, 2010), AdaBoost decision trees and neural networks (Chan and Paelinckx, 2008), or support vector machines (Pal, 2005).

Random Forest training

We implemented the RF classifier using the R-software package “randomForest” (Liaw and Wiener, 2002; Liaw and Wiener, 2008). The database is made of 100 villages, characterized by both the type of agricultural system and the 30 remote-sensing metrics and socio-environmental data listed in Table 2.

The RF algorithm has only two user-defined parameters. The number of trees to build (*ntree*) was set to 1000, and the number of descriptors that are randomly selected to split each node (*mtry*) was set to 6. The number of trees was chosen to be sufficiently large so that the OOB error was able to stabilize. There is no penalty for having “too many” trees, other than waste of computational resources, in contrast to other algorithms which require a stopping rule. This RF property was mathematically proven in Breiman, 2001) and experimentally confirmed in two other studies (Bernard *et al.*, 2009; Latinne *et al.*, 2001). The number of descriptors ranges from 1 to p , p being the total number of descriptors. We chose to set it equal to the square root of the number of descriptors, as recommended by Bernard (2009).

Classification accuracy

The classification performance was evaluated on the data set used for training. In Random Forests, there is no need for cross validation or for a separate test set to obtain an unbiased estimate of the test set error (Lawrence *et al.*, 2006). Since Random Forest only uses roughly two-thirds of the training data to build the classification model, the remaining one-third of the training data (the out-of-bag samples) can be used to estimate the error of the predictions.

The performance of the classification is evaluated through (i) the confusion matrix, and (ii) the stability of the algorithm. Here we used both the global accuracy to measure the percentage of pixels correctly classified in the training set, the “producer's accuracy” of the class to indicate the probability of a reference pixel being correctly classified, and the “user's accuracy” of the class to indicate the probability that a pixel classified on the map actually represents that category on the ground (Congalton and Mead, 1983; Story and Congalton, 1986). We also calculated the F-score measure (Chinchor, 1992), evenly balanced user and producer accuracy, which quantifies the performance of the system for each class. The stability of the algorithm is measured by the standard variations in the global accuracies and the mean class accuracies calculated on Random Forest processes repeated 50 times.

Evaluation of the importance of the metrics

As a starting point, we were interested in the metrics that actually help distinguish different types of agricultural systems. To this end, a RF-based variable importance measure was used to evaluate the metrics, in order to eliminate metrics of minor importance for the accuracy of the classification. RF offers a number of internal measures to estimate the importance of the selected metrics for the accuracy of a given classification. The properties of these measures have been intensively studied in recent years, and the permutation importance is considered to be a computationally tractable choice in datasets with (possibly) correlated variables (Nicodemus *et al.*, 2010). The permutation importance, hereafter termed variable importance, is based on the decrease in classification accuracy when the values of a variable in a node of a tree are permuted randomly (see Stumpf and Kerle, 2011) for more details). In brief, the variable importance provides a measure of the usefulness of the different metrics in distinguishing between classes A, B and C.

We estimated each variable importance from 1000 Random Forest trees using the mean decrease in accuracy when permuted, and the number of times one variable was selected over another (Breiman, 2001). In the second stage, we analyzed the contribution of NDVI, texture, phenology, and fragmentation metrics in the model and recommended a minimum set of metrics to characterize agricultural systems in Mali.

Random Forest classification

All 4000 villages in South Mali constituted our application data set. This data set had the same attributes as the training set (see Table 2). The previously obtained RF model was then applied to these village attributes in order to predict the class (type of agricultural systems) to which each of the villages in the area belongs.

Results

Accuracy assessment

As mentioned above, one of the most interesting properties of the RF models is that there is no need for cross validation or for a separate test set to obtain an unbiased estimate of the test set error rate. The accuracy assessment was thus based on the 100-village data set used for training the RF model.

The Random Forest model classified the agricultural system classes with an estimated overall accuracy of 60%, calculated on out-of-bag (OOB) observations. Class errors ranged

from 30% to 50% (Table 3). The classification process was repeated 50 times, and the standard deviations were acceptable ($< \pm 0.04$ for each class error, and $< \pm 0.02$ for OOB error), showing that the model is stable.

Table 3. Confusion matrix of the Random Forest model trained on the 100-village data set.

		Reference data				User Accuracy UA (commission error)	Producer Accuracy PA (omission error)	F-score ⁽¹⁾
		A	B	C	Total			
Classified data	A	18	1	10	29	62% (38%)	70% (30%)	0.66
	B	5	22	10	37	59% (41%)	67% (33%)	0.63
	C	4	10	20	34	59% (41%)	50% (50%)	0.54
Total		27	33	40	100	Overall accuracy:		60%

$$^{(1)} \text{ F-score} = (2 * \text{UA} * \text{PA}) / (\text{UA} + \text{PA})$$

Classes A and B were classified best (F-score equal to 0.66 and 0.63, respectively). There was significant confusion between classes B and C. Both classes contain grains and cotton, in different proportions, but grown using similar practices. Globally, commission and omission errors were reasonably balanced for each class (less than 10% difference), implying that the village agricultural systems were estimated correctly.

Ranking the metrics

Figure 2 shows the results of the variable importance analysis and the distribution of the most important metrics per class.

Among the eight most important variables, we found, in decreasing order of importance, NDVI, phenology and texture metrics. As expected, the most important variable for determining the class of the agricultural system is a spectral metric (the annual mean of NDVI). The date of greenness maximum and the date of greenness decrease also played an important role and were ranked 2 and 3. Concerning texture, the maximum of variance in March and the maximum of skewness in June were also highly ranked (4 and 7, respectively). Two other spectral metrics, the seasonal mean and the annual maximum of NDVI, played a role in the characterization of the three types of agricultural systems (ranked 5 and 6, respectively). Maximum altitude also contributed noticeably and was ranked 8th out of 30. The other metrics contributed little to the reduction in the error rates.

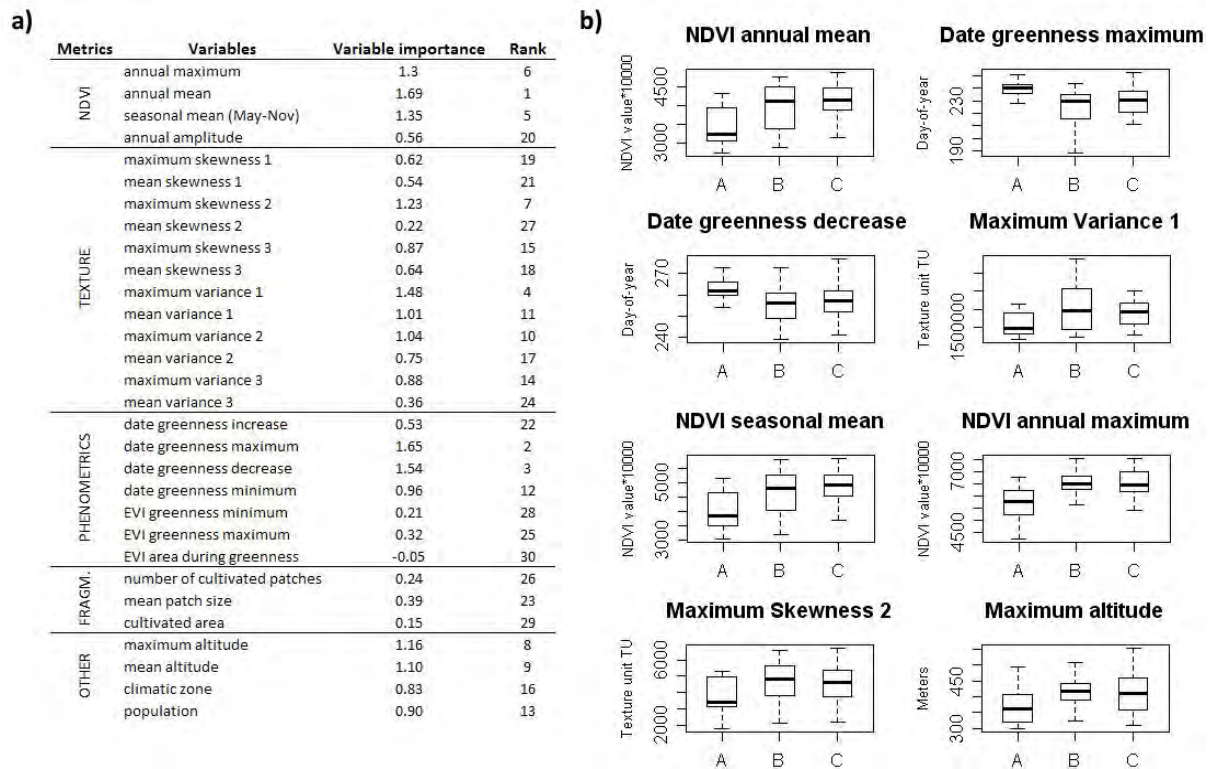


Figure 2. (a) RF-based measure of the importance of each variable, and (b) distribution of the most important metrics by class (see text for definition of class).

The results by class (Figure 2b) show that class A was accurately differentiated with phenology metrics and the NDVI annual maximum metric. Class A represents food crops, with a short monsoon season and a poor choice of species to grow (millet or sorghum), which do not usually require fertilization. Classes B and C had similar metrics, but class B generally had a wider range of spectral metrics than class C. Class B represents intensive agriculture, with a lot of different crops that can be grown.

Contribution of the different metrics

To quantify the contribution of the spectral, textural, phenological, and spatial information in distinguishing different agricultural systems, we tested 11 different sets of metrics (Figure 3) and compared them to the overall and class accuracies of the complete data set (60% overall accuracy for 30 metrics).

We first tested three sets (sets 1 to 3) -NDVI metrics, texture metrics, and phenology metrics separately- and showed that the best overall accuracy results (53%) were obtained with the texture metrics alone (set 2) with balanced class accuracies (between 51% and 55%). When NDVI metrics were used alone (set 1), the global accuracy was 45%, with a wider range of class

accuracy (from 33% for class C to 56% for class B). Using phenology metrics alone (set 3) particularly favored class A (57%), but was not suitable for class C (32%), and had a low overall rate of accuracy (46%).

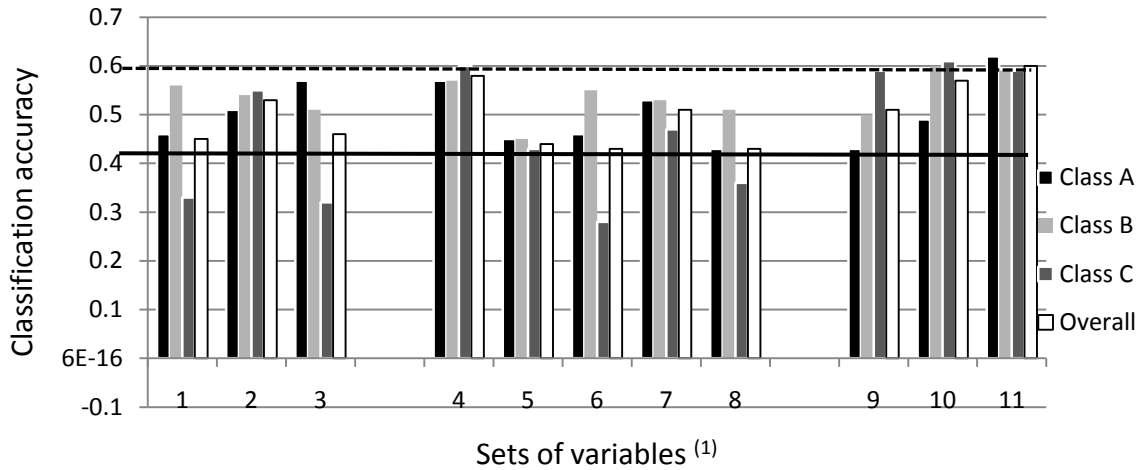
We then tested a combination of two sets of metrics, based on NDVI metrics (set 4 to 8). Using NDVI and texture metrics (set 4) increased the accuracy of class A and C by about 10% and 27% respectively compared to the use of NDVI alone, whereas the class B error remained the same. Adding other group of information to NDVI metrics (sets 5 to 8) did not increase accuracy, except when fragmentation variables (set 7) were used (+7% and +14% for class A and C respectively).

When larger combinations of metrics were compared (set 9 to 11), based on NDVI + socio-environmental data, we showed that texture (set 10 compared to set 9) still had a positive effect on class accuracy (+6% for class A and + 10% for class B). Using all metrics except phenology (set 10) led to overall accuracy comparable to that obtained with the complete data set for class B and class C (set 11), but the phenology metrics significantly improved the accuracy of class A (+13%).

Resulting map

When we applied the complete RF model (30 metrics) to the satellite and socio-economic metrics calculated for the geolocalized 4000-village data set in South Mali, we obtained a map like the one presented in Figure 4.

In the South of the Sudano-Sahelian zone (zone 4, Figure 5), class A was dominant. Sorghum and millet are well adapted to this zone because they are resistant, and have a short growth cycle of about 90 days (Kouressy *et al.*, 2008). In the traditional cotton basin (around Koutiala, zone 3, Figure 5), the dominant system is agro-forestry/pastoral agriculture mainly with rainfed crops. Agriculture is focused on cotton, the main cash crop, and corresponds to class B. The Sudanian zone (zones 1 and 2, Figure 5) is also a cotton-based system zone, but is more diversified, especially around Sikasso (Soumare, 2008), with the simultaneous presence of class B and C agricultural systems. The length of the rainy season in this region makes it possible to grow a wide range of species. Farmers usually cultivate different species and varieties to ensure a certain degree of production stability.



Set 1=NDVI; Set 2=Texture; Set 3=Pheno -

Set 4=NDVI+Texture; Set 5=NDVI+Other (pop only); Set 6=NDVI+Other (climat only); Set 7=NDVI+Frag;

Set 8=NDVI+Other (elevation only) -

Set 9=NDVI+Frag+Other (climat, elevation, pop); Set 10=NDVI+ Frag+Texture+other (climat, elevation, pop);

Set 11=all metrics.

Figure 3. Accuracy of class and overall classification of Random Forest run with different sets of metrics. The solid line corresponds to the overall accuracy obtained using spectral metrics alone (set 1), the dotted line corresponds to the overall accuracy obtained with all the metrics (set 11).

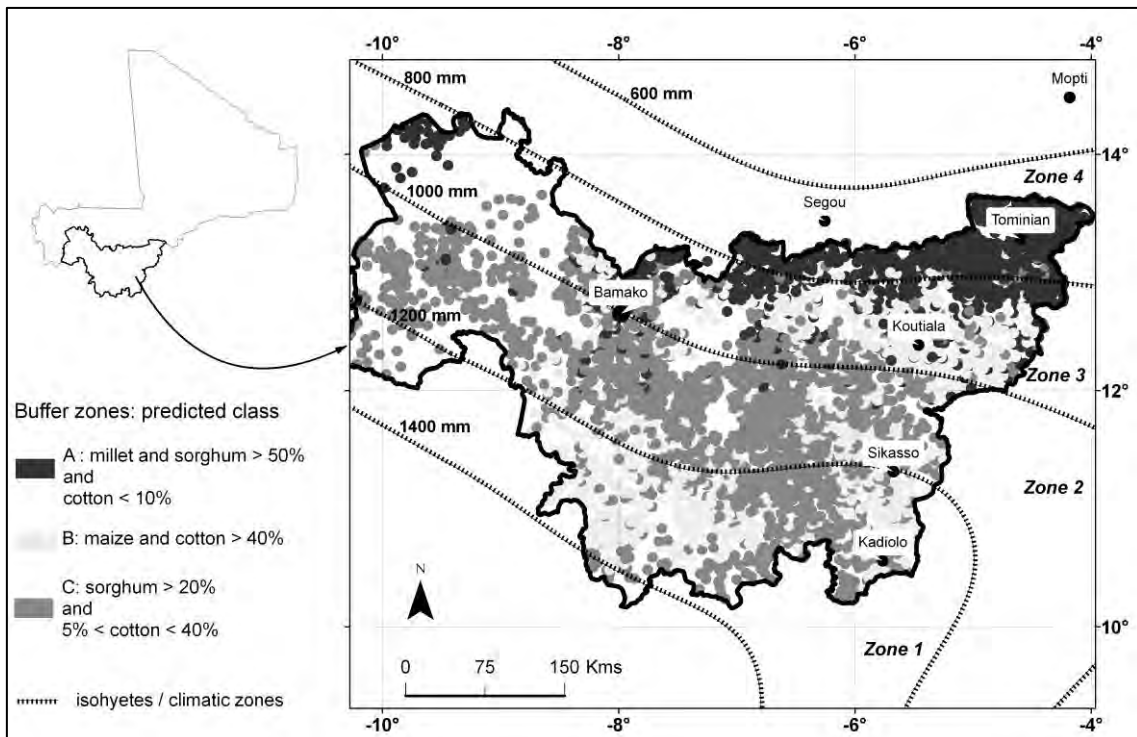


Figure 4. Village based agricultural systems in South Mali predicted by the Random Forest model.

Discussion

Relevance of the approach

We based this study on the assumption that human practices (species cropped, intensification, etc.) shape the cropped portion of the landscape, and that this landscape could be described by a set of coarse resolution satellite-derived metrics (spectral, spatial and phenological metrics). Our results show that the Random Forest (RF) model used to link these metrics to the agricultural systems typology had an overall accuracy of 60%. This accuracy, estimated internally during the run, actually underestimates the real potential of the RF model, since its calculation is based on only two thirds of its trees.

This reasonably good accuracy validates our original assumption, and shows that the RF method is suited to our objectives. We can consider that part of the misclassification was due to the quality of the ground data set given the use we made of it: i.e., a limited number of points (100 villages), old survey data (5 years old), and data originally collected for a study designed for different goals than this one. The spatial resolution of MODIS (250 m) and the fact that we considered the main type of agricultural system (mixing two or three species in different proportions) in each village is also a source of inaccuracy, but this typology was considered consistent with the coarse resolution of the images. The use of the RF classifier in this analysis was found to be valid, because, in addition to its ease of use, it accepts small training samples (100 points in this study), and authorizes correlated variables (most of the metrics used in this study were correlated).

Importance of remote sensing metrics as descriptors of the agricultural systems

The analysis of the importance of the respective variables using Random Forest revealed that the annual mean of NDVI was the most important. Cropping systems were first characterized by their spectral signature, as expected. Phenological and texture metrics also played a role in distinguishing classes A, B and C, each of which had two out of the eight most important variables. When groups of metrics (NDVI, texture, phenology and fragmentation), and not individual ones, were analyzed separately, the results showed that the texture metrics group was the most discriminating in classifying the agricultural systems (overall accuracy of 53% when only texture metrics were used, compared to 45% when only NDVI metrics were used). This result is new and unusual, as very few studies have made use of this type of metrics with coarse resolution imagery, which opens new research pathways. Mixing NDVI metrics and texture metrics also gave good results, while adding the phenology metrics was not very helpful, except for detecting class A villages.

The classification results can be analyzed through the environmental and socio-economic constraints that determine the possible diversity of farmers' practices, and hence the agricultural systems. For example, the power of discrimination of the phenology metrics for class A can be explained as this class represents the food crop class, with a short monsoon season and a poor choice of species to grow (millet or sorghum). Second, the fragmentation metrics changed the classification results for classes A and C. This variable plays an important role in the cartography of the cropped domain. Vintrou *et al.* (2011) showed that in Mali, 200 ha was a reasonable size for a patch to have a chance of being detected using MODIS imagery. The mean patch size of the 100 villages used in this study was 188 ha. The mean patch sizes for class A, B and C were 179 ha, 151 ha and 236 ha respectively. We can hypothesize that, at this scale, the crop patch was detected only for class A and class C and could be used as a discrimination metrics between the two classes. Finally, Class B is the intensive class. The presence of cash crops enables intensification (especially linked to cotton production) and probably results in a relatively homogeneous crop cover; whereas, in class A and class C, the crop cover is heterogeneous, with bare areas in class A (in the dry zone of North Mali) and with different types of species in class C, leading to abrupt breaks in the landscape and marked variance in the 7 x 7 pixel moving window. This could be the reason why the addition of texture metrics did not significantly improve the accuracy of class B, and distinguished class B from class C, which had a less homogeneous crop cover. However, Malian farmers are faced with the current cotton crisis which is resulting in practices becoming less intensive and villages in class B may be evolving towards class C (Belieres *et al.*, 2009), which would explain the confusion between the two classes.

Link with food security systems

In Mali, like in many West African countries, information is lacking on the cropped domain. The area under different crops is generally defined by the climate, elevation and soil theoretical constraints, but very few authors have attempted to stratify land according to what is actually cultivated. Recently, De Raïssac *et al.*, 2011 produced regional maps of production basins of the main rainfed crops using national statistics on yield and cropped area. However these maps are broad with basins drawn at the trans-national scale. The Mali likelihood zone map from FEWS-NET (FEWSNET, 2009) is closer to what we tried to do in this study, because it splits Mali into 13 homogeneous zones based on both production and trade/exchange. However, the zones are large and the methodology does not include any remote sensing observations.

In conclusion, the map of the main rainfed agricultural systems (food-producing, intensive and mixed agricultures classes) produced in this study is valuable because it is an original and

independent source of information for South Mali. Because local factors such as climate, soil, water availability, access to markets and fertilizers, influence the agricultural systems, mapping these systems can help to determine which region and which population may be vulnerable to different hazards. It is critical to localize different agricultural systems in order to conduct meaningful food security analyses (USAID, 2009). Furthermore, in terms of yield forecast, agricultural system maps can be helpful in agro-meteorological modelling, by reducing the number of initial conditions (cropped species and intensification practices). The identification of the major agricultural systems ensures accurate retrieval of crop specific parameters (Doraiswamy *et al.*, 2005).

Conclusion

Because the highly fragmented landscapes of West Africa are not well retrieved using existing Earth Observing Systems, and given the importance of such information for food security systems, there is an urgent need to develop new approaches to detect and map agricultural systems there using coarse resolution imagery. We worked on the assumption that, agricultural systems shaped the landscape through human cropping practices, and that the resulting landscape can be described with a set of coarse resolution satellite-derived metrics (spectral, textural, temporal, and spatial metrics). The method relied on a Random Forest statistical model that was built upon 32 such indicators and trained on a set of 100 villages, each assigned to one of three main crop production system classes, with an overall accuracy of 60%.

Three main achievements of this study were : i) information concerning agricultural systems can be retrieved at village level using coarse resolution imagery, ii) the use of the Random Forest classifier was found to be relevant for characterizing agricultural systems in West Africa because, in addition to its ease of use, it accepts small training samples, and authorizes correlated variables and iii) even at coarse resolution, the images contain textural information that can be exploited to inform on broad agricultural systems types of fragmented and heterogeneous landscapes in West Africa.

These results now need to be confirmed by extending the method to neighboring countries with similar agricultural landscapes such as Senegal, Niger or Burkina Faso, in which case, new developments in applied research on extracting textural information from coarse satellite imagery can be foreseen.

Acknowledgments

This work was supported by CIRAD funds (E. Vintrou's fellowship) and by the AMMA project (An European Community's Sixth Framework Research Programm, <http://www.amma-international.org>). The authors would like to thank Michel Passouant (CIRAD-ES) for his help with statistics, and Jean-François Belières (CIRAD-ES) for sharing his knowledge on Malian landscapes. Finally, we wish to acknowledge the anonymous reviewers and the Editor of this special issue, Dr Thenkabail, for helping us improve a previous version of the manuscript.

References

- Arvor, D., M. Margareth, V. Dubreuil, A. Bégué, and Y.E. Shimabukuro, 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, 32(2):702-713.
- Asner, G.P., M. Keller, and R. Pereira, 2002. Remote sensing of selective logging in Amazonia: Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. *Remote Sensing of Environment*, 80(3):483-496.
- Belieres, J.F., J.É. Bidou, and I. Droy, 2009. Le cas des systèmes de production cotonniers au Mali. In: B. Thibaud and A. François (Editors),. *Systèmes de production et durabilité dans les pays du Sud.*, KARTHALA Editions, pp.247-274.
- Bernard, S., 2009. *Forêts Aléatoires: De l'Analyse des Mécanismes de Fonctionnement à la Construction Dynamique*. Ph.D. dissertation, University of Rouen, Rouen (FR), 192 p.
- Bernard, S., L. Heutte, and S. Adam, 2009. On the selection of decision trees in Random Forests. *Proceedings of the International Conference on Intelligence Computing (ICIC)*, IEEE, Atlanta (Georgia, USA), pp. 302-307.
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1):5-32.
- Brown, M.E. (Editor), 2008. *Famine early warning systems and remote sensing data*. Springer-Verlag Berlin Heidelberg., 313 p.
- Brown, M.E., C.C. Funk, G. Galu, and R. Choularton, 2007. Earlier famine warning possible using remote sensing and models. *Eos*, 88:39.
- Bruzzone, L., C. Conese, F. Maselli, and F. Roli, 1997. Multisource classification of complex rural areas by statistical and neural-network approaches. *Photogrammetric Engineering & Remote Sensing*, 63(5):523-533.
- Burnett, C., and T. Blaschke, 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168(3):233-249.
- Chan, J.C.W., and D. Paelinckx, 2008. Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112(6):2999-3011.
- Chinchor, N., 1992. MUC-4 Evaluation Metrics. *Proceedings of the Fourth Message Understanding Conference*, Association for Computational Linguistics Stroudsburg, PA, USA, pp. 22-29.
- Congalton, R.G., and R.A. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, 49:69-74.
- Coulibaly, O.N., 1995. *Devaluation, new technologies, and agricultural policies in the Sudanian and Sudano-Guinean zones of Mali*. Ph.D. dissertation, Purdue University, West Lafayette (Indiana, USA), 188 p.

- Cutler, D.R., T.C. Edwards, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler, 2007. Random forests for classification in ecology. *Ecology*, 88(11):2783–2792.
- De Fries, R., M. Hansen, M. Steiner, R. Dubayah, R. Sohlberg, and J. Townshend, 1997. Subpixel forest cover in central Africa from multisensor, multitemporal data. *Remote Sensing of Environment*, 60(3):228-246.
- De Raïssac, M., N. Bricas, F. Maraux, H. Boirard, P. Remy, D. Simon, M.C. Thirion, and C. Uhder (Editors), 2011. *Les cultures vivrières pluviales en Afrique de l'Ouest et du centre : Eléments d'analyse et propositions pour l'action*. Collection "A savoir", 6. AFD, Paris (FR), 192 p.
- Doraiswamy, P.C., T.R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, and J. Prueger, 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, 97(2):192-202.
- FEWSNET, 2009. Livelihood zone map and profiles. URL: <http://www.fews.net/>, (last date accessed: December 2011).
- Fritz, S., and L. See, 2008. Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology*, 14(5):1057-1075.
- Hountondji, Y.C., N. Sokpon, and P. Ozer, 2006. Analysis of the vegetation trends using low resolution remote sensing data in Burkina Faso (1982-1999) for the monitoring of desertification. *International Journal of Remote Sensing*, 27(5):871-884.
- Hutchinson, C.F., 1991. Uses of satellite data for famine early warning in sub-Saharan Africa. *International Journal of Remote Sensing*, 12(6):1405-1421
- Ichokua, C., and A. Karnieli, 1996. A review of mixture modeling techniques for sub-pixel land cover estimation. *Remote Sensing Reviews*, 13(3-4):161-186.
- Justice, C.O., J.R.G. Townshend, B.N. Holben, and C.J. Tucker, 1985. Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, 6(8):1271-1318.
- Kayitakire, F., C. Hamel, and P. Defourny, 2006. Retrieving forest structure variables based on image texture analysis and IKONOS-2 imagery. *Remote Sensing of Environment*, 102(3-4):390-401.
- Kouressy, M. et al., 2008. Adaptation of Malian sorghums to climate variability. *Adaptation des sorghos du Mali a la variabilite climatique*, 17(2):95-100.
- Latinne, P., O. Debeir, and C. Decaestecker, 2001. Limiting the number of trees in random forests. *Multiple Classifier Systems*:178-187.
- Lawrence, R.L., S.D. Wood, and R.L. Sheley, 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). *Remote Sensing of Environment*, 100(3):356-362.
- Liaw, A., and M. Wiener, 2002. Classification and Regression by randomForest. *R news*, 2(3):18–22.
- Liaw, A., and M. Wiener, 2008. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. URL <http://CRAN.R-project.org/package=randomForest>. *R package version 4.5-28*.
- Na, X., S. Zhang, X. Li, H. Yu, and C. Liu, 2010. Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic. *Photogrammetric Engineering and Remote Sensing*, 76(7): 833-840.
- Nicodemus, K., J. Malley, C. Strobl, and A. Ziegler, 2010. The behaviour of random forest permutation-based variable importance measures under predictor correlation. *BMC bioinformatics*, 11(1):110.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217-222.
- Peddle, D.R., and S.E. Franklin, 1991. Image texture processing and data integration for surface pattern discrimination. *Photogrammetric Engineering and Remote Sensing*, 57(4):413-420.

- Prince, S.D., and S.N. Goward, 1995. Global primary production: a remote sensing approach. *Journal of Biogeography*, 22:815-835.
- Soumare, M., 2008. *Dynamique et durabilité des systèmes agraires à base de coton au Mali*. . Ph.D. dissertation, Université de Paris X Nanterre, Paris (FR), 373 p.
- Story, M., and R.G. Congalton, 1986. Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing*, 52:397–399.
- Stumpf, A., and N. Kerle, 2011. Object-oriented mapping of landslides using Random Forests. *Remote Sensing of Environment*, 115(10):2564-2577.
- Thenkabail, P., G.J. Lyon, H. Turrall, and C.M. Biradar, 2009a. *Remote Sensing of Global Croplands for Food Security*. CRC Press - Taylor and Francis group, Boca Raton, London, New York, 556 p.
- Thenkabail, P.S., 2010. Guest Editor: Special issue on “Global Croplands”. *Remote Sensing Journal*, URL: http://www.mdpi.com/journal/remotesensing/special_issues/croplands (last date accessed: 24 November 2011).
- Thenkabail, P.S. et al., 2009b. Global irrigated area map (GIAM), derived from remote sensing, for the end of the last millennium. *International Journal of Remote Sensing*, 30(14):3679-3733.
- Thenkabail, P.S., M.A. Hanjra, V. Dheeravath, and M. Gumma, 2011. Global croplands and their water use from remote sensing and non-remote sensing perspectives. In: Q. Weng (Editor), *Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications*. CRC Press, Taylor and Francis Group, pp. 383-419.
- Thenkabail, P.S., J.G. Lyon, H. Turrall, and C.M. Biradar, 2009c. *Remote sensing of global croplands for food security*. CRC Press, Taylor & Francis Group, Boca Raton (FL), London, New York, 556 p.
- Tsaneva, M.G., D.D. Krezhova, and T.K. Yanev, 2010. Development and testing of a statistical texture model for land cover classification of the Black Sea region with MODIS imagery. *Advances in Space Research*, 46(7):872-878.
- USAID, 2009. Application of the livelihood zone maps and profiles for food security analysis and early warning - Guidance for Famine Early Warning Systems Network (FEWS NET) representatives and partners. USAID, pp. 23.
- Vintrou, E., A. Desbrosse, A. Bégué, P.C.S. Traoré, C. Baron, and D. Lo Seen, 2011. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):83-93.
- Wardlow, B.D., S.L. Egbert, and J.H. Kastens, 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sensing of Environment*, 108(3):290-310.

ARTICLE 3 :

**Analysing MODIS phenometrics information content
in heterogeneous agricultural landscapes of West Africa**

Soumission prévue en Mai 2012

Analysing MODIS phenometrics information content in heterogeneous agricultural landscapes of West Africa

Elodie Vintrou*⁽¹⁾, Agnès Bégué⁽¹⁾, Christian Baron⁽¹⁾, Danny Lo Seen⁽¹⁾, Saad Alexandre⁽¹⁾, Seydou Traoré⁽²⁾

⁽¹⁾ CIRAD UMR TETIS, Maison de la Télédétection, 500 rue J.F. Breton, Montpellier, F-34093 France

⁽²⁾ Centre Régional AGRHYMET, 0425 Bld de l'Université, BP 11 011 Niamey, Niger

* Corresponding author. Tel.: +33 4 67 54 87 54; fax: +33 4 67 54 87 00

E-mail address: elodie.vintrou@cirad.fr

Abstract

In order to verify whether the phenometrics from MODIS MCD12Q2 reproduce well the phenological stages of different crops in Mali, these metrics were extracted for areas previously identified as crops and compared to simulated LAI temporal profiles generated by SARRA-H crop model for the same areas and seasons, taking into account the different types of crops grown in a given area. We both studied spatial (north-south gradient, agricultural areas) and temporal (from one year to another) differences between crop cycles and how they are returned in MODIS phenometrics. Comparing observations and simulations of phenological metrics in Mali spatially and temporally, we showed a difference of 20 to 40 days between transition dates from MODIS and from LAI, explainable by the mixture of land cover in a MODIS pixel. We conclude preconizing the use of these remotely-sensed phenometrics in crop model for yield forecasting in complement to climatic data commonly used, so as to provide a closer estimation of vegetation phenological changes that integrate both rainfall variability and land cover status.

Introduction

Phenological dynamics of crops are expected to be essential information when evaluating crop production in the food insecure regions of West Africa. As most cereals there are cultivated under rainfed conditions, the timing of a specific phenological stage of millet or sorghum can vary from year to year because of variations in the start of the planting season. The latter highly depends on soil moisture and temperature conditions, as well as on farm-level management decisions (Sakamoto *et al.*, 2005). For these reasons, vegetation conditions need to be carefully monitored by early warning systems during the critical growth stages when estimating year-end crop yields in these regions.

In the last two decades, the availability of global remote sensing data sets has provided new means of studying global vegetation patterns and dynamics (Zhang *et al.*, 2005). With the ability to detect surface phenology objectively, uniformly in time, and on a global scale, time series of low- and medium-resolution satellite images have been used to study the phenological patterns in relation to climate variability and human actions (Kathuroju *et al.*, 2007; Reed *et al.*, 1994; Zhang *et al.*, 2009). Because the observation of vegetation phenology through satellite sensors integrates signals from land surface processes in addition to reflectance from plants, satellite-based phenology is often called land surface phenology (De Beurs and Henebry, 2005). They provide information on the main transition dates of vegetation phenology and can also give indication on biomass production (Wessels *et al.*, 2010).

Thereby, a variety of methods have been developed over the past decade to detect the timing of vegetation phenology from satellite data. A review of the different methods can be found in White *et al.* (2009) and Schwartz and Hanes (2010). For example, vegetation start-of-season may be derived from characteristics of the seasonal NDVI curve (Funk and Budde, 2009; Rojas *et al.*, 2011) or precipitation data (Balme *et al.*, 2005; Heumann *et al.*, 2007; Wessels *et al.*, 2010). As vegetation phenology in arid and semiarid ecosystems is primarily controlled by water availability, a number of field studies have attempted to quantitatively link phenology to precipitation forcing. Zhang *et al.* (2005) examined how phenology changed with latitude and how it was related to the timing of seasonal rainfall in the Sahel and Sudanian regions. His conclusion was that well-defined thresholds

exist in cumulative rainfall that stimulate vegetation green-up in arid and semiarid regions of Africa. Therefore, vegetation indices are often coupled with precipitation data (e.g. Balme *et al.*, 2005; Brown and de Beurs, 2008; Heumann *et al.*, 2007; Wessels *et al.*, 2010) to study vegetation phenology. However, these climatic data are either aggregated or extrapolated from weather stations (which are few in most developing countries, and whose data are difficult to access) or estimated from low spatial resolution satellite data (whose precision is not compatible with the requirements of crop models) (Brown *et al.*, 2008). Thus, directly estimating phenological indices from time series of vegetation indices obtained at medium spatial resolution would not only be an improvement on spatial resolution, but would also provide a closer estimation of vegetation phenological changes that integrate both rainfall variability and land cover status.

The only currently available global product that includes phenological variables is the MODIS Land Cover Dynamics Yearly (MCD12Q2) product. This product provides information on the start-of-season (SOS or greenup onset), the maximum of the growing season (or maturity onset), the senescence onset (or end of the maximum) and the end of the growing season (or dormancy onset) calculated from vegetation indices time series at a spatial resolution of 500 m (Ganguly *et al.*, 2010; Zhang *et al.*, 2003). In a previous study (Vintrou *et al.*, 2012) phenological metrics from MCD12Q2 product were shown to be important variables for the characterization of agricultural systems in West Africa. In the present study, we go one step further in assessing whether the information content of these phenological metrics may be used to monitor vegetation in heterogeneous landscapes of West Africa.

In such data-scarce areas, surveys on agricultural practices (sowing and harvest dates) exist only as district-level aggregates and cannot be used directly in this assessment. Instead, we have used a crop model that is capable of reproducing the evolution of phenological stages of different crop species, while needing mainly rainfall as input data. The SARRA-H crop model (Dingkuhn *et al.*, 2003) used has already been applied to all the main crop species and varieties in West Africa, including the different sorghum, millet and maize varieties that are found in Mali, and is used by Agrhymet in its crop yield forecast system, where it can provide information on the productivity of different crops and crop varieties or level of intensification.

In order to verify whether the phenometrics from MODIS MCD12Q2 reproduce well the phenological stages of different crops that we know in Mali, these metrics were extracted for areas previously identified as crops (and where we had measured meteorological data to run the crop model), and compared to simulated LAI temporal profiles generated by the crop model for the same areas and seasons, taking into account the different types of crops grown in a given area. We both studied spatial (north-south gradient, agricultural areas) and temporal (from one year to another) differences between crop cycles and how they are returned in MODIS phenometrics. After comparing observations and simulations of phenological metrics in Mali, we conclude on the possibility of using remotely-sensed phenometrics to improve agricultural production estimation for a country as large as Mali.

1. Material

1.1. Study Area

With an area of 1,240,192 km², Mali is a land-locked country of West Africa situated between latitudes 10°N and 24°N (Figure 1). Mali exhibits a latitudinal climatic gradient that ranges from sub-humid to semi-arid and that extends further north to arid and desertic. As with other West African countries along the same latitudinal belt, food security relies on an adequate supply of rainfall during the monsoon season. Farming is the main source of income for many people of this region, with rainfed millet and sorghum as the major food crops. The vast majority of the population (80%) consists of subsistence farmers. A few larger farms produce crops for sale (cash crops), mainly cotton and peanuts. In this study, we will not consider the Saharan zone in the north of the country where the sparse rainfall amounts to less than 200 mm per year.

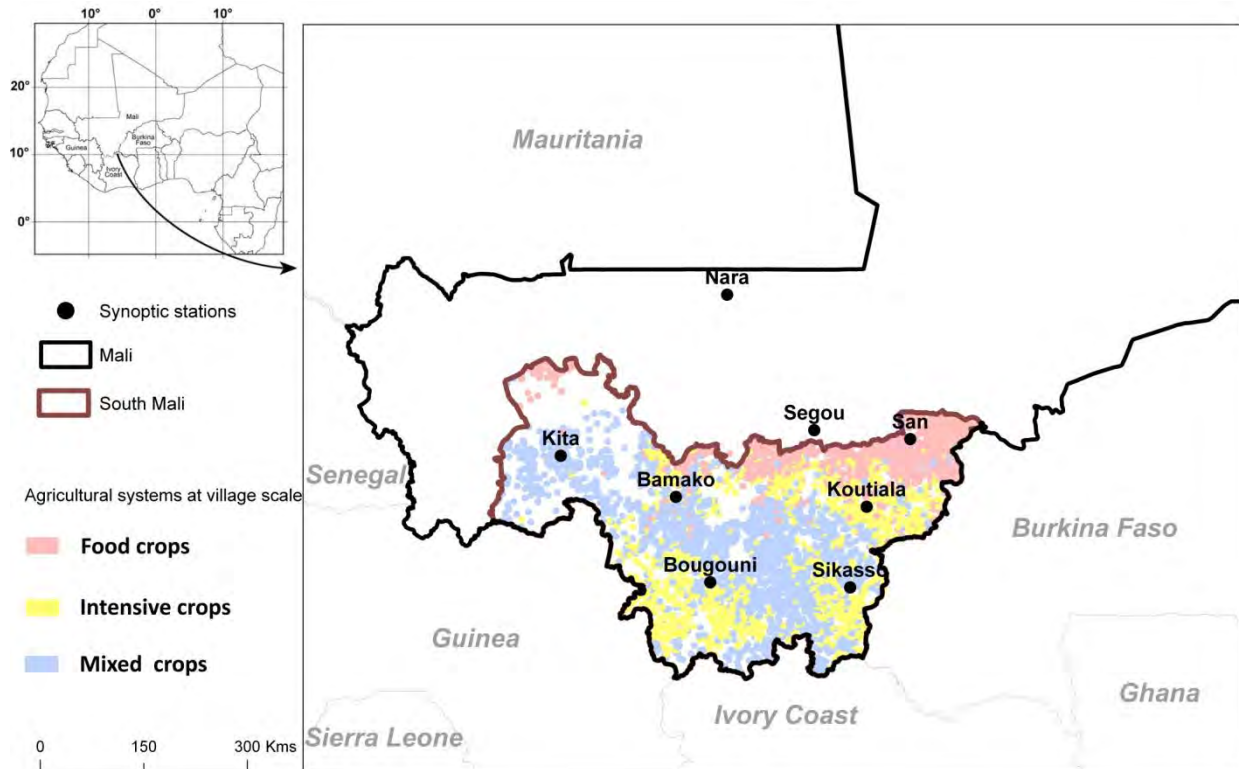


Figure 1 : Localization of Mali and the synoptic stations used in this study, and map of the crop production systems of South Mali (Vintrou *et al.*, 2011):

- **Food-producing agriculture** = area dedicated to millet and sorghum > 50% and to cotton < 10%;
- **Intensive agriculture** = area dedicated to maize and cotton > 40% ;
- **Mixed agriculture** = area dedicated to sorghum > 20% and to cotton between 5% and 40%.

1.2. Data

MODIS Land Dynamics product (MCD12Q2)

We download MCD12Q2 on Mali (two tiles) for years 2001 to 2008.

The yearly MODIS Land Cover Dynamics product (MCD12Q2; Zhang *et al.*, 2006) includes 15 layers (Table 1) and has been developed to support studies of seasonal phenology and interannual variation in land surface and ecosystem properties. The Collection 5 land cover dynamics product is described in Ganguly *et al.* (2010) and available online for the years 2000 to 2009 (accessible from <https://wist.echo.nasa.gov/>) at 500 m spatial resolution. This product is produced each year from the 8-day vegetation index EVI (Enhanced Vegetation Index) calculated from the NBAR reflectance (Nadir

Bidirectional Reflectance Distribution Function - Adjusted Reflectance). The EVI is used because it provides greater dynamic range than the normalized difference vegetation index (Huete *et al.*,2002).

For the global phenology MODIS product, Zhang *et al.* (2003) developed a method to estimate phenological events of the vegetation within each year. They developed an algorithm that models the growth and degrowth of the vegetation index with a series of logistic functions developed on 24 months of input data (i.e., the 12 months of interest data bracketed by six month's data on either side). This algorithm characterizes vegetation growth cycles using four transition dates based on the curvature-change rate of EVI from time series of MODIS data: (1) greenup: the date of onset of EVI increase or usually called Start-Of-Season (SOS); (2) maturity: the date of onset of EVI maximum, usually called Start-of-MAXimum (SMAX); (3) senescence: the date of onset of EVI decrease, usually called End-of-MAXimum (EMAX); and (4) dormancy: the date of onset of EVI minimum, usually called End-Of-Season (EOS) (Ganguly *et al.*, 2010). In addition to the transition dates, the minimum, maximum and growing season-integrated values of EVI are given for each pixel. Each of these variables are encoded on two distinct layers (n and $n + I$) in order to catch two growing seasons a year. Indeed, the band $n + I$ of each phenometric is useful if there are two rainfall seasons or two crop cycles.

Band	Phenological variable name	Units	Valid range	Scale factor
1	Onset_Greenness_Increase cycle 1	Days since January 1, 2000	0–32766	1.0
2	Onset_Greenness_Increase cycle 2			
3	Onset_Greenness_Maximum cycle 1	Days since January 1, 2000	0–32766	1.0
4	Onset_Greenness_Maximum cycle 2			
5	Onset_Greenness_Decrease cycle 1	Days since January 1, 2000	0–32766	1.0
6	Onset_Greenness_Decrease cycle 2			
7	Onset_Greenness_Minimum cycle 1	Days since January 1, 2000	0–32766	1.0
8	Onset_Greenness_Minimum cycle 2			
9	NBAR_EVI_Onset_Greenness_Min cycle 1	NBAR EVI units	[-10 000 , +10 000]	0.0001
10	NBAR_EVI_Onset_Greenness_Min cycle 2			
11	NBAR_EVI_Onset_Greenness_Max cycle 1	NBAR EVI units	[-10 000 , +10 000]	0.0001
12	NBAR_EVI_Onset_Greenness_Max cycle 2			
13	NBAR_EVI_Area cycle 1	NBAR EVI units	[-3 660 , +3 660]	0.01
14	NBAR_EVI_Area cycle 2			
15	Phenology_Quality	Concatenated flags	N/A	N/A

Table 1 : File layers of the MODIS Land Cover Dynamics product (MCD12Q2).

EVI data

We calculated EVI data set from the Nadir BRDF-Adjusted Reflectance 8-Day L3 Global 500m SIN Grid V005 (MCD43A4) for Mali, from 2001 to 2009, in order to be able to position phenometrics values on annual EVI profiles.

Crop mask and agricultural systems mapping

The map of the cultivated domain in Mali (2-classes: “crop” and “non crop”) was produced at 250 m spatial resolution by Vintrou *et al.* (2012) using MODIS time series. The map of agricultural systems was also produced for South Mali using spectral, spatial, temporal and textural indicators extracted from MODIS images, in combination with ground data (Vintrou *et al.*, 2011). On this map, each 4000 villages of the study area is given one of the 3-class agricultural system (“food-producing”, “intensive” and “mixed” agriculture), and is spatially represented on the map by a 5-km buffer region around the center of the villages (Figure 1).

Climatic data

Provided by the AGRHYMET, climatic data (rainfall, temperature, insolation) of eight synoptic stations have been collected on Mali for the year 2007. Two stations, Segou and Sikasso, covered 8 years, from 2001 to 2008 (Figure 1).

1.3. SARRA-H crop model

SARRA-H is a simple, deterministic crop model for cereals operating at daily time steps and implemented on the Ecotrop platform (<http://ecotrop.cirad.fr/>), of the Centre International de Recherche Agronomique pour le Développement (CIRAD) (Baron *et al.*, 1996; Dingkuhn *et al.*, 2003; Sultan *et al.*, 2005). This platform allows the management of different models, data management and simulation scenarios. The model chosen for this study is SARRAH version 3.2, simulating the dynamics of biomass (root, stem, leaves and grains) of grain, especially in several selected varieties of millet, maize and sorghum. The model reproduces three major processes: evolution of phenological

stages of the varieties (cycle length and photoperiodism), carbon balance (changes in biomass and distribution) and water balance (Marteau *et al.*, 2011). The simulation of the production of biomass is constrained by the availability of two main resources: the light energy and soil water resources.

Model parameters

The model uses daily climatic data (rainfall, global radiation or insolation, temperature and evapotranspiration), soil type, agricultural practices information, and crop variety. Furthermore, a number of empirical constants set the rainfall criteria that would initiate sowing, the modalities for an automatic test during seedling stage for stress-induced crop failure, and the option for automatic replanting in the case of failure (Dingkuhn *et al.*, 2003). Depending on whether the crop is traditional (photo-period sensitive) or improved (insensitive), it will reach maturity either at a relatively stable calendar date or after genotype specific growth duration. During that period, the crop will undergo variable degrees of drought, with variable effects on yield, because the phenological phases of the crop differ in stress sensitivity (Bacci and Reyniers, 1998).

The different scenarios used for the model parameterization defined a series of a Leaf Area Index (LAI) profiles. LAI represents the ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows. LAI is used to predict photosynthetic primary production and as a reference tool for crop growth.

2. Data processing

2.1. Satellite-derived phenometrics

As the MODIS Phenology product has a coarser spatial resolution than the crop mask (500 m vs 250 m), we proceeded to mask pre-processing. To identify “crop” pure pixels on MODIS Phenology product, we used a 3x3 sum filter on the crop mask (crop is coded 1, and non crop is coded 0), and keep only the pixels with a score higher or equal to 7 (out of 9). This filtered crop mask guarantees a high proportion of crop at 500 m resolution. The filtered mask was applied on the MODIS phenology product, and the resulting product is referred hereafter as the Crop Phenology product.

Finally, we calculated summary statistics (histogram, mean, standard deviation) for the four phenometrics on 10x10 km windows centred on the synoptic stations.

2.2. Modeled-derived phenometrics

SARRA-H model was used to predict crops behaviour in their original environment as a function of rainfall regimes and agricultural practices (crop specie and variety, fertilization). For each synoptic station, we fixed a set of parameters using the agricultural systems map (Figure 1) and expert knowledge.

- The choice of species and varieties (Table 2) was made according to the litterature review. The phenology of Malian sorghum varieties was previously studied to understand the impact of climate constraints on their cycle duration (Kouressy *et al.*, 2008) while the farmers sowing strategy for millet in Niger was analyzed to investigate the relationships between rainfall, the sowing date and meteorological onset dates (Marteau *et al.*, 2011; Traoré *et al.*, 2011).
- For the fertilization parameter, maize is known to be an intensified production, concentrated in the cotton zone (South-East Mali), promoted by the CMDT, whereas millet is usually not fertilized and localized in the Sudano-Sahelian zone.
- Soil types (ranging from sandy to clayey soil in Mali) did not impact the LAI dynamics in Mali. We chose a single soil type (sandy clay) without limiting the seeding depth (180mm).

The model was run and a LAI profile extracted as a yearly-output by synoptic station. The corresponding model-derived phenometrics were calculated using Zhang algorithm (Zhang *et al.*, 2003), as for MODIS MCD12Q2 product.

2.3. Comparison between satellite and model-derived phenometrics

We compare both types of metrics one by one, in each synoptic station, calculating the differences in day between the two. A linear regression analysis was then used to: i) compare both types of

metrics spatially, i.e. in each synoptic station for the year 2007, and ii) compare both types of metrics temporally, i.e in Segou and Sikasso on 8 years, from 2001 to 2008, corresponding to the availability of climatic data on these stations.

STATIONS	CROPS
Bamako	Sorghum guinea, Millet kenikeba, Millet choho, Maize
Bougouni	Sorghum guinea, Maize
Kayes	Sorghum caudatum, Millet kenikeba, Maize
Kita	Sorghum guinea, Millet kenikeba, Millet choho, Maize
Koutiala	Sorghum guinea, Millet kenikeba, Maize
Nara	Sorghum caudatum, Millet kenikeba, Millet souna
San	Sorghum caudatum, Millet souna
Segou	Sorghum guinea, Millet kenikeba, Millet choho
Sikasso	Sorghum guinea, Millet kenikeba, Millet choho, Maize

Table 2. Species and varieties inputs for the synoptic station simulations with SARRA-H.

3. Results

3.1. Satellite-derived phenometrics

Exploring the MCI2Q2 product

Here we consider data from MODIS Global Land Cover Dynamics product. Each four phenometric (Start-Of-Season = SOS, Start-of-Maximum of season = SMAX, End-Of-Maximum of season = EMAX, End-Of-Season = EOS) were encoded on two bands to ensure the possibility to catch two growing seasons. At the national scale, we found that only 5% of the cultivated pixels had two complete cycles (which is in accordance with the rainfall systems and the scarcity of irrigated crops) and that 69% of the cultivated pixels had a complete cycle on the band 1 or 2.

At the station level, Table 3 shows proportions of “pure” cultivated pixels and percentages of remaining pixels post-processing of the Crop Phenology product (complete cycle and filtered crop mask). The stations that have lost much information during pre-processing are Segou and Koutiala.

Stations (polygons of 10x10km)	Nb pure crop pixels (% area)	Nb pure crop pixels with complete cycle (% area)
Bamako	27 (5%)	20 (4%)
Bougouni	402 (76%)	382 (72%)
Kita	59 (11%)	14 (3%)
Koutiala	318 (60%)	26 (5%)
Nara	120 (23%)	85 (16%)
San	182 (34%)	129 (24%)
Segou	406 (77%)	225 (43%)
Sikasso	151 (29%)	114 (22%)

Table 3 : Percentage of “pure” cultivated pixels and percentage of “pure” cultivated pixels with useable phenology information on the MCD12Q2 product in 2007 in the synoptic stations (polygons of 10 x 10 km).

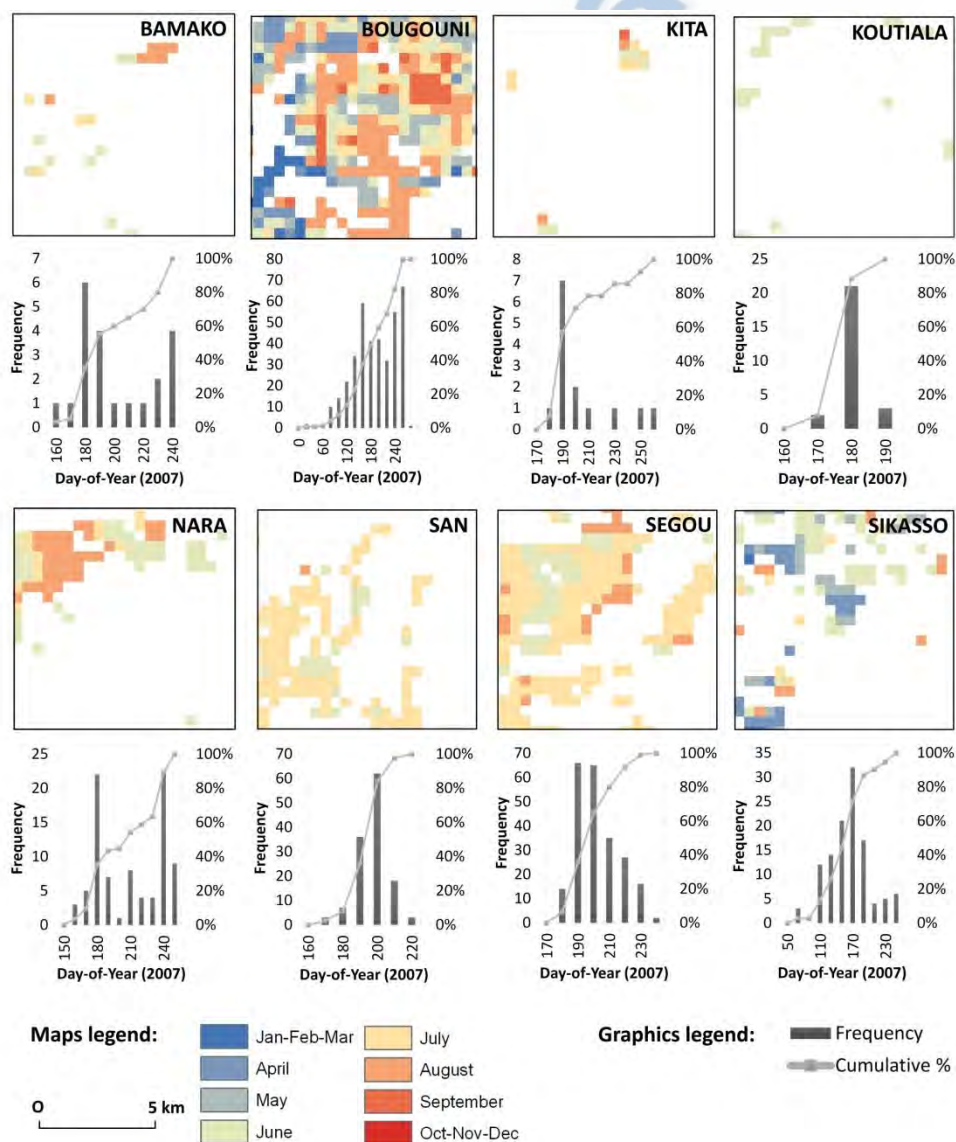


Figure 2. Start-of-season MODIS phenometric for 5 synoptic stations on crop mask in 2007.

Figure 2 shows maps of the Start-of-Season phenometric of the Crop Phenology product at station level, using the cultivated pixels with a single complete cycle in 2007. We can notice a north-south gradient of the Start-of-Season metric. The growing season begins later in the north (San, Segou, Nara), in relation to the rainfall conditions. This figure also highlighted very high standard deviations of Start-of-Season values inside the different stations.

We then analyzed the histograms of the SOS values for each station (Figure 2). For Bougouni, Sikasso, Nara, Bamako and Kita stations, we observed two peaks of SOS values. The first peak is in accordance with the rainfall regime and practices, while the second peak is inconsistent. The second peak occurs very late in the season and correspond to none of the natural or cropped vegetation phenology. We hypothesized that this phenomenon is due to missing values in the EVI profile preventing the Zhang algorithm from fitting the curve and calculating the right curvature change rate. Figure 3 illustrates this phenomenon with the example of three pixels of Bougouni station with different start-of-seasons. For the first pixel (red curve), the three missing values of EVI (syntheses 29, 30 and 31) meant that Zhang algorithm detected the start of the growing season in the synthesis 31, which is actually its maximum. The second pixel (blue curve) showed the same slow start as the red curve (unlike the green curve whose slope is steeper and therefore the algorithm detects a start-of-season earlier), but only two syntheses of EVI missing. In the case of this pixel, the algorithm of Zhang worked.

We removed the pixels corresponding to the second peak, with a threshold value fixed to the minimum frequency between the two peaks and recalculated the mean of the four phenometrics by stations.

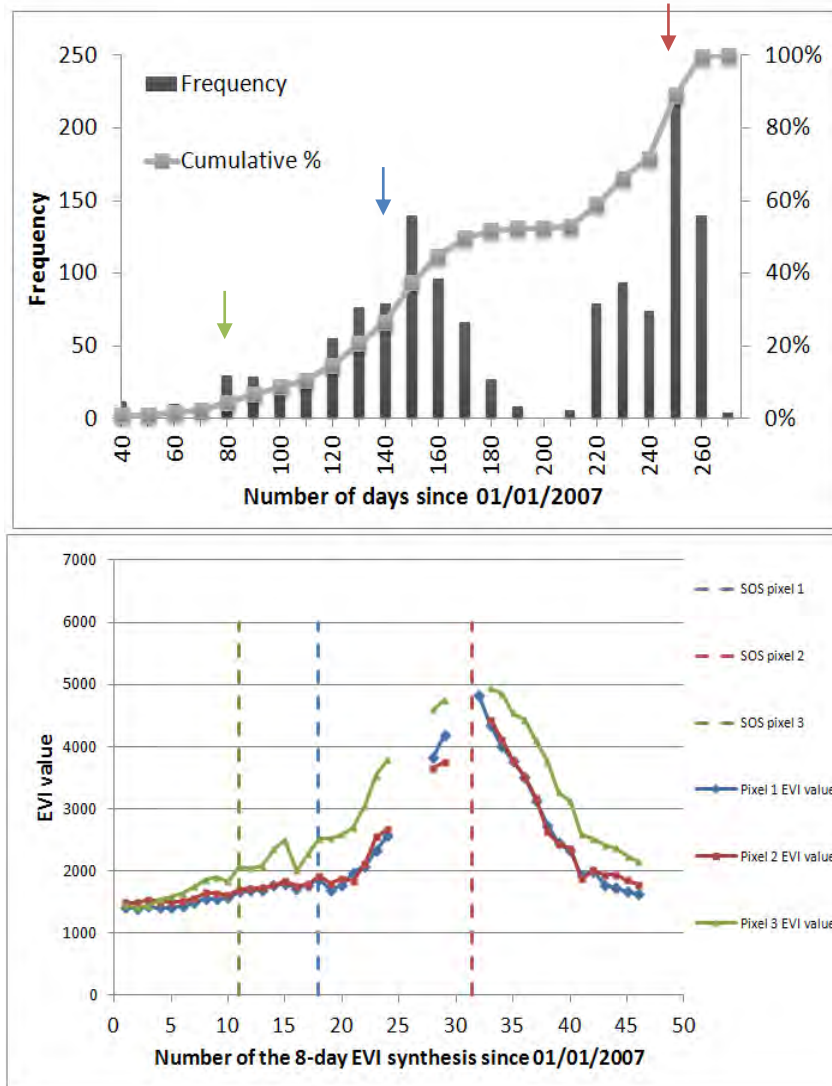


Figure 3. Start-of-Season in Bougouni station: (a) frequency and cumulated histogram, and (b) examples of three pixels EVI profile.

Phenometrics value on the synoptic stations

Table 4 shows the four phenometrics values at station level, with and without histograms corrections as the standard deviations of values inside each station. The north-south gradient of the Start-of-Season metric is confirmed, with a growing season beginning in May in Bougouni and Sikasso, in June in Koutiala and Bamako, and in July in San, Segou and Kita. The Start-of-Maximum of Season is concentrated in the second half of August while the End-of-Maximum is concentrated in September and the End-of-Season in December. The histograms corrections made the start-of-season phenometrics 15 days to one month earlier than the original start-of-season phenometrics.

Concerning the standard deviations of the phenometrics values inside a station, we noticed a wide range of values in the South (Bougouni and Sikasso), with about 30 days, while the start-of-season values are more concentrated around the mean value in the North (15 days for San, 8 days for Nara).

Station	Histograms corrections	SOS (SD in days)	SMAX (SD in days)	EMAX (SD in days)	EOS (SD in days)
Bamako	Before	July-17 (28.7)	Aug.-28 (8.9)	Sept.16 (7.7)	Nov.25 (11.9)
	After	June-26 (12.9)	Aug.25 (8.7)	Sept.15 (7.9)	Nov.29 (12.8)
Bougouni	Before	July-1 (58.7)	Aug.26 (25.7)	Sept.19 (19.5)	Nov.26 (32.4)
	After	May-12 (31.4)	Aug.14 (30.6)	Sept.11 (23.8)	Nov.29 (43.6)
Kita	Before	July-18 (29.4)	Sept.3 (15.3)	Sept.26 (9.8)	Dec.12 (9.0)
	After	July-2 (6.8)	Aug.28 (13.1)	Sept.23 (9.8)	Dec.13 (8.9)
Koutiala	None	June-26 (4.7)	Aug.11 (7.6)	Sept.10 (4.1)	Nov.22 (8.9)
Nara	Before	July-23 (32.7)	Aug.24 (18.8)	Sept.14 (15.6)	Nov.13 (29.2)
	After	June-23 (8.3)	Aug.12 (18.8)	Sept.7 (18.5)	Nov.7 (38.0)
San	None	July-12 (10.2)	Sept.7 (8.6)	Sept.26 (8.5)	Dec.7 (17.3)
Segou	None	July-16 (15.7)	Sept.5 (8.0)	Sept.28 (6.8)	Dec.12 (12.4)
Sikasso	Before	June-5 (43.4)	Aug.19 (22.4)	Sept.22 (19.4)	Dec.11 (31.8)
	After	May-23 (31.8)	Aug.15 (22.8)	Sept.21 (19.7)	Dec.10 (34.0)

Table 4 : Satellite-derived phenometrics : mean and standard deviations values (SD) before and after histogram correction for the 10x10 km polygons around the synoptic stations.

3.2. Model-derived phenometrics

LAI simulation results

LAI profiles were simulated as a yearly-output of SARRAH crop model for each synoptic station. Figure 4 shows four daily LAI profiles for the synoptic station of San in 2007 (one millet variety, one sorghum variety, two fertilizer treatment (yes/no) each). For fertilized crops, we notice earlier Start-Of-Season and maximum of growing season than for non fertilized crops. The LAI profiles also demonstrate that the curve drops sharply at the end of the growing season, on the 10th of September for Sorghum and at the 1st of October for Millet, which correspond to maturity dates.

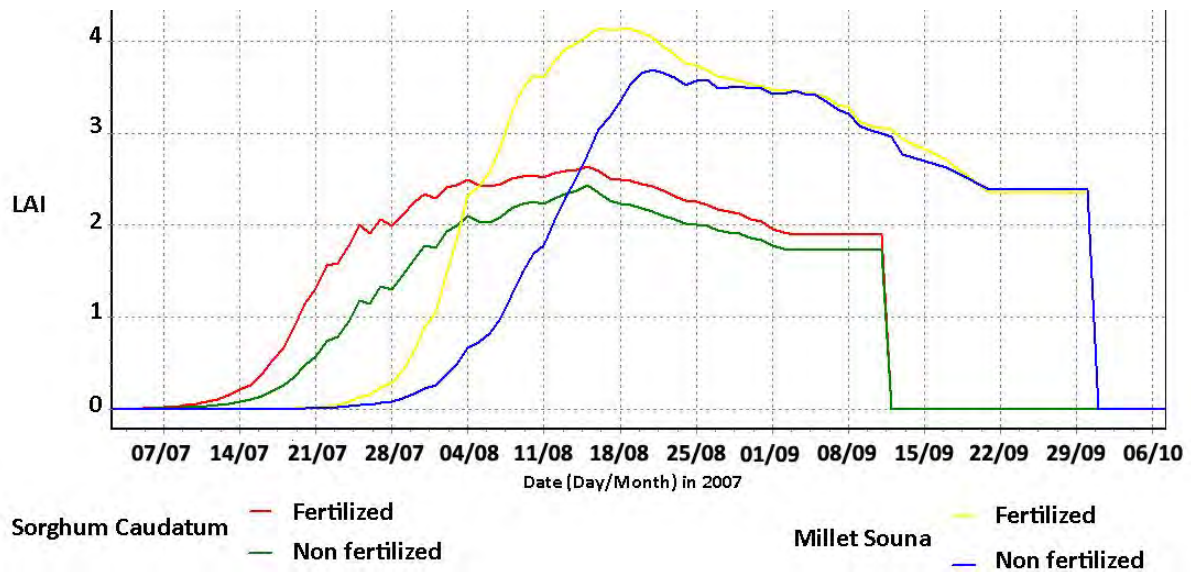


Figure 4. Profiles from SARRAH crop model for San synoptic station in 2007.

Calculation of the LAI-based phenometrics

Using the daily LAI calculated by SARRAH model for each crop of each station (Table 2), the four phenometrics of interest were calculated on R 2.9.1 version software, with the algorithm of Zhang (2003)(Figure 5). These different phenometrics were averaged in order to have an unique value of start, maximum, senescence and end-of-season for each synoptic station.

3.3. Spatial comparison

Satellite-derived phenometrics and model-derived phenometrics of the eight stations were regressed against each other, and compared, for 2007. Table 5 presents the results of the comparison for each phenometric through the RMSE (Root Mean Square Error) of the regression lines. Phenological transition dates on the eight stations have a difference of about 20 days between LAI and MODIS, for the Start-of-Season (SOS), Start-of-Maximum of Season (SMAX) and End-of-Maximum of Season (EMAX), and 40 days for the End-of-Season (EOS) (Table 5).

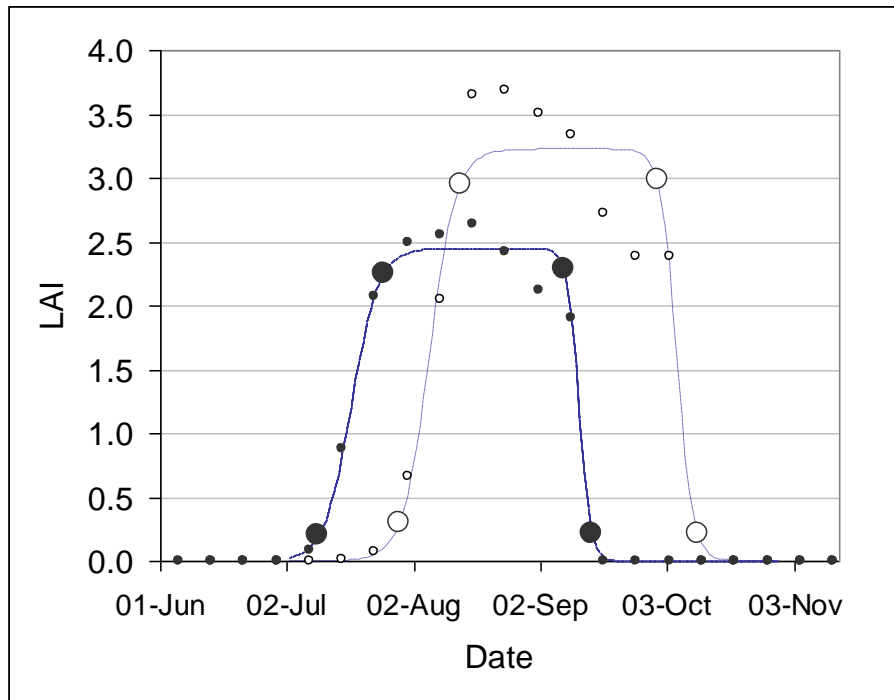
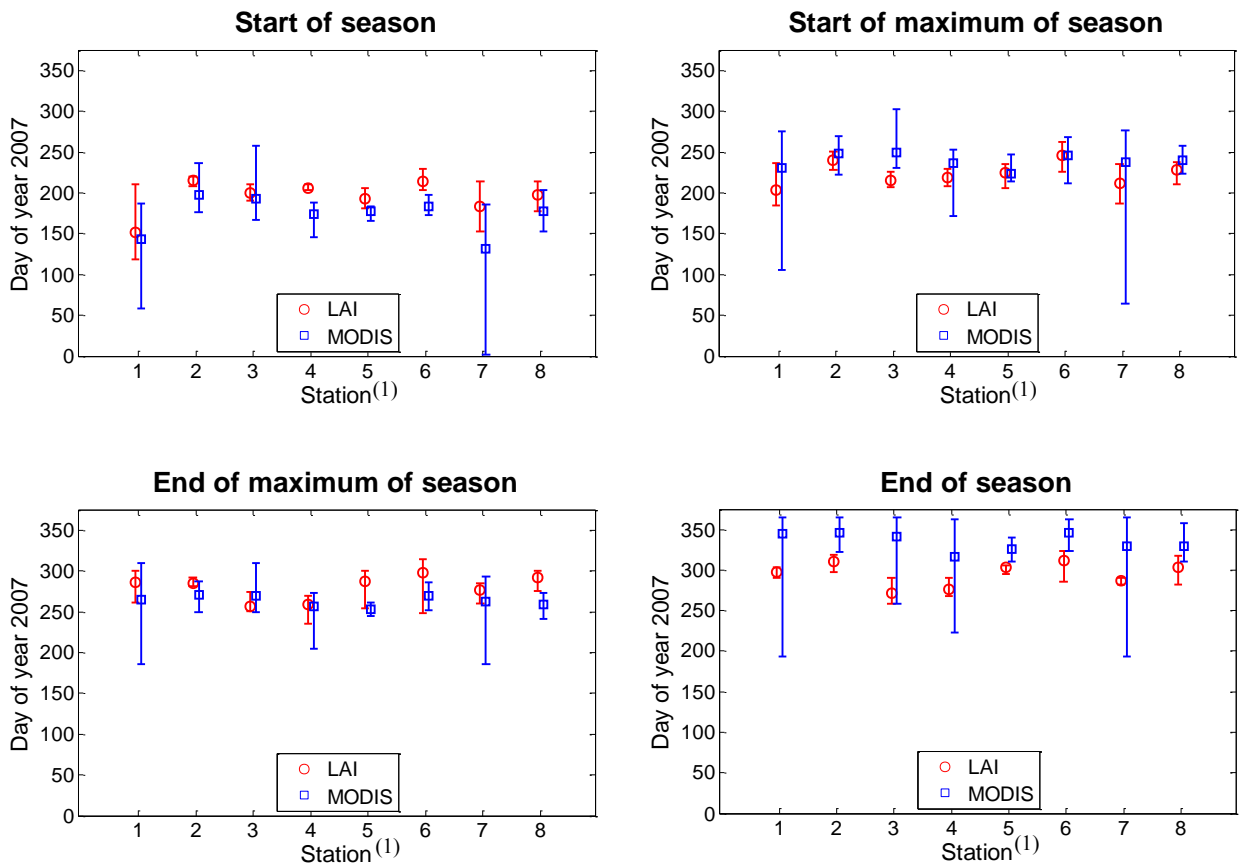


Figure 5. Examples of modeled Zhang non-linear functions calculated for two sets of LAI simulated with SARRA-H crop model on San synoptic station in 2007 (black symbols and solid line represent the Sorghum Caudatum fertilized simulation ; open symbols and dotted line correspond to Millet Souna non fertilized simulation). Small symbols correspond to 8-day sampled LAI, and large symbol correspond to the phenometrics values (SOS, SMAX, EMAX and EOS in chronological order).

Station (year)	Model-Satellite phenometrics RMSE / bias (days)			
	SOS	SMAX	EMAX	EOS
All stations (2007)	26.7 / -23.0	19.5 / 15.1	22.5 / -17.0	42 / 39.8
Segou (2001 to 2008)	21.7 / -13.4	17.0 / 15.1	18.5 / - 18.0	40.9 / 40.7
Sikasso (2001 to 2008)	27.3 / -19.9	21.9 / 20.8	27.7 / -26.5	38.4 / 35.3

Table 5. Spatial (8 stations in 2007) and temporal (8 years for 2 stations) RMSE and bias (in days) between satellite-derived and model-derived phenometrics.

$$\text{Bias} = \sum (\text{Phenometrics MODIS} - \text{phenometrics LAI}) / 8$$



(1) Station 1 = Bamako; Station 2 = Bougouni; Station 3 = Kita; Station 4 = Koutiala; Station 5 = Nara; Station 6 = San; Station 7 = Segou; Station 8 = Sikasso.

Figure 6. Minimum, maximum and mean of phenometrics values by stations in 2007. Blue lines represent MODIS phenology, and black lines represent LAI phenology.

When individual stations are observed (Figure 6), we notice that the MODIS SOS comes always sooner than the LAI SOS. The stations Bougouni, Nara and Kita appeared to have the largest SOS differences between the two data sets (51, 32 and 31 days respectively). For SMAX, the largest differences occurred in Bougouni, San and Sikasso stations (26, 35 and 27 days respectively), with a maximum usually achieved sooner for the LAI phenometric. EMAX is achieved sooner for MODIS phenometric, whereas the EOS comes sooner from LAI profile. In conclusion, Figure 6 shows that the differences between MODIS and LAI phenometrics are consistent between the stations, with a larger range of values for MODIS phenometrics compared to LAI phenometrics.

3.4. Temporal comparison

Annual values of phenometrics were compared for Segou and Sikasso stations during the years 2001 to 2008. Table 5 shows a difference of about 22 days (between 17 and 27 days) between the SOS, SMAX and EMAX phenometrics, and around 40 days for the EOS.

Figure 7 shows a very good correlation between LAI and MODIS phenometrics all together, with a slope close to 1 for both stations. For Segou station (Figure 7a), SOS and SMAX show a high inter-annual variability. The EMAX and the EOS of the different years have globally the same values. On the contrary, EMAX and EOS values from Sikasso have a high inter-annual variability on MODIS product, the different SMAX are grouped and the SOS is more variable from LAI phenometrics than from MODIS.

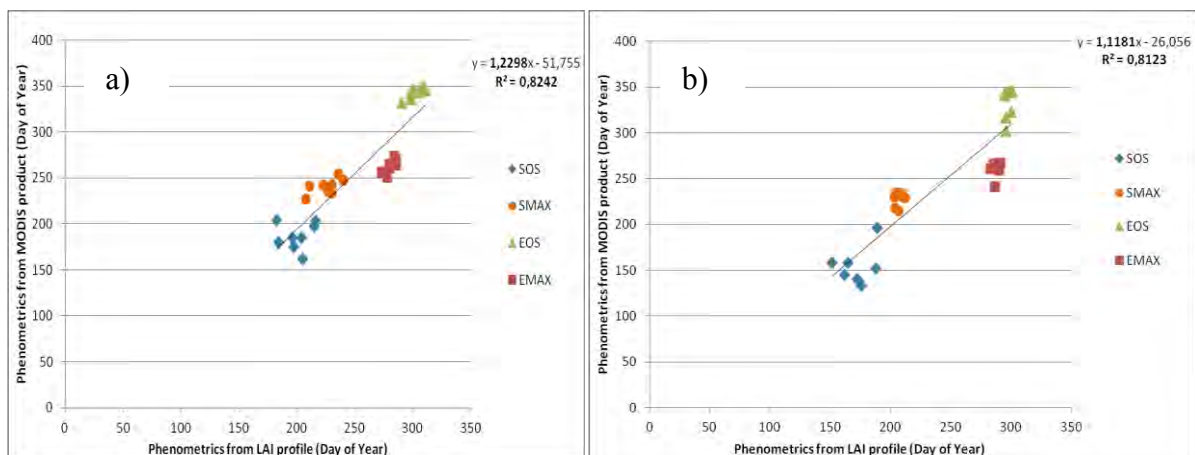


Figure 7. Linear regressions of (a) Segou and (b) Sikasso phenometrics from 2001 to 2008.

4. Discussions

In this study, we proposed to assess whether the information content of the phenological variables from MODIS product (MCD12Q2) may be used to monitor agricultural systems in heterogeneous landscapes of West Africa. To do so, we i) analyzed the data consistency of the MCD12Q2 product by comparing EVI profiles and phenometrics on 8 stations of South Mali for 2007, and ii) compared phenometrics observations from MODIS with simulations from crop SARRA-H model spatially and temporally.

In Mali, it was found that at national level 69% of the cultivated pixels had a complete cycle in 2007. Then, on eight synoptic stations (polygons of 10 km x 10 km), we found that phenological variables were consistent with EVI profile in 2007, after removing pixels with incomplete cycles, with one exception : finding two peaks of high frequency of a Start-of-Season (SOS) value in a same station was not uncommon. A main concern about the MODIS product in Mali is the SOS value problem, although this is not surprising, considering the method using to fit the Zhang algorithm on EVI profile. The "holes" in the EVI time series unfairly suggest two cycles on the MODIS MCD12Q2 product. Yet, the EVI time series are smoothed using a local median-value moving window technique using a centered three-date moving window average (Zhang *et al.*, 2003). Therefore this algorithm is not applied in cases where there are more than two consecutive missing 16-day NBAR EVI values, which is often the case in tropical countries of West Africa. Zhang *et al.*, (2009) estimated that the probability that the absolute error in phenological estimates is greater than 5 days is 20% if there are two consecutive missing values in the time series of 16-day data during vegetation growing seasons. Our evaluation also corroborate earlier findings in a study conducted by Ganguly (2010). Studying missing data of MODIS Global Land Cover Dynamics Product, Ganguly showed that areas in the tropics are characterized by persistent cloud cover, high levels of atmospheric aerosols, or weak seasonality, all of which present substantial challenges for land surface phenology algorithms. They found the higher proportion of missing value in Sahelian zone, and a higher uncertainty for EOS than SOS metric.

Considering that ground data and observations are very difficult to access in Mali, we decided to compare these remotely-sensed phenological variables with modeled-phenological estimates from SARRA-H crop model and verify whether the phenometrics from MODIS MCD12Q2 reproduce well the phenological stages of different crops that we know in Mali. Using an agricultural systems map (Vintrou *et al.*, 2011), crop LAI profiles were simulated as a yearly-output of SARRAH crop model by synoptic station. Unlike the EVI profile, the LAI profile drops sharply at the end of the growing season, corresponding to the maturity date. This stage of phenology called “maturity” can correspond to the harvesting date but in reality, it can also be:

- sooner: to avoid moisture in case of too much rain or during the “soudure” period, which is a period of dearth, when reserves of food from one season are nearly all eaten and before the next harvest.
- later: if there is a lack of working force, or because the farmers practices suggest to let the grains drying up.
- the real harvesting date for grains but with remained stems, implying that the LAI profile would not drop sharply in that case.

This is the reason why the simulated EOS was considered unusable.

On a spatial point of view, comparing the phenological observations and simulations at the station level in 2007, MODIS SOS was detected 20 days before the SARRA-H LAI SOS. Because vegetation and land cover are rarely uniform across 25 ha (500 m²), the exact meaning of remotely-sensed phenological metrics at this spatial resolution is somewhat unclear (Zhang *et al.*, 2006). We hypothesized that this SOS delay may be due to the mixture of natural vegetation and crops on a pixel, as natural vegetation starts at the beginning of the rainy season, while farmers generally wait a certain amount of rain before sowing. Considering SOS and EOS, the difference was about 20 days between LAI and MODIS metrics. SMAX is usually achieved sooner for LAI phenometrics, and EMAX later. In fact, we assumed that LAI profiles simulations suggest that the cultivated patch is homogeneous, with a simultaneous growth for all plants whereas MODIS product considers the mixture of land cover. This is the reason why SMAX is reached faster according to the LAI profile. On the contrary, LAI profiles do not show marked decrease after the maximum of the season and keeps a low slope until the maturity date. It can be an explanation for a later detection of the EOS, generally 5 days before the EOS. Whether it is from MODIS product or from SARRA-H LAI profiles, the EOS phenometric was found irrelevant.

On a temporal point of view, when comparing the MODIS and the LAI phenometrics from 2001 to 2008 in Segou and Sikasso, we verified that the correlations stayed the same, with different climatic conditions. Segou and Sikasso metrics also gave a difference of about 20 days between LAI and MODIS for the three first phenometrics, with a SOS achieved sooner for MODIS product. An interesting point is that MODIS phenometrics in Sikasso are more variable from year to year than in

Segou. Once again this suggests the importance of natural vegetation in a humid site, which varies the transition dates because of non-homogeneous land-cover.

Despite the differences between MODIS and LAI transition dates, the link between the two is clear. LAI phenological metrics is obviously a simplification of reality, an idealized crop growth whereas MODIS metrics are aggregated on 250 m, encompassing several land covers. However, the differences are understood (explainable) and the spatial and temporal comparisons between the two make us think that MODIS MCD12Q2 product can reproduce the evolution of phenological stages of different crop species. This MODIS Global Phenology Product could provide a closer estimation of phenological changes (integrating both rainfall variability and land cover status), continuously in space, and then complete the commonly used aggregated or extrapolated climatic data from weather stations.

In Mali, it has been shown that except for the humid sub-region with dense cloud cover during more than 15 days, phenometrics from MCD12Q2 product are consistent with ground knowledge and modeling. However, in the food security context, the MCD12Q2 product cannot be used because it is generally delivered the next year. Zhang algorithm (Zhang *et al.*, 2003) could be used on the EVI time series during the growing season and it was found that at the middle of the growing season, the algorithm is capable of calculating the first two curvature-change rate of EVI, i.e. the Start-of-Season and the Start-of-Maximum of season. Both metrics could be used by Early Warning Systems for crop monitoring. In fact, in semi-arid monsoonal ecosystems, they have become early warning indicators of possible future reductions in yield due to growing season length (Brown, 2008). Zhang algorithm becomes a new opportunity to monitor the crop growth and these metrics could be used in a crop model to forecast yield (Doraiswamy *et al.*, 2005).

Conclusion

In this study, we hypothesize that estimating directly the phenological indices from the time series of vegetation index obtained at medium resolution space would first improve the spatial resolution of the phenological indices when compared to commonly used climatic data (extrapolated from weather

stations or estimated from low spatial resolution satellite data). Moreover, directly measuring the phenology of the vegetation that depends not only on rainfall but also on the land cover, would provide closer estimations of field reality. This is the reason why we analyzed the recent MODIS MCD12Q2 product and its information content. It was shown to be relevant in Mali, after removing pixels with incomplete cycles or inconsistent Start-of-Season due to cloud cover. When comparing these phenological observations with simulations from SARRA-H crop model spatially and temporally, a difference of 20 to 40 days was found, depending on the phenometric. It can mainly be explained by the natural vegetation mix with crops.

MODIS MCD12Q2 was found to reproduce well the phenological stages of crops in Mali. Food security systems could benefit from a use of this new product, making use of its spatially continuous information content in agro-meteorological modeling, and improving agricultural production estimation. These results now need to be confirmed by extending the method to neighboring countries with similar agricultural landscapes such as Senegal, Niger or Burkina Faso.

References

- Bacci, L., and F.N. Reyniers, 1998. Le futur des céréales photopériodiques pour une production durable en Afrique tropicale semi-aride: actes du séminaire final du Projet STD3 TS3 93.0223" *Risques et intensification: le diagnostic hydrique et minéral des céréales dans les pays sahéliens et les zones communautaires du Zimbabwe*": 27-30 avril 1998, Florence, Italie. Ed. CeS. IA. 270 p.
- Balme, M., S. Galle, and T. Lebel, 2005. Démarrage de la saison des pluies au Sahel: variabilité aux échelles hydrologique et agronomique, analysée à partir des données EPSAT-Niger. *Sécheresse*, 16(1):15-22.
- Baron, C., A. Clopes, P. Perez, B. Muller, and F. Maraux, 1996. Manuels d'utilisation de: SARRAMET 45 p. SARRABIL.
- Brown, M.E., 2008. *Famine early warning systems and remote sensing data*. Springer Verlag, 313 p.
- Brown, M.E., and K.M. de Beurs, 2008. Evaluation of multi-sensor semi-arid crop season parameters based on NDVI and rainfall. *Remote Sensing of Environment*, 112(5):2261-2271.
- De Beurs, K.M., and G.M. Henebry, 2005. Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high latitude transects. *Global Change Biology*, 11(5):779-790.
- Dingkuhn, M. et al., 2003. Decision support tools for rainfed crops in the Sahel at the plot and regional scales. In: Struif-Bontkes, T.E., Wopereis, M.C.S. (Editors),. *A Practical Guide to Decision-support Tools for Agricultural Productivity and Soil Fertility Enhancement in Sub-Saharan Africa*, IFDC, CTA, pp.127–139.
- Doraiswamy, P.C., T.R. Sinclair, S. Hollinger, B. Akhmedov, A. Stern, and J. Prueger, 2005. Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment*, 97(2):192-202.
- Funk, C., and M.E. Budde, 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sensing of Environment*, 113(1):115-125.
- Ganguly, S., M.A. Friedl, B. Tan, X. Zhang, and M. Verma, 2010. Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, 114(8):1805-1816.
- Heumann, B.W., J. Seaquist, L. Eklundh, and P. Jonsson, 2007. AVHRR derived phenological change in the Sahel and Soudan, Africa, 1982-2005. *Remote Sensing of Environment*, 108(4):385-392.
- Huete, A., K. Didan, T. Miura, E.P. Rodriguez, X. Gao, and L.G. Ferreira, 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1-2):195-213.

- Kathuroju, N., M.A. White, J. Symanzik, M.D. Schwartz, J.A. Powell, and R.R. Nemani, 2007. On the use of the advanced very high resolution radiometer for development of prognostic land surface phenology models. *Ecological Modelling*, 201(2):144-156.
- Kouressy, M. et al., 2008. Adaptation of Malian sorghums to climate variability. *Adaptation des sorghos du Mali a la variabilite climatique*, 17(2):95-100.
- Marteau, R., B. Sultan, V. Moron, A. Alhassane, C. Baron, and S.B. Traoré, 2011. The onset of the rainy season and farmers' sowing strategy for pearl millet cultivation in Southwest Niger. *Agricultural and Forest Meteorology*, 151(10):1356-1369.
- Reed, B.C., J.F. Brown, D. VanderZee, T.R. Loveland, J.W. Merchant, and D.O. Ohlen, 1994. Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*:703-714.
- Rojas, O., F. Rembold, J. Delincé, and O. Léo, 2011. Using the NDVI as auxiliary data for rapid quality assessment of rainfall estimates in Africa. *International Journal of Remote Sensing*, 32(12):3249-3265.
- Sakamoto, T., M. Yokozawa, H. Toritani, M. Shibayama, N. Ishitsuka, and H. Ohno, 2005. A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96(3-4):366-374.
- Schwartz, M.D., and J.M. Hanes, 2010. Intercomparing multiple measures of the onset of spring in eastern North America. *International Journal of Climatology*, 30(11):1614-1626.
- Sultan, B., C. Baron, M. Dingkuhn, B. Sarr, and S. Janicot, 2005. Agricultural impacts of large-scale variability of the West African monsoon. *Agricultural and Forest Meteorology*, 128(1-2):93-110.
- Traoré, S.B. et al., 2011. Characterizing and modeling the diversity of cropping situations under climatic constraints in West Africa. *Atmospheric Science Letters*, 12(1):89-95.
- Vintrou, E., A. Begue, C. Baron, and D. Lo Seen, 2011. Mapping Fragmented Agricultural Systems in the Sudano-Sahelian region using Random Forest and Metrics of Coarse Resolution MODIS Imagery. *Photogrammetric Engineering and Remote Sensing.*, In revision.
- Vintrou, E., A. Desbrosse, A. Bégué, S. Traoré, C. Baron, and D. Lo Seen, 2012. Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation*, 14(1):83-93.
- Wessels, K., K. Steenkamp, G. Von Maltitz, and S. Archibald, 2010. Remotely sensed vegetation phenology for describing and predicting the biomes of South Africa. *Applied Vegetation Science*, 14(1):49-66.
- White, M.A. et al., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, 15(10):2335-2359.

- Zhang, X., M. Friedl, and C. Schaaf, 2009. Sensitivity of vegetation phenology detection to the temporal resolution of satellite data. *International Journal of Remote Sensing*, 30(8):2061-2074.
- Zhang, X., M.A. Friedl, and C.B. Schaaf, 2006. Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *J. Geophys. Res.*, 111(G04017):G04017.
- Zhang, X., M.A. Friedl, C.B. Schaaf, A.H. Strahler, J.C.F. Hodges, F. Gao, B.C. Reed, and A. Huete, 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3):471-475.
- Zhang, X., M.A. Friedl, C.B. Schaaf, A.H. Strahler, and Z. Liu, 2005. Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. *Journal of Geophysical Research*, 110(D12):1-14.

ARTICLE 4 :

Mining Sequential Patterns from MODIS Time Series for Cultivated Area Mapping

Advancing Geoinformation Science for a Changing World : 45-62.

Mining Sequential Patterns from MODIS Time Series for Cultivated Area Mapping

Yoann Pitarch¹, Elodie Vintrou^{2;4}, Fadi Badra^{3;4},

Agnès Bégué^{2;4} and Maguelonne Teisseire^{3;4}

¹ LIRMM - CNRS - UM2 ² CIRAD ³ Cemagref

⁴ TETIS - 500 rue Jean-François Breton, F-34093 Montpellier, France

e-mail: pitarch@lirmm.fr, {firstname.lastname}@teledetection.fr

Abstract. To predict and respond to famine and other forms of food insecurity, different early warning systems are using remote analyses of crop condition and agricultural production, using satellite-based information. To improve these predictions, a reliable estimation of the cultivated area at national scale must be carried out. In this study, we developed a data mining methodology for extracting cultivated domain patterns based on their temporal behavior as captured in time-series of moderate resolution remote sensing MODIS images.

Keywords: Knowledge Discovery, Data Mining, MODIS time series

1 Introduction

The northern fringe of sub-Saharan Africa is a region that is considered particularly vulnerable to climate variability and change, and food security remained there a major challenge.

One of the preliminary stages necessary for analyzing such impacts on agriculture and food security is a reliable estimation of the cultivated domain at national level, a scale compatible with climate change studies. For that purpose, different early warning systems such as FEWS and JRC-MARS use global land cover maps but they are generally focused on large

ecosystems, and are not suitable for fragmented and heterogeneous African landscapes. Recent moderate-resolution sensors, such as MODIS/TERRA, with spatial resolutions as low as 250 m, offer new possibilities in the study of agricultural lands. With this increase in spatial resolution, the detection of groups of fields can now be considered. The low and medium spatial resolutions do not, by themselves, provide a completely satisfactory representation of the landscape but are compensated for by a large coverage area and by an excellent temporal resolution.

This brings us to the question whether moderate-resolution satellite data, in combination with external data (fields surveys, climate etc.) can provide a correct assessment of the distribution of the cultivated domain at country level. It is expected that more consistent information on vegetation would allow monitoring Sahelian rural landscapes with better continuity, thereby providing relevant information for early warning systems.

In this study, we develop a datamining methodology to extract relevant sequential patterns to describe cultivated areas. These patterns are obtained from the static description and the temporal behavior as captured in time-series of moderate resolution remote sensing images. We applied this methodology in Mali, a representative country of the Sahel Belt of Africa. Both the temporal and spatial dimensions add substantial complexity to data mining tasks. A prioritization is needed to reduce the search space and to allow the relevant pattern extraction. We thus adopt a two-step approach: (1) identification of relevant descriptors per class (2) associated pattern mining from MODIS time series.

2 The data description

2.1. Study area

Mali is, after Senegal, the second westernmost country of West Africa around Latitude 14°N. It displays a South North climatic gradient that ranges from subtropical to semi-arid, and which extends further north to desertic. As for other West African countries along the same latitudinal belt, food security relies on an adequate supply of rainfall during monsoon season. This country can therefore be considered representative of the Soudano-Sahelian zone, where a strong dependence on rainfed agriculture implies vulnerability to major changes due to climate and human activities, and hence requires specific attention. A particular attention was paid to three zones in Bani catchment, mainly located in Southern Mali (Table 1).

Table 1. Main characteristics of the three studied sites

Site name (eco-climatic zone)	Mean annual rainfall	Main crops	Natural vegetation type
Cinzana (Soudano-Sahelian)	600 mm	Millet, sorghum	High proportion of bare soils and sparse vegetation
Koutiala (Soudano-Sahelian)	750 mm	Cotton, millet, sorghum	Large areas of semi-open and closed natural vegetation
Sikasso (Soudanian)	1000 mm	Maize, cotton, fruit crops	Dense natural vegetation

2.2. Data

Field data Fields surveys were conducted in Mali during the 2009 and 2010 cropping seasons (from May to November) in order to characterize Soudano-Sahelian rural landscapes. Three sites (Cinzana, Koutiala, Sikasso) were selected to sample the main agro-climatic regions of Central and Southern Mali (Table 1). 980 GPS waypoints were registered, and farmers were interviewed. Each waypoint was transformed into a polygon whose center has been affected a land use.

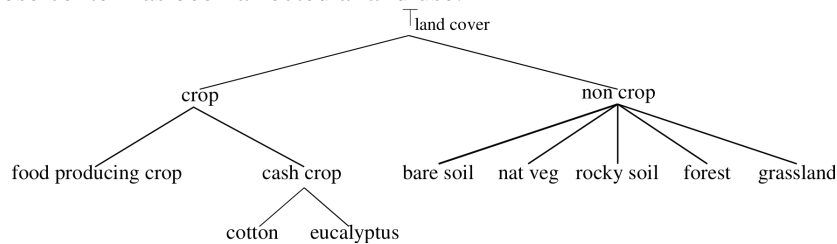


Fig. 1. Crop hierarchy

External data

Six static descriptors were also used to characterize the site surveys: soil type, distance to the village, distance to the river, rainfall, ethnic group, name of the village. The domains of associated data values are detailed in Table 2.

Images data

MODIS time series: The NASA Land Process Distributed Active Archive Center (LP DAAC) is the repository for all MODIS data. Amongst MODIS products, we selected the ‘Vegetation Indices 16-Day L3 Global

250 m SIN Grid' temporal syntheses for our study. For Mali, a set of 12 MODIS 16-days composite normalized difference vegetation index (NDVI) images (MOD13Q1/V05 product) at a resolution of 231.6 m were acquired for 2007 (we keep the best quality composite image out of two for each month).

The year 2007 was chosen to overlap with the more recent high-resolution data available. We assume that the observed classes of land use remained globally unchanged from 2007 to 2009 (fields surveys in 2009). However, Malian farmers practice "crop rotation". It is the practice of growing a series of dissimilar types of crops in the same area in sequential seasons for various benefits such as to avoid the build up of pathogens and pests that often occurs when one species is continuously cropped, improving soil structure and fertility. Thus, we decided to only consider the two higher levels of the crop hierarchy (Figure 1).

Remotely sensed indices used

- *Normalized Difference Vegetation Index*: NDVI is one of the most successful index to simply and quickly identify vegetated areas and their "condition", providing a crude estimate of vegetation health. It displays the relationship between the quantity of chlorophyll in leaves with red and near infrared wavelength, so that NDVI image is used to search vegetation as estimating biomass, plant productivity, fractional vegetation cover (Rouse, 1974).

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

where RED and NIR stand for the spectral reflectance measurements acquired in the red and near-infrared regions, respectively. In general, NDVI values range from -1.0 to 1.0, with negative values indicating clouds and water, positive values near zero indicating bare soil, and higher positive values of NDVI ranging from sparse vegetation (0.1 - 0.5) to dense green vegetation (0.6 and above). Furthermore, different land covers exhibit distinctive seasonal patterns of NDVI variation. Crops have generally a distinct growing season and period of peak greenness, which allows the discrimination with other types of land cover.

- *Texture*: Information content in a digital image is expressed by the 'intensity' of each pixel (*i.e.* tone or color) and by the spatial arrangement of pixels (*i.e.* texture, shape, and context) in the image. Traditionally, tone (*i.e.* spectral intensity) has been the primary focus for most image analysis and hence information extraction in remote-sensing studies. However, texture analysis is examined as an important contributor to scene information extraction. The majority of image classification procedures,

particularly in operational use, relies on spectral ‘intensity’ characteristics alone and thus is oblivious to the spatial information content of the image. Textural algorithms, however, attempt to measure image texture by quantifying the distinctive spatial and spectral relationships between neighboring pixels. In response to the need for extracting information based on the spatial arrangement of digital image data, numerous texture algorithms have been developed. Statistical approaches, such as those developed by Haralick et al. (1973) make use of gray-level probability density functions, which generally are computed as the conditional joint probability of pairs of pixel gray levels in a local area of the image. In this study, we used four Haralick textural indices (Haralick, 1979) calculated on the MODIS time series: “variance”, “homogeneity”, “contrast” and “dissimilarity” on ENVI®. The Haralick textural features describe the spatial distribution of gray values and the frequency of one gray tone appearing with another gray tone in a specified distance and at a specified angle. The generation of these indices is based on different orientations of pixels pairs, with specific angle (horizontal, diagonal, vertical, co-diagonal) and distance, called “patterns”. We determined empirically a size of pattern of 15 pixels for MODIS, which is the smaller patch repeated in different direction and distance.

Table 2. The domains of external data values

Dimension D_i	Values
Id-pt	{1,2, ... ,980}
Date	{1,2, ... ,12}
Site name	{cinzana, koutiala, sikasso}
Crop	{millet, sorghum, rice...}
Soil type	{clay loam, sandy loam, gravelly soils...}
Distance to the village	[200 ; 30000]
Distance to the river	[5000 ; 68000]
Rainfall (mm)	{600, 750, 1000}
Ethnic group	{bambara, senoufo, bozo...}
Name of the village	{dioforongo, tigui, sanando...}

3 Motivating Example

In order to illustrate our approach, we consider the following example that will be used throughout the paper as a running example. Let us consider a relational table T in which $NDVI$ values by field are stored. More precisely, we assume that T is defined over six dimensions (or attributes)

as shown in Table 3 and where: D is the date of statements (considering two dates, denoted by 1 and 2), I is the field identifier (considering four different fields, denoted by F1, F2, F3, and F4), C is the crop type (considering two discretized values, denoted by FP (food-producing) and NFP (non food-producing)), S is the soil type (considering three different soil types, denoted by GS (gravelly soils), SL (sandy loam) and CL (clay loam)), DV is the distance between the associated field and the nearest village (considering two discretized values, denoted by near and far), $NDVI$ stands for the NDVI value associated to each field at each timestamp (considering 4 abstract values n_1, n_2, n_3 and n_4).

We consider five sets of dimensions as follows:

- (i) the dimension D representing the date,
- (ii) the dimension I representing the identifier,
- (iii) the dimensions S and DV , that we call *static dimensions* or *descriptors* (values of these dimensions associated to a given field do not evolve over time),
- (iv) the dimension $NDVI$, that we call *dynamic dimension* or *indicators* (values of these dimensions associated to a given field evolve over time),
- (v) the dimension C that we call the class.

For instance, the first element of T (Table 3) means that the field 1 is a food-producer crop composed by CL, near to a village and that, at date 1, the NDVI value was n_1 . Observing in great details the static attribute values per class, some comments should be made. First, food-producing crops are always located near to the village whereas the soil composition is changing. Similarly, non food-producing crops are always cultivated on GS whereas the distance to the nearest village is changing.

A first interpretation to these comments is that the dimension DV appears to be decisive to identify food-producing crops whereas the dimension S appears to be decisive to identify non food-producing crops. Consequently, it is pertinent to only consider decisive dimensions per crop to mine representative rules. Once static dimensions have been filtered, the dynamic dimension ($NDVI$) is considered in order to mine sequential patterns characterizing crops.

Let us suppose that we look for sequences that are verified by all the crops in a given class. Under this condition, the pattern $\langle (near, n_1)(near, n_2) \rangle$ (meaning that fields located near to a village and where the NDVI statement are n_1 at a certain date and n_2 after) characterizes the food-producing crops and the pattern $\langle (GS, n_3) \rangle$ characterizes the non food-producing crops. It should be noted that representative rules per class could be composed by values of different dimensions. In the rest of this

paper, we describe the adopted methodology to determine the decisive attributes per class and how the table T is subdivided and mined to obtain representative rules per class.

Table 3. Table T

D (Date)	I (Id)	C (Crop)	S (Soil)	D (Distance to village)	NDVI (NDVI value)
1	F1	FP	CL	near	n_1
1	F2	FP	SL	near	n_2
1	F3	NFP	GS	far	n_3
1	F4	NFP	GS	near	n_4
2	F1	FP	CL	near	n_1
2	F2	FP	SL	near	n_2
2	F3	NFP	GS	far	n_3
2	F4	NFP	GS	near	n_4

4 Preliminary Definitions

In this section, concepts and definitions concerning multidimensional sequential patterns are presented and are inspired by the notations introduced in Plantevit et al. (2010). For each table defined on the set of dimensions D , we consider a partition of D into three sets: D_t for the temporal dimension, D_A for the analysis dimensions and D_R for the reference dimension. Each tuple $c = (d_1, \dots, d_n)$ can thus be denoted $c = (r, a, t)$ with r the restriction on D_R , a the restriction on D_A and t the restriction on D_t .

Definition 1. (Multidimensional Item) A multidimensional item e defined on $D_A = \{D_{i_1}, \dots, D_{i_m}\}$ is a tuple $e = (d_{i_1}, \dots, d_{i_m})$ such that $\forall k \in [1, m], d_{i_k} \in \text{Dom}(D_{i_k})$.

Definition 2. (Multidimensional Sequence) A multidimensional sequence S defined on $D_A = \{D_{i_1}, \dots, D_{i_m}\}$ is an ordered non empty list of multidimensional items $S = \langle e_1, \dots, e_l \rangle$ where $\forall j \in [1, l], e_j$ is a multidimensional item defined on D_A .

Considering our running example and that $D_A = \{DV, NDVI\}$, $(near, n_1)$ is a multidimensional item.

$\langle (near, n_1)(near, n_2) \rangle$ is a multidimensional sequence on D_A .

Remark In the original framework of sequential patterns (Agrawal and Srikant, 1995), a sequence is defined as an ordered non-empty list of itemsets where an itemset is a non empty set of item. Nevertheless, in the scope of this paper, we only consider sequences of item since at each date, one and only one item can occur for each field. For instance, only one NDVI statement is available per date and field.

An identifier is said to support a sequence if a set of tuples containing the items satisfying the temporal constraints can be found.

Definition 3. (Inclusion) *An identifier $r \in Dom(D_R)$ supports a sequence $S = \langle e_1, \dots, e_l \rangle$ if $\forall j \in 1 \dots l, \exists d_j \in Dom(D_i), \exists t = (r, e_j, d_j) \in T$ where $d_1 < d_2 < \dots < d_l$.*

Definition 4. (Sequence Support) *Let D_R be the reference dimension and T the table. The support of a sequence S is:*

$$support(S) = \frac{|\{r \in Dom(D_R) \text{ s.t. } r \text{ supports } S\}|}{|Dom(D_R)|}$$

Definition 5. (Frequent Sequence) *Let $minSupp \in [0, 1]$ be the minimum user-defined support value. A sequence S is said to be frequent if $support(S) \geq minSupp$.*

Considering the definitions above, an item can only be retrieved if there exists a frequent tuple of values from domains of D_A containing it. For instance, it can occur that neither $(CL, near)$ nor $(SL, near)$ nor $(GS, near)$ is frequent whereas the value $near$ is frequent. Thus, Plantevit et al. (2010) introduces the *joker* value $*$. In this case, we consider $(*, near)$ which is said to be *jokerized*.

Definition 6. (Jokerized Item)

*Let $e = (d_1, \dots, d_m)$ a multidimensional item. We denote by $e[d_i/\delta]$ the replacement in e of d_i by δ . e is said to be a jokerized multidimensional item if: (i) $\forall i \in [1, m], d_i \in Dom(D_i) \cup *$, (ii) $\exists i \in [1, m]$ such that $d_i \neq *$ and (iii) $\forall d_i = *, \exists \delta \in Dom(D_i)$ such that $e[d_i/\delta]$ is frequent.*

A *jokerized* item contains at least one specified analysis dimension. It contains a * only if no specific value from the domain can be set. A *jokerized* sequence is a sequence containing at least one *jokerized* item.

5 Method

5.1. Overview

In this paper, we aim at discovering representative rules in order to characterize crop classes and propose a four-step method to achieve this issue. It should be noticed that the crop classes depends on the user-defined interest level of the crop hierarchy displayed in Fig 1. For instance, assuming that the user would like to discover representatives rules for classes in the second level of the hierarchy, the set of classes will be $\{food-producing, non\ food-producing, other\}$. Such rules could be specific to one site or corresponding to all sites.

These four steps are illustrated in Figure 2 and are briefly presented here:

- *The raw database pretreatment.* During this phase, two actions are performed. First, since the raw database stores crops at the lowest level of the hierarchy, these attributes values must be rewritten to match with the user-defined interest level. Second, sequential pattern mining aims at discovering frequent relations in a database but is not well adapted to mine numerical attributes (*e.g.*, distance to the village, NDVI value) due to the huge domain of definition for such attributes. Consequently, numerical attributes are discretized to improve the sequential pattern mining phase.
- *The build of projected databases.* Since we would like to obtain representative rules per class, the pretreated database is projected on the different class values.
- *The decisive attribute computation.* During this step, a search is performed on each projected databases in order to find and delete non-decisive static attributes dimensions. Intuitively, a static attribute is said to be non-decisive if none of its value allows characterizing the class. More precisely, we guarantee that if it does not exist any value of a static attribute appearing in at least $minSupp\%$ in the projected database, the representative rules associated to this class won't ever contain specific values of this static attribute. Consequently, it is useless to consider it in the rest of the process and this attribute will be removed from the projected database

- *The sequential pattern mining.* Once the projected databases were cleaned up, the algorithm M^2SP is applied on each database. We obtain a set of frequent pattern for each class.

These steps are now detailed in the following subsections.

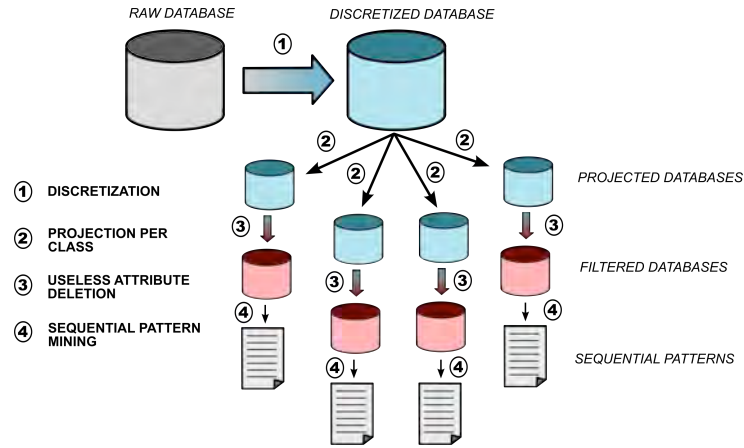


Fig. 2. Overall schema of the proposed methodology

5.2. The Database Pretreatment and Projections

The first performed treatment is the rewriting of the database in order to make the crop attribute values and the user-defined interest level match. This is motivated by two reasons. First, as mentioned in Section 2, mining representative rules for precise crop values is not consistent. As a consequence, crop attribute values must be rewritten to, at least, the above level of granularity. Second, since the hierarchy is composed by two workable levels of granularity, it is interesting to allow user to choose which level must be explored. Consequently, a user-defined parameter, *Level*, is introduced to specify which level of granularity to mine. Thus, rules representing different generalized classes can be compared. An illustration of this database rewriting is displayed in Table 3 where crop attribute values have been already generalized to the second level of granularity (*i.e.*, $Dom(Crop) = \{FP, NFP\}$).

A second pretreatment is the discretization of numerical attributes. This discretization is motivated by the use of the sequential pattern technique to mine representative rules. Indeed, sequential pattern algorithms aim at discovering frequent relation among the fields belonging to the same class.

When dealing with numerical attributes, two values can be considered as different items even if they are very close. For instance, let us consider that the distance to the nearest village is 200 m for field 1 and is 205 m for the field 2. These two distances would have been considered as different items by the M^2SP algorithm without discretization even if they are semantically closed. In our application case, numerous attributes are numerical. Thus, this discretization is necessary. Numerous discretization techniques can be found in the literature (Catlett, 1991). Section 6 details the adopted technique per numerical attribute.

Once the database was pretreated, projection per crop attribute values is performed. Indeed, this is motivated by the fact that we would like to discover representative rules per class. Thus, an intuitive way to achieve this goal is to subdivide the pretreated table into smaller ones associated to each class. Regarding our running example, Tables 4 and 5 display the result of this projection.

Table 4. T_{FP} , the FP projected table

D (Date)	I (Id)	S (Soil)	D (Distance to village)	NDVI (NDVI value)
1	F1	CL	near	n_1
1	F2	SL	near	n_1
2	F1	CL	near	n_2
2	F2	SL	near	n_2

5.3. Dimensionality Reduction

Once the projected databases were built, a search is performed on the static attributes of each database in order to identify useless static attributes. Intuitively, if values of a static attribute are very changing, this attribute is not really characteristic to this class. So, it can be deleted from the projected class. The main advantage of such a strategy is to reduce the search space during the sequential pattern mining phase.

Indeed, it is empirically shown in Plantevit et al. (2010) that the number of dimensions exponentially impacts on both the memory consumption and the extraction time. Whereas traditional applications domains often deal with few analysis dimensions, this point can be very problematic in our context since the number of both static and dynamic dimensions can be

high. For instance, experiment results presented in Section 6 concern at most 12 dimensions. Traditional multidimensional sequential pattern approaches cannot efficiently deal with such a number of analysis dimensions.

Moreover, independently to performance considerations it is important to notice that the higher the number of dimensions, the higher the number of extracted patterns. Since experts will exploit these extracted patterns, reducing the dimensionality without loss of expressivity is very relevant to improve the result analysis phase.

Table 5. T_{NFP} , the NFP projected table

D (Date)	I (Id)	S (Soil)	D (Distance to village)	NDVI (NDVI value)
1	F3	GS	far	n_2
1	F4	GS	near	n_3
2	F3	GS	far	n_4
2	F4	GS	near	n_3

To perform such a dimensionality reduction, we proceed as follows:

Let $minSupp$ be the user-defined parameter used during the sequential pattern-mining phase, T_i be a projected database and $D_j \in T_i$ be one the static dimension in T_i . It can be easily proved that if it does not exist any value of D_j appearing in at least $minSupp * |T_i| / \text{tuples in } T_i$ (where $|T_i|$ is the size of T_i), it cannot exist any sequential pattern extracted from T_i where a value of D_j appears. If so, the dimension D_j is considered as useless and is thus deleted from T_i .

A direct corollary of this propriety is that if an attribute is retained, it will exist at least one sequential pattern containing a value of D_j . To illustrate this affirmation, let us consider, T_{FP} , the projected database presented Table 4 and $minSupp = 1$. The two static attributes are DV and S . Regarding the DV attribute, all the tuples share the same value (*near*). This attribute is considered as useful for the next step and is thus retained.

Let us now consider the S attribute. Here, no value satisfies the $minSupp$ condition. As a consequence, S is deleted from this table. To attest the consequence of such a strategy, let us consider, SP_{FP} , the set of the multidimensional sequential patterns extracted from T_{FP} where $minSupp = 1$, $D_i = D$, $D_r = I$ and $D_A = \{C, S, NDVI\}$ (i.e., all the static and dynamic attributes are considered). Under these conditions, $SP_{FP} = \{ \langle (*, near, n_i) \rangle, \}$

$\langle (*,near,n_1)(*,near,n_2)\rangle$. It is readily noticeable that DV occurs in SP_{FP} but not S .

It is interesting to observe that the set of useful attributes per class can be different. As a consequence, independently to the values of these attributes, attributes themselves can be representative to one class. For instance, performing the above described dimensionality reduction technique on T_{NFP} (see Table 5), S but not DV will be retained this time.

5.4. Mining Representative Rules

Once useless attributes have been deleted, the M^2SP algorithm is applied on each projected and cleaned database T_i such that $minSupp$ is the same as defined during the previous step, $D_i=D$, $D_R=I$ and D_A is composed by the retained static attributes and the dynamic attributes. We note SP_{T_i} the set of sequential patterns extracted from T_i . For instance considering T_{FP} and $minSupp = 1 \langle (near,n_1)(near,n_2)\rangle$ is a frequent sequence meaning that NDVI values equals n_1 and then n_2 is a frequent behavior for fields cultivating food-producing crops located near to a village.

6 Experiment Study

In this section, we present experiments to evaluate the feasibility and efficiency of our approach. Throughout the experiments, we answer the following questions inherent to efficiency issues: *Does the dimensionality reduction technique allow deleting useless static attributes without loss of information? Does the mining process allow discovering discriminating patterns per class? Does the texture data allow a better discriminating pattern extraction than only considering NDVI values?* The experiments were performed on Intel(R) Xeon(R) CPU E5450 @ 3.00GHz with 2GB of main memory, running Ubuntu 9.04. The methods were written in Java 1.6. We first describe the adopted protocol and then present and discuss our results.

6.1. Protocol

The method was evaluated on the dataset described in Section 2. This dataset contains 980 distinct fields and a MODIS time series of length 12 is associated to each field. The 7 static dimensions and the 5 dynamic dimensions were the same as described in Section 2.

As mentioned in Section 5, a discretization step is necessary to efficiently mine frequent patterns. The adopted discretization methods are as follows:

- EQUI-WIDTH technique (the generated intervals have the same width) was used for distance village and distance river attributes
- EQUI-DEPTH technique (the generated intervals have the same size) was used for the other numerical attributes.

In this experiment study, two sets of classes were considered. The first set of classes, denoted by B aims at discovering patterns allowing the distinction between food-producing crops (FP), non food-producing crops (NFP) and non crops (OTHER). The second set of classes, denoted by C , aims at discovering patterns allowing the distinction of more general classes: crops (Cr) and non crops (NCr).

In order to evaluate the impact of texture data in discriminating pattern extraction, we consider a first configuration, denoted by *Default*, where all the dynamic attributes were used. On the contrary, the configuration denoted by *NDVI* is only composed by NDVI values as dynamic attribute.

Three experiment results are presented and discussed in this section:

1. A first experiment was performed to evaluate the number of retained static attributes according to two *minSupp* values.
2. A second experiment was performed to evaluate the number of discriminating patterns. Here, *discriminating* means that a pattern appears in one class but not in the others.
3. Finally, the last experiment was performed to observe the discriminating dimension values according the two above described configurations.

6.2. Results and Discussion

Table 6 displays the the retained attributes according to the two sets of classes and two *minSupp* values. First of all, it can be noticed that the *minSupp* threshold value has an obvious impact on this attribute selection. Indeed, considering *minSupp=0.5*, more than half of the attributes were deleted. Moreover, it is interesting to observe that the retained attributes per class and set of classes are roughly identical.

Table 6. Retained static attributes under default configuration

Level	Class	Static attributes (<i>minSupp</i> =0.5)						Static attributes (<i>minSupp</i> =0.3)					
		Distance village	Site name	Ethnic group	Rainfall	Soil type	Distance river	Village Distance village	Site name	Ethnic group	Rainfall	Soil type	Distance river
B	FP	x	x	x	x	x		x	x	x			
	NFP	x	x	x	x	x		x	x				
	OTHE	x	x	x	x	x		x	x				
	R												
C	Cr	x	x	x	x	x		x	x				
	NCr	x	x	x	x	x		x	x				

Table 7 displays the proportion of discriminating patterns per class with *minSupp*=0.5 and the NDVI configuration. Indeed, even if a pattern was extracted from one class, it is not enough to consider it as discriminating (*i.e.*, the same pattern can appear in different classes) Thus, queries was formulated to search which patterns appear in one class and not in the others. Two conclusions can be drawn from this figure. First, considering the set of classes *B*, most of the extracted patterns are discriminating (even if the *FP* class obtains a worse score). Second, finding discriminating patterns on the set of classes *is* more difficult.

Table 7. Proportion of discriminating patterns per class with *minSupp*=0.5 and the NDVI configuration

Level	Class	#disc.patterns	#patterns	Proportion
B	FP	6	9	66.67%
	NFP	12	12	100.00%
	OTHER	13	16	81.25%
C	Cr	3	10	30.00%
	NCr	4	11	36.36%

Table 8 displays some representative discriminating attribute values according the two configurations and the two sets of classes. An attribute value is said to be discriminating if it does not appear in any pattern of the other classes. This experiment aims at observing the impact of texture dynamic values on the extracted patterns.

Some conclusions can be drawn. First of all, the class *OTHER* does not contain discriminating value independently to the configuration. Second, a very interesting and promising result is that the default configuration contains much more discriminating values than the NDVI configuration. Moreover, these discriminating values concern the texture attributes. This result reinforces our idea that texture attributes are very useful in automatic landscape recognition.

To conclude this experiment study, we have empirically shown that:

1. The dimensionality reduction method allows reducing the search space by deleting useless attributes.
2. Most of the extracted patterns are discriminating.
3. It appears to be more difficult to distinguish between *Cr* and *NCr* classes than *FP*, *NFP* and *OTHER* classes with our approach.
4. And most of the discriminating attribute values concern the texture attributes.

Table 8. Some discriminating dimension values per class with $minSupp=0.3$ (top: default config. / bottom: NDVI config.)

Level	Class	Attribute	Value
B	FP	modis homogeneity 1km	0.48-0.52
		modis variance 1km	3.27-4.34
		modis dissimilarity 1km	1.36-1.51
	NFP	distance village	3150-6205
		modis contrast 1km	5.35-6.54
	OTHER	NONE	1.51-1.66
C	Cr	modis dissimilarity 1km	1.21-1.36
	NCr	modis variance 1km	3.27-4.34
		modis variance 1km	10.28-14.15
	Class	Attribute	Value
B	FP	NONE	
	NFP	rainfall	800
		distance village	3149.3-
	OTHER	NONE	6205.6
C	Cr	NONE	
	NCr	distance village	3149.3-
			6205.6

7 Related Work

Applications of sequential pattern mining methods to Satellite Image Time Series (SITS) include Julea et al. (2006, 2008, 2011) and Petitjean et al. (2010). Interest in these methods to study change detection on satellite images come from the fact that they are (i) multi-temporal, (ii) robust to noise, (iii) able to handle large volumes of data, and (iv) capable of capturing local evolutions without the need for prior clustering.

In Julea et al. (2008), sequential pattern mining is applied to study change in land cover over a 10 months period on a rural area of east Romania. Pattern extraction is used to group together SPOT pixels that share the same spectral evolution over time. The SITS data is thus processed at the pixel level, by taking the values of the pixels on each of the SPOT bands. A method is proposed to visualize the extracted patterns on a single image. Petitjean et al. (2010) presents a similar approach but pixel values are computed from four SPOT bands instead of a single band. The SITS period coverage is also much longer: a 20-year time image series is mined in order to study urban growth in the south west of France. A visualization technique is proposed to locate areas of evolution. Results show that mining all pixels of the images leads to the generation of a huge number of non-evolution patterns. Additional strategies are then required to filter out all non informative patterns.

To the best of our knowledge, sequential pattern mining has only been applied at the pixel level on high resolution images without taking into account external data or texture information in the mining process. In this paper, we have shown that sequential pattern mining can help to characterize cultivated areas from moderate resolution remote sensing images MODIS.

8 Conclusions

The objective of this study was to propose an original method to extract sets of relevant sequential patterns from MODIS time series that can be used for cultivated area mapping.

We have developed a data mining method based on two steps and applied it in Mali. The algorithm we used was selected on the basis of its efficiency to spatio-temporal data and its scalability.

Experiment study conducted on this data set reinforces our intuition about the importance of texture attributes to improve the automatic landscape recognition.

Our future work will be aimed at validating the extracted patterns per class. After which, we can go a step further to build the classifier based on these patterns and evaluate the predictions of the cultivated area at national scale. It would be interesting to compare such a classifier with classic prediction approaches in order to evaluate the interest of datamining methods in the remote sensing domain

References

- Agrawal, R., Srikant, R. (1995) Mining sequential patterns, in: Proceedings of the Eleventh International Conference on Data Engineering, pp. 3-14.
- Catlett, J. (1991) On changing continuous attributes into ordered discrete attributes, in Machine Learning EWSL-91, Springer, pp. 164-178.
- Haralick, R. (1979) Statistical and structural approaches to texture image type analysis, in: Proceedings of IEEE, 67, pp. 786-804.
- Haralick, R., Shanmugam, K., Dinstein, I. (1973) Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, 3(6), pp. 610-621.
- Julea, A., Méger, N., Bolon, P. (2008) On mining pixel based evolution classes in satellite image time series, in: Proceedings of the 5th Conf. on Image Information Mining: pursuing automation of geospatial intelligence for environment and security (ESA-EUSC 2008), 6 pgs.
- Julea, A., Méger, N., Bolon P., Rigotti, C., Doin, M.P., Lasserre, C., Trouvé, E., Lazarescu, V. (2011) Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns, IEEE Transactions on Geoscience and Remote Sensing, 49(4), 14 pgs.
- Julea, A., Méger, N., Trouvé, E. (2006) Sequential patterns extraction in multi-temporal satellite images, in : Proceedings of the 17th European Conference on Machine Learning and 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2006) - Berlin (Germany), Berlin (Allemagne), September 2006, pp. 94-97.
- Petitjean, F., Gançarski, P., Maseglia, F., Forestier, G. (2010) Analysing satellite image time series by means of pattern mining, in: Proceedings of Intelligent Data Engineering and Automated Learning - IDEAL 2010, 11th International Conference, Paisley (UK), September 1-3, 2010, Springer, Lecture Notes in Computer Science, 6283, pp. 45-52.
- Plantevit, M., Laurent, A., Laurent, D., Teisseire, M., Chong, Y.W. (2010) Mining Multi-Dimensional and Multi-Level Sequential Patterns, in : ACM Transactions on Knowledge Discovery from Data (TKDD), January 2010, 1), 37 pgs.
- Rouse, I. (1974) The explanation of culture change, Science, 185, pp. 343-344.

ARTICLE 5 :

**Recherche de motifs et cartographie des surfaces agricoles:
des relevés terrain aux données satellitaires
(application au Mali)**

Revue Internationale de Géomatique. In press.

Fouille de données et cartographie des surfaces agricoles

Des relevés terrain aux données satellitaires : application au Mali

Elodie Vintrou * — **Yoann Pitarch** ** — **Agnès Bégué** * — **Maguelonne Teisseire** ***

* CIRAD, UMR TETIS F-34093 Montpellier, France

** LIRMM-CNRS-UM2 F-34095 Montpellier, France

*** Cemagref, UMR TETIS F-34093 Montpellier, France

{nom}@teledetection.fr; pitarch@lirmm.fr

RÉSUMÉ. Dans cet article, nous présentons une approche préliminaire de caractérisation des paysages ruraux et de leurs systèmes de culture à partir de techniques de fouille de données (recherche d'itemsets fréquents). Cette méthode permet de coupler des données de relevé terrain aux indicateurs de texture extraits des images satellites. Sa mise en œuvre sur des données associées au Mali pose les premières bases d'une méthode originale d'extraction de motifs séquentiels à partir de données complexes.

ABSTRACT. Remote sensing mining

MOTS-CLÉS : Fouille de données, itemsets fréquents, image satellite, occupation du sol

KEYWORDS: Data Mining, Frequent Itemsets, Satellite Image, Land Cover

1. Introduction

Motivés par des problèmes d'Aide à la Décision, les chercheurs de différentes communautés (Intelligence Artificielle, Statistiques, Bases de Données ...) se sont intéressés à la conception et au développement d'une nouvelle génération d'outils permettant d'extraire automatiquement de la connaissance de grandes bases de données. Ces outils, techniques et approches sont le sujet d'un thème de recherche connu sous le nom de *Knowledge Discovery in Databases* ou KDD (Extraction de Connaissances dans les Bases de Données) ou Data Mining (Fouille de Données). Elles sont utilisées dans de nombreux domaines d'applications. Les exemples les plus courants sont les compagnies d'assurance, les compagnies bancaires (crédit, prédiction du marché, détection de fraudes), le marketing (comportement des consommateurs, mailing personnalisé), la recherche médicale (aide au diagnostic, au traitement, surveillance de population sensible), les réseaux de communication (détection de situations alarmantes, prédiction d'incidents), l'analyse de données spatiales, etc.

La fouille de données peut être définie par «*Processus non trivial permettant l'extraction automatique de connaissances d'une base de données pour obtenir de nouvelles données, valides, potentiellement utiles et compréhensibles*» (Fayyad *et al.*, 1996). Bien que le terme de fouille de données représente la découverte de connaissances, il ne constitue en fait qu'une seule des étapes du KDD, qui comprend globalement trois étapes : la préparation des données, l'extraction des données (Data Mining) et leur interprétation.

La première étape consiste à sélectionner uniquement les données potentiellement utiles de la base (opération de filtrage), sur lesquelles on effectue une phase de pré-traitement (gestion des données manquantes ou invalides). Ensuite, les données obtenues passent par une phase de formatage, afin de les préparer au processus de Data Mining. Finalement, la dernière étape est une étape d'analyse et d'interprétation de la connaissance extraite par la fouille de données, pour la rendre lisible et compréhensible par l'utilisateur. Les besoins variés nécessitent des approches différentes telles que la classification, la recherche de corrélations, la segmentation ou encore la détection de déviation.

Les travaux couplant la télédétection et la fouille de données sont récents et correspondent essentiellement à des approches utilisant les arbres de décision (dos Santos Silva *et al.*, 2005; Aksoy *et al.*, 2004) ou la détection de changement au sein de séries temporelles (Romani *et al.*, 2010; dos Santos Silva *et al.*, 2005). Il existe également des propositions de recherche de motifs comme outil d'exploration des données satellitaires, nous pouvons citer (Julea *et al.*, 2011; Petitjean *et al.*, 2010) qui extraient des patrons de comportement ou identifient des profils de changement.

Notre objectif est d'utiliser des techniques de fouille de données permettant de coupler des données hétérogènes pour faire une cartographie de l'occupation des sols (culture / non culture) à partir de relevés terrains, de données environnementales et

d'informations issues d'images satellites. Ce projet est donc une alternative pertinente pour définir un mécanisme d'apprentissage basé sur des données multi-sources (données spectrales, texturales et temporelles), des données environnementales (climat, relief, type de sol, ...) et des données de terrain, et mettre en évidence des critères d'implication qui n'auraient pas pu être identifiés autrement. Pour cela, nous proposons d'adapter un mécanisme de recherche de motifs séquentiels multidimensionnels, comme proposé dans (Plantevit *et al.*, 2010), à la fouille de séries d'images satellitaires et des données pouvant être mises en corrélation (informations externes).

Un des enjeux de l'application de ces techniques en télédétection réside dans le fait que les données relatives à une même réalité terrain sont par nature multi-sources, puisqu'elles résultent du croisement de données issues de différents capteurs, de relevés terrain et bases de données externes. Leur description se fait par ailleurs dans plusieurs dimensions et combine des informations spectrales, spatiales et temporelles. Pour cette raison, nous nous tournons vers des algorithmes d'extraction d'itemsets multidimensionnels (Pinto *et al.*, 2001), qui peuvent prendre en compte le caractère multidimensionnel des données.

Le jeu de données utilisé dans cette étude a été acquis dans la région soudano-sahélienne du Mali, qui présente l'intérêt d'être une zone particulièrement complexe pour laquelle la cartographie de l'occupation du sol par télédétection donne des résultats mitigés (Hansen *et al.*, 2000; Fritz *et al.*, 2010). Les principales raisons évoquées pour expliquer cette difficulté est que la zone soudano-sahélienne est une zone de transition eco-climatique présentant une grande variabilité spatiale des systèmes agricoles et des calendriers culturels. De plus, le parcellaire agricole y est généralement petit, l'hétérogénéité inter et intra-parcellaire est grande et les couverts végétaux, naturels et cultivés, ont le même cycle de croissance en raison du régime des pluies (Vintrou *et al.*, 2011). Nous proposons de tester deux jeux de données d'images satellitaires : un jeu de données monodate à haute résolution spatiale (2.5 m) et un jeu de données mensuelles à faible résolution spatiale. Ces deux jeux de données sont représentatifs des jeux de données actuellement disponibles en Observation de la Terre.

Les résultats présentés dans cet article mettent en œuvre un jeu de données restreint d'imagerie satellitaire et de données exogènes. Cette approche préliminaire, centrée sur l'apprentissage du système, nous a semblée indispensable pour assurer des bases communes aux deux communautés engagées dans ce travail, les agronomes et les spécialistes de fouille de données, et ainsi initier des collaborations de recherche plus pointues.

Dans les sections suivantes, nous décrivons tout d'abord les données étudiées (section 2) pour ensuite présenter les définitions associées aux motifs multidimensionnels (section 3). Puis, nous décrivons en détail le processus d'extraction mis en œuvre (section 4) ainsi que l'analyse des premiers résultats obtenus (section 5).

Site d'étude (éco-climat)	Précipitations moyennes annuelles	Culture principale	Végétation naturelle majoritaire	surface cultivée	Date SPOT
Cinzana (soudano-sahélien)	600 mm	mil et sorgho	végétation dégradée et sol nu	43%	14 nov. 2007
Koutiala (soudano-sahélien)	750 mm	mil, sorgho et coton	végétation ouverte et fermée	52%	20 nov. 2007
Sikasso (soudanien)	1 000 mm	maïs, coton et fruits	végétation dense	40%	20 nov. 2007

Tableau 1. Principales caractéristiques des trois zones d'étude et dates d'acquisition des images SPOT

2. Description des données

2.1. Contexte et zones d'étude

Le Mali est un pays d'Afrique de l'Ouest, autour de la latitude 14°N. Ce pays possède un gradient climatique Nord-Sud, qui varie de régions subtropicales à semi-arides, et s'étend plus au Nord vers des zones arides et désertiques.

Le Mali peut être considéré comme représentatif de la zone soudano-sahélienne, où la forte dépendance à l'agriculture pluviale entraîne une vulnérabilité aux changements climatiques et anthropiques. Par conséquent, afin de mieux connaître le phénomène de mousson en Afrique de l'Ouest et sa variabilité, et d'améliorer les prévisions des impacts de cette variabilité sur l'agriculture et la sécurité alimentaire, une des premières étapes nécessaires est une estimation fiable du domaine cultivé. Une attention particulière a été portée sur 3 zones, le long du gradient climatique malien (tableau 1).

2.2. Les données terrain

Des missions de terrain ont été effectuées au Mali de mai à novembre 2009, dans le but de caractériser les paysages agricoles soudano-sahéliens. Au total, 744 points GPS ont été enregistrés, et des paysans de chacune des régions étudiées ont été interrogés. Les relevés ont été effectués par transects, du centre du village vers l'extérieur. Chaque point relevé a été transformé en un polygone dont il est le centre et auquel a été affecté un type d'occupation du sol ("mil", "sorgho", "maïs", "coton", "végétation naturelle" principalement), comme indiqué figure 8.

2.5. Les données dérivées des images

À partir de ces images sont calculés un certain nombre de descripteurs spectraux (indice de végétation) et spatiaux (indices de texture) des états de surface (occupation du sol) :

– *L'indice de végétation NDVI*. Les indices de végétation sont des combinaisons, linéaires ou non, de réflectances dans les bandes spectrales R (rouge), PIR (proche infrarouge) et MIR (moyen infrarouge). Ils permettent de caractériser le couvert végétal en terme de vigueur de la végétation. Parmi ces indices, le « *Normalized Difference Vegetation Index* » (Rouse, 1974) est de loin le plus utilisé. Il permet de caractériser l'activité photosynthétique de la surface, et donc de discriminer facilement des surfaces végétalisées (0,9 pour de la végétation verte et dense) des surface de sols nus (environ 0,1 pour sol nu). Ces propriétés font que le NDVI est fréquemment utilisé comme une mesure indirecte de la biomasse. La formule du NDVI est la suivante :

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

– *Les indices de texture*. La variabilité spatiale d'une image est représentée par le concept de texture. Haralick élargit dans (Haralick, 1979) la définition en décrivant une texture comme un phénomène à deux dimensions : la première concernant la description d'éléments de base ou primitives (le motif) à partir desquels est formée la texture ; la deuxième dimension est relative à la description de l'organisation spatiale de ces primitives.

La matrice de co-occurrences (ou matrice de dépendance spatiale) est une des approches les plus connues et les plus utilisées pour extraire des caractéristiques de textures. Elle effectue une analyse statistique de second ordre de la texture, par l'étude des relations spatiales des couples de pixels (Haralick *et al.*, 1973; Haralick, 1979). Quatorze indices (définis par Haralick) qui correspondent à des caractères descriptifs des textures peuvent être calculés à partir de cette matrice.

Dans cette étude, chaque indice d'homogénéité, de variance, de dissimilarité et de contraste a été calculé sur une fenêtre de de 15×15 pixels.

3. Extraction d'itemsets séquentiels multidimensionnels

Le problème de la recherche de motifs séquentiels a été introduit par R. Agrawal dans (Agrawal *et al.*, 1995) et appliqué avec succès dans de nombreux domaines comme la biologie (Wang *et al.*, 2004; Salle *et al.*, 2009), la fouille d'usage du Web (Pei *et al.*, 2000; Masseglia *et al.*, 2008), la détection d'anomalie (Rabatel *et al.*, 2010), la fouille de flux de données (Marascu *et al.*, 2006) ou la description des comportements au sein d'un groupe (Perera *et al.*, 2009).



Figure 2. Carte du Mali sur fond de série temporelle MODIS 2007 et emprise géographique des trois images SPOT de validation (Cinzana, Koutiala, Sikasso).

Des approches plus récentes (Julea. *et al.*, 2008) utilisent les motifs séquentiels pour décrire les évolutions temporelles des pixels au sein des séries d'images satellites. Néanmoins, à notre connaissance, l'étude de la littérature ne fait état d'aucuns travaux sur l'application de techniques de recherche de motifs séquentiels couplant données externes et télédétection.

Dans cette section, nous introduisons les définitions relatives à la fouille d'itemsets séquentiels multidimensionnels et décrivons l'algorithme de fouille de données utilisé.

3.1. Itemsets et séquences multidimensionnels

L'approche M^3SP (Plantevit *et al.*, 2010) permet l'extraction de motifs séquentiels multidimensionnels. Cette méthode est adaptée à notre contexte car elle permet d'analyser à la fois la dimension temporelle des séries d'images MODIS ainsi que les informations terrains associées. Les concepts proposés sont définis dans cette section.

Soit un ensemble \mathcal{D} de dimensions et $\{\mathcal{D}_R, \mathcal{D}_A, \mathcal{D}_T, \mathcal{D}_I\}$ une partition de \mathcal{D} dans laquelle \mathcal{D}_R désigne les dimensions de référence, qui permettent de déterminer si une séquence est fréquente, \mathcal{D}_A les dimensions d'analyse, sur lesquelles les corrélations sont extraites, et \mathcal{D}_T les dimensions permettant d'introduire une relation d'ordre (généralement le temps). Les dimensions \mathcal{D}_I sont les dimensions ignorées lors de la fouille.

Pour chaque dimension $D_i \in \mathcal{D}$, on note $Dom(D_i)$ son domaine de valeurs. À chaque domaine de valeurs $Dom(D_i)$ est associé une hiérarchie H_i , et l'on suppose

que $Dom(D_i)$ contient une valeur particulière notée \top_i (la racine de la hiérarchie). Lorsqu'aucune hiérarchie de valeurs n'est définie sur une dimension D_i , nous considérons H_i comme un arbre de profondeur 1 dont la racine est \top_i et dont les feuilles sont les éléments de $Dom(D_i) \setminus \{\top_i\}$. La figure 3 présente un exemple de hiérarchie de valeurs H_i pour la dimension site-étude et la figure 8 celle de la dimension type-utilisation-sol.

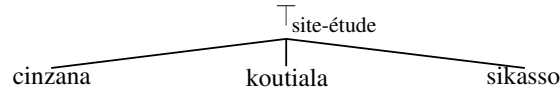


Figure 3. La hiérarchie de valeurs H_i pour la dimension site-étude.

Un *item multidimensionnel* $e = (d_1, d_2, \dots, d_m)$ est un m -uplet défini sur les dimensions d'analyse \mathcal{D}_A , c'est-à-dire tel que $\forall i \in [1 \dots m], d_i \in Dom(D_i)$ avec $D_i \in \mathcal{D}_A$ et $\exists d_i \in [1, \dots, m]$ tel que $d_i \neq \top_i$. Par exemple, $e = (sorgho, cinzana)$ et $e' = (sorgho, \top_{site-étude})$ sont des items multidimensionnels qui décrivent des points terrain sur les dimensions d'analyse $\mathcal{D}_A = \{type - utilisation - sol, site - étude\}$. On définit une relation d'inclusion \subseteq entre items multidimensionnels :

un item multidimensionnel $e = (d_1, d_2, \dots, d_m)$ est inclus dans un item multidimensionnel $e' = (d'_1, d'_2, \dots, d'_m)$ (noté $e \subseteq e'$) si $\forall i \in [1, \dots, m], d_i = d'_i$ ou est une spécialisation de d'_i dans H_i .

Dans l'exemple précédent, on a l'inclusion $(sorgho, cinzana) \subseteq (sorgho, \top_{site-étude})$ car *cinzana* est une spécialisation de $\top_{site-étude}$ dans la hiérarchie $H_{site-étude}$.

Un *itemset multidimensionnel* $i = (e_1, e_2, \dots, e_m)$ est un ensemble non vide d'items multidimensionnels non deux à deux comparables par rapport à \subseteq (c.-à-d., $\forall i, j \in [1, \dots, m], e_i \not\subseteq e_j$ et $e_i \not\supseteq e_j$). On définit une relation d'inclusion \subseteq entre itemsets multidimensionnels : un itemset i est inclus dans un itemset i' (noté $i \subseteq i'$) si pour chaque item a de i , il existe un item a' de i' tel que $a \subseteq a'$.

Une *séquence multidimensionnelle* $s = \langle i_1, \dots, i_n \rangle$ est une liste ordonnée non vide d'itemsets multidimensionnels.

Une relation de généralisation (ou spécialisation) entre séquences multidimensionnelles est définie. Une séquence $s = \langle i_1, i_2, \dots, i_n \rangle$ est plus spécifique qu'une séquence $s' = \langle i'_1, i'_2, \dots, i'_m \rangle$ s'il existe des entiers $1 \leq j_1 \leq \dots \leq j_m \leq n$ tels que $s_{j_1} \subseteq s'_1, s_{j_2} \subseteq s'_2, \dots, s_{j_m} \subseteq s'_m$.

Étant donnée une table relationnelle DB , on appelle *bloc* l'ensemble des n -uplets qui ont la même projection sur \mathcal{D}_R . Par exemple, le tableau 3 donne le bloc formé en ne gardant que les n -uplets de la table relationnelle DB donnée au tableau 2 dont la projection sur $\mathcal{D}_R = \{type - utilisation - sol, site - étude\}$ est $(mil, koutiala)$. Le *support* d'une séquence est le nombre de blocs qui contiennent cette séquence.

pt-id	date	type-utilisation-sol	site-étude	SPOT-NDVI-100
1	1	arachide	sikasso	très faible
2	1	mil	koutiala	faible
2	2	mil	koutiala	modéré
3	1	sorgho	koutiala	élevé

Tableau 2. Base de données DB.

pt-id	date	type-utilisation-sol	site-étude	SPOT-NDVI-100
2	1	mil	koutiala	faible
2	2	mil	koutiala	modéré

Tableau 3. Bloc $B_{(mil,koutiala)}$.

Étant donné un seuil σ_{min} de support minimum, le but de la recherche d'itemsets séquentiels multidimensionnels est de trouver toutes les séquences dont le support est supérieur ou égal à σ_{min} .

4. Mise en œuvre

Un processus d'extraction de connaissances a été mis au point (figure 4). Il est constitué de quatre étapes : (1) chargement des données, (2) préparation des données, (3) fouille de données et (4) interprétation et validation des résultats.

Dans l'étape de chargement des données, les données sont collectées et placées dans une base de données (données "brutes"). Cinq indices images sont calculés à partir des images SPOT et MODIS : l'indice de végétation NDVI et les indices de texture variance, homogénéité, contraste et dissimilarité. À chaque point terrain est associée une valeur de chacun de ces indices, qui est calculée en prenant la moyenne de la valeur de l'indice sur les pixels contenus dans un polygone carré centré sur ce point (figure 5). Deux jeux d'indices sont ainsi calculés avec des tailles de polygones de 100 m et 200 m de côté. Concernant le jeu de données MODIS, la même démarche est suivie avec des polygones d'un kilomètre de côté.

L'étape de préparation des données a pour but de constituer l'ensemble d'apprentissage et de formater les données pour pouvoir être traitées par l'algorithme de fouille.

L'ensemble d'apprentissage est constitué en sélectionnant parmi l'ensemble des points terrain les points qui se situent sur une des images SPOT et qui correspondent

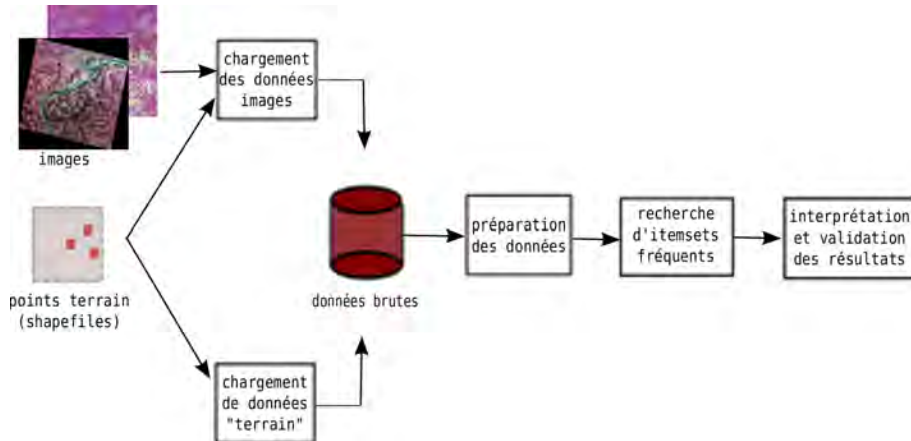


Figure 4. Les différentes étapes du processus d'extraction de connaissances.

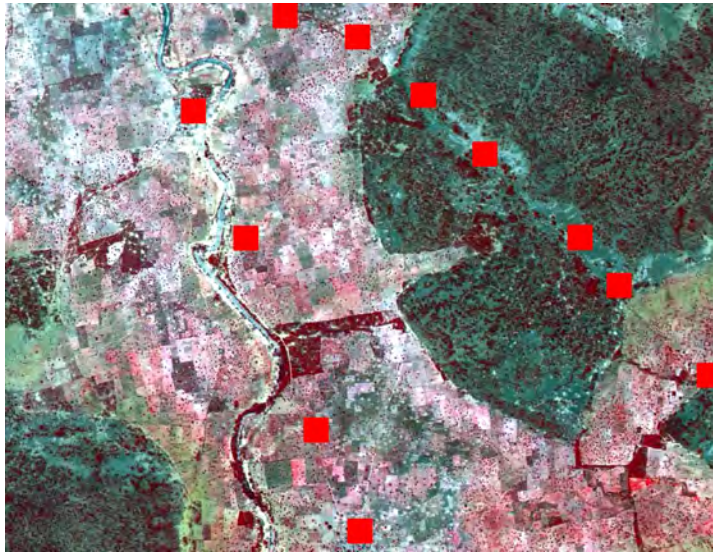


Figure 5. Extraction des indices images pour chaque point terrain sur un polygone carré centré sur ce point (zoom sur une image SPOT du site de Cinzana, les carrés sont des polygones de 100 m×100 m centrés sur chacun des points terrain). © CNES 2009, Distribution SPOT Image

à des relevés en zone de culture. On obtient 498 points dont la répartition par site d'étude et en culture / non culture est donnée par le tableau [4](#).

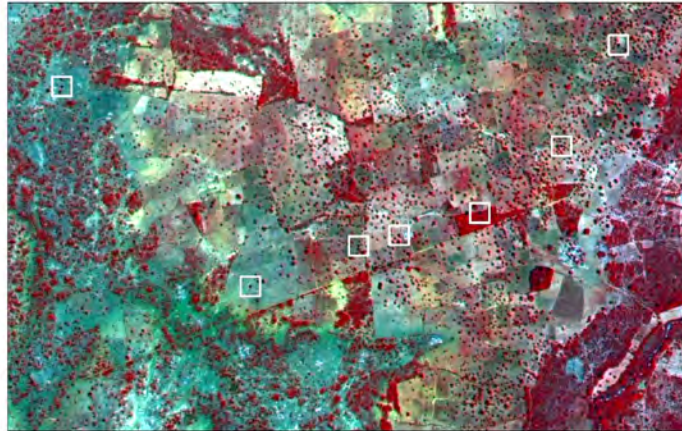


Figure 6. Extraction des indices images pour chaque point terrain sur un polygone carré centré sur ce point (zoom sur une image SPOT du site de Koutiala, les carrés sont des polygones de $100\text{m} \times 100\text{m}$ centrés sur chacun des points terrain). © CNES 2009, Distribution SPOT Image

Site d'étude	Culture	Non culture	Total
Cinzana	138	85	223
Koutiala	105	78	183
Sikasso	46	46	92

Tableau 4. Les points terrain de l'ensemble d'apprentissage.

Chaque point de l'ensemble d'apprentissage est décrit par les valeurs qu'il prend dans un ensemble de dimensions D_i :

- id-pt : entier qui identifie chaque point terrain de manière unique
- date : estampille temporelle (constante pour les images SPOT et de 1 à 12 pour les images MODIS)
- site-étude : le nom du site d'étude
- type-occupation-sol : "culture" ou "non culture"
- nom-village : le nom du village le plus proche, pris parmi les relevés village
- distance-village : la distance du point au village le plus proche
- SPOT-NDVI-100, SPOT-variance-100, SPOT-homogénéité-100, SPOT-dissimilarité-100 et SPOT-contraste-100 : les valeurs des descripteurs images SPOT calculés avec des polygones carrés de 100 m de côté

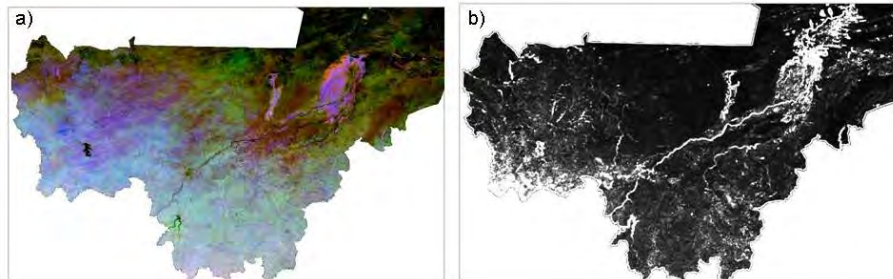


Figure 7. Série temporelle 2007 a) MODIS NDVI, b) MODIS TEXTURE variance

– SPOT-NDVI-200, SPOT-variance-200, SPOT-homogénéité-200, SPOT-dissimilarité-200 et SPOT-contraste-200 : les valeurs des descripteurs images SPOT calculés avec des polygones carrés de 200 m de côté

– la série temporelle (12 valeurs) de MODIS-NDVI, MODIS-variance, MODIS-homogénéité, MODIS-dissimilarité et MODIS-contraste : les valeurs des descripteurs images MODIS calculés avec des polygones carrés d'un kilomètre de côté pour chacun des mois concerné.

Les domaines $Dom(D_i)$ de chaque dimension D_i sont ensuite discrétisés en découpant leurs intervalles de valeurs. Pour l'indice de végétation NDVI, le découpage de l'intervalle de valeurs $[-1, 1]$ est donné par l'expert. Les domaines de valeurs des indices de texture sont découpés en cinq classes de même effectif. Le tableau 5 résume l'ensemble des dimensions D_i utilisées pour décrire les points de l'ensemble d'apprentissage et le découpage de leurs domaines de valeurs.

Une hiérarchie de valeur H_i est construite pour chaque dimension D_i . Les hiérarchies considérées sont toutes de profondeur 1 sauf pour l'attribut type-utilisation-sol (figure 8).

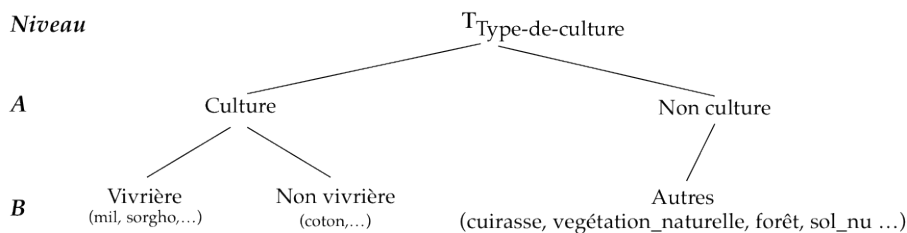


Figure 8. La hiérarchie de valeurs H_i pour la dimension type-utilisation-sol.

dimension D_i	intervalles de valeurs				
id-pt	{1}, {2}, ..., {498}				
date	{1}, {2}, ..., {12}				
site-étude	{cinzana, koutiala, sikasso}				
type-utilisation-sol	{riz, sorgho, maïs, ...}				
nom-village	{dioforongo, tigui, sanando, ...}				
distance-village	proche	éloigne			
	[0,3000]	[3001,+∞[
SPOT-NDVI-100	très faible	faible	modéré	élevé	
	[-1,0.2]	[0.2,0.3]	[0.3,0.5]	[0.5,1]	
SPOT-NDVI-200	très faible	faible	modéré	élevé	
	[-1,0.2]	[0.2,0.3]	[0.3,0.5]	[0.5,1]	
SPOT-variance-100	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.56]	[0.56, 1.22]	[1.22, 2.38]	[2.38, 4.00]	[4.00, +∞[
SPOT-variance-200	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.67]	[0.67, 0.74]	[0.74, 0.80]	[0.80, 0.87]	[0.87, +∞[
SPOT-homogénéité-100	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.26]	[0.26, 0.40]	[0.40, 0.54]	[0.54, 0.72]	[0.72, +∞[
SPOT-homogénéité-200	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[
SPOT-dissimilarité-100	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.26]	[0.26, 0.40]	[0.40, 0.54]	[0.54, 0.72]	[0.72, +∞[
SPOT-dissimilarité-200	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[
SPOT-contraste-100	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[
SPOT-contraste-200	très faible	faible	modéré	élevé	très élevé
] - ∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[
MODIS-NDVI	[-∞, 2288]]2288, 2725]]2725, 3497]]3497, 4692]]4692, +∞[
Modis_contrast_1km] - ∞, 2.35]]2.35, 3.73]]3.73, 5.35]]5.35, 9.11]]9.11, +∞[
Modis_variance_1km] - ∞, 3.26]]3.26, 5.44]]5.44, 8.15]]8.15, 14.15]]14.15, +∞[
Modis_dissimilarite_1km] - ∞, 1.08]]1.08, 1.36]]1.36, 1.66]]1.66, 2.11]]2.11, +∞[
Modis_homogeneite_1km] - ∞, 0.37]]0.37, 0.45]]0.45, 0.51]]0.51, 0.57]]0.57, +∞[

Tableau 5. Les dimensions D_i et le découpage de leurs domaines de valeurs $Dom(D_i)$ en intervalles.

Dans l'étape de fouille, l'algorithme de recherche d'itemsets séquentiels fréquents M^3SP est appliqué sur les données d'apprentissage formatées, en choisissant un seuil de support σ_{min} , un ensemble de dimensions de référence \mathcal{D}_R et un ensemble de dimensions d'analyse \mathcal{D}_A .

Les itemsets séquentiels extraits sont présentés à l'analyste pour interprétation et validation.

5. Résultats et discussion

Dans cette section, nous analysons principalement les résultats obtenus à partir des images SPOT. Nous examinons ensuite les séquences obtenues sur les séries temporelles MODIS dans notre objectif de cartographie "culture" et "non culture", selon les valeurs associées à la dimension type-de-culture représentées figure 8.

5.1. Itemsets obtenus à partir des indicateurs sur SPOT

Toutes les expériences ont été effectuées en prenant $\mathcal{D}_T = \{1\}$ et $\mathcal{D}_R = \{id - pt\}$. Comme la même estampille temporelle est associée à chaque point, les séquences obtenues ne sont constituées que d'un seul itemset. De plus, la dimension de référence étant l'identifiant des points, l'ensemble d'apprentissage est divisé en autant de blocs

qu'il y a de points d'apprentissage. Par conséquent, dans ces expériences chaque itemset résultat est composé d'un seul item et son support correspond au nombre de points terrain qui le partagent.

Plusieurs expériences ont été menées en faisant varier les dimensions d'analyse \mathcal{D}_A et en ne fouillant que les points situés sur un site d'étude donné (par filtrage de l'ensemble d'apprentissage avant la fouille).

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_1 = \langle \{(culture, \top_{distance-village})\} \rangle$	138 (62%)
	$s_2 = \langle \{(culture, proche)\} \rangle$	121 (54%)
Koutiala (183 points)	$s_3 = \langle \{(culture, \top_{distance-village})\} \rangle$	105 (57%)
	$s_4 = \langle \{(culture, proche)\} \rangle$	80 (44%)
Sikasso (92 points)	$s_5 = \langle \{(culture, \top_{distance-village})\} \rangle$	46 (50%)
	$s_6 = \langle \{(culture, proche)\} \rangle$	27 (29%)

Tableau 6. Itemsets extraits pour $\mathcal{D}_A = \{ type-occupation-sol, distance-village \}$.

Les itemsets présentés dans le tableau 6 ont été extraits en prenant comme dimensions d'analyse le type de culture et la distance au village le plus proche. Les résultats montrent que dans les 3 sites étudiés, les cultures sont généralement cultivées autour des villages, dans un rayon de 2 à 3 km pour la majorité. En effet, 88% des cultures de Cinzana (121/138, itemsets s_1 et s_2), 77% de celle de Koutiala (80/105, itemsets s_3 et s_4) et 59% de celles de Sikasso (27/46, itemsets s_5 et s_6) sont dans la couronne de 3 km autour des différents villages. Il n'apparaît pas possible de faire un lien entre la distance au centre du village et le type de culture au vu du nombre trop faible de points terrain disponibles par site. Il a cependant été déjà observé dans plusieurs villages d'Afrique de l'Ouest un aménagement en auréoles. Le village et les jardins occupent une position centrale. Une première auréole (soforo) est constituée par les champs « de case » cultivés en rotation annuelle. Une seconde auréole (kongo foro) est formée par les champs de brousse (mil, sorgho, arachides, coton...). Enfin, la brousse (kongo) fournit les produits de la chasse, de la cueillette, le bois d'œuvre et de feu. La distance entre ces 3 auréoles varie entre les villages.

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_7 = \langle \{(culture, très faible)\} \rangle$	74 (33%)
	$s_8 = \langle \{(culture, modéré)\} \rangle$	56 (31%)
Koutiala (183 points)	$s_9 = \langle \{(culture, faible)\} \rangle$	33 (18%)
	$s_{10} = \langle \{(culture, faible)\} \rangle$	25 (28%)
Sikasso (92 points)	$s_{11} = \langle \{(culture, modr)\} \rangle$	20 (22%)

Tableau 7. Itemsets extraits pour $\mathcal{D}_A = \{ type-utilisation-sol, SPOT-NDVI-100 \}$.

Les itemsets présentés dans le tableau 7 ont été extraits en prenant comme dimensions d'analyse le type de culture et le descripteur NDVI calculé avec des polygones

carrés de 100 m de côté. Nous pouvons signaler que le cycle de production végétale est particulier au Mali : les cultures étant pour la majorité des cultures pluviales, la croissance des plantes est étroitement liée à la pluviométrie (quantité et répartition). Les supports des itemsets séquentiels sont plus fréquemment faibles concernant le NDVI à Cinzana, qu'à Koutiala, qu'à Sikasso. Ceci reflète bien le gradient bioclimatique au Mali. Il pleut moins au Nord qu'au Sud, et donc les plantes ont une activité photosynthétique inférieure à Cinzana qu'à Sikasso, en moyenne. D'autre part, pour le NDVI du mois de novembre, les cultures sont déjà entièrement récoltées à Cinzana, et partiellement récoltées à Koutiala et à Sikasso. Ceci explique les 54% de culture avec un NDVI « très faible » à Cinzana (74/138, itemsets s_1 et s_7), contre 98% et 94% de cultures avec un NDVI « faible » ou « medium » (itemsets s_3 , s_5 et s_8 à s_{11}) à Koutiala et Sikasso respectivement.

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_{12} = \langle \{(culture, \top_{variance-100}, \top_{homognit-100}, \top_{dissimilarit-100}, \acute{e}lev\acute{e})\} \rangle$	56 (25%)
Koutiala (183 points)	$s_{13} = \langle \{(culture, \top_{variance-100}, \top_{homognit-100}, \top_{dissimilarit-100}, faible)\} \rangle$	35 (19%)
Sikasso (92 points)	$s_{14} = \langle \{(culture, \top_{variance-100}, trslev, trsfaible, trsfaible)\} \rangle$	23 (25%)
	$s_{15} = \langle \{(culture, \top_{variance-100}, \top_{homognit-100}, \top_{dissimilarit-100}, faible)\} \rangle$	18 (20%)

Tableau 8. Itemsets extraits pour $\mathcal{D}_A = \{type-utilisation-sol, SPOT-variance-100, SPOT-homogénéité-100, SPOT-dissimilarité-100, SPOT-contraste-100\}$.

Les itemsets présentés dans le tableau 8 ont été extraits en prenant comme dimensions d'analyse le type de culture et les quatre descripteurs de texture calculés avec des polygones carrés de 100 m de côté. La présence de quatre indices de texture induit une difficulté à interpréter la présence des itemsets fréquents. Si l'on analyse seulement le contraste, on observe dans 20 à 25% des cas, un contraste « élevé » pour Cinzana (itemset s_{12}), « faible » pour Koutiala (itemset s_{13}) et « très faible » pour Sikasso. Ce contraste qui diminue du Nord au Sud peut s'expliquer par une différence de densité d'arbres dans les champs cultivés. Il serait en effet plus commun de trouver des arbres comme le Balanzan, le Néré ou le Karitier dans des champs de la région de Cinzana, qu'à Koutiala ou Sikasso, ce qui expliquerait les brusques changements de radiométrie, et donc un indice de contraste élevé.

Enfin, les différentes tailles de polygones n'induisent pas de changements notables dans les résultats.

À la suite de cette étude, il apparaîtrait intéressant de faire un lien entre la distance au centre du village et le type de culture. Il doit exister un lien entre les espèces cultivées et la distance au village, que nous essaierons de déterminer à partir de données terrain supplémentaires. D'autre part, le NDVI et la texture pourraient être mis

Niveau	Classe	Nb séquence discriminantes	Nb séquences total	Proportion
B	Vivrière	6	9	66.67%
	Non vivrière	12	12	100%
	Autres	13	16	81.25%
A	Culture	3	10	30%
	Non Culture	4	11	36.36%

Tableau 9. Proportion de séquences discriminantes par classe avec $min.Supp = 0.5$

en relation avec des images du mois de septembre ou d'octobre. C'est la période pendant laquelle les cultures sont dans des phases de croissance différentes suivant les régions, et entre elles également, puisque le maïs par exemple, est récolté, alors que les autres cultures céréalières restent sur pied pour encore un mois voire deux. Nous essaierons également d'utiliser pour l'extraction d'indices des polygones plus grands, pour un calcul de texture optimum (la faible taille des polygones ne permet pas de détecter beaucoup de motifs de texture répétés) et enfin, d'utiliser des séries temporelles d'images MODIS pour prendre en compte le cycle de croissance de chaque occupation du sol et ainsi mieux les différencier.

5.2. Séquences obtenues à partir des indicateurs sur MODIS

Toutes les expériences ont été effectuées en prenant $\mathcal{D}_T = \{date\}$ et $\mathcal{D}_R = \{id - pt\}$. Ayant différentes estampilles temporelles, nous obtenons des séquences d'itemsets. De façon identique à la section précédente, les supports correspondent au nombre de points terrain qui partagent la séquence obtenue.

Le tableau 9 indique la proportion de séquences extraites qui sont *discriminantes* c'est-à-dire qui apparaissent dans une classe mais qui n'apparaissent pas dans les autres classes considérées selon le niveau des valeurs dans la hiérarchie. Il est intéressant de constater que le nombre de motifs discriminants est le plus important pour les valeurs vivrière et non vivrière. De plus, il apparaît difficile d'identifier à partir des indicateurs MODIS une réelle distinction entre les valeurs culture et non culture.

Le tableau 10 décrit quelques séquences obtenues selon les valeurs culture et non culture, vivrière et non vivrière. Les séquences obtenues selon les valeurs culture et non culture donnent des informations sur l'évolution du NDVI au cours du temps. Nous avons montré à travers la fouille sur les images SPOT que le NDVI reflétait le gradient bioclimatique au Mali. De par l'étude MODIS, nous pouvons suivre la phénologie des cultures et de la végétation naturelle à travers les motifs. Pour pou-

Motifs séquentiels	Classe
<(distance_village=eloigne,modis_ndvi=2182.6-2927.0) (distance_village=-eloigne,modis_ndvi=2927.0-3671.5)>	Culture
<(modis_ndvi=2182.6-2927.0) (modis_ndvi=2182.6-2927.0) (modis_ndvi=2927.0-3671.5)>	Non culture
<(modis_variance=-inf-7.585,modis_contrast=-inf-6.245) (modis_variance=-inf-7.585,modis_contrast=-inf-6.245)>	Culture
<(site=cinzana,distance_village=proche, modis_contrast=-inf-6.245392, modis_homogeneite=0.550197-0.628796) >	Vivrière
< (site=koutiala,modis_dissimilarite=1.852252-2.321403, modis_contrast=6.245392-11.962048) >	Non_Vivrière

Tableau 10. Quelques séquences extraites à partir des séries temporelles d'indicateurs des images MODIS : $\mathcal{D}_A = \{ \text{distance_village}, \text{modis_ndvi}, \text{modis_variance}, \text{modis_homogénéité}, \text{modis_dissimilarité}, \text{modis_contrast} \}$.

voir différencier par site, nous envisageons l'utilisation d'une base de données plus importante.

Nous pouvons souligner les deux séquences obtenues selon les valeurs « vivrière » et « non vivrière » qui implique les dimensions texturales. Tout comme pour SPOT, la distance au village inférieure à 3km apparaît sur le site de Cinzana. Concernant la texture, les deux sites de Koutiala et Cinzana présentent des motifs faisant apparaître la dimension texturale. Le motif de Cinzana possède un contraste moins élevé pour une culture vivrière que celui de Koutiala pour une culture non vivrière. Le motif de Cinzana présente également un item d'homogénéité, alors que celui de Koutiala présente un item de dissimilarité. Nous pouvons donc en conclure que la texture est différente suivant le type de culture considéré, mais également suivant le site d'étude. Cette prise en compte de la texture est donc très importante à des fins de classification.

Il est intéressant de constater des valeurs discriminantes des dimensions texturales sur les sites étudiés. Par ailleurs, l'analyse préliminaire des motifs a montré qu'il n'y a aucune variation temporelle suffisante de texture. Ceci nous pose question, car la texture devrait, tout comme le NDVI, apparaître avec des valeurs différentes dans un motif. Nous émettons l'hypothèse que les variations sont minimales, et que les intervalles sortants sont donc toujours les mêmes. Ceci ne signifie pas qu'elle est inutile, puisque comme nous l'avons montré auparavant, les valeurs de textures sont différentes d'un site à un autre, ou d'une occupation du sol à une autre.

6. Conclusion

Nous avons présenté la première étape d'extraction d'itemsets et de séquences multidimensionnels d'une méthode de caractérisation des paysages ruraux et de leurs systèmes de culture. Nous avons mis en œuvre ce processus d'extraction sur des données du Mali et avons pu fouiller des données hétérogènes comme les relevés terrains avec des indicateurs obtenus des images satellitaires. Les premiers résultats sont prometteurs et laissent présager des perspectives d'analyse pertinentes. Il s'agit à présent de poursuivre ces travaux (1) afin de prendre en compte la partie séquentielle des données (séquences extraites des images MODIS) qui seront le support du mécanisme de classification des types de culture puis (2) de proposer une évaluation (rappel, confiance) afin de mesurer l'efficacité d'une telle approche.

7. Bibliographie

- Agrawal R., Srikant R., « Mining Sequential Patterns », in P. S. Yu, A. L. P. Chen (eds), *Proceedings of the Eleventh International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan*, IEEE Computer Society, p. 3-14, 1995.
- Aksoy S., Koperski K., Tusk C., Marchisio G., « Interactive training of advanced classifiers for mining remote sensing image archives », *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '04*, ACM, New York, NY, USA, p. 773–782, 2004.
- dos Santos Silva M. P., Câmara G., de Souza R. C. M., Valeriano D. M., Escada M. I. S., « Mining Patterns of Change in Remote Sensing Image Databases », *ICDM*, p. 362-369, 2005.
- Fayyad U. M., Piatetsky-Shapiro G., Smyth P., « From Data Mining to Knowledge Discovery : an Overview », *Advances in knowledge discovery and data mining*, vol. 1, p. 1–34, 1996.
- Fritz S., See L., Rembold F., « Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa », *International Journal of Remote Sensing*, vol. 31, n° 9, p. 2237 - 2256, 2010.
- Hansen M., Reed B., « A comparison of the IGBP DISCover and University of Maryland 1km global land cover products », *International Journal of Remote Sensing*, vol. 21, n° 6-7, p. 1365-1373, 2000.
- Haralick R. M., « Statistical and Structural Approaches to Texture », *Proceedings of the IEEE*, vol. 67, n° 5, p. 786-804, 1979.
- Haralick R. M., Shanmugam K., Dinstein I., « Textural Features for Image Classification », *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, n° 6, p. 610–621, November, 1973.
- Julea. A., Meger N., Bolon P., « On mining pixel based evolution classes in satellite image time series », *Proc. of the 5th Conf. on Image Information Mining : pursuing automation of geospatial intelligence for environment and security (ESA-EUSC 2008)*, p. 6, 2008.
- Julea A., Méger N., Bolon P., Rigotti C., Doin M.-P., Lasserre C., Trouve E., Lazarescu V., « Unsupervised Spatiotemporal Mining of Satellite Image Time Series using Grouped

- Frequent Sequential Patterns », *IEEE Transactions on Geoscience and Remote Sensing*, 2011. To appear, vol. 49, issue 4, 2011, 14 pages.
- Marascu A., Massegli F., « Mining sequential patterns from data streams : a centroid approach », *Journal of Intelligent Information Systems*, vol. 27, n° 3, p. 291-307, 2006.
- Massegli F., Poncelet P., Teisseire M., Marascu A., « Web usage mining : extracting unexpected periods from web logs », *Data Mining and Knowledge Discovery (DMKD)*, vol. 16, n° 1, p. 39-65, 2008.
- Pei J., Han J., Mortazavi-Asl B., Zhu H., « Mining Access Patterns Efficiently from Web Logs », in T. Terano, H. Liu, A. L. P. Chen (eds), *Knowledge Discovery and Data Mining, Current Issues and New Applications, 4th Pacific-Asia Conference, PADKK 2000, Kyoto, Japan, April 18-20, 2000, Proceedings*, Lecture Notes in Computer Science, Springer, p. 396-407, 2000.
- Perera D., Kay J., Koprinska I., Yacef K., Zaïane O. R., « Clustering and Sequential Pattern Mining of Online Collaborative Learning Data », *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, n° 6, p. 759-772, 2009.
- Petitjean F., Gançarski P., Massegli F., Forestier G., « Analysing Satellite Image Time Series by Means of Pattern Mining », *Intelligent Data Engineering and Automated Learning - IDEAL 2010, 11th International Conference, Paisley, UK, September 1-3, 2010. Proceedings*, vol. 6283 of *Lecture Notes in Computer Science*, Springer, p. 45-52, 2010.
- Pinto H., Han J., Pei J., Wang K., Chen Q., Dayal U., « Multi-dimensional sequential pattern mining », *CIKM '01 : Proceedings of the tenth international conference on Information and knowledge management*, ACM, New York, NY, USA, p. 81-88, 2001.
- Plantevit M., Laurent A., Laurent D., Teisseire M., Choong Y. W., « Mining multidimensional and multilevel sequential patterns », *ACM Transactions on Knowledge Discovery from Data TKDD*, 2010.
- Rabatel J., Bringay S., Poncelet P., « Aide à la décision pour la maintenance ferroviaire préventive », in S. B. Yahia, J.-M. Petit (eds), *Extraction et gestion des connaissances (EGC'2010), Actes, 26 au 29 janvier 2010, Hammamet, Tunisie*, Revue des Nouvelles Technologies de l'Information, Cépaduès-Éditions, p. 363-368, 2010.
- Romani L. A. S., de Ávila A. M. H., Jr. J. Z., Jr. C. T., Traina A. J. M., « Mining Relevant and Extreme Patterns on Climate Time Series with CLIPSMiner », *JIDM*, vol. 1, n° 2, p. 245-260, 2010.
- Rouse I., « The explanation of culture change », *Science*, vol. 185, p. 343-344, 1974.
- Salle P., Bringay S., Teisseire M., « Mining Discriminant Sequential Patterns for Aging Brain », in C. Combi, Y. Shahar, A. Abu-Hanna (eds), *Artificial Intelligence in Medicine, 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings*, Lecture Notes in Computer Science, p. 365-369, 2009.
- Vintrou E., Desbrosse A., Bégué A., Traoré S., Baron C., LoSeen D., « Can MODIS landscape stratification improve the accuracy of cultivated area mapping in Mali compared to existing global land products ? », *Journal of Applied Earth Observation and Geoinformation - In revision*, 2011.
- Wang K., Xu Y., Yu J. X., « Scalable sequential pattern mining for biological sequences », *CIKM '04 : Proceedings of the thirteenth ACM International Conference on Information and Knowledge Management*, ACM, New York, NY, USA, p. 178-187, 2004.

Annexe 6 : Les systèmes d'alerte précoce en Afrique de l'Ouest

Les systèmes d'alerte précoce visent à fournir une information fiable et préventive sur les risques potentiels de crise alimentaire ou d'insécurité alimentaire localisée (Brown *et al.*, 2007). Ces systèmes ont à la fois un rôle d'alerte en cas de crise (sécheresses, conflits) mais également de surveillance, sur le plus long terme, des besoins liés à des facteurs environnementaux (gestion de la fertilité des sols, démographie en hausse etc), et réfléchir sur la viabilité des systèmes agricoles actuels.

Dans le premier cas (alerte en cas de crise), l'information doit arriver le plus précocement possible, pour pouvoir diffuser des alertes et réagir à temps. L'estimation précoce des productions par type de culture est également importante pour anticiper les risques potentiels en lien avec l'évolution des marchés nationaux et internationaux. Il ne suffit pas, en effet, pour les populations rurales, de produire assez pour nourrir un nombre défini de personnes, mais il faut aussi arriver à vendre la production à un bon prix. Ainsi, les variations de prix des produits alimentaires sont très souvent à l'origine de l'insécurité alimentaire. De plus, les pays en développement ont théoriquement les ressources suffisantes pour atteindre l'autosuffisance alimentaire, mais les prix peu élevés des importations alimentaires ont conduit les décideurs politiques à encourager la production des cultures d'exportation. C'est pourquoi dans ces pays la sécurité alimentaire dépend à la fois de la production intérieure, mais aussi des conditions sur le marché mondial et des facteurs politiques.

Dans le deuxième cas (surveillance sur le long terme), les observations sur les évolutions des systèmes agricoles (l'extension des surfaces cultivées, le recours de plus en plus important à l'irrigation...), doivent permettre une évaluation de la vulnérabilité des agro-écosystèmes et des propositions de changements.

Le système d'alerte précoce sert donc à prendre les mesures nécessaires concernant les problèmes de déficits et excédents de production, de disettes et de famines. Pour prévoir la production, et éventuellement détecter des anomalies par rapport à d'autres années, le système d'alerte précoce estime une superficie cultivée au sol, puis un rendement de la culture. La multiplication de ces deux facteurs donne la production de la culture, et participe donc à l'évaluation de la sécurité alimentaire. Les estimations se font à l'échelle régionale et globale, et sont fournies le plus tôt possible durant la saison de croissance des plantes, et mises à jour périodiquement jusqu'à la récolte.

Les premiers systèmes d'alerte précoce nationaux, mis en place au Mali et au Tchad en 1984, l'ont été à l'initiative d'une O.N.G. de droit belge, AEDES. Suite aux épisodes de grande sécheresse de 1973-1974 et de 1984-1985, à leur impact dramatique sur les populations du Sahel et aux difficultés d'y apporter des réponses, le Ministère chargé des Collectivités Territoriales et les bailleurs de fonds ont fondé le Système d'Alerte Précoce (SAP) de la République du Mali. Placé sous la tutelle du Commissariat à la Sécurité Alimentaire (CSA), le SAP est toujours opérationnel et est en charge d'identifier les zones et les populations risquant de connaître une crise alimentaire et d'estimer les besoins.

Les méthodes développées dans ces premières expériences ont ensuite servi à la mise en place d'autres systèmes nationaux, notamment au Niger et au Burkina Faso. Le Comité permanent Inter-Etats de Lutte contre la Sécheresse dans le Sahel (CILSS) a mis en place en 1998 le Système Intégré d'Alerte Précoce (SIAP) qui repose sur quatre éléments : la gestion d'une banque de données (SGBD), un système d'analyse territorial (SAT), une procédure de représentation de la vulnérabilité structurelle (PRVS) et un système d'analyse conjoncturelle (SAC). Ses bénéficiaires finaux sont les SAP nationaux. Dans certains pays, ces SAP sont submergés, car ils doivent à la fois gérer l'information, aider aux décisions et mettre en œuvre des stratégies d'assistance. Dans d'autres pays, les SAP nationaux ont rencontré tant de difficultés qu'ils ont été fermés, c'est le cas au Burkina Faso par exemple. Face à ces problèmes, certaines organisations internationales et ONG impliquées dans la gestion de l'aide alimentaire ont développé leur propre méthodologie. C'est le cas de la méthode VAM (Vulnerability Assessment Method) du Programme Alimentaire Mondial ou du centre régional Agrhymet qui privilégie le suivi des zones à risques en milieu rural, et qui propose des cartes de vulnérabilité, ainsi qu'un système d'analyse de cette vulnérabilité. D'autres organisations telles que le Famine Early Warning Systems Network (FEWS NET), qui dépend de l'Agence des Nations-Unies pour le Développement International (USAID), développent une capacité de synthèse des multiples sources d'information existant au niveau national, et une capacité d'analyse de cette information (voir section 4.2). Il existe aussi :

- la Surveillance Globale pour la Sécurité Alimentaire (GMFS), qui est un projet de l'Agence Spatiale Européenne (ESA), faisant partie de l'élément de Surveillance Globale de l'Environnement et de la Sécurité (GSE).
- le Système Mondial d'Informations et d'Alerte Rapide sur l'alimentation et l'agriculture (SMIAR, ou en anglais Global Information and Early Warning System – GIEWS), qui a été créé par la FAO au début de la crise mondiale de l'alimentation au début des années 1970.
- la mission FOODSEC du Joint Research Centre (JRC) de la Commission Européenne, développée depuis 2001 dans le cadre de l'initiative de la Global Monitoring for Environment and Security (GMES).

Nous décrivons ici le SAP, le FEWS NET Mali, le GIEWS, le GMFS, et l'Agrhymet qui sont les cinq structures d'envergure analysant les besoins des populations vulnérables avec un suivi systématique d'un ou plusieurs pays d'Afrique de l'Ouest, et différents niveaux d'analyses et informations divulguées.

1. Le Système d'Alerte Précoce au Mali (SAP)

Le SAP se base sur une collecte permanente de données liées à la situation alimentaire et nutritionnelle des populations. Ces données couvrent des domaines très divers tels que la pluviométrie et la crue des fleuves, les déprédateurs, les campagnes agricoles, l'élevage et la pêche, les prix sur les marchés, les migrations de populations, leurs habitudes et réserves alimentaires, ainsi que leur état de santé .

Ces informations sont recueillies auprès des services administratifs et techniques du

gouvernement, des élus locaux et société civile depuis les communes vers les chefs-lieux de cercles, les chefs-lieux de régions et enfin Bamako. Avant d'être transmises sous forme de rapport régional mensuel vers Bamako, ces informations sont examinées par un groupe de travail SAP du Comité Régional de Développement qui se réunit mensuellement et qui regroupe les structures concernées par les questions de sécurité alimentaire sous la présidence du Conseiller aux Affaires Economiques et Financières du Gouverneur. À Bamako les rapports régionaux, les résultats d'enquêtes et les informations collectées sont analysés et rassemblés dans un rapport national mensuel qui est examiné puis adopté par le groupe de travail SAP, avant d'être publié et distribué sous forme de bulletin national, lequel est destiné aux autorités nationales, régionales, locales et aux organismes internationaux afin de leur permettre de prendre les mesures susceptibles de prévenir une crise alimentaire (Figure 1). Les zones où une dégradation de la situation alimentaire est suspectée, une équipe composée d'agents de la santé et des affaires sociales mène une enquête socio-médico-nutritionnelle approfondie afin de préciser l'importance des problèmes suspectés.

Le suivi des campagnes s'opère en combinant les connaissances et les observations de terrain de multiples services, en s'appuyant sur les apports de la télédétection.

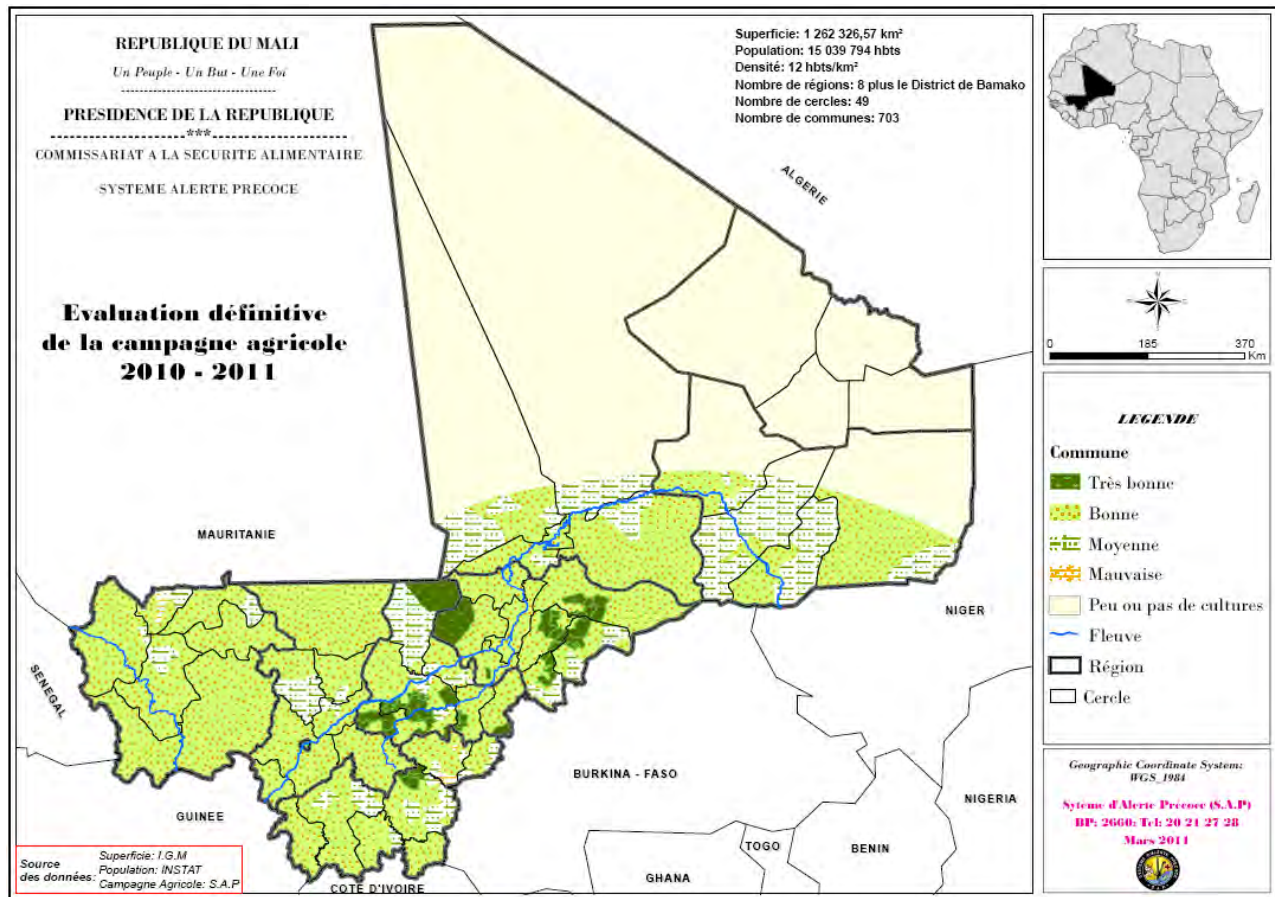


Figure 1 : Bulletin SAP (mars 2011) : évaluation définitive de la campagne agricole 2010-2011.

2. Le Famine Early Warning Systems Network (FEWS NET)

Le FEWS NET est une activité financée par l'USAID, qui collabore avec des partenaires internationaux, régionaux et nationaux pour fournir des informations d'alerte et de vulnérabilité en matière de sécurité alimentaire. FEWS NET utilise un ensemble de produits d'aide à la décision. Ces produits sont mis à jour chaque mois pour 25 pays (dont le Mali). Quatre agences gouvernementales interagissent en ce sens : l'Observatoire Géologique des Etats-Unis (USGS), l'Administration Spatiale et Aéronautique Nationale (NASA), l'Administration Océanographique et Atmosphérique Nationale (NOAA) et le Département de l'Agriculture des Etats-Unis (USDA). Le FEWS NET gère ainsi des informations et des données variées, incluant des données satellitaires sur les conditions météorologiques (pluies), des données de récoltes, sur les marchés de céréales et de bétail, de nutrition, et d'autres indicateurs d'alerte sur les revenus, les stratégies de vie et de survie des ménages. Les outils et services développés sont fournis aux décideurs : il s'agit de bulletins réguliers sur le suivi de la sécurité alimentaire, d'analyse des informations agro-hydro-météorologiques provenant d'imagerie satellitaire, tout au long des saisons agricoles, d'évaluation de la vulnérabilité (à la sécheresse, aux inondations, aux prix des produits...), et de renforcement des capacités avec une assistance technique fournie aux systèmes nationaux.

3. Le Global Information for Early Warning Systems (GIEWS)

Le GIEWS prépare des bulletins réguliers sur la production et sur les marchés des différentes cultures vivrières aux niveaux mondial, sous régional et national. Les données concernant les estimations de pluies et de rendements lui sont fournis par le programme « Global Monitoring for Food Security » (GMFS) de l'Agence Spatiale Européenne (ESA). Il suit continuellement l'offre et la demande de céréales dans tous les pays du monde. Les données statistiques sur la production, le commerce, l'aide alimentaire, sont analysés par des agroéconomistes sur le long terme. Les images satellites reçues en temps réel servent à prévoir les sécheresses. Le GIEWS organise, en commun avec le PAM, des missions d'évaluation rapide dans les pays affectés, et en publie des rapports ou alertes spéciales diffusées à la communauté internationale.

4. Le Global Monitoring for Food Security (GMFS)

Le GMFS est financé par l'ESA. Elle se concentre initialement sur l'Afrique sub-saharienne, et dirige son attention sur les utilisateurs finaux des organisations régionales et nationales dont le mandat consiste en la surveillance agricole pour la sécurité alimentaire et l'alerte rapide des crises alimentaires. Le GMFS a pour but d'appuyer et de compléter les systèmes existants d'information et d'alerte rapide sur l'alimentation et l'agriculture (Figure 2). Elle fournit des services d'alerte rapide, de cartographie agricole et d'évaluation des récoltes et de biomasse. Au niveau national, les activités du GMFS se concentrent sur l'Ethiopie, le Soudan, le Sénégal, le Zimbabwe, le Mozambique et le Malawi.

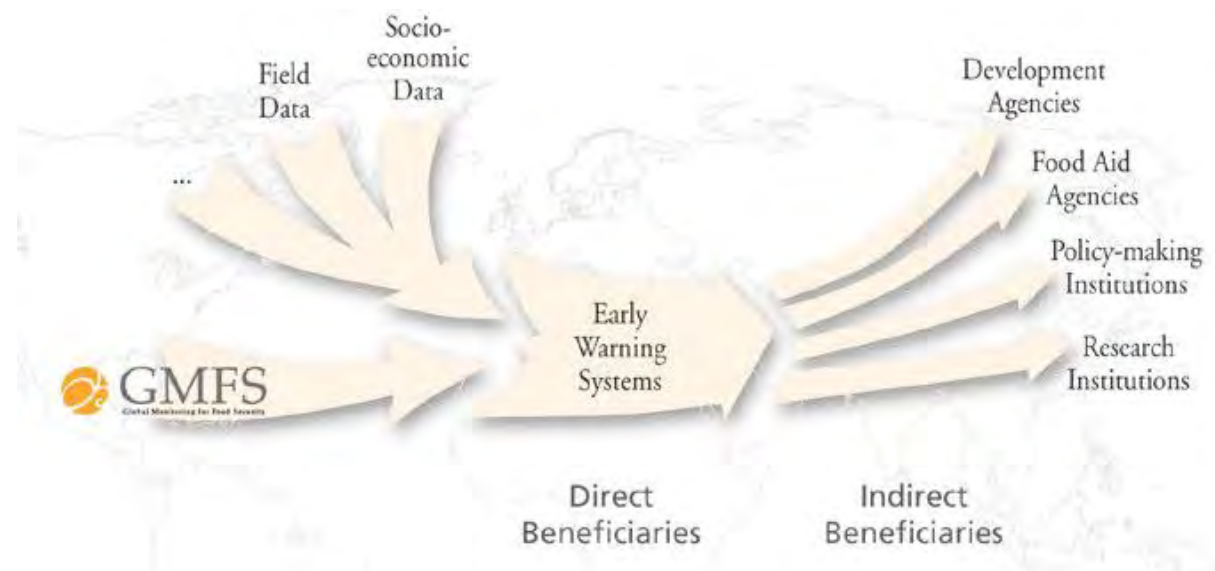


Figure 2 : Le GMFS en appui aux systèmes d'alerte précoce régionaux, nationaux et internationaux existants.

5. L'AGRHYMET

L'Agrhymet est un Centre Régional créé dans les années 70. C'est une institution spécialisée du CILSS qui regroupe neuf Etats membres (Mauritanie, Mali, Niger, Tchad, Burkina Faso, Sénégal, Gambie, Guinée, Guinée Bissau). Ses objectifs principaux sont de contribuer à la sécurité alimentaire et à l'augmentation de la production agricole dans les pays membres du CILSS, ainsi que d'aider à l'amélioration de la gestion des ressources naturelles de la région du Sahel. Le Centre Agrhymet assure le suivi agro-météorologique et hydrologique au niveau régional, la production de statistiques agricoles et de suivi des cultures, la gestion et la diffusion de l'information sur le suivi des ressources, et le renforcement de la coopération entre les Etats à travers l'échange de méthodologies et technologies. Il assure aussi la maintenance des instruments météorologiques et les relevés de plus de plus de 1000 postes d'observations dans les pays du CILSS (Figure 3). Concrètement, pour le suivi des cultures, le Centre propose par exemple des cartes des dates de semis basées sur l'estimation des pluies à partir de données METEOSAT, à une échelle spatiale de 5 km x 5 km. Le modèle « Diagnostic Hydrique des Cultures » (DHC) a été développé pour fournir les dates de semis, le développement phénologique des cultures, la condition hydrique des sols, et l'estimation de rendement, ceci tous les 10 jours pendant la saison des pluies.

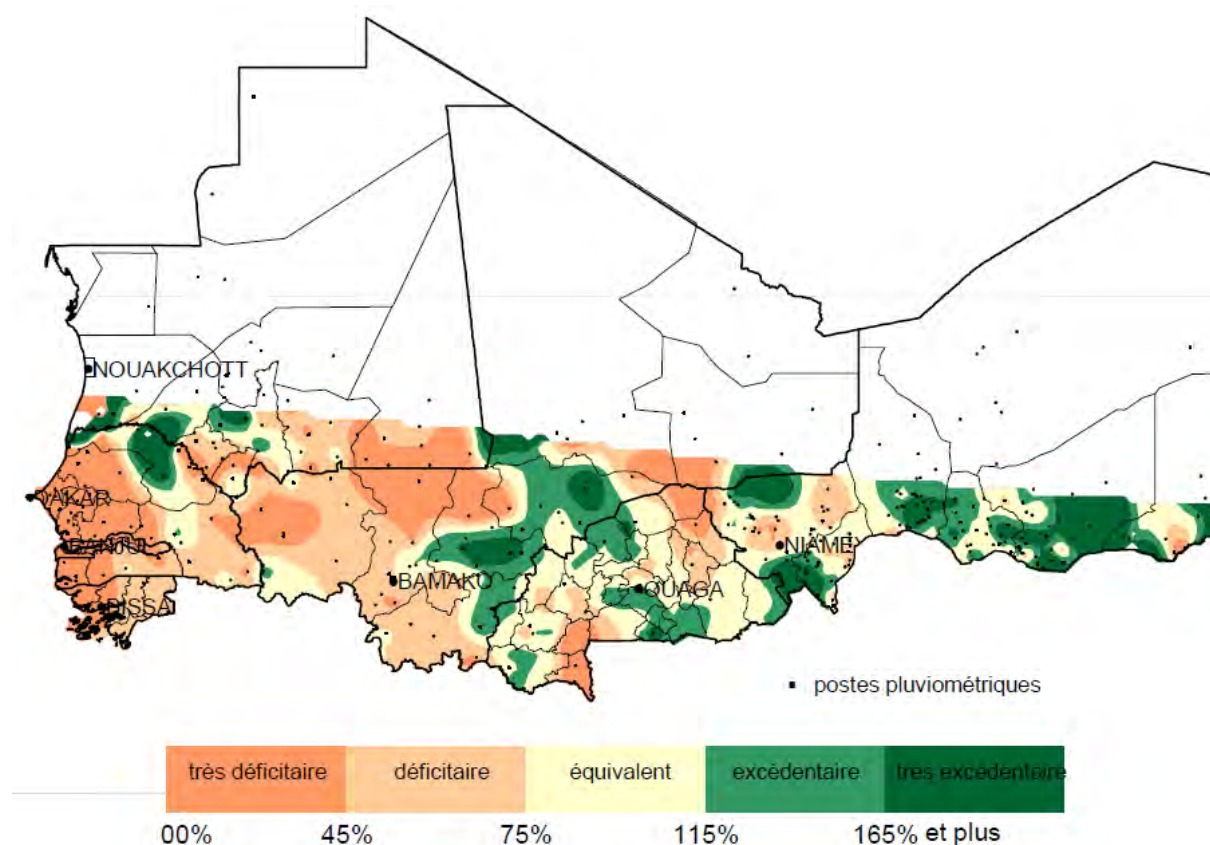


Figure 3 : Comparaison en pourcents des hauteurs pluviométriques du mois de juin 2010 à celles de l'année 2009 au Sahel. (Source : Agrhymet, bulletin mensuel Juin 2010).

Les systèmes d'alerte précoce régissent la surveillance de la sécurité alimentaire dans les pays vulnérables. Ils utilisent des méthodes différentes, sont associés dans des réseaux d'échanges des informations, et essaient de divulguer aussi rapidement que possible les informations alarmantes. Tous ces systèmes n'abordent pas tous les aspects de la sécurité alimentaire. Certains s'intéressent à l'accès aux marchés, d'autres à la stabilité politique, d'autres encore à la disponibilité alimentaire. Les activités du GMFS se limitent par exemple aux facteurs influençant la disponibilité alimentaire et pouvant être surveillés par télédétection alors que GIEWS s'intéresse davantage aux marchés, à l'offre et la demande de céréales.