Etude physico-chimique des surfaces continentales par imagerie Vis-PIR (400 – 2500 nm).

Cécile Gomez

Université de Montpellier

Ecole Doctorale GAIA

Mémoire déposé en vue de l'obtention de l'Habilitation à Diriger des Recherches

Etude physico-chimique des surfaces continentales par imagerie Vis-PIR (400 – 2500 nm).

Cécile Gomez

Soutenu le 09 Juin 2017

Devant un jury composé de :

R. Escadafal	Directeur de Recherche, IRD, UMR CESBIO	Rapporteur
S. Chabrillat	Senior Scientist, GFZ, Posdam	Rapporteur
B. Van Wesemael	Professeur, Université Catholique de Louvain	Rapporteur
V. Bellon-Maurel	Ingénieur IPEF, IRSTEA, UMR ITAP	Examinatrice
C. Walter	Professeur, AgroCampus Ouest, UMR SAS	Examinateur
P. Lagacherie	Ingénieur de recherche, INRA, UMR LISAH	Examinateur

UMR LISAH
(Laboratoire d'étude des Interactions entre Sol - Agrosystème – Hydrosystème)

IRD
(Institut de Recherche pour le Développement)

PREAMBULE

Ce mémoire est structuré en deux parties. Une première partie présente de manière factuelle le résumé de ma carrière, incluant ma formation et mon parcours, mes activités de gestion, animation de recherche, encadrement, et enfin la liste de mes publications. Une seconde partie synthétise mes travaux de recherche menés depuis mon recrutement en 2007 à l'IRD, et présente mes perspectives de recherches.

Ce mémoire est également accompagné d'une annexe consignée dans un document séparé, qui contient une sélection de tirés à part de 6 articles de revue à comité de lecture, dont je suis co-auteure.

Je voudrais remercier sincèrement toutes les personnes qui ont contribué de près ou de loin à la réalisation de ces travaux de recherche. Il est difficile de citer tout le monde. Je pense aux collègues avec qui j'ai (et j'ai eu) la chance de travailler, aux étudiants qui m'ont fait confiance en partageant un bout de leur parcours universitaire à mes côtés, aux staffs administratifs (antérieurs et actuels) sans qui mes recherches et encadrements d'étudiants ne pourraient pas se dérouler dans d'aussi bonnes conditions.

A vous tous, Merci.

Partie 1

Résumé de Carrière

I. CURRICULUM VITAE

CECILE GOMEZ

Chargée de Recherche 1ère classe à l'Institut de Recherche pour le Développement (IRD)
Docteur en Sciences de la Terre / Télédétection
Née le 18 Novembre 1978 à Pau (France).

ADRESSE PROFESSIONNELLE

UMR LISAH (Laboratoire d'étude des Interactions entre Sol - Agrosystème - Hydrosystème)
Campus SupAgro – Bâtiment 24
34060 Montpellier, FRANCE
E-mail : cecile.gomez@ird.fr
Tel : +33 (0)4 99 61 28 41

FORMATION

2004 : Doctorat en Sciences de la Terre / Télédétection
Etude des potentiels des données de télédétection multi sources pour la cartographie géologique : Application à la région de Rehoboth (Namibie)
Laboratoire des Sciences de la Terre (Lyon, France)

2001 : Diplôme d'Etudes Approfondies en Mécanique des Milieux Géophysiques et Environnement. Université de Grenoble, France.

2000 : Maîtrise de Mathématiques Appliquées, spécialité Analyse Numérique et Calcul Scientifique. Université de Pau et des Pays de l'Adour, France

CURSUS PROFESSIONNEL

Depuis 2011 : Chargée de Recherche IRD (CR1), UMR LISAH, Montpellier, France.
2009 - 2011 : Chargée de Recherche IRD (CR2), UMR LISAH, Tunis, Tunisie.
2007 - 2008 : Chargée de Recherche IRD (CR2), UMR LISAH, Montpellier, France.
2006 : Post Doctorante au Australian Centre for Precision Agriculture (Sydney, Australie)
2005 : Post Doctorante à l'Institut d'Astrophysique Spatiale (IAS, Orsay, France)
2001 - 2004 : Doctorante au Laboratoire des Sciences de la Terre (Lyon, France)

II. ANIMATION ET GESTION DE LA RECHERCHE

COORDINATION DE PARTENARIAT AVEC LE SUD	
Responsable convention d'application IRD / Direction Générale de l'Aménagement et de la Conservation des Terres Agricoles (DG/ACTA).	2009 – 2018
Co-responsable conventions de Recherche : IRD / Centre National de Cartographie et de Télédétection (CNCT) IRD / Ecole Nationale d'Ingénieur de Tunis (ENIT)	2009 – 2014 2011 – 2013

ANIMATION DE LA RECHERCHE	
Coordination de Projets	
ANR Blanc. **DIGISOL-HYMED**. Coordinatrice du projet pour la partie Tunisie. Coordinatrice du WP "soil properties mapping by spatial spectroscopy". *4 partenaires Tunisiens. 3 partenaires Français.*	2009 – 2012
INSU MISTRAL. **SICMED Lebna**. Coordinatrice du WP "Spatialisation des propriétés fonctionnelles des sols". *6 partenaires Tunisiens. 4 partenaires Français.*	2010 – 2015
CNES TOSCA. **MiHySpecSol**. Coordinatrice du projet. *1 partenaire Tunisien. 2 partenaires Français.*	2014 – 2015
Participation à des Projets	
CNES TOSCA. Mission **HYPXIM**. *1 partenaire Tunisien. 2 partenaires Français.*	2013 – 2015
ANR TRANSMED. **ALMIRA**. *3 partenaires Tunisiens. 4 partenaires Marocains. 6 partenaires Français.*	2014 – 2017
CNES TOSCA. **CES THEIA Cartographie Numérique des Sols**. *7 partenaires Français.*	2016 – 2018

ÉVALUATION & EXPERTISE	
Review pour des revues internationales de rang A. ~8 articles /an, pour e.g., Remote Sensing of Environment, IEEE - Transactions on Geosciences and Remote Sensing, International Journal of Remote Sensing, European Journal of Soil Science, Geoderma.	Depuis 2008

Sélection de communications en congrès Internationaux : IGARSS	2008 – 2009
Evaluation de projets : ANR, CNES TOSCA, FP6-EUFAR, FP7-EUFAR, FRS Belge, GRF German.	Depuis 2010
Membre de comité de pilotage de thèse : A. Blavet (IPGP/ONERA)	2015 – 2018
Membre du groupe Mission HYPXIM. Conception de la mission spatiale Française Hyperspectrale HYPXIM.	Depuis 2014

FORMATION & TRANSFERT

Formations scientifiques organisées et dispensées :

- Personnels scientifiques du CNCT (Tunisie). Etude de données Hyperspectrales pour l'étude pédologique et minéralogique (2010, 5 ½ jrs) et Traitement de données Hyperspectrales (logiciel ENVI) (2008, 6 ½ jrs)
- Equipe cartographique du Geological Survey of Namibia, pour l'initiation au traitement d'images (ENVI) et aux traitements de signal (IDL). (2004, 2 jrs)

Formations scientifiques co-organisées :
Rencontres HélioSPIR sur le thème "NIRS hors du labo" (2015, 1jr), "Aux frontières du proche infrarouge " (2014, 1jr) et "Spectrométrie InfraRouge et hétérogénéité » (2013, 1jr)

Chairman en congrès : GeoHyper (Montpellier, 2015), SFPT-GH (Grenoble, 2016), TeanGeo (Tunis, 2016)

RESPONSABILITÉS ET ACTIVITÉS ADMINISTRATIVES

Membre élu de la **Commission Scientifique Sectorielle (CSS) 3** à l'IRD	Depuis 01/09/2016
Membre élu du **Conseil d'unité** de l'UMR LISAH	Depuis 2011
Trésorière adjointe de l'Association **HELIOSPIR** http://www.heliospir.net/	2012-2015
Membre de la **Commission Documentation** à l'UMR LISAH	Depuis 2012
Coordinatrice principale d'une **mission aéroportée** Hyperspectrale en Tunisie (Gouvernorat du Cap Bon, projet ANR DIGISOL-HYMED)	2009-2010
Responsable du **laboratoire de spectroscopie Visible-Proche InfraRouge** à l'UMR LISAH	Depuis 2009

III. ENCADREMENTS

ENCADREMENT DE THESE

Walid Ouerghemmi. 2011-2014. Bourse IRD et SCAC de France.

Co-directeurs : S. Nacer (ENI Tunis) et P. Lagacherie (UMR LISAH, Montpellier).

Développement d'une approche de Double Séparation de Sources, pour la cartographie de propriétés de sol de surface, sur des surfaces semi-végétalisées, par imagerie Vis-PIR hyperspectrale. Cas du taux d'argile en milieu viticole Languedocien.

ENCADREMENT DE POST-DOCTORANTS

Karine Adeline. 2014. Contrat CNES TOSCA.

Co-encadrement avec J.M. Roger et N. Gorretta (UMR ITAP, Montpellier).

Etude de l'impact de la résolution spectrale, sur l'estimation de l'argile texturale par spectroscopie de laboratoire Vis-PIR (0.4-2.5 µm).

Asa Gholizadeh. 2014. Bourse de mobilité Européenne.

Co-encadrement avec P. Lagacherie (UMR LISAH, Montpellier).

Comparison of soil properties maps obtained by multivariate models and spectral indexes.

Rosa Oltra Carrio. 2013. Contrat CNES TOSCA.

Co-encadrement avec X. Briottet (ONERA, Toulouse).

Etude de l'impact des effets atmosphériques et de la résolution spatiale, sur l'estimation de l'argile texturale par imagerie hyperspectrale Vis-PIR (0.4-2.5 µm).

SUPERVISION DE CDD INGENIEURS

Azziza Bouzine. 2016. Contrat ANR-TRANSMED ALMIRA.

Cartographie Numérique des Sols du bassin versant du Tleta (Maroc), à partir de données multispectrales et de laboratoire VIS-PIR (0.4-2.5 µm).

Imene Majoub. 2009 et 2010. Contrat ANR DIGISOL-HYMED.

Aide à la mise en place de la campagne aéroportée hyperspectrale en Tunisie (Gouvernorat du Cap Bon), et de la campagne de terrain associée.

ENCADREMENT DE STAGES

Oumayma Sakri. 2016. Master 2 Signal, Imagerie et Applications parcours Télédétection (Université Paul Sabatier, Toulouse).

Co-encadrant : X. Briottet (ONERA, Toulouse).

Étude de sensibilité de l'approche de « Double-Séparation de Sources » pour la cartographie du taux d'argile sur surfaces semi-végétalisées

Paulina Michura. 2016. Master 2 CLUES (AgroParisTech, Grignon)

Co-encadrante : E. Vaudour (UMR ECOSYS, Grignon).

Prediction of soil properties through the use of Sentinel2 multispectral satellite imagery.

Loubna Abdelhak. 2015. Ecole d'Ingénieur en Agronomie, Option : Management des Ressources en Sols et en Eaux (IAV, Hassan II, Maroc).

Co-encadrant : M. Chikhaoui (IAV Hassan II).

Spatialisation de propriétés primaires du sol sur le bassin versant du Tleta (Maroc)

Bart Driessen. 2015. Master 2 Geo-Information Science (Université de Wageningen, Pays-Bas).

Etude de l'impact de la résolution spectrale sur l'estimation de teneur en argile par imagerie hyperspectrale Vis-PIR (0.4-2.5 µm). Simulation de données satellitaires multispectrales SENTINEL 2 et hyperspectrales HyXim.

Maroua Nouri. 2014. Master 2 AgroTIC Technologie de l'information et de la communication (SupAgro, Montpellier).
Encadrants principaux : J.M. Roger et N. Gorretta (UMR ITAP, Montpellier).
Prédiction du taux d'argile du sol sur une image hyperspectrale aéroportée par transfert des modèles d'étalonnage construits au laboratoire

Arthur Drost. 2013. Master 2 Geo-Information Science (Université de Wageningen, Pays-Bas).
Co-encadrant : J.M. Roger (UMR ITAP, Montpellier).
Analyse des incertitudes associées aux prédictions de teneur en argile obtenues par imagerie hyperspectrale Vis-PIR aéroportée (0.4-2.5 µm).

Mahmoud Sidi Youssef. 2012. Master 2 Sup'Com Tunis (Tunisie).
Co-encadrants : N. Chehata (UMR LISAH, Tunis) et S. Nejeh (CNCT Tunis).
Etude d'une classification de types de sol par Support Vector Machine.

Anis Guesmi. 2011. Master 2 Géologie appliquée à l'environnement (Faculté des Sciences, Tunisie).
Co-encadrant : S. Nejeh (CNCT Tunis).
Classification de type de sols par méthodes d'unmixing à partir de données hyperspectrales Vis-PIR.

Walid Ouerghemmi. 2010. Master 2 Instrumentation et mesures (ENI Tunis, Tunisie).
Co-encadrant : S. Nacer (ENI Tunis).
Etude du concept de séparation de sources en aveugle, pour l'extraction de spectre de sol à partir d'un spectre de mélange Sol/Végétation.

Lilian Baup. 2010. 2ème année IUT (France).
Co-encadrant : G. Coulouma (UMR LISAH, Montpellier).
Mise en place, étude et comparaison de deux protocoles de travail pour l'estimation de propriétés de sol : la spectrométrie Vis-PIR et l'expertise terrain.

Katia Statiokis. 2003. Licence de Géologie. Université Claude Bernard, LYON I (France).
Co-encadrant : P. Allemand (LST Lyon).
Etude de la relation entre la topographie et les données de réflectance ASTER.

Delphine Deslis. 2002. Master 1. ENS LYON (France).
Co-encadrant : P. Allemand (LST Lyon).
Cartographie géologique du massif des Coirons (Ardèche) par imagerie spatiale.

IV. PUBLICATIONS

ARTICLES DANS DES REVUES SCIENTIFIQUES INTERNATIONALES (INDEXÉES AU WEB OF SCIENCE)

24 au total. En 1er auteur : 9. En 2ème auteur : 5.
Avec partenaires du Sud : 7 (soit **47% des ACL avec le Sud** depuis mon recrutement à l'IRD en 2007).
H-Index : **12** (WOS), **13** (SCOPUS), **15** (GoogleScholar)

Code :
– en gras : signalement de ma présence dans la liste des auteurs
– en souligné : auteurs IRD
– avec une étoile : partenaires du Sud
– en italique : stagiaire ou doctorant que j'ai co-encadré

AC2005.1. **Gomez C.,** Delacourt C., Allemand P., Ledru P., Wackerle R. (**2005**). "Using ASTER remote sensing data set for geological mapping, in Namibia". *Physics and Chemistry of the Earth***,** Parts A/B/C, 30, 1-3, 97-108.

AC2005.2. Poulet F., Bibring J.P., Mustard J., Mangold, N., Gendrin A., Langevin Y., Ardvison, R., Gondet B., **Gomez C.** and the OMEGA Team (**2005**). "Phyllosilicates on Mars and Implications for Early Mars". *Nature*. 438, 623-627.

AC2006.1. Bibring J.P., Langevin Y., Mustard J.F., Poulet F., Arvidson R., Gendrin A., Gondet B., Mangold N., Pinet P., Forget F., Berthe M., and the OMEGA Team: Bibring J.P., Gendrin A., **Gomez C.,** Gondet B., Jouglet D., Poulet F., Soufflot A., Vincendon M., Combes M., Drossart P., Encrenaz T., Fouchet T., Merchiorri R., Belluci G., Altieri F., Formisano V., Capaccioni F., Cerroni P., Coradini A., Fonti S., Korablev O., Kottsov V., Ignatiev N., Moroz V., Titov D., Zasova L., Loiseau D., Mangold N., Pinet P., Doute S., Schmitt B., Sotin C., Hauber E., Hoffmann H., Jaumann R., Keller U., Arvidson R., Mustard J.F., Duxbury T., Forget F., Neukum G. (**2006**). "Global mineralogical and aqueous mars history derived from OMEGA/Mars Express data". *Science*. 312(5772): 400-4.

AC2007.1. Poulet F., **Gomez C.,** Bibring J.-P., Langevin Y., Gondet B., Mustard J., Belluci G. (**2007**) "Martian Surface Mineralogy from OMEGA/MEx: Global Mineral Maps" *Journal of Geophysical Research – Planet***,** VOL. 112, 2156-2202.

AC2007.2. **Gomez C.,** LeBorgne H., Allemand P., Delacourt C., Ledru P. (**2007**). "N-FindR method versus Independent Component Analysis method for lithological identification in hyperspectral imagery". *International Journal of Remote Sensing*, 28, 23, 5315-5338.

AC2007.3. Loizeau D., Mangold N., Poulet F., Bibring J.P., Gendrin A., Ansan V., **Gomez C.,** Langevin Y., Gondet B., Masson P., Neukum G. (**2007**). "Phyllosilicates in the Mawrth Vallis region of Mars". *Journal of Geophysical Research – Planet***,** Vol. 112.

AC2007.4. Jouglet D., Poulet F., Milliken R.E., Mustard J.F., Bibring J.P., Langevin Y., Gondet B., **Gomez C.** and the OMEGA team. (**2007**). "Hydration state of the Martian surface as seen by Mars Express OMEGA". *Journal of Geophysical Research – Planet*. vol.112, E08S06.

AC2008.1. Poulet F., Arvidson R., **Gomez C.,** Bibring J.P., Morris R.V., Langevin Y., Gondet B., Griffes J. L. (**2008**) "Mineralogy of Terra Meridiani and Western Arabia Terra from OMEGA/MEx and implications for formation". *Icarus*. 195(1), pp. 106-130.

AC2008.2. **Gomez C.,** Rossel R.A.V., McBratney A.B. (**2008**) "Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: an Australian case study". *Geoderma*, 2008, 146 (3-4), pp. 403-411.

AC2008.3. **Gomez C.,** Lagacherie P., Coulouma G. (**2008**) "Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements". *Geoderma*, 2008, 148 (2), pp. 141-148.

AC2009.1. Cecillon L., Barthes B., **Gomez C**., Ertlen D., Genot V., Hedde M., Stevens A., Brun J. J. (**2009**) "Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS)". *European Journal of Soil Science*, 60 (5), pp. 770-784.

AC2011.1. *Ouerghemmi* W.,* **Gomez C.,** Nacer* S., Lagacherie P. (**2011**) "Applying Blind Source Separation on hyperspectral data for Clay content estimation over partially vegetated surface". *Geoderma*, 163, 3-4, pp. 227-237.

AC2012.1. Lagacherie P., Bailly J.S., Monestiez P. and **Gomez C**. (**2012**). "Using scattered soil sensing field surveys to map soil properties over a region. An example with airborne hyperspectral imagery". *European Journal of Soil Science*. 63, pp. 110–119.

AC2012.2. **Gomez C.,** Coulouma G., Lagacherie P. (**2012**). « Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data", *Geoderma*, 189–190, pp. 176-185*.*

AC2013.1. Ciampalini R., Lagacherie P., **Gomez C.**, Grunberger O., Hamrouni* H., Mekki* I., Richard A. (**2013**). "Detecting and correcting biases of measured soil profiles data. A case study in the Cap Bon Region (Tunisia)". *Geoderma*. 192, pp. 68-76.

AC2013.2. **Gomez C.**, Le Bissonnais Y., Annabi* M., Bahri* H., Raclot D. (**2013**) "Laboratory Vis–NIR spectroscopy as an alternative method for estimating the soil aggregate stability indexes of Mediterranean soils", *Geoderma*, 209–210, pp. 86-97.

AC2013.3. Lagacherie P., Ruth A., **Gomez C.,** Bacha* S., Coulouma G., Hamrouni* H, Mekki* I. and Bacha* S., (**2013**) "Combining hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia)". *Geoderma*, 209–210, pp 168-176.

AC2014.1. Goge F., **Gomez C.**, Jolivet C., Joffre R. (**2014**) "Which strategy is best to predict soil properties of a local site from a national Vis–NIR database?". *Geoderma*, 213, pp. 1-9.

AC2015.1. **Gomez C.**, *Drost A.P.A.,* Roger J-M. (**2015**) "Analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data", *Remote Sensing of Environment*, 156, pp. 58-70.

AC2015.2. **Gomez C.**, *Oltra Carrio R.,* Lagacherie P., Bacha* S., Briottet X. (**2015**) "Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery". *Remote Sensing of Environment*, 164, pp. 1–15.

AC2016.1. Viscarra Rossel R.A., Behrens T., Ben-Dor E., Brown D.J., Demattê* J.A.M., Shepherd K.D., Shi Z., Stenberg B., Stevens A., Adamchuk V., Aïchi* H., Barthès B.G., Bartholomeus H.M., Bayer A.D., Bernoux M., Böttcher K., Brodsky L., Changwen D., Chappell A., Fouad Y., Genot V., **Gomez C.,** Grunwald S., Gubler

A., Guerrero Maestre C., Hedley C.B., Knadel M., Morras H.J.M., Nocita M., Ramirez-Lopez L., Roudier P., Rufasto Campos E.M., Sanborn P., Sellitto V.M., Sudduth K.A., Rawlins B.G., Walter C., Winowiecki L.A., Young Hong S., Ji W. (**2016**) "A global spectroscopic library to characterize the world's soil". *Earth Sciences Reviews*, 155, pp. 198-230.

AC2016.2. *Ouerghemmi*, W.,* **Gomez C.,** Nacer*, S., Lagacherie P. (**2016**). "Semi-blind source separation for estimation of clay content over semi-vegetated areas, from VNIR/SWIR hyperspectral airborne data". *Remote Sensing of Environment*. 181, pp. 251-263.

AC2016.3. **Gomez C.,** *Gholizadeh A.,* Borůvka L., Lagacherie P., (**2016**). "Using legacy data for predicting soil surface clay content from VNIR/SWIR hyperspectral airborne images". *Geoderma*, 276, pp. 84–92.

AC2016.4. *Gasmi* A.,* **Gomez C.,** Zouari* H., Masse A., Ducrot D. (**2016**). PCA and SVM as geo-computational methods for geological mapping in the southern of Tunisia, using ASTER remote sensing data set, *Arabian Journal of Geosciences*, December 2016, 9:753.

AC2017.1. *Adeline, K.,* **Gomez, C.,** Gorretta, N., Roger JM. (**2017**). "Sensitivity of soil property predictions to instrumental spectral configurations from laboratory Vis-NIR spectroscopy". *Geoderma*, 288, pp.143–153.

ARTICLES DANS DES REVUES SCIENTIFIQUES INTERNATIONALES – EN COURS

AC20XX.1. Walker, E., Monestiez, P., **Gomez, C.,** Lagacherie, P. Combining measured sites, soilscapes map and soil sensing for mapping soil properties of a region. *Geoderma*, En revision.

AC20XX.2. *Maroua, N.,* **Gomez, C.** Gorretta, N., Roger, JM. Clay content mapping from airborne hyperspectral Vis-NIR data by transferring Laboratory regression model. *Geoderma*, En revision.

AC20XX.3. **Gomez, C.,** *Adeline K.,* Bacha* S., *Driessen B.,* Gorretta N., Lagacherie P., Roger J.M., Briottet X. Sensitivity to spectral resolution of VNIR/SWIR imaging data for clay content mapping. *To be submitted to Remote Sensing of Environment*. Soumis.

AC20XX.4. *Gasmi* A.*, **Gomez C.,** Lagacherie P. and Zouari* H. Clay content mapping at large scale using VNIR / SWIR multispectral ASTER data. *To be submitted to Remote Sensing*. Soumis.

ARTICLES DANS DES REVUES SCIENTIFIQUES INTERNATIONALES (NON INDEXEES WEB OF SCIENCE)

A2014.1. *Gasmi* A.,* **Gomez C.,** Zouari* H., Masse A., Ducrot D. (**2014**). "Using Vis-NIR hyperspectral HYPERION data for bare soil properties mapping over Mediterranean area: plain of the Oued Milyan, Tunisia". *European Academic Research*. Vol. II, Issue 9.

CHAPITRES D'OUVRAGES (INDEXEES WEB OF SCIENCE)

Cw2014.1. Lagacherie P. and **Gomez C.** (**2014**). "What can GlobalSoilMap expect from Vis-NIR hyperspectral imagery in the near future?", In "*GlobalSoilMap: Basis of the global spatial soil information system". Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, Alex B. McBratney (Eds), January 27, 2014 by CRC Press.*

CHAPITRES D'OUVRAGES (NON INDEXEES WEB OF SCIENCE)

C2010.1. Lagacherie P., **Gomez C.,** Bailly J.S., Baret F., Coulouma G. (**2010**). "Chapter 8: The use of Hyperspectral Imagery for Digital Soil Mapping in Mediterranean areas" In *"Digital Soil Mapping: Bridging Research, Environmental Application, and Operation"* Boettinger J., Howell, D., Moore, A., Hartemink A., Kienast-Brown, S. (ed.). Springer Verlag. P. 93-102

C2012.1. **Gomez C.,** Lagacherie P., Bacha* S. (**2012**). "Using an Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia)." *Digital Soil Assessments and Beyond*. CRC Press 2012, Minasny, Malone & McBratney (eds), p. 387-392.

C2012.2. Ciampalini R., Lagacherie P., Monestiez P, Walker E., **Gomez C.** (**2012**). "Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia)." *Digital Soil Assessments and Beyond*. CRC Press 2012, Minasny, Malone & McBratney (eds), p. 393-398.

C2016.1. Gorretta N. and **Gomez C.,** (2016). "Chapter 18: Spectral–Spatial Unmixing Approaches in Hyperspectral VNIR/SWIR Imaging", In: Cyril Ruckebusch, Editor(s), *Data Handling in Science and Technology*, Elsevier, 2016, Volume 30, p. 579-611.

C2016.2. **Gomez C,** and Lagacherie P. (2016). "Chapter 1: Mapping of Primary Soil Properties Using Optical Visible and Near Infrared (Vis-NIR) Remote Sensing". Edited by N. Baghdadi and M. Zribi in *Land Surface Remote Sensing in Agriculture and Forest*, Remote Sensing Observations of Continental Surfaces Set, ISBN: 9781785481031, 2016, p. 1-36.

CHAPITRES D'OUVRAGES – EN COURS

C20XX.1. Lagacherie P. and **Gomez C.** "Vis-NIR-SWIR remote sensing products as new soil data for Digital Soil Mapping". Pedometric book. Soumis.

CONFERENCES INVITEES

CI2005-1. **Gomez C.**, Delacourt C., Allemand P., Ledru P. (**2005**). "Utilisation combinée de données de télédétection multi source: Potentiels en cartographie géologique". *Workshop International en Information Spatiale et développement Durable*. Rabat, Maroc. 14-16 Novembre 2005.

CI2006-1. **Gomez C.** (**2012**) "Cartographie Numérique des Sols par imagerie hyperspectrale pour la modélisation environnementale en région méditerranéenne ; Projet DIGISOL_HYMED", *Comité Mixte pour la Coopération Universitaire*, 22 Octobre 2012, Université Toulouse II Le Mirail, France.

CI2013-1. Lagacherie P. and **Gomez C.** (**2013**). "What can GlobalSoilMap expect from Vis-NIR hyperspectral imagery in the near future?" *Global Soil Mapping Conference*, 7-9 of October 2013, Orléans, France.

COMMUNICATIONS ORALES AVEC ACTES

COA2004-1. **Gomez C.,** Delacourt C., Allemand P., Ledru P. (**2004**). "Potentials and limitations of coupling ASTER and airborne geophysical data for improvement of geological mapping in arid region (Namibia, Rehoboth region)". *In Remote Sensing for Environmental Monitoring, GIS Applications, and Geology III,* edited by Manfred Ehlers, Hermann J. Kaufmann, Ulrich Michel, *Proceedings of SPIE* Vol. 5239 (SPIE, Bellingham, WA, 2004), 98-108. 2004.

COA2005-1. Forni O., Poulet F., Bibring J.-P., Erard S., **Gomez C.,** Langevin Y., Gondet B., the OMEGA Science Team, (**2005**) "Component Separation of OMEGA spectra with ICA". *36th Lunar and Planetary Science Conference*, March 14–18, 2005, League City, Texas.

COA2006-1. Loizeau, D., Mangold, N., Poulet, F., Bibring, J.P., Gendrin, A., **Gomez C.,** Langevin, Y., Gondet, B., Ansan, V., Masson, P., Neukum, G., the Omega Team and the HRSC Team (**2006**)."Phyllosilicates Rich Terrains in Mawrth Vallis Region, Mars, as Seen by OMEGA and HRSC/Mars Epress".*37nd Annual Lunar and Planetary Science Conference*, March 13-17, 2006, League City, Texas, abstract no.1658.

COA2006-2. **Gomez C.,** Poulet, F., Bibring, J.-P., Langevin, Y., Gondet, B., Pelkey, S. M., Mustard, J. F., Bellucci, G. and the Omega Science Team (**2006**) "Global Mineral Maps on Mars". *37nd Annual Lunar and Planetary Science Conference*, March 13-17, 2006, League City, Texas, abstract no.1405.

COA2007-1. Poulet F., Arvidson, R., **Gomez C.,** Bibring, J.P., Morris, R.V., Langevin, Y. and Gondet, B. (**2007**) "Surface composition of Terra Meridiani and Western Arabia Terra from OMEGA" *38nd Annual Lunar and Planetary Science Conference*, March 12-17, 2007, League City, Texas.

COA2008-1. Lagacherie P., **Gomez C.,** Bailly J.S., Baret F., Coulouma G. (**2008**). " The use of hyperspectral imagery for digital soil mapping in Mediterranean areas". *3rd workshop on Digital Soil Mapping*. Utah, USA.

COA2010-1. Lagacherie P., Le Martret H., Bendaya* L., Blanca Y., Chedly Derouiche* M., **Gomez C.**, Hamrouni* H., Mc Bratney A.B., Mekki* I., Minasny B. (**2010**) "Processing legacy soil data for global digital soil mapping : A Tunisian experience". *4th Global Workshop on Digital Soil Mapping*. Rome, Italy, 24-26 May 2010.

COA2012-1. Walker E., Monestier P., **Gomez C.,** Lagacherie P., Ciampalini R. (**2012**). "Cokriging for Large Spatial Datasets: Mapping Soil Properties at a Region Scale from Airborne Hyperspectral Imagery". *Ninth International Geostatistics Congress*, Oslo, Norway.

COA2014-1. **Gomez C.,** *Oltra-Carrió R.,* Bacha* S., Lagacherie P., Briottet X. (**2014**). « Clay contents predicted from Hyperspectral VNIR/SWIR imagery, under different atmospheric conditions and spatial resolutions". *WHISPERS*, Lausanne, Suisse, 24-27 Juin 2014.

COMMUNICATIONS ORALES SANS ACTES

56 au total. En tant que 1er auteur : 24.

2002-1 **Gomez C.,** Ledru P., Delacourt C., Allemand P. (**2002**). “Improvement of geological map from remote sensing and airborne geophysical data in arid region in Namibia”. *Workshop « geological interpretation of airborne geophysical data sets »,* Toulouse, France

2003-1 **Gomez C.,** Ledru P., Delacourt C., Allemand P., Wackerle R. (**2003**), “Integration of remote sensing and radiometric data set to improve geologic map: application to Namibia”. *EGS-AGU-EUG Joint Assembly,* Nice, France

2003-2 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (**2003**), “Potentials and limitations of ASTER and radio-element data for improvement of geological mapping in arid regions: application to Namibia”. *SPIE 10th international Symposium Remote Sensing*, Barcelona, Spain

2004-1 **Gomez C.,** Ledru P., Delacourt C. (**2004**) “A case history of supervised classification for geological mapping in semi-arid environment using multi-source remote sensing data”. *Congrès de Géologie Africaine*, Orléans.

2005-1 Poulet F., Mustard J., Langevin Y., Bibring J.P., Gendrin A., Gondet B. and **the Team OMEGA** (**2005**). “Variation of the 3µm absoprtion feature on Mars”. *1st Mars Express Science Conference*. European Space Research and Technology Centre (ESTEC) Noordwijk, The Netherlands, 21-25 February 2005.

2005-2 Poulet F., Mangold N., Bibring J., Loizeau D., Mustard J. F., Head J.W., Gendrin A., Ansan V., **Gomez C.,** Masson P., Neukum G. and the Omega-HRSC Co-I Teams, (**2005**) “Evidence for Phyllosilicates in the Early Mars Crust at Nili Fossae and Mawrth Valles Using OMEGA and HRSC Data”. *European Geosciences Union General Assembly*, April 25-29, 2005. Vienna, Austria.

2005-3 Poulet F., Bibring J.P., Gendrin A., Mangold N., Mustard J., Langevin Y., Gondet B., Loizeau D., **Gomez C.,** Chevrier V., Neukum G., OMEGA Team and the HRSC Team (**2005**) “Phyllosilicates as revealed by OMEGA and its implication for the early martian environment”. *37th annual meeting of the Division for Planetary Sciences, of the American Astronomical Society*. Cambridge, UK. 4-8 September 2005.

2005-4 Loizeau D., Mangold N., Poulet F., Bibring J.P., Gendrin A., **Gomez C.,** Langevin Y., Gondet B., Ansan V., Masson P., Neukum G., OMEGA Team and the HRSC Team (**2005**) “Outcrops of clay rich crust in Mawrth Vallis, Mars, revealed by OMEGA and HRSC/MEX”. *37th annual meeting of the Division for Planetary Sciences, of the American Astronomical Society*. Cambridge, UK. 4-8 September 2005.

2005-5 Poulet F., Bibring J.P., Gondet B., **Gomez C.,** Langevin Y., Gendrin A., Mangold N. (**2005**) “The phyllosilicate global pixture”. *OMEGA 6th Science team Meeting*. IAS, Orsay, France. 11-13 October 2005.

2005-6 Jouglet D., Poulet F., Bibring J.P., Langevin Y., Gondet B., **Gomez C.,** Mustard J., and the OMEGA Team (**2005**). “Hydration of the 3 µm feature: the IAS approach”. *OMEGA 6th Science team Meeting*. IAS, Orsay, France. 11-13 October 2005.

2006-1 Poulet F., Arvidson R., **Gomez C.,** Bibring J.P., Gondet B., Langevin Y., Gendrin A. (**2006**) "MEx/OMEGA characterization of Terra Meridiani, Mars". *36th COSPAR Scientific Assembly*. Beijing, China, 16 – 23 July 2006.

2006-2 Poulet F., **Gomez C.,** Bibring J.P., Langevin Y., Gondet B., Mustard J. (**2006**) “Martian Surface Mineralogy from OMEGA/MEx: Global mineral maps" *Eos Trans. AGU, 87*(52), Fall Meet. Suppl., 11-15 Décembre 2006.

2007-1 Lagacherie P. and **Gomez C.** (**2007**). “Imaging spectrometry : a new perspective for the Digital Soil Mapping of mediterranean areas”. *9th International Meeting on Soils with Mediterranean Type of Climate*. Aix-en-Provence, France, 22-25 October 2007.

2007-2 **Gomez C.** and Lagacherie P. (**2007**). “Spatialisation de propriétés du sol par Spectroscopie Visible et Proche Infrarouge”. *Colloque Sciences du Sol,* Grenoble, France, 25-26 Octobre 2007.

2008-1 **Gomez C.,** Lagacherie P., Coulouma G. (**2008**). “Continuum removal versus PLSR method for digital mapping of Clay and Calcium carbonate contents, using laboratory and airborne hyperspectral measurements”. Euro*pean Geosciences Union General Assembly*, April 14-18, 2008. Vienna, Austria.

2008-2 **Gomez C.,** Viscarra Rossel, R.A., McBratney, A.B. (**2008**) “Comparing predictions of soil organic carbon by field vis-NIR spectroscopy and hyperspectral remote sensing”. Euro*pean Geosciences Union General Assembly*, April 14-18, 2008. Vienna, Austria.

2008-3 Cécillon, L., Barthès, B., **Gomez C.,** Ertlen, D., Genot, V., Hedde, M., Stevens, A. and Brun, J. (**2008**) “Assessment and monitoring of soil conditions using indexes based on near infrared reflectance (NIR) spectroscopy”. *EUROSOIL* 2008. 25 - 29 August 2008, Vienna, Austria

2009-1 Lagacherie P., **Gomez C.**, Bacha* S., Ben Mechlia* M., Baret, F., Blanca Y., Coulouma G., Derouiche* M.G., Follain S., Hamrouni* H , Jacob F., Le Bissonnais Y., Le Martret H, Masmoudi* M.M., McBratney A.B., Minasny B., Monestiez P., Raclot D., Saber* N. (**2009**) « Cartographie Numérique des propriétés des sols par imagerie hyperspectrale pour la modélisation environnementale en zone méditerranéenne: le projet DIGISOL-HYMED» *Journées d'Etude des sols*, Strasbourg (France), Avril 2009.

2009-2 **Gomez C.,** Lagacherie P., Coulouma C. (**2009**) « Comparaison des méthodes « Continuum Removal » et « Partial Least-Squares Regression » pour la spatialisation du taux d’argile et de carbonate de calcium à partir de mesures de réflectance de laboratoire et aéroportées (400-2500nm) » *Journées d'Etude des sols*, Strasbourg (France), Avril 2009.

2009-3 Lagacherie P., **Gomez C**., Baret F. (**2009**) Cartographie des propriétés des sols par télédétection hyperspectrale en milieu méditerranéen. *Rencontre Télédétection* INRA, Bordeaux, France. 17-18 novembre 2009.

2010-1 Lagacherie P., Le Martret H., Bendaya* L., Blanca Y., Derouiche* M., **Gomez C.,** Hamrouni* H., Mc Bratney A.B., Mekki* I., Minasny B. (**2010**) "Vers une cartographie Numérique des Sols en Tunisie" *Centre internationaux de recherche agronomique* (GCARD) Montpellier, France. 28-31 Mars 2010.

2010-2 **Gomez C.**, *Ouerghemmi* W., Nacer S.* (**2010**) “Blind Signal Separation approach for soil properties mapping in partially vegetated pixels" *Workshop on Quantitative Applications of Soil Spectroscopy*. Potsdam, Germany. 15-16 Avril 2010.

2010-3 *Ouerghemmi* W.,* **Gomez C.,** Nacer* S. (**2010**) “Blind Signal Separation contribution for in soil properties mapping in partially vegetated pixels" *6th Astronomical Data Analysis Conference*. Monastir, Tunisie. 3-6 May 2010.

2010-4 Lagacherie P. and **Gomez C**. (**2010)** « Utilisation de l’imagerie hyperspectrale pour la cartographie numérique des propriétés des sols en zone méditerranéenne ». *10e Rencontres HélioSpir*, Montpellier, Agropolis international, 24 septembre 2010.

2010-5 *Ouerghemmi* W.,* **Gomez C.**, Nacer* S. (**2010**) “Adaptation du concept de séparation aveugle de sources pour la cartographie de propriétés de sol sur des

surfaces partiellement végétalisées" *XIIème Journées Scientifiques du Réseau Télédétection de l'A.U.F*, 23-25 Novembre 2010, Monastir, Tunisie.

2010-6 **Gomez C.** and Lagacherie P. (**2010).** "Utilisation de l'imagerie hyperspectrale pour la cartographie numérique des propriétés des sols en zone méditerranéenne". *XIIème Journées Scientifiques du Réseau Télédétection de l'A.U.F*, 23-25 Novembre 2010, Monastir, Tunisie.

2011-1 **Gomez C.**, Lagacherie P., Coulouma G. (**2011).** « Utilisation de l'imagerie hyperspectrale pour la cartographie numérique des propriétés des sols en zone méditerranéenne ». *Colloque d'inauguration de la Société Française de Télédétection Hyperspectrale (SFTH)*, 7 & 8 avril 2011 – Université Paris Diderot, France.

2011-2 Lagacherie P., **Gomez, C.,** Bacha* S., Ciampalini, R., Hamrouni* H. (**2011**). "A proof-of-concept of GlobalsoilMap.net in Tunisia: First advances and perspectives". *Annual meeting of the GlobalSoilmap.net consortium,* June 20-15 at JRC in Ispra, Italy

2011-3 **Gomez C.,** *Ouerghemmi* W.,* Naceur* S., Lagacherie P. (**2011**). "Applying Blind Source Separation on hyperspectral data for clay content estimation over partially vegetated surfaces". *Pedometrics*. August 30 – September 3, 2011. Czech Republic, Prague.

2011-4 Lagacherie P., **Gomez C.,** Coulouma, G. (**2011**). "Can hyperspectral images help us to gain insight into the real spatial structures of soil properties?: some learnings from a case study in southern France". *Pedometrics*. August 30 – September 3, 2011. Czech Republic, Prague.

2011-5 Monestiez, P., Walker E., **Gomez, C.**, Lagacherie P. (**2011**). "Multivariate geostatistical model to map soil properties at a region scale from airborne hyperspectral imagery and scattered soil field surveys: dealing with large dimensions". *Spatial Data Methods for Environmental and Ecological Processes – 2nd Edition* – Foggia and Gargano (FG), Puglia, Italy, on September 1-2, 2011.

2012-1 **Gomez, C.,** Lagacherie, P., Bacha* S. (**2012**). « Spatialisation de propriétés de sols par données aéroportées hyperspectrales sur le bassin versant du Lebna (300km², Cap Bon, Tunisie) ». *Journées d'Etude des sols*, Versailles (France), Mars 2012.

2012-2 Lagacherie P., **Gomez C.,** Bacha* S., Ciampalini R., Hamrouni* H., Monestiez P. (**2012**). "Digital soil mapping from legacy data and hyperspectral imagery in CapBon (Tunisia) First results and perspectives". *Global Soil Partnership Workshop*. 20-23 March 2012. Rome, Italia.

2012-3 **Gomez C.,** Lagacherie P., Bacha* S. (**2012**). "Using an Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia)". *5th Global Workshop on Digital Soil Mapping*. 10-13 April 2012. Sydney, Australie.

2012-4 Ciampalini R., Lagacherie P., Monestiez P, Walker E., **Gomez C.** (**2012**). "Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia)". *5th Global Workshop on Digital Soil Mapping*. 10-13 April 2012. Sydney, Australie.

2012-5 **Gomez C.,** Lagacherie P., Bacha* S. (**2012**). "Using a Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia)". *European Geosciences Union General Assembly*, 22-27 April 2012. Vienna, Austria.

2012-6 **Gomez C.**, *Ouerghemmi* W.,* Lagacherie P., Saber* N. (**2012**). Spatialisation du taux d'Argile sur surfaces semi-végétalisées, par traitement de données

hyperspectrales Hymap. *2ème colloque scientifique de la SFTH*, ONERA Toulouse, France, 18 et 19 juin 2012.

2012-7 *Ouerghemmi[*] W.,* Saber[*] N., **Gomez C.,** Lagacherie P., (**2012**) "Soil properties mapping using hyperspectral data and Blind Source Separation". *International Conference on Computer Related Knowledge ICCRK' 2012* , July 5-7, Sousse, Tunisia

2013-1 **Gomez C.,** Bissonnais Y., Annabi* M., Bahri* H., Raclot D. (**2013**). "Laboratory Vis-NIR spectroscopy as an alternative method for estimating the soil aggregate stability indexes of Mediterranean soils". in NIR 2013 - *16th International Conference on Near Infrared Spectroscopy*, 2 - 7 June 2013, La Grande-Motte, France.

2013-2 **Gomez C.,** Gogé F., Jolivet C., Joffre R. (**2013**). "Which strategy is best to predict soil properties of local sites from a national Vis-NIR database?" *in NIR 2013 - 16th International Conference on Near Infrared Spectroscopy*, 2 - 7 June 2013, La Grande-Motte, France.

2013-3 Lagacherie P. and **Gomez C.** (**2013**). "What can GlobalSoilMap expect from Vis-NIR hyperspectral imagery in the near future?" *GlobalSoilMap Conference* 2013, 7-9 October 2013, Orleans, France.

2013-4 **Gomez C.,** *Drost A.,* Roger J.M**.** (**2013**). « Analyse des incertitudes associées aux prédictions de la teneur en argile obtenues par imagerie hyperspectrale aéroportée ». *14èmes rencontres Héliospir*, 13th November 2013, Montpellier, France.

2013-5 Bahri* H., Annabi* M., Raclot D., **Gomez C.,** Le Bissonnais Y. (**2013**). « Etude la variabilité spatiale de la stabilité structurale des sols à l'échelle du bassin versant de Lebna en Tunisie ». *In: International Seminar on Sustainable Management of Water and Soil Resources: Context, Challenges and Prospects*, INRGREF, Hammamet, 19-20 Novembre 2013.

2014-1 *Oltra-Carrió R.,* **Gomez C.,** Bacha* S., Lagacherie, P., Briottet, X. (**2014**) "Sensitivity of soil property prediction obtained from Hyperspectral VNIR-SWIR imagery to atmospheric effects and degradation in image spatial resolutions". *EGU General Assembly 2014*, 27 April – 02 May 2014, Vienna, Autriche.

2014-2 **Gomez C.,** *Oltra-Carrió R.,* Bacha* S., Lagacherie P., Briottet X. (**2014**). « Etude de sensibilité aux effets atmosphériques et à la résolution spatiale, pour l'estimation d'Argile par Imagerie Hyperspectrale VNIR-SWIR ». *3ème colloque scientifique SFPT-GH*, 15 et 16 mai 2014, IGESA Porquerolles.

2014-3 *Oltra-Carrió R.,* **Gomez C.,** Bacha* S., Lagacherie P., Briottet X. (**2014**). "Analysis of the impact of the spatial resolution on soil property prediction from hyperspectral VNIR-SWIR imagery". *4th International Symposium on Recent Advances in Quantitative Remote Sensing: RAQRS'IV,* 22 – 26 September 2014, Universitat de València, Spain.

2014-4 *Adeline K.,* **Gomez, C.,** Gorretta N., Roger J.M. (**2014**). "Sensitivity of soil property prediction obtained from VNIR/SWIR data to spectral configurations". *4th International Symposium on Recent Advances in Quantitative Remote Sensing: RAQRS'IV,* 22 – 26 September 2014, Universitat de València, Spain.

2014-5 *Adeline K.,* **Gomez C.,** Gorretta N., Roger J.M. (**2014**). "Sensibilité des estimations de propriétés de sol par spectroscopie Vis-PIR de laboratoire, à la configuration spectrale". *15èmes Rencontres HélioSPIR*, 14 octobre 2014, Montpellier (Agropolis International).

2014-6 **Gomez C.,** Bacha* S., Ciampalini R., Coulouma G., Hamrouni* M.H., Jacob F., Lebissonais Y., Mekki* I., Monestiez P., Raclot D. Ruth A., Walker E., Lagacherie P. **(2014**). « Cartographie de propriétés de sol par imagerie Hyperspectrale Vis-PIR, sur le Bassin Versant du Lebna (Tunisie). » *Colloque International du programme SICMED*.

2015-1 *Nouri M.*, **Gomez C.,** Gorretta N., Roger JM (**2015**). "Prediction of soil clay content on an airborne hyperspectral image by transferring the calibration models constructed in the laboratory". *IGARSS 2015*, July 26-31 2015, Milan, Italia

2015-2 **Gomez C.,** *Gholizadeh A.,* Borůvka L., Lagacherie P. (**2015**). "Using legacy soil data for standardizing predictions of topsoil clay content obtained from VNIR/SWIR hyperspectral airborne images". *PEDOMETRICS 2015*, 14-18 September 2015, Córdoba, Spain.

2015-3 **Gomez C.,** *Gholizadeh A.,* Borůvka L., Lagacherie P. (**2015**) "Using legacy soil data for standardizing predictions of topsoil clay content obtained from VNIR/SWIR hyperspectral airborne images". *GeoHyper 2015*, 01-02 Octobre 2015, La Grande Motte, France.

2015-4 *Ouerghemmi* W.,* **Gomez C.,** Nacer* S., Lagacherie P. (**2015**) "Using legacy soil data for standardizing predictions of topsoil clay content obtained from VNIR/SWIR hyperspectral airborne images". *GeoHyper 2015*, 01-02 Octobre 2015, La Grande Motte, France.

2015-5 *Nouri M.,* **Gomez C.,** Gorretta N., Roger J.M. (**2015**). « La prédiction du taux d'argile du sol sur une image hyperspectrale aéroportée par transfert des modèles d'étalonnage construits au laboratoire ». *16èmes rencontres Héliospir*, 3th December 2015, Montpellier, France.

2016-1 **Gomez C.,** *Adeline K.,* Bacha* S., *Driessen B.,* Lagacherie P., Briottet X. (**2016**). "Impact of the spectral resolution of VNIR/SWIR imaging sensors on clay content prediction". *4e colloque scientifique du groupe thématique hyperspectral de la SFPT*, Grenoble, 11-13 mai 2016.

2016-2 Lefevre-Fonollosa M.J., Briottet X., Gastellu-Etchegorry J.P., Weber C., Feret J.B., **Gomez C.,** Delacourt C., Le Dantec N., Bajjouck T., Petit T., Carrère V., Marion R. (**2016**). "Preparing the future: the HYPXIM Mission". *36th EARSeL Symposium,* 2016, 20-24 June 2016, Bonn, Germany

COMMUNICATIONS SOUS FORME DE POSTER SANS ACTES

2009-1 Lagacherie P., **Gomez C.,** Bacha* S., Ben Mechlia* M., Baret, F., Blanca Y., Coulouma G., Derouiche* M.G., Follain S., Hamrouni* H , Jacob F., Le Bissonnais Y., Le Martret H, Masmoudi* M.M., McBratney A.B., Minasny B., Monestiez P., Raclot D., Saber* N. (**2009**) "Digital Soil Mapping by hyperspectral Imagery for Environmental modelling in Mediterranean regions : the DIGISOL_HYMED project" , *9ème congrès de la fédération des associations d'étude des sols d'Asie de l'Est et du Sud Est (ESAFS9)*, Seoul, 27-28 octobre 2009

2014-1 Carrère V., Briottet X., Jacquemoud S., Marion, R., Bourguignon A., Chami M., Chanussot J., Chevrel S., Deliot P., Dumont M., Foucher P.-Y., **Gomez C.,** Minghelli-Roman A., Sheeren D., Weber C., Lefevre-Fonollosa M.J., Mandea M. (**2014**). The French proposal for a high spatial resolution Hyperspectral mission. . *EGU General Assembly 2014*, 27 April – 02 May 2014, Vienna, Autriche.

2014-2 **Gomez C.,** *Drost A.,* Roger J.M. (**2014**). « Analyse des incertitudes associées aux prédictions de la teneur en argile obtenues par imagerie hyperspectrale Vis-PIR aéroportée (0.4-2.5 µm) ». *3ème colloque scientifique SFPT-GH*, 15 et 16 mai 2014, IGESA Porquerolles.

2015-1 **Gomez C,** Jacob F, Mekki* I., Molénat J, Zitouna-Chebbi* R., Raclot D, Annabi* M., S. B, Bahri* H., Ben Mechlia* N. Ben Slimane* A., Bouhlila* R., Boussema* M.R., Bouteffeha* M., Ciampalini R., Coulouma G., Dagès C., El Maaoui* M.A., Evrard O., Feurer D., Hamrouni* H., Lagacherie P., Le Bissonnais Y., Lefèvre I., Masmoudi* M., Nacer* S., Ouerghemmi* W., Pierrot-Deseilligny M.P., Planchon O., Prévot L., Voltz M. (**2015**). ”Characterization of landscape physical properties and biophysical processes for hydrological cycle over the Cap Bon region (Tunisia)”. *Conference internationale MISTRALS 2015*. 20-22 Octobre 201(, Marseille, France.

2015-2 Feurer D, **Gomez C,** Bailly J-S, Coulouma G, Lagacherie P, Jacob F, Prévot L, Raclot D. (**2015**). Advances in remote sensing of cultivated landscapes at LISAH. Les technologies de l'observation de la Terre pour l'agriculture et les risques naturels. *Seminaire CVT Allenvi*, Agropolis International, Montpellier, FRANCE.

SEMINAIRES

2003-1 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (Octobre **2003**). “Potentiels et Limitations de la télédétection pour la cartographie géologique en milieu semi aride”. BRGM, Orléans

2004-1 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (Février **2004**). “Potentials and limitations of remote sensing dataset for the geological mapping in arid climate”. *Geological Survey of Namibia*. Windhoek, Namibie

2005-1 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (**2005**). “Utilisation combinée de données de télédétection multi source : Etude des potentiels en cartographie géologique et perspectives en Sciences de l’Eau”. Maison des Sciences de l’Eau. Montpellier. 18 Octobre 2005.

2005-2 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (**2005**). “Utilisation combinée de données de télédétection multi source : Etude des potentiels en cartographie géologique”. Maison de la Télédétection, Montpellier, 13 Décembre 2005.

2006-1 **Gomez C.,** Le Borgne H., Delacourt C., Allemand P., (**2006**). “Extraction d’information lithologique par Analyse en Composantes Indépendantes ”. Laboratoire de Géodynamique et Planétologie, Nantes, 19 Janvier 2006.

2013-1 **Gomez C.** (**2013**). Cartographie Numérique des sols par Imagerie hyperspectrale Visible-Proche Infrarouge (VIS-PIR). UMR ITAP, 18 Février 2013, UMR ITAP, IRSTEA Montpellier, France.

2013-2 Lagacherie P., **Gomez C.** (**2013**). Utilisation de la télédétection Vis-PIR (400-2500nm) hyperspectrale pour cartographier des propriétés des sols. Séminaire de Télédétection INRA, Bordeaux, 17-19 Juin 2013.

2015-1 **Gomez C.** (**2015**). Cartographie Numérique des sols par Imagerie hyperspectrale Visible-Proche Infrarouge (0.4-2.5µm), en région Méditerranéenne. Séminaire à l'UMR SAS, 5 Mars 2015.

COMMUNICATIONS A VOCATION DE TRANSFERT VERS LA SOCIETE CIVILE

2004-1 **Gomez C.,** Delacourt C., Allemand P., Ledru P. (**2004**). “Potentials and limitations of remote sensing dataset for the geological mapping in arid climate”. *Geological Survey of Namibia*. Windhoek, Namibie.

2015-1 **Gomez C,** Jacob F. (**2015**). Présentation des activités conduites par le LISAH et ses partenaires sur le thème de l’adaptation des mosaïques paysagères. Formation "Climat et Éducation au développement durable" pour le ministère de l’éducation nationale. Organisée à l’IRD de Montpellier.

Partie 2

Travaux Scientifiques

I. POSITIONNEMENT SCIENTIFIQUE PERSONNEL

J'ai suivi une formation universitaire en Mathématiques Appliquées prolongée d'un DEA de Mécaniques des milieux Géophysiques et Environnement, ponctuées d'expériences de recherche à la fois courtes et enrichissantes. Ces premières expériences, effectuées au Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE) de Grenoble, étaient dédiées à des études de modélisation d'écoulement de calottes polaires. Ces travaux ont soulevé chez moi la question de l'obtention et de la précision de paramètres nécessaires au calage et à la validation de ces modèles (tels que la morphologie, topographie et température des calottes). Cette étape de caractérisation des surfaces de calottes polaires me semblait primordiale. Depuis ce moment, autrement dit depuis ma thèse, j'ai toujours travaillé à la ***caractérisation de surfaces continentales***, qu'elles soient un désert rocheux Terrestre, la surface Martienne ou un agrosystème, dans le but 1) d'accroitre nos connaissances sur le milieu et ainsi notre compréhension de son évolution et 2) d'améliorer la modélisation de processus environnementaux.

Enjeux scientifiques autour des surfaces continentales

Les surfaces continentales terrestres sont classées en cinq grands groupes (d'après la nomenclature de Corin Land Cover) : 1) les surfaces artificialisées (e.g., zones urbanisées), 2) les zones agricoles (e.g., vignes), 3) les forêts et zones semi-arides (e.g., déserts rocheux), 4) les zones humides (e.g., marais salants) et 5) les surfaces en eau (e.g., calottes glaciaires). ***Les enjeux environnementaux, sociaux et économiques de chacune de ces zones sont multiples***. Sans faire un recensement exhaustif de ces enjeux, on peut citer l'importance, pour les *zones agricoles*, de la préservation et de l'utilisation raisonnée des ressources en eau et en sol, le maintien d'une activité économique autour de l'agriculture, l'augmentation de zones agricoles dédiées à la production biomasse forestière ou de cultures à finalité énergétique ou bio-industrielle (pesant sur la capacité globale de mobilisation de terres à des fins alimentaires). En ce qui concerne les *zones semi-arides*, on peut citer l'importance d'étudier la Terre primitive (roches et minéraux reliques du premier milliard d'années) pour caractériser les processus à l'origine de la vie. Et l'étude de la *surface Martienne*, comme analogue de la surface Terrestre primitive, via une caractérisation de sa composition physico-chimique, permettrait également de mieux comprendre les conditions de développement de la vie.

Enjeux scientifiques autour de l'acquisition de données

Pour analyser et comprendre certains de ces enjeux, la modélisation est de plus en plus utilisée (modélisation hydrologique, modélisation de l'érosion, modélisation d'évolution de scénarios d'occupation du sol…). Néanmoins ces modèles requièrent (entre autre) des données précises et spatialisées sur les surfaces, qui s'avèrent trop souvent indisponibles et/ou imprécises.

Les surfaces continentales se caractérisent par leur topographie, composition physico-chimique (en surface et sub-surface), occupation végétale (naturelle ou anthropique) et biologique (faune), et leur utilisation par l'homme (évidemment pas sur Mars…). Pour caractériser de tels objets, plusieurs voies peuvent être envisagées dont l'observation de

terrain, les prélèvements et analyses associées, et l'utilisation d'outils tels que les mesures géophysiques (e.g., la résistivité électrique, gravimétrie) et la télédétection. Les campagnes de terrain ne sont pas toujours réalisables (par exemple sur Mars) et offrent difficilement la densité d'information nécessaire à la modélisation de processus environnementaux. La ***télédétection*** me semblait être un outil en adéquation avec mon objet (parfois difficile d'accès, souvent de large superficie) et les besoins de la finalité envisagée de modélisation environnementale (haute densité d'information, potentiellement à différentes dates). Plusieurs outils de télédétection existent, qui diffèrent par leur domaine spectral utilisé (Ultraviolet, Visible, Proche Infrarouge, Thermique, Micro-onde) et le système du capteur (actif ou passif). Et aucun capteur ne peut, à lui seul, nous permettre de caractériser une surface en termes de topographie, composition physico-chimique et occupation (anthropique ou naturelle). Il est souvent nécessaire de jongler avec les caractéristiques de différents capteurs (résolution spatiale et spectrale, répétitivité temporelle, qualité du signal), le signal en lui-même (Ultraviolet, Visible, Proche Infrarouge, Thermique, Micro-onde), et les composants présents entre la cible et le capteur (atmosphère, végétaux en surface). De plus la résolution spatiale d'acquisition de données varie d'un capteur à un autre, et n'est pas forcément en adéquation avec la variabilité spatiale des caractéristiques recherchées. Ainsi, la télédétection, bien qu'adaptée à l'étude d'un objet tel que la surface continentale, n'en est pas moins un outil complexe qu'il s'agit de mieux appréhender.

Mes premiers pas en cartographie de surfaces continentales

Mes travaux de thèse au LST (Lyon) et de post doctorat à l'IAS (Paris) ont été consacrés à la ***caractérisation de zones semi-arides et arides*** dans le but d'accroitre nos connaissances sur le milieu et ainsi notre compréhension de son évolution.

Durant mes travaux de thèse j'ai cherché à identifier le capteur ou la combinaison de capteurs en adéquation avec des besoins de cartographie géologique d'une région semi-aride inhabitée et non cultivée (Namibie), support de ressources minérales. Trois types de capteurs ont été étudiés : un imageur satellitaire Multispectral Visible, Proche Infrarouge (Vis-PIR, 400 – 2500) (capteur ASTER), un imageur satellitaire Hyperspectral Vis-PIR (capteur HYPERION), et un imageur aéroporté Gamma-ray. Chacun de ces 3 capteurs avaient fait ses preuves pour la cartographie minéralogique et/ou géologique (concernant ASTER : Rowan and Mars, 2002 ; Rowan et al., 2003 ; Ninomiya, 2004 ; concernant HYPERION : Cudahy et al., 2001 ; Kruse et al., 2003 ; concernant des données gamma-ray : Wilford et al., 1997 ; Tourlière et al., 2003). Bien que la combinaison de ces capteurs offre une classification d'unité géologique plus fine qu'un seul capteur, l'imagerie hyperspectrale Vis-PIR est la technologie offrant la plus grande source d'information sur la composition géologique d'une surface continentale. En particulier, dans un cas de mélange géologique intra-pixellaire, situation couramment rencontrée dans la nature, la résolution spectrale du capteur Vis-PIR HYPERION permet des cartographies sub-pixellaires de différents matériaux géologiques (Bayliss et al., 1997 ; Winter, 1999 ; AC2007.2).

Mon passage à l'Institut d'Astrophysique Spatial (IAS, Orsay, France) a été l'occasion d'étudier la caractérisation minéralogique de la surface Martienne, par le biais de l'imagerie hyperspectrale Vis-PIR. Ce travail fut une occasion unique d'étudier la composition d'une surface continentale sans aucune donnée de calibration ni validation, mais en utilisant nos connaissances du milieu terrestre comme référence (e.g., [AC2007.1]). Cette expérience « Martienne » m'a ainsi aidé à construire mes recherches d'aujourd'hui et de demain en

m'interrogeant sur l'interprétation et la validation des cartes issues de données de télédétection.

Un centrage sur la caractérisation des sols

Dans ce mémoire, j'ai pris le parti de ne présenter que mes travaux effectués depuis mon recrutement à l'IRD en 2007 et qui sont consacrés à la caractérisation des sols sur des agrosystèmes pluviaux méditerranéens. Ces agrosystèmes pluviaux méditerranéens sont représentatifs de plusieurs agrosystèmes pluviaux en Afrique du Sud et de l'Est, au Proche Orient et en Amérique du Sud, via les problématiques rencontrées : concentration des précipitations, urbanisation, surexploitation des ressources en eau et en sol, et fragilisation des agricultures familiales. L'UMR LISAH, unité dont je fais partie depuis mon recrutement à l'IRD, s'intéresse aux relations et interactions entre les processus biophysiques (i.e. hydrologiques, bio-géochimiques et écologiques) et les propriétés des paysages (propriétés naturelles telles que le sol et le relief, et propriétés en lien avec l'activité agricole telles que le parcellaire, les cultures, les pratiques agricoles...), en vue de répondre aux enjeux liés aux problématiques précitées (e.g., préservation et utilisation raisonnée des ressources en eau et en sol). La caractérisation des sols et de leurs changements d'états en vue de l'étude du fonctionnement hydrologique et pédologique des paysages cultivés est ainsi un axe fort du projet d'unité. Et c'est par l'utilisation de données hyperspectrales Vis-PIR que je tente d'accroitre nos connaissances sur les sols de ces agrosystèmes pluviaux méditerranéens.

Hypothèses et démarche scientifique

Le sol se caractérise par des propriétés que je considère comme pérennes à l'échelle de la décennie (dont la texture, le carbone organique, le carbonate de calcium $CaCO_3$, le fer, le pH), des propriétés variables dans le temps (humidité et rugosité de surface) et des matériaux partiellement présents à sa surface (cailloux et végétaux) qui varient également dans le temps. Je qualifie une propriété de sol comme étant « variable dans le temps » si celle-ci est susceptible de varier entre la date d'acquisition de mes données de télédétection et la date de collecte et/ou d'observation sur le terrain de mes données de références. Ces propriétés et matériaux précités influent sur le spectre Vis-PIR (Lobell et Asner, 2002 ; Ben-Dor et al., 2003 ; Denis et al., 2014 ; Bartholomeus et al., 2010). Ainsi un spectre Vis-PIR de sol acquis par télédétection contient une grande richesse d'information, mais l'étude de chacune de ces propriétés et matériaux s'avère complexe de par leurs interactions au sein de la réponse spectrale (e.g., l'humidité affecte l'estimation de la matière organique, Minasny et al., 2011).

Ayant choisi de me concentrer sur l'***estimation de propriétés du sol pérennes***, l'humidité, la rugosité et les matériaux partiellement présents à sa surface sont pour moi des facteurs de perturbations dans la mesure spectrale de ma cible. Considérant la jeunesse de la discipline, il était difficile de me lancer sur la prise en compte et correction de chacun de ces facteurs. Dans mon milieu d'étude (climat de type Méditerranéen subhumide avec une saison sèche prolongée), et compte tenu de l'heure d'acquisition nécessaire des données Vis-PIR (à midi solaire + ou – 2 heures), j'ai considéré l'humidité comme un facteur de perturbation négligeable. En milieu cultivé la rugosité de surface dépend entre autre des pratiques culturales et peut varier d'une parcelle à une autre, voire même à l'intérieur d'une parcelle (en cas de fissures par exemple). Malgré ce fort risque de perturbation dans mes mesures spectrales, j'ai travaillé sur des spectres non corrigés de cet effet de rugosité. Lever

ce verrou fait partie de mes perspectives de recherche (cf Chapitre V). Finalement, les matériaux partiellement présents à la surface du sol (cailloux et végétaux) sont nombreux sur mes terrains d'étude. L'impact de ces matériaux au sein du spectre Vis-PIR est donc potentiellement important et j'ai choisi de prendre en compte ce facteur de perturbation et de le corriger pour rendre possible l'estimation de propriétés pérennes de sol sur de telles surfaces.

Ainsi, partant d'études ayant montré que la spectroscopie Vis-PIR en laboratoire constitue une alternative aux analyses physico-chimiques des sols (e.g., Viscarra Rossel el al., 2006), et m'appuyant sur les choix précédemment expliqués, j'ai développé des travaux de spatialisation de propriétés pérennes de sol 1) sur des surfaces de sol nu, 2) en prenant en compte la présence partielle de végétation à la surface et 3) en associant un calcul d'incertitude à l'estimation de ces propriétés pérennes. Ces travaux sont exposés en Chapitre II (Figure 1.1, Axe 1).

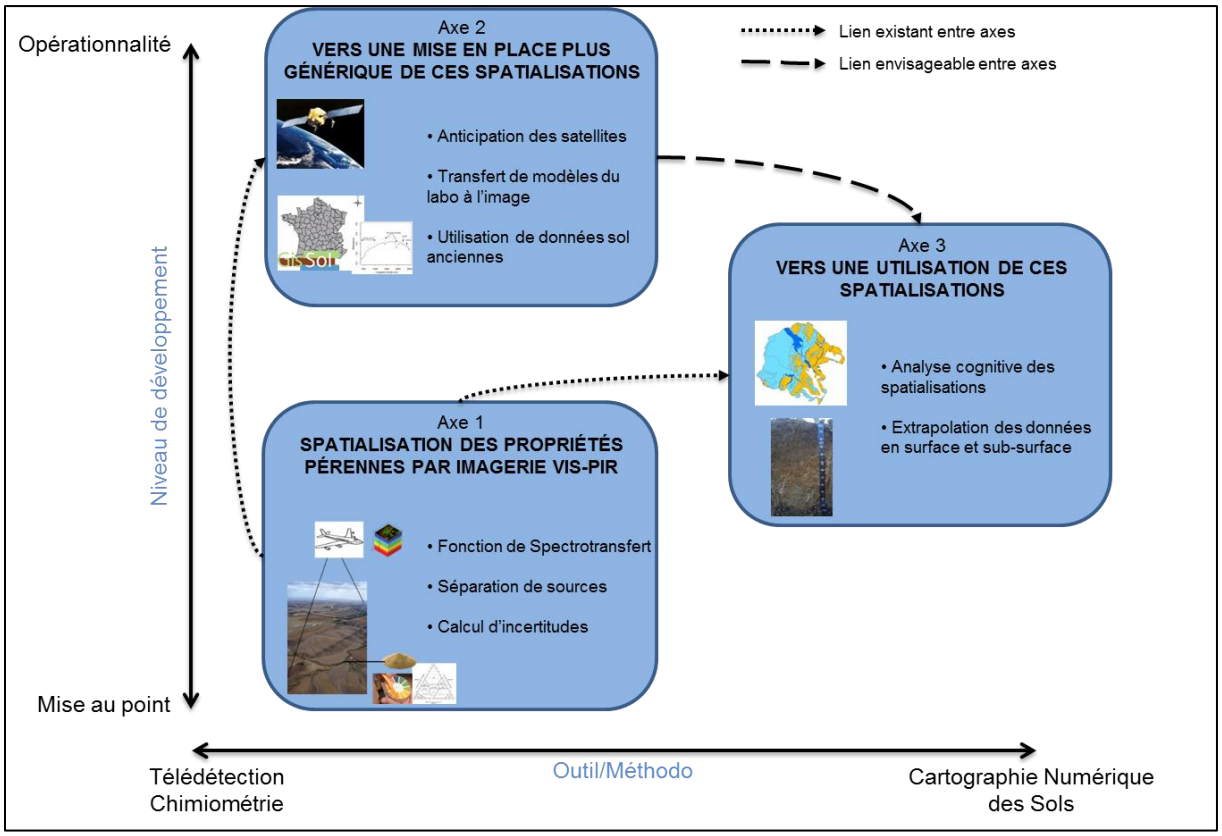


Figure 1.1 : Démarche scientifique personnelle

Une partie de la communauté scientifique travaillant sur la spectroscopie des sols (de laboratoire ou de télédétection) s'intéresse à la comparaison de méthodes pour une meilleure caractérisation de ses composants physico-chimiques (e.g., Stevens et al., 2010 ; Bayer et al., 2012). En effet le développement de méthodes de prédiction adaptées à chaque type de données (variables explicatives et à expliquer) pourrait permettre d'accroitre les précisions de prédictions. Pour ma part, j'ai préféré développer mes travaux principalement autour de la Partial Least Squares Regression (PLSR), qui est la méthode la plus fréquemment utilisée en spectroscopie car adaptée à des variables explicatives en grand nombre et corrélées entre elles (bandes spectrales). Car au-delà de cet enjeu d'amélioration

des précisions de ces estimations de propriétés pérennes de sol, réside un enjeu qui me semble plus prégnant qui est de tendre vers une ***démocratisation des spatialisations de propriétés pérennes de sol*** en tentant de développer des approches plus génériques d'utilisation de l'imagerie Vis-PIR.

Cet enjeu est en effet particulièrement important dans les pays du Sud qui constituent mon cadre d'étude privilégié, en tant que chercheuse IRD. Dans ces pays, les difficultés d'accès aux données Vis-PIR hyperspectrales et aux données de terrain nécessaires à la calibration des modèles de prédiction de propriétés pérennes de sol sont un frein important à leur utilisation, de par le coût et le temps nécessaires à leurs acquisitions. De plus, la plupart des travaux de spatialisation de propriétés pérennes de sol ont été menés sur des zones entre 10 km² et 600 km² (e.g., Stevens et al., 2006 ; Selige et al., 2006), et moins de dix sites au monde semblent avoir bénéficié de plus d'une campagne hyperspectrale d'après la littérature. Une mise en place plus générique des spatialisations de propriétés pérennes de sol permettrait de rendre l'accès à l'utilisation de données d'imagerie Vis-PIR moins couteuse en temps et en argent, afin de répondre à cette difficulté d'accès aux données fréquemment retrouvée au Sud, et permettrait une extension spatiale et temporelle des cartographies.

Ainsi, j'ai développé des travaux de spatialisation de propriétés pérennes de sol 1) utilisant des bases de données sol anciennes pour calibrer les modèles d'estimations de propriétés pérennes de sol, 2) associant des bases de données spectrales de laboratoire avec des images Vis-PIR hyperspectrales, et 3) anticipant l'arrivée de futurs satellites Vis-PIR. Ces travaux sont exposés en Chapitre III (Figure 1.1, Axe 2).

Au-delà de la mise en place méthodologique permettant l'estimation spatialisée de ces propriétés pérennes de sol de surface, je me suis interrogée avec l'appui de plusieurs collègues, sur ***l'utilisation possible de ces produits issus de l'imagerie vis-PIR***. Ces produits peuvent, à priori, être utilisés dans 1) des approches cognitives, telles que l'amélioration de cartes pédologiques existantes (e.g., par une modification des contours d'unité pédologique), 2) des approches de Cartographie Numérique des Sols qui se définit comme « la création et l'enrichissement de systèmes d'information pédologiques à références spatiales par des modèles numériques inférant les variations spatiales et temporelles des sols et de leurs propriétés à partir d'observations de sol et de données spatiales d'environnement des sols » (Lagacherie and McBratney, 2007), via l'utilisation de ces produits comme variables d'entrée de modèles de prédiction des sols et 3) des approches de modélisations (hydrologiques, agronomiques …) via l'assimilation de ces produits dans les modèles. Et mon environnement de travail au sein de l'UMR LISAH est propice à la réflexion collective autour de l'utilisation possible de ces produits issus de l'imagerie vis-PIR, grâce aux travaux réalisés par mes collègues en Cartographie Numérique des Sols et modélisation de processus biophysiques.

Sans être parvenu à aller jusqu'à l'assimilation de ces produits dans des modèles, j'ai développé avec l'appui de collègues 1) une analyse cognitive des estimations spatialisées pour améliorer nos connaissances sur le milieu (par exemple, par une mise à jour de carte pédologique) et 2) des travaux d'extrapolation des estimations en surface et sub-surface. Ces travaux sont exposés en Chapitre IV (Figure 1.1, Axe 3).

J'expose ensuite, en Chapitre V, mes perspectives de recherche pour les prochaines années. Un volet de mon travail de recherche se fera dans la continuité de ma démarche

actuelle, étendue à une problématique de spatialisation d'éléments de surface variables dans le temps (rugosité, couvert végétal, cailloux) qui sont à la fois *i)* des éléments « perturbateurs du signal » pour l'estimation de propriétés pérennes, et *ii)* des déterminants de processus hydrologiques (e.g., ruissellement, infiltrabilité). De ces travaux découleront des recherches autour de la spatialisation des propriétés dites « fonctionnelles » des sols (e.g., infiltrabilité des sols, réserve utile).

Mes terrains d'étude

Les travaux de recherche que je présente dans les trois chapitres qui suivent, ont été réalisés sur deux terrains d'étude : le bassin versant (BV) de La Peyne (France) et le bassin versant du Lebna (Tunisie). Le climat de ces sites d'étude, de type Méditerranéen sub-humide à saison sèche prolongée, rend favorable les acquisitions de données de télédétection optiques sur de larges surfaces de sol nu. Ainsi ces sites sont d'excellentes zones tests pour une étude de données optiques pour la caractérisation des sols. Le ***BV de la Peyne***, situé en Languedoc-Roussillon (~80 km²), est caractérisé par une culture viticole (Figure 1.2a). Les sols de ce site, soumis à l'exploitation viticole, se caractérisent par de faibles teneurs en matière organique, et inversement de fortes variations de teneur en carbonate de calcium, argile, et fer. Au cœur d'enjeux de gestion durable des sols pour le maintien de l'activité viticole, ce site est étudié par l'UMR LISAH depuis 1992, et est fortement décrit et caractérisé en terme pédologique, agronomique et hydrologique. Le ***BV du Lebna***, situé dans le gouvernorat de Nabeul (~210 km²), concentre une large variété de paysages (des plaines du littoral jusqu'aux reliefs montagneux) et de cultures pluviales (cultures de céréales, légumes et oliviers, prairies de parcours, …) (Figure 1.2b). Les sols de ce site se caractérisent par de faibles teneurs en limon et carbonate de calcium, et inversement de fortes variations de teneur en argile, sable et fer. Soumise à des problèmes d'érosion et de déficit hydrique, cette région a été l'objet de nombreux programmes de recherche menés conjointement par l'IRD et des partenaires tunisiens tels que la DG/ACTA Sol (Direction Générale des Sols du Ministère de l'Agriculture), l'INGREF (Institut National de la Recherche en Génie Rural, Eaux et Forêts) et l'INAT (Institut National Agronomique de Tunisie).

Sur ces 2 terrains d'étude, je bénéficie de jeux de données acquis en grande partie grâce à l'obtention de contrats de recherche (ANR, TOSCA-CNES) comprenant : des échantillons de sol associés à leurs propriétés physico-chimiques, des bases de données de sol anciennes en lien avec des campagnes de description de fosses pédologiques, des cartes de sols à grande échelle (e.g., carte au 1/5000 sur le BV de La Peyne produite par G. Coulouma, 2008), des spectres Vis-PIR de sol acquis en laboratoire, et des données hyperspectrales Vis-PIR aéroportées (acquises en 2003 sur le BV de la Peyne et en 2010 sur le BV du Lebna).

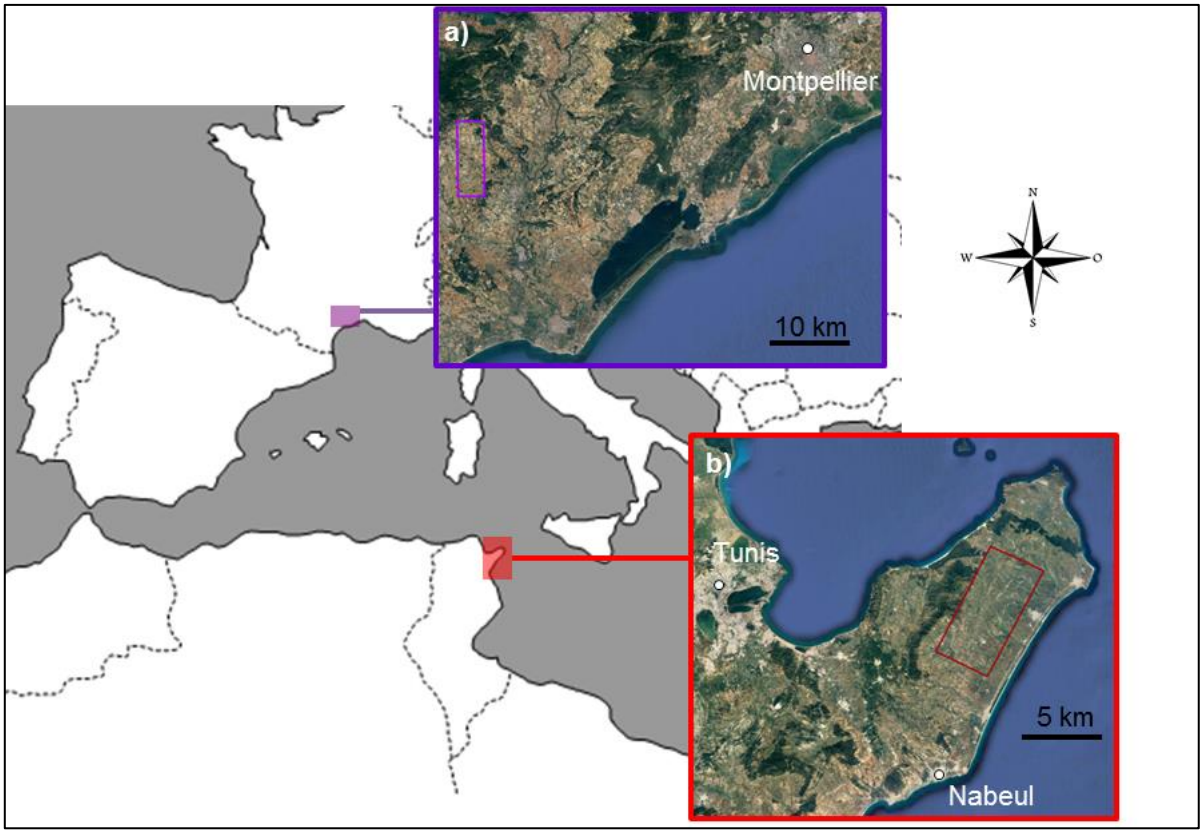


Figure 1.2 : Localisation des deux sites d'études : a) bassin versant de la Peyne en France (le polygone violet localise l'emprise de l'image aéroportée hyperspectrale Vis-PIR que j'utilise dans mes travaux) et b) bassin versant du Lebna en Tunisie (le polygone rouge localise l'emprise de l'image aéroportée hyperspectrale Vis-PIR que j'utilise dans mes travaux).

II. SPATIALISATION DE PROPRIETES PERENNES DE SOL PAR IMAGERIE VIS-PIR

2.1 INTRODUCTION

L'estimation de propriétés pérennes de sol peut être réalisée via le développement de fonctions de spectrotransfert (FST). Une FST *f* relie la variable *y* à prédire (propriété pérenne) pour une cible *C* (surface de sol), à des variables descriptives *X* (mesures spectrométriques Vis-PIR) mesurées sur cette même cible *C* (e.g., Zornoza et al., 2008), suivant : *y=f(X)*. Cette FST *f* peut être calibrée, offrant des estimations en valeur absolue, ou non calibrée, offrant alors des estimations en valeur relative. En télédétection, la calibration de ces FST repose sur la constitution d'une base de données [*X,y*] composée de spectres de réflectance Vis-PIR acquis par spectrométrie sur des surfaces de sol nu (variables descriptives *X*), associés à la propriété de sol d'intérêt mesurée en laboratoire sur un échantillon de surface prélevé sur ces mêmes surfaces de sol nu (variables à prédire *y*).

Des FST pour l'estimation de propriétés physico-chimiques de surfaces continentales (principalement pour une étude minéralogique) ont initialement été développées via l'analyse de bandes d'absorption et de géométrie du spectre. Le principe général repose sur le calcul d'indices spectraux (Clark and Roush, 1984) basés soit sur le rapport entre (à minima) deux valeurs de réflectance, soit sur différents modes d'analyse d'une bande d'absorption tels que le calcul de l'aire de la bande d'absorption, la pente entre deux points, le calcul de la profondeur d'absorption maximale, l'asymétrie de la bande d'absorption, etc (e.g., Escadafal, 1994 ; Madeira Netto et al., 1997 ; Mathieu et al., 1998). De telles fonctions peuvent ou non être calibrées.

Compte tenu de la complexité d'un spectre acquis sur une surface continentale, en particulier le sol, l'utilisation d'indices spectraux peut s'avérer insuffisante pour l'estimation de certaines propriétés. Afin de prendre en compte le spectre dans son ensemble, et non une partie spécifique du spectre, une seconde approche consiste à faire appel à la chimiométrie qui se définit comme l'ensemble des méthodes statistiques, graphiques ou symboliques permettant de mieux comprendre les informations obtenues dans le domaine de la chimie (notion introduite en 1972 par Wold and Kowalski ; e.g, Geladi and Esbensen, 1990). L'approche chimiométrique consiste ici à rechercher les relations statistiques entre l'ensemble du spectre *X* (et non une partie spécifique comme précédemment) et la propriété d'intérêt *y* au moyen de modèles de régression multivariés.

Comme introduit en Chapitre I, la présence d'éléments (tels que des éléments grossiers, des résidus de végétation sèche ou encore de végétation verte), une humidité significative ou une rugosité forte au sein de la cible *C* perturbe la mesure spectrale *X* acquise sur cette cible *C* (e.g., Ben-Dor et al., 2009). La présence de végétation en surface est considérée comme un facteur de perturbation de premier ordre pour l'estimation de propriétés physico-chimiques sur les surfaces continentales (e.g., Palacios-Orueta et al., 1999). Bartholomeus et al. (2010) estime qu'au-delà de 5 % de végétation dans les pixels, l'estimation du taux de carbone organique dans le sol de surface est impossible. Les FST *f* calibrées sur des cibles de sol nu, ne peuvent donc être appliquées qu'aux surfaces de sol nu.

Bien que les spatialisations obtenues à l'issu du développement de ces FST offrent une vision synoptique des propriétés physico-chimiques des surfaces imagées, aucun indice

de qualité ne permet de juger la qualité de ces estimations spatialisées. Les FST sont calibrées et validées sur des bases de données comprenant la plupart du temps moins de 200 sites, pour être appliquées ensuite à des millions de sites (une surface de 12 x 12 km imagée avec une résolution de 5 m, contiendra plus de 5 millions de pixels). Et seuls les indices de qualité des FST (e.g., coefficients de détermination de validation, R^2_{val} ou Erreur Standard de Prédiction / Standard Error of Prediction, *SEP*) servent de « garant » à la qualité des spatialisations.

Dans ce contexte, mes travaux ont porté successivement sur les aspects suivant : cartographie de propriétés pérennes de sol sur des surfaces continentales nues (décrite en section 2.2) et des surfaces continentales semi-végétalisées (décrite en section 2.3), puis cartographie de l'incertitude associée à ces estimations (décrite en section 2.4).

2.2 CAS DES SURFACES CONTINENTALES NUES

Suite à la mise en évidence du fort potentiel de la spectroscopie Vis-PIR de laboratoire pour l'estimation de propriétés pérennes des sols (e.g., Ben-Dor and Banin, 1995a et 1995b ; Viscarra Rossel et al., 2006), différents travaux de cartographie de ces mêmes propriétés des sols ont été initiés à partir de données hyperspectrales aéroportées Vis-PIR sur des surfaces de sol nu (e.g., Ben-Dor et al., 2002 ; Selige et al. 2006 ; Stevens et al., 2010). Pour notre part nous avons mené une étude de spatialisation de huit propriétés pérennes[1] de sol (argile, sable, limon, Carbonate de Calcium $CaCO_3$, fer, pH, carbone organique et Capacité d'Echange Cationique CEC) à partir des données hyperspectrales Vis-PIR AISA-DUAL aéroportées collectées sur le bassin versant du Lebna en Tunisie (300 km²) à une résolution spatiale de 5 m (Figure 1.2b). L'intérêt de nos travaux repose sur trois points : *i)* la superficie couverte par les données hyperspectrales est une des plus importantes jusqu'ici étudiées et donc couvre une grande variabilité pédologique, *ii)* la gamme de propriétés de sol d'intérêt est une des plus larges jusqu'alors étudiée et *iii)* les FST calibrées avec succès ont été appliquées à l'ensemble des surfaces de sol nu de la zone d'étude permettant ainsi une analyse détaillée du bénéfice et des limites de cartes d'estimation de propriétés pérennes de sol.

Pour chacune des huit propriétés de sol d'intérêt, une FST reliant les variables *y* à prédire (propriété de sol) aux variables descriptives *X* (mesures Vis-PIR aéroportées acquises par le capteur AISA-DUAL ; 400 - 2500 nm), a été construite à partir de modèles Partial Least Square regression (PLSR, Tenenhaus, 1998) et en utilisant une base de 129 données de calibration *[X,y]* et sans données de validation. D'après l'analyse des coefficients de détermination de calibration R^2_{cal} et l'Erreur Standard de Calibration (Standard Error of Calibration – *SEC*), seules quatre propriétés de sol peuvent être estimées sur notre zone d'étude, par des FST calibrées à partir des données Vis-PIR aéroportées AISA-DUAL : l'argile, le sable, le fer et la CEC.

[1] Je rappelle que je qualifie une propriété de sol comme étant « pérennes» si celle-ci n'est pas susceptible de varier entre la date d'acquisition de mes données de télédétection et la date de collecte et/ou d'observation sur le terrain de mes données de références.

Résultat Marquant :

Cette étude nous a permis de confirmer les deux règles identifiées initialement par Ben-Dor et al. (2002), que doivent respecter les propriétés de sol pour pouvoir être prédites par imagerie hyperspectrale Vis-PIR [C2012.1] :

1) la propriété de sol d'intérêt doit avoir une signature spectrale marquée, liée à un composant chimique ou une structure physique associée (condition vérifiée dans notre cas pour l'argile et le fer), et/ou être corrélée à une propriété de sol ayant une signature spectrale marquée (condition vérifiée dans notre cas pour le sable et la CEC).

2) et la propriété de sol d'intérêt doit avoir une large gamme de valeur sur le site donné (condition vérifiée dans notre cas pour les quatre propriétés prédictibles).

Appliquées à l'ensemble des pixels de sol nu de l'image AISA-DUAL du bassin versant du Lebna, ces FST ont permis l'obtention de cartes d'argile texturale, de sable, fer et CEC sur un large territoire (48 % de 300 km²) et à haute résolution spatiale (5 m). L'analyse de la carte d'estimation d'argile texturale a permis de mettre en évidence un modèle de sol régional complexe sur la zone d'étude, principalement dû à des variations de la lithologie (Figure 2.1). En milieu cultivé, le travail du sol entraine une redistribution du sol sur le premier horizon cultivé. Ainsi, malgré la faible pénétration en surface des ondes électromagnétiques dans le Vis-PIR (pénétration de l'ordre de leur longueur d'onde, i.e. autour du µm), chaque estimation de propriétés physico-chimiques de sol de surface faite par imagerie Vis-PIR peut être considérée comme représentative du premier horizon cultivé, élargissant ainsi encore un peu plus, la portée de ces résultats.

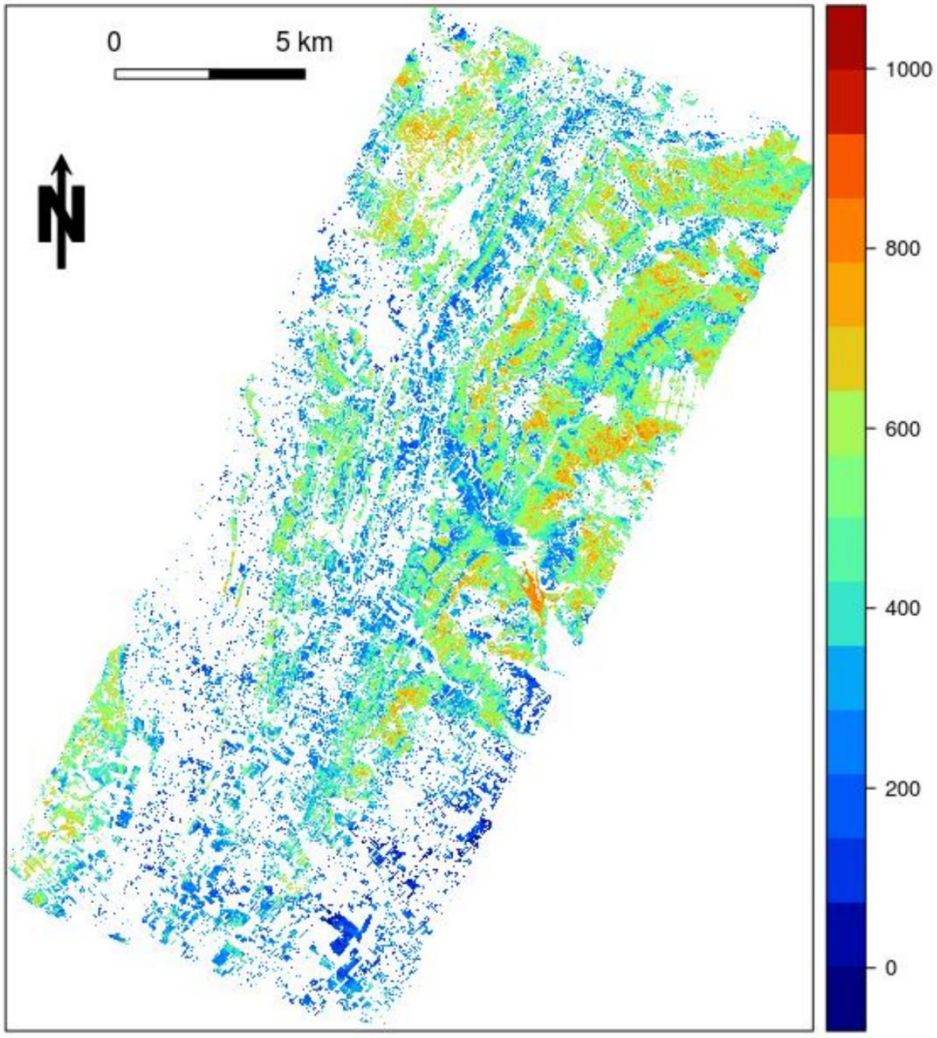


Figure 2.1 : Cartographie de la teneur en argile texturale, obtenue par traitement de données hyperspectrales aéroportées AISA-DUAL (résolution spatiale de 5 m) à partir d'un modèle d'estimation PLSR [C2012.1]. Les zones en blancs correspondent aux pixels masqués (zones de végétation et urbaines).

2.3 CAS DES SURFACES CONTINENTALES SEMI-VEGETALISEES

Les méthodes développées jusqu'à présent en télédétection hyperspectrale Vis-PIR pour l'estimation de propriétés pérennes de sol de surface restent inopérantes en présence de végétation (e.g., Selige et al., 2006 ; Stevens et al., 2010 ; AC2008.2 ; AC2012.2 ; C2012.1). Dans ce contexte, nous avons mis en place une approche dédiée à la spatialisation de propriétés de sols sur les surfaces semi-végétalisées. Cette approche, appelée « Double Séparation de Sources » (DSS), consiste à *1)* estimer le spectre Vis-PIR de sol $\hat{X}_{sol}$ à partir de *n* spectres Vis-PIR « mixtes » X_{mix} en utilisant une méthode de Séparation Aveugle de Sources, puis *2)* estimer une propriété physico-chimique du sol, à partir du spectre de sol estimé $\hat{X}_{sol}$ en utilisant une FST. Ces travaux ont été menés dans le cadre de la thèse de W. Ouerghemmi que j'ai co-encadré entre 2011 et 2014.

Cette approche a été développée et testée sur des sites tests semi-végétalisés extraits de données hyperspectrales HyMap aéroportées collectées sur le bassin versant de La Peyne (Languedoc-Roussillon, France) avec une résolution spatiale de 5 m (Figure 1.2a). Ces sites tests semi-végétalisés sont couverts à minima de vigne verte et de sol. Et des éléments grossiers et de la végétation basse (herbe sèche et/ou verte, litière) peuvent également être présents en surface. Chaque site test correspond à une grille de 3x3 pixels HyMap, soit 15x15 m, pour laquelle 1) la propriété physico-chimique « vraie » est connue au niveau du pixel central de la grille, et 2) les propriétés physico-chimiques du sol (et donc les spectres X_{sol} correspondant), sont supposées invariables à l'échelle de 15x15 m. Cette approche a été testée pour l'estimation du taux d'argile texturale. La première étape de l'approche développée est basée sur la séparation aveugle de sources (ou Blind Source Separation – BSS, Ans et al, 1985 ; Jutten & Hérault, 1991) qui consiste à estimer *p* signaux sources inconnus à partir de *n* signaux observés « mixtes », sans connaissance à priori de l'abondance des signaux sources au sein des *n* mélanges (Etape 1, Figure 2.2). La seconde étape de l'approche consiste à estimer le taux d'argile textural à partir du spectre de sol estimé $\hat{X}_{sol}$ en utilisant une FST *f* calibrée sur pixels de sol nu [AC2008.3 ; AC2012.2] (Etape 2, Figure 2.2).

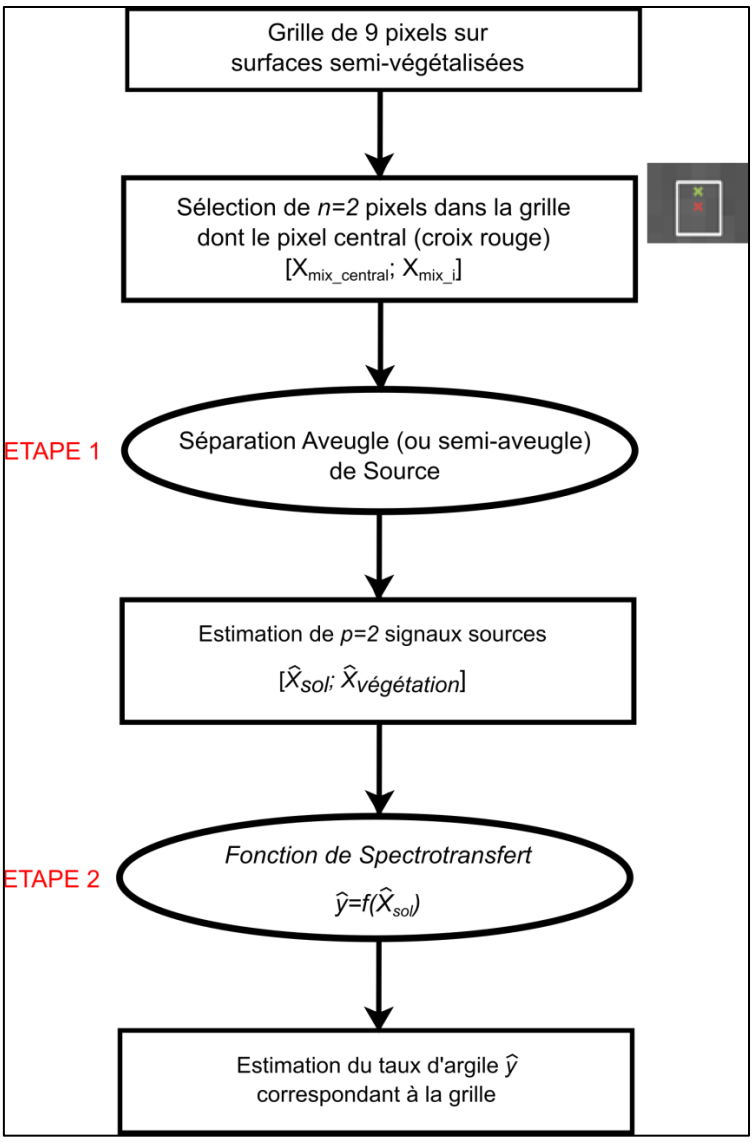


Figure 2.2 : Schéma de l'approche de « Double Séparation de Sources » (DSS).

Les premiers travaux ont donné lieu à des résultats encourageants [AC2011.1], mais leurs portées étaient faibles de par certains de nos choix. En effet l'approche a été initialement testée :

- sur un trop faible nombre de sites (seulement 10 sites), ne permettant pas de considérer l'ensemble de la variabilité pédologique et de la variabilité de couverture végétale du site d'étude.
- en réalisant l'étape d'estimation du spectre Vis-PIR de sol $\hat{X}_{sol}$, à travers un algorithme appartenant à la famille de l'Analyse en Composantes Indépendantes (Independent Component Analysis - ICA, Common, 1994). Or la résolution de problèmes de Séparation Aveugle de Sources par l'ICA est basée sur des hypothèses non vérifiées dans notre contexte. En effet l'ICA est basée sur les hypothèses suivantes (Common, 1994) : *1)* les signaux sources $\hat{X}_{sol}$ à extraire à partir des *n* spectres de mélanges X_{mix} sont supposés être statistiquement indépendants entre eux et *2)* au plus un des signaux sources $\hat{X}_{sol}$ peut suivre une distribution normale (gaussienne). Or les spectres de réflectance de sol et de végétation sont très corrélés entre eux, violant l'hypothèse d'indépendance statistique entre eux. De plus il est fréquent que les spectres de réflectance de sol et de végétation suivent des distributions normales.

Par la suite, des améliorations ont été apportées à cette approche en prenant en compte à la fois le contexte de l'étude et la connaissance a priori que nous en avons [AC2016.2]. Deux améliorations ont porté sur l'étape d'estimation du spectre Vis-PIR de sol $\hat{X}_{sol}$ (Etape 1, Figure 2.2) :

- L'utilisation d'un algorithme appartenant à la famille de la Factorisation en Matrice Non-Negatives (Non-negative Matrix Factorisation – NMF, Lee and Seung, 2001) et reposant sur des hypothèses admissibles vis à vis de notre contexte en lieu et place de l'algorithme ICA. Cet algorithme NMF suppose la positivité des sources $\hat{X}_{sol}$ (la réflectance est bien une grandeur physique positive) et de leurs abondances.
- Une initialisation de l'algorithme NMF à travers l'utilisation de spectres de terrain représentatifs des matériaux recherchés (un spectre de sol et un spectre de feuille de vigne verte) qui remplace l'initialisation aléatoire standard et permet une meilleure convergence vers la solution.

De plus, une amélioration a également porté sur l'étape d'estimation de la propriété de sol recherchée, à partir du spectre de sol estimé $\hat{X}_{sol}$ (Etape 2, Figure 2.2) :

- Le biais des estimations d'argile obtenues à partir des spectres de sol estimés $\hat{X}_{sol}$ a été corrigé, en aval de la fonction de spectrotransfert.

Finalement cette approche a été testée sur un nombre plus important de sites tests (54 sites), permettant ainsi de considérer une plus grande variabilité en terme pédologique et végétale [AC2016.2].

Résultat Marquant :

L'approche décrite ci-dessus et appliquée aux pixels semi-végétalisés dont l'indice spectral NDVI est inférieur à 0.55, nous a permis de spatialiser environ 63 % de la surface de la zone d'étude avec un *SEP* de 49 g/kg (Figure 2.3b, [AC2016.2]) (contre 4% de la surface avec la fonction de spectrotransfert classique *f* applicable uniquement sur surfaces de sol nu). Cette

performance d'estimation de taux d'argile par approche DSS peut paraitre modeste, mais est néanmoins presque équivalente à la performance de la FST appliquée sur sol nu (*SEP* de 44.5 g/kg) (Figure 2.3a, [AC2008.3] et [AC2012.2]). Finalement, la structure spatiale de l'argile texturale prédite par DSS sur surfaces semi-végétalisées dont l'indice spectral NDVI est inférieur à 0.55, est cohérente vis à vis de la structure spatiale de l'argile texturale prédite sur sol nu.

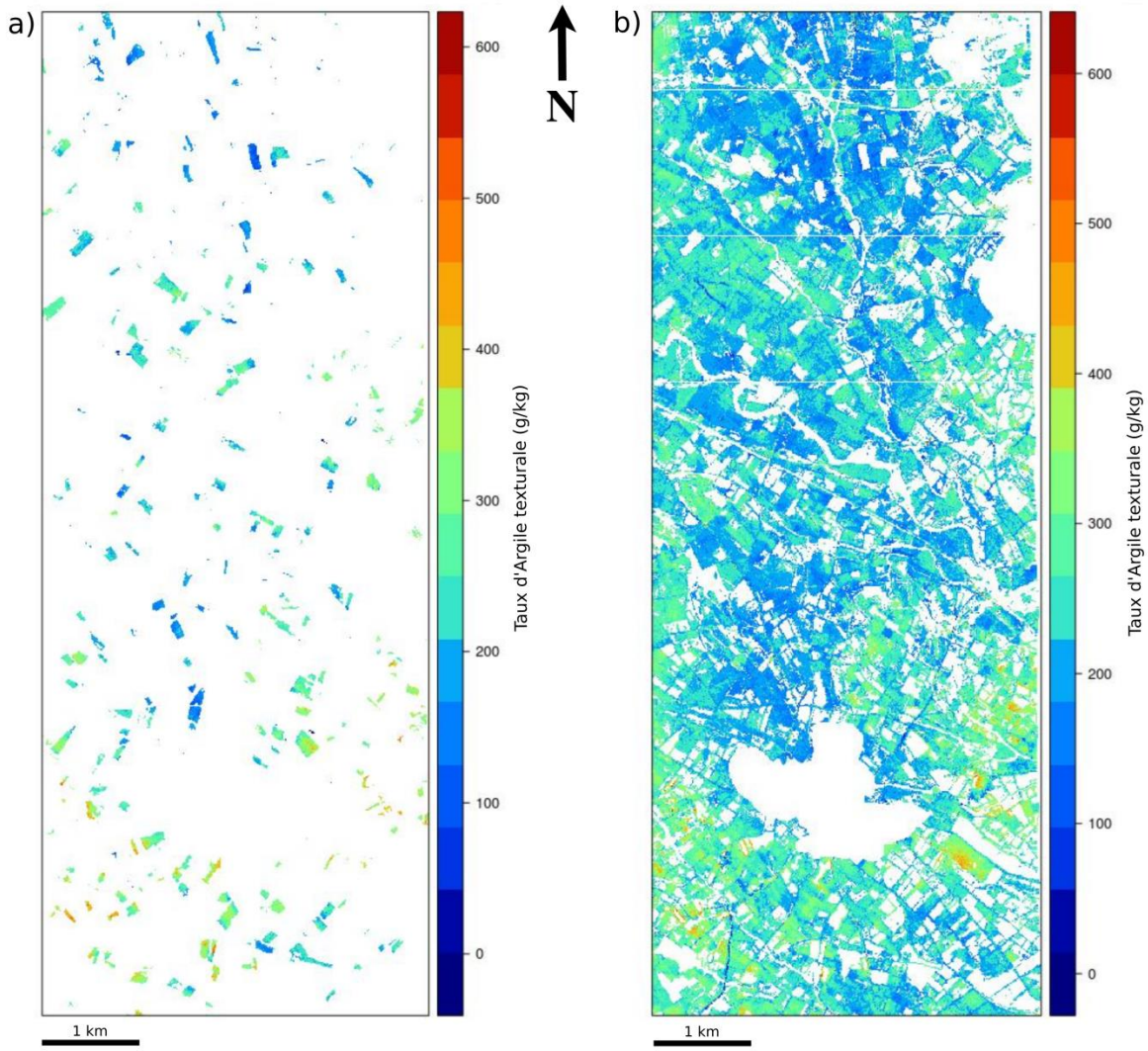


Figure 2.3 : Carte de taux d'argile estimés *a)* sur les surfaces de sol nu par application de la FST mise en place dans [AC2008.3 ; AC2012.2] (soit 4% de la zone d'étude) *b)* sur les surfaces semi-végétalisées dont le NDVI est inférieur à 0.55 par application de la DSS (soit 63% de la zone d'étude) [AC2016.2].

2.4 INCERTITUDE ASSOCIEE AUX ESTIMATIONS

Bien que ces spatialisations (Figures 2.1 et 2.3) offrent une vision synoptique des propriétés physico-chimiques des surfaces imagées, aucun indice de qualité ne permet de juger la qualité de ces estimations spatialisées. Sur le bassin versant du Lebna, 129 pixels ont été utilisés pour calibrer des FST, et les FST ont alors été appliquées à plus de 4,9 millions de

pixels sur sol nu (Figure 2.1). Et seuls les indices de qualités (coefficients de détermination de calibration R^2_{cal} et l'Erreur Standard de Calibration ou Standard Error of Calibration – *SEC*) des FST servent de « garant » à la qualité de ces spatialisations.

J'ai ainsi cherché à associer à chaque taux d'argile texturale estimé sur le bassin versant du Lebna à partir de données hyperspectrales AISA-DUAL aéroportées (Figure 2.1, [C2012.1]), une carte d'incertitude. L'incertitude se définit comme étant « le paramètre associé au résultat de l'action de mesurer qui caractérise la dispersion des valeurs qui pourraient raisonnablement être attribuées à l'action de mesurer ». Chaque estimation est issue d'une chaine d'opérations, entrainant chacune une source d'incertitude :

- L'incertitude des spectres, due à une éventuelle répétition d'acquisition instrumentale (I_d) et au positionnement spatial de l'acquisition de ces spectres (I_s). Cette incertitude des spectres est supposée identique entre les spectres de calibration et de validation et l'incertitude I_d est supposée négligeable dans notre étude, en comparaison de l'incertitude I_s.
- L'incertitude des analyses physico-chimiques de référence (I_y). Cette incertitude I_y est supposée négligeable dans notre étude.
- L'incertitude de la FST (I_m) est causée par deux facteurs principaux : l'incertitude I_c intègre le choix des données de calibration lors de la construction du modèle et l'incertitude I_d intègre le choix de la dimension du modèle (e.g., nombre de valeurs latentes choisies dans un modèle PLSR).

Plusieurs calculs d'incertitude ont été proposés et analysés au cours de ces travaux [AC2015-1] :

- Une procédure de bootstrap a été utilisée pour développer *n* fonctions de spectrotransfert *f*, utilisant le modèle PLSR, et donnant lieu à *n* estimations d'argile $\hat{y}$ pour chaque spectre Vis-PIR étudié. Chaque FST *f* diffère par la sélection de la base de données de calibration (tirage avec remise dans l'ensemble des 129 couples [X,y] disponibles sur la zone d'étude). La variance des *n* estimations $\hat{y}$ associées à un pixel est supposée représenter la somme de toutes les incertitudes et est considérée comme la *variance « vraie »* dans notre étude $var(\hat{y})$.
- Deux *distances spectrales* entre l'espace des spectres de calibration (utilisés pour construire les FST *f*) et le spectre Vis-PIR pour lequel le taux d'argile et l'incertitude associée sont recherchés, ont été calculées : la distance de Mahalanobis, utilisant l'espace ACP (Analyse en Composantes Principales) des spectres de calibration, et le Leverage, utilisant l'espace du modèle multivarié PLS des spectres de calibration. Ces calculs d'incertitudes représentent l'incertitude de la FST (I_m).
- Une modélisation de la variance des estimations, mise en place par Fernandez-Ahumada et al. (2012), a été utilisée et se décompose en trois termes, dont les plus importants sont :
    - Un terme *T1* reflète l'incertitude liée au positionnement géographique du pixel Vis-PIR étudié au sein de l'image et de l'échantillon de sol qui lui est associé lors de l'utilisation de couples $[X_{cal},y_{cal}]$. Ce calcul d'incertitude représente l'incertitude liée au positionnement spatial de l'acquisition de ces spectres (I_s).
    - Un terme *T2* reflète l'incertitude liée à la variance des b-coefficients des *N* FST *f*. Ce calcul d'incertitude représente l'incertitude liée à la construction de la FST (I_m).
    -

Résultats Marquants :

Cette étude nous a permis d'identifier les sources d'incertitudes et de les hiérarchiser [AC2015-1]. En particulier, la construction des modèles d'estimation PLSR (par exemple le choix de la base de données de calibration) s'avère une cause d'incertitude plus importante que le positionnement géographique des pixels Vis-PIR étudiés au sein de l'image et des échantillons de sol utilisés pour la calibration. Les fortes incertitudes liées au positionnement géographique (terme *T1*) sont localisées aux limites de parcelles et dans les zones urbaines (pixels de non-sol qui auraient dû être masqués) (Figure 2.4c). Les fortes incertitudes liées à la construction des fonctions PLSR (terme *T2*) sont localisées au sein de certaines parcelles et dans les zones urbaines (pixels de non-sol qui auraient dû être masqués) (Figure 2.4d). Enfin la spatialisation de la variance « vraie » permet également de localiser des pixels de non-sol qui devraient être masqués (zones urbaines) (Figure 2.4b).

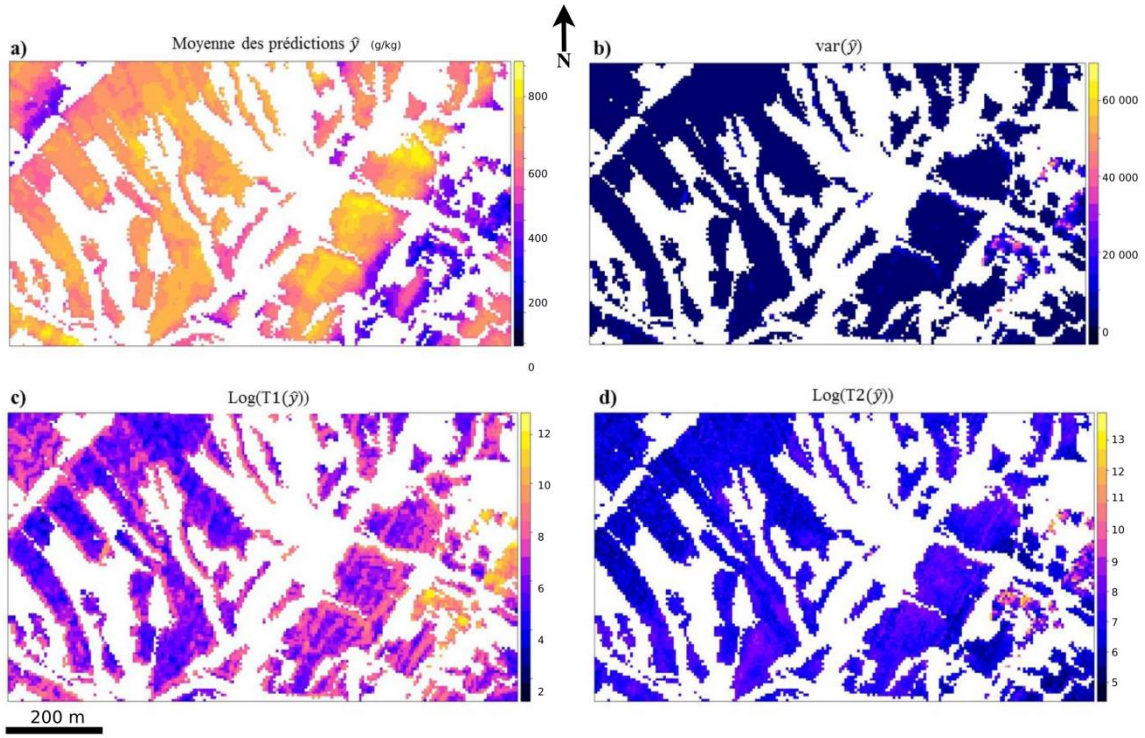


Figure 2.4 : Cartes centrées sur le bassin versant élémentaire de Kamech (sous-bassin du Lebna, Tunisie) a) des valeurs estimées d'argile $\hat{y}$, b) de la "vraie" variance $var(\hat{y})$, c) du terme *T1*, et d) du terme *T2* [AC2015-1]. Les zones en blanc correspondent à des pixels préalablement masqués.

Cette étude a montré le bénéfice des cartes d'incertitude liée aux prédictions de propriétés de sol, pour *i)* effectuer une meilleure identification des surfaces de sol nu en amont de l'étape de spatialisation des taux d'argile (diminution d'erreur dans le masque de non-sol, jusque-là simplement basé sur des indices de végétation verte et sèche, e.g., Madeira Netto et al., 2006), *ii)* mieux définir l'échantillonnage de sol de surface et donc la base de données de calibration (e.g., augmenter la densité de l'échantillonnage sur les surfaces associées à de fortes incertitudes) et iii) mieux caractériser la qualité des spatialisation de propriétés de sol.

2.5 SYNTHESE

A travers les travaux présentés dans ce chapitre, j'ai contribué à démontrer le potentiel de l'imagerie Visible-Proche Infra-Rouge hyperspectrale pour la cartographie de propriétés pérennes des sols, à des résolutions spatiales fines et sur de grandes étendues, dans un contexte d'agrosystèmes pluviaux Méditerranéens. J'ai également contribué à développer des approches raisonnées d'estimation de propriétés pérennes des sols par imagerie Vis-PIR en contexte semi-végétalisé, et d'estimation des incertitudes associées aux estimations de propriétés pérennes des sols. Ces développements ont bénéficié de grand nombre de travaux antérieurs en spectroscopie Vis-PIR de laboratoire et ont été réalisés dans un contexte de forte demande de données physico-chimiques à haute densité spatiale. L'ensemble de ces travaux permet d'envisager ainsi l'imagerie Vis-PIR hyperspectrale comme un outil de prédilection pour la spatialisation de propriétés pérennes des sols de surface.

III. VERS UNE MISE EN PLACE OPERATIONNELLE DE CES SPATIALISATIONS

3.1 INTRODUCTION

A l'heure actuelle, les FST mises en place à partir d'images hyperspectrales Vis-PIR pour la cartographie de propriétés pérennes de sol sont construites en utilisant des spectres de pixels de l'image (variables descriptives *X*) associés aux propriétés physico-chimiques (variables prédictives *y*) mesurées sur des échantillons recueillis sur le terrain sur ces mêmes pixels. Et les collectes d'échantillons de terrain et d'acquisition Vis-PIR ne sont jamais espacées de plus de 8 ans afin de rester dans des conditions similaires. Bien que les potentialités de ces données Vis-PIR couplées à des techniques telles que présentées en section 2.2, aient été démontrées, plusieurs facteurs limitent la démocratisation de ces techniques et l'expansion de ces cartographies. Ces facteurs sont les coûts (en argent et en temps) de collecte d'échantillons de sol de surface sur les sites imagés, de mesures physico-chimiques de ces mêmes échantillons, et d'acquisition de données hyperspectrales Vis-PIR. Pour pallier ces facteurs limitants, trois approches ont été développées :

- L'approche (1) consiste à calibrer les FST en utilisant des mesures physico-chimiques *y* de sol réalisées dans le cadre d'anciens programmes de cartographie. Ces données pédologiques anciennes sont le produit de plusieurs décennies de campagnes pédologiques dans le monde. De telles données sol anciennes ont déjà été utilisées de manière opérationnelle pour des travaux de cartographie de sol à l'échelle régionale (Vaysse and Lagacherie, 2015), continentale (Viscarra Rossel, 2015) et globale (Hengl et al, 2015). Dans notre contexte, ces données pédologiques anciennes, localisées sur ou proche de l'emprise des images Vis-PIR considérées, pourraient être utilisées en substitution des mesures physico-chimiques réalisées sur des échantillons prélevés spécialement sur le site imagé.
- L'approche (2) consiste à calibrer les FST en utilisant des bases de données spectrales Vis-PIR acquises en laboratoire couplées à des mesures physico-chimiques [X_{lab} , *y*], puis à appliquer ces mêmes FST sur des données spatiales X_{im} (e.g., Schwanghart and Jarmer, 2011 ; Jarmer, 2013). Des bases de données [X_{lab} , *y*] locales, nationales (base de données RMQS en France, Arrouays et al., 2002, et base de données du Danemark, Knadel et al., 2012), continentale (base de données Européenne LUCAS ; Toth et al., 2013) et mondiale (Global Spectral Library, [AC2016.1]) pourraient ainsi se substituer aux campagnes de collecte d'échantillons et de mesures physico-chimiques réalisées jusqu'à présent sur les sites imagés.
- L'approche (3) consiste à étudier le potentiel des données Vis-PIR de capteurs satellitaires actuellement en préparation. Les capteurs hyperspectraux Italien PRISMA (PRecursore IperSpettrale della Missione Applicativa) (Stuffler et al., 2007) et Allemand EnMAP (Environmental Mapping and Analysis Program, http://www.enmap.org/) (Lopinto and Ananasso, 2013 ; Giampaolo et al., 2008) sont planifiés pour un lancement en 2018. Par ailleurs, trois autres capteurs hyperspectraux sont actuellement à l'étude : le capteur Américain HyspIRI (Hyperspectral Infrared Imager, hyspiri.jpl.nasa.gov), le capteur Italo-Israélien SHALOM (Spaceborne Hyperspectral Applicative Land and Ocean Mission) (Bussoletti, 2012 ; Ben-Dor et al., 2014) et le capteur français HYPXIM (HYPerspectral X

Imagery) (Carrere et al., 2013 ; Briottet et al., 2013). De plus, les capteurs multispectraux WV-3 (Kruse and Perry, 2013) et SENTINEL 2 (ESA, 2011 ; Drusch et al., 2012) sont désormais disponibles.

Dans la suite de ce chapitre, des travaux relatifs aux trois approches précitées sont successivement exposés.

3.2 UTILISATION DE BASE DE DONNEES PEDOLOGIQUES ANCIENNES

Les surfaces continentales sont étudiées depuis des décennies via des campagnes de terrain donnant lieues à des prélèvements, des carottages, ainsi que des analyses physico-chimiques et minéralogiques associées. Au-delà de l'utilisation de ces données de terrain pour une cartographie pédologique et/ou géologique des sites étudiés, ces données sont archivées mais rarement ré-exploitées.

L'approche (1) a consisté à utiliser ces données sol anciennes localisées sur ou proche de l'emprise des images Vis-PIR considérées, pour construire des FST applicables sur ces mêmes images Vis-PIR [AC2016.3]. Ces données sol anciennes comportent deux inconvénients qu'il s'agit de prendre en compte dans cette approche. Tout d'abord, les sites anciennement échantillonnés (et pour lesquels une mesure physico-chimique y est disponible) ne sont localisés que très rarement à la fois sur l'emprise des images Vis-PIR considérées et sur des surfaces de sol nu au moment de l'acquisition de ces mêmes images Vis-PIR. Donc il parait peu probable de pouvoir constituer une base de données de calibration [X_{im} , y] constituée de spectres de pixels de l'image (X_{im}) associés aux propriétés physico-chimiques (y) mesurées sur des échantillons recueillis sur le terrain lors d'anciennes campagnes sur ces mêmes pixels. D'autre part, ces bases de données anciennes de mesures y sont bien souvent affectées par des biais générés par des différences de date de collecte, de protocoles d'échantillonnage ou encore de méthodes de mesures physico-chimiques (e.g., Baume et al., 2011).

L'approche mise au point pour utiliser des données sol anciennes y (en substitution des traditionnelles mesures physico-chimiques réalisées sur des échantillons prélevés spécialement sur le site imagé), consiste en 3 étapes successives :

- Un indice spectral est utilisé pour estimer en valeur relative la propriété de sol d'intérêt $\hat{y}_1$
- Les valeurs estimées $\hat{y}_1$ de la propriété de sol d'intérêt sont alors normalisées afin d'obtenir des valeurs estimées $\hat{y}_2$. Pour cela, une transformation Box-Cox est appliquée aux valeurs estimées $\hat{y}_1$ afin de transformer leur distribution en distribution normale (Legendre and Legendre, 1998).
- Les valeurs estimées $\hat{y}_2$ sont alors centrées et réduites en utilisant les propriétés physico-chimiques d'une base de données sol ancienne (y_{cal}), afin d'obtenir des valeurs estimées $\hat{y}_3$ ayant la même moyenne et le même écart type que la base de données sol ancienne utilisée.

Cette approche a été testée pour l'estimation du taux d'argile texturale à partir des données

hyperspectrales AISA-DUAL aéroportées acquises sur le bassin versant du Lebna (rectangle rose, Figure 3.1), une base de données « de référence » (points rouges, Figure 3.1) et de 3 bases de données sol que nous considérons comme « anciennes » (points verts, jaunes et bleus, Figure 3.1). Ces 3 bases de données sol dites « anciennes » comprennent des analyses physico-chimiques (dont le taux d'argile texturale) réalisées sur le premier horizon de sol lors d'étude de profils pédologiques, et diffèrent par leur date de campagne de terrain, le protocole d'échantillonnage, la méthode d'analyse physico-chimique et la couverture géographique couverte. L'indice spectral utilisé est l'indice d'argile proposé par Levin et al. (2007) utilisant 3 bandes spectrales à $\lambda_1 = 2133$ nm, $\lambda_2 = 2225$ nm and $\lambda_3 = 2209$ nm.

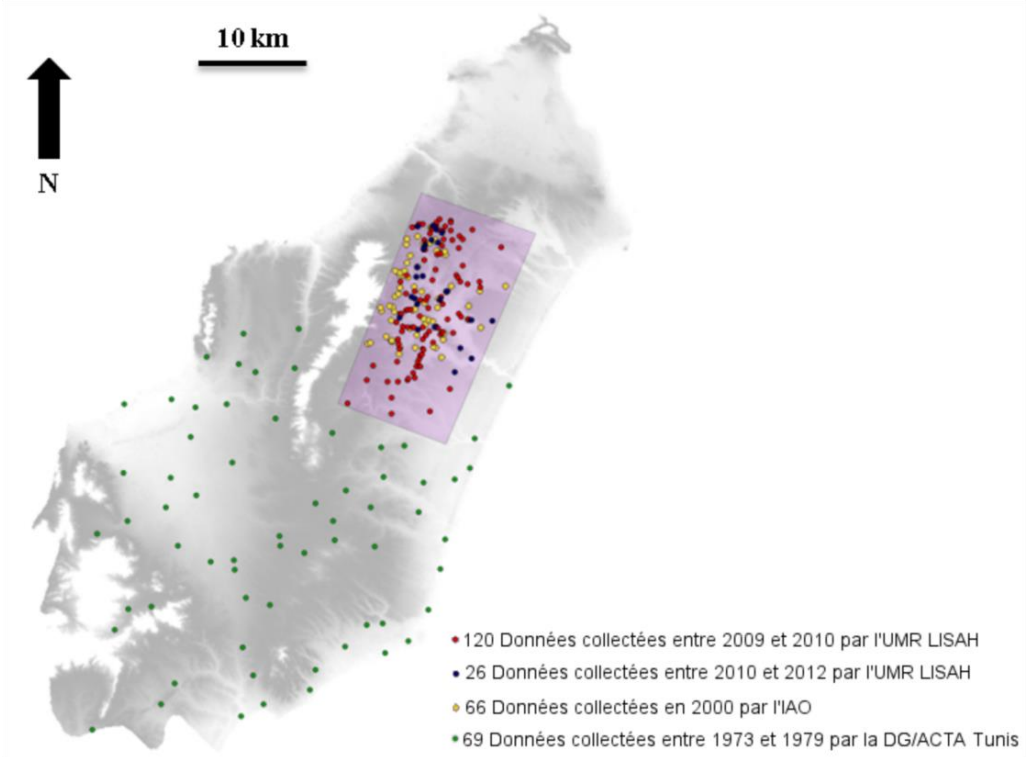


Figure 3.1 : Localisation des 3 bases de données « anciennes » issues de descriptions de profils de sol (seul le premier horizon est utilisé et est assimilé à la donnée de surface), de la base de données de référence (en rouge) et de l'emprise de la donnée hyperspectrale AISA-DUAL (rectangle rose), sur un Modèle Numérique de Terrain.

Résultat Marquant :

L'approche décrite ci-dessus nous a permis de montrer que les estimations d'argile texturale issues de l'indice spectral de Levin et al. (2007) sans calibration (valeurs relatives de prédiction), ont une faible précision (faible R^2, associé à un fort *biais* et *SEP*) [AC2016.3]. En effet la performance de prédiction (autrement dit la précision des valeurs relatives des estimations) est dépendante du biais contenu dans les bases de données anciennes utilisées.

La correction de la distribution et de la moyenne des valeurs d'argile estimées, en utilisant les bases de données anciennes, permet une amélioration de la précision de prédiction (augmentation du R^2 et diminution du *biais* et *SEP*). Et il est préférable d'utiliser un faible nombre de données anciennes localisées sur l'emprise de l'image hyperspectrale, plutôt qu'un grand nombre de données anciennes dispersées en dehors de cette emprise.

Enfin, quel que soit la base de données anciennes utilisée pour corriger la distribution et de la moyenne des valeurs d'argile estimées $\hat{y}_2$, le résultat de spatialisation offre une mise en évidence des structures pédologiques (alternance de bandes de grès et de marne). Autrement dit, les cartes d'estimations $\hat{y}_3$ constituent de nouvelles sources d'information spatiale pédologique.

3.3 UTILISATION DE BASE DE DONNEES SPECTRALES DE SOL DE LABORATOIRE

Pour un nombre d'échantillons de sol équivalent, les FST calées sur des spectres X_{im} extraits d'une image Vis-PIR hyperspectrale sont moins performantes que les FST calées sur des spectres X_{lab} acquis en laboratoire dans des conditions expérimentales contrôlées. De plus, des bases de données [X_{lab} , y] nationales (RMQS, Arrouays et al., 2002 ; Danemark, Knadel et al., 2012), continentales (Européenne LUCAS, Toth et al., 2013) et mondiales (Global Spectral Library, [AC2016.1]) ont récemment vu les jour, venant compléter les bases de données locales.

Afin de bénéficier à la fois de la haute qualité des mesures spectrales de laboratoire X_{lab} et de la disponibilité de large bases de données (nationales, continentales et mondiales), nous avons mis en place une série de travaux (approche (2)) dans l'optique d'utiliser des bases de données spectrales Vis-PIR acquises en laboratoire**,** pour prédire les propriétés pérennes de sol à partir d'imagerie Vis-PIR hyperspectrale, en suivant un processus en trois étapes successives :

Etape 1 (Figure 3.2, Flèche Verte) : Des données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude A (*BD_Lab_A*), sont utilisées pour caler une FST, qui est ensuite appliquée à des données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude B (*BD_Lab_B*), où B est inclus dans A et où la superficie de B est bien inférieure à la superficie de A [AC2014.1]. Ce travail est décrit en section 3.3.1.

Etape 2 (Figure 3.2, Flèche bleue) : Des données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude B (*BD_Lab_B*), sont utilisées pour caler une FST, qui est ensuite appliquée à des données hyperspectrales aéroportées X_{im} acquises sur cette même surface d'étude B (*BD_Hyper_B*) [AC20XX.2]. Ce travail est décrit en section 3.3.2.

Etape 3 (Figure 3.2, Flèche violette) : Des données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude A (*BD_Lab_A*), sont utilisées pour caler une FST, qui est ensuite appliquée à des données hyperspectrales aéroportées X_{im} acquises sur une surface d'étude B (*BD_Hyper_B*), où B est inclus dans A et où la superficie de B est bien inférieure à la superficie de A. Ces travaux ont débuté fin 2015 via l'encadrement d'un stage de Master 2 et ne sont pas encore aboutis. Par conséquent, ces travaux ne sont pas décrits dans ce mémoire.

De façon générale, l'utilisation de données spectrales acquises dans des conditions données, pour prédire des propriétés physico-chimiques y à partir de données spectrales acquises dans des conditions différentes, suppose la mise en place d'un transfert d'étalonnage. La base de données (BD) spectrales utilisée pour étalonner une FST est appelée "BD maître" (dans mon cas, à terme, la BD de laboratoire). Et la BD sur laquelle est appliquée cette FST (dans mon cas, à terme, l'image hyperspectrale) est appelée "BD

esclave". Concrètement, il s'est agi, au cours de ces 3 étapes précitées, de compenser l'effet d'une ou plusieurs grandeurs d'influence G dont la variation *dg* (entre la BD maître et la BD esclave) provoque soit une variation de spectre *dx* entre la BD maitre et la BD esclave (cas d'une différence de type de spectroradiomètre, de condition atmosphérique, d'angle de visé, de rugosité de cibles, d'intensité de la source lumineuse, d'hétérogénéité des échantillons, de dimension des supports de mesure…) soit un manque de représentativité de la BD maitre vis-à-vis de la BD esclave (cas d'une différence de contexte pédologique).

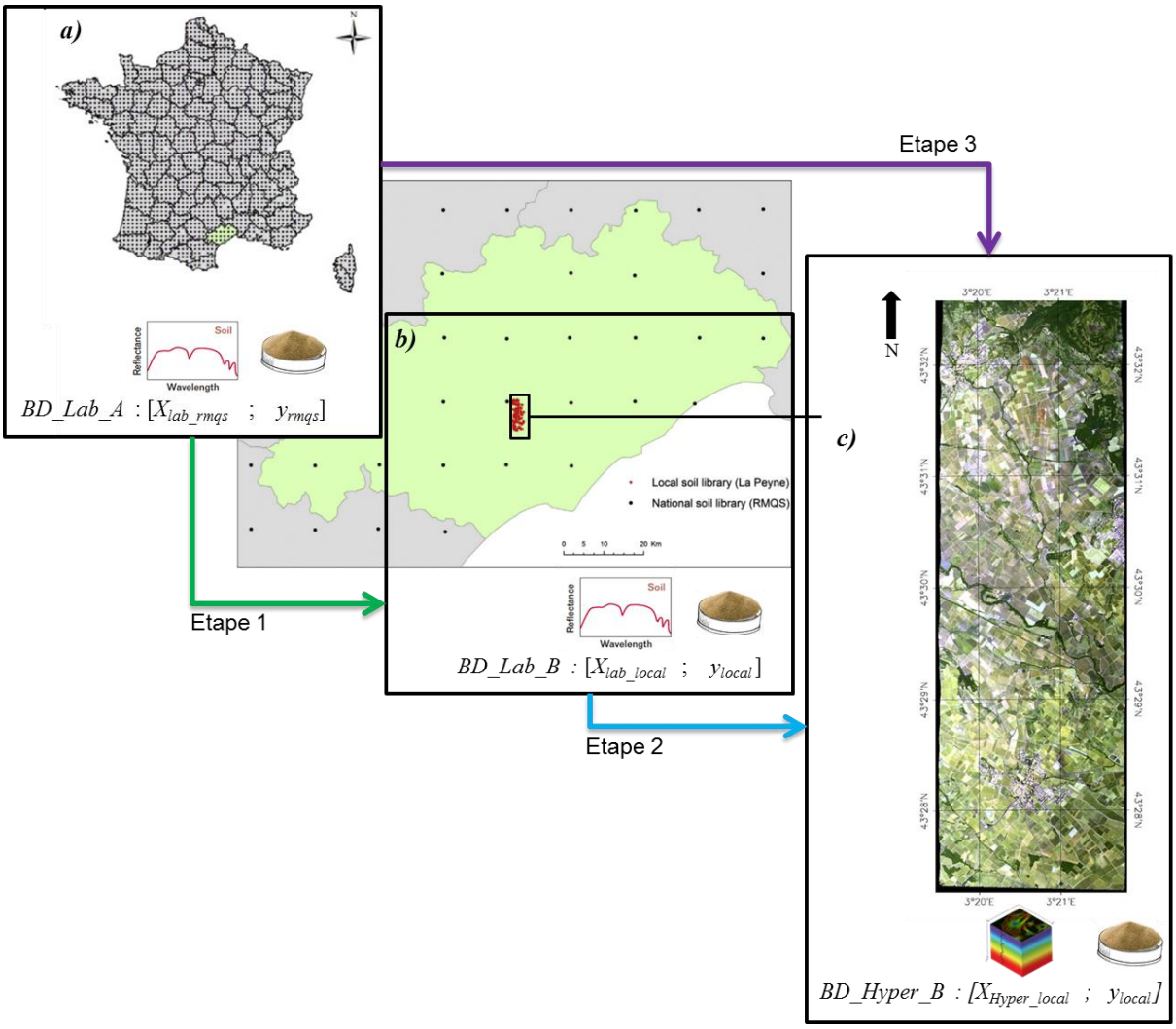


Figure 3.2 : Représentation schématique de la démarche adoptée utilisant : *a)* la base de données labo RMQS, notée *BD_Lab_A*, *b)* la base de données labo acquises sur le bassin versant de La Peyne notée *BD_Lab_B* (points rouges), *c)* l'image hyperspectrale HYMAP pour laquelle correspond la base de données *BD_Hyper_B.*

3.3.1. Transfert de FST calibrée sur une Base de données Labo Nationale vers une base de données Labo Locale

En spectroscopie de laboratoire, la démarche classique consiste à construire une FST à partir de données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude A, pour l'appliquer sur des données spectrales labo acquis sur des échantillons

prélevés sur une même surface d'étude A (e.g., Viscarra Rossel et al., 2006). Et les conditions d'acquisition des spectres labo X_{lab} et les analyses physico-chimiques *y* sont réalisées de manière identique (même spectroradiomètre, angle de visée, intensité de la source lumineuse, préparation des échantillons …).

Nous avons cherché à construire une FST à partir de données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface d'étude A ("BD maître" *BD_Lab_A*), pour l'appliquer sur des données spectrales labo acquises sur des échantillons prélevés sur une surface d'étude B ("BD esclave" *BD_Lab_B*). Dans notre cas, B est inclus dans A et la superficie de B est bien inférieure à la superficie de A (Figure 3.2, Flèche Verte) [AC2014.1]. Cette étude a été réalisée en utilisant *i)* une base de données sol collectées dans le cadre du Réseau de Mesures de Qualité des Sols (RMQS) comprenant 2126 sites de 20 x 20 m répartis sur la France métropolitaine en suivant un maillage régulier de 16 km de côté (*BD_Lab_A*) (Figure 3.2a), et *ii)* une base de données sol collectées sur le bassin versant de la Peyne comprenant 144 échantillons collectés sur des sites de 10 x 10 m (*BD_Lab_B*) (Figure 3.2b). Pour ces deux bases de données sol, nous disposons de spectres Vis-PIR de laboratoire X_{lab} et d'analyses physico-chimiques (argile texturale, $CaCO_3$, CEC et fer)[2] ([X_{lab_rmqs} ; y_{lab_rmqs}] et [X_{lab_local} ; y_{lab_local}]) réalisés dans des conditions identiques.

Dans ce cas, deux principales grandeurs d'influence peuvent perturber l'application d'une FST étalonnée sur la "BD maître" *BD_Lab_A,* sur la "BD esclave" *BD_Lab_B* : le protocole d'échantillonnage de terrain et le contexte pédologique couvert par les deux BD. Dans notre étude, nous avons considéré l'impact du protocole d'échantillonnage de terrain comme négligeable, en comparaison de l'impact du contexte pédologique. La méthode de transfert de modèle d'étalonnage sélectionnée pour ce travail est une méthode de mise à jour du modèle, dite *spiking* ou *Update Model* (Brown, 2007). Cette méthode consiste à utiliser conjointement les spectres maîtres (donc des spectres X_{lab} de la *BD_Lab_A*) et *Nc* spectres esclaves (donc des spectres X_{lab} de la *BD_Lab_B*, dits « standards ») dans la construction de la FST. Deux modèles de prédictions ont été testés pour la construction des FST : 1) le modèle PLSR, classiquement utilisé en spectroscopie des sols (e.g., AC2012.2), et 2) le modèle Fast Fourier Transform Local Weighted (FFT-LW, Gogé et al., 2012) appartenant à la famille des régressions locales et basé sur la construction de calibrations spécifiques à chaque échantillon à prédire.

Chacune de ces FST a été construit à partir i) de m_b échantillons de *BD_Lab_B* (m_b*=10,..,94*), ii) de la totalité des échantillons de *BD_Lab_A* (m_{a_t} *= 2126*) associés à m_b échantillons de *BD_Lab_B* (m_b *=0,10,..,94*) et iii) de m_{a_e} échantillons de *BD_Lab_A* (m_{a_e} = *300*) associés à m_b échantillons de *BD_Lab_B* (*m=0,10,..,94*). Et la sélection des m_{a_e} échantillons de *BD_Lab_A* s'effectue en utilisant l'algorithme de Kennard and Stone (Kennard and Stone, 1969) qui recherche les spectres de la *BD_Lab_A* les plus proches des spectres de *BD_Lab_B* selon une distance Euclidienne. Cette sélection a pour objectif de calibrer les FST sur les spectres « les plus proches » spectralement des spectres de *BD_Lab_B*.

[2] Les travaux exposés ici ont été testés sur 5 propriétés physico-chimiques (argile texturale, $CaCO_3$, CEC, fer et carbone organique) (cf AC2014-1). Mais le carbone organique n'étant jamais prédictible, quel que soit le modèle et/ou la base données, je préfère exclure cette propriété de la présentation de ce travail.

Résultat Marquant :

Cette étude, malgré la richesse des bases de données et l'exploration de deux modèles de régression distincts, n'a pas permis de dégager une tendance commune aux 4 propriétés pérennes testées [AC2014.1].

Ainsi, concernant le fer et la CEC, les meilleures performances sont obtenues avec un modèle PLSR calé sur des échantillons de la *BD_Lab_B* (même en utilisant un faible nombre d'échantillons). Autrement dit, l'utilisation de la BD nationale ne suffit pas pour estimer ces propriétés à l'échelle locale. Et l'utilisation de la BD nationale n'apporte pas de plus-value avec ou sans spiking d'échantillons de la *BD_Lab_B* dite « locale ».

Concernant l'argile, les meilleures performances sont obtenues avec le modèle local FFT-LW calé sur la *BD_Lab_A* Nationale associée à un minimum de 20 échantillons de la *BD_Lab_B*. Autrement dit l'utilisation de la BD nationale associée à un minimum de 20 échantillons de la BD locale permet l'estimation de cette propriété à l'échelle locale, lorsque le modèle local FFT-LW est utilisé.

Enfin, concernant le $CaCO_3$, de très bonnes performances sont obtenues avec le modèle local FFT-LW calé sur la *BD_Lab_A* Nationale seule. L'ajout d'échantillons de la *BD_Lab_B* à ce modèle local n'améliore pas les performances de prédiction de $CaCO_3$. Autrement dit l'utilisation de la BD nationale suffit à estimer le $CaCO_3$ sur le terrain d'étude local et aucune acquisition supplémentaire de données « locales » n'est nécessaire (Figure 3.3).

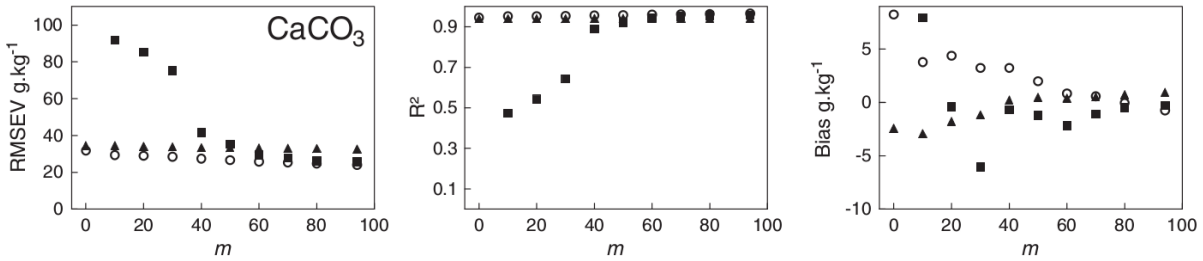


Figure 3.3 : *RMSEV* (gauche), R^2 (centre) et *biais* (droite) de prédiction obtenus pour le $CaCO_3$, en utilisant la PLSR calibrée sur m_b échantillons de *BD_Lab_B* (carrés), la PLSR calibrée sur $m_{a\_e} = 300$ échantillons de la *BD_Lab_A* associés à m_b échantillons de *BD_Lab_B* (triangle), et la FFT-LW calibrée sur $m_{a\_e} = 300$ échantillons de la *BD_Lab_A* associés à m_b échantillons de *BD_Lab_B* (cercle).

3.3.2. Transfert de FST calibrée sur une Base de données Labo Locale vers une image hyperspectrale locale

En imagerie hyperspectrale Vis-PIR, la démarche classique consiste à construire une FST à partir de spectres X_{im} extraits d'une image Vis-PIR hyperspectrale acquise sur une surface d'étude B, pour l'appliquer à des spectres extraits de cette même image (e.g., Selige et al., 2006 ; AC2012.2). Nous avons cherché à construire une FST à partir de données spectrales labo X_{lab} acquis sur des échantillons prélevés sur une surface d'étude B (*BD_Lab_B*), pour l'appliquer sur des données hyperspectrales aéroportées X_{im} acquises sur cette même surface d'étude B (*BD_Hyper_B*) (Figure 3.2, Flèche bleue) [AC20XX.2]. Cette étude a été réalisée en utilisant *i)* une base de données sol collectées sur le bassin versant de la Peyne (France) en 2009 comprenant 144 échantillons collectés sur des sites de 10 x 10 m (Figure 3.2b), ii) des données hyperspectrales HyMap aéroportées acquises en 2003 sur ce bassin

versant de La Peyne (24 km²) (Figure 3.2c). La *BD_Lab_B* comprend ainsi 147 spectres de laboratoire et le taux d'argile textural associé [X_{lab} , y]. La *BD_Hyper_B* comprend 96 spectres HYMAP sur sol nu (correspondant aux pixels sur lesquels ont été collectés les échantillons de sol en 2009) et le taux d'argile textural associé [X_{im} , y].

Dans ce cas, les grandeurs d'influence empêchant l'application directe d'une FST construite à partir de *BD_Lab_B* sur la *BD_Hyper_B* sont : le type de capteur (spectromètre de laboratoire / capteur aéroporté), l'altitude d'acquisition de l'image (influençant principalement l'angle d'acquisition des photons) et l'état du sol (broyé et séché au laboratoire / sol brut sur le terrain). Chaque méthode de transfert testée dans notre étude traite ces trois facteurs d'influence simultanément, comme un seul "effet bloc". Et ces méthodes de transfert de modèle d'étalonnage sont couplées à un modèle de régression PLSR afin de prédire le taux d'argile texturale. Trois méthodes de transfert de FST ont été testées :

1) la méthode *Piecewise Direct Standardisation* (PDS, Bouveresse and Massart, 1996) est une standardisation optique consistant à enregistrer le spectre de *Nc* échantillons à partir de l'appareil maître (spectromètre de laboratoire) et l'appareil esclave (capteur aéroporté) (*Nc* échantillons dits « standards »), afin d'estimer le *dx* (variation des spectres entre X_{lab} et X_{im}) et de l'appliquer aux spectres maîtres X_{lab} afin de construire une FST adaptée aux spectres esclaves X_{im} (Fearn, 2001).

2) la méthode de mise à jour du modèle, aussi appelée *Spiking* ou *Update Model* (Brown, 2007) consiste à utiliser conjointement les spectres maîtres X_{lab} et *Nc* spectres esclaves X_{im} (*Nc* échantillons dits « standards ») dans la construction de la FST (méthode également utilisée dans les travaux décrits en section 3.3.1).

3) la méthode *Repfile* (Shenk and Westerhaus, 1991), consiste à utiliser conjointement les spectres maîtres X_{lab} (associés à la valeur d'argile vraie y) et *Nc* exemples de perturbations spectrales entre des spectres maîtres X_{lab} et esclaves X_{im} (associés à des valeurs d'argile y nulles) (*Nc* échantillons dits « standards »).

L'impact du nombre de « standards » *Nc* utilisés dans ces calibrations a été étudié, ainsi que l'impact d'une sélection de « standards » contrainte par la pédologie (autrement dit, en s'assurant ou non, que les *Nc* standards choisis sont uniformément répartis sur les types de sols présents dans la zone d'étude couverte par l'image hyperspectrale).

Résultat Marquant :

Dans notre cas d'étude, la FST « classique » (appelée *FST_Image*) construite à partir des données hyperspectrales HyMap X_{im} de la *BD_Hyper_B* et appliquée à des spectres de pixels de cette même image, offre des performances satisfaites ($R^2_{val} \simeq 0.72$; SEP $\simeq$ 42 g/kg) [AC2012.2]. Comme attendu, la FST (appelée *FST_Lab*) construite à partir des données spectrales labo X_{lab} acquis sur des échantillons prélevés sur la surface d'étude B (*BD_Lab_B*) et appliquée à des spectres X_{im} des données hyperspectrales HyMap, offre des performances faibles ($R^2_{val} \simeq 0.41$; SEP $\simeq$ 103 g/kg) [AC20XX.2].

Partant de ces résultats préliminaires, nos travaux ont permis de montrer le bénéfice de l'utilisation d'une base de données Labo (*BD_Lab_B*) couplée à une méthode de transfert, pour l'estimation d'argile sur les données hyperspectrales HyMap [AC20XX.2] :

- Les trois méthodes de transfert testées (Update, repfile et PDS) améliorent les performances de la *FST_Lab*, quel que soit le nombre de standards *Nc* utilisé.

- Quel que soit la méthode de transfert utilisée, plus le nombre de standards *Nc* pris en compte est important, plus la performance de prédiction s'améliore, et mieux vaut tenir compte de la pédologie dans la sélection de ces *Nc* standards.
- L'utilisation de seulement *Nc=10* standards sélectionnés en tenant compte de la pédologie associée à la *BD_Lab_B* dans un modèle de transfert Update ou Repfile, permet d'offrir des performances de prédiction très correctes ($R^2_{val} \simeq 0.65$; SEP $\simeq$ 48 g/kg).
- L'utilisation de seulement *Nc=35* standards sélectionnés en tenant compte de la pédologie associée à la *BD_Lab_B* dans un modèle de transfert Update, permet d'offrir des performances de prédiction supérieures à la FST « classique » (*FST_Image*) ($R^2_{val} \simeq 0.73$; SEP $\simeq$ 40 g/kg).

3.4 ANTICIPER L'ARRIVEE DES SATELLITES VIS-PIR

Aucun des capteurs satellitaires Vis-PIR en préparation (EnMap, PRISMA, SHALOM, SENTINEL 2…) n'est dédié à une seule thématique d'observation de la terre, et chacun est caractérisé par des configurations spectrales (largeur de bandes à mi-hauteur, nombre de bandes spectrales, et pas d'échantillonnage spectral), spatiales et des niveaux de bruit différents. Les différences notables entre des données hyperspectrales Vis-PIR aéroportées et satellitaires sont les suivantes : 1) altitude des capteurs donc différences d'atmosphère lors de l'acquisition des données, 2) résolution spatiale : entre 2 et 5 m en aéroporté, entre 8 et 60 m pour les futurs satellites hyperspectraux et 3) résolution spectrale : autour de 4 à 6 nm en aéroporté, entre 10 et 100 nm en satellite.

Face à l'arrivée imminente de nouveaux capteurs, des travaux ont été menés sur l'impact de la dégradation de résolution spatiale et spectrale sur l'étude de paramètres liés à la végétation (e.g., Rahman et al., 2003 ; Nijland et al., 2009 ; Schaaf et al., 2011 ; Thorp et al., 2013), au milieu urbain (Jensen and Cowen, 1999 ; Kruse et al., 2012 ; Roberts et al., 2012) à la température de sol de surface lors d'incendie (Lugassi et al., 2010) et à la qualité des eaux côtières (Giardino et al., 2014). A l'image de ces différents travaux, un travail d'analyse du potentiel de ces capteurs pour la spatialisation de propriétés pérennes de sol, a été réalisé à partir des données aéroportées hyperspectrales Vis-PIR AISA-DUAL acquises sur le bassin versant du Lebna (Tunisie) (Figure 1.2b).

3.4.1. Sensibilité à l'atmosphère

Dans un premier temps, la sensibilité à l'atmosphère de FST pour l'estimation d'argile texturale a été étudiée [AC2015.2]. Pour cela, des données Vis-PIR satellites ont été simulées, en gardant les mêmes résolutions spectrales et spatiales que les données aéroportées Vis-PIR AISA-DUAL. Les données AISA-DUAL acquises en luminance par avion ont été *1)* simulées en haut de l'atmosphère, en utilisant le modèle de transfert radiatif COMANCHE (Miesch et al., 2005), puis *2)* converties en données de réflectance corrigées des effets atmosphériques en utilisant le modèle de correction atmosphérique COCHISE (Miesch et al., 2005). Enfin, des FST construites à partir des données aéroportées AISA-DUAL et des données satellites simulées, en utilisant un modèle PLSR, ont été calibrées à partir de base de données de 90 couples [X_{im} ; y], et validées à partir de base de données de 30 couples [X_{im} ; y].

Résultat Marquant :

Les performances de ces FST construites à partir des données aéroportées AISA-DUAL et des données satellites simulées donnent des performances de prédiction similaires en calibration, avec un R^2_{cal} de 0,74 et un *SEC* de 86 g/kg. En validation, la FST construite à partir des données aéroportées AISA-DUAL se révèle légèrement plus performante que la FST construite à partir des données satellites simulées (R^2_{val} de 0,75 et 0.71, et un *SEP* de 86 et 94 g/kg respectivement). Outre ces indicateurs de performance, la structure spatiale obtenue pour chacune des cartes d'estimation d'argile texturale est similaire.

Ainsi, il a été confirmé que les effets atmosphériques ne semblent pas avoir d'impact prononcé sur la qualité de la spatialisation d'argile texturale, sous la condition d'une correction atmosphérique de qualité [AC2015.2].

3.4.2. Sensibilité à la résolution spatiale

Dans un second temps, la sensibilité à la résolution spatiale de FST pour l'estimation d'argile texturale a été étudiée [AC2015.2]. Pour cela, les données AISA-DUAL simulées en haut de l'atmosphère ont été agrégées à différentes résolutions spatiales (10, 15, 20, 30 et 60 m, par moyenne des spectres) afin de simuler les futurs capteurs satellitaires hyperspectraux. Nous avons également simulé la résolution spatiale de 90 m, qui correspond aux préconisations du projet de cartographie de propriétés de sol mondiale, GlobalSoilMap.net (Sanchez et al., 2009 ; Hempel et al., 2014). Des FST utilisant le modèle PLSR ont été construites à partir des données aéroportées AISA-DUAL et des données satellites simulées pour l'estimation du taux d'argile.

Résultat Marquant :

La précision de nos FST lors d'une dégradation de la résolution spatiale est dépendante de deux paramètres [AC2015.2] :

- Les variations spatiales à courte échelle de la propriété de sol étudiée. La précision des modèles reste invariable tant que la résolution spatiale est inférieure à la variation spatiale à courte échelle de la propriété de sol étudiée. Ainsi sur notre zone, les capteurs satellitaires ayant des résolutions spatiales entre 5 et 30 m offrent des performances correctes (R^2_{cal} > 0.66, Figure 3.4) permettant d'envisager une spatialisation de l'argile texturale.
- La taille des parcelles sur la zone d'étude. La précision des modèles reste invariable tant que la résolution spatiale est inférieure à largeur moyenne des parcelles de la zone d'étude. Dans notre cas, la faible performance des FST construites à partir de données aux résolutions de 60 et 90 m (Figure 3.4) peut s'expliquer par une hétérogénéité spectrale intra-pixel très forte. En effet, la superficie moyenne des parcelles de notre zone est de 0,56 ha (Jenhaoui et al., 2008), donc un pixel de 60 m peut être à cheval sur deux parcelles ayant des humidités différentes, des rugosités différentes, ou ayant subi un travail du sol différent (Figure 3.5), impactant le spectre acquis sur ces surfaces.

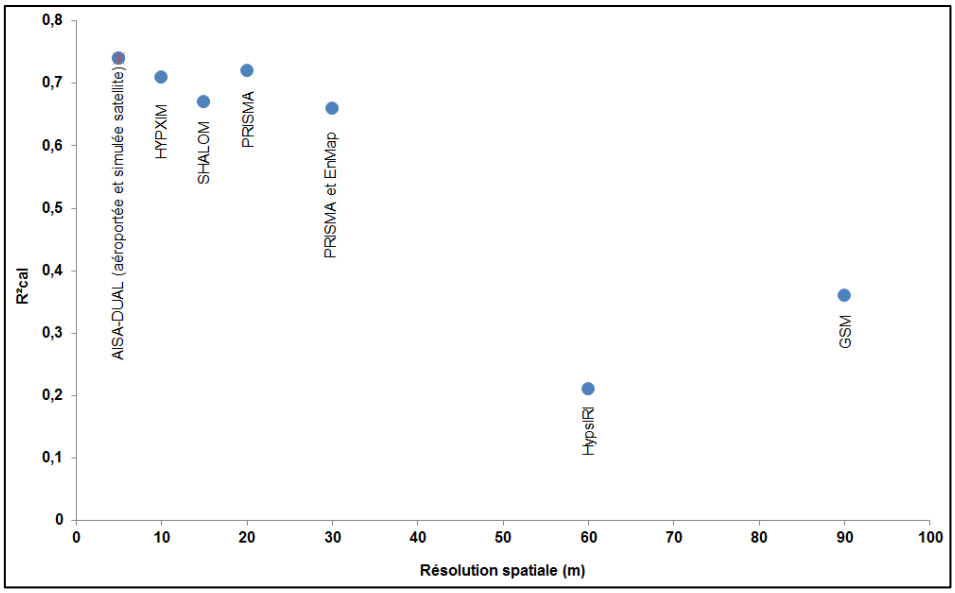


Figure 3.4 : Coefficient de détermination de calibration, obtenu pour chaque FST construite à partir des données AISA-DUAL aéroportée, et des données satellites simulées aux résolutions spatiales suivantes : 10 m (capteur HYPXIM), 15 m (capteur Shalom), 20 m (capteur PRISMA), 30 m (capteur PRISMA et EnMap), 60 m (capteur HyspIRI) et 90 m (résolution fixée par le projet GlobalSoilMap.net).

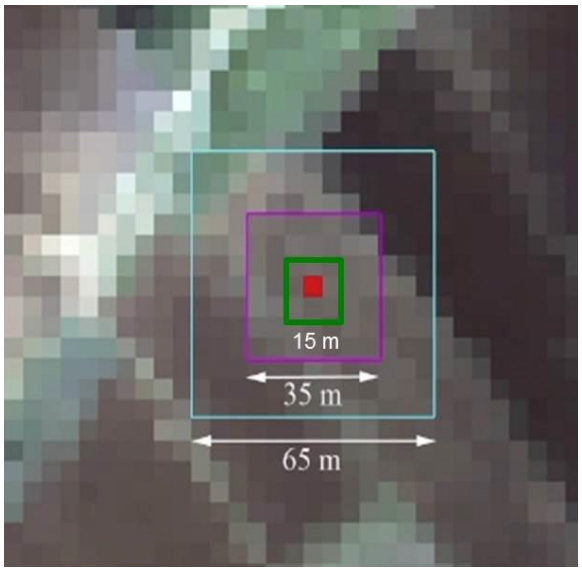


Figure 3.5 : Illustration de pixels de 15 m (carré vert), 35 m (carré violet) et 65 m (carré bleu clair) sur une image AISA-DUAL (bassin versant du Lebna, Tunisie). Le pixel en rouge indique la localisation d'un échantillon de sol dont la teneur en argile *y* a été mesurée, et utilisée dans la base données de calibration $[X_{im},y]$.

3.4.3. Sensibilité à la résolution spectrale

Dans un troisième et dernier temps, la sensibilité à la résolution spectrale de FST pour la l'estimation d'argile texturale a été étudiée. Pour cela, les données AISA-DUAL ont été rééchantillonées à différentes résolutions spectrales afin de simuler une série de capteurs caractérisés par des résolutions spectrales régulières (RSR) (Tableau 3.1) et une série de capteurs caractérisés par des résolutions spectrales irrégulières (RSI) (Tableau 3.1). Et des

FST ont été construites à partir des données aéroportées AISA-DUAL et des données satellites simulées pour l'estimation du taux d'argile textural, en utilisant les modèles PLSR.

	Nom de la configuration simulée ou Nom du capteur simulé	**FWHM [en nm]** *entre 4 – 1000 nm*	**FWHM [en nm]** *entre 1000 – 2500 nm*	**Nombre de bandes spectrales**
Capteurs à résolutions spectrales régulières (RSR)	AISA-DUAL	4.68	6.28	247
	~5/10 (EnMAP)	4.68	12.56	169
	~10/10 (HYPXIM /HyspIRI / HYPERION)	9.36	12.56	123
	~37/37	37.44	37.68	36
	~60/60	60.84	62.8	21
	~100/100	98.28	100.48	12
	~200/200	201.24	200.96	4
Capteurs à résolutions spectrales irrégulières (RSI)	ASTER	Entre 60 et 100	Entre 40 et 100	9
	SENTINEL 2	Entre 15 et 115	Entre 90 et 180	11
	LANDSAT 7	Entre 60 et 150	Entre 200 et 260	6
	LANDSAT 8	Entre 28 et 60	Entre 84 et 187	6

Table 3.1 : Résolutions spectrales des différents capteurs simulés. La FWHM (Full Width at Half Maximum, ou largeur à mi-hauteur) est une mesure de la résolution spectrale.

Résultat Marquant :

L'analyse des performances des FST construites à partir des données aéroportées et des données simulées à différentes résolutions spectrales a permis de mettre en évidence les éléments suivants :

- les performances de FST sont correctes et globalement stables lorsque les variables explicatives X_{im} (spectres de reflectance) ont une résolution spectrale régulière entre 5 et 100 nm ($0.73 < R^2_{cal} < 0,78$) [AC20XX.3].
- parmi les capteurs RSI, seul le capteur multispectral ASTER permettrait de construire une FST avec une précision correcte pour l'estimation d'argile [AC20XX.3].

En complément de cette étude basée sur des données aéroportées, une étude de sensibilité à la résolution spectrale de FST a été réalisée sur des spectres Vis-PIR acquis en laboratoire. Ces travaux complémentaires se sont intéressés à la prédiction de *(i)* 3 propriétés de sol ayant une signature spectrale marquée liée à un composant chimique ou une structure physique associée (cas du fer, argile texturale et $CaCO_3$) et *(ii)* 1 propriété de sol n'ayant pas de réponse spectrale marquée (pH). Ces travaux ont permis de montrer que la performance des FST pour l'estimation de propriétés de sol est d'autant plus sensible à la résolution spectrale que la propriété d'intérêt n'a pas de signature spectrale marquée et n'est pas corrélée à une propriété de sol ayant une signature spectrale marquée [AC2017.1].

3.5 SYNTHESE

Par ces travaux, j'ai contribué au développement de méthodes vouées à rendre plus générique et plus accessible la spatialisation des propriétés pérennes des sols de surface. Pour cela, j'ai choisi de m'attaquer aux trois sources de données nécessaires à l'obtention de ces spatialisations : les données physico-chimiques y et les données spectrales X nécessaires à la calibration des FST, et les données d'imagerie Vis-PIR hyperspectrale. Ces travaux semblent offrir des perspectives intéressantes pour envisager des spatialisations plus étendues et sur des territoires encore peu décrits. Il serait par exemple envisageable de croiser l'approche (1) (utilisation de données sols anciennes) et l'approche (3) (utilisation de données satellites). De même il est possible d'envisager de croiser l'approche (2) (utilisation de base de données spectrales de Labo) et l'approche (3) (utilisation de données satellites).

IV. VERS UNE UTILISATION DE CES SPATIALISATIONS

4.1 INTRODUCTION

A l'approche de l'année 2010, l'imagerie hyperspectrale Vis-PIR (400 - 2 500 nm) apparaît déjà comme un outil prometteur pour cartographier des propriétés physico-chimiques de sol. Malgré la reconnaissance de ce potentiel par la communauté des Sciences du Sol, Ben-Dor et al. (2009) souligne très justement la sous-utilisation de l'imagerie Vis-PIR pour des applications thématiques et le report fréquent de telles intentions. Ainsi le passage du développement méthodologique de ces spatialisations à une étude plus cognitive de ces produits a du mal à voir le jour.

Or ces cartes de propriétés physico-chimiques, issues de l'imagerie Vis-PIR, sont caractérisées par une haute résolution spatiale (autrement dit une forte densité spatiale d'information sur un terrain donné) et une large couverture spatiale (bien plus grande que les territoires pouvant être couverts par l'homme). Ainsi ces produits, d'une richesse inégalée, peuvent à leur tour être utilisés afin de fournir de nouvelles données d'intérêt.

Dans cette optique, j'ai participé à plusieurs travaux voués à l'utilisation de ces produits issus de l'imagerie Vis-PIR (ex : carte d'argile texturale) i) pour une meilleure connaissance des surfaces de sol et la mise à jour de cartes de sol, et ii) comme substitut de données sol classiques pour une extension de ces propriétés pérennes en tout point d'une zone et pour renseigner la sub-surface en terme de propriétés pérennes de sols.

4.2 VERS UNE MEILLEURE CONNAISSANCE DES SURFACES DE SOL

Une analyse des cartes d'estimation des propriétés pérennes de sol produites à haute résolution spatiale (~5 m) (section 2.2) permet d'identifier des motifs d'organisation de la couverture pédologique, d'ouvrir de nouvelles perspectives de compréhension des processus de mise en place des sols dans les milieux étudiés, et d'envisager une amélioration des cartes pédologiques actuelles.

Résultat Marquant :

L'analyse des cartographies de 4 propriétés pérennes des sols (argile texturale, sable, fer et CEC) réalisées à l'échelle du bassin versant du Lebna (300 km^2) (section 2.2) a permis de mettre en évidence un modèle de sol régional complexe sur la zone d'étude, principalement dû à des variations de la lithologie [C2012.1]. Des différences à la fois de teneur en argile texturale et de modèle d'organisation des sols sont apparues entre la zone située dans le coin sud-est de l'image, où les sols sont développés sur des dépôts marins pliocène, et le reste de l'image où les sols sont développés sur des dépôts marin miocènes plus anciens (Figure 2.1). En effet la zone des dépôts marins Pliocène présente de faibles teneurs en argile texturale et celles-ci sont faiblement variables. A l'inverse, la zone de dépôts marins Miocène présente une large gamme de valeurs de teneur en argile texturale dont les variations spatiales sont très bien révélées par la carte d'argile estimée. Elles correspondent aux alternances de bancs de grès et de marnes caractérisant l'étage miocène, entraînant

une alternance de faibles et fortes valeurs de teneur en argile, respectivement. Ces motifs de sol (alternances) varient en direction sud-est / nord-ouest, avec une diminution de la distance entre les affleurements de grès successifs et l'apparition d'un grand affleurement de grès au milieu de la zone d'étude. La carte d'estimation d'argile texturale du bassin versant élémentaire de Kamech (6.6 km^2) est caractéristique des alternances de bancs de grès et de marne (Figure 4.1). On remarque des dépôts de sable (dû à l'érosion des zones de grès dans les vallées) perpendiculaires aux affleurements, qui viennent s'ajouter à la complexité des variations de taux d'argile préalablement décrites.

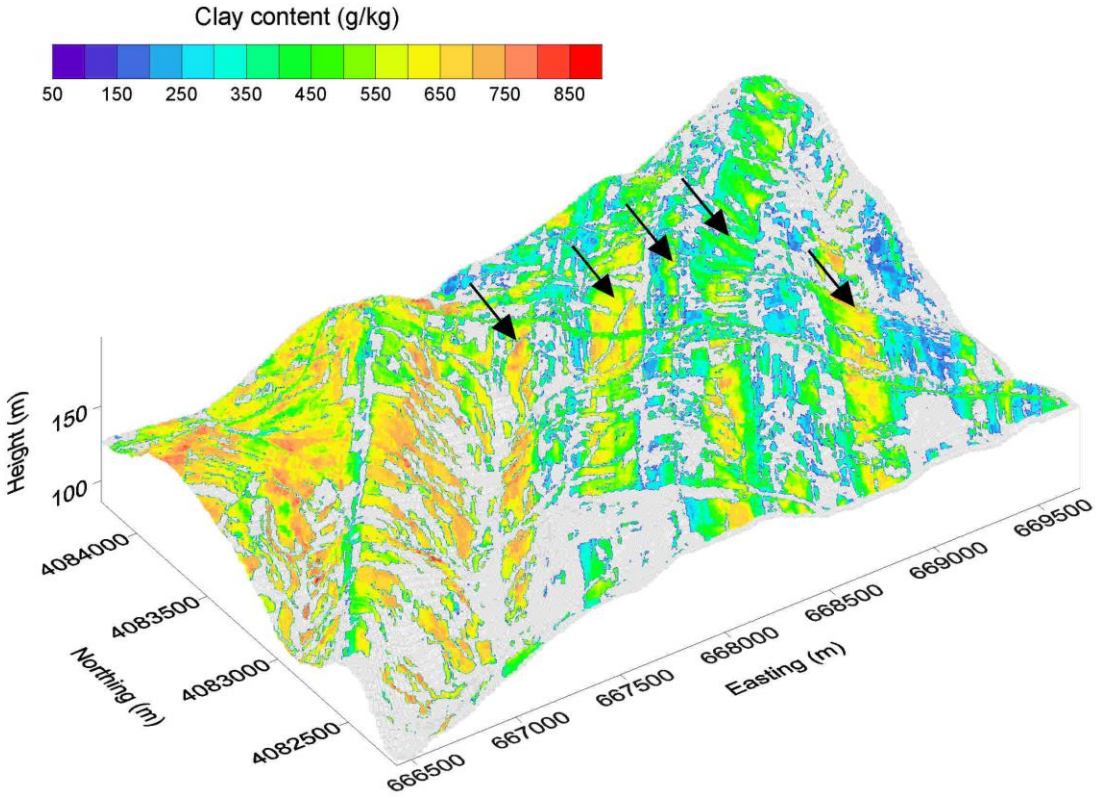


Figure 4.1 : Spatialisation de l'argile par imagerie Vis-PIR [C2012.1] sur un Modèle Numérique de Terrain ASTER, pour le bassin versant élémentaire de Kamech (6.6 km²) inclus dans le bassin versant du Lebna. Les flèches indiquent les bancs de marne.

4.3 LES PRODUITS DE L'IMAGERIE VIS-PIR COMME SUBSTITUTS DE DONNEES SOL EN CNS

La cartographie numérique des sols (CNS), aussi appelée Digital Soil Mapping (DSM), se définit comme « la création et l'enrichissement de systèmes d'information pédologiques à références spatiales par des modèles numériques inférant les variations spatiales et temporelles des sols et de leurs propriétés à partir d'observations de sol et de données spatiales d'environnement des sols » (Lagacherie et al., 2007). Le principe général de la CNS consiste à prédire des classes de sol ou des propriétés de sol (terme *S* à gauche de l'équation [4.1]) en utilisant une fonction de prédiction *f* reliant ce terme *S*, à des données spatiales sur les sols disponibles sur la zone à étudier (terme *S* à droite de l'équation [4.1]) et des données spatiales représentant des éléments du paysage en relation (de causalité ou

non) avec le sol ou l'une de ses propriétés (termes *C, O, R, P, A, N* à droite de l'équation [4.1]), selon l'équation suivante (McBratney et al., 2003) :

$$S = f(S, C, O, R, P, A, N) + \varepsilon \qquad [4.1]$$

Où *C* correspond au climat, *O* aux êtres vivants, *R* au relief, *P* à la roche mère, *A* au temps, *N* à la localisation géographique et ε représente l'erreur de prédiction. Ces données spatiales (termes *C, O, R, P, A, N)* appelées aussi « covariables du sol », sont généralement issues de modèle numérique de terrain, de carte d'occupation du sol, de carte géologique, etc.

Dans le cadre de l'équation dite de « Scorpan » [4.1], l'utilisation des spatialisations de propriétés pérennes des sols, issues de l'imagerie Vis-PIR, peut s'envisager selon deux approches :

- Dans le cas où le produit de l'imagerie Vis-PIR (spatialisation de propriétés pérennes des sols) utilisé concerne spécifiquement la propriété à cartographier (terme *S* à gauche de l'équation [4.1]), chaque pixel du produit de l'imagerie Vis-PIR est considéré comme un nouveau site où une mesure de cette propriété est disponible (terme *S* à droite de l'équation [4.1]). Un exemple de cette approche est décrit en section 4.3.1.
- Dans le cas où le produit de l'imagerie Vis-PIR (spatialisation de propriétés pérennes des sols) utilisé concerne une autre propriété que celle à cartographier (terme *S* à gauche de l'équation [4.1]), chaque pixel du produit de l'imagerie Vis-PIR est considéré comme une nouvelle covariable du sol (nouveau terme à droite de l'équation [4.1], pouvant être apparentée aux termes *C, O, R, P, A, N)*. Un exemple de cette approche est décrit en section 4.3.2.

4.3.1. Les produits de l'imagerie Vis-PIR comme variables d'entrée *S* du modèle Scorpan

Dans le cas où le produit de l'imagerie Vis-PIR (spatialisation de propriétés pérennes des sols) utilisé concerne spécifiquement la propriété à cartographier (terme *S* à gauche de l'équation [4.1]), chaque pixel du produit de l'imagerie Vis-PIR est considéré comme un nouveau site où une mesure de cette propriété est disponible (terme *S* à droite de l'équation [4.1]). L'intérêt de cette approche, par rapport à l'utilisation de mesures sur sites ponctuels (données classiquement utilisées en CNS), repose sur la densité spatiale considérablement plus élevée qui permet d'appréhender de façon très détaillée et sur de grandes superficies les variations de la propriété *S*. A l'inverse, le désavantage de cette approche repose sur la plus faible précision des estimations réalisées par imagerie Vis-PIR, par rapport aux analyses de laboratoires classiques associées sur les sites ponctuels.

Un exemple de spatialisation de propriétés pérennes de sol de surface en tout point de la zone d'étude a été étudié sur le bassin versant du Lebna (Tunisie), à partir *i)* des estimations de ces mêmes propriétés de sol obtenues par imagerie hyperspectrale Vis-PIR sur sol nu (Figure 2.1), et *i)* de mesures de laboratoire de ces mêmes propriétés de sol faites sur des échantillons de sol collectés sur le Lebna. Le modèle d'estimation utilisé est une fonction d'interpolation spatiale basée sur le co-krigeage (Wackernagel, 1995). Cette démarche a été

testée sur les quatre propriétés pérennes spatialisables par imagerie hyperspectrale Vis-PIR sur le bassin versant du Lebna (taux d'argile, de sable, de fer et CEC) [C2012.1].

Résultat Marquant :

La performance d'estimation des propriétés pérennes des sols par la fonction géostatistique de co-krigeage varie, en termes de R^2_{cv} entre 0.41 (pour le fer) et 0.66 (pour l'argile texturale) [C2012.2]. Ainsi le recours à cette approche de CNS permet de renseigner des zones dépourvues d'estimations de propriétés pérennes des sols (Figure 4.2a), car affectées par un facteur de perturbation (par exemple, la végétation et les zones urbaines). Cette approche permet de passer de 41.6 % de zone cartographiée par imagerie hyperspectrale Vis-PIR seule (Figure 2.1) à 100 % de la zone (Figure 4.2a) [C2012.2]. Et la carte des taux d'argile finalement obtenue offre une bien meilleure résolution spatiale que les cartes pédologiques disponibles dans cette même région du Gouvernorat de Nabeul, qui sont pourtant parmi les plus précises que l'on puisse trouver sur le continent africain (Figure 4.2b).

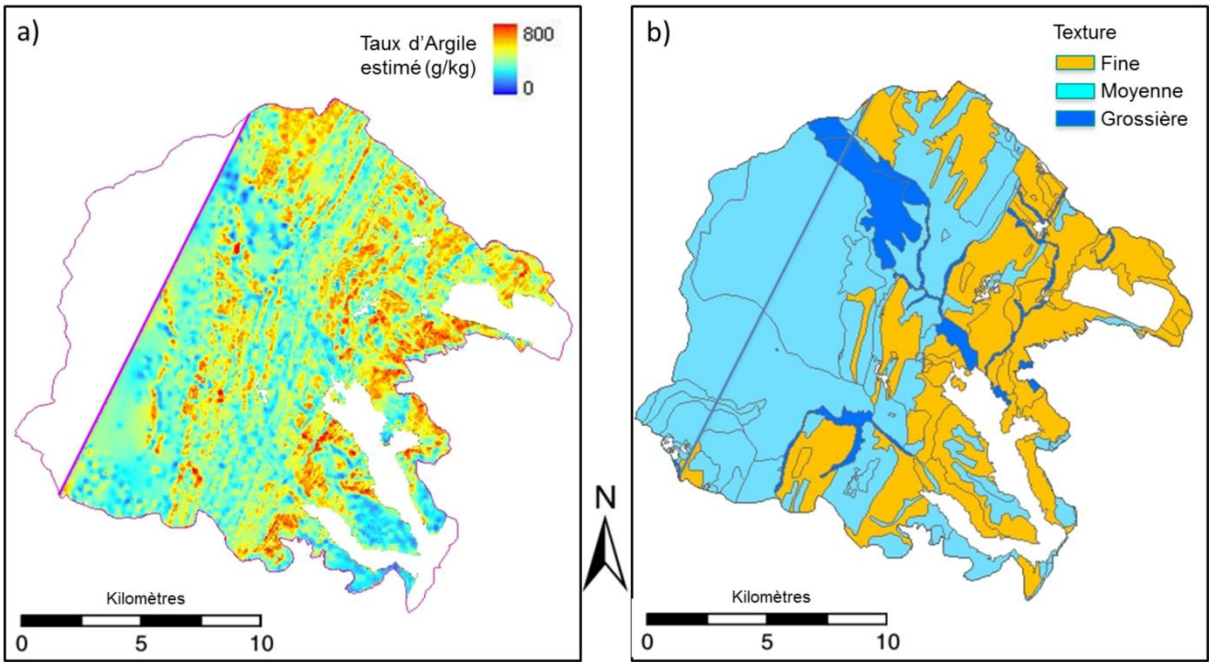


Figure 4.2 : a) Cartographie du taux d'argile par co-krigeage et b) Carte agricole de la Tunisie au 1: 20 000, restreinte aux contours stricts du bassin versant du Lebna.

4.3.2. Les produits de l'imagerie Vis-PIR comme co-variables du modèle Scorpan

Dans le cas où le produit de l'imagerie Vis-PIR (spatialisation de propriétés physico-chimiques des sols) utilisé concerne une autre propriété que celle à cartographier (terme *S* à gauche de l'équation [4.1]), chaque pixel du produit de l'imagerie Vis-PIR est considéré comme une nouvelle covariable du sol (nouveau terme à droite de l'équation [4.1], pouvant être apparentée aux termes *C, O, R, P, A, N)*. L'intérêt de cette approche repose sur la disponibilité de produits de l'imagerie Vis-PIR pouvant être considérés comme des covariables a priori corrélées avec la propriété de sol *S* à cartographier.

En particulier on peut citer le cas de corrélations entre une propriété physico-chimique de sol en surface (sur 0-15 cm de profondeur) et cette même propriété à une couche profonde de sol (par exemple sur 60-100 cm de profondeur). Ainsi, la propriété physico-chimique de sol en surface peut être considérée comme une covariable de la propriété à une couche profonde de sol.

Un exemple de prédiction de propriétés pérennes de couches profondes du sol, à partir d'estimations de surface par imagerie hyperspectrale Vis-PIR de ces mêmes propriétés de sol a été étudié sur le bassin versant du Lebna (Tunisie) [AC2013.3]. La démarche a consisté à calibrer des régressions linéaires *f*, à partir d'un ensemble de profils pédologiques de la région d'étude pour lesquels des propriétés pérennes de sol de laboratoire étaient disponibles, afin d'estimer les différences de valeurs de ces propriétés de sol entre la surface (0-5 cm) et un intervalle de profondeur donné (0-15 cm, 15-30 cm, 30-60 cm ou 60-100 cm) (exemple en Figure 4.3a). Cette démarche a été testée sur les quatre propriétés pérennes spatialisables par imagerie hyperspectrale Vis-PIR sur le bassin versant du Lebna (taux d'argile, de sable, de fer et CEC) [C2012.1].

Les covariables utilisées pour prédire ces différences ont été, d'une part, les cartes de propriétés de sol de surface obtenues par imagerie hyperspectrale Vis-PIR et, d'autre part, trois indicateurs de relief dérivés d'un Modèle Numérique de Terrain, traduisant les processus d'érosion-redéposition des sols susceptibles d'impacter les relations entre propriétés de surface et propriétés profondes (Altitude, Courbure verticale du relief, indice d'appartenance à l'unité géomorphologique « vallée à fond plat »). Une fois ces fonctions de prédiction *f* « surface-profondeur » établies, elles ont été appliquées aux estimations de propriétés de surface obtenues par télédétection hyperspectrale Vis-PIR, produisant alors des cartes de propriétés à différentes profondeurs (Figure 4.3b).

Résultat Marquant :

L'analyse des performances des fonctions de prédiction *f* « surface-profondeur » établies en utilisant entre autres, la spatialisation de propriétés pérennes des sols sur surfaces nues (Figure 2.1) issue de l'imagerie Vis-PIR hyperspectrale, permet de dégager les conclusions suivantes [AC2013.3] :

- Les fonctions de prédiction *f* « surface-profondeur » pour les quatre propriétés testées (taux d'argile, de sable, de fer et CEC) offrent des précisions acceptables jusqu'à la profondeur 30-60 cm (R^2 entre 0.55 et 0.81), puis des précisions faibles aux profondeurs 60-100 cm (R^2 entre 0.38 and 0.43).
- Les cartes de propriétés pérennes des sols en subsurface estimées par les fonctions *f* « surface-profondeur » capturent 1/3 à 2/3 de la variabilité totale, due à l'erreur de propagation.
- Ces cartes de propriétés pérennes des sols en subsurface apportent une source d'information pédologique d'une densité inégalée et mettent en évidence des structures spatiales pédologiques en cohérence avec les connaissances pédologiques de la zone d'étude.

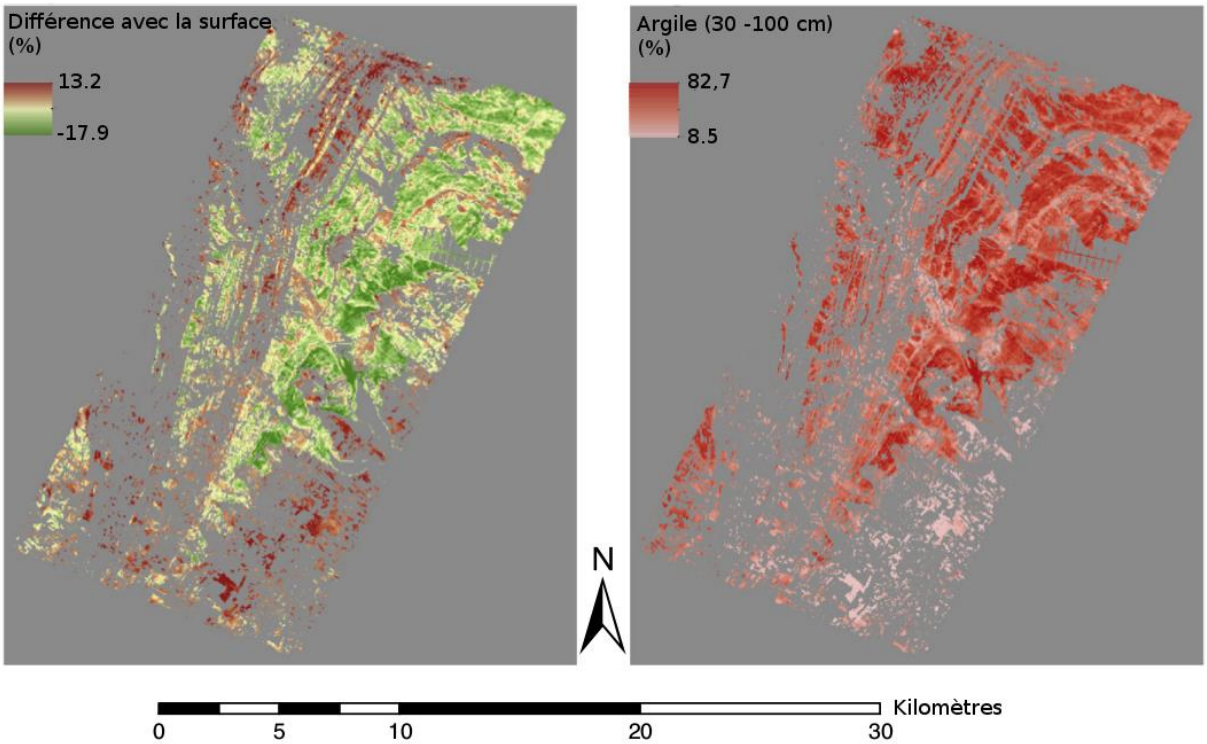


Figure 4.3 : a) Différences d'estimations de taux d'argile entre la surface (0-15 cm de profondeur) et la couche de 30-100 cm de profondeur, par approche CNS en utilisant un produit de l'imagerie Vis-PIR (Figure 2.1). b) Spatialisation du taux d'argile pour la couche de 30-100 cm de profondeur.

4.4 SYNTHESE

A travers ces travaux, je montre que les cartographies de propriétés pérennes des sols obtenues par imagerie hyperspectrale Vis-PIR ne sont pas seulement un produit et donc une finalité, mais peuvent à leur tour devenir des « données de bases ». Au-delà des exemples cités dans ce chapitre, une utilisation de ces données pourrait être envisagée pour des études agronomiques (e.g., étude du lien entre le choix des cultures et la composition physico-chimique des sols), des estimations spatialisées de la stabilité structurale ou encore des caractérisations hydrodynamiques des sols.

V. PERSPECTIVES DE RECHERCHE

Dans cette synthèse de travaux de recherche, j'ai tenté d'exposer mon positionnement scientifique autour de ***la spatialisation des propriétés physico-chimiques de sol de surface*** que je qualifie de pérennes (à l'échelle d'une décennie). L'originalité scientifique de mes travaux repose entre autre sur les deux aspects suivants : 1) le couplage entre données de laboratoire et données d'imagerie Vis-PIR, pour ouvrir une voie vers la démocratisation des estimations spatialisées de propriétés de sol et 2) l'utilisation de produits d'imagerie Vis-PIR dans des approches de cartographie numérique des sols pour une extension spatiale de ces estimations.

Mon projet de recherche s'inscrit dans la continuité de ces travaux et tient compte des ***avancées scientifiques et technologiques*** en spectroscopie et imagerie. Il est actuellement raisonnable d'espérer le lancement du capteur satellitaire hyperspectral Allemand EnMap pour 2018 (Lopinto and Ananasso, 2013 ; Giampaolo et al., 2008). Ce capteur offrira des données hyperspectrales Vis-PIR adaptées à la cartographie de propriétés pérennes de sol, y compris dans des contextes de type Méditerranéen pluvial caractérisés par des parcelles agricoles souvent de petites superficies, grâce à sa résolution spatiale de 30 m [AC2015.2]. De plus, des données satellitaires multispectrales Vis-PIR à haute répétitivité temporelle sont d'ores et déjà disponibles grâce au capteur Sentinel-2A (ESA, 2011, Drusch et al., 2012). Ces données, à priori peu adaptées à la cartographie de propriétés pérennes de sol de par leur résolution spectrale (entre 15 et 115 nm) [AC20XX.3], offrent par contre une série temporelle de données spectrales jusqu'ici inégalée (acquisition tous les 10 jours actuellement, et tous les 5 jours après le lancement du capteur Sentinel-2B) pouvant potentiellement permettre l'étude de changements d'états de surface du sol (e.g., rugosité ou humidité) entre deux dates. Finalement les récents développements de bases de données spectrales de sol (nationales, e.g., Arrouays et al., 2002 ; continentales, e.g., Toth et al., 2013 ; mondiales [AC2016.1]) permettent d'envisager des développements d'utilisation conjointe de données spectrales de laboratoire et d'imagerie à différentes échelles spatiales.

Mon projet de recherche tient également compte des ***enjeux sociétaux*** auxquels nous devons faire face. La *sécurité alimentaire* reste un enjeu toujours croissant, principalement dans les pays du Sud où la concurrence entre différentes utilisations des sols (productions alimentaires *versus* non-alimentaires, e.g., bio-carburants) et la dégradation des sols (qui menace plus de 40% des terres émergées) ne cessent de s'amplifier. Et à cela s'ajoute une augmentation de la population. Un des leviers d'action pour répondre à cet enjeu de sécurité alimentaire réside dans l'étude de la dégradation des sols afin d'inverser la tendance et préserver la ressource en sol. La préservation de la ressource en sol peut être envisagée, entre autre, à travers la mise en place de stratégies d'adaptation des mosaïques paysagères relatives à l'occupation du sol et aux systèmes de culture en vue d'optimiser les fonctions écosystémiques du milieu (Jacob, 2015). Elle peut également être envisagée via la réhabilitation des sols dégradés (dégradation par salinisation, par érosion). La *lutte contre le changement climatique* est également un enjeu croissant affectant une fois de plus et de manière prégnante les pays du Sud où les populations sont les plus vulnérables. Un des leviers d'action pour répondre à cet enjeu de lutte contre le changement climatique réside dans l'étude de la séquestration de carbone dans les milieux (sol, océan, forêt). Dans ce

sens, l'initiative « 4 pour 1000 » vise à favoriser les pratiques agricoles permettant la séquestration du carbone organique dans les sols afin de compenser les émissions carbonées d'origine anthropique (http://4p1000.org/enjeux). Finalement, l'étude et l'optimisation des services écosystémiques fournis par le sol (dont les services d'approvisionnement en produits destinés à l'alimentation et de séquestration du carbone) reposent bien souvent sur des modélisations de processus biophysiques (e.g., modèles d'érosion, de dynamique de carbone) pour lesquelles la spatialisation d'éléments du paysage est nécessaire (e.g., composition physico-chimique des sols, fossés, topographie, climat…).

Dans ce contexte et compte tenu des résultats obtenus jusqu'à présent, je projette de poursuivre les développements initiés en spatialisation de propriétés physico-chimiques pérennes des sols (Figure 5.1, ***Axe 1***). Je mettrai l'accent sur l'évaluation de la généricité de ces approches et sur le développement d'approches utilisant conjointement des données spectrales de laboratoire et d'imagerie satellitaire.

Au-delà de leurs propriétés pérennes, les sols cultivés se caractérisent également par des propriétés de surface variables dans le temps telles que la rugosité (dûe à un éventuel travail du sol, à la présence de mottes ou la présence de croutes) et l'humidité qui sont des propriétés physiques du sol de surface, et les éléments grossiers (cailloux) ou organiques (débris de végétaux secs ou chlorophylliens et cultures). Ainsi je chercherai à accroitre notre connaissance des sols en caractérisant ces propriétés de surface variables dans le temps (Figure 5.1, ***Axe 2***).

La caractérisation spatialisée de ces deux types de propriétés des sols (pérennes et variables) ouvre de nouvelles perspectives constituant deux autres axes de mon projet de recherche.

Les propriétés de surface variables dans le temps peuvent, d'une part, être considérées comme des facteurs de perturbations de la mesure de télédétection Vis-PIR, impactant la qualité des fonctions d'estimation de propriétés pérennes de sol (e.g., Stevens et al., 2008 ; Ben-Dor et al., 2009 ; Lagacherie et al., 2008). La prise en compte de ces propriétés dans la mesure de réflectance, constitue désormais un verrou scientifique majeur qu'il s'agira de lever dans les prochaines années, en vue d'améliorer les performances de modèles de prédiction calibrés sur des données de télédétection (e.g., Denis et al., 2014) (Figure 5.1, ***Axe 3***).

Les propriétés de surface variables dans le temps peuvent, d'autre part, être considérées comme des composantes d'intérêt du paysage pour la caractérisation de propriétés fonctionnelles des sols, telles que l'infiltrabilité, la réserve utile en eau ou encore la stabilité structurale des sols. Il s'agira ainsi d'accéder à la spatialisation de ces propriétés fonctionnelles dont les caractérisations relèvent d'une combinaison de propriétés des sols pérennes et variables dans le temps abordées respectivement dans les axes 1 et 2 (Figure 5.1, ***Axe 4***).

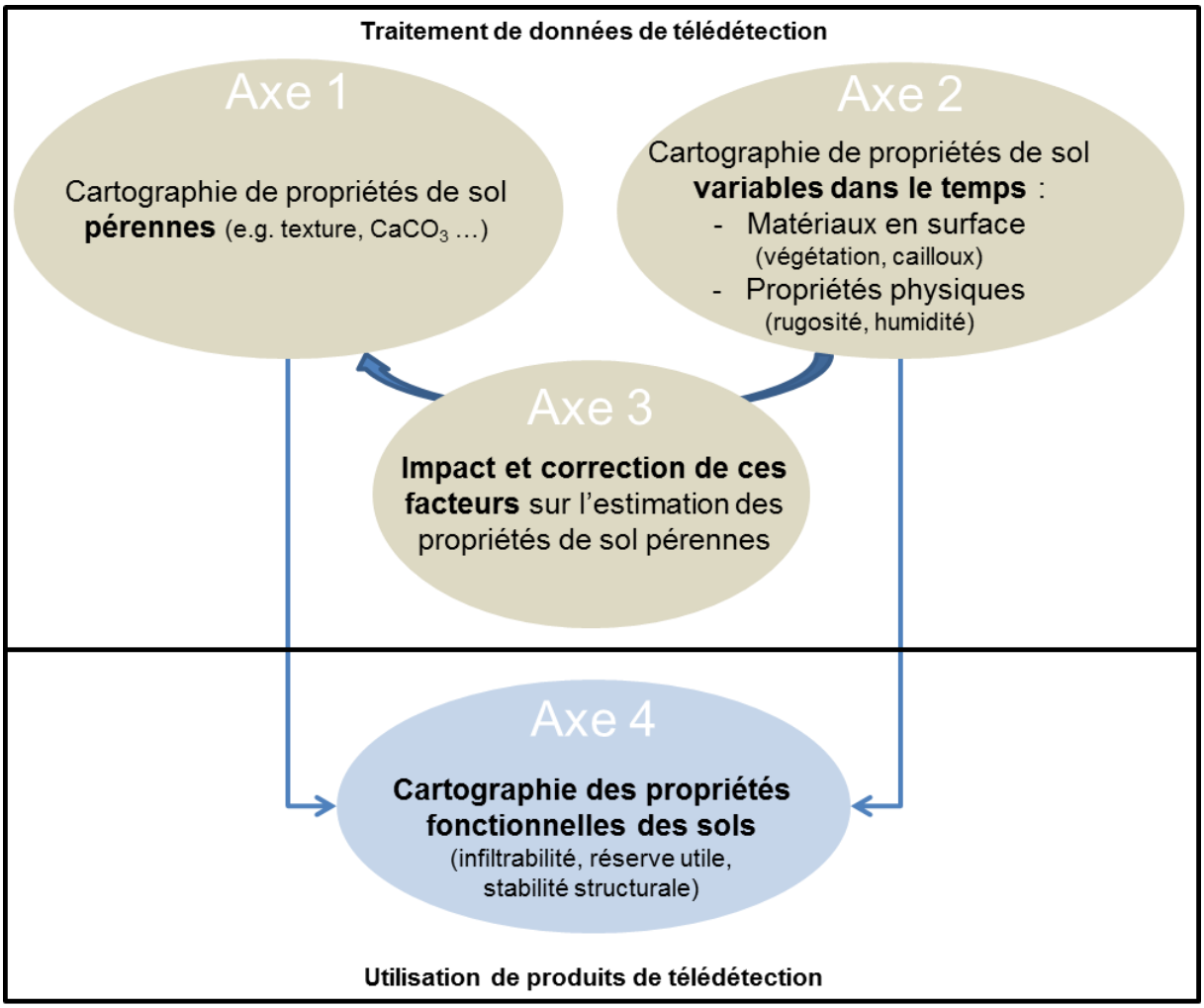


Figure 5.1 : Démarche scientifique envisagée.

Axe 1 : Cartographie de propriétés pérennes des sols. Poursuite des travaux.

Sur les surfaces de sol nu, trois approches innovantes de spatialisation de propriétés pérennes de sol ont été développées afin de limiter les coûts (en argent et en temps) de collecte d'échantillons de sol de surface sur les sites imagés, de mesures physico-chimiques de ces mêmes échantillons, et d'acquisition de données hyperspectrales Vis-PIR (Chapitre III).

Une de ces approches consiste à calibrer les FST en utilisant des bases de données spectrales Vis-PIR acquises en laboratoire couplées à des mesures physico-chimiques [X_{lab} , y], puis à appliquer ces mêmes FST sur des données spatiales X_{im} (e.g., Schwanghart and Jarmer, 2011 ; Jarmer, 2013) (section 3.3). Mes travaux sur ce thème nécessitent de plus amples développements. En premier lieu, il s'agirait de mener à bien les travaux initiés sur l'utilisation de données spectrales labo X_{lab} acquises sur des échantillons prélevés sur une surface A et utilisées pour caler une FST, qui est ensuite appliquée à des données hyperspectrales aéroportées X_{im} acquises sur une surface d'étude B. Dans cette situation, B est inclus dans A et la superficie de B est bien inférieure à la superficie de A (section 3.3, Figure 3.2, Flèche violette). A terme, je souhaiterais poursuivre cette approche en utilisant la Base de Données Globale Mondiale [AC2016.1] pour calibrer des FST et les appliquer ensuite à des données d'imagerie Vis-PIR. En effet, au Sud, il n'existe que très peu de BD régionales et nationales composées à la fois de spectres de labo et d'analyses physico-

chimiques associées. On peut citer les BD constituées par Shepherd and Walsh (2002) sur plus de 1000 sols de surface provenant de l'Afrique de l'Est et du Sud, et par Bellinaso et al. (2012) sur 223 profils de sol au Brésil. Ainsi l'objectif est d'utiliser la BD Globale Mondiale en tout point du Globe. Cela nécessiterait une optimisation de l'utilisation de cette BD Globale (car extrêmement hétérogène spectralement et pédologiquement) en fonction du terrain d'application. La mise en place de modèles dits « locaux » pourrait alors être envisagée (e.g., Stevens et al., 2010).

Les deux autres approches consistent pour l'une à calibrer les FST en utilisant des mesures physico-chimiques *y* de sol réalisées dans le cadre d'anciens programmes de cartographie (section 3.2), et pour l'autre à identifier les données Vis-PIR de futurs capteurs satellitaires ayant le potentiel d'estimer des propriétés pérennes de sol (section 3.4). Ces deux approches pourraient être développées conjointement. Des travaux développés jusqu'alors ont montré que les données multispectrales ASTER permettent de cartographier le taux d'argile texturale en surface (section 3.4.3 ; Nawar et al., 2015 ; AC20XX.5). Il pourrait s'agir d'utiliser conjointement des données multispectrales ASTER et des mesures physico-chimiques *y* de sol réalisées dans le cadre d'anciens programmes de cartographie, pour calibrer des FST d'estimation d'argile texturale.

Ces travaux, qui sont dans la continuité de mes recherches actuelles, pourront être réalisés grâce aux collaborations déjà nouées avec mes collègues du Nord et du Sud.

Axe 2 : Cartographie de propriétés des sols variables dans le temps.
Comme exprimé précédemment, le sol se caractérise également par des propriétés variables dans le temps : une rugosité et une humidité qui sont des propriétés physiques du sol de surface, et les éléments grossiers et organiques.

Une cartographie de ces propriétés de sol variables dans le temps, réalisée en simultané de l'acquisition de données Vis-PIR utilisées pour estimer les propriétés pérennes des sols, amènerait à considérer ces propriétés comme des facteurs « de perturbations » des estimations de propriétés pérennes, et permettrait de tester des corrections de ces facteurs sur les estimations de propriétés pérennes (***Axe 3***). La réalisation de ces cartographies dépendra de la disponibilité en données auxiliaires (e.g., données radar) et du type de données Vis-PIR utilisées pour l'estimation de propriétés pérennes (données multispectrales ou hyperspectrales). Des données Vis-PIR hyperspectrales peuvent à elles seules nous permettre de caractériser la présence d'éléments organiques (végétation verte et sèche) et l'humidité (indice Normalized Soil Moisture Index et Water Index SOIL, Fabre et al., 2015). La rugosité du sol n'est pas, pour le moment caractérisable par des données Vis-PIR hyperspectrales, mais pourrait l'être via l'utilisation de données radar acquises en simultané des données Vis-PIR. Cette action nécessitera la mise en place de nouvelles collaborations, telles que M. Zribi de l'UMR CESBIO (Toulouse).

Des cartographies multi-dates de ces propriétés de sol variables dans le temps, cette fois-ci dissociées des acquisitions de données Vis-PIR utilisées pour l'estimation de propriétés pérennes, serviraient à cartographier des propriétés fonctionnelles des sols (***Axe 4***). Ces cartographies multi-dates pourraient être réalisées en utilisant les capteurs à haute répétitivité temporelle, type Sentinel 1 (radar) et Sentinel 2 (multispectral). Le couvert végétal

chlorophyllien pourrait être estimé et spatialisé via l'utilisation des données multispectrales Sentinel 2 à travers des indices spectraux tels que l'indice de végétation NDVI et de couverture foliaire LAI. Le couvert végétal non-chlorophyllien (résidu de végétaux secs) pourrait être estimé et spatialisé via l'utilisation de données multispectrales Sentinel 2 via l'utilisation i) de la bande spectrale à 2190 nm, correspondant à la longueur d'onde de la cellulose et ii) de la répétitivité temporelle dont bénéficient les données Sentinel 2. En effet, l'évolution temporelle de la réflectance à cette longueur d'onde vis-à-vis des actes techniques au champ (connus à dire d'expert) pourrait être analysée afin de permettre une classification des pixels. En effet, il est attendu que la réflectance à 2190 nm d'un pixel, aux temps t et t+1, varie si ce pixel couvert de résidus végétaux secs à un temps t a subi un labour entre la date d'acquisition t et la date d'acquisition t+1 (particulièrement vrai en Automne au moment des labours).

Enfin, les propriétés physiques de surface variables dans le temps telles que la rugosité et l'humidité pourraient être estimées et spatialisées via l'utilisation de données radar Sentinel 1 (e.g., Zribi et al., 2014 ; Gorrab et al., 2015).

***Axe 3 : Impact et correction de ces propriétés variables dans le temps*, pour l'estimation de propriétés pérennes de sol.**

Dans un premier temps, il s'agirait d'estimer l'impact des propriétés et matériaux en surface variables dans le temps, sur l'estimation des propriétés pérennes de sol. Par exemple, l'humidité du sol a un impact sur le spectre dans le domaine spectral du Visible et les bandes liées aux ions OH. Donc la présence d'humidité dans les sols perturbe l'estimation du carbone organique qui a lui-même une réponse spécifique dans le domaine du Visible (Nocita et al., 2013 ; Minasny et al. 2011), et l'estimation de l'argile texturale qui a une réponse spécifique liée aux ions OH autour de 2206 nm (Minasny et al. 2011). Par contre l'humidité ne perturbe à priori que très peu l'estimation du $CaCO_3$ qui a une réponse spécifique lié aux ions CO^{3+} autour de 2340 nm.

Ces travaux, nécessitant des expérimentations in-situ, pourront être développés en collaboration avec l'ONERA de Toulouse, ou encore l'UMR ITAP de Montpellier.

Dans un deuxième temps il s'agirait de corriger l'effet de ces propriétés et matériaux en surface variables dans le temps, afin d'améliorer l'estimation des propriétés pérennes de sol. Cette correction peut être envisagée selon deux voies.

Une première voie consisterait à corriger l'estimation de propriétés pérennes, sans estimation du (ou des) facteur(s) de perturbation, via l'utilisation de transferts de modèle (section 3.3). A titre d'exemple, Ackerson et al. (2015) ont cherché à corriger l'effet de l'humidité affectant des spectres enregistrés sur le terrain sur des échantillons de sol mouillés, en utilisant des spectres de laboratoire obtenus sur ces mêmes échantillons séchés. Ces travaux étaient dédiés à la prédiction de l'argile texturale.

Une deuxième voie consisterait à corriger l'estimation de propriétés pérennes, après estimation du (ou des) facteur(s) de perturbation. Mes travaux de spatialisation d'argile texturale sur surfaces semi-végétalisées appartiennent à cet axe (section 2.3). Alors que seulement 4 % d'une zone d'étude test (surfaces de sol nu) peut être cartographiée en terme de taux d'argile par FST classique [AC2012.2], l'approche DSS permet la spatialisation de près de 63 % de la surface de cette zone d'étude test [AC2016.2]. A ce stade il s'agirait :

- d'étudier la robustesse de l'étape de séparation aveugle de sources à travers une analyse de sensibilité de l'algorithme au nombre de matériaux (=sources) supposés dans les pixels mixtes et à leurs natures.
- d'étudier le comportement de l'approche DSS lorsque celle-ci est dédiée à la prédiction de propriétés de sol ayant une signature spectrale moins marquée que l'argile texturale (par exemple le fer ou la matière organique) et dans des contextes pédologiques différents.
- d'améliorer l'approche DSS en utilisant le maximum d'information à disposition (connaissance pédologique, géographique de la zone, occupation du sol...). Il pourrait s'agir d'adapter l'approche DSS (à la fois l'étape de séparation aveugle de sources et l'étape de prédiction d'argile) à chaque unité pédologique des zones d'étude.

Notons que cette approche pourrait être transposée à d'autres applications telles que l'estimation de propriétés biophysiques des plantes ou l'identification d'espèces végétales par imagerie Vis-PIR, en présence de matériau non végétal (par exemple le sol) au sein des pixels étudiés, comme initié par Tits et al. (2014).

Une approche complémentaire à la DSS consisterait à classer les facteurs de perturbations et à construire des FST (ou des DSS) par classe de facteurs de perturbations. Dans cette optique, Nocita et al. (2013) proposent de classer l'humidité en classes en utilisant l'indice spectral NSMI (Normalized Soil Moisture Index) puis de construire une FST pour l'estimation de carbone organique des sols, par classe d'humidité.

Axe 4 : Cartographie des propriétés fonctionnelles des sols.

L'estimation et la spatialisation de propriétés fonctionnelles des sols (e.g., infiltrabilité, réserve utile et stabilité structurale) pourraient reposer sur trois approches, via l'utilisation conjointe des propriétés de sol pérennes et variables dans le temps estimées par télédétection. En effet, l'infiltrabilité des sols cultivés peut être expliquée et raisonnablement prédite à partir d'un nombre limité i) de caractéristiques observables de la surface du sol et ii) de leurs modalités : présence/absence de travail du sol, caractéristiques structurales (rugosité ou taille des mottes, présence/absence de croutes), caractéristiques de couverture du sol (éléments minéraux –cailloux- ou organiques -débris de végétaux et couverture végétale-) et caractéristiques texturales (répartition entre l'argile, le limon et le sable) (e.g., Tighe et al., 2012). La réserve utile en eau des sols dépend quant à elle principalement de la texture du sol, de la teneur en éléments grossiers et de la profondeur du sol (e.g., Cousin et al., 2014). Et la stabilité structurale dépend principalement de la matière organique et de la texture du sol (e.g., Le Bissonnais et al, 2007).

Une première approche consisterait à développer des fonctions de pédotransfert (Bouma, 1989) qui sont établies le plus souvent par régression multilinéaire mettant en relation des propriétés physico-chimiques du sol et la propriété fonctionnelle recherchée. A titre d'exemple, Le Bissonnais et al (2007) ont développé des fonctions de pédotransfert visant à estimer la stabilité structurale des sols Méditerranéens à partir de la texture, le $CaCO_3$ et le carbone organique de ces sols. Un premier test encourageant d'estimation de stabilité structurale par spectroscopie Vis-PIR de laboratoire a été mené avec des collègues de l'UMR LISAH et des partenaires Tunisiens de l'INRAT et l'INRGREF [AC2013.2]. De tels travaux pourraient être étendus à la spatialisation de cette même propriété fonctionnelle, via l'utilisation de données d'imagerie Vis-PIR.

Une seconde approche consisterait à développer des modèles de type système expert (aussi appelés règles de pédotransfert), sous forme d'un arbre de décision afin de définir des

classes qualitatives des propriétés fonctionnelles du sol. A titre d'exemple, Leonard and Andrieux (1998) ont développé un modèle de type système expert visant à estimer l'infiltrabilité des sols Méditerranéens, à partir du couvert végétal, de la texture et de la rugosité des sols. La spatialisation de ces composants du sol réalisée au cours des axes 1 et 2, pourrait être utilisée dans de tels modèles afin de retrouver une infiltrabilité des sols.

Une troisième approche consisterait à inverser des modèles de culture pour en déduire des propriétés fonctionnelles de sol, en utilisant des produits de télédétection. A titre d'exemple, l'utilisation de données de végétation et d'humidité des sols issues de traitement d'images de télédétection dans des modèles de végétation et de culture a récemment permis d'estimer des propriétés hydriques de surface et de la zone racinaire (e.g., Shin et al., 2013 ; Mohanty 2013). Sur la base de ces travaux, il pourrait être envisagé d'estimer la réserve utile des sols par inversion de modèles de culture, en utilisant des cartes d'humidité, de texture du sol et de couverture d'éléments grossiers obtenues par traitement de données Vis-PIR hyperspectrales EnMap et multispectrales Sentinel-2 (en axes 2 et 3).

Travaux « Trans-Axes » :

Sur la base du travail d'estimation des incertitudes liées aux prédictions de propriétés pérennes de sol (section 2.4), il serait nécessaire d'étudier les incertitudes associées à chaque spatialisation de propriétés de sol pérennes, variables dans le temps et fonctionnelles. Ces incertitudes devront intégrer l'incertitude liée *i)* au modèle de prédiction (e.g., modèle de régression, fonction de pédotransfert, arbre de décision…) et *ii)* à la mesure des propriétés et matériaux utilisés pour la calibration de ces modèles.

Enfin, les cartes de propriétés de sol pérennes, variables dans le temps et fonctionnelles pourraient être utilisées comme paramètres d'entrée de modèles de processus biophysiques tels que les modèles des cultures (e.g., Brisson et al., 2010), de dynamique du carbone (e.g., Falloon et Smith, 2002), modèles hydrologiques (e.g., Moussa et al, 2002), d'érosion (e.g., Le Bissonnais et al., 2005), afin d'aider à la décision en matière de préservation des ressources en sol et de production agricole. Mon environnement de recherche au sein de l'UMR LISAH est propice à la construction de collaboration sur ce thème. Sans pour autant porter ce travail, je pourrais collaborer avec des « processologues », afin d'estimer le bénéfice de l'utilisation de ces produits de télédétection dans les modèles de processus biophysiques. Il s'agira alors d'étudier la sensibilité des modèles à ce type de données et la propagation d'erreur qui résulte de l'utilisation de telles données dans ces modèles.

REFERENCES

Ackerson, J.P., Demattê, J.A.M. and Morgan, C.L.S. (2015). Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma*, 259–260, p. 196–204.

Ans B., Hérault J. and Jutten C. (1985). Adaptive neural architectures: Detection of primitives. *in: Proceedings of COGNITIVA'85,* Paris, France, p. 593–597.

Arrouays D., Jolivet C., Boulonne L., Bodineau G., Saby N. and Grolleau E. (2002). A new initiative in France: a multi-institutional soil quality monitoring network. *Comptes Rendus de l'Academie d' Agriculture de France,* 88, p. 93–105.

Bartholomeus H., Kooistra L., Stevens A., Van Leeuwen M., Van Wesemael B., Ben-Dor E. and Tychon B. (2010). "Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy". *International Journal of Applied Earth Observation and Geoinformation*, 13(1), p. 81–88.

Baume O., SkÃ¸ien J.O., Heuvelink G.B.M. and Pebesma E.J. (2011). A geostatistical approach to data harmonization - application to radioactivity exposure data. *International Journal of Applied Earth Observation and Geoinformation,* 13 (3), p. 409–419.

Bayer A., Bachmann M., Müller A. and Kaufmann H. (2012). "A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem," *Applied and Environmental Soil Science,* 2012, Article ID 971252, 20 pages.

Bayliss J., Gualtieri J.A. and Cromp R.F. (1997). "Analysing hyperspectral data with independent component analysis". In *Applied Image and Pattern Recognition Workshop*, 3240, p. 133–143.

Bellinaso H., Dematte J.A.M. and Romeiro S.A. (2010). Soil Spectral Library and Its Use in Soil Classification. *Revista Brasileira de Ciencia do Solo*, 34(3): 861-870.

Ben-Dor E. and Banin A. (1995a). Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of American Journal, 59, p. 364–372

Ben-Dor E. and Banin A. (1995b). Near infrared analysis (NIRA) as a simultaneously method to evaluate spectral featureless constituents in soils. *Soil Science*, 159, p. 259–269

Ben-Dor E., Patkin K., Banin A. and Karnieli A. (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data-a case study over clayey soils in Israel. *International Journal of Remote Sensing,* 23 (6), p. 1043-1062.

Ben-Dor E., Goldlshleger N., Benyamini Y., Agassi M. and Blumberg D.G. (2003). The Spectral Reflectance Properties of Soil Structural Crusts in the 1.2- to 2.5-µm Spectral Region

Ben-Dor E., Chabrillat S., Demattê J.A.M., Taylor G.R., Hill J., Whiting M.L. and Sommer S. (2009). "Using imaging spectroscopy to study soil properties". *Remote Sensing of Environnment,* 113, p. 38-55.

Ben-Dor E., Kafri A. and Varacalli G. (2014). SHALOM: An Italian-Israeli Hyperspectral orbital Mission – Update. *International Geoscience and Remote Sensing Symposium*, Quebec, Canada, July 13-18, 2014.

Brisson N., Gate P., Gouache D., Charmet G., Oury F.X. and Huard F. (2010). "Why are wheat yields stagnating in Europe? A comprehensive data analysis for France". *Field Crops Research*. 119 : 201- 212.

Bouma J., (1989). "Using soil survey data for quantitative land evaluation". *Advances in Soil Science*, 177–213.

Bouveresse E. and Massart D.L. (1996). Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics and intelligent laboratory systems*, *32*(2), p. 201-213.

Briottet X., Marion R., Carrere V., Jacquemoud S., Bourguignon A., Chami M, Chanussot J., Chevrel S., Deliot P., Dumont M., Foucher P.Y., Minghelli-Roman A., Sheeren D., Weber C., Prastault P, Hosford S. and Lefevre M.J. (2013). "HYPXIM: HYPXIM: a second generation high spatial resolution hyperspectral satellite for dual applications" 5th *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing* (WHISPERS), June 2013, Gainesville, Florida, USA.

Brown D.J. (2007). Using a globalVNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma* 140 (4), p. 444–453.

Bussoletti E. (2012). "Space Observations for Agriculture and Food Support". Inter-Agency *Meeting on Outer Space Activities: 2012*. Thirty-second session , 7-9March 2012. Rome, Italy.

Carrere V., Bourguignon A., Briottet X., Chami M., Chevrel S., Jacquemoud S. and Marion R. (2013). The French Hyperspectral Earth Observation Science/Defense mission HYPXIM - A second generation high spectral and spatial resolution imaging spectrometer, *IGARSS*, 21-26 July 2013, Melbourne, Australia.

Clark R.N. and Roush T.L. (1984). Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), p. 6329–6340.

Common P. (1994). Independent component analysis, a new concept ? Signal processing 36, p. 287–314.

Cousin I., Nicoullaud B., Tetegan D. M., Richer de Forges A., Arrouays D. and Bouthier, A. (2014). Estimating the Available Water Content of highly heterogeneous soils including stony soils at the regional scale. In: Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, Alex McBratney, dir., *GlobalSoilMap: Basis of the global soil information system*, Leiden, NLD : CRC Press, p. 221-225.

Cudahy T., Hewson R., Huntington J., Quigley M. and Barry P.S. (2001). The Performance of the Satellite-Borne HYPERION Hyperspectral VNIR-SWIR Imaging System for Mineral Mapping at Mount Fitton, South Australian, *Proceedings IEEE 2001 International Conference on Geoscience and Remote Sensing*.

Denis A.A., Stevens A., Wesemael B.V., Udelhoven T. and Tychon B. (2014). Soil organic carbon assessment by field and airborne spectrometry in bare croplands: accounting for soil surface roughness, *Geoderma*, 226–227, p. 94-102,

Dobos E., Carré F., Hengl T., Reuter H.I. and Tóth G. (2006). Digital Soil Mapping as a support to production of functional maps. EUR 22123 EN, *Office for Official* Publications of the European Communities, Luxemburg, 68 pp

Drusch M., Del Bello U., Carlier S., Colin O., Fernandez V., Gascon F., Hoersch B., Isola C., Laberinti P., Martimort P., Meygret A., Spoto F., Sy O., Marchese F., Bargellini P. (2012) Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment, 120, p. 25-36.

ESA. 2011. GMES Overview. http://www.esa.int/esaLP/SEMRRI0DU8E_LPgmes_0.html.

Escadafal R. (1994). Soil spectral properties and their relationships with environmental parameters - Examples from arid regions." in: Hill und Mégier (Ed.): *Imaging Spectrometry - A Tool for Environmental Observations*, ECSC, EEC, EAEC, p. 71-87.

Fabre S., Leasignoux A. and Briottet X. (2015). Estimation of Soil Moisture Content on Spectral Reflectance of Bare Soils in the 0.4 -2.5 m Domain. Sensors, MDPI, p.1-18.

Falloon P. and Smith P. (2002). “Simulating SOC changes in long-term experiments with RothC and CENTURY: model evaluation for a regional scale application”. *Soil Use and management*, 18: 101–111.

Fearn T. (2001). Standardisation and calibration transfer for near infrared instruments: a review. *Journal of Near Infrared Spectroscopy*, *9*(4), p. 229-244.

Fernandez-Ahumada E., Roger J.M. and Palagos B. (2012). “A new formulation to estimate the variance of model prediction. Application to near infrared spectroscopy calibration”. *Analytica Chimica Acta*, 721, p. 28–34.

Geladi P. and Esbensen K.J. (1990). The start and early history of chemometrics: selected interviews. Part 1. *Journal of Chemometrics* , 4, p. 337-354.

Giampaolo P., Cisbani A., De Cosmo V., Galeazzi C., Labate D. and Melozzi M. (2008). Hyperspectral Instruments for Earth Observation. *International conference on Space Optics*, October 14-17 Toulouse, France.

Giardino C., Bresciani M., Matta E. and Brando V. (2014). Hyperspectral observations of optical properties in lakes in Perspectives of futures satellite sensors - a case study in Italy. *In Whispers*, Lusanne, SWISS, 24-27 June 2014.

Gogé F., Joffre R., Jolivet C., Ross I. and Ranjard L. (2012). Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. *Chemometrics and Intelligent Laboratory Systems* 110, p. 168–176

Gorrab A., Zribi M., Baghdadi N. and Lili Chabaane Z. (2015). Mapping of bare soil surface parameters from TerraSAR-X radar images over a semi-arid region, in *Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII*, Christopher M. U. Neale; Antonino Maltese, Editors, *Proceedings of SPIE* Vol. 9637 (SPIE, Bellingham, WA 2015), 96371F.

Hempel J.W., McBratney A.B., Arrouays D., McKenzie N.J. and Hartemink A.E. (2014). GlobalSoilMap project history. In: Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, Alex McBratney, dir., *GlobalSoilMap: Basis of the global soil information system*, Leiden, NLD : CRC Press, p.3-8.

Hengl T., Heuvelink G.B.M., Kempen B., Leenaars J.G.B., Walsh M.G., Shepherd K.D., Sila A., MacMillan R.A., De Jesus J.M., Tamene L. and Tondoh J.E. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One, 10, p. 1–26.

Jacob F. (2015) ALMIRA : adapter les mosaïques paysagères pour mieux gérer la production, les sols et l’eau agricoles. *Les dossiers d'Agropolis International Changement climatique : impacts et adaptations*, 20, 11.

Jarmer T. (2013). Spectroscopy and hyperspectral imagery for monitoring summer barley. *International journal of remote sensing*, *34*(17), p. 6067-6078.

Jenhaoui J., Raclot D. and Lamachère J.M. (2008). Le parcellaire et l'occupation du sol entre 2004 et 2008 sur le bassin versant de Kamech (Cap Bon, Tunisie). Tunis: UMR LISAH, IRD (15 pp. + annexes).

Jensen J.R. and Cowen D.C. (1999). Urban/Suburban Infrastructure and Socio-Economic Attributes, *Photogrammetric Engineering and Remote Sensing,* 65(5), p. 611-622.

Jutten C. and Herault J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24, p. 1–10.

Kennard R.W. and Stone L.A. (1969). Computer aided design of experiments. *Technometrics*, 11 (1), p. 137–148.

Knadel M., Deng F., Thomsen A., and Greve M.H. (2012). Development of a Danish national Vis-NIR soil spectral library for soil organic carbon determination. *Digital Soil*

Assessments and Beyond. CRC Press 2012, Minasny, Malone McBratney (eds). p. 403-408.

Kruse F.A., Boardman J. and Huntington J. (2003). Comparison of Airborne Hyperspectral Data and EO-1 HYPERION for Mineral Mapping, *IEEE Transactions on Geoscience and Remote Sensing*, 41(6).

Kruse F.A., Clasen C.C., Kim A.M. and Carlisle S.C. (2012). "Effects of spatial and spectral resolution on remote sensing for disaster response," *Proceedings of the IEEE Geoscience and Remote Sensing Symposium* (IGARSS 2012), 22 - 27 July 2012, Munich, Germany.

Kruse F.A. and Perry S. (2013). Mineral mapping using simulated Worldview-3 short-wave-infrared imagery, *Remote Sensing*, 5, p. 2688–2703.

Lagacherie P. and McBratney A.B. (2007). « Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping», In P. Lagacherie, McBratney A.B., & Voltz M. (Eds.), *Digital Soil Mapping: an Introductory perspective*, Elsevier, p. 3–22.

Lagacherie P., Baret F., Feret J-B, Madeira Netto J. and Robbez-Masson J-M (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, 112 (3), p. 825-835.

Le Bissonnais Y., Daroussin J., Jamagne M., Lambert J.-J., Le Bas C., King D., Cerdan O., Léonard J., Bresson L.-M. and Jones R.J.A. (2005). “Pan-European soil crusting and erodibility assessment from the European Soil Geographical Database using pedotransfer rules”. *Advances in Environmental Monitoring and Modelling*, 2, 1, p.1-15.

Le Bissonnais Y., Blavet D., De Noni G., Laurent J.Y., Asseline J. and Chenu C. (2007). Erodibility of Mediterranean vineyard soils: relevant soil aggregate stability methods and significant soil variables. *European Journal of Soil Scienc*e 58, p. 188–195.

Lee D.D. and Seung H.S. (2001). “Algorithms for non-negative matrix factorization”. *Advances in Neural Information Processing Systems*, 13: MIT Press, p. 556–562.

Legendre L. and Legendre P. (1998). Numerical Ecology. *2nd English ed. Elsevier. Masson*, Amsterdam, p. 39–45.

Leonard J. and Andrieux P. (1998). Infiltration characteristics of soils in Mediterranean vineyards in Southern France. *Catena*, 32: 209-223. doi:10.1016/S0341-8162(98)00049-6.

Levin N., Kidron G.J. and Ben-Dor E. (2007). Surface properties of stabilizing coastal dunes: combining spectral and field analyses. *Sedimentology*, 54, p. 771-788

Lobell D.B. and Asner G.P. (2002) “Moisture effects on soil reflectance”. *Soil Science Society of America Journal*.,66, p. 722–727.

Lopinto E. and Ananasso C. (2013). "The Prisma Hyperspectral Mission". *33rd EARSeL Symposium*, *Towards Horizon 2020: Earth Observation and Social Perspectives*, Matera (Italy), 3-6 June 2013.

Lugassi R., Ben-Dor E. and Eshel G. (2010). A spectral-based method for reconstructing spatial distributions of soil surface temperature during simulated fire events. *Remote Sensing of Environment*, 114(2), p. 322–331.

McBratney A.B., Mendonca Santos M.L., Minasny B. (2003). « On digital soil mapping », *Geoderma*, 117(1-2), p. 3–52.

Madeira Netto J.S., Bedidi A., Pouget M. and Cervelle B. (1997). “Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: the application of a Thematic Mapper (TM) image for soil-mapping in Brasilia, Brazil”, *International Journal of Remote Sensing,* 18, p. 2835–2852.

Madeira Netto J.S.R., Robbez-Masson J.M. and Martins E. (2006). Visible-NIR hyperspectral imagery for discriminating soil types in the La Peyne watershed (France). In: Lagacherie, P., Mc Bratney, A.B., Voltz, M. (Eds.), *Digital Soil Mapping: an Introductory Perspective*. Elsevier.

Mathieu R., Pouget M., Cervelle B. and Escadafal R. (1998). «Relationships between satellite based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment», *Remote Sensing of Environment*, 66, p. 17-28.

Miesch C., Poutier L., Achard V., Briottet X., Lenot X. and Boucher Y. (2005). "Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery". *IEEE Transactions on Geoscience and Remote Sensing*, 437(7), p. 1552–1562.

Minasny B., McBratney A.B., Bellon-Maurel V., Roger J.-M., Gobrecht A., Ferrand L. and Joalland S. (2011). Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma*, p. 167–168, p. 118–124.

Mohanty B.P. (2013). Soil Hydraulic Property Estimation Using Remote Sensing: A Review. Vadose Zone Journal, 12.

Moussa R., Voltz M. and Andrieux P. (2002). "Effects of the spatial organization of agricultural management on the hydrological behaviour of a farmed catchment during flood events", *Hydrological Processes*, 16, p.393-412.

Nawar S., Buddenbaum H and Hill J. (2015). Digital Mapping of Soil Properties Using Multivariate Statistical Analysis and ASTER Data in an Arid Region. *Remote Sensing*. 7(2), p. 1181-1205

Nijland W., Addink E.A., De Jong S.M. and Van der Meer F.D. (2009). Optimizing spatial image support for quantitative mapping of natural vegetation, *Remote Sensing of Environment*, 113, p. 771-780.

Ninomiya Y (2004). Lithologic Mapping with Multispectral ASTER TIR and SWIR Data, Sensor, Systems, and Next generation Satellites VII. *SPIE* 5234, p. 180–190.

Nocita M., Stevens A., Tóth G., Van Wesemael B., Montanarella L. (2013). Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biology Biochemisty*, 68, p. 337-347.

Palacios-Orueta A., Pinzón J.E., Ustin S.L. and Roberts D.A. (1999). Remote Sensing of Soils in the Santa Monica Mountains: II. Hierarchical Foreground and Background Analysis, *Remote Sensing of Environment*, 68, 2, p. 138-151.

Rahman A.F., Gamon J.A., Sims D.A. and Schmidts M. (2003). Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland. *Remote Sensing of Environment*. 84, p. 192–207.

Roberts D.A., Quattrochi D.A., Hulley G.C., Hook S.J. and Green R.O. (2012). Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. *Remote Sensing of Environment*, 117, p. 83–101.

Rowan L. and Mars J. (2002). Lithological Mapping in the Mountain Pass, California Area Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Data, *Remote Sensing of Environment,* 3, p. 350–366.

Rowan L., Abrams M. and Mars J. (2003). Mapping Hydrothermally Altered Rocks at Cuprite, Nevada, Using the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), a New Satellite-Imaging System, *Economic Geology and the Bulletin of the Society of Economic Geologists*, 98(5), p. 1019–1027.

Sanchez P., Ahamed, S., Carré, F., Hartemink, A., Hempel, J., Huising, J., et al. (2009). "Digital soil map of the world". *Science*, 325, p. 680–681.

Schaaf A.N., Dennison P.E., Fryer G.K., Roth K.L. and Roberts D.A. (2011). Mapping Plant Functional Types at Multiple Spatial Resolutions Using Imaging Spectrometer Data. *GIScience and Remote Sensing*. 48(3), p. 324-344.

Schwanghart W. and Jarmer T. (2011). "Linking spatial patterns of soil organic carbon to topography—a case study from south-eastern Spain". *Geomorphology*, 126(1), p. 252-263.

Selige T., Bohner J. and Schmidhalter U. (2006). High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, 136(1-2), p. 235-244.

Shenk J.S. and Westerhaus M.O. (1991). New standardization and calibration procedures for NIRS analytical systems. Crop Science, 31(6), p. 1694-1696.

Shepherd K.D. and Walsh M.G. (2002). Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal. 66 (3), 988 – 998.

Shin Y., Mohanty B.P. and Ines A.V.M (2013). Estimating Effective Soil Hydraulic Properties Using Spatially Distributed Soil Moisture and Evapotranspiration. Vadose Zone Journal, 12(3).

Stevens A., Van Wesemael B.,Vandenschrick G., Touré S. and Tychon B. (2006). Detection of Carbon Stock Change in Agricultural Soils Using Spectroscopic Techniques, *Soil Science Society of America journal*, 70, p. 844–850.

Stevens A., Van Wesemael B., Bartholomeus H., Rosillon D., Tychon B. and Ben-Dor E. (2008). Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils, *Geoderma*, 144(1–2), p. 395-404.

Stevens A., Udelhoven T., Denis A., Tychon B., Lioy R., Hoffmann L. and Van Wesemael B. (2010). "Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy", *Geoderma*, 158, 1–2, p. 32-45.

Stuffler T., Kaufmann H., Hofer S., Förster K.-P., Schreier G., Müller A., Eckardt A., Bach H., Penne B., Benz U. and Haydn R. (2007). The EnMAP hyperspectral imager- An advanced optical payload for future applications in Earth observation programmes**,** *Acta Astronautica*, 61, 1-6, p. 115-120.

Tenenhaus M. (1998). "La régression PLS". *Paris: Editions Technip* (254 pp.).

Tighe M., Muñoz Robles C., Reid N., Wilson B. and Briggs S. V. (2012). Hydrological thresholds of soil surface properties identified using conditional inference tree analysis. *Earth Surface Processes and Landforms*, 37(6), 620-632.

Thorp K.R., French A.N. and Rango A. (2013). Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (MESMA). *Remote Sensing of Environment*. 132, p. 120–130.

Tits L., Somers B., Saeys W. and Coppin P. (2014) Site-Specific Plant Condition Monitoring Through Hyperspectral Alternating Least Squares Unmixing. IEEE *Journal of Selected topics in Applied Earth Observations and remote Sen*sing. 7 (8), p. 3606-3618.

Toth G., Jones A. and Montanarella L. (2013). LUCAS Topsoil Survey: Methodology, Data and Results. *Publications Office of the European Union*, Luxembourg, p.141. 10. 2788/97922.

Tourlière B., Perrin J., LeBerre P. and Pasquet J. (2003). Use of Airborne Gamma-Ray Spectrometry for Kaolin Exploration, *Journal of Applied Geophysics* 53(2-3), p. 91–102.

Vaysse K. and Lagacherie P. (2015). Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional,* 4, p. 20–30.

Viscarra Rossel R.A., Walvoort D.J.J., Mcbratney A.B., Janik L.J. and Skjemstad J.O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma,* 131, p. 59–75.

Viscarra Rossel R.A. and Bui E.N. (2015). A new detailed map of total phosphorus stocks in Australian soil, *Science of The Total Environment*, 542, Part B, 15, p. 1040-1049.

Wackernagel H. (1995). Multivariate geostatistics. Springer—Verlag. p. 255.

Wilford J., Bierwirth P. and Craig M. (1997). Application of Airborne Gamma-Ray Spectrometry in Soil/Regorith Mapping and Applied Geomorphology, AGSO *Journal of Geology and geophysics,* 17(2), p. 201–216.

Winter E.M. (1999). "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data". In Imaging Spectrometry V, July 1999, Denver, Colorado, *Proceeding SPIE*, 3753, p. 266–277.

Zornoza R., Guerrero V., Mataix-Solera J., Scow K.M., Arcenegui V. and Mataix-Beneyto J. (2008). Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biology and Biochemistry*, 40, p. 1923–1930.

Zribi M., Gorrab A., and Baghdadi N., (2014). A new soil roughness parameter for the modelling of radar backscattering over bare soil, *Remote Sensing of Environment*, 152, p. 62-73.

ANNEXE

(Tirés à part)

Tirés à part relatifs au Chapitre II :

C2012.1 Gomez C., Lagacherie P., Bacha S. (2012). “Using an Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia).” Digital Soil Assessments and Beyond. CRC Press 2012, Minasny, Malone & McBratney (eds), 387-392.

AC2016.2 Ouerghemmi, W., Gomez C., Nacer, S., Lagacherie P. (2016). “Semi-blind source separation for estimation of clay content over semi-vegetated areas, from VNIR/SWIR hyperspectral airborne data”. Remote Sensing of Environment. 181, 251-263.

Tirés à part relatifs au Chapitre III :

AC2015.2 Gomez C., Oltra Carrio R., Lagacherie P., Bacha S., Briottet X. (2015) “Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery”. Remote Sensing of Environment, 164, 1–15.

AC2016.3 Gomez C., Gholizadeh A., Borůvka L., Lagacherie P., (2016). “Using legacy data for predicting soil surface clay content from VNIR/SWIR hyperspectral airborne images”. Geoderma, 276, pp 84–92.

Tirés à part relatifs au Chapitre IV :

C2012.2 Ciampalini R., Lagacherie P., Monestiez P, Walker E., Gomez C. (2012). “Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia).” Digital Soil Assessments and Beyond. CRC Press 2012, Minasny, Malone & McBratney (eds), 393-398.

AC2013.3 Lagacherie P., Ruth A., Gomez C., Bacha S., Coulouma G., Hamrouni H, Mekki I. and Bacha S., (2013) “Combining hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia)”. Geoderma, 209–210, pp 168-176.

Digital Soil Assessments and Beyond – Minasny, Malone & McBratney (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62155-7

Using Vis-NIR hyperspectral data to map topsoil properties over bare soils in the Cap Bon region, Tunisia

C. Gomez
IRD, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

P. Lagacherie
INRA, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

S. Bacha
Centre National de Cartographie et de Télédétection (CNCT), El Aouina, Tunis, Tunisia

ABSTRACT: The aim of this work was to examine whether Vis-NIR airborne spectroscopy could be used for mapping topsoil properties in a 300 km² Mediterranean cultivated landscape (Lebna catchment, Tunisia) that includes contrasting pedological patterns and a large proportion of bare soil surfaces. This work employed AISA-Dual Vis-NIR hyperspectral airborne data acquired with a fine spatial resolution (5 m) and fine spectral resolution (260 spectral bands from 450 to 2500 nm). Partial Least Square Regressions were applied to model the relations between soil properties and AISA-Dual spectra. The results showed that four out of the eight soil properties (clay, sand, iron, and cation-exchange capacity) were satisfactorily mapped with good precisions both for estimating local values and for capturing the spatial structures. This study highlights the complexity of the North African soil patterns and opens up the possibility of more extensive use of hyperspectral data for digital soil mapping of the successfully mapped soil properties.

1 INTRODUCTION

Among the large set of possible emerging technologies that could be used in Digital Soil Mapping (DSM), visible and near infrared (Vis-NIR, 350–2500 nm) hyperspectral imaging spectroscopy is one of the most promising methods because (i) it is derived from reflectance spectroscopy, a laboratory technique that has been proven as a good alternative to costly physical and chemical laboratory soil analysis for the estimation of a large range of soil properties (Ben-Dor & Banin, 1995); (ii) it can benefit from the increasing number of methodologies developed for Vis-NIR hyperspectral airborne imaging in soil property mapping (e.g., Selige et al., 2006, Gomez et al., 2008); (iii) it can provide a global view of the area under study at spatial resolutions appropriate for DSM (Gomez et al., under review in Geoderma); and (iv) it is particularly well adapted to Mediterranean and semi-arid areas, where bare soil surfaces are common and where dry periods allow for avoiding soil moisture perturbations of the spectrum (Lagacherie et al., 2008).

The aim of this work was to examine whether Vis-NIR hyperspectral airborne imaging spectroscopy could be used for mapping eight of the most common soil properties (clay, sand, silt, calcium carbonate, free iron, cation-exchange capacity (CEC), organic carbon (OC) and pH) over a 300 km² Mediterranean area. This work investigated the use of the partial least-square regression (PLSR) to construct the models necessary to estimate the soil properties from the Vis-NIR data. The high spatial resolution (5 m) of the imaging data used in this research is expected to provide detailed pattern recognition of the soil's heterogeneity. In addition, the large coverage (300 km²) of the imaging data used in this research is expected to provide a global view of main soil patterns.

2 MATERIAL AND METHODS

2.1 *Study area*

The study area is located in the Cap Bon region in northern Tunisia (36°24′N to 36°53′N; 10°20′E to 10°58′E), 60 km east of Tunis, Tunisia (Figure 1).

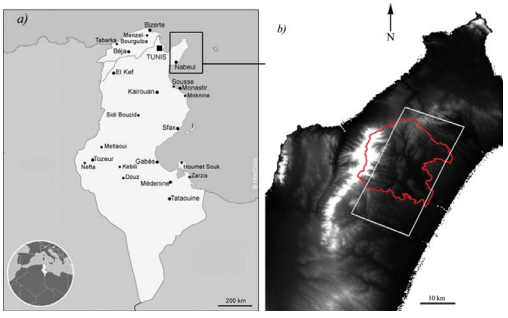


Figure 1. a) Location of the Cap Bon region in the Northern Tunisia, b) limits of the Lebna Catchment (in red) and the AISA hyperspectral image (in grey) plotted over the STRM DEM of the Cap Bon.

This 300 km² area includes the Lebna catchment (Figure 1b), which is mainly rural (>90%) and dominated by cereals in addition to legumes, olive trees, natural vegetation for breeding and vineyards. It is characterized by relief areas, with an altitude between 0 and 226 m. The main soil types are Regosols, Eutric Regosols (9.6%) predominantly associated with sandstone outcrops, Calcic Cambisol, and Vertisol predominantly formed on marl outcrops and lowlands. The southeastern region of the study area represents a more flat landscape with sandy Pliocene deposits yielding Calcosol and Rendzina.

2.2 *AISA-Dual Vis-NIR hyperspectral airborne data*

On November 2, 2010, Vis-NIR AISA-Dual hyperspectral data were acquired over the study area (12 × 25 km) with a spatial resolution of 5 m (Figure 2). The AISA-Dual airborne imaging spectrometer measures the reflected radiance in 359 non-contiguous bands covering the 400- to 2450-nm spectral domain, with 4.6 nm bandwidths between 400 and 970 nm and 6.5 nm bandwidths between 970 and 2450 nm. The instantaneous field of view (IFOV) is 24 degrees. The radiance units were converted to reflectance units using ASD spectrometer measurements of uniform surfaces (parking lots, asphalt, concrete) that were collected at the same time during the over flight. An empirical line correction method was used to calibrate each flight line to the reflectance. Topographic corrections were performed using a 30 m digital elevation model built from ASTER data and ground control points. In this study, we removed: 1) the spectral bands in the blue part of the spectral domain (between 400 and 484 nm) due to noise in these bands and 2) the spectral bands

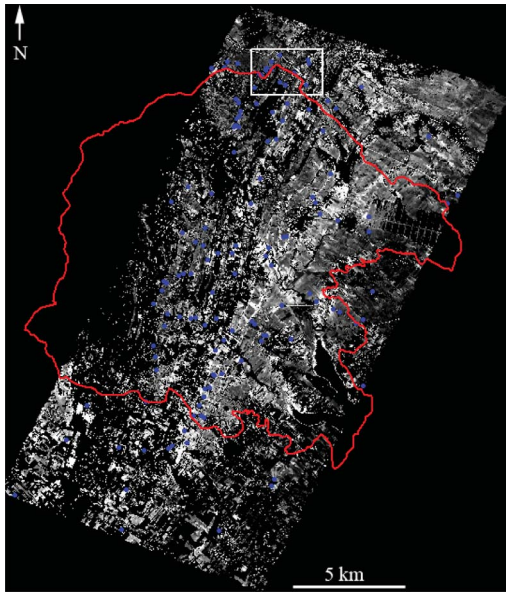


Figure 2. Location of the 129 soil samples (blue points) collected over bare soils and the Kamech catchment (full white square) plotted over the AISA-Dual image (12 × 25 km). Only the bare soils are represented (water, urban, vegetation area and mixed pixels are represented in black).

between 1339 and 1464 nm as well as between 1772 and 2004 nm due to vibrational-rotational H2O absorption bands. Consequently, 280 AISA-Dual spectral bands were retained.

When the image was acquired (November 2010), a minor part of the soil surface was covered by green vegetation, consisting mainly of olive trees, native forests, green plants and vineyards. To isolate the bare soil areas, pixels with normalized difference vegetation index (NDVI) values over an expert-calibrated threshold were masked: a value of 0.20 was determined after considering twenty parcels that had been visually inspected on the field. Water areas were also masked using an expert-calibrated threshold: pixels with a reflectance of less than 8% at 1665 nm were removed. Finally, urban areas were masked using a map of urban areas.

Based on the AISA-Dual data with a spatial resolution of 5 m, the bare soils represent 46.3% of our study area and 5 889 847 AISA-Dual pixels (Figure 2).

2.3 *Field sampling*

129 soil samples were collected on the Lebna catchment: 58 were collected in October 2008, 30 in

October 2009, and 41 in November 2010. All of these soil samples were collected in fields that were bare during the hyperspectral data acquisition in November 2010 (Figure 2). All of the samples were composed of five sub-samples collected to a depth of 5 cm at random locations within a 10 × 10 m square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument.

The samples were air-dried and sieved with a 2 mm sieve. After homogenizing the sample, approximately 20 g was used for soil property analysis. The determination of eight soil properties was performed using classical physico-chemical soil analysis. These properties were free iron, cation-exchange capacity (CEC), clay (granulometric fraction <2 µm), silt (granulometric fraction between 2 and 50 µm), sand (granulometric fraction between 0.05 and 2 mm), calcium carbonate ($CaCO_3$), pH and organic carbon (OC).

2.4 *Prediction models*

The Partial Least Square Regression (PLSR) was used to establish relationships between the soil variables and the AISA-Dual spectra. Before this multivariate analysis, the reflectance was converted into "absorbance" (log [1/reflectance]) and a noise reduction was achieved through standard pretreatments: a Savitzky–Golay filter with second-order polynomial smoothing and window widths of 30 nm, mean centering and variance scaling. The spectroscopic and chemometric analyses were implemented in R. The maximum number of latent variables (LV) in the PLSR was defined as 10. The optimal number of LVs was determined using prediction residual error sum of squares (PRESS) analysis, taking care to avoid under—and over-fitting. A prediction model was built for each soil property. Because a limited number of samples was available, a leave-one-out cross-validation procedure was adopted to verify the prediction capability of the PLS models. Two types of outliers were rejected from the calibration set: (i) concentration outlier, when the predicted value has a residual difference significantly greater (>2.5) than the mean of the predicted values and (ii) spectral outlier, when the sample is spectrally different from the rest of the samples. An H value of 3 based on the Mahalanobis distance, calculated on PCA-reduced data, was selected for the identification of spectral outliers.

2.5 *Models evaluation*

The prediction models were evaluated using the root mean squared error of cross-validation (RMSECV), the coefficient of determination of cross-validation (R^2cv) and the ratio of performance deviation (RPD). The RPD is the ratio between the standard deviation of the entire data set against the RMSECV. Chang and Laird (2002) defined three classes of RPD: category A (RPD > 2) describes models that can accurately predict the soil property, category B (2 > RPD > 1.4) describes models with limited predictive power and category C (RPD < 1.4) describes models that have no prediction ability.

3 RESULTS

3.1 *Soil samples study*

The soil properties exhibited contrasting variations over the study area (Table 1). The clay, sand, CEC and iron were very variable, whereas silt, $CaCO_3$, pH and carbon exhibited much smaller variations with sometimes asymmetric distributions (pH and $CaCO_3$). Most of the soil properties of these 129 soil samples were not correlated to each other, with the exception of: (i) high positive correlation between the clay content and CEC (R^2 = 0.9), (ii) high negative correlation between the clay and sand contents

Table 1. Statistical parameters of soils properties for the 129 soils samples.

	Min	Max	Mean	SD*
Iron (g/100 g)	0,3	3,3	1,6	0,5
CEC (cmol + /kg)	2,8	34,1	19,5	6,6
Clay (g/kg)	46	772	463,7	175,9
Sand (g/kg)	32	896	326	202,3
Silt (g/kg)	58	429	210,2	58,7
$CaCO_3$ (g/kg)	1	346	36,9	52,3
pH	5,8	8,8	8,2	0,5
OC (g/kg)	2,7	21,8	8,9	2,8

* *Standard deviation*.

Table 2. Correlation coefficient between the soil properties, calculated from the 129 soil samples.

	Clay	Silt	Sand	OC	pH	$CaCO_3$	CEC
Silt	0,31						
Sand	**–0,96**	**–0,56**					
OC	0,14	0,21	–0,18				
pH	0,34	0,23	–0,37	–0,11			
$CaCO_3$	–0,10	0,00	0,09	0,22	0,26		
CEC	**0,96**	0,37	**–0,94**	0,22	0,34	–0,09	
Fer	**0,77**	0,31	**–0,76**	0,07	0,08	–0,25	**0,71**

(R^2 = 0.9) and (iii) positive correlation between iron and sand contents as well as between iron and clay contents (R^2 = 0.6) (Table 2). The ranges of clay, sand, and iron contents and CEC are large (Table 1) and exhibit a centered normal distribution. Whereas the ranges of the four other properties were quite small (Table 1) and exhibited Poisson distribution.

3.2 *Prediction models results*

PLSR-based prediction models were built using the 129 AISA-Dual spectra corresponding to the location of the soil samples collected over bare soils (Figure 2). Two spectral outliers were identified and removed, and the number of concentration outliers depended on the soil property and varied between 0 and 7. The elimination of outliers from the soil database modified significantly the concentration range of the soil properties for $CaCO_3$, silt, pH and OC (Table 3).

Correct prediction models (category B), with R^2 and RPD values greater than 0.6 and 1.4 respectively, were obtained for four soil properties: iron, CEC, clay and sand contents (Figure 3). The prediction models were inaccurate for silt, $CaCO_3$, pH and OC, with R^2 values less than 0.35 (Figure 3).

3.3 *Predicted maps*

The prediction of soil properties was performed for all the bare soils of the AISA-Dual image. Only the soil properties, for which correct local predictions (Figure 3) were obtained, were mapped. Thus we created digital maps for four soil properties: free iron, CEC, clay and sand.

The predicted clay map of the entire study area showed a complex regional soil pattern (Figure 4), with predominant variations in lithology. Differences both in values and soil patterns appeared between the Pliocene area, located in the southeast corner of the image, and the Miocene area, covering the rest of the image. Pliocene area exhibits low and weakly variable topsoil clay contents, whereas the Miocene area shows a large range of clay content values. Variations within the Miocene area are also visible. They follow the geological pattern formed by the alternating sandstone and marl outcrops, yielding low (blue) and high (red) values of clay content, respectively (Figure 4). The soil patterns vary across the southeast/northwest direction, with a decreasing distance between successive sandstone outcrops and the occurrence of a large sandstone outcrop in the middle. The deposition of sandy material from the erosion of sandstone areas in the valleys that

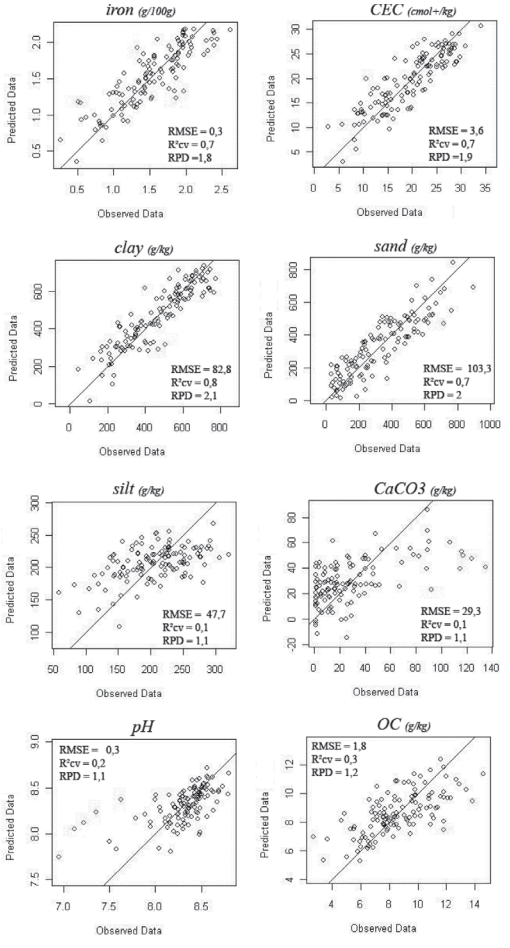


Figure 3. Plots of observed versus predicted soil properties obtained by PLSR using AISA-Dual spectra of the 129 soil samples location.

Table 3. Statistical parameters of soils properties for the soil samples used in the PLSR models after outlier removal.

	Outlier[#]	Min	Max	Mean	SD*
Iron (g/100 g)	7	0,26	2,62	1,5	0,48
CEC (cmol + /kg)	0	2,8	34,1	19,5	6,6
Clay (g/kg)	1	46	772	467	176
Sand (g/kg)	3	32	896	327	204
Silt (g/kg)	3	58	321	206	51,3
$CaCO_3$ (g/kg)	5	1	135	29,3	31
pH	6	6,9	8,8	8,3	0,3
OC (g/kg)	2	2,7	14,6	8,6	2,2

[#] *Number of concentration outliers.*

[*] *Standard deviation.*

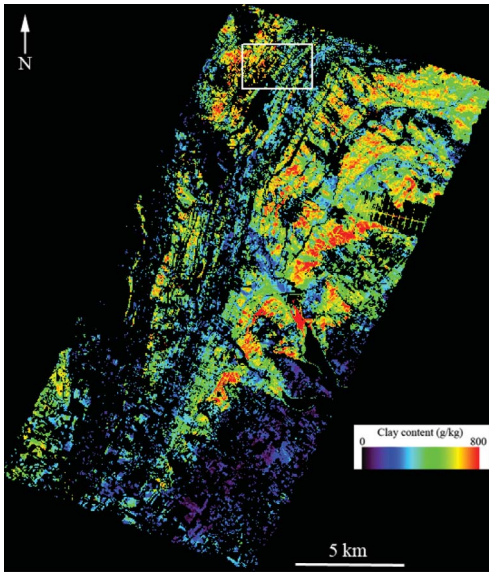


Figure 4. Clay content predicted over bare soils from AISA—Dual spectra (black areas correspond to mixed surfaces). The white square delimits the Kamech catchment.

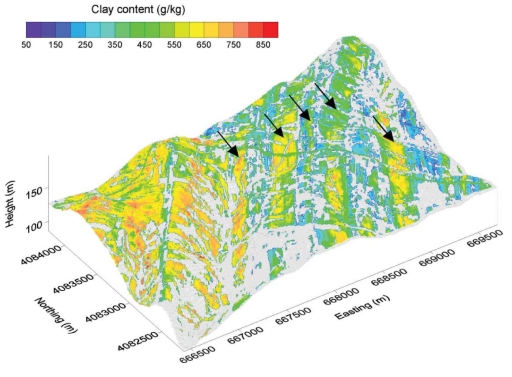


Figure 5. Zoom on predicted clay content map over bare soils of the Kamech catchment (grey areas correspond to mixed surfaces). Black arrows represent clay-rich areas. Coordinates are in UTM WGS 84.

are perpendicular to the outcrops add to the complexity of the regional soil pattern.

The predicted clay map over the Kamech catchment (Figure 5) is representative of the alternating sandstone and marl outcrops (highlighted by arrows in Figure 5). Some mixed areas also appear in transition areas between these outcrops and in shoal areas. In the Northeast corner of the study area (Figure 4) similar successions of Marl and sandstone outcrops were observed with larger distances between successive outcrops.

4 DISCUSSION

The local soil property prediction models, built from AISA-Dual Vis-NIR spectra using the 129 soil samples, allowed the estimation of four out of the eight soil properties with respect to common quality indicators (Figure 3). The accuracy of the prediction model of clay contents (R^2_{cv} = 0.8, Figure 3) is in agreement with that presented by Selige et al. (2006) and Gomez et al. (2008). The accuracy of the prediction model of sand contents (R^2_{CV} = 0.7, Figure 3) is in agreement with that reported by Selige et al. (2006). The inaccuracy of the prediction models of $CaCO_3$ contents (R^2_{CV} = 0.1, Figure 3) differs from that of Gomez et al. (2008). Our lower accuracy may be due to the low $CaCO_3$ contents in the calibration data set (between 1 to 135 g/kg after removing the outliers, Table 3) as compared to $CaCO_3$ contents of Gomez et al. (2008) (between 0.26 to 472 g/kg). The inaccurate results obtained for the prediction of OC contents differ from those reported by Stevens et al. (2010). Our lower accuracy may be due to the low variability in OC content (between 2 and 14.6 g/kg after removing the outliers, Table 3) in the calibration data set compared to the OC contents of Stevens et al. (2010) (between 5 to 50 g/kg). These poorest results can also be explained by the lower number of samples used to calibrate our model (129) compared to Stevens et al. (2010) (306 samples). Finally, no predictions based on hyperspectral data for free iron, CEC, pH and silt contents are available in the literature, so literature references cannot be used to compare and evaluate our results for these soil properties.

The large coverage (300 km²) of our AISA-Dual imaging data provides a global view of the main soil patterns (Figure 4). Successions of sandstone outcrops and marl outcrops are mapped, with a decrease of the distance between these successive outcrops from East to West. A sandy area is mapped in the Southeast part of the study area, which corresponds to Pliocene. At a local scale, the high spatial resolution (5 m) of our AISA-Dual imaging data provides detailed pattern recognition of the soil's heterogeneity, in particular for the alternance of sandstone and marl outcrops.

As correlations exist between clay, CEC, iron and sand properties (Table 2), predicted maps of these four soil properties are highly correlated as well. Nevertheless, in addition to these soil properties maps, a map of the textural class could be

obtained from the synthesis of the predicted clay and sand maps.

5 CONCLUSION

This study demonstrated that Vis-NIR hyperspectral imaging data can be used to map several key topsoil properties over large areas of bare soil. In the future, this new spatial information on topsoil properties should be used in Digital Soil Mapping both for generating complete maps of soil properties (Ciampalini et al., Submitted in DSM2012) and for improving the digital soil mapping of related subsoil properties. Moreover, diverse surface conditions including partially vegetated surfaces should be considered and treated to increase the surface of key soil properties mapping. A first way could be to use source separation methods as shown by Ouerghemmi et al. (2011). Finally the development of Vis-NIR hyperspectral sensors which are planned to be launched on board satellites within the next two years, such as PRISMA and EnMap, will extend the use of Vis-NIR hyperspectral imaging data in Digital Soil Mapping.

ACKNOWLEDGEMENTS

The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National de Cartographie et de Télédétection, Tunisia), for providing the AISA-Dual images for this study. This hyperspectral data acquisition was granted by IRD, INRA and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01) ». We are also indebted to Yves Blanca (IRD-UMR LISAH Montpellier), Zakia Jenhaoui (IRD-UMR LISAH Tunis) for the soil sampling in 2009 and 2010 over the Lebna catchment and to Hedi Hamrouni (DG/ACTA Sol, Tunis) for his significant support to this study.

REFERENCES

Ben-Dor, E. & Banin, A. (1995). Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate, several soil properties. *Soil Science Society of American Journal*, 59, 364–372.

Chang, C.-W. & Laird, D.A. 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, 167 (2), 110–116.

Ciampalini, R., Lagacherie, P., Monestier, P., Walker E. & Gomez, C. (2012). Using a Vis-NIR hypersepctral image as a covariate for mapping properties in Northern Tunisia. Submitted at the DSM 2012 Conference Proceedings.

Gomez, C., Lagacherie, P. & Coulouma, G. 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, 148 (2), 141–148.

Gomez, C., Lagacherie, P. & Coulouma, G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Under Review in *Geoderma*.

Lagacherie, P., Baret, F., Feret, J-B, Madeira Netto, J. & Robbez-Masson, J.-M. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, 112 (3), 825–835.

Ouerghemmi, W., Gomez, C. Nacer, S. & Lagacherie, P. (2011). Applying Blind Source Separation on hyperspectral data for clay content estimation over partially vegetated surfaces. *Geoderma*, 163 (3–4), 227–237.

Selige, T., Bohner, J. & Schmidhalter, U. 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, 136, no1–2, pp. 235–244.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L. & Wesemael, B. (2010). Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy, *Geoderma*, 158, 1–2.

Remote Sensing of Environment 181 (2016) 251–263

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Semi-blind source separation for the estimation of the clay content over semi-vegetated areas using VNIR/SWIR hyperspectral airborne data

W. Ouerghemmi [a,b,*], C. Gomez [a,**], S. Naceur [b], P. Lagacherie [c]

[a] IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France
[b] LTSIRS, Laboratoire de Télédétection et Systèmes d'Information à Référence Spatiale, ENIT, Tunisia
[c] INRA, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form 8 April 2016
Accepted 16 April 2016
Available online xxxx

Keywords:
Blind source separation
Non-negative matrix factorization
Partial least squares regression
Clay content
Semi-vegetated pixels
Hyperspectral imaging
Spectroscopy

ABSTRACT

Visible, near-infrared and short wave infrared (VNIR/SWIR) hyperspectral imagery has proven to be a useful technique for mapping the soil surface properties over bare soils pixels. Multivariate regression models are usually built linking a set of soil surface properties (response Y-variables) to a set of imaging reflectance spectra over bare soil pixels (predictor X-variables), and then, they are applied to all bare soil pixels to map the soil surface properties. The applicability of VNIR/SWIR hyperspectral imagery for soil properties mapping decreases when surfaces are partially covered by vegetation. The objective of this research was to develop a "Double-Extraction" approach for clay content estimation over semi-vegetated surfaces and to evaluate its performance using VNIR/SWIR HyMap airborne data acquired in a Mediterranean region over an area of 24 km^2. The "Double-Extraction" approach consists of *1*) an extraction of a soil reflectance spectrum, $\hat{s}_{soil}$, using a semi-blind source separation (SBSS) technique applied to couples of semi-vegetated spectra and *2*) an extraction of clay content from the soil reflectance spectrum $\hat{s}_{soil}$ using a multivariate regression method. The source separation approach is semi-blind due to the use of available knowledge about expected soil and vegetation spectra. The multiplicative algorithm of Lee & Seung, belonging to the family of non-negative matrix factorization (NMF) methods, is used to solve the blind source separation (BSS) problem. The multivariate regression method used in this study is the partial least squares regression (PLSR) method. The "Double-Extraction" approach was compared to a "Direct" approach consisting of the application of the multivariate regression model built from bare soil spectra over the semi-vegetated area.

Our results showed poor prediction performances for both approaches when applied to all pixels; however, a slight improvement was observed when correcting the bias prediction that occurs when using the PLSR model. Conversely, satisfactory prediction performances were obtained by restricting the prediction to the weakly vegetated area (NDVI < 0.55) that covered 63% of the study area. The resulting clay map over this restricted vegetated area exhibited patterns of variations that matched the previous expertise acquired on the spatial structures of soils in this area.

1. Introduction

The uses of visible, near-infrared and short wave infrared (VNIR/SWIR) hyperspectral imaging spectroscopy in soil science have been widely developed in recent years; in particular, an increasing number of soil properties estimation techniques have been developed for laboratory spectroscopy, including partial least square regression (e.g., Gomez, Viscarra Rossel, & McBratney, 2008a), support vector machine regression (e.g., Stevens et al., 2010), multiple linear regression (e.g., Bayer, Bachmann, Müller, & Kaufmann, 2012) and stepwise multiple linear regression (e.g., Lu, Wang, Niu, Li, & Zhang, 2013). Several soil properties, such as organic carbon (e.g., Stevens et al., 2010), clay content (e.g., Chabrillat, Goetz, Krosley, & Olsen, 2002), soil salinity (e.g. Dehaan & Taylor, 2001, Taylor & Dehaan, 2000), and carbonate calcium (e.g., Lagacherie, Baret, Feret, Madeira Netto, & Robbez-Masson, 2008), have been successfully mapped using VNIR/SWIR imaging data. Nevertheless, mapping of these soil properties were limited to pure pixels of bare soil surfaces (e.g., Gomez, Lagacherie, & Coulouma, 2012). Currently, none of the soil properties can be

* Correspondence to: W. Ouerghemmi, IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France.
** Corresponding author.
E-mail addresses: walid.ouerg@gmail.com (W. Ouerghemmi), cecile.gomez@ird.fr (C. Gomez).

http://dx.doi.org/10.1016/j.rse.2016.04.013
0034-4257/

estimated using VNIR/SWIR imaging data over partially vegetated (or mixed) surfaces. The "no-soil" components of the pixels affect the spectral signatures, thereby affecting the performances of the soil properties estimation models (Bartholomeus et al., 2011).

In the remote sensing community, unmixing problems are frequent, and several unmixing approaches have been developed to find the spectra and abundances of pure component present in each of the mixed pixels. These unmixing approaches can be classified into three families: *i*) geometric, *ii*) sparse regression, and *iii*) statistical methods. Geometric methods are based on a projection of the mixed spectra into a simplex, for which the vertices correspond to the pure endmembers. Some of the popular geometrical approaches are the Pixel Purity Index (PPI) (Boardman, 1993), Vertex Component Analysis (VCA) (Nascimento & Bioucas-Dias, 2005) and *N*-FINDR (Winter, 1999). The efficiency of the geometrical approaches depends on the existence of pure pixels corresponding to the different endmembers existing in the study scene (Bioucas-Dias et al., 2012). Sparse regression methods are based on the modeling of mixed spectra, using a large reference spectral database (e.g., Iordache, Bioucas-Dias, & Plaza, 2011). Some of the popular sparse regression methods are the least angle regression (Efron, Hastie, Johnstone, & Tibshirani, 2004), the basis pursuit (Chen, Donoho, & Saunders, 2001), and the matching pursuit (Mallat & Zhang, 1993). The efficiency of the sparse regression methods depends on the constitution of the reference spectral database. Finally, statistical methods are based on the blind source separation concept, which was initially designed for signal processing (Comon & Jutten, 2010) and was then adapted to a multidimensional imagery application (e.g.Caiafa, Salerno, & Proto, 2007, Meganem, Deville, Hosseini, Deliot, & Briottet, 2014). These statistical methods aim to recover the spectra of the pure components from the mixed spectra without any information regarding the proportion of the pure components in the mixture or the original spectra that composed the mixed spectra. Some of the popular statistical methods are independent component analysis (ICA, e.g., Jutten & Herault, 1991; Comon, 1994; Hyvärinen & Oja, 2000), non-negative matrix factorization (NMF, e.g., Leggett, 1977; Paatero & Tapper, 1994; Lee & Seung, 1999, 2001) and sparse component analysis (SCA, e.g., Lin, Grier, & Cowan, 1997; Lee, Lewicki, Girolami, & Sejnowski, 1999; Zibulevsky & Pearlmutter, 2001; Starck, Moudden, Bobin, Elad, & Donoho, 2005).

Until now, few studies have addressed the estimation of soil properties over mixed pixels, including semi-vegetated pixels, by overcoming the problem of mixed spectra. Bartholomeus et al. (2011) and Ouerghemmi, Gomez, Naceur, and Lagacherie (2011) proposed an original approach of soil properties estimation from mixed spectra. The term "Double-Extraction" was introduced by Ouerghemmi et al. (2011), to characterize this approach which consisting of 1) a first extraction of an estimated soil spectrum, $\hat{s}_{soil}$, from mixed VNIR/SWIR spectra *X* and 2) a second extraction of soil property content from the estimated soil spectrum $\hat{s}_{soil}$. Bartholomeus et al. (2011) estimated soil organic carbon (SOC) content from spectral mixtures of soil and maize. The first extraction of the estimated soil spectra $\hat{s}_{soil}$ was performed by a linear inversion of the mixed VNIR/SWIR spectra *X*, and the second extraction of SOC was performed by applying a PLSR model to the estimated soil spectra $\hat{s}_{soil}$. The linear inversion process was done assuming the presence of two endmembers (soil and maize) in the studied mixtures, and thanks to the knowledge of the spectrum of the maize and the endmembers abundance fractions. Thus, this unmixing method requires a representative spectrum of a "no-soil" endmember and reliable abundance endmember fractions for each of the mixed pixels. Ouerghemmi et al. (2011) estimated the clay content from spectral mixtures of soil and vineyard. The first extraction of the estimated soil spectra $\hat{s}_{soil}$ was performed by using an ICA method, which is one of the most popular statistical methods for unmixing, and the second extraction of clay contents was performed by applying PLSR models to the estimated soil spectrum $\hat{s}_{soil}$. The major deficit of this unmixing development is due to the ICA hypothesis (i.e., independence of the endmembers spectra), which is not in accordance with reality. Finally, the performances of both studies (Bartholomeus et al., 2011Ouerghemmi et al., 2011) were encouraging but should be improved by more adequate algorithm developments.

In this paper, we propose a new "Double-Extraction" approach to estimate the clay content from the spectral mixtures of soil and vineyard. The first extraction of the estimated soil spectra $\hat{s}_{soil}$ is performed by using an NMF method which was previously successfully used in different source separation applications (e.g.Lu, Wu, & Yuan, 2014, Meganem et al., 2014). The second extraction of the clay contents is performed by applying PLSR models to the estimated soil spectrum $\hat{s}_{soil}$. Due to the positivity constraint imposed on the sources and the abundances, which is in accordance with hyperspectral imagery characteristics (spectral signatures and their relative abundances are positive), the NMF methods are more suitable than the ICA ones (e.g., Pauca, Piper, & Plemmons, 2006) to estimate the endmembers in the hyperspectral context. Moreover, to avoid poor convergences of the Lee & Seung NMF algorithm due to its random initialization (e.g.Cichocki, Zdunek, Phan, & Amari, 2009, Donoho & Stodden, 2003), we use field knowledge to initialize the algorithm and makes it "semi-blind". The HyMap hyperspectral data and the field soil data used in this paper were previously studied for the prediction of clay content over bare soils (Gomez et al., 2012Lagacherie et al., 2008).

The study area, airborne, laboratory and field data are described in Section 2. The methodologies, including the "Double-Extraction" approach, are presented in Section 3. Finally, the results are presented in Section 4 and discussed in Section 5.

2. Data

2.1. Study area

The study area was located on the southern part of the La Peyne catchment (43°29′ N and 3°22′ E), 60 km west of Montpellier in France (Fig. 1). The acreage of the study area is 24.6 km^2, and the area is largely devoted to vineyards. The climate is typical of the Mediterranean region, characterized by sub-humid to prolonged dry seasons. The annual rainfall median over 20 years is 634 mm, and the annual evapotranspiration average is 1102 mm. In the upstream region of the catchment, the rate of cultivated land is poor due to steep terrain slopes and Mediterranean maquis shrubs, whereas in the downstream region, moderate terrain slopes contribute to the predominant agricultural land use devoted to vineyards. At the surface, the soil appears crusted due to tillage practices and strong episodic rainfalls. Beneath the surface, the soil substrate is largely influenced by heterogeneous Miocene marine sediments, i.e., marl, sandy loam, and calcareous sandstone with low carbon content (<2%). Thin Miocene lacustrine limestones compose the structure of the hillslopes in the middle of the La Peyne valley, which is representative of cuesta topography. Hill backslopes are partially overlaid by successive alluvial deposits, ranging from Pliocene to Holocene, and differ in their initial nature and in the duration of weathering conditions. The presence of clay minerals is dominated by illite and kaolinite, with a weak abundance of an illite-smectite mixture. The soils of the study area have been classified into five soilscapes that were defined from lithological variations (IUSS Working Group WRB, 2006). The summary description of the five soilscapes can be found in Gomez et al. (2012).

2.2. VNIR/SWIR HyMap data

A HyMap hyperspectral image with a spatial resolution of 5 × 5 m was acquired over the study area on 13 July 2003 at 12:15 pm from

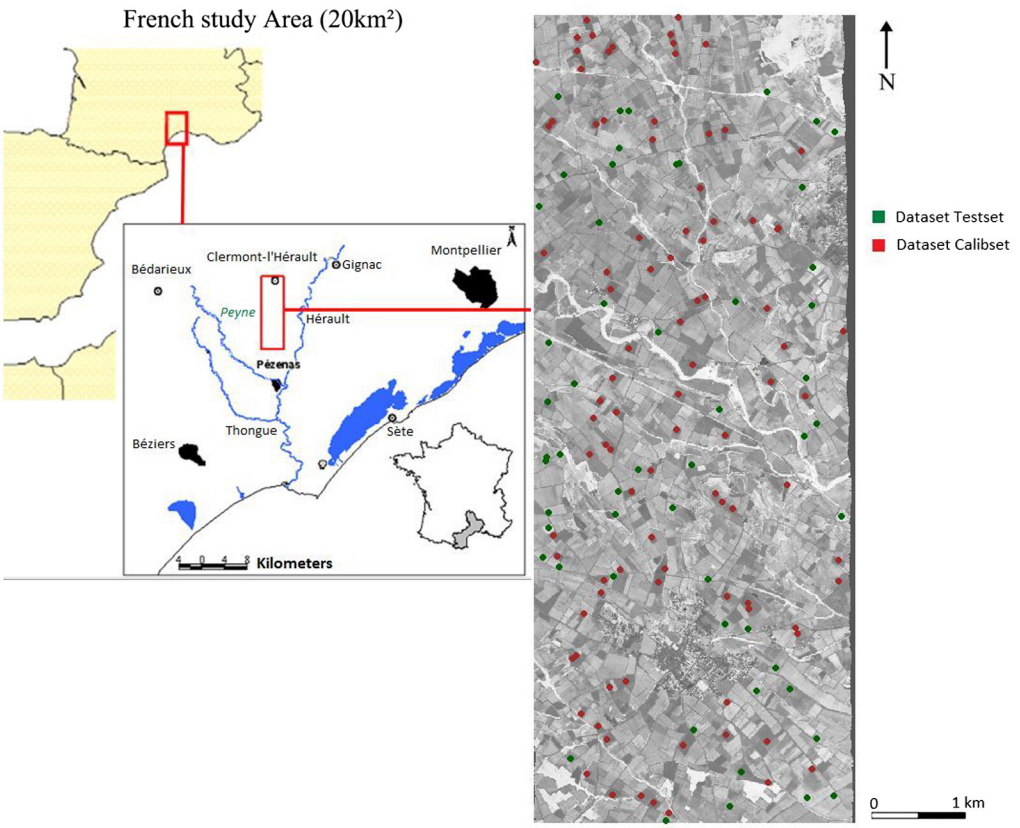


Fig. 1. HyMap hyperspectral image over the La Peyne Catchment in France (in red the 95 soil samples, in green the 54 soil samples).

an altitude of 2680 m (Fig. 1). The HyMap airborne imaging spectrometer measures the reflected radiance in 126 non-contiguous bands covering the 400–2500 nm spectral domain with a 19 nm bandwidth and average sampling intervals of 17 nm between 1950 and 2480 nm. The instantaneous field of view (IFOV) is 2.5 mrad along the track and 2.0 mrad across the track for a 60° field-of-view over flat terrain. HyMap provides a signal-to-noise ratio (SNR) >500:1. This image has been geometrically, atmospherically and topographically corrected by the German Aerospace Center. Radiometric HyMap calibration was performed in-flight (Richter, 1996) using the noon image and nadir ground measurements (Beisl, 2001). The ATCOR4 program for airborne sensors was used to convert the at-sensor radiance to the surface reflectance and for atmospheric corrections of the image (Richter, 2000). Topographic corrections were performed using a high-resolution digital elevation model (25 m) from the French Institut Géographique National (IGN) and ground control points furnished by the Faculty of Geographical Sciences – Utrecht University. In this study, a total of 12 spectral bands were removed from the 126 original bands: 2 noisy bands (at 438 and 2483 nm), 8 bands from 1404 to 1503 nm due to H_2O band absorption and 2 bands at 1951 and 1970 nm, also due to H_2O band absorption. A total of 114 spectral bands from 450 to 2468 nm were retained.

At the time the image was acquired (July 2003), most of the soil surface was covered by green vegetation, consisting mainly of vineyards and native forests or crop residues. To mask the bare soil pixels, we used the process described by Madeira Netto, Robbez-Masson, and Martins (2007). First, the pixels with normalized difference vegetation index (NDVI) values under an expert-calibrated threshold were masked. This "expert-calibrated" threshold was determined after considering 1) more than thirty visually inspected parcels in the field during the flight acquisition and 2) the "salt-and-pepper" patchiness of the resulting mask. A value of 0.26 was used for the threshold. Second, we used the absorption band centered at 2100 nm, which is related to *O*—H and *C*—O bonds in cellulose (Kokaly & Clark, 1999), to retain the areas with significant dry vegetation cover. Finally, the semi-vegetated areas represent 88% of our study area.

Finally, the NDVI varied between 0.26 and 0.91 over the semi-vegetated pixels. In addition, the cellulose absorption index (CAI), which described the depth of the lignocellulose absorption feature in the shortwave infrared region (2.0–2.2 μm) (Daughtry, McMurtrey, Nagler, Kim, & Chappelle, 1996), varied between −628 and 530 over the semi-vegetated pixels.

2.3. Field soil and vegetation data

One hundred and forty-nine soil samples were collected in the study area. The soil samples were composed of five sub-samples that were collected within a 10-m wide square centered on the geographical position of a sampling plot. A subset of data including 95 of these soil samples (further denoted by *CalibSet*) were sampled in fields that were bare during the hyperspectral data acquisition (red points in Fig. 1). A second dataset (further denoted as the *TestSet*) included the remaining 54 samples, which were collected in fields that were partially covered by vineyard during the hyperspectral data acquisition (green points in Fig. 1).

The 54 semi-vegetated sites were characterized by a normalized difference vegetation index (NDVI, Rouse, Haas, Schell, & Deering, 1973) and a cellulose absorption index (CAI, Daughtry et al., 1996) between 0.28 and 0.72 and −247.5 and 31.5, respectively, and both followed a distribution close to normality (Fig. 2b and c, respectively). The clay content (granulometric fraction <2 μm) was determined using a pipette method according to the method

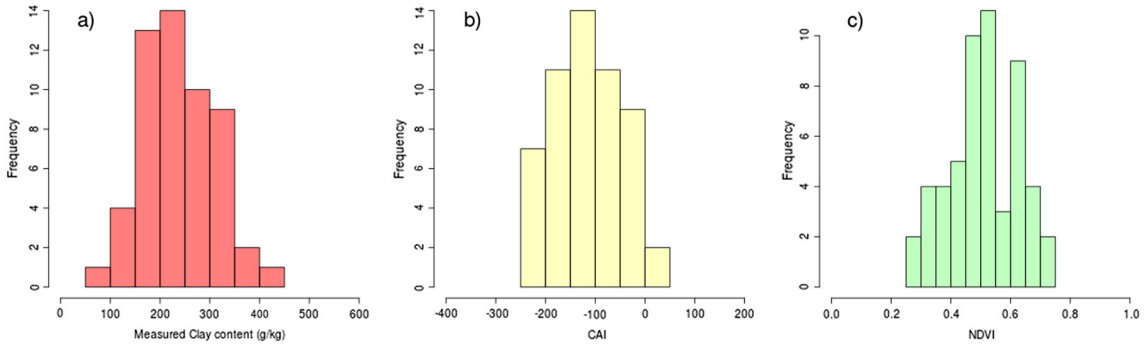


Fig. 2. Distribution of the a) measured clay content (g/kg) of the 54 soil samples (*TestSet*), and b) CAI and c) NDVI of the 54 semi-vegetated sites.

NF × 31–107 (Baize & Jabiol, 1995). The clay content of the *CalibSet* and the *TestSet* varied between 73 and 397 g/kg and 85 and 418 g/kg, respectively, and both followed a distribution close to normality (Fig. 2a).

In addition, vineyard leaves were collected over the study area. Two reflectance spectra of the vineyard leaves were obtained in the laboratory in the 350–2500 nm wavelength domain using an ASD pro FR portable spectroradiometer. The spectral resolution ranged from 3 nm (at 700 nm) to 10 nm (in the SWIR). The spectral sampling interval varies between 1.4 nm (from 350 nm to 1000 nm) and 2 nm (from 1000 to 2500 nm). The data were resampled to 1 nm output values. The ASD measurements had an SNR better than 1000 at its original spectral resolution. The measurements were performed using a black room to control the irradiance conditions. Four 50-W tungsten halogen light sources with aluminum reflectors (24 V, ~3000 K color temperature, and a DC power supply) were placed on each side of the sample, with the light beam directed 45° from vertical. The sensor was positioned from the nadir at a distance of 22 cm from the sample, providing an 8 cm diameter spot using the 8° lens. A white Spectralon panel (12 × 12 cm Labsphere, North Sutton, USA) provided the absolute reflectance factors for the laboratory measurements. Both resulting vineyard spectra were then resampled to the 114 HyMap spectral bands (from 450 to 2468 nm) instead of the 2151 lab ASD spectral bands to be used as simulated HyMap spectra of vineyard.

3. Methodology

This section first presents the implementation of the "Double-Extraction" scheme for clay content prediction over semi-vegetated pixels (Section 3.1). Next, the statistical methods of signal processing used in this study are presented from Section 3.2 to Section 3.5. Finally, the performance criteria selected to evaluate the "Double-Extraction" approach are presented (Section 3.6).

3.1. Methodology implementation via a "double extraction" scheme

In this study, we propose a "Double-Extraction" approach that combines a Semi-Blind Source Separation (SBSS) method with a Partial Least Squares Regression (PLSR) method. The goal of this approach was to predict clay content over semi-vegetated areas. The "Double-Extraction" approach of the information for clay content prediction was tested on the 54 semi-vegetated sites corresponding to the *TestSet*. For each semi-vegetated site, the methodology considered a grid of 3 × 3 pixels, which was supposed to be homogeneous in terms of the vegetation type and the clay content and centered on the location of the field sample.

The "Double-Extraction" of information for clay content prediction is described in Fig. 3. The approach consists first, for each site, of a selection of two spectra, including the central pixel spectrum (Fig. 3a, step 1). Next, these two spectra were used as the input of the SBSS method to estimate two new spectra, one of them being the estimated soil spectrum (Fig. 3a, step 2, and described in Sections 3.1 and 3.2). After identification of this soil spectrum among outputs of the SBSS method (Fig. 3a, step 3, and described in Section 3.2), the clay content was predicted using a PLSR model f_{clay} (Fig. 3a, step 4, and described in Section 3.4) and was corrected from the bias (Fig. 3a, step 5, and described in Section 3.5). This process was repeated for the eight couples from each site (Fig. 3a). As a result, eight clay content predictions for each semi-vegetated site were obtained, averaged and attributed to the central pixel.

We also tested the clay content prediction from the central pixel (Fig. 3b, step 1), using only the PLSR model f_{clay} (Fig. 3b, step 2, and described in Section 3.4) and a bias correction process (Fig. 3b, step 3, and described in Section 3.5). This prediction process was called the "Direct" Approach.

3.2. Blind source separation approach

Blind source separation (BSS) is a popular thematic area related to the signal and image processing fields and was initially formulated in the 1980s (Ans, Hérault, & Jutten, 1985; Jutten & Herault, 1991). The goal of the BSS technique is to recover the original signals from their mixtures. The term "blind" refers to the fact that we have no knowledge of how the sources are mixed and we have no knowledge about the sources themselves. Considering a linear mixture, m observed signals x_i are linear combinations of n signals s_j (with $m \geq n$), which can be written as follows:

$$x_i(\lambda) = \sum_{j=1}^{n} a_{ij} s_j(\lambda) + \alpha(\lambda). \quad (1)$$

In the matrix notation, this equation becomes

$$X(\lambda) = AS(\lambda) + \alpha(\lambda) \quad (2)$$

where λ is the index of the spectral band, $X(\lambda) = [x_1(\lambda)...x_m(\lambda)]^T$ are the reflectance spectra of a mixture, $S(\lambda) = [s_1(\lambda)...s_n(\lambda)]^T$ are the reflectance spectra of the original components, A is an m x n mixing matrix and $\alpha(\lambda) = [\alpha_1(\lambda)... \alpha_m(\lambda)]^T$ is the noise component. The aim of the BSS method is to solve (1) when S and A are unknown, where $S = A^{-1}X$. The noise term α is neglected to consider a simple model.

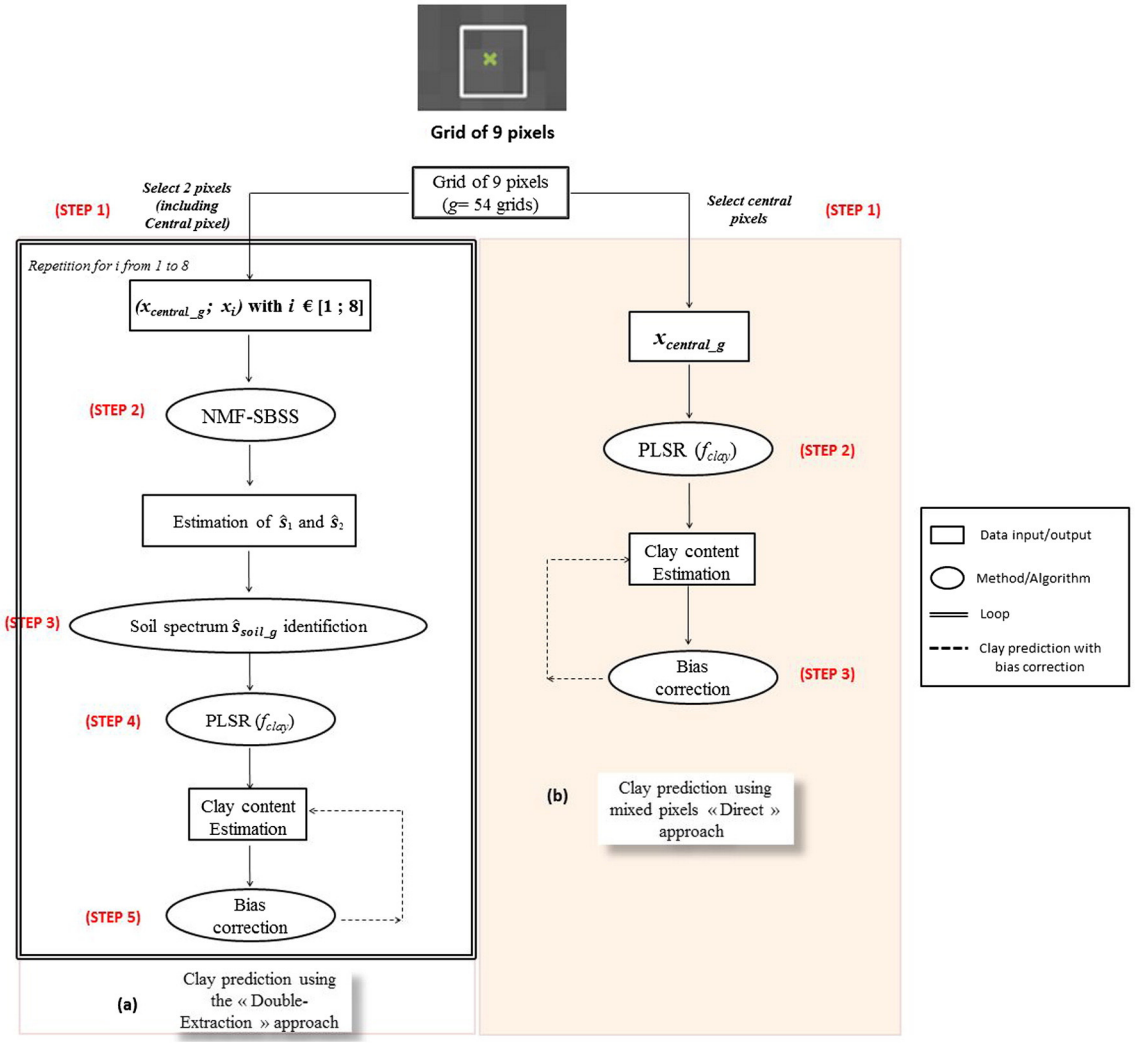


Fig. 3. Procedure overview for clay content prediction using (a) the "Double-Extraction" approach and (b) the "Direct" approach.

Depending on the mathematic hypothesis given on the sources signals, the BSS methods can be divided in 3 families of methods: independent component analysis (ICA), non-negative matrix factorization (NMF) and sparse component analysis (SCA) methods. ICA techniques require the independence of the sources s_j, NMF methods require the positivity of the sources s_j and the abundances coefficients a_{ij}, and SCA techniques require the sparseness of the sources s_j.

3.3. NMF semi-blind source separation algorithm

In this study, the NMF method will be considered. Particularly, the NMF of Lee & Seung's multiplicative algorithm (Lee & Seung, 2001) was used to solve the BSS problem. The goal of the NMF is to approximate 2 positives matrices A and S that best fit the mixing matrix X, so that $X \approx AS$. Towards that goal, we consider the cost function J_E, which minimizes the Euclidian distance between X and AS:

$$J_E = D_E(X;AS) = \frac{1}{2}\|X - AS\|^2. \tag{3}$$

The NMF was popularized by the first report of the multiplicative Lee & Seung algorithm (Lee & Seung, 2001), which can simultaneously estimate S and A using a multiplicative scheme:

$$\begin{aligned} s_{jt} &= s_{jt}\frac{\left[A^T X\right]_{jt}}{\left[A^T AS\right]_{jt}} \\ a_{ij} &= a_{ij}\frac{\left[XS^T\right]_{ij}}{\left[ASS^T\right]_{ij}} \end{aligned} \tag{4}$$

where j is the j-th observation, i is the i-th source, t is the t-th sample and $()^T$ is the matrix transpose.

The NMF algorithms are known to be sensitive to the initial conditions (S_{init} and A_{init}) and are known to give a non-unique solution (Cichocki et al., 2009; Moussaoui, Brie, & Idier, 2005). To overcome these limitations, we used a supervised initial condition similar to the one used in Meganem et al. (2014). We considered the existence of 2 components in the mixtures ($n = 2$), so the matrix S_{init} was initialized with 2 spectra, $S_{init} = [s_{1\_init}, s_{2\_init}]^T$, where s_{1_init} is a HyMap soil spectrum taken from the calibration dataset *CalibSet*; the spectrum is considered as the most correlated spectrum with all the other soil spectra of the *CalibSet*. The spectrum s_{2_init} is a laboratory vegetation spectrum resampled to the HyMap spectral bands (see Section 2.3). A_{init} is identified using a

non-negative least square method. We also added a sum to one constraint to the abundances matrix as $\sum_{j=1}^{n} a_{ij} = 1$ to avoid the scale factor that could affect the estimated sources $\hat{s}_j$ with ($j = 1$ *and* 2). The outputs of our SBSS method were 2 spectra $\hat{s}_1$ and $\hat{s}_2$, corresponding to the components existing in the studied sites with $\hat{S} = [\hat{s}_1, \hat{s}_2]^T$.

The stopping criteria was controlled via the connectivity matrix $\hat{C} = [\hat{c}_{xy}] \in \mathbb{R}^{n \times n}$, which measures the stationarity of the estimated matrix $\hat{S}$ with regard to the previous iterations. The convergence was assumed to be reached when $\hat{C}$ remained unchanged for 40 convergence tests; the entry $\hat{c}_{xy}$ was assigned to 1 if samples x and y belonged to the same sample-cluster and was assigned to 0 if they belonged to different sample-clusters.

The identification of the soil spectrum among the estimated $\hat{s}_1$ and $\hat{s}_2$ spectra was performed by computing the coefficient of correlation R_j between the estimated spectrum $\hat{s}_j$ and a reference soil spectrum s_{soil} (Fig. 3a, step 3). The estimated spectrum $\hat{s}_j$ with the higher R_j was identified as the estimated soil spectrum.

Treatments were conducted using R software version 2.14.1 with Package *MASS* for matrix inversion.

3.4. Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is a multivariate model commonly used for soil properties estimation. The PLSR method specifies a linear relationship between a set of dependent (response) variables (Y-variables, the clay content in our case) and a set of predictor variables (X-variables, the spectra) (Tenenhaus, 1998). A detailed description of the PLSR model can be found in Wold, Sjostrom, and Eriksson (2001). The general concept of PLSR is to extract the orthogonal or latent predictor variables, accounting for the maximum amount of the variation of the Y-variables. A PLSR model is developed from a training set of N observations (the number of spectra in the calibration dataset) with K X-variables (the number of wavelengths in the spectra) and M Y-variables (the number of soil properties).

A PLSR model was built using the 95 VNIR/SWIR soil spectra associated with their clay content corresponding to the *CalibSet*. Prior to the quantitative statistical analysis, the reflectance was converted into a "pseudo absorbance" (log [1/reflectance]). Noise reduction was achieved through standard pre-treatments, including a Savitzky–Golay filter with second-order polynomial smoothing and window widths of 30 nm (Savitzky & Golay, 1964) for noise removal and a standard normal variate correction (Barnes, Dhanoa, & Lister, 1993) for additive and multiplicative effect removal. Moreover, an analysis was performed to detect the spectral outliers in the calibration dataset. Spectral outliers are commonly defined as observations that are spectrally different from the remainder of the samples (Chiang, Pell, & Seasholtz, 2003; Pearson, 2002). To identify these spectral outliers, we used the principle of the Mahalanobis distance (Mark & Tunnell, 1985) and applied it to principal component analysis reduced data. The samples with a Mahalanobis distance >3 were identified as outliers.

A leave-one-out cross-validation procedure was adopted to verify the prediction capability of the PLSR model for the calibration set (Wold, 1978). Each time, N − 1 samples were used to build the regression model from all N samples within the dataset. Based on this model, the value for the soil property of the sample that was not used to develop the model was predicted. This procedure was repeated for all N samples, resulting in predictions for all of the calibration samples.

The number of latent predictor variables was set to 6, as determined using the predicted residual sum of squares (PRESS) statistic, taking care to avoid under- and over-fitting. Before building the PLSR model, four spectral outliers were extracted from to the *CalibSet*. The performance for clay content estimation in Cross-Validation was evaluated using the root mean square error of cross-validation (RMSECV) and the coefficient of determination (R^2). This PLSR model has correct performances with a RMSECV of 44.5 g/kg and a R^2 of 0.63, as previously analyzed by Gomez et al. (2012). This PLSR model, called f_{clay}, was used to estimate the clay content from each mixed spectrum (central pixel of the 54 semi-vegetated sites) (Fig. 3b, step 2) and each estimated soil spectrum $\hat{s}_{soil}$ (Fig. 3a, step 4).

Treatments were conducted using R software version 2.14.1 with the *signal* (Short, 2006) and *pls* (Mevik & Wehrens, 2007) packages for Savitzky–Golay filter and PLS regressions, respectively.

3.5. Bias correction

Differences between bare soil spectra (i.e., used to calibrate the PLSR model) and semi-vegetated spectra are due to differences of samples composition. Differences between bare soil spectra (i.e., used to calibrate the PLSR model) and estimated spectra $\hat{s}_j$ are expected, following the source separation accuracy. So we assumed the presence of bias of clay predictions obtained after applying the PLSR model f_{clay} on both semi-vegetated and estimated spectra and we corrected it. This bias, b_0, of predictions was calculated as the difference between the mean of measurement of the clay content in the *CalibSet* and the mean of estimated clay content of the *TestSet*:

$$b_0 = \sum_{j=1}^{m} \frac{\hat{y}_j}{m} - \sum_{i=1}^{n} \frac{y_i}{n} \quad (5)$$

where y_i is the measurement clay content of the sample i of the *CalibSet*, $\hat{y}_j$ is the estimated clay content of the sample j of the *TestSet*, and n and m are the number of soil samples in the *CalibSet* and the *TestSet*, respectively. Subsequently, the following equation was applied to correct the biased predictions:

$$z_j = \hat{y}_j + b_0 \quad (6)$$

where z_j is the corrected clay content of the sample j of the *TestSet*.

3.6. Performance criteria

The SBSS performances for soil spectrum extraction were evaluated using the index $R_{\hat{s}_{soilg}/s_{soil}}$, which is the coefficient of correlation between the estimated soil spectrum $\hat{s}_{soilg}$ (where g is the selected mixed site, see Fig. 3) and the reference soil spectrum s_{soil}. This reference soil spectrum s_{soil} belongs to the *CalibSet* and is one of the most correlated with all the soil spectra of the dataset.

The PLSR performances for clay content estimation were evaluated using the root mean square error of prediction (RMSEP) of the *Testset*. The expected RMSEP should be higher than the RMSE obtained over bare surfaces (RMSECV, see Section 3.4) and should not exceed 72.54 g/kg, which is the standard deviation of the clay contents of the *TestSet* (denoted $RMSEP_{Max}$). The ratio of the performance to the deviation (RPD), which is the ratio between the standard deviation in the *TestSet* and the RMSEP, was used. The RPD value must be >1.4 to consider the PLSR model performance as correct (Chang & Laird, 2002). The coefficient of determination (R^2) between the predicted clay content and observed clay content was used to measure the accuracy of the PLSR model to determine whether the PLSR model performance is adequate. In addition to these three criteria, we measured the prediction bias (see Eq. (5)).

Table 1
Clay prediction performance obtained from the "Direct" approach.

		R^2	RMSEP (g/kg)	RPD	Bias (g/kg)
Mixed spectra	$Test_{mix}$	0.15	66.25	1.11	25.53
	$Test_{mix\_b}$	0.27	61.47	1.19	−6.47

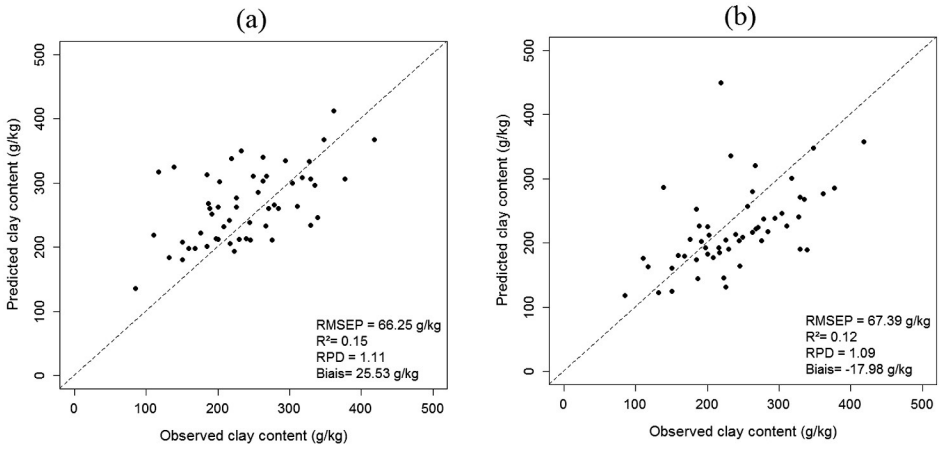


Fig. 4. Predicted clay content versus observed clay content, using a) the direct approach applied to semi-vegetated HyMap spectra *X*, b) the "Double-Extraction" approach applied to estimated spectra S_{soil}.

4. Results

4.1. Prediction using mixed spectra ("direct" approach)

The PLSR model f_{clay} was applied to the HyMap mixed spectra $X_{central} = [x_{central\_1} \ldots x_{central\_g}]^T$ of the *Testset* with $g = 54$ test grids (Fig. 3b). The set of predictions of clay content obtained from these mixed HyMap spectra was called $Test_{mix}$ (Fig. 3b). These predictions were unsuccessful. Although the RMSEP of 66.25 g/kg was slightly inferior to the standard deviation of the clay contents of the *TestSet* ($RMSEP < RMSEP_{Max}$), the R^2 and RPD values were very low, with values of 0.15 and 1.11, respectively (Table 1 and Fig. 4a). Furthermore, the estimated clay predictions showed an important bias (25.5 g/kg).

The clay content predictions obtained directly from the mixed HyMap spectra were corrected from the bias and called the $Test_{mix\_b}$ (Fig. 3b, step 3). These predictions were still unsuccessful (Table 1). Although the RMSEP of 61.47 g/kg was slightly better than the RMSEP obtained without bias correction, the R^2 and RPD were still low, with values of 0.27 and 1.19, respectively.

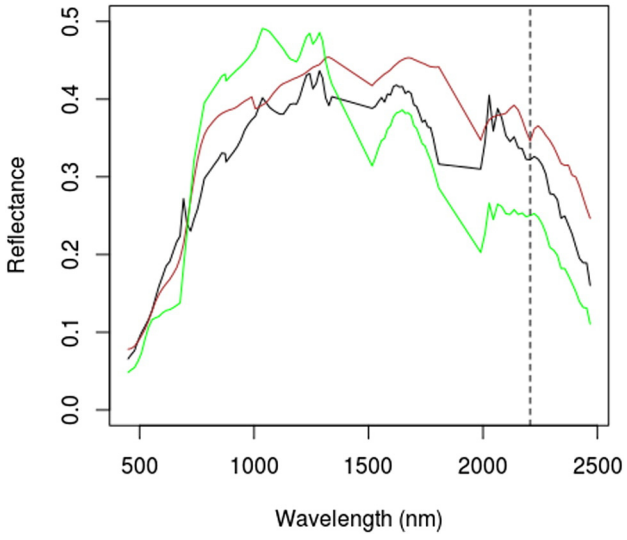


Fig. 5. A HyMap spectrum x_i over a semi-vegetated pixel *i* (green line), the corresponding soil spectrum s_{soil} measured in Lab condition (black line, resampled to the HyMap resolution) and a soil spectrum s_{soil} estimated by SBSS method using the HyMap spectrum x_i and a second HyMap spectrum among the eight neighbors of pixel *i* (brown points). The dotted vertical line indicates the spectral absorption band at 2206 nm.

4.2. Prediction with the "double extraction" approach

a) Extraction of soil spectrum $\boldsymbol{s_{soilg}}$ using the SBSS method
The SBSS method was applied to 8×54 couples of mixed spectra ($X_{central}$, X_i, with $i = 8$). Following the identification process of the soil spectra among each couple of estimated spectra, we have an estimation of 8×54 soil spectra $\hat{S}_{soil}$. To evaluate the accuracy of the estimated soil spectra obtained for the 54 test sites, we calculated the mean coefficient of correlation $R_{\hat{s}_{soilg}/s_{soil}}$ for each site. The soil spectra estimation was accurate, with $R_{\hat{s}_{soilg}/s_{soil}}$ having a value between 0.66 and 0.99 and a mean of 0.96. An example of an estimated soil spectrum is shown in Fig. 5.

The PLSR model f_{clay} was applied to each soil spectrum $\hat{s}_{soilg}$ estimated using the SBSS method (Fig. 3a, step 4). As the estimation of the soil spectrum $\hat{s}_{soilg}$ was realized from 8 combinations of pixels (always including the central pixel) for each of the 54 grids centered on the semi-vegetated sites (54×8), clay contents were predicted from the PLSR model f_{clay}. For each grid, the 8 clay content predictions were averaged; thus, a set of 54 averaged clay content predictions was finally considered and was called the $Test_{sbss}$.

These predictions of clay content, $Test_{sbss}$, after the SBSS method were found to be unsuccessful (Table 2 and Fig. 4b). Although the RMSEP of 67.39 g/kg was slightly inferior to the standard deviation of the clay contents of the *TestSet*, the R^2 and RPD values were very low, with values of 0.12 and 1.09, respectively ($R^2 < 0.5$ and RPD < 1.4). The "Double-Extraction" using the SBSS method showed no improvements in the performance of the clay content estimation compared to the results directly obtained from the spectra $X_{central}$ (Fig. 4a and b).

The SBSS method combined with PLSR appeared to not be efficient for clay estimation; however, the prediction bias of −17 g/kg is better than that obtained using the direct prediction from $X_{central}$, which is 25 g/kg.

The clay content predictions $Test_{sbss}$ were corrected from bias and were called the $Test_{sbss\_b}$ (Fig. 3a, step 5). The bias correction process

Table 2
Clay prediction performance from soil spectra estimated by the "Double-Extraction" approach.

		R^2	RMSEP (g/kg)	RPD	Bias (g/kg)
Soil spectra	$Test_{sbss}$	0.12	67.39	1.09	−17.98
Estimated by SBSS	$Test_{sbss\_b}$	0.17	65.32	1.12	−6.98

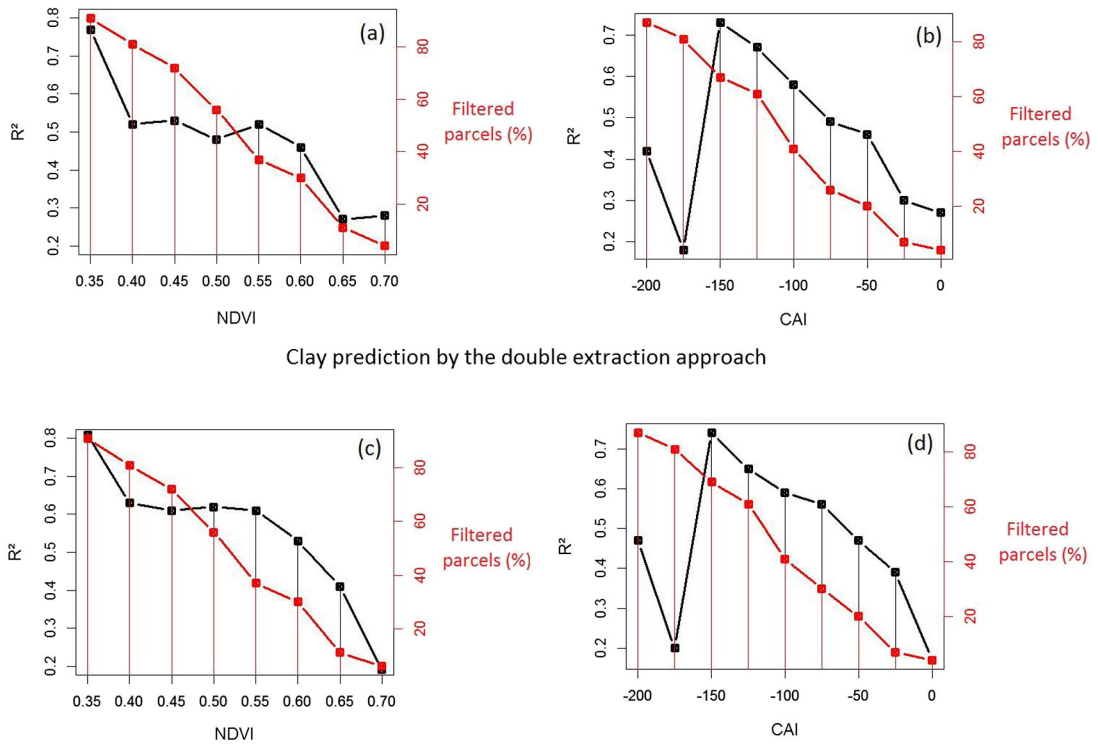


Fig. 6. Clay prediction performance (R^2) and filtered parcels proportions, a) versus NDVI (dataset $Test_{mix\_b\_f}$), b) versus CAI (dataset $Test_{mix\_b\_f}$), c) versus NDVI (dataset $Test_{sbss\_b\_f}$), d) versus CAI (dataset $Test_{sbss\_b\_f}$).

did not provide successful predictions. Although the bias decreased due to the bias correction, the R^2 and RPD values were still low (Table 2).

4.3. Sensitivity to the vegetation proportion and best compromises

The DSS approach consists first, for each site, of a selection of two spectra, including the central pixel spectrum (Fig. 3a, step 1). And the "Direct" approach consists first, for each site, of a selection of the central pixel spectrum (Fig. 3b, step 1). But we may reasonably expect to have highly vegetated pixels over the study area, for which a soil spectrum may not be extracted by source separation and so for which both approaches would give poor performances.

So we studied the sensitivity of both approaches to the presence of vegetation in the mixed pixels (green and dry vegetation). We used the green vegetation index NDVI and the dry vegetation index CAI. The goal was to find the best compromises between accurate predictions of the clay content, as measured by R^2, and the minimum numbers of unpredictable (filtered) sites as measured by the proportions of sites exceeding a vegetation index threshold. We applied the "Double-Extraction" and "Direct" approaches with bias correction (Fig. 3a and b, respectively) on the *Testset* pixels selected in the step 1 of the DSS approach (Fig. 3) under NDVI and CAI thresholds.

Decreases of R^2 were observed according to an increase of the NDVI and CAI thresholds, regardless of the approach used (Fig. 6). The only exception was for the lowest thresholds of the CAI values (left of Fig. 6b and d) that did not predict enough sites to obtain significant results. Meanwhile, as expected, the number of filtered parcels continuously decreased with the increases of the NDVI and CAI thresholds, regardless of the approach.

For each of the approaches and each vegetation index used, it was possible to define the best compromises that provided both acceptable predictions and a small amount of unpredicted sites (Table 3). Considering these compromises revealed that the "Double-Extraction" approach systematically outperformed the "Direct" approach by providing more accurate clay predictions (for NDVI-based thresholds), predicting more sites (for CAI-based thresholds), or providing the two former types of improvement (for combined NDVI and CAI thresholds). Thresholding NDVI provided better compromises than thresholding CAI, and the combination of the two vegetation indices did not provide significant improvements.

Finally, the best compromise among the approaches and vegetation indices was obtained by applying the NDVI < 0.55 threshold to the "Double-Extraction" results, which provided fairly accurate predictions ($R^2 = 0.61$) with limited numbers of unpredicted sites (37%).

4.4. Clay content mapping

The PLSR model f_{clay} built using the 95 spectra associated with the bare soil pixels of the HyMap image was applied to all the spectra over bare soil of the HyMap image in a previous study (Gomez et al.,

Table 3
Best compromises of clay prediction performance by (a) the "Direct" approach and (b) the "Double-Extraction" approach, using the best ranges in terms of CAI and/or NDVI and a bias correction processes. (In bold are indicated performances of prediction with $R^2 > 0.5$ and RPD > 1.4).

	Vegetation index threshold	RMSEP (g/kg)	R^2	RPD	Bias (g/kg)	Filtered parcels (%)
Clay prediction by the "Direct" approach $Test_{mix\_b\_f}$ (a)	CAI ≤ −100	52.33	0.58	1.4	−8.75	41
	NDVI ≤0.55	54.28	0.52	1.35	−0.08	37
	CAI ≤ −100 and NDVI ≤0.55	53.51	0.58	1.37	−4.95	46
Clay prediction by the "Double −Extraction" approach $Test_{sbss\_b\_f}$ (b)	CAI ≤ −75	**51.66**	**0.56**	**1.42**	**−7.98**	**30**
	NDVI ≤0.55	**49.06**	**0.61**	**1.5**	**−0.09**	**37**
	CAI ≤ −75 and NDVI ≤0.55	**50.28**	**0.6**	**1.46**	**−3.43**	**41**

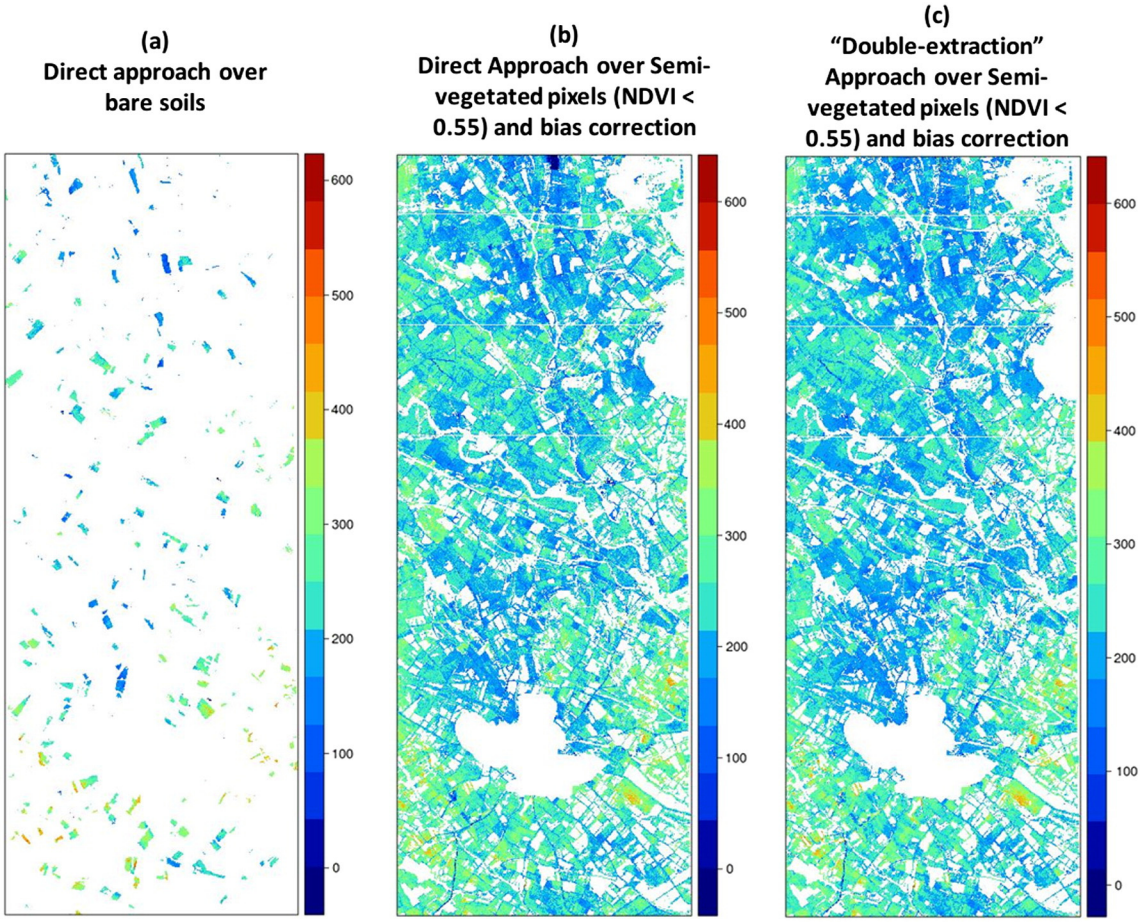


Fig. 7. Maps of clay contents estimated (a) over bare soils by the PLSR model f_{clay} (*Clay_Bare* map), (b) over semi-vegetated areas with NDVI <0.55 by the "Direct" approach and bias correction (*Clay_MixDirect* map) and (c) over semi-vegetated areas with NDVI <0.55 by the "Double-Extraction" approach and bias correction (*Clay_MixDSS* map).

2012) generating a predicted clay map called *Clay_Bare* (Fig. 7a). This same PLSR model f_{clay} was also applied to all the HyMap spectra covering semi-vegetated pixels with NDVI inferior or equal to 0.55 generating a predicted clay map called *Clay_MixDirect* (Fig. 7b). In addition, the PLSR model f_{clay} was applied to all the soil spectra $\hat{S}_{soil}$ estimated by SBSS method over semi-vegetated pixels with NDVI inferior or equal to 0.55 generating a predicted clay map called *Clay_MixDSS* (Fig. 7c).

In a second step, the *Clay_Bare* and *Clay_MixDirect* maps were joined to produce an extensive mapping of predicted clay contents, over pixels with NDVI from 0 to 0.55 and called *Clay_BareMixDirect*. And the *Clay_Bare* and *Clay_MixDSS* maps were joined to produce an extensive mapping of predicted clay contents, over pixels with NDVI from 0 to 0.55 and called *Clay_BareMixDSS*.

It was difficult to discern any differences of predicted spatial distribution over the entire study region according to both approaches ("Direct" or "Double-Extraction") and the studied surfaces (bare or semi-vegetated). The estimated clay content values of the three maps follow normal laws with no significant biases between each of them (Fig. 8).

The two predicted maps (*Clay_MixDirect* and *Clay_MixDSS*, Fig. 7b and c) exhibited spatial variations of clay that appeared to be in accordance with our knowledge of the soil of the study area (Coulouma, Barthes, Masson, & Robbez, 2008). The lowest clay contents were found both in the north of the image and in the north of the Alignan du Vent village (the southern "urban" polygon), for which soils are developed over Miocene loose sediments that are poor in clay. Conversely, the highest clay contents were found in the south of the study area,

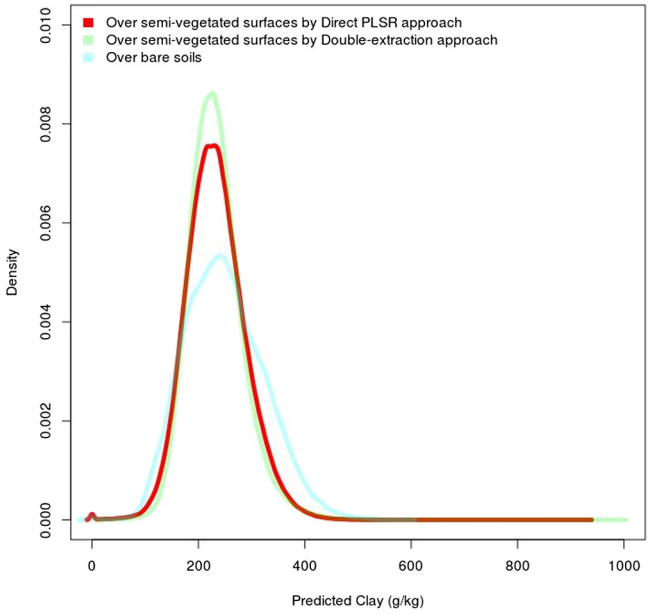


Fig. 8. Histograms of clay contents estimated over bare soils by the PLSR model f_{clay} (blue curve), over semi-vegetated areas with NDVI <0.55 by the "Direct" approach and bias correction (red curve) and by the "Double-Extraction" Approach and bias correction (green curve). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

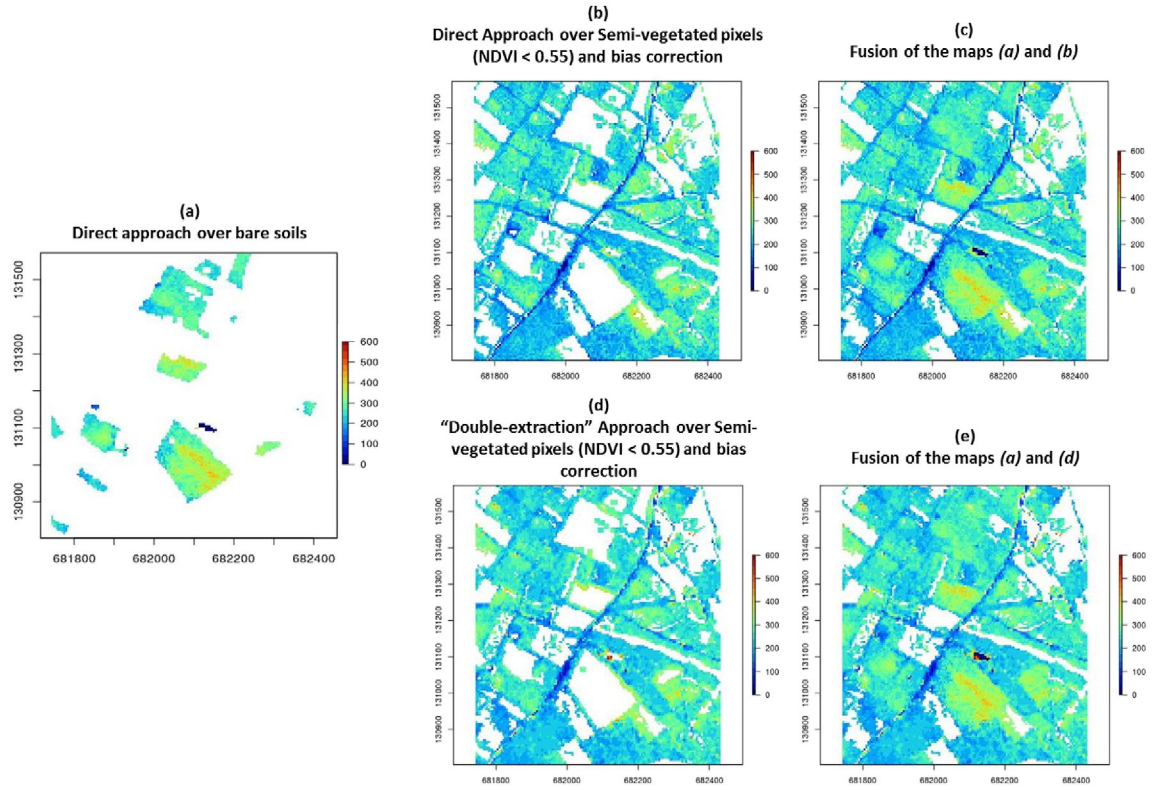


Fig. 9. Maps of clay contents estimated over a 1st selected area: (a) over bare soils by the PLSR model f_{clay} (*Clay_Bare* map), (b) over semi-vegetated areas with NDVI < 0.55 by the "Direct" approach and bias correction (*Clay_MixDirect* map), (c) over bare soils and semi-vegetated areas with NDVI < 0.55 by the fusion of the maps (a) and (b) (*Clay_BareMixDirect* map), (d) over semi-vegetated areas with NDVI < 0.55 by the "Double-Extraction" approach and bias correction (*Clay_MixDSS* map) and (e) over bare soils and semi-vegetated areas with NDVI < 0.55 by the fusion of the maps (a) and (d) (*Clay_BareMixDSS* map).

characterized by Pliocene marls outcrops, which are clay-rich soils. Intermediate values of the predicted clay contents were found along the left bank of the Peyne River and on the center-east and center-west of the study area; these areas correspond to soils developed on old fluviatile alluvium.

However, differences between approaches were visible when focusing on a small number of fields (Fig. 9). A "salt and pepper" effect was clearly visible in the results of the "Direct" approach, which could be due to the green vegetation (Figs. 9b-c and 10b-c). This salt and pepper effect was much less visible using the "Double-Extraction" approach, which resulted in a smoother clay map than the one obtained using the "Direct" approach (Figs. 9d-e and 10d-e). The clay content estimated by the "Double-Extraction" approach is better distributed over the parcels than the one estimated by the "Direct" approach. In addition, focusing on a small number of fields allowed for verification of the good concordance and continuity of short scale spatial patterns between the clay contents estimated over bare soils by the PLSR model f_{clay}, (Gomez et al., 2012) (Figs. 9a and 10a) and the neighboring clay contents over semi-vegetated pixels estimated by both approaches (Figs. 9b-d and 10b-d).

Finally, the "Double-Extraction" approach allowed the estimation of the clay content over 63% of the pixels of our HyMap image, which increased the mapped surface by 59.5% compared to the previous study of Gomez et al. (2012).

5. Discussion

5.1. Spectral unmixing

The statistical method of spectral unmixing must be carefully chosen regarding the mathematical hypotheses given to the sources. Regarding the ICA method, the hypothesis of independence is difficult to satisfy, especially in the spectral dimension, which is characterized by a high dependence between the endmembers that hamper the applicability of the method to a spectral endmembers determination context (Bioucas-Dias et al., 2012). Regarding the SCA method, the hypothesis of sparseness is not verified in the spectral domain; the sources must be transformed in a sparse domain to use the method. The NMF method is characterized by a positivity constraint imposed on the sources and the abundances, which is in accordance with the hyperspectral imagery characteristics (spectral signatures and their relative abundances are positive). Indeed, the NMF method appears to be more suitable than the other 2 methods to estimate the endmembers in a hyperspectral context (e.g., Pauca et al., 2006; Berné et al, 2009; Meganem et al., 2014).

The presence of vegetation could introduce some non-linearities to the mixture model due to the multiscattering phenomenon that occurs in presence of vegetation. For example, non-linear mixtures models, such as the quadratic model (Hosseini & Deville, 2003; Meganem et al., 2014), could be integrated into the NMF algorithm and tested.

5.2. Introduction of prior knowledge

The "Double-Extraction" approach was first developed and used without any prior knowledge on the sources of interest (Ouerghemmi et al., 2011). In this study, we wanted to take advantage of our ground knowledge; therefore, we modified the approach to allow it to consider this knowledge. A simple way to achieve this approach is to initialize the NMF algorithm used to unmix the observed data with some known reference spectra. In this study, we considered the simple case of two materials, i.e., soil and vegetation (vineyard), so the algorithm was

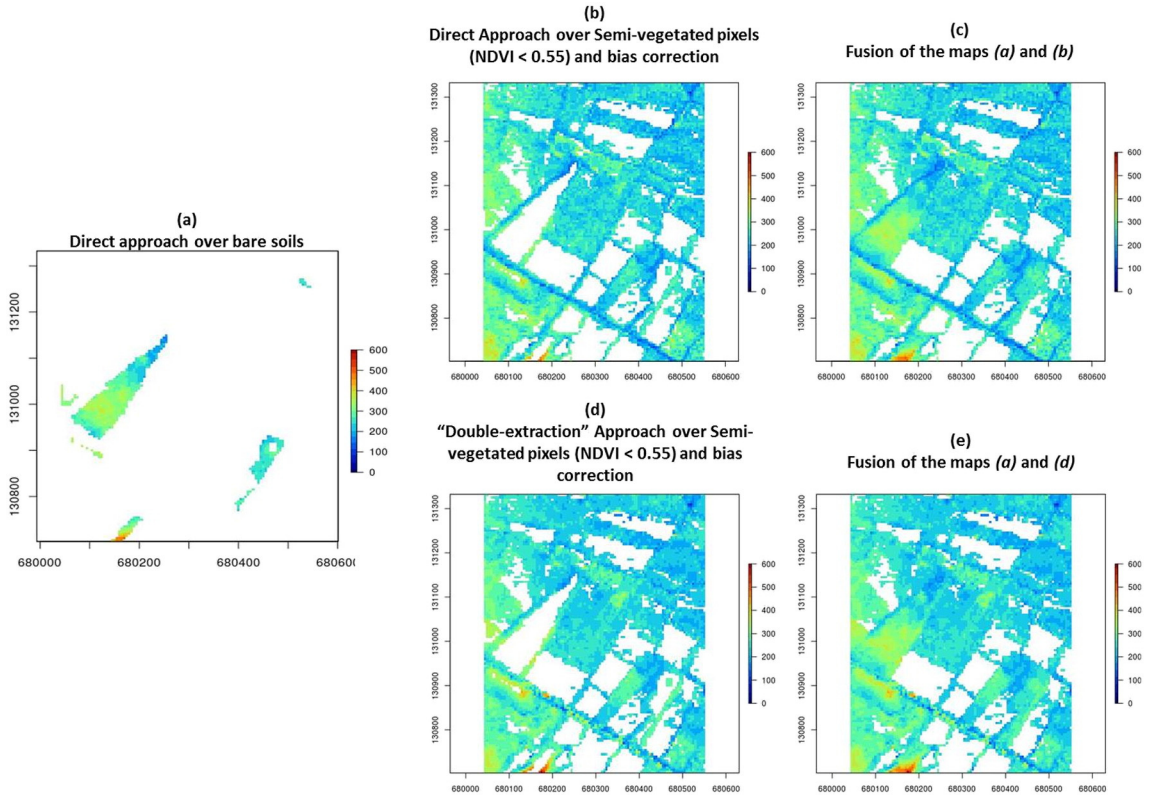


Fig. 10. Maps of clay contents estimated over a 2[sd] selected area: (a) over bare soils by the PLSR model f_{clay} (*Clay_Bare* map), (b) over semi-vegetated areas with NDVI < 0.55 by the "Direct" approach and bias correction (*Clay_MixDirect* map), (c) over bare soils and semi-vegetated areas with NDVI < 0.55 by the fusion of the maps (a) and (b) (*Clay_BareMixDirect* map), (d) over semi-vegetated areas with NDVI < 0.55 by the "Double-Extraction" approach and bias correction (*Clay_MixDSS* map) and (e) over bare soils and semi-vegetated areas with NDVI < 0.55 by the fusion of the maps (a) and (d) (*Clay_BareMixDSS* map).

initialized with a representative soil spectrum (i.e., taken from the HyMap soil spectra used in the calibration of the PLSR method) and vineyard spectrum (i.e., measured in the laboratory and resampled to the HyMap spectral resolution). This supervised initialization enables the enhancement of both the separation performance and the prediction performance. The use of vineyard spectra measured on the ground (instead of the actual laboratory vineyard spectra) in the initialization process might also improve the prediction performances of the "Double-Extraction" approach. Indeed, the same condition of reflectance acquisition as that in the airborne sensor acquisition would be taken into account (integration of atmospheric effects and solar conditions). In addition, to enhance the separation performance by better modeling the mixed pixels over the La Peyne catchment, one can take into account the greater number of materials in the mixtures (i.e., >2), particularly in fields with maintenance practices favoring the presence of green herbs between the rows of a vineyard. A prerequisite would then be to develop very high resolution remote sensing approaches to identify such field peculiarities (Corbane, 2006).

Few studies have focused on the mixture of soil and vegetation for the mapping of the soil properties. Bartholomeus et al (2011) addressed this problem in the context of maize fields. Under the assumption of the presence of 2 sources (soil and maize), the authors used a simulated pure spectrum of maize and estimated the abundance fractions of the sources to extract the soil spectra from the mixtures. Our "Double-Extraction" approach offers more flexibility and fewer requirements to process the mixed data than the inversion method used by Bartholomeus et al (2011), which requires knowledge regarding the non-soil spectra and their abundances; in addition, a simulated spectral library is required to calculate the abundances. Indeed, the "Double-Extraction" approach requires only n (in our case $n = 2$) representative pure spectra to be initialized.

5.3. Raw performances of the clay predictions over vegetated pixels

The results obtained by both approaches ("Direct" and "Double-Extraction") on semi-vegetated areas (regardless of the NDVI) were unsuccessful ($R^2 = 0.15$ and RMSEP = 66 g/kg for the "Direct" approach; $R^2 = 0.15$ and RMSEP = 66 g/kg for the "Double-Extraction" approach) (Table 1 and Table 2, respectively). Thus, whatever the approach, the clay prediction is not possible. The poor performance of the "Direct" approach was not surprising. Indeed, the spectra used to calibrate the PLSR model (i.e., soil spectra), were spectrally different from the test spectra (i.e., mixed spectra of soil/vegetation). In this case study, the PLSR, which is a multivariate chemometric regression method, cannot manage the spectral difference by itself.

An alternative approach could be to develop a transfer method of implementing regression models built from spectra X_{cal} (measured by an instrument A) for the prediction from spectra X_{new} (measured by an instrument B) (e.g., Fearn, 2001).

5.4. Bias correction process

Biases of predictions are a frequent consequence of the differences between calibration and test spectra (because of the instrument, acquisition conditions, nature of the target, etc.), as already demonstrated by e.g., Bouveresse and Massart (1996); Roger, Chauchard, and Williams (2008) and Fearn (2001). Using the "Direct" approach, the bias was positive, i.e., the PLSR model over-estimated the clay content. A part of this bias might be due to differences between the calibration spectra (spectra collected over bare soils pixels) and the test spectra (spectra collected over mixed pixels). Moreover, the presence of vegetation in the mixed spectra involved an amplification of OH ions absorption band at 2200 nm, which might involve an over-estimation of the clay content, as this absorption band is linked to the presence of clay minerals and

used in PLSR models (Gomez, Lagacherie, & Coulouma, 2008b). Using the "Double-Extraction" approach, the bias was negative, which means that the PLSR model under-estimated the clay content. This bias can be due to lower absorption band of OH ions in soil spectra estimated by SBSS than in the HyMap soil spectra (see Fig. 5), which might, as previously noted, involve an under-estimation of the clay content.

These contrasted absorption bands might be due to the unrealistic vegetation spectrum that served for initializing the SBSS (see Section 3.3) because this spectrum was measured on a single vineyard leaf in laboratory and then resampled to the HyMap spectral resolution. More realistic vineyard spectra retrieved from a hyperspectral image would require finer spatial resolution.

Our bias correction process could be classified as a posteriori correction to the PLSR model development, following the classification of Chauchard, Cogdill, Roussel, Roger, & Bellon-Maurel, 2004. This correction process slightly differs from the slope-bias correction (SBC), which is one of the standardization methods based on correcting the predicted values of an existing calibration model (Osborne & Fearn, 1983; Bouveresse and Massart, 1996), largely used in the chemometric community to calibration transfer (or standardization) between instruments (e.g., Salguero-Chaparro, Palagos, Peña-Rodríguez, & Roger, 2013). The SBC correction uses a pool of samples similar in both calibration and test databases, which allow for the estimation of slope and bias values, while our correction method cannot calculate a slope because no sample belongs to both calibration and test databases. Thus, our bias correction process is as a reduction of the SBC correction.

5.5. Impact of the vegetation coverage

The clay prediction performance decreased according to an increase in the vegetation proportion (green or dry) at the mixtures level, regardless of the approach used ("Direct" or "Double-Extraction"). Our results show that the estimations of the vegetation proportion using NDVI or CAI were precise enough to predict the behavior of the approaches with regard to the presence of vegetation. Therefore, these vegetation indices could be used to detect and exclude mixed parcels that are difficult to handle using "Direct" or "Double-Extraction" approaches.

Once we fixed the indices to be used for estimating vegetation proportion, the evaluation of the approaches should consider not only the prediction precision (models performance) but also the number of characterized sites. It seems important to find a compromise between the proportion of characterized sites and the performance of the models when choosing a vegetation proportion range to be considered for each approach.

For our case study, the compromises were chosen by deduction, after crossing the performance indicators for both approaches. For an NDVI threshold of 0.55, the "Double-Extraction" approach offers reasonable performances ($R^2 = 0.61$ and RMSEP = 49 g/kg) (Table 3) and permits the characterization of 63% of the semi-vegetated sites. For a CAI threshold of −100, the "Direct" approach offers reasonable performances ($R^2 = 0.58$ and RMSEP = 52.3 g/kg) (Table 3) and permits the characterization of 59% of the semi-vegetated sites.

6. Conclusion

Due to a lack of methodologies adapted to mixed surfaces, the mapping of primary soil properties by hyperspectral data are estimated, until now, only over bare soils. This lack of methodologies is a severe restriction, especially when processing images over a natural landscape with a high percentage of vegetation. This study demonstrated that an exhaustive use of the "Double-Extraction" approach, regardless of the vegetation conditions, did not succeed in predicting the clay content. In contrast, acceptable predictions were obtained using the "Double-Extraction" approach for a significant part of the study area (63%) that was moderately covered by vegetation (NDVI <0.55). Furthermore, this approach outperformed a "Direct" approach consisting of the use of a classical multivariate regression model (PLSR) on unprocessed pixels ("Direct" approach). The performance results may be considered as modest, even using the "Double-Extraction" approach restricted to moderate vegetation coverage (NDVI <0.55). Nevertheless, the "Double-Extraction" approach performance is close to the one of the multivariate regression model (PLSR, f_{clay}) in cross-validation, which is initially modest ($R^2 = 0.63$, Gomez et al., 2012). So this multivariate regression model is a limiting factor in the "Double-Extraction" approach. And higher performances of PLSR model f_{clay} in cross-validation would involve higher performances of "Double-Extraction" approach.

Finally, the "Double-Extraction" approach could be adapted for other purposes, depending on the material of interest. In this study, we were interested in the extraction of soil spectra from mixtures to estimate a soil physico-chemical property using a multivariate regression model, and the estimated vegetation spectra were abandoned. However, future studies could use these estimated vegetation spectra to predict e.g., biophysical parameters of plants, as initiated by Tits, Somers, Saeys, and Coppin (2014), or to identify vegetation species.

Acknowledgments

This research was supported by a PhD funding from the Institut de la Recherche pour le Développement (IRD) 2011-2014. We are indebted to Dr Steven M. de Jong, Utrecht University in The Netherlands and to Dr Andreas Mueller of the German Aerospace Establishment (DLR) in Wessling, Germany for providing the 2003 Hymap images for this study. We are also indebted to Yves Blanca, IRD-UMR LISAH Montpellier, for the soil samples sampling in 2009 over the vineyard plain of Languedoc.

References

Ans, B., Hérault, J., & Jutten, C. (1985). Adaptive neural architectures: Detection of primitives. *Proceedings of COGNITIVA'85* (pp. 593–597). Paris: France.

Baize, D., & Jabiol, B. (1995). *Guide pour la description des sols.* Paris: INRA édition.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1993). Correction to the description of standard normal variate (snv) and de-trend transformations in practical spectroscopy with applications in food and beverage analysis — 2nd edition. *Journal of Near Infrared Spectroscopy*, *1*, 185–186.

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., Van Wesemael, B., Ben-Dor, E., & Tychon, B. (2011). Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *13*, 81–88.

Bayer, A., Bachmann, M., Müller, A., & Kaufmann, H. (2012). A comparison of feature-based MLR and PLS regression techniques for the prediction of three soil constituents in a degraded South African ecosystem. *Applied and Environmental Soil Science*, *2012*.

Beisl, U. (2001). Correction of bidirectional effects in imaging spectrometer data. *Remote Sensing Series*. (pp. 37)Zurich: Remote Sensing Laboratories.

Berné, O., Joblin, C., Tielens, A., Deville, Y., Puigt, M., Guidara, R., Hosseini, S., Mulas, G., & Cami, J. (2009). Source separation algorithms for the analysis of hyperspectral observations of very small interstellar dust particles. *Proceedings of the 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (IEEE WHISPERS 2009)* (pp. 26–28). Grenoble, France, August.

Bioucas-Dias, J. M., Plaza, A., Dobigeon, N., Parente, M., Du, Qian, Gader, P., & Chanussot, J. (2012). Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*(2), 354–379.

Boardman, J. W. (1993). Automating spectral unmixing of AVIRIS data using convex geometry concepts. *Summaries of the 4th Annual JPL Airborne Geoscience Workshop* (pp. 11–14).

Bouveresse, E., & Massart, D. L. (1996). Improvement of the piecewise direct standardization procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, *32*, 201–213.

Bouveresse, E., & Massart, D. L. (1996). Standardization of near-infrared spectrometric instruments: a review. *Vibrational spectroscopy*, *11*(1), 3–15.

Caiafa, C., Salerno, E., & Proto, A. (2007). Blind source separation applied to spectral unmixing: Comparing different measures of nongaussianity. In B. Apolloni, R. Howlett, & L. Jain (Eds.), *Knowledge-based Intelligent Information and Engineering System. 4694.* (pp. 1–8). Berlin Heidelberg: Springer.

Chabrillat, S., Goetz, A. F. H., Krosley, L., & Olsen, H. W. (2002). Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sensing of Environment*, *82*, 431–445.

Chang, C. -W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, *167*(2), 110–116.

Chauchard, F., Cogdill, R., Roussel, S., Roger, J. M., & Bellon-Maurel, V. (2004). Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust

and portable sensor for acidity prediction in grapes. *Chemometrics and Intelligent Laboratory Systems*, *71*(2), 141–150.
Chen, S., Donoho, D., & Saunders, M. (2001). Atomic decomposition by basis pursuit. *SIAM Review*, *43*(1), 129–159.
Chiang, L. H., Pell, R. J., & Seasholtz, M. B. (2003). Exploring process data 735 with the use of robust outlier detection algorithms. *Journal of Process Control*, *13*(5), 437–449.
Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S. (2009). *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation.* Wiley Publishing.
Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*, 287–314.
Comon, P., & Jutten, C. (2010). *Handbook of blind source separation: Independent component analysis and applications*. Academic press.
Corbane, C. (2006). *Detection of soil surface characteristics in Mediterranean agricultural areas by remote sensing at very high spatial resolution.* (PhD Dissertation) Montpellier, France: Université de Montpellier 2 Sciences et Techniques du Languedoc (251 pp.).
Coulouma, G., Barthes, J. P., Masson, J. M., & Robbez, J. M. (2008). *Carte des sols de la basse vallée de la Peyne — Notice*.
Daughtry, C. S. T., McMurtrey, J. E., III, Nagler, P. L., Kim, M. S., & Chappelle, E. W. (1996). Spectral reflectance of soils and crop residues. In A. M. C. Davies, & Phil Williams (Eds.), *Near infrared spectroscopy: The future waves* (pp. 505–511). Chichester, UK: NIR Publications.
Dehaan, R. L., & Taylor, G. R. (2001). Mapping irrigation-induced salinity with hyperspectral imagery. *International Geoscience and Remote Sensing Symposium IGARSS '01* (pp. 293–295). Proceedings, IEEE International.
Donoho, D. L., & Stodden, V. (2003). When does non-negative matrix factorization give a correct decomposition into parts? *Proc. NIPS 2003*. MIT Press.
Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499.
Fearn, T. (2001). Standardisation and calibration transfer for near infrared instruments: A review. *Journal of Near Infrared Spectroscopy*, *9*, 229–244.
Gomez, C., Lagacherie, P., & Coulouma, G. (2008b). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, *148*(2), 141–148.
Gomez, C., Lagacherie, P., & Coulouma, G. (2012). Regional predictions of eight common soil properties and their spatial structures from hyperspectral vis–NIR data. *Geoderma*, *189–190*(0), 176–185.
Gomez, C., Viscarra Rossel, R. A., & McBratney, A. B. (2008a). Soil organic carbon prediction by hyperspectral remote sensing and field vis–NIR spectroscopy: An Australian case study. *Geoderma*, *146*, 403–411.
Hosseini, S., & Deville, Y. (2003). Blind separation of linear-quadratic mixtures of real sources using a recurrent structure. In J. Mira, & J. Álvarez (Eds.), *Artificial neural nets problem solving methods. 2687*. (pp. 241–248). Berlin Heidelberg: Springer.
Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications.Neural. *Networks*, *13*(4–5), 411–430.
Iordache, M. D., Bioucas-Dias, J. M., & Plaza, A. (2011). Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(6), 2014–2039.
IUSS Working Group WRB (2006). World reference base for soil resources. *World soil resources report 103* (2nd ed.). Rome, Italy: FAO (133 pp.).
Jutten, C., & Herault, J. (1991). Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, *24*, 1–10.
Kokaly, R. F., & Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sensing of Environment*, *67*(3), 267–287.
Lagacherie, P., Baret, F., Feret, J. -B., Madeira Netto, J., & Robbez-Masson, J. -M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, *112*(3), 825–835.
Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791.
Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in neural information processing systems. 13*. (pp. 556–562). MIT Press.
Lee, T. W., Lewicki, M. S., Girolami, M., & Sejnowski, T. J. (1999). Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, *6*, 87–90.
Leggett, D. J. (1977). Numerical analysis of multicomponent spectra. *Analytical Chemistry*, *49*, 276–281.
Lin, J., Grier, D., & Cowan, J. (1997). Faithful representation of separable distributions. *Neural Computation*, *9*, 1305–1320.
Lu, P., Wang, L., Niu, Z., Li, L., & Zhang, W. (2013). Prediction of soil properties using laboratory VIS–NIR spectroscopy and Hyperion imagery. *Journal of Geochemical Exploration*, *132*, 26–33.
Lu, X., Wu, H., & Yuan, Y. (2014). Double constrained NMF for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(5), 2746–2758.
Madeira Netto, J. S. R., Robbez-Masson, J. M., & Martins, E. (2007). Visible–NIR hyperspectral imagery for discriminating soil types in the La Peyne watershed (France). In P. Lagacherie, A. B. Mc Bratney, & M. Voltz (Eds.), *Developments in soil science. 31*. (pp. 219–234). Elsevier.
Mallat, S., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, *41*(12), 3397–3415.
Mark, H. L., & Tunnell, D. (1985). Qualitative near infrared reflectance analysis using Mahalanobis distances. *Analytical Chemistry*, *57*(7), 1449–1456.
Meganem, I., Deville, Y., Hosseini, S., Deliot, P., & Briottet, X. (2014). Linear-quadratic blind source separation using NMF to unmix urban hyperspectral images. *IEEE Transactions on Signal Processing*, *62*(7), 1822–1833.
Mevik, B. -H., & Wehrens, R. (2007). The PLS package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*(2), 1–24.
Moussaoui, S., Brie, D., & Idier, J. (2005). Non-negative source separation: Range of admissible solutions and conditions for the uniqueness of the solution. *Proceedings of IEEE international conference on acoustics, speech, and signal processing, ICASSP, 5* (pp. 289–292).
Nascimento, J. P. M., & Bioucas-Dias, J. M. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(4), 898–910.
Osborne, B. G., & Fearn, T. (1983). Collaborative evaluation of near-infrared reflectance analysis for the determination of protein, moisture and hardness in wheat. *Journal of the Science of Food and Agriculture*, *34*(9), 1011–1017.
Ouerghemmi, W., Gomez, C., Naceur, S., & Lagacherie, P. (2011). Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces. *Geoderma*, *163*(3–4), 227–237.
Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, *5*(2), 111–126.
Pauca, V. P., Piper, J., & Plemmons, R. J. (2006). Nonnegative matrix factorization for spectral data analysis. *Linear Algebra and its Applications*, *416*(1), 29–47.
Pearson, R. K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, *10*(1), 55–63.
Richter, R. (1996). Atmospheric correction of DAIS hyperspectral image data. *Computers and Geosciences*, *22*, 785–793.
Richter, R. (2000). *A unified approach to parametric geocoding and atmospheric topographic correction for wide FOVairborne imagery. Part 2: Atmospheric topographic correction. Proceedings of the Second EARSeL Workshop on Imaging Spectroscopy* (Enschede, The Netherlands).
Roger, J. M., Chauchard, F., & Williams, P. (2008). Removing the bloc effects in calibration by means of dynamic orthogonal projection. Application to the year effect correction for wheat protein prediction. *Near infrared spectroscopy*, *16*, 311–315.
Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. *Prog. Rep. RSC 1978-1*. College Station: Remote Sensing Center, Texas A&M Univ (93 pp.).
Salguero-Chaparro, L., Palagos, B., Peña-Rodríguez, F., & Roger, J. M. (2013). Calibration transfer of intact olive NIR spectra between a pre-dispersive instrument and a portable spectrometer. *Computers and Electronics in Agriculture*, *96*(0), 202–208.
Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627–1639.
Short, T. (2006). Signal processing package in R, original octave version by Paul Kienzle pkienzle@user.sf.net. *Conversion to R by Tom Short*.
Starck, J. L., Moudden, Y., Bobin, J., Elad, M., & Donoho, D. (2005). Morphological component analysis. *Proceedings of the SPIE conference wavelets, San Diego, USA* (pp. 5914).
Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., & Wesemael, B. (2010). Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, *158*(1–2), 32–45.
Taylor, G. R., & Dehaan, R. L. (2000). Mapping soil salinity with hyperspectral imagery. *Proceedings of the 14th international conference applied geologic remote sensing, Veridan ERIM International, 6th November, Las Vegas, Nevada* (pp. 512–519).
Tenenhaus, M. (1998). *La régression PLS, Edition TECHNIP.*
Tits, L., Somers, B., Saeys, W., & Coppin, P. (2014). Site specific plant state monitoring through hyperspectral alternative least squares unmixing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(8), 3606–3618.
Winter, M. E. (1999). N-FINDR: Algorithm for fast autonomous spectral endmember determination in hyperspectral data. *Proceedings of SPIE, 3753* (pp. 266–277).
Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, *20*, 397–405.
Wold, S., Sjostrom, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.
Zibulevsky, M., & Pearlmutter, B. A. (2001). Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, *13*, 863–882.

Remote Sensing of Environment 164 (2015) 1–15

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery

C. Gomez [a,*], R. Oltra-Carrió [b], S. Bacha [c], P. Lagacherie [d], X. Briottet [e]

[a] IRD, UMR LISAH (INRA–IRD–SupAgro), F-34060 Montpellier, France
[b] CESBIO, Toulouse, France
[c] CNCT, Tunis, Tunisia
[d] INRA, UMR LISAH (INRA–IRD–SupAgro), F-34060 Montpellier, France
[e] ONERA, The French Aerospace Lab, 31055 Toulouse, France

ARTICLE INFO

Article history:
Received 2 July 2014
Received in revised form 16 February 2015
Accepted 19 February 2015
Available online xxxx

Keywords:
VNIR/SWIR spectroscopy
Hyperspectral satellite
Airborne remote sensing
Spatial resolution
Clay content mapping
PLSR
Mediterranean context

ABSTRACT

Visible, near-infrared and short wave infrared (VNIR/SWIR, 0.4–2.5 μm) hyperspectral satellite imaging is one of the most promising tools for topsoil property mapping for the following reasons: i) it is derived from a laboratory technique that has been demonstrated to be a good alternative to costly physical and chemical laboratory soil analysis for estimating a large range of soil properties; ii) it can benefit from the increasing number of methodologies developed for VNIR/SWIR hyperspectral airborne imaging; and iii) it provides a synoptic view of the study area. Despite the significant potential of VNIR/SWIR hyperspectral airborne data for topsoil property mapping, the transposition to satellite data must be evaluated. The objective of this study was to test the sensitivity of clay content prediction to atmospheric effects and to degradation of spatial resolution. This study may offer an initial analysis of the potential of future hyperspectral satellite sensors, such as the HYPerspectral X Imagery (HYPXIM), the Spaceborne Hyperspectral Applicative Land and Ocean Mission (SHALOM), the PRecursore IperSpettrale della Missione Applicativa (PRISMA), the Environmental Mapping and Analysis Program (EnMAP) and the Hyperspectral Infrared Imager (HyspIRI), for soil applications. This study employed VNIR/SWIR AISA-DUAL airborne data acquired in a Mediterranean region over a large area (300 km^2) with an initial spatial resolution of 5 m. These hyperspectral airborne data were simulated at the top of the atmosphere and aggregated at six spatial resolutions (10, 15, 20, 30, 60 and 90 m) to correlate with the future hyperspectral satellite sensors. The predicted clay content maps were obtained using the partial least squares regression (PLSR) method. The large area of the studied region allows analysis of different pedological patterns of soil composition and spatial structures. Our results showed the following: (i) when a correct compensation of atmosphere effects was performed, only slight differences were detected between clay maps retrieved from the airborne imagery and those from spaceborne imagery (both at 5 m of spatial resolution); (ii) the PLSR models, built from data with 5 to 30 m spatial resolutions, performed well, and allowed clay mapping, although variations in clay content related to short scale succession of parent material was imperfectly captured beyond 15 m of spatial resolution; (iii) the PLSR models built from data with 60 and 90 m spatial resolutions were inaccurate, and did not enable clay mapping; and (iv) the two latter results could be explained by the combination of a small short-scale clay content variability and small field sizes observed in the study area. Therefore, in the Mediterranean and under the spectral specifications of the AISA-DUAL airborne sensor, most of the future hyperspectral satellite sensors (four of the five sensors examined in this study) will be potentially useful for clay content mapping.

1. Introduction

In soil science, laboratory VNIR/SWIR spectroscopy has been demonstrated to be an alternative to costly physical and chemical soil analysis for the estimation of a large range of soil properties (e.g., Ben-Dor & Banin, 1995; Cécillon et al., 2009; Viscarra Rossel, Walvoort, McBratney, Janik, & Skjemstad, 2006). One way to retrieve soil properties is to use known specific absorption features (e.g., Clark & Roush, 1984; Lagacherie, Baret, Feret, MadeiraNetto, & Robbez-Masson, 2008). For example, illite, kaolinite and montmorillonnite clay materials have an absorption peak around 2.206 μm, corresponding to the combination of OH stretch and OH-Al bending modes (e.g., Chabrillat, Goetz, Krosley, & Olsen, 2002). Illite clay material also has absorption features near

* Corresponding author.
E-mail address: cecile.gomez@ird.fr (C. Gomez).

http://dx.doi.org/10.1016/j.rse.2015.02.019
0034-4257/

2.3 and 2.4 μm (Clark, 1999), and CaCO3 has an absorption peak around 2.348 μm corresponding to CO3 overtone vibrations (Gaffey, 1987). Moreover, various chemometric methods have been used to relate soil VNIR/SWIR spectra to soil properties, including multiple regression analysis (Ben-Dor & Banin, 1995), stepwise multiple linear regression (Shibusawa, Imade Anom, Sato, Sasao, & Hirako, 2001), multivariate adaptive regression splines (Shepherd & Walsh, 2002), principal components regression (e.g., Chang, Laird, Mausbach, & Hurburgh, 2001) and support vector machine regression (Stevens et al., 2010). The partial least-squares regression (PLSR) method is the most common multivariate statistical technique used for spectral calibration and prediction of soil properties (e.g., Chang & Laird, 2002; McCarty, Reeves, Reeves, Follett, & Kimble, 2002). These chemometric methods allow for the prediction of soil properties, independent of whether they have a spectral response. For example, pH which is not expected to have a direct spectral response, has been fairly predicted in several cases using chemometric methods, due to covariation to spectrally active soil constituents, such as organic matter and clay (e.g., Chang et al., 2001; Islam, Stingh, & McBratney, 2003).

Because knowledge of laboratory spectroscopy has improved, the number of studies using VNIR/SWIR hyperspectral airborne imaging in topsoil property mapping has also increased (e.g., Ben-Dor, Patkin, Banin, & Karnieli, 2002; Gomez, Lagacherie, & Coulouma, 2008; Selige, Bohner, & Schmidhalter, 2006; Stevens et al., 2010). VNIR/SWIR hyperspectral airborne imaging has been considered a promising technology for increasing the accuracy of digital mapping of topsoil properties targeted by the Global Soil Map (GSM) project (www.globalsoilmap.net) (Lagacherie & Gomez, 2014). The GSM project has proposed the construction of a new digital soil map of the world at a spatial resolution of 90 m to assist in the decision-making process for a range of global issues, such as food production, climate change and environmental degradation (Sanchez et al., 2009). In this context, VNIR/SWIR hyperspectral airborne imaging can provide a synoptic view of the area under study at spatial resolutions appropriate for topsoil property mapping (Gomez, Coulouma, & Lagacherie, 2012) and is particularly well-adapted to semi-arid areas in which bare soil surfaces are common, and dry periods enable avoidance of soil moisture perturbations of the spectrum (Lagacherie et al., 2008).

Two spaceborne sensors, PRISMA (PRecursore IperSpettrale della Missione Applicativa) and EnMAP (Environmental Mapping and Analysis Program, http://www.enmap.org/) are expected to be launched in the near future. PRISMA is an Italian hyperspectral sensor to be launched in 2017 (Lopinto & Ananasso, 2013). The spatial resolution initially considered for this sensor was between 20 and 30 m over the spectral range of 0.4 to 2.5 μm, and between 2.5 and 5 m in the panchromatic band (Giampaolo et al., 2008). Recently, the spatial resolution of PRISMA has been highlighted as 30 m in the 0.4 to 2.5 μm interval, and 5 m in the panchromatic band (The Prisma Mission, Nota Informativa, Agenzia Spaziale Italiana, DC-OST-2009-124-04/09/09/ Retrieved 23 August 2013). EnMAP is a German hyperspectral sensor, set for launch in 2017, with a spatial resolution of 30 m over the spectral range of 0.4 to 2.5 μm (Stuffler et al., 2007). Moreover, the following three new advanced hyperspectral sensors are under study: HyspIRI (Hyperspectral Infrared Imager), SHALOM (Spaceborne Hyperspectral Applicative Land and Ocean Mission) and HYPXIM (HYPerspectral X Imagery). HyspIRI is an American hyperspectral sensor with a spatial resolution of 60 m over the spectral range of 0.38 to 2.5 μm (hyspiri.jpl.nasa.gov). SHALOM is an Italy–Israel initiative with a spectral range of 0.4 to 2.5 μm, and a spatial resolution initially defined at 15 m (Bussoletti, 2012), which recently evolved from 15 m to 10 m (Ben-Dor, Kafri, & Varacalli, 2014). HYPXIM is a French hyperspectral sensor with a spatial resolution better than 8 m over the spectral range of 0.5 to 2.5 μm, and of 1.8 m in the panchromatic band (Briottet et al., 2013; Carrere et al., 2013).

Despite the significant potential of the VNIR/SWIR hyperspectral airborne data for mapping several soil properties, the application of methods developed for airborne data (such as the PLSR method) to satellite data is not a straightforward process. Several factors can affect topsoil property mapping via satellite sensors, including a lower signal to noise ratio, atmospheric effects, different spectral characteristics (varying band centers and spectral resolutions) and coarser spatial resolutions. At the moment, the future spaceborne sensors plan to acquire reflectance data with spectral resolution of approximately 10 nm (Staenz, Mueller, & Heiden, 2013), which is close to the resolution used by airborne sensors HYMAP (approximately 15 nm, Cocks, Jenssen, Stewart, Wilson, & Shields, 1998; Kruse et al., 1999) or AISA-DUAL (approximately 5 nm and 7 nm, Specim Society, http://www.channelsystems.ca/). So the application of methods to satellite data for mapping several soil properties should not be affected by the spectral resolution. Conversely, spatial resolutions planned for future spaceborne sensors (between 8 to 60 m) are coarser than the ones used by airborne sensors, such as HYMAP or AISA-DUAL (approximately 5 m). The various effects of the spatial resolution degradation on multispectral remote sensing data have been studied, including flux prediction (e.g., Kustas & Norman, 2000), crop area estimation and crop growth monitoring (e.g., Duveiller & Defourny, 2010) and agricultural land monitoring (e.g., Ismail, Mutanga, Kumar, & Urmilla, 2008; Pax-Lenney & Woodcock, 1997). Some spatial resolution degradation studies have also focused on VNIR/SWIR hyperspectral data for vegetation studies (e.g., Nijland, Addink, De Jong, & Van der Meer, 2009; Rahman, Gamon, Sims, & Schmidts, 2003; Schaaf, Dennison, Fryer, Roth, & Roberts, 2011; Thorp, French, & Rango, 2013; Zhang, Middleton, Gao, & Cheng, 2012). However, to our knowledge, only a few studies on the impact of degraded spatial resolution have used VNIR/SWIR hyperspectral data for fire detection, temperature retrieval (Lugassi, Ben-Dor, & Eshel, 2010) and urban studies (Jensen & Cowen, 1999; Roberts, Quattrochi, Hulley, Hook, & Green, 2012), with none focused on soil science applications.

Moreover, few studies have reported the use of VNIR/SWIR hyperspectral satellite data for topsoil property mapping (e.g., Gomez, Viscarra Rossel, & McBratney, 2008; Weng, Gong, & Zhu, 2008; Zhang, Li, & Zheng, 2009). This scarcity of studies using satellite data in soil science, compared with studies using airborne data, is due to the existence of only one VNIR/SWIR hyperspectral satellite sensor (the Hyperion sensor on board the EO-1 satellite launched in 2002), which has a low signal to noise ratio (e.g., ~50:1 from 2.1 to 2.4 μm) (Folkman, Pearlman, Liao, & Jarecke, 2001).

The objective of this study was to evaluate the sensitivity of VNIR/SWIR hyperspectral-based soil property prediction to atmospheric effects and degradation in spatial resolution. The selected property was the topsoil clay content, which is one of the basic soil properties used by soil surveyors to describe soil types. It is also a key driver of soil erosion processes (Le Bissonnais, 1996), and one of the primary soil properties included in the GlobalSoilMap specifications (Arrouays et al., 2014). This study employed the VNIR/SWIR AISA-DUAL hyperspectral airborne data acquired over a large area (300 km^2) in a Mediterranean region, with an initial spatial resolution of 5 m. The predicted clay content maps were obtained using the PLSR method.

The study area is described in Section 2, and the airborne, simulated satellite and field data are described in Section 3. The methodology of the clay content estimation and statistical issues are presented in Section 4. Finally, the results are presented in Section 5 and are discussed in Section 6.

2. Study area

The study area is located in the Cap Bon region in northern Tunisia (36°24′N to 36°53′N; 10°20′E to 10°58′E), 60 km east of Tunis (Fig. 1a). This 300 km^2 area includes the Lebna catchment, which is primarily rural (>90%), and devoted to cereals, legumes, olive trees, natural vegetation for breeding and vineyards. This area is characterized by rolling land, with an elevation between 0 and 226 m. The climate varies

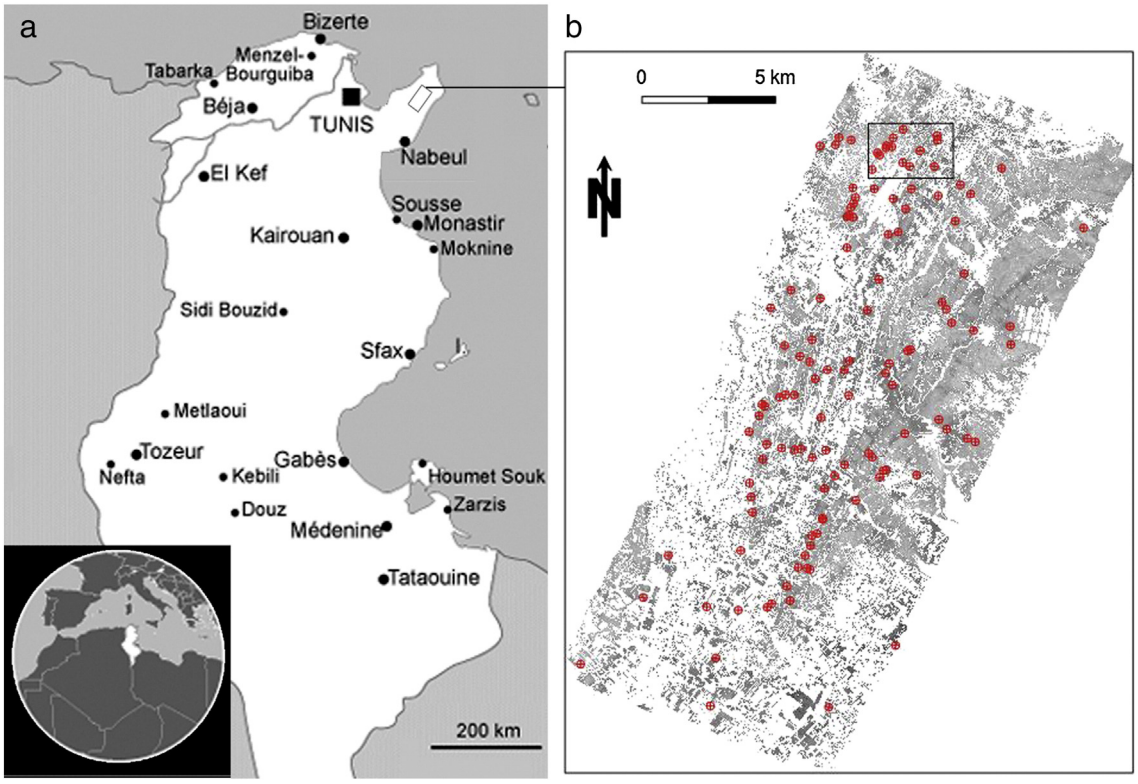


Fig. 1. a) Map of Tunisia and location of the study area over the Cap Bon region (black polygon) and b) the AISA-DUAL hyperspectral image at 2.206 μm, with location of the Kamech catchment (in the rectangle). In white are the masked areas (vegetated, urban, and water areas). Red targets represent locations of collected soil samples of *CalBase_5m*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from humid to semi-arid, with an inter-annual precipitation of 600 mm and an annual potential evapotranspiration of 1500 mm. The soil pattern of the Lebna catchment primarily arises from variations in the lithology. The changes in the landscape between the Miocene sandstone and the marl outcrops induce significant variations in the physical and chemical soil properties (Zante, Collinet, & Pepin, 2005). Furthermore, the distance between the successive sandstone outcrops decreases in the direction of sea-mountains, which results in variations in the soil property patterns. The soil materials were redistributed laterally along the slopes during the Holocene, which adds to the complexity of the soil patterns. The main soil types are Regosols, Eutric Regosols (9.6%) preferentially associated with sandstone outcrops, Calcic Cambisol and Vertisol, preferentially formed on marl outcrops and lowlands. The southeastern region of the study area has a flatter landscape, with sandy Pliocene deposits yielding Calcosol and Rendzina.

A sub-catchment of the study area was used to examine the short scale variations in the clay contents. This sub-catchment is a 6.67 km^2 area centered on the Kamech catchment (Fig. 1b). The Kamech experimental catchment belongs to a long-term environmental research observatory called OMERE (Mediterranean observatory of water and rural environment), which aims to study the anthropogenic effects on water and sediment budgets at the catchment scale (e.g., Mekki, Albergel, Ben Mechlia, & Voltz, 2006; Raclot & Albergel, 2006). The Kamech catchment represents an area (15 km from the coast) in which the soil pattern shows strong variations on a small scale, with a rapid succession of sandstone and shale. This area had a high percentage of bare soils during the image acquisition (43.4%), and exhibited contrasting soil patterns.

3. Airborne, simulated satellite and field data

3.1. Hyperspectral airborne radiance data

On November 2, 2010, an AISA-DUAL hyperspectral image was acquired over the study area (12 × 25 km) via plane, at an altitude of 3600 m, with a field of view (FOV) of 24° and a spatial resolution of 5 m (Fig. 1b). The AISA-DUAL airborne imaging spectrometer measures the at-flight radiances ($L_{flight\_AISA\_5m}$) in 359 non-contiguous bands, covering the 0.40 μm to 2.45 μm spectral domain, with 0.0046 μm bandwidths between 0.40 μm and 0.97 μm, and 0.0065 μm bandwidths between 0.97 μm and 2.45 μm. The field of view of the instrument is 24°. On November 2, 2010, the visibility was estimated at 40 km, using weather reports and visual estimations from the airplane. The cloud cover was 0% during the flight-time, and neither the in-situ atmospheric profiles nor the aerosol data were available during the campaign.

3.2. Simulation of the satellite radiances

3.2.1. Simulation of the "Top of Atmosphere" radiances

The Top of Atmosphere (TOA) synthetic data were obtained by considering the entire atmosphere layer, requiring the definition of an atmosphere profile with a given water vapor content, and the type and abundance of the aerosols. From these inputs, the TOA radiances (L_{TOA_5m}) were simulated from the at-flight radiances ($L_{flight\_AISA\_5m}$) (Fig. 2), according to the following equation:

$$L_{TOA\_5m} = \left(\tau_{dir} + \tau_{dif}\right) L_{flight\_AISA\_5m} + L_{atm} \quad (1)$$

where τ_{dir} and τ_{dif} are the atmospheric direct and diffuse transmittance, respectively, and L_{atm} is the upwelling atmospheric radiance (without the contribution of the land). COMANCHE is a radiative transfer model used to solve Eq. (1) (Miesch et al., 2005). COMANCHE uses an analytical formulation of the upwelling radiance at the sensor level, in which the atmospheric parameters are independent of the ground parameters. The MODTRAN 4 radiative transfer code (Berk et al., 1999) was used to compute most of the atmosphere parameters (e.g., the columnar water vapor amount), except for the Earth atmosphere coupling irradiance

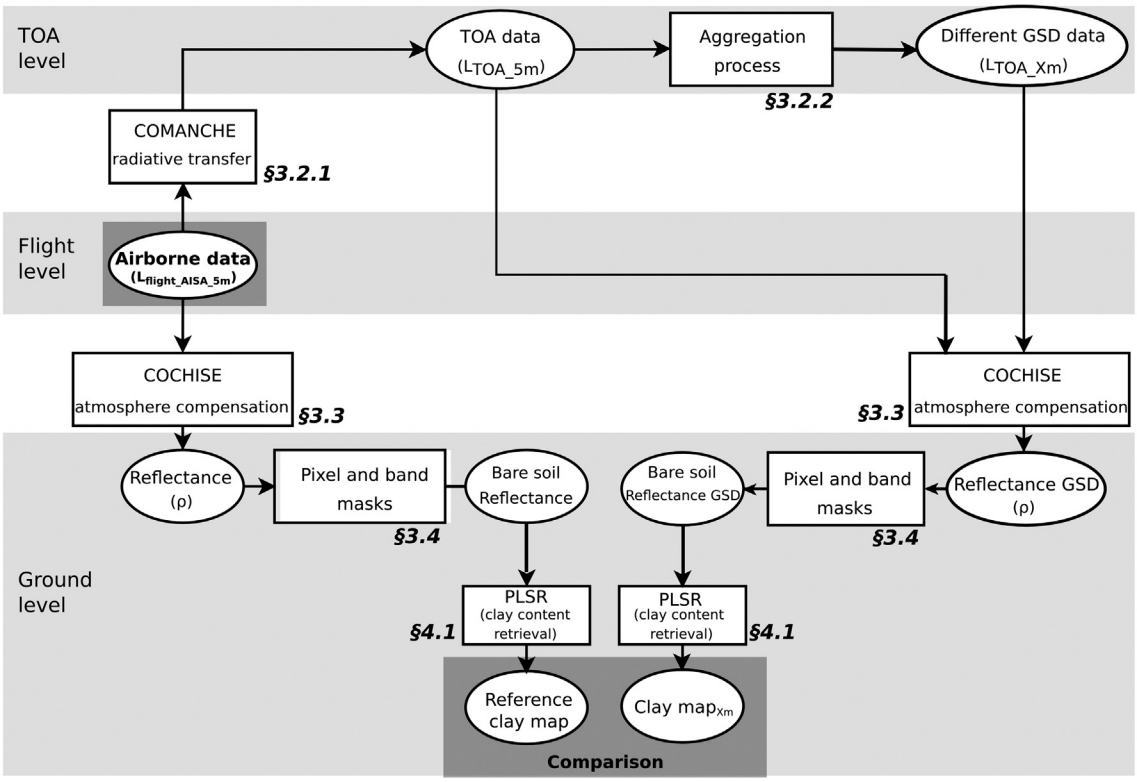


Fig. 2. Procedure overview.

and the environment upwelling radiance, which were estimated with a Monte Carlo kernel.

In the COMANCHE tool, a rural aerosol model was used with a 40 km visibility, and the United States Standard 1976 atmosphere model from MODTRAN (Berk et al., 1999). This model was chosen because it had the best fit with the value of the total water vapor content, estimated at 1.5 g/cm^2 by MOD07 (https://lpdaac.usgs.gov/).

3.2.2. Spatial resolution degradation

An aggregation process was applied to the geometrically corrected TOA radiances (L_{TOA_5m}) (Fig. 2). As proposed in the following Eq. (2), the flux of the aggregated surface is equal to the addition of the flux of each *i* element forming the entire aggregated surface:

$$\Omega_{agg} L_{TOA\_xm} = \sum_{i=1}^{N} \Omega_i L_{TOA\_5m,i} \tag{2}$$

where Ω is the solid angle corresponding to the instantaneous FOV related to the different ground sampling distance (GSD) values, and the index *agg* refers to the aggregated surface, which is composed of *N i* elements. If the *N* solid angles, Ω_i, are equivalent, then the total aggregated solid angle is given by Eq. (3). Finally, the aggregated radiance is defined by the average of the *i* radiances, as shown in Eq. (4). Therefore, the new aggregated pixels were obtained by averaging all of the pixel values that contributed to the output pixel. This exercise was performed using the Resampling Aggregate tool of the ENVI application using the following equations:

$$\Omega_{agg} = N\Omega_i \tag{3}$$

$$L_{TOA\_xm} = \frac{1}{N}\sum_{i=1}^{N} L_{TOA\_5m,i}. \tag{4}$$

The selected ground sampling distances were driven by the GSD of the airborne acquired data (5 m). Then, six new TOA radiance images (L_{TOA_xm}) were produced at the following GSD values: 10, 15, 20, 30, 60 and 90 m. These values were close to those from the spaceborne missions, as follows: HYPXIM (10 m), SHALOM (15 m), PRISMA (20 m, 30 m), EnMAP (30 m) and HyspIRI (60 m), and from the value of the Global Soil Map (GSM) project (www.globalsoilmap.net), set at 90 m.

3.3. Hyperspectral reflectance data

From the airborne radiances ($L_{flight\_AISA\_5m}$), the simulated TOA radiances (L_{TOA_xm}) described in Sections 3.1 and 3.2 and the knowledge of the atmospheric conditions, the surface reflectance, ρ, was retrieved using the atmosphere compensation tool COCHISE, which is an inverse radiative transfer model (Miesch et al., 2005) (Fig. 2), based on the following equation:

$$L_{TOA\_xm}(i,j) = L_{atm} + \tau_{dir}\cdot\left[\frac{E_0\cdot\rho_{xm\_SAT}(i,j)}{\pi.(1-\rho_G(i,\ j).S)}\right] + \tau_{dif}\cdot\left[\frac{E_0\cdot\rho_F(i,\ j)}{\pi.(1-\rho_G(i,\ j).S)}\right] \tag{5}$$

where E_0 is the distance corrected extra-terrestrial solar irradiance, $\rho_{xm\ \_\ SAT}(i,j)$ is the reflectance to be retrieved at pixel (i,j), $\rho_F(i,j)$ and $\rho_G(i,j)$ are the average albedo values given by the convolution with the environmental functions to consider the contribution of the environmental upwelling radiance (F) and the earth atmosphere coupling irradiance (G), respectively. *S* is the atmospheric spherical albedo. These terms are estimated using a Monte Carlo approach. The remaining atmospheric parameters are computed from the MODTRAN 4 radiative transfer code (Berk et al., 1999). The inversion process is iterative. In the initial step, the adjacency effects are neglected, and a first estimate of $\rho_{xm\ \_\ SAT}(i,j)$ is obtained. Then, Eq. (5) is used to retrieve the surface reflectance from the simulated TOA radiances (L_{TOA_xm}). Eq. (5) is used

again to retrieve the surface reflectance from the airborne radiances ($L_{flight\_AISA\_5m}$), and in this case, $\rho_{xm\_SAT}(i,j)$ is replaced by $\rho_{flight\_5m}(i,j)$.

Finally, to remove the noise effects introduced by this process, the reflectance spectrum of each pixel was smoothed using the following filter:

$$\rho_i^{smooth} = \frac{1}{N}\sum_{j=i-m}^{i+m}\rho_j, \quad m = \frac{N-1}{2} \tag{6}$$

where ρ_i^{smooth} is the surface reflectance in the channel number i after the application of the filter, and N is an odd integer indicating the number of wavelength neighbors to be averaged. In our case, N is equal to 3.

Thus, eight reflectance images were obtained. One image was obtained at a 5 m resolution from the acquired airborne radiances and was called 5*m_AISA*. Seven images were obtained from the simulated TOA radiances at the different GSD values (see Section 3.2.2) and were called 5*m_HYPXIM*, 10*m_HYPXIM-SHALOM*, 15*m_SHALOM*, 20*m_PRISMA*, 30*m_PRISMA-EnMAP*, 60*m_HyspIRI* and 90*m_GSM*.

3.4. Band selections and urban, vegetation and water masks

Before the implementation of the clay content retrieval procedure, noisy or troublesome spectral bands and pixels must be removed (Fig. 2). We removed the following: a) 20 spectral bands in the blue part of the spectral domain from 0.4 to 0.484 μm due to instrumental noise, b) 5 bands from 0.747 μm to 0.766 μm, which correspond to O_2 band absorption, c) 12 bands from 0.952 μm to 1.019 μm due to H_2O band absorption and instrument defect, d) 14 bands from 1.094 μm to 1.176 μm due to H_2O band absorption, e) 21 bands from 1.339 μm to 1.465 μm due to H_2O band absorption and f) 38 bands from 1.773 μm to 2.005 μm due to H_2O band absorption. Therefore, 249 spectral bands between 0.489 μm and 2.451 μm were retained.

Considering that vegetation coverage is a limiting factor for soil property mapping, the current hyperspectral data treatments do not allow comprehensive mapping of an extensive region (e.g., Gomez, Lagacherie, et al., 2008; Schwanghart & Jarmer, 2011; Selige et al., 2006; Stevens et al., 2010). Only two studies in remote sensing soil mapping have focused on diversified surface conditions, including partially vegetated surfaces (Bartholomeus et al., 2010; Ouerghemmi, Gomez, Naceur, & Lagacherie, 2011), and performed marginally well. Therefore, in the absence of effective techniques for predicting soil properties over semi-vegetated pixels, having a suitable mapping surface remains an important criterion of hyperspectral image quality. To isolate the bare soil areas, pixels with normalized difference vegetation index (NDVI) values over an expert-calibrated threshold were masked. A value of 0.20 was determined after considering twenty parcels, which were visually inspected in the field. The NDVI was retrieved using bands at 0.672 μm and 0.799 μm. Areas of water were also masked using an expert-calibrated threshold. Pixels with a reflectance of less than 8% at 1.665 μm were removed. Finally, 13 urban areas were identified by visual inspection and were also masked.

When the hyperspectral airborne image was acquired (November 2010), a major portion of the soil surface was covered by green vegetation, primarily consisting of olive trees, native forests, green plants and vineyards. Based on the hyperspectral image, with a spatial resolution of 5 m, bare soils represented 43.4% of our study area. A degradation of spatial resolution leads to a slight decrease in the suitable surface area that can be mapped by VNIR/SWIR spectroscopy. From 5 to 30 m, this decrease in suitable surface is approximately 4.4% of suitable surface area, and from 30 to 90 m, this decrease is approximately 6.4%. Based on the images at 10 m, 15 m, 20 m, 30 m, 60 m and 90 m of spatial resolution, bare soils represent 42.6%, 41.5%, 40.6%, 39%, 35.2% and 32.6%, respectively.

3.5. Field data

One hundred twenty soil samples were collected over the study area (Fig. 1b). Among this sample set, 50 were collected in June 2008, 30 in October 2009 and 41 in November 2010. All of these soil samples were collected in fields that were bare during the hyperspectral airborne data acquisition in November 2010. All of the samples were composed of five sub-samples, which were collected to a depth of 5 cm at random locations within a $10 \times 10\ m^2$ square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. Because soil samples were collected over plowed fields, the collected soils from 0 to 5 cm can be considered as having mixed and homogeneous soil properties. After homogenizing the sample, approximately 20 g was allocated for clay content analysis. The initial samples were air-dried and sieved with a 2 mm sieve prior to being transported to the laboratory for analysis. The determination of the clay content (granulometric fraction < 2 μm) was determined using a pipette method (method NF X 31-107, particle size distribution by sedimentation, Baize & Jabiol, 1995). The clay content of the 120 soil samples varied between 108 and 772 g/kg, and followed a normal distribution (Table 1).

A degradation in the spatial resolutions leads to a slight decrease in the suitable surface area that can be mapped by VNIR/SWIR spectroscopy (Section 3.4), and leads to a decrease in the number of soil samples associated with bare soil pixels. At 5 m, 120 soil samples were available over the bare soil pixels. This 120 sample database is called *CalBase-5m*. At 60 and 90 m, only 67 and 65 soil samples, respectively, were available over the bare soil pixels. These databases are called *CalBase-60m* and *CalBase-90m*, respectively. The *CalBase-5To30m* database consists of 82 soil samples, which correlate with the images from 5 to 30 m of resolution. Although the number of reference soil samples varied depending on the spatial resolution, the distribution and statistical characteristics of the soil datasets did not vary. The clay content of the *CalBase-5To30m* database varied from 122 to 772 g/kg, and followed a normal distribution (Table 1). The clay content of the *CalBase-60m* and *CalBase-90m* database varied from 122 to 772 g/kg and to 764 g/kg, respectively, and both followed a normal distribution (Table 1).

4. Methods

4.1. Clay content prediction

Partial least squares regression (PLSR) is a multivariate model commonly used for soil property estimation. The PLSR method specifies a linear relationship between a set of dependent (response) variables (Y-variables, the clay content in our case), and a set of predictor variables (X-variables, the spectra) (Tenenhaus, 1998). A detailed description of the PLSR model can be found in Wold, Sjöström, and Eriksson (2001). The general concept of PLSR is to extract the orthogonal or latent predictor variables, accounting for the maximum amount of the variation of the Y-variables. A PLSR model is developed from a training set of N observations (the number of spectra in the calibration dataset), with K X-variables (the number of wavelengths in the spectra), and M Y-variables (the number of soil properties). The training data form matrices X and Y, of dimensions $(N \times K)$ and $(N \times M)$, respectively. As with all of the factorial methods, the main goals of PLSR are the following: i) to locate a subspace of the spectral space $\mathbb{R}^K$ on which the spectra are projected, yielding a matrix of N scores $\mathbf{T}$ $(N \times k)$; and ii) to perform

Table 1
Main descriptive statistics of the soil dataset used.

Dataset	Number of data	Mean ± st.dev	Min	Max	Skewness
CalBase-5m	120	466 ± 174	108	772	−0.11
CalBase-5To30m	82	481 ± 173	122	772	−0.19
CalBase-60m	67	489 ± 173	122	772	−0.15
CalBase-90m	65	484 ± 170	122	764	−0.23

a linear regression between T and Y. The maximum number of latent predictor variables was set to 10 and was determined using the Predicted Residual Sum of Squares (PRESS) statistic, taking care to avoid under- and over-fitting.

An analysis was performed to detect the outliers in the calibration dataset. Outliers are commonly defined as observations that are inconsistent with the majority of the data (Chiang, Pell, & Seasholtz, 2003; Pearson, 2002), such as observations that deviate significantly from normal values. An outlier can be defined as the following: (i) a spectral outlier when the sample is spectrally different from the remainder of the samples or (ii) a concentration outlier when the predicted value has a residual difference significantly greater than the mean of the predicted values. To identify the spectral outliers, we used the principle of the Mahalanobis distance (Mark & Tunnell, 1985) and applied it to principal component analysis reduced data. The samples with a Mahalanobis distance greater than 3 were identified as outliers and removed from the calibration dataset. For the case of normally distributed data, 99% of the data will be included in an interval of $\pm 3\sigma$, which will be less than a threshold of 3.

Prior to the quantitative statistical analysis, the reflectance was converted into a "pseudo absorbance" (log [1/reflectance]), and a standard normal variate correction was achieved for removal of additive and multiplicative effects (Barnes, Dhanoa, & Lister, 1993).

The prediction capability of the PLSR was analyzed using two different procedures, which were dependent on the size of the dataset (the number of spectra associated with the clay content). *Type* 1 was applied to the *CalBase*-5*m* and *CalBase*-5*To*30*m* datasets, which contain more than 80 variables. In this case, the dataset was divided into two groups, one group for calibration of the PLSR model (3/4 of the dataset), and a second group used for validation (1/4 of the database). The Y values were sorted in an ascending order. The method starts by selecting the sample with the lower clay content, and placing it in a validation set. Then, the next three samples are placed in the calibration set, and the procedure is continued by alternately placing the following sample in the validation set and the next three samples in the calibration set. Such a process would ensure a relatively equal distribution of the samples in both sub-datasets. *Type* 2 is applied to datasets with less than 80 reference samples (*CalBase*-60*m* and *CalBase*-90*m*). In this case, a subdivision in calibration and validation datasets would provide too little data in the calibration set. So all of the data were used in the calibration dataset, and no independent validation set was used.

In both types of procedures, a leave-one-out cross-validation procedure was adopted to verify the prediction capability of the PLSR model for the calibration set. $N - 1$ samples were used to build the regression model from all of the N samples within the dataset. Based on this model, the value for the clay content of the sample not used in the development of the model was predicted. This procedure was repeated for all of the N samples, resulting in predictions for all of the samples.

4.2. Performance of the clay content prediction

The performance of the prediction was evaluated using the coefficient of correlation (R^2_{cal}) of the predicted values against the measured values, and the root mean square errors of calibration (RMSEC). Moreover, for *Type* 1, the correlation coefficient of validation (R^2_{val}) and the root mean square errors of the prediction in the validation set (RMSEP) were also measured. The ratio of the performance to the deviation (RPD), which is the ratio between the standard deviation in the validation set and the RMSEP, was used. Thresholds of the RPD values used to assess the accuracy of prediction were proposed by Chang et al. (2001). RPD values greater than 2 indicate models with excellent prediction capability, values between 2 and 1.4 indicate intermediate models and values below 1.4 indicate unreliable models. However, although RPD is largely used in soil spectroscopy, these thresholds were not determined statistically, and RPD does not describe the range of variation correctly, particularly for data with a non-normal distribution. Therefore, the ratio of performance to interquartile (RPIQ), which is the ratio of the interquartile (IQ = Q3 − Q1) to the RMSEP recently proposed to represent the spread of the population (Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010), was also used.

Finally, the Variable Importance in the Projection (VIP) (Chong & Jun, 2005; Wold et al., 2001) and the PLSR b-coefficients (Haaland & Thomas, 1988) were used to study the significant wavelengths used in PLSR. A wavelength is considered to be significant when both its b-coefficient and VIP value are sufficiently large (Wold et al., 2001). The thresholds for the VIP were set to 1 (following the recommendations by Chong & Jun, 2005), and the thresholds for the b-coefficients were based on their standard deviations (Viscarra Rossel, Jeon, Odeh, & McBratney, 2008). The chemometric developments were implemented in R (version 1.17) using the pls package (Mevik & Wehrens, 2007).

4.3. Spatial structure issues

A comparison between two digital images can be performed using a simple method, which compares the two images on a pixel by pixel basis. However, this method does not allow for analysis of the structure of the predicted values in a spatial dimension. Thus, variograms were used to compare the ability of the prediction models to accurately reproduce the spatial structures of the soil property in the study area. Variograms measure the spatial dependence of the soil properties using semi-variance. The average variance between any pair of sampling points (i.e., the semivariance) for a soil property, Y, at any vector of distance h apart can be determined using the following formula (Webster & Oliver, 1990):

$$\gamma(h) = \frac{1}{2m}\sum_{i=1}^{m}\{Y(x_i) - Y(x_i + h)\} \quad (7)$$

where $\gamma(h)$ is the average semi-variance of the soil property, m is the number of pairs of sampling points, Y is the soil property content, x is the coordinate of the point and h is the lag (the distance that the pairs are apart). Variograms provide evidence of spatial autocorrelation when the semi-variances are lower at smaller lags than at larger lags, i.e., the sampling locations located close to each other have similar values. The autocorrelation of clay distribution can be studied by means of nugget (unexplained variability referring to noise in the data), partial sill (structural component of the total variance) and range (range of the variogram) analysis. The geostatistical operations were performed with R (version 1.17), using the gstat package (Pebesma, 2004).

To evaluate the rate of soil property variability that was theoretically lost as the spatial resolution increased, the variability in the predicted clay contents for each aggregated image was estimated from the map of clay content predicted using the 5*m_AISA* airborne data. For this purpose, grids corresponding to square surfaces, centered on the observed sites with sizes corresponding to the tested spatial resolutions, were created, and the inner-grid variability, V, was calculated according to the following equation:

$$V = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{p}\sum_{j=1}^{p}\sqrt{\left(z_i\left(x_j\right) - \overline{z}_i\right)} \quad (8)$$

where, N is the number of calibration soil samples, p is the number of pixels in the grids, $z_i(x_j)$ is the predicted soil property, i.e., the clay content of the pixel x_j in the grid centered on the *ith* calibration sample, and $\overline{z}_i$ is the mean of the predicted clay contents inside the grid centered on the *ith* calibration sample. The number of pixels, p, depends on the studied spatial resolution. Because the original spatial resolution was 5 m and the grids must be centered in the calibration pixels, the variability could not be measured exactly for the GSD values proposed in this study. The following five sizes were constructed: 3×3 pixels ($p = 9$,

spatial resolution of 15 m), 5 × 5 pixels (p = 25, spatial resolution of 25 m), 7 × 7 pixels (p = 49, spatial resolution of 35 m), 13 × 13 pixels (p = 169, spatial resolution of 65 m) and 19 × 19 pixels (p = 361, spatial resolution of 95 m) (Fig. 3).

5. Results

5.1. Preliminary results

A reference PLSR model, following *Type* 1 (Section 3.4), was built from the 5*m_AISA* airborne image and the 120 clay content values associated with the 120 available soil samples. Ninety samples were assigned to the calibration database, and thirty to the validation database. The performance of this model was relatively accurate, with an R^2_{cal} value of approximately 0.74, an R^2_{val} value of approximately 0.75, a RPIQ of 3.2 and an RMSEP value of approximately 86 g/kg (Table 2), as previously shown in Gomez, Lagacherie, and Bacha (2012). The inner-pixel variability, *V* (Eq. 8), ranges from 27 g/kg (with $N = 82$ grids of 15 × 15 m^2 and $p = 9$) to 66 g/kg (with $N = 65$ grids of 95 × 95 m^2 and $p = 361$) (Fig. 4).

This reference PLSR model was applied to all the spectra over bare soil from the 5*m_AISA* airborne image to produce a predicted clay content map. The model is referred to as the 5*m_AISA_Clay* map (Fig. 5a).

Contrasting clay content and soil patterns appear between the Pliocene area, located in the southeast corner of the image, and the Miocene area, covering the remainder of the image. The Pliocene area exhibits low and weakly variable topsoil clay content, whereas the Miocene area shows a large range of clay content values. Variations within the Miocene area are also visible, and follow the geological pattern formed by the alternating sandstone and marl outcrops, yielding low (blue) and high (red) values of clay content, respectively (Fig. 5a). The predicted clay content values of the 5*m_AISA_Clay* map followed a normal distribution, were centered on 447 g/kg and had a standard deviation of 160 g/kg, and a skewness of 0.17 (Fig. 5c in blue).

The empirical variogram of the predicted clay content values exhibited spatial structures with clear increases in the semi-variances as distance between pairs of points increased (Fig. 5d). An exponential model was sufficient to fit the variogram of the predicted clay content values (blue line on Fig. 5d). The theoretical variogram is characterized by a nugget of approximately 3900, a sill of approximately 16,000 and a range of approximately 380.

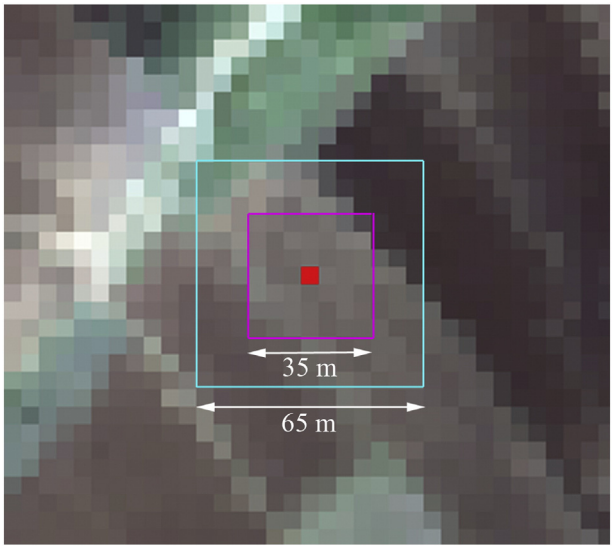


Fig. 3. Grid of 7 × 7 pixels (pink square) and 13 × 13 pixels (cyan square) which correspond respectively to 35 m and 65 m of spatial resolution. The pixel in red indicates the center location a calibration site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Sensitivity to atmospheric effects

A PLSR model was built following *Type* 1, using the 120 spectra associated with the bare soil pixels of the 5*m_HYPXIM* image, for which soil samples were collected and clay contents were measured. The models built from the 5*m_HYPXIM* and 5*m_AISA* images were compared (Table 2). The best model was obtained from the 5*m_AISA* image (Table 2, $R^2_{val} \approx 0.75$ and RPIQ ≈ 3.2). However, the performance of the model obtained from the satellite-simulated data at 5 m is only slightly less accurate (Table 2, $R^2_{val} \approx 0.71$ and RPIQ ≈ 3). Therefore, the atmosphere appears to slightly affect the performance of the PLSR-based prediction models.

No significant difference in the b-coefficients appears between the models (Fig. 6a). Only the b-coefficients between 0.932 and 1.056 μm and 1.182 and 1.25 μm were more important for the PLSR-model built from the satellite-simulated data than from the airborne data. No significant difference in the VIP values appears between both models (Fig. 6b). Only the VIP values between 1.5 and 1.77 μm were more important for the PLSR model built from the airborne data than from the satellite-simulated data. However, the same important wavelengths were identified regardless of the PLSR-model used. Finally, the analysis of the combination of the VIP and the b-coefficients shows that the most important bands are located from 2.2 to 2.4 μm in both images, including the spectral absorption band of clay at 2.206 μm (Fig. 6a and b).

The PLSR model built following *Type* 1, and using the 120 spectra associated with the bare soil pixels of the 5*m_HYPXIM* image, was then applied to all the spectra over bare soil in the 5*m_HYPXIM* image. Therefore, a predicted clay content map was obtained for the study area from this PLSR model and was called the 5*m_HYPXIM_Clay*_1 map (Fig. 5b). No significant difference can be observed between maps 5*m_HYPXIM_Clay*_1 and 5*m_AISA_Clay* via visual inspection (Fig. 5a and b). The same large pedological units can be identified, indicating that both of the PLSR models made the same relative predictions. Moreover, no difference occurs in the absolute prediction, considering the distribution of the predicted clay content over the entire area (Fig. 5c). The predicted clay content values of the 5*m_HYPXIM_Clay* map, follow a normal law centered on 449 g/kg, a standard deviation of 155 g/kg and a skewness of 0.19 (Fig. 5c). Finally, no difference of prediction appears, and the same pedological variations occur in the semi-variogram of the study area (Fig. 5d). An exponential model was sufficient to fit the variogram of the predicted clay content values (pink line on Fig. 5c). Moreover, the theoretical variogram is characterized by a nugget of approximately 3800, a sill of approximately 15,000 and a range of approximately 380 m.

5.3. Sensitivity to degradation of spatial resolutions

The analysis of the PLSR models, built from the seven simulated-satellite images, allowed the identification of two groups of sensors with specific behaviors.

A first group of spatial resolutions, from 5 to 30 m, corresponding to HYPXIM, SHALOM, PRISMA and EnMAP sensors, offers accurate PLSR models with $R^2_{cal} > 0.65$, RMSEC < 100 g/kg and $R^2_{val} > 0.6$ (Table 2). The number of outliers remains stable (between 1 and 3) regardless of the spatial resolution of this first group of sensors. Slight differences in the prediction performance were observed in the validation dataset. The spatial resolutions of 5 and 30 m yielded similar prediction performances, whereas the spatial resolution of 20 m provided a better performance, with an R^2_{val} value of 0.81 and RPIQ value of 3.7. The spatial resolution of 15 m provided the worst performance, with an R^2_{val} value of 0.7 and RPIQ value of 3 (Table 2). Therefore, these results indicate that between 5 and 30 m of spatial resolution, the various model performances do not decrease when the spatial resolution becomes coarser, and the degradation in spatial resolutions of the images does not linearly affect model performances. Finally, no significant difference in the VIP and b-coefficient values appears between the first group

Table 2
Results of the PLSR models.

	Type of PLSR model	Number of latent variables	Number of calibration data	R^2_{cal}	RMSEC (g/kg)	Number of validation data	R^2_{val}	RMSEP (g/kg)	RPD	RPIQ	Name of predicted clay map
5m_AISA	Type_1	5	90	0.74	86	30	0.75	86	2	3.2	5*m_AISA_Clay*
5m_HYPXIM	Type_1	6	90	0.74	86	30	0.71	94	1.9	3	5*m_HYPXIM_Clay*_1
5m_HYPXIM	Type_1	6	62	0.68	99	20	0.71	90	1.9	3	5*m_HYPXIM_Clay*_2
10m_HYPXIM-SHALOM	Type_1	6	62	0.71	94	20	0.6	105	1.6	2.6	10*m_HYPXIM_Clay*
15m_SHALOM	Type_1	6	62	0.67	99	20	0.7	91	1.9	3	15*m_SHALOM_Clay*
20m_PRISMA	Type_1	6	62	0.72	91.22	20	0.81	73	2.3	3.7	20*m_PRISMA_Clay*
30m_PRISMA-ENMAP	Type_1	6	62	0.66	100	20	0.7	90	1.9	3	30*m_PRISMA-ENMAP_Clay*
60m_HyspIRI	Type_2	4	67	0.21	156						
90m_GSM	Type_2	4	65	0.36	135						

models (Fig. 7a and b). The most important bands are located from 1.6 to 1.8 μm and from 2.2 to 2.4 μm for all of the models, including the spectral absorption band of clay at 2.206 μm (Fig. 7a and b). The spectral range from 1.6 to 1.8 μm can include information on soil moisture at the acquisition time, which may be related to clay. The visualization of three spectra, extracted from the 5*m_HYPXIM* and 30*m_PRISMA-EnMAP* images, and corresponding to 3 different clay contents (122, 489 and 772 g/kg), show only a slight difference of reflectance (Fig. 8). The most important difference between spectra at both spatial resolutions comes from the albedo. Nevertheless, this difference of albedo is not the same for each target. A higher albedo exists for the 30*m_PRISMA-EnMAP* image than for the 5*m_HYPXIM* image, for targets with clay contents of 489 and 772 g/kg. An inversely lower albedo exists for the 30*m_PRISMA-EnMAP* image than for the 5*m_HYPXIM* image, for targets with clay content of 122 g/kg (Fig. 8).

A second group of spatial resolutions corresponding to the HyspIRI sensor and GSM requirement, and greater than 30 m, produced inaccurate PLSR models, with an $R^2_{cal} < 0.4$ and RMSEC > 130 g/kg (Table 2). The analysis of the combination of the VIP and b-coefficients shows that only approximately ten spectral bands, centered at 2.206 μm, were significant for the model at 60 m, and no significant spectral bands were identified at 90 m (Fig. 9a and b).

Each PLSR model built from the 82 soil samples of the common dataset was applied to all the spectra over bare soil for its corresponding simulated-satellite image, generating predicted clay maps called 5*m_HYPXIM_Clay*_2, 10*m_HYPXIM-SHALOM_Clay*, 15*m_SHALOM_Clay*, 20*m_PRISMA_Clay* and 30*m_PRISMA-EnMAP_Clay*. Because the PLSR models built from the simulated-satellite images at 60 and 90 m were inaccurate, no predicted clay maps were built at these resolutions. Regardless of the spatial resolution from 5 to 30 m, no large difference occurs in the absolute prediction in regard to the distribution of the predicted clay content over the entire area (Fig. 10).

The Kamech catchment is characterized by strong variations in soil patterns on a small scale, with a close succession of clay-rich areas and clay-poor areas, oriented northwest/southeast, corresponding to marl and sandstone outcrops, respectively. The 5*m_HYPXIM_Clay*_2 map, which was restricted to the Kamech catchment, shows these variations in the soil pattern (Fig. 11a). Certain mixed areas also appear in transition areas between these outcrops, and in shoal areas. The reduction in the spatial resolution resulted in a decrease in the number of bare soil pixels that can be used in VNIR/SWIR spectrometry, providing only a partial view of the succession of the clay-rich areas and clay-poor areas (Fig. 11). Particularly, clay-rich areas with a width less than 60 m disappear when a spatial resolution of 30 m is used (Fig. 11e). For the intermediate spatial resolutions (10 to 20 m), marl and sandstone outcrops can still be observed, but the boundaries between these outcrops are unclear (Fig. 11b, c and d).

6. Discussion

6.1. Impact of atmosphere compensation

From our simulations of the satellite data, the atmosphere compensation appears to only slightly affect the performances of the

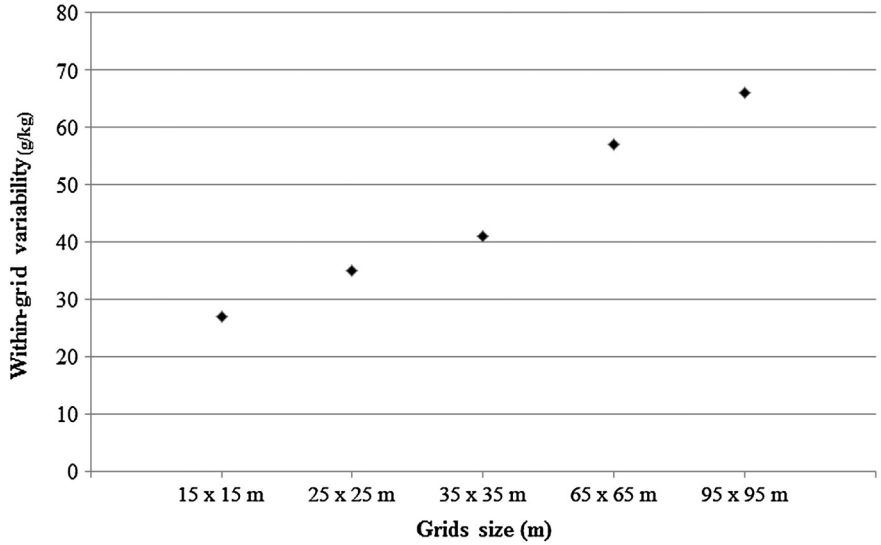


Fig. 4. Within-grid variability *V* of the clay contents predicted from the *AISA_5m* image over surfaces of 15 × 15 m, 25 × 25 m, 35 × 35 m, 65 × 65 m and 95 × 95 m.

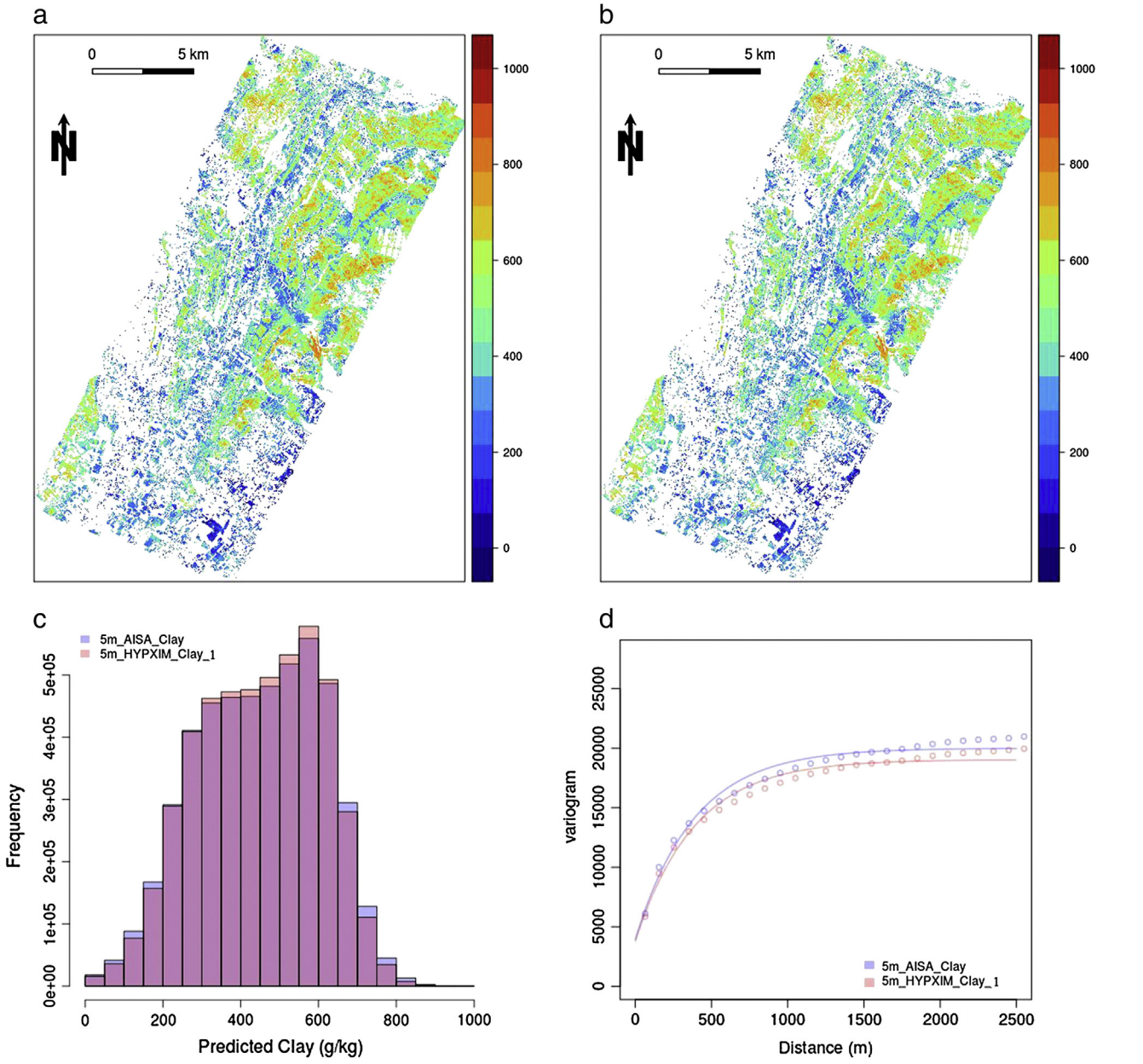


Fig. 5. a) 5*m_AISA_Clay* map, b) 5*m_HYPXIM_Clay*_1 map, c) histograms and d) empirical variograms (points) and theoretical variograms fitted (lines) of clay predictions of the 5*m_AISA_Clay* (blue) and 5*m_HYPXIM_Clay*_1 (red) maps (with lag of 100 m). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

PLSR-based prediction models. The best model was obtained from the AISA-DUAL airborne image. The performance of the PLSR model obtained from the satellite-simulated data at 5 m is only slightly less accurate. The most important difference in the performance may occur from a misunderstanding of the atmosphere. However, in our case study, the water vapor content is very low because such an error in this parameter will not induce a significant difference in the PLSR performance. Furthermore, the visibility during the flight was very high (40 km, close to Rayleigh conditions). Therefore, the effect of the aerosols on the simulated satellite and the retrieved reflectance data can both be considered negligible. For such a low aerosol load, the type of aerosol has no effect. Considering the previous and numerous validation studies conducted using COMANCHE (Miesch et al., 2005) and COCHISE (using Modtran kernel), the errors introduced by the models themselves can be neglected. The results obtained from the satellite-simulated data at 5 m correctly reflect the performances that will be obtained from future satellite images at 5 m of spatial resolution, and under these atmospheric conditions.

6.2. Short-scale clay content variations as a driver of spatial resolutions impact

The spatial resolutions from 5 to 30 m appear to only slightly affect the performances of the PLSR-based prediction models (Table 2). The results of the calibration and validation datasets indicate that the PLSR model performances do not decrease when the spatial resolution becomes coarser, and the degradation in spatial resolution of the image does not linearly affect the model performance. The spatial resolutions of 5 and 30 m yielded similar prediction performances, whereas the spatial resolution of 20 m provided better performance. The slight differences in the performances may arise from the calibration dataset, which contains an acceptable, but not a significant number of data to

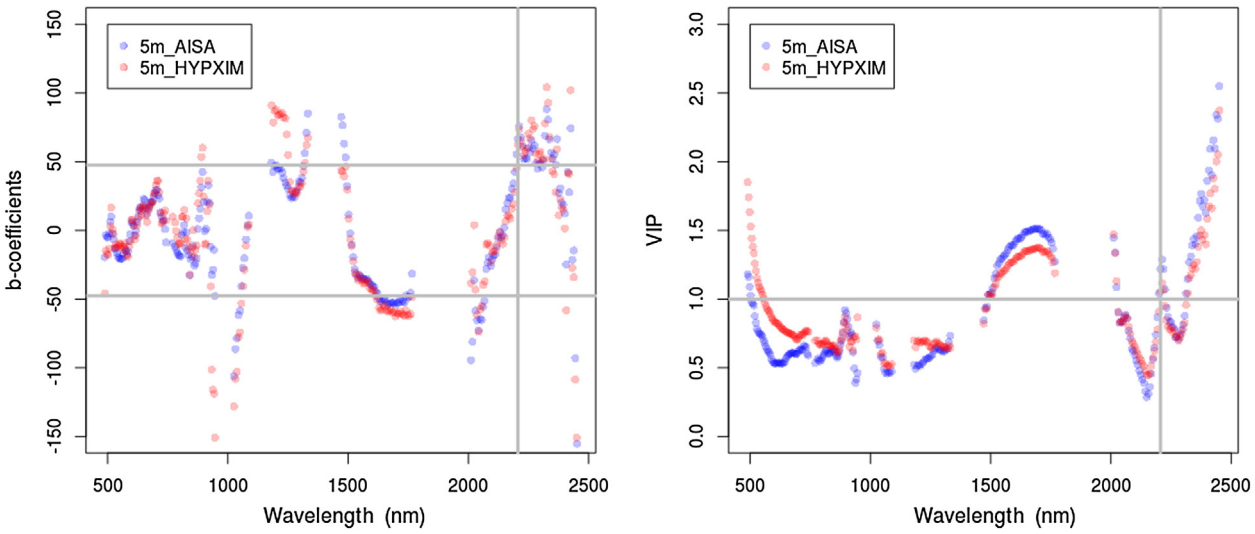


Fig. 6. b-Coefficients (left) and VIP (right) values for both PLSR models built from the 5*m_HYPXIM* and 5*m_AISA* images, using 120 soil samples. The spectral absorption band of clay at 2.206 μm is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of b-coefficients (their standard deviation) and VIP (threshold at 1).

calibrate the models. The slight effect of the spatial resolutions on the PLSR model performances may be explained by the low rate of short-scale clay content variations, which represents only 22% of the whole variance of clay content in our study area. Indeed, the clay samples collected for characterizing surfaces of 10×10 m^2 were also representative of surfaces of 30×30 m^2 because differences in clay contents remain low within such small areas.

6.3. Field size as a driver of spatial resolutions impact

The results revealed a failure in the performances of the PLSR-based prediction models beyond 30 m of spatial resolution, which can be attributed to two cumulative errors. The first error may arise from the Y-variables of the PLSR models (clay contents), and the second error may be due to the X-variables (soil spectra). The error due to the Y-variables (clay contents) is linked to the natural pedological variations in the study area. The soil samples collected over a 10×10 m^2 surface would not be representative of surfaces larger than 30 m. The clay content measured over the 10×10 m^2 surface would be excessively different from the clay content of the 90×90 m^2 surface. However, the error in the Y-variables cannot completely explain the poor performances observed beyond 30 m because the inner-grid variability of clay content, discussed above, increases no more than linearly beyond 30 m (Fig. 4). The error due to the X-variables (soil spectra) is mainly driven by the changes in the topsoil surface conditions, which add a "noise" to the calibration spectra. In the study area, the mean area of the cultivated fields is 0.56 ha (Jenhaoui, Raclot, & Lamachère, 2008). Therefore, a pixel of 60×60 m^2 or 90×90 m^2 can straddle two fields, and can include two types of surface roughness and soil humidity due to the different types of plowing (Fig. 3).

It is also worth noting that degradation in spatial resolution leads to a slight decrease in the suitable surface area that can be mapped using VNIR/SWIR spectroscopy (Section 3.4), and leads to a decrease in the number of soil samples associated with bare soil pixels. When the pixels

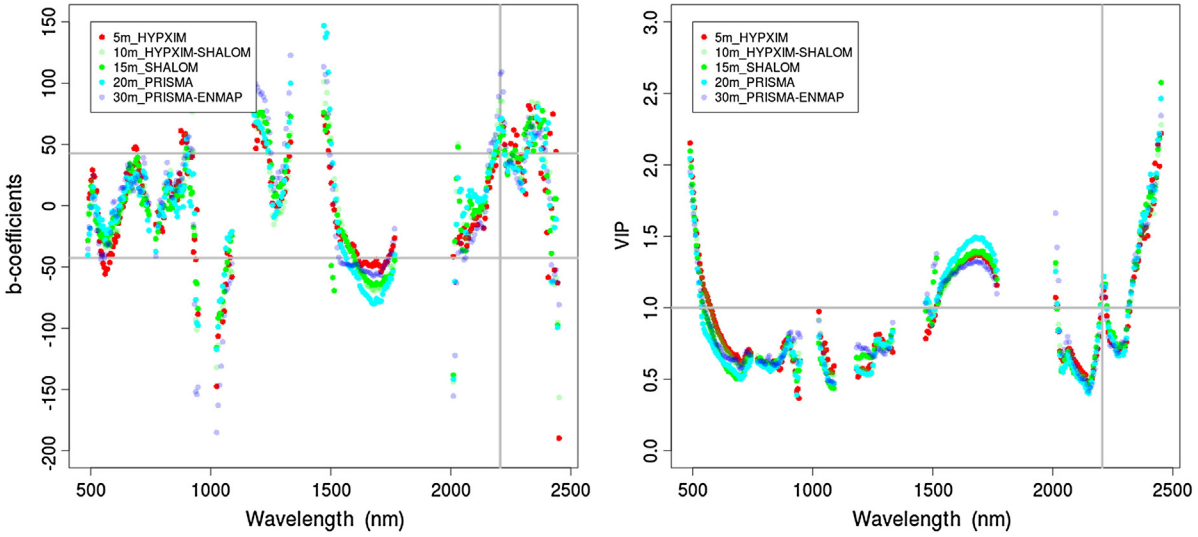


Fig. 7. b-Coefficients (left) and VIP (right) values for PLSR models built from simulated-satellite images, using the 82 soil samples of the *CalBase-5To30m* dataset. The spectral band of clay at 2.206 μm is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of VIP (threshold at 1) and b-coefficients (calculated from the b-coefficients of the model based on the 5*m_HYPXIM* data).

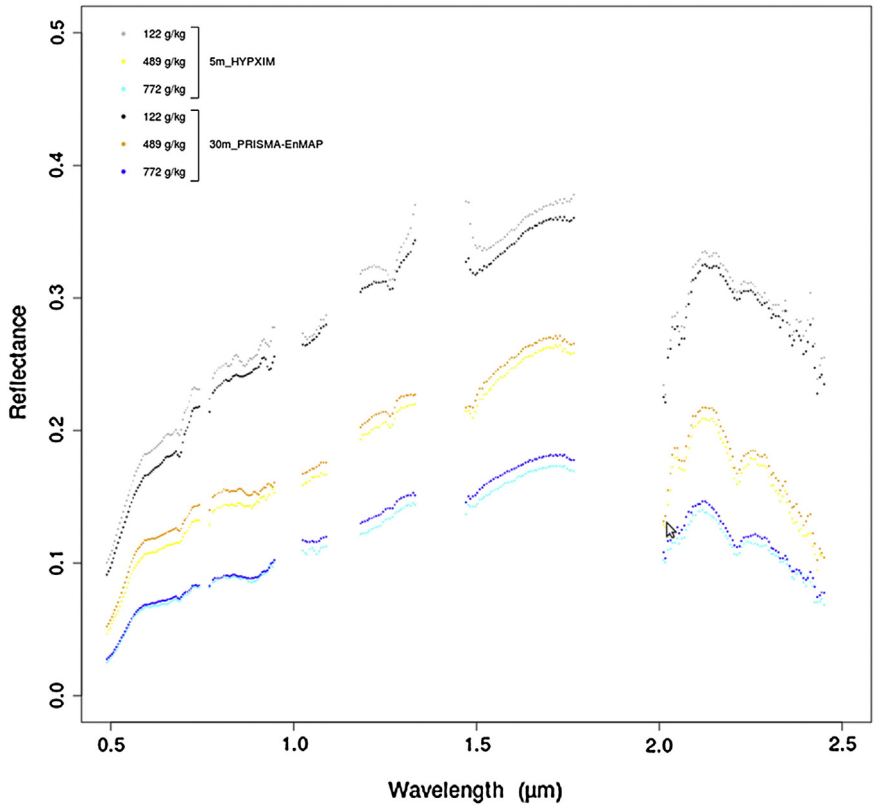


Fig. 8. Reflectance spectra of 3 pixels corresponding to 3 different clay contents (122, 489 and 772 g/kg) extracted from the 5*m_HYPXIM* and 30*m_PRISMA-EnMAP* images.

of the study area become coarser, an increasing number of pixels become inadequate for calibration and prediction because they more often include heterogeneous land use. The smaller the field size, the more prominent the phenomenon is. VNIR-SWIR airborne spectroscopy has proven to be a powerful tool for estimating soil properties over pixels covered by bare soils, whereas the applicability of spectroscopic techniques decreases when the pixels are mixed, such as when a soil surface is partially covered with vegetation (Bartholomeus et al., 2010; Ouerghemmi et al., 2011). However, this limitation would not be insurmountable if the development of the unmixing methods becomes an extensive research area.

6.4. Exploration of results for future hyperspectral-satellite-based soil predictions

The relation between spatial resolutions and performance of clay content predictions found in this study is specific to this area, and might differ for other areas, soil properties or pedological contexts.

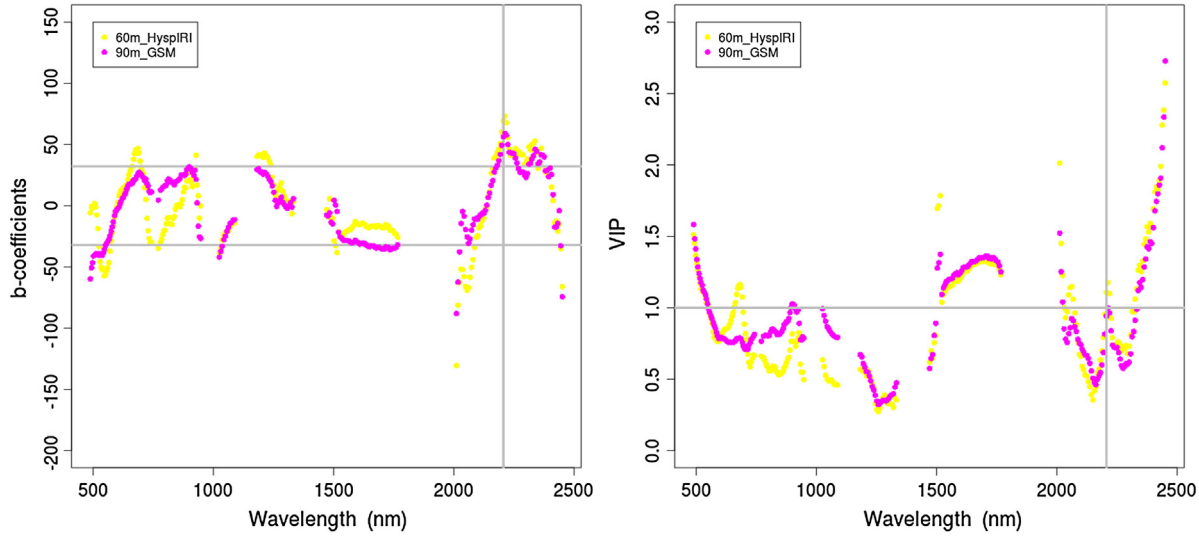


Fig. 9. b-Coefficients (left) and VIP (right) values for PLSR models built from the simulated-satellite images at 60 and 90 m of spatial resolution, using the soil samples of the *CalBase*-60*m* and *CalBase*-90*m* datasets respectively. The spectral band of clay at 2.206 µm is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of VIP (threshold at 1) and b-coefficients (calculated from the b-coefficients of the model based on the 60*m_HysplRI* data).

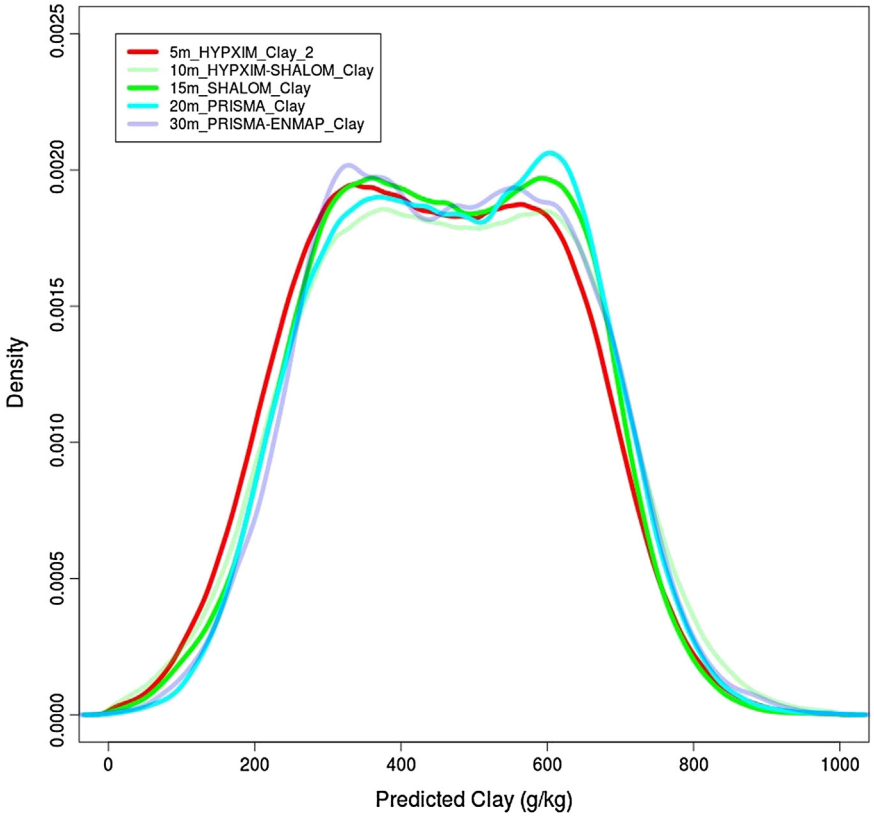


Fig. 10. Histograms of clay predictions obtained over the entire area, from the 5*m_HYPXIM*, 10*m_HYPXIM-SHALOM*, 15*m_SHALOM*, 20*m_PRISMA* and 30*m_PRISMA-ENMAP* images using the 82 soil samples of the *CalBase-5To30m* dataset.

Nevertheless, the slight effect of the spatial resolutions from 5 to 30 m on the PLSR model performances can reasonably be extrapolated to areas where the soil properties exhibit moderate spatial variabilities at short distances, as was observed in this case study. An analysis of the variograms of the clay contents gathered by McBratney and Pringle (1999) (Fig. 2b, p 132) showed that 77% of these variograms had a smaller semi-variance at a lag value of 30 m than the variogram in our case study. This finding suggests that the results obtained in this study could be extrapolated over a significant portion of the planet. Similarly, the decrease in performances observed beyond 30 m is closely related with the small sizes of fields in our study area (mean size = 0.56 ha). However, these may represent important surfaces, such as those covered by traditional and familial agrosystems.

Finally, by identifying two main drivers of spatial resolution impact, namely short scale soil variability and field size, this study paves the way to an a priori assessment of the performances of hyperspectral-satellite-based soil predictions from measurable properties of continental surfaces. More studies similar to the one presented in this paper will be necessary to strengthen this assessment.

6.5. Comparisons with other targeted properties of continental surfaces

Based on a literature review, although different spatial resolutions are estimated to be optimal depending on the application, these optimal resolutions are all included from 6 to 20 m. Rahman et al. (2003) found that a pixel size of 6 m or less would be optimal for studying a canopy- to watershed-level ecosystem, and a pixel size of 20 m or more would imply a failure to identify the variability of the plants. Thorp et al. (2013) showed that the optimal spatial resolution for studying vegetation features in their study area was less than 15 m. Therefore, the soil does not present high specificities regarding spatial resolutions of future hyperspectral satellite sensors.

6.6. Future research

Residue cover, soil roughness and moisture content are environmental factors that are known to affect soil spectra collected in the field and in laboratory (e.g., Liu et al., 2002; Ouerghemmi et al., 2011). Therefore, these environmental factors are expected to also affect the soil spectra recorded by imaging sensors, including proxy, Unmanned Aerial Vehicles, airborne or satellite sensors. Currently, few publications address the consideration of these environmental factors in spectra collection and prediction models (e.g., Haubrock et al., 2008; Lagacherie et al., 2008). Futures research should focus on how spectral information from surface soils and soil prediction models is affected by disturbances in soil surface conditions, such as roughness, residues and variations in soil moisture, and at different measurement scales (i.e., different spatial resolutions). Taking into account environmental factors may help the research community to build more performant prediction models, in waiting the hyperspectral satellite data.

7. Conclusions

Aerial Vis–NIR hyperspectral imaging sensors have previously shown their effectiveness for use in soil surface mapping. Therefore, the availability of future hyperspectral satellite data might offer considerable opportunities for mapping topsoil properties over large areas. In this study, we found that spatial resolutions from 5 to 30 m provided suitable surface clay contents maps in our Mediterranean study area. We also demonstrated that the impact of spatial resolution is dependent on the short-scale spatial variability of the soil property and field size. Thorp et al. (2013) found that the spatial resolution is of secondary importance to the spectral resolution for mapping green vegetation, nonphotosynthetic vegetation and bare soil using a Multiple Endmember Spectral Mixture (MESMA). All the future hyperspectral spaceborne sensors are expected

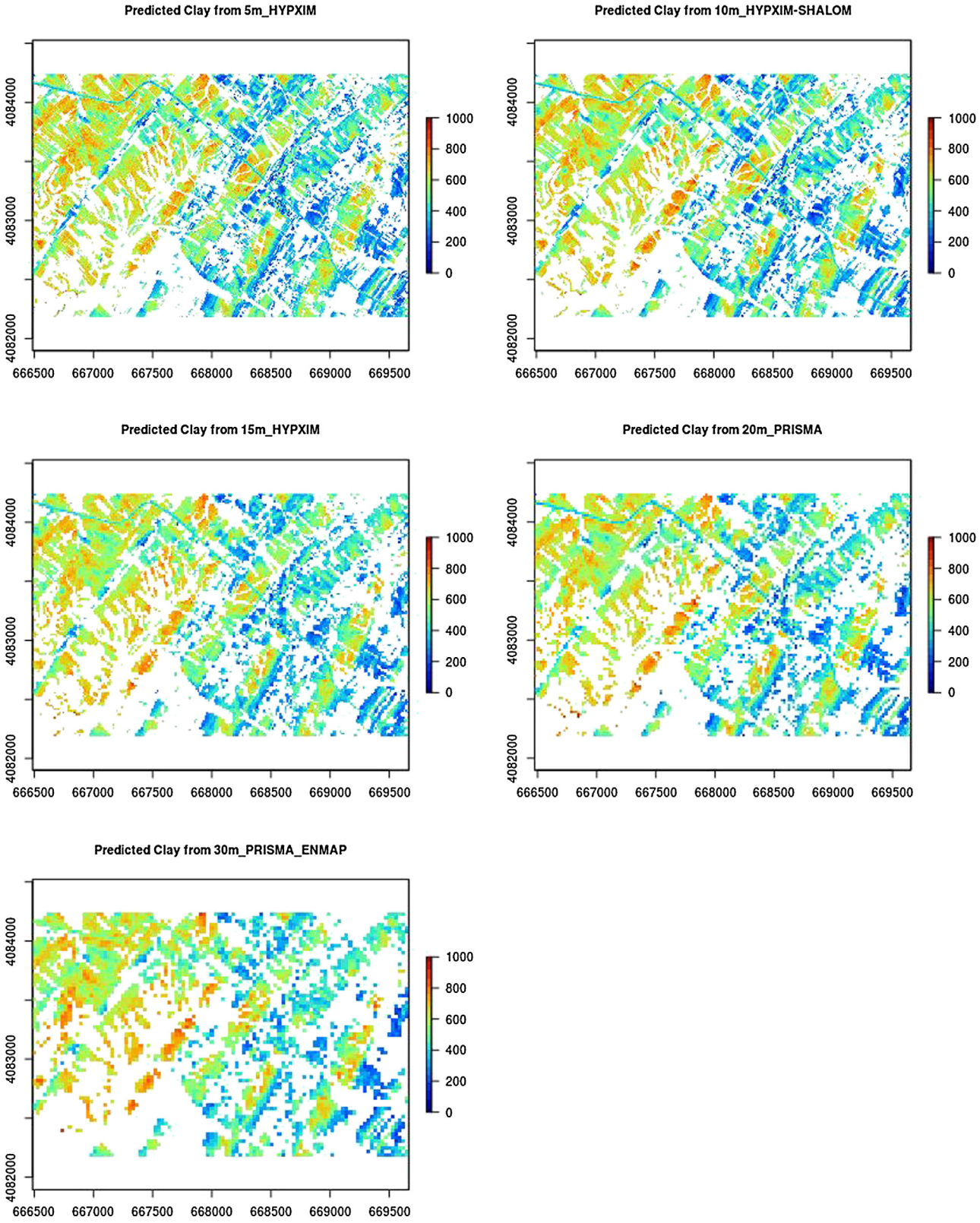


Fig. 11. a) 5*m_HYPXIM_Clay*_2 map, b) 10*m_HYPXIM-SHALOM_Clay* map, c) 15*m_SHALOM_Clay* map, d) 20*m_PRISMA_Clay* map and e) 30*m_PRISMA-ENMAP_Clay* map over the Kamech catchment (upper rectangle in Fig. 1b).

to acquire reflectance data with similar spectral resolution of approximately 10 nm (Staenz et al., 2013), which is close to the resolution used by Vis–NIR hyperspectral imaging airborne sensors, such as HYMAP or AISA-DUAL. As these airborne sensors have already shown their potentials for soil surface property mapping, we can be confident in the performances of future hyperspectral spaceborne sensors. Nevertheless, it would be interesting to study the sensitivity of soil property prediction results to degradation in spectral resolutions, as has been initiated by Casa, Castaldi, Pascucci, Palombo, and Pignatti (2013), as well as the signal to noise ratio. This could potentially improve the design of

future spaceborne sensors, in terms of spectral resolutions and signal to noise ratio. Finally, the optimal spatial resolutions found in this study are specific to this area, and likely differ in other areas with different soil properties. To reinforce these conclusions and the guideline, this study could be enlarged to include additional soil properties and additional pedological settings.

Acknowledgments

This research was granted by the TOSCA-CNES project «HUMPER-Mission HYPXIM : Apport de la résolution spatiale de la mission HYPXIM pour l'étude des propriétés pérennes des sols et de leur humidité de surface» (2013–2014). This research was also supported by the French National Research Agency (ANR) through the ALMIRA project (ANR-12-TMED-0003). The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National de Cartographie et de Télédétection, Tunisia), for providing the AISA-DUAL hyperspectral images for this study. This hyperspectral data acquisition was granted by IRD (Institut de Recherche pour le Developpement), INRA (Institut National de la Recherche Agronomique) and the French National Research Agency (ANR) (ANR-08-BLAN-C284-01).

References

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., et al. (2014). Chapter three — GlobalSoilMap: Toward a fine-resolution global grid of soil properties. In Donald L. Sparks (Ed.), *Advances in Agronomy*, *Volume 125*. (pp. 93–134). Academic Press.

Baize, D., & Jabiol, B. (1995). *Guide pour la description des sols.* Paris: INRA édition.

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1993). Correction to the description of standard normal variate (SNV) and de-trend transformations in practical spectroscopy with applications in food and beverage analysis — 2nd edition. *Journal of Near Infrared Spectroscopy*, *1*, 185–186.

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., VanWesemael, B., Ben-Dor, E., et al. (2010). Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 81–88.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J. M., & McBratney, A. (2010). Prediction of soil attributes by NIR spectroscopy. A critical review of chemometric indicators commonly used for assessing the quality of the prediction. *TrAC Trends in Analytical Chemistry*, *29*(9), 1073–1081.

Ben-Dor, E., & Banin, A. (1995). Near infrared analysis (NIRA) as a simultaneously method to evaluate spectral featureless constituents in soils. *Soil Science*, *159*, 259–269.

Ben-Dor, E., Kafri, A., & Varacalli, G. (2014). SHALOM: An Italian–Israeli hyperspectral orbital mission — Update. *International Geoscience and Remote Sensing Symposium, Quebec, Canada, July 13–18, 2014*.

Ben-Dor, E., Patkin, K., Banin, A., & Karnieli, A. (2002). Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—A case study over clayey soils in Israel. *International Journal of Remote Sensing*, *23*, 1043–1062.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., et al. (1999). MODTRAN 4 radiative transfer modeling for atmospheric correction. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, *3756*, 348–353.

Briottet, X., Marion, R., Carrere, V., Jacquemoud, S., Bourguignon, A., Chami, M., et al. (2013). HYPXIM: HYPXIM: A second generation high spatial resolution hyperspectral satellite for dual applications. *5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), June 2013, Gainesville, Florida, USA.*

Bussoletti, E. (2012). Space observations for agriculture and food support. *Inter-Agency Meeting on Outer Space Activities: 2012. Thirty-second session, 7–9 March 2012. Rome, Italy*.

Carrere, V., Bourguignon, A., Briottet, X., Chami, M., Chevrel, S., Jacquemoud, S., et al. (2013). *The French Hyperspectral Earth Observation Science/Defense mission HYPXIM — A second generation high spectral and spatial resolution imaging spectrometer. IGARSS*, 21–26 July 2013, Melbourne, Australia.

Casa, R., Castaldi, F., Pascucci, S., Palombo, A., & Pignatti, S. (2013). A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma*, *197–198*(2013), 17–26.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., et al. (2009). Assessment and monitoring of soil conditions using indexes based on near infrared reflectance (NIR) spectroscopy. *European Journal of Soil Science*, *60*, 770–784.

Chabrillat, S., Goetz, A. F. H., Krosley, L., & Olsen, H. W. (2002). Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sensing of Environment*, *82*(2), 431–445.

Chang, C. -W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, *167*(2), 110–116.

Chang, C. -W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R., Jr. (2001). Near-infrared reflectance spectroscopy—Principal components regression analysis of soil properties. *Soil Science Society of America Journal*, *65*, 480–490.

Chiang, L. H., Pell, R. J., & Seasholtz, M. B. (2003). Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control*, *13*(5), 437–449.

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, *78*(2005), 103–112.

Clark, R. N. (1999). Chapter 1: Spectroscopy of rocks and minerals, and principles of spectroscopy. In A. N. Rencz (Ed.), *Manual of Remote Sensing. Remote Sensing for the Earth Sciences*, *Volume 3*. (pp. 3–58). New York: John Wiley and Sons.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, *89*(B7), 6329–6340.

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., & Shields, T. (1998). The HyMap Airborne Hyperspectral Sensor: The system, calibration and performance. In M. Schaepman, D. Schläpfer, & K. I. Itten (Eds.), *Proc. 1st EARSeL Workshop on Imaging Spectroscopy. 6–8 October 1998, Zurich, EARSeL, Paris.* (pp. 37–42).

Duveiller, G., & Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, *114*, 2637–2650.

Folkman, M., Pearlman, J., Liao, L., & Jarecke, P. (2001). EO1/Hyperion hyperspectral imager design, development, characterization and prediction. In W. L. Smith, & Y. Yasuoka (Eds.), *Hyperspectral Remote Sensing of the Land and Atmosphere. SPIE Proc*, *Vol. 4151*. (pp. 40–51).

Gaffey, S. J. (1987). Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 um): Anhydrous carbonate minerals. *Journal of Geophysical Research*, *92*.

Giampaolo, P., Cisbani, A., De Cosmo, V., Galeazzi, C., Labate, D., & Melozzi, M. (2008). Hyperspectral instruments for earth observation. *International conference on Space Optics, October 14–17 Toulouse, France*.

Gomez, C., Coulouma, G., & Lagacherie, P. (2012). Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. *Geoderma*, *189–190*, 176–185.

Gomez, C., Lagacherie, P., & Bacha, S. (2012). Using an VNIR/SWIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia). In B. Minasny, B. P. Malone, & A. B. McBratney (Eds.), *Digital Soil Assessments and Beyond* (pp. 387–392). Springer.

Gomez, C., Lagacherie, P., & Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, *148*(2), 141–148.

Gomez, C., Viscarra Rossel, R. A., & McBratney, A. B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field VNIR/SWIR spectroscopy: An Australian case study. *Geoderma*, *146*(3–4), 403–411.

Haaland, D. M., & Thomas, E. V. (1988). Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. *Analytical Chemistry*, *60*(11), 1202–1208.

Haubrock, S. -N., Chabrillat, S., Kuhnert, M., Hostert, P., & Kaufmann, H. (2008). Surface soil moisture quantification and validation based on hyperspectral data and field measurements. *Journal of Applied Remote Sensing*, *2*(1).

Islam, K., Stingh, B., & McBratney, A. B. (2003). Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research*, *41*, 1101–1114.

Ismail, R., Mutanga, O., Kumar, L., & Urmilla, B. (2008). Determining the optimal spatial resolution of remotely sensed data for the detection of *Sirex noctilio* infestations in pine plantations IN KwaZulu-Natal, South Africa. *South African Geographical Journal*, *90*(1), 22–30.

Jenhaoui, J., Raclot, D., & Lamachère, J. M. (2008). *Le parcellaire et l'occupation du sol entre 2004 et 2008 sur le bassin versant de Kamech (Cap Bon, Tunisie)*. Tunis: UMR LISAH, IRD (15 pp. + annexes).

Jensen, J. R., & Cowen, D. C. (1999). Urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, *65*(5), 611–622.

Kruse, F. A., Boardman, J. W., Lefkoff, A. B., Young, J. M., Kierein-Young, K. S., & Cocks, T. (1999). The 1999 USA HyMap Group Shoot. *Proceedings, ISSSR'99, 31 October–4 November 1999, Las Vegas, Nevada*.

Kustas, W. P., & Norman, J. M. (2000). Evaluating the effects of subpixel heterogeneity on pixel average fluxes. *Remote Sensing of Environment*, *74*, 327–342.

Lagacherie, P., Baret, F., Feret, J. -B., MadeiraNetto, J., & Robbez-Masson, J. -M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sensing of Environment*, *112*(3), 825–835.

Lagacherie, P., Ruth, A., Gomez, C., Bacha, S., Coulouma, G., Hamrouni, H., Mekki, I., & Bacha, S. (2013). Combining hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia). *Geoderma*, *209–210*, 168–176.

Lagacherie, P., & Gomez, C. (2014). What can GlobalSoilMap expect from Vis–NIR hyperspectral imagery in the near future? In Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, & Alex B. McBratney (Eds.), *GlobalSoilMap: Basis of the global spatial soil information system* (pp. 387–392). CRC Press.

Le Bissonnais, Y. (1996). Soil aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. *European Journal of Soil Science*, *47*, 425–431.

Liu, L., Baret, F., Xingfa, G., Quingxi, T., Lanfen, Z., & Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, *81*, 238–246.

Lopinto, E., & Ananasso, C. (2013). The Prisma Hyperspectral Mission. *33rd EARSeL Symposium, Towards Horizon 2020: Earth Observation and Social Perspectives, Matera (Italy), 3–6 June 2013*.

Lugassi, R., Ben-Dor, E., & Eshel, G. (2010). A spectral-based method for reconstructing spatial distributions of soil surface temperature during simulated fire events. *Remote Sensing of Environment*, *114*(2), 322–331.

Mark, H. L., & Tunnell, D. (1985). Qualitative near-infrared reflectance analysis using Mahalanobis distances. *Analytical Chemistry*, *57*, 1449–1456.
McBratney, A. B., & Pringle, M. J. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture*, *1*, 219–236.
McCarty, G. W., Reeves, J. B., III, Reeves, V. B., Follett, R. F., & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. *Soil Science Society of America Journal*, *66*, 640–646.
Mekki, I., Albergel, J., Ben Mechlia, N., & Voltz, M. (2006). Assessment of overland flow variation and blue water production in a farmed semi-arid water harvesting catchment. *Physics and Chemistry of the Earth*, *31*, 1048–1061.
Mevik, B. -H., & Wehrens, R. (2007). The pls Package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, *18*(2), 1–24.
Miesch, C., Poutier, L., Achard, V., Briottet, X., Lenot, X., & Boucher, Y. (2005). Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *437*(7), 1552–1562.
Nijland, W., Addink, E. A., De Jong, S. M., & Van der Meer, F. D. (2009). Optimizing spatial image support for quantitative mapping of natural vegetation. *Remote Sensing of Environment*, *113*, 771–780.
Ouerghemmi, W., Gomez, C., Naceur, S., & Lagacherie, P. (2011). Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces. *Geoderma*, *163*(3–4), 227–237.
Pax-Lenney, M., & Woodcock, C. E. (1997). The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote Sensing of Environment*, *61*, 210–220.
Pearson, R. K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, *10*(1), 55–63.
Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, *30*, 683–691.
Raclot, D., & Albergel, J. (2006). Runoff and water erosion modelling using WEPP on a Mediterranean cultivated catchment. *Physics and Chemistry of the Earth*, *31*, 1038–1047.
Rahman, A. F., Gamon, J. A., Sims, D. A., & Schmidts, M. (2003). Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland. *Remote Sensing of Environment*, *84*, 192–207.
Roberts, D. A., Quattrochi, D. A., Hulley, G. C., Hook, S. J., & Green, R. O. (2012). Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HysplRI) Decadal Survey mission. *Remote Sensing of Environment*, *117*(2012), 83–101.
Sanchez, P., Ahamed, S., Carré, F., Hartemink, A., Hempel, J., Huising, J., et al. (2009). Digital soil map of the world. *Science*, *325*, 680–681.
Schaaf, A. N., Dennison, P. E., Fryer, G. K., Roth, K. L., & Roberts, D. A. (2011). Mapping plant functional types at multiple spatial resolutions using imaging spectrometer data. *GIScience & Remote Sensing*, *48*(3), 324–344.
Schwanghart, W., & Jarmer, T. (2011). Linking spatial patterns of soil organic carbon to topography — A case study from south-eastern Spain. *Geomorphology*, *126*, 252–263.
Selige, T., Bohner, J., & Schmidhalter, U. (2006). High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, *136*(1–2), 235–244.
Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*, *66*, 988–998.
Shibusawa, S., Imade Anom, S. W., Sato, S., Sasao, A., & Hirako, S. (2001). Soil mapping using the real-time soil spectrophotometer. In G. Grenier, & S. Blackmore (Eds.), *ECPA 2001. Third European Conference on Precision Agriculture*, *Vol. 1*. (pp. 497–508). Agro Montpellier.
Staenz, K., Mueller, A., & Heiden, U. (2013). Overview of terrestrial imaging spectroscopy missions. *Geoscience and Remote Sensing Symposium (IGARSS), 2013* (pp. 3502–3505). IEEE International.
Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., et al. (2010). Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, *158*, 1–2.
Stuffler, T., Kaufmann, H., Hofer, S., Förster, K. -P., Schreier, G., Müller, A., et al. (2007). The EnMAP hyperspectral imager — An advanced optical payload for future applications in Earth observation programmes. *Acta Astronautica*, *61*(1–6), 115–120.
Tenenhaus, M. (1998). *La régression PLS*. Paris: Editions Technip (254 pp.).
Thorp, K. R., French, A. N., & Rango, A. (2013). Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (MESMA). *Remote Sensing of Environment*, *132*, 120–130.
Viscarra Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., & McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Soil Research*, *46*, 1–16.
Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, *131*, 59–75.
Webster, R., & Oliver, M. A. (1990). *Statistical methods in soil and land resource survey*. Oxford: Oxford University Press (316 pp.).
Weng, Y., Gong, P., & Zhu, Z. (2008). Soil salt content estimation in the Yellow River delta with satellite hyperspectral data. *Canadian Journal of Remote Sensing*, *34*(3), 259–270.
Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.
Zante, P., Collinet, J., & Pepin, Y. (2005). Caractéristiques pédologiques et hydrométéorologiques du bassin versant de Kamech, Cap Bon, Tunisie. *UMR LISAH IRD Tunis, DG ACTA Direction des Sols Tunis, INRGREF Tunis* (21 pp. + 6 annexes).
Zhang, T., Li, L., & Zheng, B. (2009). Partial least squares modeling of Hyperion image spectra for mapping agricultural soil properties. *Proceedings of SPIE — The International Society for Optical Engineering*, *7454*.
Zhang, Q. Y., Middleton, E. M., Gao, B. C., & Cheng, Y. B. (2012). Using EO-1 Hyperion to simulate HysplRI products for a coniferous forest: The fraction of PAR absorbed by chlorophyll (fAPAR(chl)) and leaf water content (LWC). *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1844–1852.

Geoderma 276 (2016) 84–92

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Using legacy data for correction of soil surface clay content predicted from VNIR/SWIR hyperspectral airborne images

C. Gomez [a,*], A. Gholizadeh [a,b], L. Borůvka [b], P. Lagacherie [c]

[a] IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France
[b] Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, 165 21, Prague 6, Suchdol, Czech Republic
[c] INRA, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form 8 March 2016
Accepted 20 April 2016
Available online 17 May 2016

Keywords:
VNIR/SWIR spectroscopy
Airborne remote sensing
Clay content mapping
Legacy soil data
Spectral indexes
Linear regression model
Mediterranean context

ABSTRACT

Visible, near-infrared and short-wave infrared (VNIR/SWIR, 0.4–2.5 μm) hyperspectral airborne imaging has been demonstrated to be a potential tool for topsoil property mapping (such as free iron, clay, and organic matter) over bare soils of large areas. Nevertheless, one of the limiting factors of hyperspectral airborne data use for soil property mapping is the need for a set of soil spectra extracted from bare soils pixels of the VNIR/SWIR airborne data and the corresponding soil property values measured over soil samples collected over the bare soils pixels for which soil spectra are extracted. We propose to test a new approach which uses legacy soil data collected over and/or around the study site, instead of soil property values measured over soil samples collected over bare soils pixels. As legacy soil samples can be inaccurately localized or can be located out of bare soils of hyperspectral airborne data or out of the study area, these data could be unusable as calibration data for classical predictive models (such as the partial least-squares regression method). So the proposed approach first uses a spectral clay index to estimate clay contents (in relative values as it is done without calibration) and then transform these estimated clay contents thanks to a correction of the distribution and range of clay content estimations using legacy soil data. This procedure is compared to a linear model built from the spectral clay index and calibrated using a reference database. The spectral index was proposed by Levin et al. (2007) using spectral bands at 2.205, 2.13, 2.224 μm. This study employs the VNIR/SWIR AISA-DUAL hyperspectral airborne data acquired over an area of 300 km^2 in a Mediterranean region. Our results show that *1*) this spectral index offers predictions with low accuracy in terms of the coefficient of determination, R^2, which is associated with high *bias* and *SEP*; *2*) the distribution and range correction made using legacy soil data allows for both an increase of accuracy (R^2) and an improvement of *bias* and *SEP*; *3*) it is better to have a small number of legacy ground measurements focused on the study area than a high number of legacy ground measurements dispersed on and far from the study area; *4*) the correction of the prediction bias is highly dependent on the legacy soil data quality; and *5*) regardless of which legacy soil database is used, the soil pattern is discriminated. With the coming availability of the next generation of hyperspectral VNIR/SWIR satellite data for the entire globe, this study may open a new way toward accessible and cheap methods for the delivery of soil property maps to the geoscience community.

1. Introduction

Visible, near-infrared and short-wave infrared (VNIR/SWIR, 0.4–2.5 μm) spectroscopy is a physical nondestructive, rapid, reproducible, and inexpensive method that allows for the characterization of materials, including soils. VNIR/SWIR spectroscopy has been proven to be accurate for the estimation of a large range of soil properties such as Soil Organic Carbon (SOC) (e.g., Chang and Laird, 2002), pH (e.g. Shepherd and Walsh, 2002), clay (e.g. Chang et al., 2001), iron (e.g. Richter et al., 2009), potassium (K) (e.g., Cozzolino and Moron, 2003) (see also Viscarra Rossel et al., 2006; Ben-Dor et al., 2009).

Following these studies, VNIR/SWIR hyperspectral airborne imaging has been demonstrated to be a potential tool for topsoil property mapping over large areas (e.g., Selige et al., 2006; Gomez et al., 2008; Stevens et al., 2010). Various methods have been applied to relate soil spectra extracted from the VNIR/SWIR image, to soil properties, including spectral indexes (e.g., Gomez et al., 2008; Lagacherie et al., 2008) and partial least-squares regression (PLSR) (e.g., Selige et al., 2006; Stevens et al., 2010). Spectral indexes are based on the physical analyses of spectral reflectance, such as the slope or absorption band depth value, which allow for estimating mineral, rock, and soil properties. For example, the absorption band depth values centered at 2.206 and 2.341 μm can be used to estimate clay (Chabrillat et al., 2002) and Calcium Carbonate ($CaCO_3$) (Gaffey, 1986) contents, respectively. In the PLSR approach, the full spectrum is used to establish a linear regression model in which the

* Corresponding author.
E-mail address: cecile.gomez@ird.fr (C. Gomez).

http://dx.doi.org/10.1016/j.geoderma.2016.04.019
0016-7061/

significant information contained in the VNIR/SWIR spectra is concentrated in a few latent variables that are optimized to produce the best correlation with the desired soil property (Wold et al., 2001). In spite of its demonstrated interest for topsoil property mapping, VNIR/SWIR hyperspectral airborne imaging is still much too costly for being used in operational soil mapping.

A first perspective of cost reduction is the launch of new orbiting hyperspectral VNIR/SWIR sensors, including the German EnMAP sensor (Stuffler et al., 2007), the Italian PRISMA (Lopinto and Ananasso, 2013), the French HypXim sensor (Carrere et al., 2013; Briottet et al., 2013) and the Italy-Israel SHALOM (Ben-Dor et al., 2014) scheduled between 2017 and 2021. These sensors will routinely deliver high-resolution spectral images for the entire globe and so it can be hoped that the cost of future satellite hyperspectral images will be decreased. However, this expected cost decrease will still be hampered by the cumbersome field data collection that is required by the calibration step that is necessary for converting the images of spectra measurements into images of estimated topsoil properties. This step needs to visit a set of sites covered by bare soils during the hyperspectral flight, that is sized large enough for accurately representing the distribution of soil properties over the study area (e.g., Selige et al., 2006; Gomez et al., 2008; Stevens et al., 2010). At these sites, topsoil properties are measured through wet laboratory analysis (Baize and Jabiol, 1995) and the corresponding soil spectra are extracted from the VNIR/SWIR hyperspectral data. These laboratory analysis and corresponding soil spectra are further used as inputs for calibrating a regression model that infer predictions of a soil property from the VNIR/SWIR hyperspectral spectral measurements. Although substantial cost reductions are expected by replacing wet laboratory analysis by the less expensive laboratory VNIR/SWIR spectroscopy analysis, this field data collection will still remain prohibitive for the operational use of VNIR/SWIR hyperspectral imaging in digital soil mapping.

To overcome this problem, it could be highly profitable to mobilize the soil legacy data in the calibration process. These data are the products of several decades of soil surveys all over the world and are thus available in a lot of study areas. They have been already largely used in operational Digital Soil Mapping at global (Hengl et al., 2015), continental (Viscarra-Rossel and Bui, 2016) and regional (Vaysse and Lagacherie, 2015) scale. However the use of legacy soil data in VNIR/SWIR spectrometry has two drawbacks. Firstly, the locations of the sampled sites cannot be controlled for matching both the area covered by the hyperspectral image and the bare soil surfaces included in this area. Consequently, it would be hard to obtain enough couples of legacy soil property values and associated VNIR/SWIR imagery spectra for calibrating the above-evoked regression models. Furthermore, the obtained couples would be highly affected by georeferencing errors since the legacy data were usually georeferenced without any GPS system. Secondly, the legacy soil data are often affected by biases generated by the differences in sampling dates, sampling procedures and soil analysis protocols and practices (Ciampalini et al., 2013). These biases could propagate toward the calibration process and then generate biases in the estimation of soil properties from VNIR/SWIR imaging spectrometry. Several solutions have been proposed for correcting such biases (e.g., Baume et al., 2011; Ciampalini et al., 2013).

In this paper, we propose a new calibration approach that allows the use of legacy data by relaxing the need of controlling the location of the sampling points. The approach combines the use of well-established spectral indexes whose outputs are further corrected by a standardization process that uses the distributions – and not the exact locations – of the locally available soil data. This approach is tested for the mapping of topsoil clay content, which is one of the basic soil properties used by the soil surveyors to describe soil types, a key driver of soil erosion processes (Le Bissonnais, 1996), and one of the primary soil properties included in the GlobalSoilMap specifications (Arrouays et al., 2014). This approach is compared to the above evoked classical approach using a linear regression model. Different actual and legacy soil datasets are considered as input to explore the influence of different rates and sources of biases. This study employs the VNIR/SWIR AISA-DUAL hyperspectral airborne data acquired over a large area (300 km^2) in northern Tunisia (Cap Bon region), in a Mediterranean context.

2. Materials

2.1. Study area

The study area is located in the Cap Bon region in northern Tunisia (36°24′N to 36°53′N; 10°20′E to 10°58′E), 60 km east of Tunis (Fig. 1). This 300 km^2 area includes the Lebna catchment, which is primarily rural (> 90%) and devoted to cereals, legumes, olive trees, natural vegetation for breeding and vineyards. This area is characterized by rolling areas, with an elevation between 0 and 226 m. The climate varies from humid to semi-arid, with an inter-annual precipitation of 600 mm and an annual potential evapotranspiration of 1500 mm. The soil pattern of the Lebna catchment primarily arises from variations in the lithology. The changes in the landscape between the Miocene sandstone and the marl outcrops induce significant variations in the physical and chemical soil properties (Zante et al., 2005). Furthermore, the distance between the successive sandstone outcrops decreases steeply along a sea-mountain direction, which results in variations in the soil property patterns. The soil materials were redistributed laterally along the slopes during the Holocene, which adds to the complexity of the soil patterns. The main soil types are Regosols, Eutric Regosols (preferentially associated with sandstone outcrops), Calcic Cambisol and Vertisol (preferentially formed on marl outcrops and lowlands). The southeastern region of the study area has a flatter landscape, with sandy Pliocene deposits yielding Calcosol and Rendzina.

2.2. AISA-DUAL VNIR/SWIR hyperspectral airborne data

In November 2010, VNIR/SWIR AISA-DUAL hyperspectral data were acquired over the study area (12 × 25 km), with a spatial resolution of 5 m. The AISA-DUAL airborne imaging spectrometer measures the reflected radiance in 359 non-contiguous bands covering the 0.4 to 2.45 μm spectral domain, with 4.6 nm bandwidths between 0.4 and 0.97 μm and 6.5 nm bandwidths between 0.97 and 2.45 μm. The instantaneous field of view (IFOV) was 24°. The radiance units were converted to reflectance units by using the atmosphere compensation tool COCHISE, which is an inverse radiative transfer model (Miesch et al., 2005). To remove the noisy effects introduced by this process, the reflectance spectrum of each pixel *i* was smoothed using a filter that averaged the reflectance using three wavelength neighbors. Noisy or troublesome spectral bands and pixels have also been removed. We removed the following: a) 20 spectral bands in the blue part of the spectral domain from 0.4 to 0.484 μm due to instrumental noise; b) five bands from 0.747 μm to 0.766 μm, which correspond to the O_2 band absorption; c) 12 bands from 0.952 μm to 1.019 μm due to the H_2O band absorption and an instrument defect; and d) 14 bands from 1.094 μm to 1.176 μm, 21 bands from 1.339 μm to 1.465 μm, and 38 bands from 1.773 μm to 2.005 μm due to H_2O band absorption. Therefore, 249 spectral bands between 0.489 μm and 2.451 μm were retained. To isolate the bare soil areas, pixels with normalized difference vegetation index (NDVI) values over an expert-calibrated threshold were masked; a value of 0.20 was determined after considering 20 parcels that were visually inspected in the field. The NDVI was retrieved using bands at 0.672 μm and 0.799 μm. Water areas were also masked using an expert-calibrated threshold; pixels with a reflectance of <8% at 1.665 μm were removed. Finally, 13 urban areas were identified through visual inspection and were masked. Thus, bare soils represent 46.3% of our study area and 5,889,847 AISA-DUAL pixels.

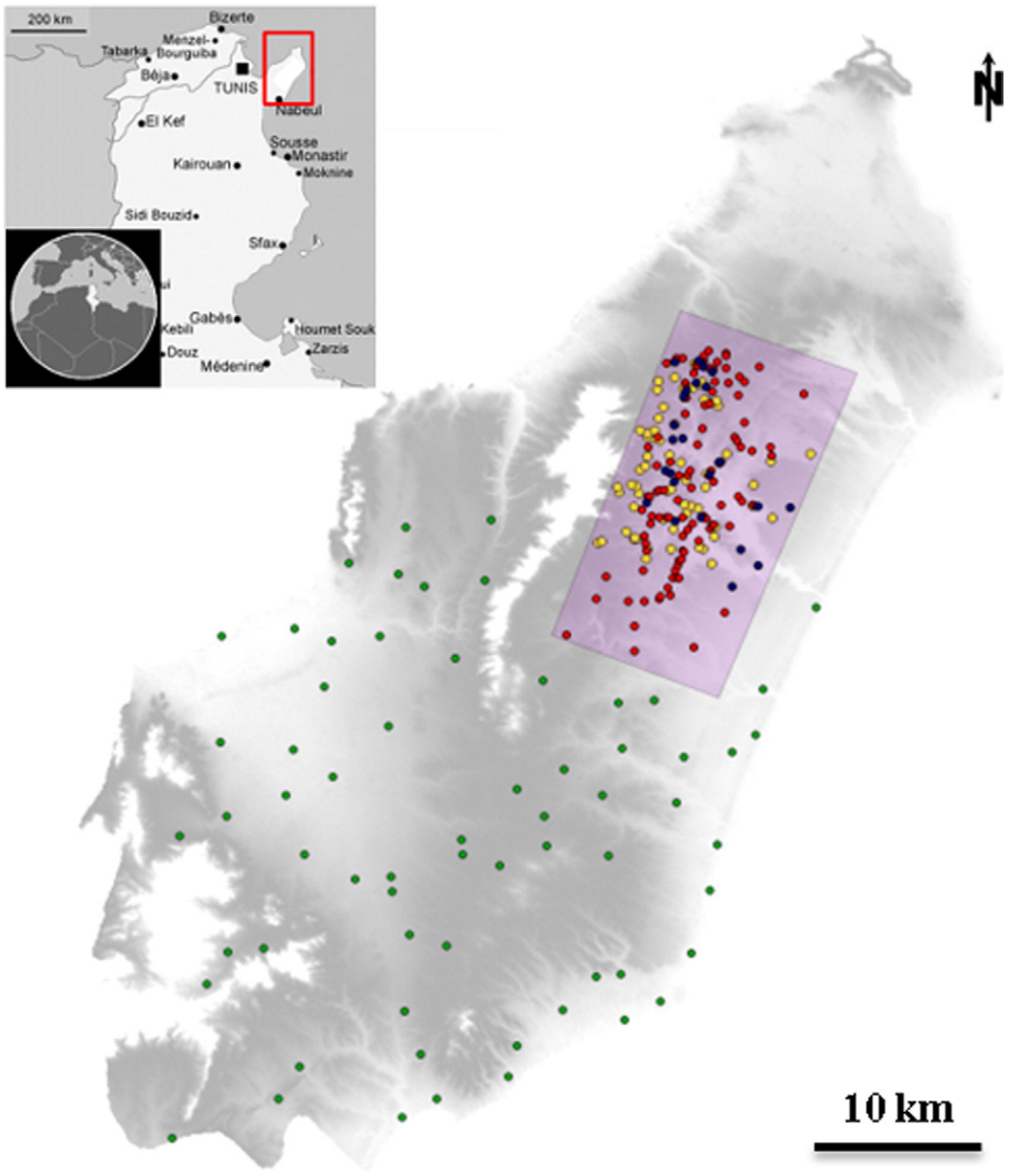


Fig. 1. Hyperspectral image print (purple rectangle) and soil samples location of the *REF* (red), *Focused-Legacy2000* (yellow), *Focused-Legacy2010* (blue), and *No-Focused-Legacy1973* (green) databases, plotted over a DEM of the Cap Bon (Tunisie).

2.3. Soil field sampling and laboratory analysis

The three following soil datasets differed in the dates of collections, the sampling protocols, the georeferencing methods, the laboratory analysis protocols and the data storage methods. They mostly refer to arable land.

2.3.1. Reference soil database

One hundred and twenty soil samples were collected over the study area covered by the AISA-DUAL data. Among this sample set, 50 were collected in June 2008, 30 in October 2009, and 40 in November 2010 (Fig. 1). All of these soil samples were collected in fields that were bare during the hyperspectral airborne data acquisition in November 2010. All of the samples were composed of five sub-samples collected to a depth of 5 cm at random locations within a 10×10 m^2 square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. As soil samples have been collected over plowed fields, the collected soils from 0 to 5 cm can be considered to be composite and homogeneous in terms of soil properties. After homogenizing the sample, approximately 20 g was allocated for clay content analysis. The initial samples were air-dried and sieved with a 2-mm sieve prior to being transported to the laboratory for analysis. The clay content (granulometric fraction <2 μm) was determined using a pipette method (method NF × 31–107, particle size distribution by sedimentation) (Baize and Jabiol, 1995). The clay content of the 120 soil samples varied between 108 and 772 g/kg and followed a normal distribution (Table 1). This dataset will be denoted as *REF*.

2.3.2. Legacy soil databases

Twenty-six soil samples were collected between 2010 and 2012 over the study area covered by the AISA-DUAL data (Fig. 1). They belonged to a spatial set of soil profiles that were purposively selected by skilled soil surveyors to cover as much as possible the soil variability of the area. The soil profiles were geo-referenced using GPS with

Table 1
Statistical description of clay content distributions retrieved from the different soil databases (in g/kg).

	Number of samples	Min	Max	Mean	Std.	Skewness
REF	120	108	772	466	173	−0.11
Focused-Legacy2010	26	106	762	431	210	0.1
Focused-Legacy2000	66	39	762	270	169	0.8
No-Focused-Legacy1973	64	10	425	173	105	0.4

latitude and longitude coordinates in grade and WGS84 as the projection. The soil profiles were described and sampled following current soil observations protocols (Baize and Jabiol, 1995). Among these descriptions, 26 topsoils ranging between 5 and 30 cm in thickness were identified and sampled, which constituted the first legacy database, denoted as "*Focused-Legacy2010*".

Eighty-six soil samples were collected in 2000 over the Cap Bon region, including the study area. Among these samples, 66 samples were selected for their inclusion in the study area covered by the AISA-DUAL, which constituted a second database denoted as the "*Focused-Legacy2000*" database (Fig. 1). These soil samples were collected in the topsoil horizons (between 0 and 35 cm in thickness) of the profiles described by the IAO (Instituto Agronomico per l'Oltremare) 20th course professional master "remote sensing and natural resources evaluation" field survey staff from 2 to 28 April 2000 according to the IAO framework (IAO, 2002). The soil profiles were selected to characterize the soil mapping unit of a 1:100,000 soil map. They were described following the Soil Taxonomy of the USDA (Soil Survey Division Staff, 1993) and the World Reference Base for Soil Resources (FAO, 1998). The soil samples were located initially on topographic maps georeferenced in CGE (Geodetic system Carthage Tunisia) with Lambert coordinates and the Greenwich Meridian as the origin. An important bias between clay content values of the *Focused-Legacy2000* database and those of the *REF* database was revealed by Ciampalini et al. (2013). This bias was assumed to be mainly caused by differences in laboratory analysis protocols.

Seventy-two soil samples were collected between 1973 and 1979 over the Cap Bon region, including the study area. Among these samples, the 64 soil samples located outside of the study area covered by the AISA-DUAL data were considered, which constituted a third soil database denoted as "*No-Focused-Legacy1973*" (Fig. 1). These soil samples corresponded to the topsoil horizons (no information was available regarding thickness) of profiles described following a Tunisian Ministry of Agriculture protocol derived from protocols defined by the French overseas research institute (ORSTOM – Maignien, 1980). The soil profiles were located initially on topographic maps georeferenced in CGE (Geodetic system Carthage Tunisia) with latitude and longitude coordinates in grade and the Paris Meridian as the origin. As for the previous database, an important bias between clay content values of the *No-Focused-Legacy1973* database and those of the *REF* database, related with laboratory protocols, was revealed by Ciampalini et al. (2013).

The clay content (granulometric fraction <2 μm) of the *Focused-Legacy2010* database was determined using a pipette method (method NF × 31–107, particle size distribution by sedimentation, Baize and Jabiol, 1995) at the Laboratory for Soil Analysis (LAS) INRA Arras. The clay content of the *Focused-Legacy2000* and *No-Focused-Legacy1973* databases was determined using the hydrometer method (Baize and Jabiol, 1995) at, respectively, the soil analysis laboratory of IAO and the soil analysis laboratory of the Ministry of Agriculture (DG-ACTA, Tunisia). The clay content of the *Focused-Legacy2010* database varied between 106 and 762 g/kg with a 0.1 skew (Table 1). The clay content of the *Focused-Legacy2000* database varied between 39 and 762 g/kg with a positive skew (Table 1). The clay content of the *No-Focused-Legacy1973* database varied between 10 and 425 g/kg with a positive skew (Table 1).

3. Methods

Different clay content predictions were derived from the AISA-DUAL spectra according to the process flow summarized in Fig. 2. The successive steps are detailed below.

3.1. Spectral index

Clay content predictions $\hat{Z}$ associated with a pixel x were obtained directly from the Clay Index (denoted as the $SWIR_{FI}$ index) proposed by Levin et al. (2007) using reflectance values R initially at spectral bands $\lambda_1 = 2.133$ μm, $\lambda_2 = 2.225$ μm and $\lambda_3 = 2.209$ μm (Fig. 2) as follows:

$$\hat{Z}(x) = \frac{R_{\lambda_1}(x)^2}{R_{\lambda_2}(x) - R_{\lambda_3}(x)^3}. \quad (1)$$

In our case, because of the spectral resolution of the AISA-DUAL data, the $SWIR_{FI}$ index used the spectral bands $\lambda_1 = 2.13$ μm, $\lambda_2 = 2.224$ μm and $\lambda_3 = 2.205$ μm. The clay content predictions $\hat{Z}$ are relative clay content values, as the Eq. (1) is independent of observed clay contents. The HYperspectral SOil Mapper (HYSOMA) software interface was used to derive semi-quantitative soil property maps from the hyperspectral data. HYSOMA was developed at the GZF German Research Center for Geosciences in the Remote Sensing section and is a software package written in the IDL language (www.gfz-potsdam.de/hysoma, Chabrillat et al., 2011).

3.2. Normalization of clay index values

The clay content predictions Z obtained from the $SWIR_{FI}$ index were normalized (Fig. 2). A boxcox transformation was applied to transform the distribution of the clay content predictions to a normal distribution (Legendre and Legendre, 1998), defined as:

$$BC\left(\hat{Z}(x)\right) = \frac{\left(\hat{Z}(x)^a - 1\right)}{a} \quad (2)$$

where $\hat{Z}(x)$ is the clay content prediction obtained from the $SWIR_{FI}$ index for the pixel x, $BC(\hat{Z}(x))$ is the transformed prediction and α is the transformation parameter. The assumption is that among all transformations with an α value between −5 and +5, transformed data have the highest likelihood to be normally distributed – but are not guaranteed – when the standard deviation is the smallest (Legendre and Legendre, 1998). The clay content predictions $BC(\hat{Z}(x))$ are still relative clay content values, as the Eqs. (1) and (2) are independent of observed clay contents. The normalization step was implemented using R software with the *Boxcox* function (Box and Cox, 1964).

3.3. Linear regression model of clay index values

A linear model between the 120 observed clay content of the *REF* database and the 120 normalized clay content predictions $BC(\hat{Z}(x))$ calculated for the bare soil pixels x corresponding to the soil samples sites of the *REF* database, was built (Fig. 2) as follows:

$$LM(Z(x)) = a * BC(Z(x)) + b \quad (3)$$

where a and b are, respectively, the slope and the y-intercept of the linear model, and $LM(\hat{Z}(x))$ is the newly predicted clay contents associated with the pixel x. The linear model was implemented using R software with the *lm* function (Chambers, 1992; Wilkinson and Rogers, 1973).

This linear model was used as a reference approach close to the one follows in literature, which consists to calibrate a spectral index using a set of soil spectra extracted from bare soils pixels of the VNIR/SWIR airborne data and the corresponding soil property values measured over

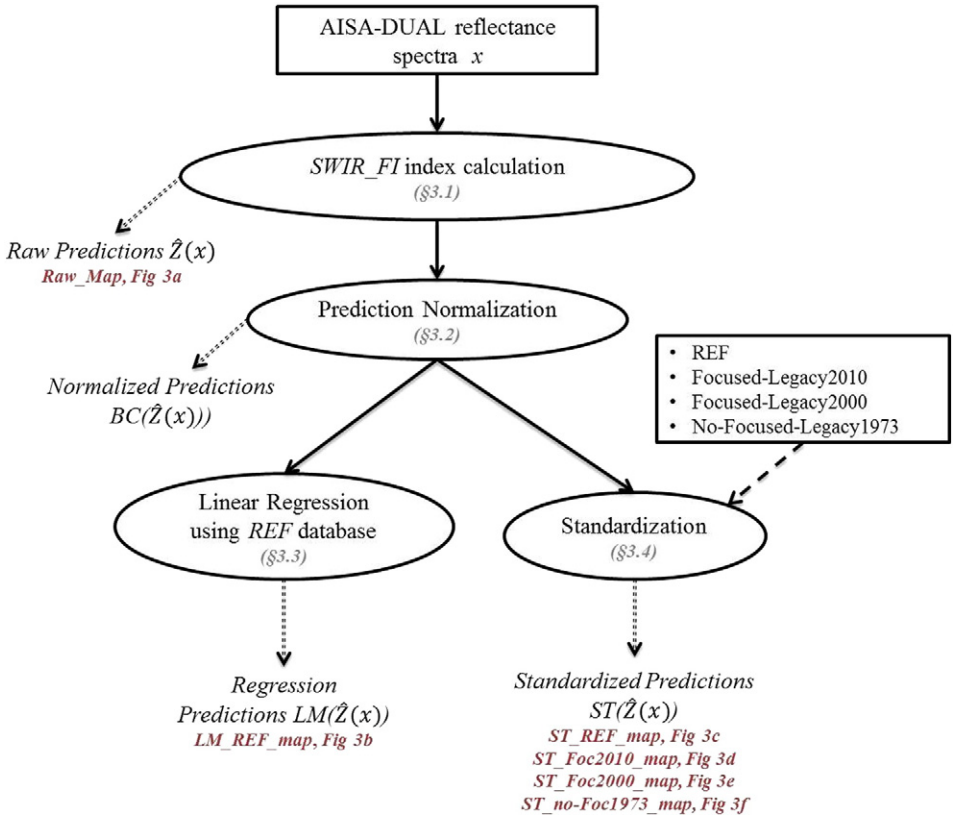


Fig. 2. Process Flow. In grey are indicated the numbers of section where is described the corresponding method. In dark red are indicated the names of resulting map and the corresponding figure in this paper.

soil samples collected over the bare soils pixels for which soil spectra is extracted (following previous researches, e.g., Lagacherie et al., 2008 and Gomez et al., 2008).

3.4. Standardization of clay index values

The clay content predictions $BC(\hat{Z}(x))$ were scaled and centered from a calibration database (x_{cal}) (Fig. 2), as follows:

$$ST\left(\hat{Z}(x)\right) = BC\left(\hat{Z}(x)\right) * c + d \text{ with } c = \frac{sd(Z(x_{cal}))}{sd\left(BC\left(\hat{Z}\right)\right)} \text{ and } d = Z(x_{cal}) - BC\left(\hat{Z}(x)\right) * c \quad (4)$$

where $sd(Z(x_{cal}))$ and $\overline{Z}(x_{cal})$ are, respectively, the standard deviation and the mean of all clay content of the calibration database $Z(x_{cal})$ and $sd(BC(\hat{Z}))$ is the standard deviation of all clay content predictions $BC(\hat{Z}(x))$ over bare soils of the image. This process was implemented using R software. Four calibration databases, which corresponded to the reference database and the three legacy databases described above, were considered.

3.5. Performances indicators

As recommended by Davies and Fearn (2006), prediction performance analyses have to be guided by the standard error of prediction (*SEP*) or root mean square error of prediction (*RMSEP*). The *SEP* is the parameter commonly used in the VNIR spectroscopy literature to describe the prediction ability of a model. *SEP* appears as an averaged error recorded on the validation-sample set and is calculated as follows:

$$SEP = \sqrt{\frac{\sum_{i=1}^{N}(Z(x_i) - \hat{Y}(x_i))^2}{N}} \quad (5)$$

where $Y(x_i)$ is the predicted clay contents [$\hat{Z}(x_i)$, $BC(\hat{Z}(x_i))$, $LM(\hat{Z}(x_i))$ or $ST(\hat{Z}(x_i))$] associated with the pixel x_i, $Z(x_i)$ is the measured clay contents of the sample of the *REF* database corresponding to the pixel x_i and N is the number of soil samples in the *REF* database. The *SEP* value can be decomposed as follows (Davies and Fearn, 2006):

$$SEP^2 = bias^2 + SEP_C^2 \quad (6)$$

where

$$bias = \sum_{i-1}^{N}\frac{\hat{Y}(x_i)}{N} - \sum_{i=1}^{N}\frac{Z(x_i)}{N} \quad (7)$$

$$SEP_c = \sum_{i=1}^{N}\frac{\left(\hat{Y}(x_i) - bias - Z(x_i)\right)}{N} \quad (8)$$

As explained in detail by Bellon-Maurel et al. (2010), the *bias* and SEP_c (for "*SEP* corrected for *bias*") appear as, respectively, the error of means and the residual variance. SEP_c and the *bias* are independent.

The prediction performances of the clay predictions were also evaluated using the coefficient of multiple determination, R^2. This coefficient R^2 is defined as the proportion of variation in the response that can be explained by the regression model and can be expressed as a function of the sum of squares due to regression (SSR), the total sum of squares (SST) and the sum of squares of error (residuals) (SSE), as follows:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{N}\left(Z(x_i) - \hat{Y}(x_i)\right)}{\sum_{i=1}^{N}(Z(x_i) - \overline{Z}(x_i))} = 1 - \frac{SEP^2}{sd(Z(x))}. \quad (9)$$

Because this coefficient of multiple determination R^2 can also be expressed by the *SEP*, its values should be inversely proportional to the *SEP*.

4. Results

The process flow presented in Fig. 2 and Section 3 was followed. The performances of the clay content predictions in each step of the process are detailed below.

4.1. Spectral index predictions with and without calibration using REF database

The clay contents $\hat{Z}$ were predicted directly from the clay spectral index $SWIR_{FI}$ according to Eq. (1). The predictions performance was very poor, with a *SEP* of 439 g/kg, a *bias* of −414 g/kg and a SEP_c of 146 g/kg (Table 2). Because the *SEP* indicator was so high, the coefficient of multiple determination, R^2, made no sense and thus was not calculated. The predicted values $\hat{Z}$ obtained directly by using the clay spectral index $SWIR_{FI}$ ranged from 10 to 305 g/kg with a mean of 52 g/kg (Table 3), which is small compared to that of the 120 reference values (Table 1); the distribution of these clay estimations $\hat{Z}$ was not normal (high skewness, Table 3). In conclusion and as expected, the use of the clay spectral index without calibration provided poor accuracy of predictions.

The normalization of the clay estimations $\hat{Z}$ using the Box Cox process, with an α value of 0.7 in Eq. (2), slightly increased the *SEP*, *bias* and SEP_c (Table 2). Thus, the prediction accuracy remained inaccurate, even after normalization of the predicted value $\hat{Z}$.

New clay contents were predicted from the linear model built from the $BC(\hat{Z})$ clay content predictions and the observed clay contents of the *REF* database (Fig. 2 and Eq. (3)). This linear model provided modest performances, with a *SEP* of 111 g/kg, an R^2 of 0.58, a *bias* of 0 g/kg and a SEP_c of 111 g/kg (Table 2). Moreover, the standardization of the $BC(\hat{Z})$ clay content predictions using the clay contents of the *REF* database provided modest performances, with a *SEP* of 116 g/kg, an R^2 of 0.55, a *bias* of 8 g/kg and a SEP_c of 115 g/kg (Table 2). Thus, the linear model provided performances that were only slightly superior to the standardization process.

4.2. Spectral index predictions calibrated with legacy database

The performances of the standardization of $BC(\hat{Z})$ clay content predictions using the clay contents of the legacy database were poorer than those of the linear model and the standardization using the *REF* database (Table 2). Nevertheless, the $ST(\hat{Z})$ predictions obtained with the legacy database increased the performance of predictions with a decrease in the *SEP*, *bias* and SEP_c compared to those obtained with the raw spectral index and with the normalization step (Table 2). The *SEP*, *bias* and SEP_c were <311 g/kg, −288 g/kg and 117 g/kg, respectively (Table 2). The standardization of the $BC(\hat{Z})$ clay content predictions using the *Focused-Legacy2010* database provided modest performances, with a *SEP* of 131 g/kg, an R^2 of 0.43, a *bias* of −25 g/kg and a SEP_c of 128 g/kg (Table 2). Meanwhile, the new $ST(\hat{Z})$ predictions obtained with the *Focused-Legacy2000* and *no-Focused-Legacy1973* databases could not be considered accurate, as the *SEP* and *bias* still remained high (the *SEP*, *bias* and SEP_c were >221 g/kg, −188 g/kg and 116 g/kg, respectively).

Table 3
Distribution of $\hat{Z}$ clay content predictions (in g/kg).

	Mean	Std.	Min	Max	Skewness
Raw clay content prediction by using the $SWIR_{FI}$ index	52.05	40.78	10.87	305.28	2.41

4.3. Estimated clay maps

The spectral index $SWIR_{FI}$ was applied to all AISA-DUAL spectra of bare soils over the study area, and the clay prediction map was called *Raw_Map* (Fig. 3a). As previously seen in Section 4.1, clay contents $\hat{Z}$ estimated by Eq. (1) over this study area have a small range (mean of 50 g/kg over the study area), so the *Raw_Map* is scaled from 0 to 200 g/kg to highlight the spatial structure of estimated soil property $\hat{Z}$. Linear regression (Eq. (3)) was applied to all $BC(\hat{Z})$ predicted clay contents over bare soils, and the clay prediction map was called *LM_REF_Map* (Fig. 3b). Finally, standardization (Eqs. (4) and (5)) was applied to all $BC(\hat{Z})$ predicted clay contents over bare soils using the *REF*, *Focused-Legacy2010*, *Focused-Legacy2000* and *no-Focused-Legacy1973* databases and were called, respectively, *ST_REF_map* (Fig. 3c), *ST_Foc2010_map* (Fig. 3d), *ST_Foc2000_map* (Fig. 3e) and *ST_no-Foc1973_map* (Fig. 3f).

We focused our attention on a test area of 6.67 km^2 that was centered on the Kamech catchment, which contained a high percentage of bare soils during image acquisition (49.2% of the test area). The Kamech experimental catchment belongs to a long-term environmental research observatory called OMERE (Mediterranean observatory of water and rural environment), which aims to study the anthropogenic impacts on water and sediment budgets at the catchment scale (e.g., Mekki et al., 2006; Raclot and Albergel, 2006). The *LM_REF_Map*, *ST_REF_map* and *ST_Foc2010_map* (Fig. 3b, c and d) provided similar clay content patterns that were also similar to those obtained by Gomez et al. (2012) using the PLSR approach. These maps showed close successions of clayey and non-clayey areas. In spite of the decreasing contrast, these successions were still mapped in *Raw_Map*, *ST_Foc2010_map* and *ST_no-Foc1973_map* (Fig. 3a, e and f).

5. Discussion

5.1. Spectral index performances

The clay spectral index proposed by Levin et al. (2007), when used without calibration data to estimate relative soil property values, provides low estimation performances with high *SEP* (439 g/kg), *bias* (−414 g/kg) and SEP_c (146 g/kg) values (Table 2). The linear regression method, which specifies a linear relationship between the clay contents of the *REF* database (set of dependent Y-variables) and the normalized spectral index values (set of predictor X-variables), provides modest estimation performances, with an R^2 of 0.58 associated with a *SEP* of

Table 2
Performance of predictions (SEP, bias and SEP_c are expressed in g/kg). *n.c.* means *not calculable.*

			SEP (R^2)	Bias	SEP_c
Raw Spectral Index $\hat{Z}$			439 (n.c.)	−414	146
Normalization	$BC(\hat{Z})$		475 (n.c.)	−445	166
	$LM(\hat{Z})$		111 (0.58)	0	111
	Standardized values $ST(\hat{Z})$ using:	*REF*	116 (0.55)	8	115
		Focused-Legacy2010	131 (0.43)	−25	128
		Focused-Legacy2000	221 (n.c.)	−188	116
		No-Focused-Legacy1973	311 (n.c.)	−288	117

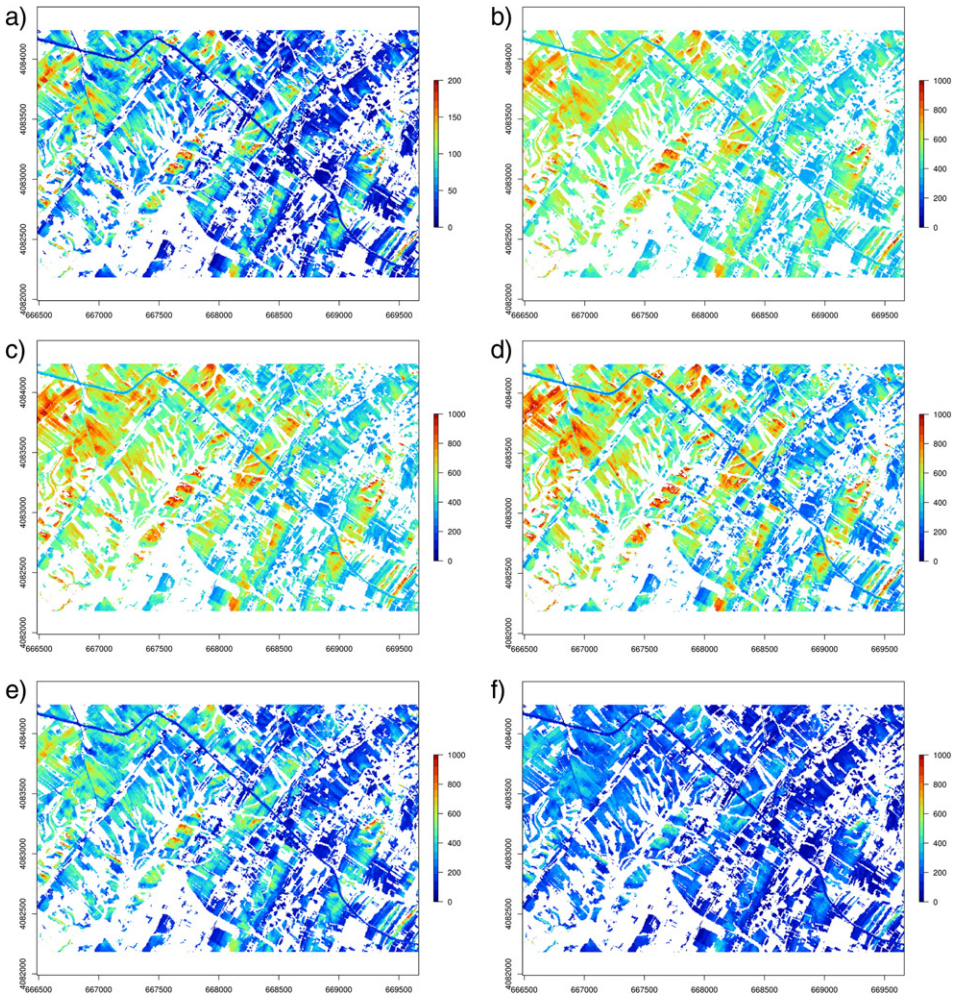


Fig. 3. a) *Raw_Map*, b) *LM_REF_map*, c) *ST_REF_map*, d) *ST_Foc2010_map*, *e*) *ST_Foc2000_map* and f) *ST_no-Foc1973_map* of clay contents (g/kg). *Raw_Map* is scaled from 0 to 200 g/kg, and the five other maps are scaled from 0 to 1000 g/kg.

118 g/kg and no *bias* (Table 2). This performance is comparable to that obtained by Lagacherie et al. (2008) using a clay spectral index and Hymap VNIR/SWIR airborne data ($R^2 = 0.6$ and $SEP = 130$ g/kg) and lower than those obtained by Gomez et al. (2012) ($R^2_{cv} = 0.8$ and $SEP = 82$ g/kg) using the PLSR technique to estimate clay content from the same datasets as in this study (VNIR/SWIR AISA-DUAL images and *REF* soil data). Moreover, the clay content map obtained from the linear regression provides a similar clay content pattern to that observed by Gomez et al. (2012).

5.2. Standardization process

The transformation applied to the raw clay content predictions ($SWIR_{FI}$) is composed of a boxcox transformation for normalization of the predicted data and a scaling and centering of the normalized data for standardization. This standardization process of raw clay content predictions ($SWIR_{FI}$) using the *REF* database provides similar results to the linear model calibrated from the *REF* database (Table 2). This standardization process is comparable to one of the most widely used transfer methods for correcting predicted values, i.e., the univariate slope and bias correction (SBC) (Bouveresse et al., 1996). Transfer methods have been developed to enable a calibration model to be effectively transferred between two "systems" (e.g., two spectroscopic instruments), thus eliminating the need for a full recalibration (Fearn, 2001; Feudale et al., 2002). In our case, we do not have a calibration model that has to be adjusted to different "systems"; rather, we have a model that has to be calibrated to an approximate system (legacy soil dataset).

5.3. Sensitivity to the choice of legacy soil database

Three legacy soil datasets, which differed with respect to the dates of collections, the standards of soil sampling and descriptions, the georeferencing methods, the laboratory analysis protocols, the data storage methods and their locations, were used to standardize the clay content predicted by the spectral index. The prediction performance after standardization using these three legacy soil datasets differs with respect to bias values. Thus, the prediction performance depends on the initial bias of the legacy database with the *REF* database, which is propagated to predictions. As explained by Ciampalini et al. (2013), the bias of the old legacy databases (*Focused-Legacy2000* and *no-Focused-Legacy1973* databases) results from two main causes:

1) differences in laboratory soil analysis protocols. This cause converges toward a general problem of digital soil mapping and should receive more attention in future studies (Ciampalini et al., 2013, Baume et al., 2011).
2) sampling locations. The sites should preferably be located inside the study area, and there is a benefit in purposely sampling to best represent the location.

Conversely, it is noticeable that the number of samples does not play a significant role because the difference between performances using the *REF* and the *Focused-Legacy2010* databases is moderate.

Apart from these two causes, one may cite the impact of the soil changes with time, which may introduce biases if aged legacy soil data

are used. We could not observe this impact in our case study since clay content is known for not changing significantly within some decades. However, the age of the legacy data should be taken into account for soil properties that are impacted by soil tillage and fertilization techniques such as Organic Carbon and pH.

5.4. Combining modern and legacy soil sampling

In theory, the new calibration approach proposed in this paper should allow to replace the expensive ad hoc field sampling that is usually involved in the calibration step of VNIR/SWIR imaging spectrometry by a legacy soil dataset which is much less expensive to collect. The use of legacy data is made possible because our approach is much less sensitive to the field sampling. Indeed this approach does not require that the sites are located on bare soil, are all located on the study area and are precisely georeferenced.

However, our results showed that the legacy data had biases that are propagated to the estimations of soil property with VNIR/SWIR imaging spectroscopy. A prior checking and correction of the possible bias are therefore necessary before using them in our calibration approach. This implies not to fully eliminate any up-to-date soil sampling since it can be indispensable for constituting the reference dataset that serve for bias correction (Ciampalini et al., 2013). This is why, rather than allowing the replacement of ad-hoc field sampling by legacy soil data, our approach paves the way to the combination use of legacy and ad-hoc soil sampling for the calibration of VNIR/SWIR imaging spectroscopy models with acceptable costs.

6. Conclusion

One of the limiting factors of actual hyperspectral airborne data use for soil property mapping is the need for measurements over the study site to calibrate predictive models of the soil properties. In this work, we proposed using a spectral index to predict clay contents, which were standardized with soil legacy data for soil properties mapping. Using an “ad hoc” soil dataset (collected over the study area at the same time as the hyperspectral airborne acquisition), the standardization process provided modest performances comparable to the use of the linear model. The main performance limitations were due to the bias of clay measurements, which affected the legacy soil databases. Thus, the legacy soil databases need to be standardized before they can be used for such an approach. This work could be extended to other clay spectral indexes, such as those proposed by Chabrillat et al. (2002) and Lagacherie et al. (2008), and to other topsoil properties, such as iron or $CaCO_3$ by using iron and carbonate indexes (e.g., Madeira et al., 1997; Lagacherie et al., 2008). This may confirm the independence of our actual observations from the selected spectral index and the topsoil property. Finally, this work creates hope for large soil properties mapping, as expected by the International Global Soil Map project (http://www.globalsoilmap.net/), thanks to the use of the legacy soil data associated to future hyperspectral satellites sensors (such as the HYPerspectral X Imagery – HYPXIM, the Spaceborne Hyperspectral Applicative Land and Ocean Mission – SHALOM, the PRecursore IperSpettrale della Missione Applicativa –PRISMA, and the Environmental Mapping and Analysis Program – EnMAP).

Acknowledgments

This research was supported by EC Operational Program (Project No.: ESF/ MEYS CZ.1.07/2.3.00/30.0040), the French National Research Agency (ANR) through the ALMIRA project (ANR-12-TMED-0003) and MISTRAL 2011 project Sicmed-Lebna “Biophysical and socio-economical analysis of water management within the Tunisian Cap Bon Peninsula: the Lebna study area”. The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National de Cartographie et de Télédétection, Tunisia), for providing the AISA-DUAL hyperspectral images for this study. This hyperspectral data acquisition was granted by IRD (Institut de Recherche pour le Developpement), INRA (Institut National de la Recherche Agronomique) and the French National Research Agency (ANR) (ANR-08-BLAN-C284-01). Soil profiles selections and descriptions of the *Focused-Legacy2010* database were made by G. Coulouma with the assistance of Y. Blanca, P. Lagacherie and R. Ciampalini. Soil profiles selections and descriptions of the “*Focused-Legacy2000*” database were made by the IAO (Instituto Agronomico per l'Oltremare). And soil profiles selections and descriptions of the “*No-Focused-Legacy1973*” were made by the DG/ACTA Sol of Tunisia.

References

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. In: Sparks, Donald L. (Ed.), Chapter Three - GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil PropertiesAdvances in Agronomy Vol. 125. Academic Press, 2014, pp. 93–134.

Baize, D., Jabiol, B., 1995. Guide Pour la Description Des Sols. INRA édition, Paris.

Baume, O., SkÃjen, J.O., Heuvelink, G.B.M., Pebesma, E.J., 2011. A geostatistical approach to data harmonization - application to radioactivity exposure data. Int. J. Appl. Earth Obs. Geoinf. 13 (3), 409–419.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Prediction of soil attributes by NIR spectroscopy. A critical review of chemometric indicators commonly used for assessing the quality of the prediction. TrAC Trends Anal. Chem. 29 (9), 1073–1081.

Ben-Dor, E., Chabrillat, S., Dematte, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. 113, S38–S55.

Ben-Dor, E., Kafri, A., Varacalli, G., 2014. SHALOM: An Italian-Israeli Hyperspectral Orbital Mission – Update. International Geoscience and Remote Sensing Symposium, Quebec, Canada, July 13–18, 2014.

Bouveresse, C., Hartmann, D.L., Massart, I.R. Last, Prebble, K.A., 1996. Standardisation of near-infrared spectrometric instruments. Anal. Chem. 68, 982–990.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26 (2), 211–252 (5. Goldberger, A.S. (1968)).

Briottet, X., Marion, R., Carrere, V., Jacquemoud, S., Bourguignon, A., Chami, M., Chanussot, J., Chevrel, S., Deliot, P., Dumont, M., Foucher, P.Y., Minghelli-Roman, A., Sheeren, D., Weber, C., Prastault, P., Hosford, S., Lefevre, M.J., 2013. HYPXIM: HYPXIM: a second generation high spatial resolution hyperspectral satellite for dual applications. 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), June 2013, Gainesville, Florida, USA.

Carrere, V., Bourguignon, A., Briottet, X., Chami, M., Chevrel, S., Jacquemoud, S., Marion, R., 2013. The French Hyperspectral Earth Observation Science/Defense Mission HYPXIM - a Second Generation High Spectral and Spatial Resolution Imaging Spectrometer, IGARSS, 21–26 July 2013, Melbourne, Australia.

Chabrillat, S., Goetz, A.F.H., Krosley, L., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sens. Environ. 82, 431–445.

Chabrillat, S., Eisele, A., Guillaso, S., Rogaß, C., Ben-Dor, E., Kaufmann, H., 2011. HYSOMA: An easy-to-use software interface for soil mapping applications of hyperspectral imagery. Proceedings of the 7th EARSeL SIG Imaging Spectroscopy Workshop, Edinburgh, Scotland, UK, 11–13 April 2011.

Chambers, J.M., 1992. Linear models. In: Chambers, J.M., Hastie, T.J. (Eds.), Chapter 4 of Statistical Models in S. Wadsworth & Brooks/Cole.

Chang, C.-W., Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Sci. 167 (2), 110–116.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh Jr., C.R., 2001. Near-infrared reflectance spectroscopy—principal components regression analysis of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Ciampalini, R., Lagacherie, P., Gomez, C., Grunberger, O., Hamrouni, H., Mekki, I., Richard, A., 2013. Detecting and correcting biases of measured soil profiles data. A case study in the Cap Bon Region (Tunisia). Geoderma 192, 68–76 (January 2013).

Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. J. Agric. Sci. 140, 65–71.

Davies, A., Fearn, T., 2006. Back to basics: calibration statistic. Spectrosc. Eur. 18, 31.

FAO, 1998. World reference base for soil resources. World Soil Resources Reports Vol. 84. FAO, Rome, p. 88.

Fearn, T., 2001. Standardisation and calibration transfer for near infrared instruments: a review. J. Near Infrared Spectrosc. 9 (4), 229–244.

Feudale, R.N., Woody, N.A., Tan, H., Miles, A.J., Brown, S.D., Ferre, J., 2002. Transfer of multivariate calibration models: a review. Chemom. Intell. Lab. Syst. 64, 181–192.

Gaffey, S.J., 1986. Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 μm): calcite, aragonite and dolomite. Am. Mineral. 71, 151–162.

Gomez, C., Lagacherie, P., Coulouma, G., 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma 148 (2), 141–148.

Gomez, C., Lagacherie, P., Bacha, S., 2012. Using an Vis-NIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon Region (Tunisia). In: Minasny, Malone, McBratney (Eds.), Digital Soil Assessments and beyond. CRC Press 2012, pp. 387–392.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, 1–26.

IAO, 2002. Land Resources of the Oued Lebna Catchment (Tunisia). 20th Course Professional Master: Remote Sensing and Natural Resources Evaluation. Istituto Agronomico per l'Oltremare, Florence, Italie, p. 132.

Lagacherie, P., Baret, F., Feret, J.-B., Madeira Netto, J., Robbez-Masson, J.-M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sens. Environ. 112 (3), 825–835.

Le Bissonnais, Y., 1996. Soil aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. Eur. J. Soil Sci. 47, 425–431.

Legendre, L., Legendre, P., 1998. Numerical Ecology. 2nd English ed. Elsevier. Masson, Amsterdam, pp. 39–45.

Levin, N., Kidron, G.J., Ben-Dor, E., 2007. Surface properties of stabilizing coastal dunes: combining spectral and field analyses. Sedimentology 54, 771–788.

Lopinto, E., Ananasso, C., 2013. The Prisma Hyperspectral Mission. 33rd EARSeL Symposium, Towards Horizon 2020: Earth Observation and Social Perspectives, Matera (Italy), 3–6 June 2013.

Madeira, J., Bedidi, A., Cervelle, B., Pouget, M., Flay, N., 1997. Visible spectrometric indices of hematite (Hm) and goethite (Gt) content in lateritic soils: the application of a thematic mapper (TM) image for soil-mapping in Brasilia, Brazil. Int. J. Remote Sens. 18 (13), 2835–2852.

Maignien, R., 1980. Manuel Pour la Description Des Sols Sur le Terrain. ORSTOM, Paris, p. 112.

Mekki, I., Albergel, J., Ben Mechlia, N., Voltz, M., 2006. Assessment of overland flow variation and blue water production in a farmed semi-arid water harvesting catchment. Phys. Chem. Earth 31, 1048–1061.

Miesch, C., Poutier, L., Achard, V., Briottet, X., Lenot, X., Boucher, Y., 2005. Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 437 (7), 1552–1562.

Raclot, D., Albergel, J., 2006. Runoff and water erosion modelling using WEPP on a Mediterranean cultivated catchment. Phys. Chem. Earth 31, 1038–1047.

Richter, N., Jarmer, T., Chabrillat, S., Oyonarte, C., Hostert, P., Kaufmann, H., 2009. Free iron oxide determination in Mediterranean soils using diffuse reflectance spectroscopy. Soil Sci. Soc. Am. J. 73, 72–81.

Selige, T., Bohner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136 (1–2), 235–244.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66, 988–998.

Soil Survey Division Staff, 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158 (1–2), 32–45.

Stuffler, T., Kaufmann, H., Hofer, S., Förster, K.-P., Schreier, G., Müller, A., Eckardt, A., Bach, H., Penne, B., Benz, U., Haydn, R., 2007. The EnMAP hyperspectral imager-an advanced optical payload for future applications in earth observation programmes. Acta Astronaut. 61 (1–6), 115–120.

Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Reg. 4, 20–30.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Viscarra Rossel, R.A., Bui, E.N., 2016. A new detailed map of total phosphorus stocks in Australian soil. Sci. Total Environ. 542, 1040–1049. Part B.

Wilkinson, G.N., Rogers, C.E., 1973. Symbolic descriptions of factorial models for analysis of variance. Appl. Stat. 22, 392–399.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Zante, P., Collinet, J., Pepin, Y., 2005. Caractéristiques pédologiques et hydrométéorologiques du bassin versant de Kamech, Cap Bon, Tunisie. UMR LISAH IRD Tunis, DG ACTA Direction Des Sols Tunis, INRGREF Tunis (21 p. + 6 annexes).

Digital Soil Assessments and Beyond – Minasny, Malone & McBratney (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62155-7

Co-kriging of soil properties with Vis-NIR hyperspectral covariates in the Cap Bon region (Tunisia)

R. Ciampalini & P. Lagacherie
INRA, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

P. Monestiez & E. Walker
INRA, Unité de Biostatistique et Processus Spatiaux (BioSP), Avignon, France

C. Gomez
IRD, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

ABSTRACT: Visible and Near Infrared (Vis-NIR) hyperspectral airborne spectroscopy can be used for predicting soil surface properties but its use is constrained to bare soil surfaces. To extend its use to larger areas, a co-kriging procedure was applied across a 339 km² area located in the Cap Bon Region (northern Tunisia). The study used 262 soil surface analysed samples and, as covariates, the bare soil estimates of four topsoil properties (Clay, Sand, Iron contents and Cation Exchange Capacity) derived from a 30 meter resolution Vis-NIR AISA-Dual hyperspectral image. The resulting co-regionalisation models and co-kriged maps allowed to capture the soil pattern of the study area which indicates a promising opportunity for using Vis-NIR hyperspectral covariates for Digital Soil Mapping. However, the co-kriging performances were limited by the short-scale soil variations of this lithology-driven Mediterranean study area.

1 INTRODUCTION

In many parts of the world, the available legacy soil data are too sparse for the Digital Soil Mapping (DSM) of soil properties with acceptable precision (e.g. Ciampalini et al., 2012). It is therefore necessary to collect new input soil data while maintaining acceptable costs.

It has been showed recently (Gomez et al., submitted, Gomez et al., 2012) that Vis-NIR hyperspectral airborne spectroscopy could provide acceptable estimates of some key topsoil properties. However, prediction of the soil properties is limited to bare soil surfaces. In recognition of this, Lagacherie et al. (2012) used co-kriging and block-cokriging for extending the hyperspectral estimates to an entire study area.

From the encouraging results of Lagacherie et al. (2012), Vis-NIR hyperspectral imagery is expected to be largely used for DSM in the future, especially in Mediterranean and semi-arid areas where bare soil surfaces are common and where dry periods can allow to avoid soil moisture perturbations of the spectrum (Lagacherie et al., 2008). Further Vis-NIR hyperspectral sensors are planned to be launched on board satellites within the next two years, PRISMA (Giampaolo et al., 2008) and EnMap (Stuffler et al., 2007).

In this paper, a cokriging procedure is applied to map four topsoil properties (Clay, Sand, iron contents and CEC) in the Cap Bon region (Northern Tunisia) using as covariates the accurate but incomplete estimates derived from an Vis-NIR hyperspectral image (Gomez et al., 2012).

2 THE CASE STUDY

2.1 *Study area*

The study area is located in the Cap Bon region in northern Tunisia (36°24′N to 36°53′N; 10°20′E to 10°58′E), 60 km east of Tunis, Tunisia (Figure 1a). This 300 km² area includes the Lebna catchment (Figure 1b), which is mainly rural (>90%) and devoted to cereals in addition to legumes,

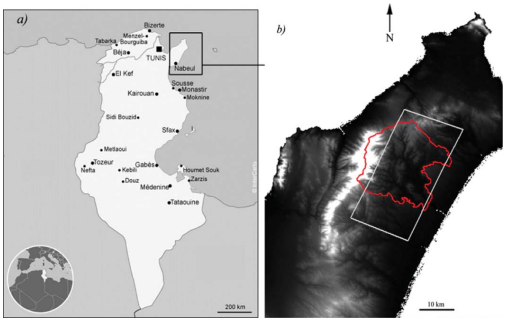


Figure 1. a) Location of the Cap Bon region in the Northern Tunisia, b) limits of the Lebna Catchment (in red) and the AISA hyperspectral image (in grey) plotted over the STRM DEM of the Cap Bon.

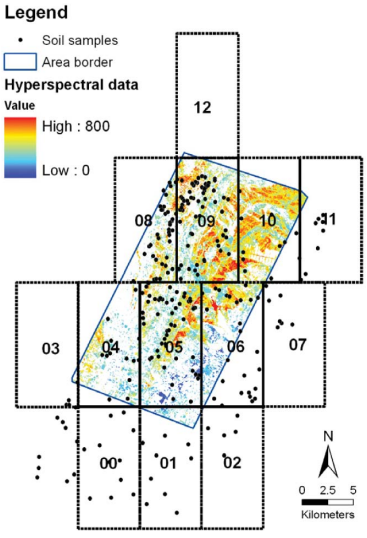


Figure 2. Locations of the samples and delimitation of computing zones for the co-kriging.

olive trees, natural vegetation for breeding and vineyards. It is characterized by relief areas, with an altitude between 0 and 226 m. The main soil types are Regosols, Eutric Regosols (9.6%) preferentially associated with sandstone outcrops, Calcic Cambisols, and Vertisols preferentially formed on marl outcrops and lowlands. The southeastern region of the study area has a flatter landscape with sandy Pliocene deposits yielding Calcosols and Rendzina.

2.2 *Field data*

262 soil samples were collected between October 2008 and November 2010 (Figure 2). 129 of these soil samples were collected within the hyperspectral image perimeter in fields that were bare during the hyperspectral data acquisition in November 2010. The remaining 133 samples were located either outside this image perimeter (73 samples) or in vegetated areas (60 samples). The former 129 were used in Gomez et al. (2012) for calibrating the hyperspectral estimation functions. The samples were composed of five sub-samples collected to a depth of 5 cm at random locations within a 10 × 10 m square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. Soil properties were measured using classical physico-chemical soil analysis. In this study only the four soil properties successfully predicted from hyperspectral data ($R^2 > 0.6$) were considered: clay, sand, free iron and cation exchange capacity (CEC).

2.3 *Hyperspectral covariates*

The covariates used in this study, called hereon "hyperspectral covariates", were the maps of topsoil properties derived from Vis-NIR AISA-Dual hyperspectral data which were acquired on 2nd November 2010 over the study area with a spatial resolution of 5 m and 359 non-contiguous bands covering the 400 to 2500 nm spectral domain. For computational efficiency, a pixel aggregation resampling method was used to provide Vis-NIR hyperspectral image at 30 m of spatial resolution. It must be noted that 30 meters is the resolution of the Vis-NIR hyperspectral sensors that are planned to be launched on board satellites (Giampaolo et al., 2008, Stuffler et al., 2007). After artifact filtering, it comprised 280 spectral bands covering the 450 to 2450 nm spectral domain. Vegetated surfaces, urban areas and water areas were masked to consider only bare soil surfaces for further soil predictions. The methodology for obtaining maps of predicted topsoil properties is fully detailed in (Gomez et al., 2012). Only a short summary is provided here.

The Partial Least Square Regression (PLSR) was used to establish relationships between the topsoil properties and the AISA-Dual spectra. PLSR-based prediction models were built using the 129 AISA-Dual spectra corresponding to the location of the soil samples collected over bare soils. Correct prediction models, with R^2 and RPD values greater than 0.6 and 1.4 respectively, were obtained for four soil properties: free iron, CEC, clay and sand content.

Figure 3 shows an example of the resulting 30 meter resolution images of soil properties that covered 41.6% of the studied area.

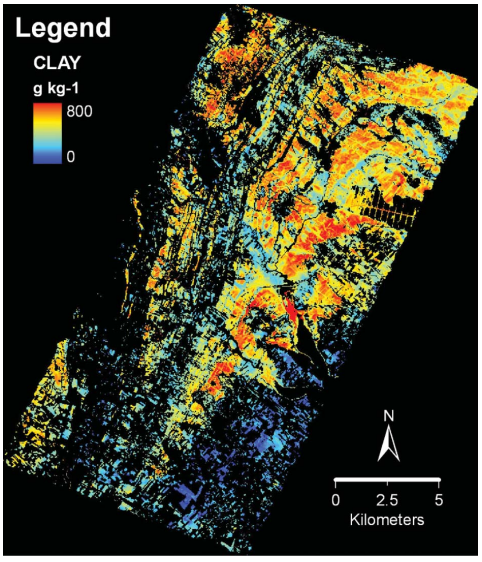


Figure 3. Clay content predicted over bare soils from AISA—Dual spectra (black areas correspond to non-predicted surfaces).

3 METHOD

The set of measured soil samples and the images of estimated soil properties were co-kriged to obtain new images that covered the whole study area. We applied a procedure that was recently developed for mapping topsoil properties from hyperspectral data in southern France (Lagacherie et al., 2012). In this study however; block-cokriging was replaced by punctual co-kriging.

All the analyses in this study were performed using R (R Development Core Team, 2007).

3.1 *Modelling multivariate spatial correlations*

In this study, the measured soil properties and their hyperspectral covariates are denoted as Z_1 and Z_2, respectively. Suppose that u is a location in two-dimensional space and $Z_1(u)$ and $Z_2(u)$ are spatial random functions. Assuming that the soil property (Z_1) is spatially cross-correlated with the hyperspectral covariate (Z_2), the spatial cross correlation between Z_1 and Z_2 can be quantified by a cross-variogram (Wackernagel, 1995). In univariate or bivariate frameworks, the covariance and variogram functions can be estimated as follows:

$$\hat{\gamma}z_i z_j(h) = \frac{1}{2\mathrm{N}(h)} \sum_{\alpha=1}^{N(h)} (z_i(u_\alpha + h) - z_i(u_\alpha)) \times (z_j(u_\alpha + h) - z_j(u_\alpha)) \quad (1)$$

In Equation (1) i and j belong to $\{1, 2\}$. When $i = j$, Equations (1) denotes the variogram estimates. When $i \neq j$, Equations (1) denotes the cross variogram estimates. $\boldsymbol{h}$ is the separation vector between the data locations u_α and $u_\alpha + \boldsymbol{h}$ (the translated of $\boldsymbol{h}$ from u_α) $Z_i(u_\alpha)$ and $Z_j(u_\alpha + \boldsymbol{h})$ are observations of the variable z_i and z_j at spatial locations u_α and $u_\alpha + \boldsymbol{h}$, respectively, and $\mathrm{N}(\boldsymbol{h})$ is the number of distinct pairs of observations at distance $\boldsymbol{h}$.

To undertake the co-kriging, a variogram matrix in which the diagonal entries are variograms and the off-diagonal entries are cross variograms must be strictly conditionally negative definite. To ensure this condition, one can use intrinsic or linear co-regionalisation models. The formulation of the latter in the bivariate case with two nested spatial structures is (Wackernagel, 1995).

$$\Gamma(\boldsymbol{h}) = B_1 g_1(\boldsymbol{h}) + B_2 g_2(\boldsymbol{h}) \quad (2)$$

where $g_1(\boldsymbol{h})$ and $g_2(\boldsymbol{h})$ are two normalized variograms, one for each spatial structure, and $\mathbf{B}_1$ and $\mathbf{B}_2$ are positive semi-definite 2×2 matrices.

3.2 *Co-kriging*

The co-kriging estimator is a best linear unbiased estimator (BLUE) and has minimum estimation error variance (Wackernagel, 1995). In the two variables case, the ordinary co-kriging estimator is a linear combination of weights $w^1{}_\alpha$ and $w^2{}_\alpha$ with data from the two variables Z_1 *and* Z_2 located at sample points in the neighborhood of a spatial location u_0. Each variable is defined on a set of samples of possibly different sizes n_1 and n_2, and the estimator is defined as:

$$\hat{Z}_1(u_0) = \sum_{\alpha=1}^{n1} w_\alpha^1 Z_1(u_\alpha) + \sum_{\alpha=1}^{n2} w_\alpha^2 Z_2(u_\alpha) \quad (3)$$

where the weights w_α^1 and w_α^2 are solutions of a co-kriging system and sum to 1 and 0, respectively.

The co-kriging variance of the estimation error of $\hat{Z}_1$ in the two variables case can be estimated from the variogram γ_{z1z1} and the crossvariogram γ_{z1z2} (Wackernagel, 1995) by using the following expression:

$$\sigma_E^2(u_0) = \sum_{\alpha=1}^{n1} w_\alpha^1 \gamma_{z1z1}(u_\alpha - u_0) + \sum_{\alpha=1}^{n2} w_\alpha^2 \gamma_{z1z2}(u_\alpha - u_0) + \mu_{z1} \quad (4)$$

where μ_{z1} is the lagrange multiplier of the co-kriging system and $u_\alpha - u_0$ denotes the distance between u_α and u_0 locations.

3.3 *Co-kriging with large datasets*

The large number of sites with hyperspectral covariates made it difficult to co-krige from the whole set of these sites at every prediction location. Subsets were therefore considered at the neighborhood of the prediction locations by randomly selecting at most 500 sites within a 300 m × 300 m area centered on the prediction location and at most 1000 additional sites within a 1000 m × 1000 m area centered on the prediction location.

3.4 *Validation*

The validation of the co-kriging outputs was done on the 189 soil samples located in the image perimeter by comparing the true measurements with the predicted values obtained from a leave-one-out cross-validation. To evaluate the prediction quality, the classical indicators R^2 and RMSE were calculated. To match GlobalSoilMap.net specifications, we also calculated the observed fractions A of the true values falling into the estimated 95%-probability interval (PI_{95}) bounded by 0.025 and 0.975 quantiles. It was expected that the computed fractions A were close to 95% and that the PI_{95} were as narrow as possible.

4 RESULTS

4.1 *Co-regionalisation models*

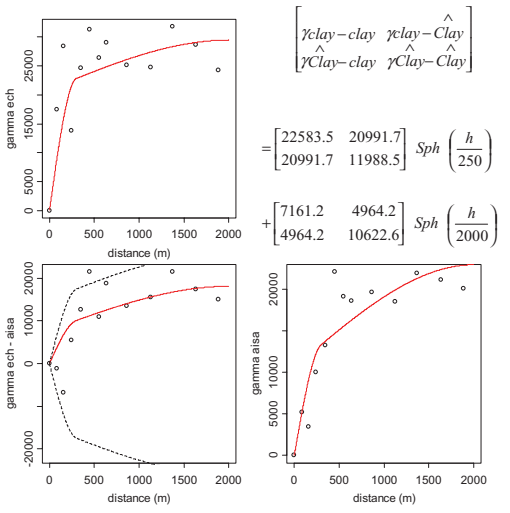
Linear co-regionalisation models were built for the pairs "soil properties-hyperspectral covariates" from the set of 129 sites at which these two variables were available. The two direct variograms were first modeled as linear combinations of two selected basic structures. The same basic structures were then fitted to the cross-semi-variograms under the positive semi-definite constraints (Goovaerts, 1997).

Figures 4a-c show the fitted co-regionalisation models for clay content, CEC, and iron content as well as their analytical expressions. In general the models fitted adequately to the data. None of the models had a nugget effect. The ranges were very similar across the soil properties with a short range of 250 and large range between 1500 and 2000 m. Short range semi-variances were clearly larger than the large range ones.

4.2 *Validation results*

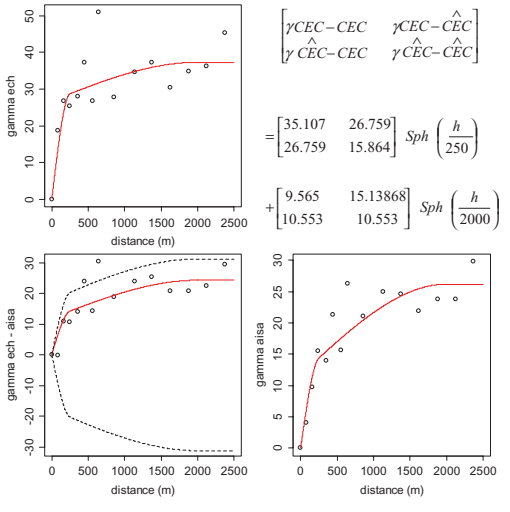
Table 1 shows the summary of the performance parameters of the cross validation procedure. RMSE and R^2 values showed moderate accuracies with clear decreases at the sites located in the gaps of the hyperspectral image (digits between brackets in table 1). At such sites, the distances from any available data were often greater than the shortest ranges at which most of the variances were observed, especially for iron.

This result was fairly well anticipated by the model by estimating 95%-probability interval widths. These widths were however slightly underestimated as shown by the fractions of the true values falling in the PI-95 that were less than the expected 0.95 values.



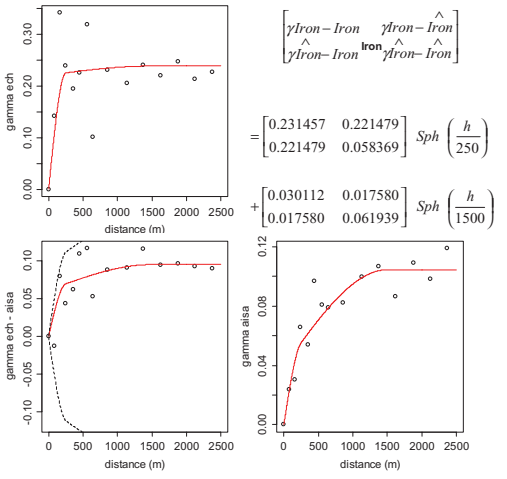
$$\begin{bmatrix} \gamma clay-clay & \gamma clay-\hat{Clay} \\ \gamma\hat{Clay}-clay & \gamma\hat{Clay}-\hat{Clay} \end{bmatrix} = \begin{bmatrix} 22583.5 & 20991.7 \\ 20991.7 & 11988.5 \end{bmatrix} Sph\left(\frac{h}{250}\right) + \begin{bmatrix} 7161.2 & 4964.2 \\ 4964.2 & 10622.6 \end{bmatrix} Sph\left(\frac{h}{2000}\right)$$

Figure 4a. The linear model of co-regionalization for clay content (clay) and its hyperspectral estimate $(\hat{Clay})$.



$$\begin{bmatrix} \gamma CEC-CEC & \gamma CEC-\hat{CEC} \\ \gamma\hat{CEC}-CEC & \gamma\hat{CEC}-\hat{CEC} \end{bmatrix} = \begin{bmatrix} 35.107 & 26.759 \\ 26.759 & 15.864 \end{bmatrix} Sph\left(\frac{h}{250}\right) + \begin{bmatrix} 9.565 & 15.13868 \\ 10.553 & 10.553 \end{bmatrix} Sph\left(\frac{h}{2000}\right)$$

Figure 4b. The linear model of co-regionalization for Cation Exchange Capacity (CEC) and its hyperspectral estimate and its hyperspectral estimate $(\hat{CEC})$.



$$\begin{bmatrix} \gamma Iron-Iron & \gamma Iron-\hat{Iron} \\ \gamma\hat{Iron}-Iron & \gamma\hat{Iron}-\hat{Iron} \end{bmatrix}_{Iron} = \begin{bmatrix} 0.231457 & 0.221479 \\ 0.221479 & 0.058369 \end{bmatrix} Sph\left(\frac{h}{250}\right) + \begin{bmatrix} 0.030112 & 0.017580 \\ 0.017580 & 0.061939 \end{bmatrix} Sph\left(\frac{h}{1500}\right)$$

Figure 4c. The linear model of co-regionalization for iron content (Iron) and its hyperspectral estimate $(\hat{Iron})$.

Table 1. Validations results.

Soil property	RMSEcv*	R^2 cv	PI95*** width*	A****
Sand	163 (207)	0.51 (0.32)	464	0.86
CEC	4.9 (6.2)	0.54 (0.32)	14.2	0.88
Iron	0.39 (0.47)	0.41 (0.31)	1.3	0.87

* in g/kg for clay and sand, in meq/100 g for CEC and g/kg for iron.
** between brakets: results for validation points located in the gaps of the hyperspectral image.
*** PI95 is the estimated 95%-probability interval (PI_{95}) bounded by 0.025 and 0.975 quantiles.
**** A is the fraction of the true values of soil property falling in 95% PI.

4.3 *Mapping results*

Figures 5 show the estimated maps of the four studied soil properties and their associated uncertainty maps that were obtained from co-kriging with the hyperspectral covariates.

The soil property maps exhibit short scaled and striped patterns that are similar across the set of studied properties. Areas with high clay and iron contents, low sand contents and high CEC contrasted with areas with moderate clay and iron contents, high sand contents and moderate CEC. These patterns strongly resemble those of the hyperspectral covariate (Figure 3) with the noticeable exception of the south-east area where the low extreme values of clay contents disappeared in the co-kriged maps (Figure 5a).

The uncertainty maps show variations that were influenced by the locations of the measured sites—around which the predicted uncertainties were minimal—and by the presence of hyperspectral covariates—as shown by the north-south increase of uncertainty with the decrease of hyperspectral covariate availability. The maximal uncertainty was predicted for a small area located at the east of the study area due to the distance from measured sites and hyperspectral covariate data (Figure 2).

5 DISCUSSION & CONCLUSION

5.1 *Modelling a Mediterranean soil pattern*

The motivation of this study was to deal with soil variations within a sample area of the North Africa region which is poor in available legacy soil data. However, throught the use of Vis-Nir hyperspectral imagery that covers the study area (Gomez et al., 2012), the complex soil pattern of this region could be quantitatively modeled in great detail (30 meter resolution). The co-regionalization models (Figures 4) and the co-kriging maps (Figures 5) of four correlated soil properties (clay, sand, iron and CEC) respectively revealed a multi-scaled and anisotropic soil pattern that we interpret as mainly driven by successions of tectonised marl and sandstone outcrops with a decrease of the distances between these successive outcrops from East to West. This lithology driven soil pattern is expected to frequently occur within Mediterranean and semi-arid regions having a geology as complex as the one of the Cap Bon Region.

5.2 *Handling incomplete spatial covariates*

Spatially incomplete landscape covariates have been rarely handled in DSM studies. Yet it may be of great interest to mask the covariate layers in locations where the local conditions weaken the correlation with a soil property as e.g. vegetated areas for hyperspectral covariates. In such situations, we confirmed after Lagacherie et al. (2012) that co-kriging is a suitable DSM procedure provided that a co-regionalization model can be fitted to the data. We observed however that the co-kriging accuracy could rapidly decrease for sites located in the gaps of a covariate layer if short scale soil variations predominate. Co-kriging may also introduce a bias if the soil of the masked areas differ significantly from the not-masked ones. This may occur if soil properties are correlated with the local conditions that are considered for masking the landscape covariates.

AKNOWLEDGMENTS

The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National

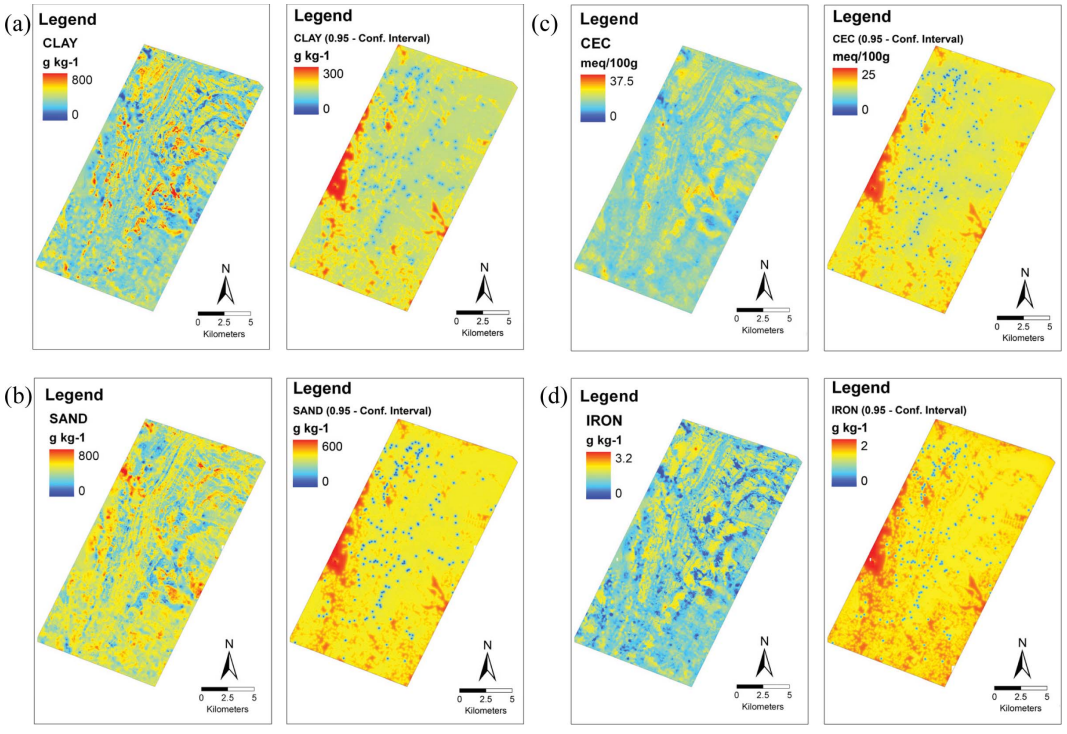


Figure 5. Maps of predicted soil properties and of their associated uncertainty (width of 95%-probability interval).

de Cartographie et de Télédétection, Tunisia), for providing the AISA-Dual images for this study. This hyperspectral data acquisition was granted by IRD, INRA and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01) ». We are also indebted to Yves Blanca (IRD-UMR LISAH Montpellier), Zakia Jenhaoui (IRD-UMR LISAH Tunis) for the soil sampling in 2009 and 2010 over the Lebna catchment and to Hedi Hamrouni (DG/ACTA Sol, Tunis) for his significant support to this study.

REFERENCES

Ciampalini, R., Lagacherie, P. & Hamrouni, H. 2012. Documenting GlobalSoilMap.net grid cells from legacy measured soil profile and global available covariates in Northern Tunisia in *DSM2012*, Minasny et al. (eds). Sydney.

Preti, G., Cisbani, A., De Cosmo, V., Galeazzi, C., Labate, D. & Melozzi, M. (2008). Hyperspectral Instruments for Earth Observation. International conference on Space Optics, October 14–17 Toulouse, France.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press.

Gomez, C., Lagacherie, P. & Bacha, S. 2012. Using Vis-NIR hyperspectral data to map topsoil properties over bare soils in the Cap Bon region, Tunisia in *DSM2012*, Minasny et al. (eds). Sydney.

Gomez, C., Lagacherie, P. & Coulouma, G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Under Review in *Geoderma.*

Lagacherie, P., Bailly, J.S., Monestiez, P. & Gomez, C. 2012. Using scattered hyperspectral imagery data to map the soil properties of a region. *Eur.J. Soil Science;* 63:110–119.

Ouerghemmi, W., Gomez, C. Nacer, S. & Lagacherie, P. (2011). Applying Blind Source Separation on hyperspectral data for clay content estimation over partially vegetated surfaces. *Geoderma*, 163(3–4), 227–237.

Stuffler, T., Kaufmann, H., Hofer, S., Förster, K.-P., Schreier, G., Müller, A., Eckardt, A., Bach, H., Penne, B., Benz, U. & Haydn, R. 2007. The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes, Acta Astronautica, 61, 1–6, 115–120.

Wackernagel, H. 1995. Multivariate geostatistics. Springer—Verlag. p. 255.

Geoderma 209–210 (2013) 168–176

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Combining Vis–NIR hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia)

Philippe Lagacherie [a,*], Anne-Ruth Sneep [b], Cécile Gomez [b], Sinan Bacha [c], Guillaume Coulouma [a], Mohamed Hédi Hamrouni [d], Insaf Mekki [e]

[a] INRA, UMR — LISAH, "Laboratoire d'étude des Interactions Sol–Agrosystème–Hydrosystème", INRA–IRD — SupAgro, Bat. 24. 2, place Viala, 34060, Montpellier, France
[b] IRD, UMR — LISAH, "Laboratoire d'étude des Interactions Sol–Agrosystème–Hydrosystème", INRA–IRD — SupAgro, Bat. 24. 2, place Viala, 34060, Montpellier, France
[c] Centre National de Cartographie et de Télédétection (CNCT), route de la Marsa, El Aouina, Tunis, Tunisia
[d] Ministry of Agriculture — Direction of Soils, 17, Hédi Karray Street, 2080 Ariana, Tunisia
[e] INRGREF, Institut National de Recherche en Génie Rural, Eaux et Forêts, BP 10, 2080 Ariana, Tunisia

ARTICLE INFO

Article history:
Received 28 February 2013
Accepted 9 June 2013
Available online xxxx

Keywords:
Remote sensing
Pedotransfer functions
Legacy soil data
Uncertainty

ABSTRACT

Previous studies have demonstrated that Visible Near InfraRed (Vis–NIR) hyperspectral imagery is a cost-efficient way to map soil properties at fine resolutions (~5 m) over large areas. However, such mapping is only feasible for the soil surface because the effective penetration depths of optical sensors do not exceed several millimeters. This study aims to determine how Vis–NIR hyperspectral imagery can serve to map the subsurface properties at four depth intervals (15–30 cm, 30–60 cm, 60–100 cm and 30–100 cm) when used with legacy soil profiles and images of parameters derived from digital elevation model (DEM). Two types of surface–subsurface functions, namely linear models and random forests, that estimate subsurface property values from surface values and landscape covariates were first calibrated over the set of legacy measured profiles. These functions were then applied to map the soil properties using the hyperspectral-derived digital surface soil property maps and the images of landscape covariates as input. Error propagation was addressed using a Monte Carlo approach to estimate the mapping uncertainties.

The study was conducted in a pedologically contrasted 300 km^2-cultivated area located in the Cap Bon region (Northern Tunisia) and tested on three soil surface properties (clay and sand contents and cation exchange capacity). The main results were as follows: i) fairly satisfactory (cross-validation R^2 between 0.55 and 0.81) surface–subsurface functions were obtained for predicting the soil properties at 15–30 cm and 30–60 cm, whereas predictions at 60–100 cm were less accurate (R^2 between 0.38 and 0.43); ii) linear models outperformed random-forest models in developing surface–subsurface functions; iii) due to the error propagations, the final predicted maps of the subsurface soil properties captured from 1/3 to 2/3 of the total variance with a significantly decreasing performance with depth; and iv) these maps brought significant improvements over the existing soil maps of the region and showed soil patterns that largely agreed with the local pedological knowledge. This paper demonstrates the added value of combining modern remote sensing techniques with old legacy soil databases.

1. Introduction

Implementing sustainable agricultural, hydrological and environmental management requires an improved understanding of the soil at increasingly finer scales. The current soil databases that exist are neither exhaustive nor precise enough to be efficiently used for this purpose. An alternative is digital soil mapping (DSM), which can be defined as "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables" (Lagacherie and McBratney, 2006).

Remote sensing images are major sources of input data for digital soil mapping. Until now, these images have mainly been used as spatial inputs for representing the landscape variables that are related to the soils, such as vegetation and parent material (the soil covariates) (McBratney et al., 2003). Boettinger et al. (2008) reviewed the main indicators that could be retrieved from a multispectral image for estimating these soil covariates. Airborne gamma-radiometry was also demonstrated as a suitable source of data for mapping soil

* Corresponding author. Tel.: +33 499 612 161; fax: +33 467 632 614.
E-mail address: lagache@supgro.inra.fr (P. Lagacherie).

0016-7061/$ – see front matter
http://dx.doi.org/10.1016/j.geoderma.2013.06.005

parent materials and their alteration rates (Wilford, 2012). After a spatial overlay with the sparse sets of observed and measured sites collected in a given area, the indicators derived from remote sensing have been used as independent variables in regression-like models or as external drift in geostatistic models (McBratney et al., 2003).

Alternatively, remote sensing can be considered a cost-efficient way to acquire dense spatial sets of soil property measurements and a means to overcome the lack of soil data that still severely limits the digital soil mapping performances (Lagacherie, 2008). A recent review (Mulder et al., 2011) reported successful attempts to directly map some key soil properties, including soil texture, soil organic carbon, iron content, carbonate content, and soil salinity, from multispectral, hyperspectral and radar images. Some initial promising results have been obtained in describing the complex spatial patterns of soil properties using airborne hyperspectral imagery (Gomez et al., 2012a; Gomez et al., 2012b; Schwanghart and Jarmer, 2011). However, such mapping is only feasible for soil surface properties because the effective penetration depths of optical and radar sensors do not exceed several millimeters (Liang, 1997) and several centimeters (Owe and Van de Griend, 1998), respectively.

Despite this limitation, the aforementioned remote sensing measurements of soil surface property variations can be valuable sources of information for predicting the variations of sub-surface properties. Eighty years of soil surveying have shown significant relations between surface and subsurface properties because i) most of the soils are formed from a single parent material that impacts all of the soil horizons and ii) the soil forming processes that drive the changes in soil properties with depth are themselves impacted by parent materials. Therefore, it should be possible to obtain valuable predictions of sub-surface soil properties from surface measurements when accounting for other landscape drivers that may also influence the pedogenic processes that result in soil property variations with depth, e.g., relief, climate, land use, and water table regime. The soil legacy data available in a growing number of soil databases around the world (Rossiter, 2004) are repositories of knowledge that could be useful in calibrating such predictions, following the approach of Gray et al. (2011) for depicting global soil–landscape relations.

This paper presents an approach for extending the successful mapping of three soil surface properties (clay content, sand content and CEC) to greater depths using hyperspectral imagery in the Cap-Bon Region, Northern Tunisia (Gomez et al., 2012a). Statistical functions inferring the sub-surface soil properties from surface property measurements and landscape variables were first calibrated from a local database of legacy measured and geo-referenced soil profiles. These functions were then coupled with hyperspectral imagery outputs to derive the maps of the subsurface properties with estimations of their associated uncertainties.

2. Methods

2.1. Surface–subsurface predictive functions

The value of a soil property S at the ith interval of depth (S_i) can be expressed as follows:

$$S_i = S_1 + \Delta_i \quad (1)$$

$$\text{with } \Delta_i = S_i - S_1 = f(\{S_1, L\}) + \varepsilon_i$$

where S_1 is the estimate of the soil surface property S that is assumed to be obtained independently, e.g., from remote sensing, and {L} is a set of easily available landscape variables. The surface–subsurface predictive function f is a statistical function that estimates the differences in the soil property values between the surface and the ith interval of depth (Δ_i) with a random error ε_i. The function requires a prior calibration from a dataset of existing measured soil profiles retrieved from a soil database. By considering such relative values as input instead of absolute values, the f functions' sensitivity to the measurement biases that frequently occur in soil databases (Batjes, 2009; Ciampalini et al., 2013) is expected to decrease. Furthermore predicted relative values are related with the pedogenic processes that induce differentiations across depths (e.g., clay illuviation), which makes the interpretation of results easier.

2.2. Computing soil property values at fixed depth intervals from legacy measured profiles

Eq. (1) expresses the relations between soil values at fixed depth intervals. Conversely, legacy soil databases usually contain soil property values that are determined for variable depth intervals that correspond to the spatially variable boundaries of the soil horizons. The depth intervals must therefore be equalized over the set of legacy measured profiles prior to using this set to calibrate Eq. (1).

This equalization was performed by interpolating the soil property values at the central depths of each horizon across a range of soil depths, varying between 0 and 200 cm. This interpolation was performed using equal-area spline functions (Bishop et al., 1999; Malone et al., 2009) that were fitted to each soil profile from the set of central horizon values. The spline function respects the average values of the soil properties over the horizons and assumes a continuous variation of soil properties across depths. The interpolated values given by the spline functions were then averaged to compute the new soil property values for fixed depth intervals.

2.3. Statistical inference models

The soil database with the soil depth equalized as described above was used to calibrate the aforementioned surface–subsurface predictive functions f. Two models were used: a linear model and a random forest, which addressed possible non-linearities that could affect the functions.

For linear models, the input variables were selected a priori. The correlations between the property at a certain depth and the explanatory variables were assessed using Pearson's product moment correlation coefficients. The correlation tests were performed using the function *cor.test* in R. To select the covariates, a significant p-value of 0.05 was considered the threshold.

Random Forest (RF) is an example of a machine learning method that has been already used in digital soil mapping (Grimm et al., 2008). RF consists of an ensemble of randomized classification and regression trees (CART) (Breiman, 2001). CART is itself a well-known learning algorithm (Breiman et al., 1984) that generates a tree structure by partitioning the data of a learning sample recursively into a number of groups, where each division is chosen to maximize some measure of difference among the response variables in the resulting two groups. RF combines many CART-like trees that are obtained randomly from learning samples and independently sampled values with the same distribution for all trees in the forest. Because of the large number of trees, the error estimation converges, and overfitting is expected not to occur (Breiman, 2001). The algorithm does not use all samples to calculate a tree; approximately one-third is not used and is called the out-of-bag (oob) data.

The Random Forest analysis was performed using the R package RandomForest (Liaw and Wiener, 2002). The selected RF parameters were chosen as follows: ntree (the number of trees in the forest) = 2000, and mtry (number of variables tried at each split) = 5.

2.4. Uncertainty estimations

It is important to accurately estimate the uncertainties associated with the soil predictions. Regarding the prediction method presented

above, two sources of uncertainty were identified: i) the uncertainty in the estimated value of the soil surface property S_1 and ii) the uncertainty associated with the surface–subsurface predictive functions (ε_i). McBratney et al. (2002) denoted these two components as input uncertainty and model uncertainty, respectively.

In this study, estimations of input uncertainty were assumed to be available from the statistical models used in estimating the values of the surface soil properties. The model uncertainty components were estimated differently according to the type of model, i.e., leave-one-out cross validations for the linear models and oob uncertainty estimates for Random Forest models. To compare the two models, a validation with an independent set of sites was also performed.

The two uncertainty components, i.e., input and model, were then combined following the method proposed by McBratney et al. (2002). A Monte Carlo simulation was performed to calculate the uncertainties resulting from the input and model uncertainties of the predicted subsurface soil property values as follows:

- Sample an S_1 value from a normal distribution with mean $= S_1$ and $\sigma = RMSE_h$ (the estimate of the S_1 uncertainty).
- Sample an ε_i value from a normal distribution with mean $= 0$ and $\sigma = RMSE_h$ (the estimate of ε_i).
- Calculate S_i using Eq. (1) with these new values.
- Repeat steps 1–3 n times, $n > 100$.
- Calculate the combined uncertainty (RMSEt) from the S_i distribution.

The two error components were reasonably assumed to be independent, which avoided the need to address the joint multivariate distributions. Furthermore, simulations were performed from the mean S_1 value only in the absence of any information showing variations in the uncertainty components across the learning sets.

3. Case study

3.1. Study area

The study area is located in the Cap Bon region in northern Tunisia (36°24′N to 36°53′N; 10°20′E to 10°58′E), 60 km east of Tunis, Tunisia (Fig. 1). This 300-km^2 area is mainly rural (>90%) and devoted to cereals in addition to legumes, olive trees, vineyards, and natural vegetation for grazing. It is a hilly area, with elevations ranging from 0 to 226 m. The main soil types are Regosols, Eutric Regosols preferentially associated with sandstone outcrops, Calcic Cambisols, and Vertisols preferentially formed on marl outcrops and lowlands. The southeastern region of the study area has a flatter landscape with sandy Pliocene deposits yielding Calcosols and Rendzina. The previous studies in this region (Ciampalini et al., 2012b; Gomez et al., 2012a; IAO, 2002) revealed a complex regional soil pattern, with a predominance of short range variations governed by the succession of contrasted outcrops (marl and sandstones) in the landscape.

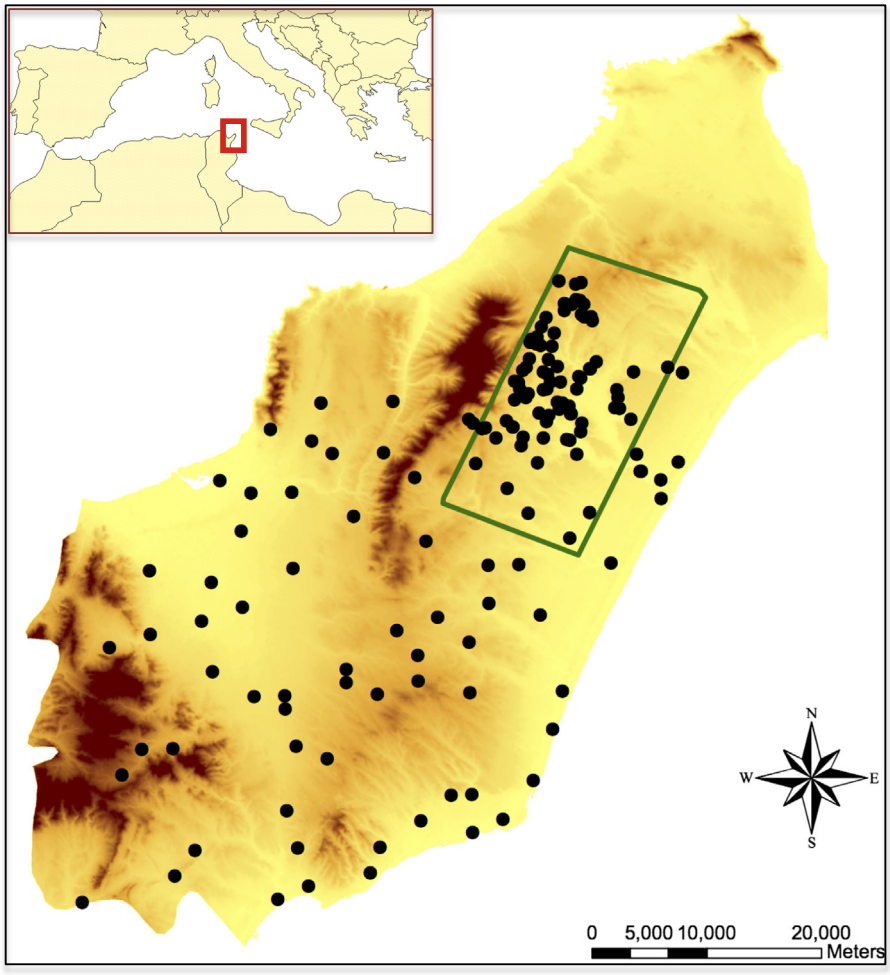


Fig. 1. Locations of the study area (green rectangle) and of the legacy measured soil profiles (black dots) plotted over the digital elevation model of the Cap Bon. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Legacy measured soil profiles

The soil database included three different sets of measured profiles (Fig. 1):

- Sixty-nine profiles with 293 soil horizons were described by Tahar Alaoui between 1973 and 1979 within the framework of the soil survey referenced as number 006. The analysis was performed by the soil analysis laboratory of the Ministry of Agriculture located in Tunis following the protocols used by the Ministry of Agriculture of Tunisia (Holderbach, 1964).
- Sixty-six profiles with 323 soil horizons were described by the IAO (Instituto Agronomico per l'Oltremare) 20th course professional master "remote sensing and natural resources evaluation" field survey staff from April 2 to 28, 2000, according to the IAO framework (IAO, 2002). The IAO laboratory performed the soil analyses following current soil analysis protocols. Eighteen of the 84 profiles initially included in this database were excluded because they were located outside the cultivated area that constitutes the study region.
- Seventeen profiles with 68 soil horizons were newly described in November 2010 and 2011. The analyses were performed by the Arras (France) INRA laboratory of soil analysis, which has obtained a COFRAC® certification. The laboratory followed current soil analysis protocols

The soil profiles of the two first datasets were initially located on topographic maps georeferenced in the Geodetic system "Carthage Tunisia". The profiles were then converted into the WGS84 Geodetic system, in which the third set of measured profiles, the landscape covariate geodatasets and the hyperspectral image were recorded. The conversion was made by the freely available geographic translator MSP GEOTRANS (National Geospatial-Intelligence Agency, Geographic Translator Version 2.4.2).

We used the above-described equal-area spline method to obtain the average soil property values over four fixed-depth intervals (0–15 cm, 15–30 cm, 30–60 cm, and 60–100 cm). These intervals were selected for having common boundaries with those specified in the world soil map (GlobalSoilMap.net, 2011). The differences with the latter were two-fold. First, the first two intervals specified by GSM (0–5 cm and 5–15 cm) were merged into a single 0–15 cm interval to better represent the soil surface characterized by the hyperspectral image. The purpose was to increase the chances of including a real soil property measurement in the interval and, thus, to decrease the spline interpolation errors due to side effects. The merging was possible in this case study because the predominance of plowed surfaces induced negligible differences in soil property values between 0 and 5 cm and 5 and15 cm and, thus, few expected biases in soil surface property estimations due to an increase in the measurement depth. Second, the deepest interval specified by GSM (100–200 cm) was not considered in this study due to the lack of measurements in the soil profile database for this depth. Finally, the pooled depth interval of 30–100 cm was also considered to explore the impact of soil layer thickness on soil prediction uncertainty.

Finally, the legacy soil dataset used in this study comprised 152 measured soil profiles with values of clay content, sand content and CEC at regular interval of depths. We used 122 soil profiles to calibrate the surface–subsurface prediction functions and 30 soil profiles to independently validate these functions.

3.3. Landscape variables

The landscape variables were all derived from a 30 m ASTER digital elevation model (DEM) with specific orthorectification and mosaicking made by gisxpert®. This DEM can be considered as a substitute for the high-quality 30 m DEM that will be available soon worldwide. The classical geomorphometric indicators found in the DSM literature were calculated from this DEM: elevation, slope, total curvature, profile curvature, multiresolution valley bottom flatness (MRVBF), multiresolution ridge top flatness (MRRTF), flow accumulation, and wetnex index. Each landscape variable was calculated for each soil profile location and added to the profile descriptions to serve as input for the statistical inference models.

3.4. Surface soil property maps obtained from Vis–NIR hyperspectral imagery

In this paper, we only provide a brief description of the surface soil property mapping using hyperspectral imagery that is fully detailed in an earlier paper (Gomez et al, 2012a). The mapping was based on AISA-Dual Vis–NIR hyperspectral airborne data acquired with a fine spatial resolution (5 m). The AISA-Dual airborne imaging spectrometer measured the reflected radiance in 359 non-contiguous bands covering the 400- to 2450-nm spectral domain. The radiance units were converted to reflectance units using the ASD spectrometer measurements of uniform surfaces (e.g., parking lots, asphalt, and concrete) that were collected at the same time during the flyover. An empirical line correction method was used to calibrate each flight line to the reflectance. Topographic corrections were performed using a 30-m digital elevation model built from ASTER data and ground control points. In this study, we removed the following: 1) the spectral bands in the blue part of the spectral domain (between 400 and 484 nm) due to noise in these bands and 2) the spectral bands between 1339 and 1464 nm and between 1772 and 2004 nm due to the vibrational–rotational H_2O absorption bands. Consequently, 280 AISA-Dual spectral bands were retained. Finally, to conserve only the bare soil surface, urban and vegetated surfaces and water were masked; these masked surfaces accounted for approximately 50% of the study area.

Partial least square regressions were applied to model the relations between the surface soil properties and AISA-Dual spectra. Three soil

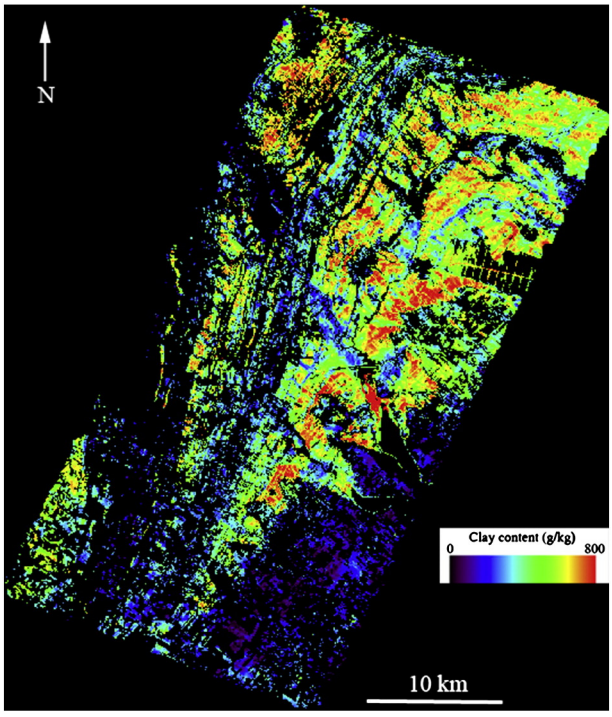


Fig. 2. Clay content predicted over bare soils from AISA-Dual spectra. Black areas correspond to unmapped surfaces (watered, urban or vegetated surfaces at the fly date).

properties that are currently available in legacy soil databases, namely clay content, sand contents and CEC, were successfully predicted with cross-validation R^2 values equal to 0.8, 0.7 and 0.7, respectively (Gomez et al, 2012a, Fig. 3). These three soil properties were then mapped over the bare soils of the AISA-Dual image (Fig. 2), thus providing realistic patterns of variations.

4. Results

4.1. Soil property variations with depths

Fig. 3 shows the distributions of the three properties for the depth intervals of 0–15 cm, 15–30 cm, 30–60 cm and 60–100 cm for the whole dataset of 152 soil profiles. For all of the considered depths, the soil properties were characterized by a high spatial variation. We observed opposite distribution trends of clay and sand content as a consequence of the highly significant negative correlation between these two variables (Gomez et al., 2012a). The general trend was an increase of clay and a decrease of sand with depth, and the two first depth intervals showed the greatest contrasts. This trend could be interpreted as the result of the deposition of the eroded materials from the uphill sandstone outcrops along the slopes that may increase topsoil sand contents. Conversely, the CEC values did not exhibit a clear variation with depth. This suggests that the clay increase with depth may be compensated by a decrease in organic matter and mineral fertilizers, which are expected to be positively correlated with CEC.

Table 1 shows the correlations between the subsurface soil properties, including both the surface soil properties and the potential landscape covariates. The former presented a strong relationship with the corresponding surface soil properties regardless of the depth interval. Significant relationships were also observed with elevation and, to a lesser extent, profile curvature and MRVBF. These landscape variables were themselves closely related to the erosion and deposition processes and to the soil water regime, which are the main drivers of the soil property variations with depth in this area.

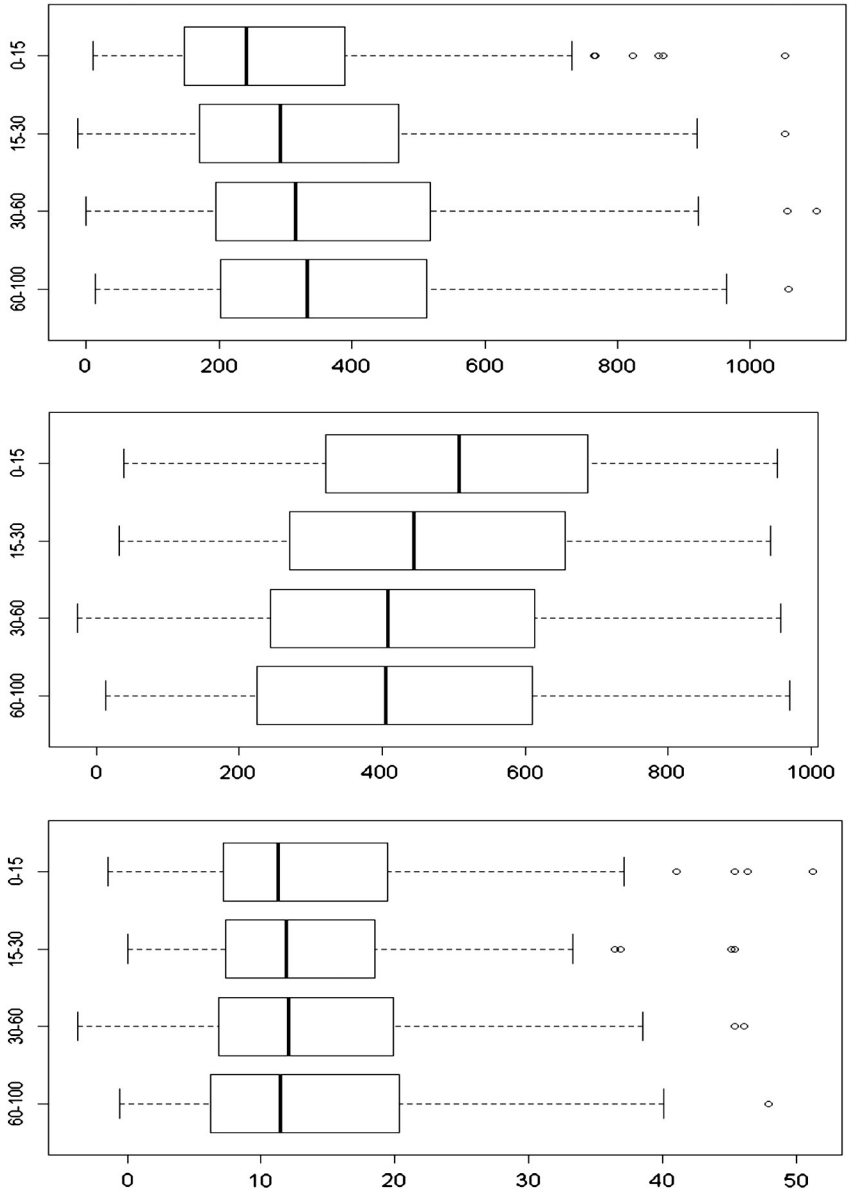


Fig. 3. Distributions of a) clay content (in g/kg), b) sand content (in g/kg) and c) cation exchange capacity (in Meq/100 g) for the intervals of depths 0–15 cm, 15–30 cm, 30–60 cm and 60–100 cm (Y axis of the figures).

Table 1
Significance of correlations between sub-surface properties and potential explanatory variables: surface soil properties and landscape properties. +++: p values < 10^{-11}, ++ p values > 10^{-11} and <10^{-2}, +: p values > 10^{-2} and <0.1. No significant correlations (p values > 0.1) were found for slope, aspect, wetness index and MRRTF at all depth intervals.

Property	Surface soil property 0–15 cm	Elevation	MRVBF	Profile curve
Clay				
15–30 cm	+++			
30–60 cm	+++	+		
60–100 cm	+++	++	+	+
Sand				
15–30 cm	+++	+		
30–60 cm	+++	++		
60–100 cm	+++	++	+	+
CEC				
15–30 cm	+++			
30–60 cm	+++			
60–100 cm	+++			

4.2. Surface–subsurface soil functions results

Table 2 shows the performances of the surface–subsurface functions built using linear models and random forest for different depth intervals. The performances of the linear model-based functions were fairly variable across soil properties and depths. Leave-one-out cross-validations showed fairly satisfactory results (R^2 between 0.55 and 0.81) at 15–30 cm and 30–60 cm but were less accurate (R^2 between 0.38 and 0.43) at 60–100 cm. The prediction uncertainties tended to increase with depth for all soil properties but significantly dropped for the last soil depth interval (60–100 cm). Grouping the two last soil depths intervals into a single thicker interval (30–100 cm) merely resulted in uncertainty values that were intermediate between the two former ones; therefore, an averaging effect was not observed. Predictions of clay and sand contents were less uncertain than those of CEC for the upper layers (15–30 cm and 30–60 cm), whereas much smaller performance differences between soil properties were observed for the deepest layers.

The results above described seemed to be fairly realistic when comparing with the R^2 values obtained from independent validation samples, except for those predicting the sand content for the deepest soil layers. However, the differences in the R^2 and RMSE values did not always vary in accordance because of the differences in the total variances; therefore, a lack of representativeness of this small validation set is revealed.

Using a more sophisticated statistical inference model, such as Random Forest, did not bring any real improvements to the soil property predictions. Linear model-based functions generally outperformed Random Forest-based ones regardless of the method for computing the uncertainty estimates. Furthermore random forest showed substantial differences between cross-validation and independent validation results, which revealed overfitting. The aforementioned variations in performance with soil depth intervals and properties were also observed for Random Forest-based functions.

4.3. Mapping subsurface soil properties

The surface–subsurface prediction functions obtained from linear models were then applied over the bare soil surfaces of the study region. The surface soil property values used as input for these functions were provided by the hyperspectral-predicted images of soil properties (Fig. 2). As expected, the image of subsurface clay content (Fig. 4, on the left) is fairly similar to the surface image (Fig. 2). However, the map of their differences (Fig. 4, on the right) reveals some variations that are governed by soil and landscape variables. The order-one driver seems to be the clay content at the soil surface because its low and high values induced large clay content increases and decreases, respectively. This relation can be partly interpreted as the result of the short scale successions across the landscape of marl and sandstones that form the parent materials of the soils. Indeed, when extremely low or high clay content values are observed at the soil surface, which correspond to pure sandstone and marl outcrops, the bottom of the soil profile may be at least partially influenced by the opposite parent material and thus results in a more balanced soil texture. However, this major trend is partly mitigated by the landscape position of the soil profile because only slight increases in clay content were mapped in valleys (black circles in Fig. 4), whereas these valleys had low clay content at the soil surface (Fig. 2). Therefore, the surface–subsurface function accounted for the deposition of materials in lowlands and generated weakly contrasted soil textures with depth. This feature, however, may cause some artifacts in the opposite landscape positions (the upper areas; the blue circle in Fig. 4), where the attenuation of the decrease of clay with depth has no clear pedological interpretation.

Table 2
Performances of the surface–subsurface functions: R^2 and RMSE between brackets in g/kg for clay and sand and in Meq/100 g for CEC.

	N samples		Linear model validation		Random-Forest validation	
			R^2 (RMSE)		R^2 (RMSE)	
	CAL	VAL	Prior[a]	Post[b]	Prior	Post
Clay						
15–30 cm	122	30	0.78 (67)	0.81 (87)	0.77 (69)	0.78 (93)
30–60 cm	122	30	0.56 (103)	0.62 (140)	0.54 (105)	0.56 (150)
60–100 cm	107	23	0.46 (115)	0.38 (138)	0.47 (114)	0.19 (158)
30–100 cm	107	23	0.55 (98)	0.54 (115)	0.56 (98)	0.46 (125)
Sand						
15–30 cm	122	30	0.83 (97)	0.75 (119)	0.78 (107)	0.78 (111)
30–60 cm	122	27	0.61 (148)	0.61 (135)	0.59 (153)	0.68 (121)
60–100 cm	107	23	0.56 (164)	0.38 (139)	0.56 (164)	0.14 (163)
30–100 cm	107	23	0.62 (146)	0.49 (126)	0.62 (145)	0.37 (140)
CEC (Meq/100 g)						
15–30 cm	117	30	0.79 (3.3)	0.71 (4.6)	0.79 (3.4)	0.72 (4.6)
30–60 cm	116	30	0.48 (5.4)	0.55 (5.7)	0.50 (5.4)	0.53 (5.8)
60–100 cm	92	22	0.41 (6.2)	0.53 (5.0)	0.35 (6.5)	0.29 (5.5)
30–100 cm	92	22	0.47 (5.5)	0.53 (4.5)	0.44 (5.7)	0.49 (4.7)

[a] From leave-one-out cross validation.
[b] From validation with independent sample.

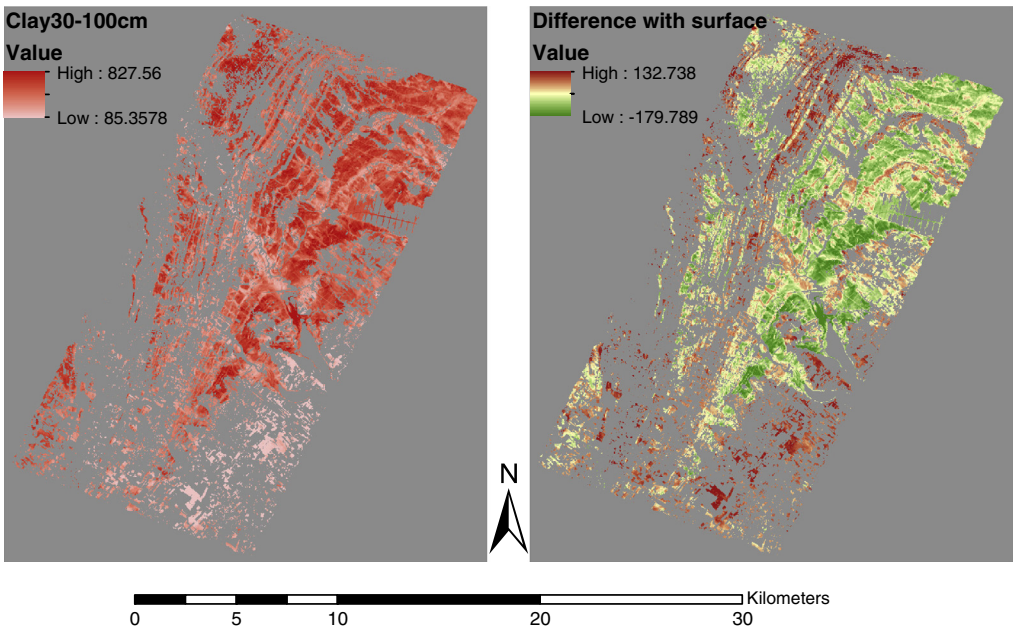


Fig. 4. Map of clay content (g/kg) at 30–100 cm depth (left) and map of differences of clay content between 0–15 cm and 30–100 cm (right). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

4.4. Uncertainties of the subsurface property spatial prediction

A priori estimates of the total uncertainty that combine the uncertainties of the inputs, the hyperspectral estimates of surface soil property values and the uncertainties of the models, i.e., the surface–subsurface functions, were computed using the Monte Carlo approach presented above. These estimates were performed for the linear models only because they showed the best results in the previous step (Table 2). These estimates could not be validated with an independent set of soil profiles because fewer than 20 soil profiles with soil property measurements, were actually located in areas with hyperspectral image outputs.

The errors estimated a priori using the Monte-Carlo approach (Table 3) were higher than those of the surface–subsurface functions (Table 2) but conserved the above-described similar hierarchy between depths and properties. This was expected because hyperspectral estimates of the input uncertainties that were fairly similar across soil properties (see Case study) were merely added into the Monte Carlo simulation of soil surface properties. Finally, it can be concluded that using linear surface–subsurface functions in association with the hyperspectral image allowed us to capture between one-third and two-thirds of the total variance of the subsurface properties.

Table 3
Estimated uncertainties of digital maps of subsurface soil properties.

Property	Depth	RMSE	R^2
Clay (g/kg)	15–30 cm	101	0.51
	30–60 cm	122	0.38
	60–100 cm	128	0.33
	30–100 cm	119	0.37
Sand (g/kg)	15–30 cm	133	0.67
	30–60 cm	170	0.49
	60–100 cm	181	0.46
	30–100 cm	165	0.51
CEC (Meq/100 g)	15–30 cm	4.4	0.69
	30–60 cm	6.0	0.44
	60–100 cm	6.4	0.35
	30–100 cm	6.0	0.37

5. Discussion

5.1. Added value of hyperspectral images

The hyperspectral-based DSM approach presented in this paper can be first compared with the classical DSM approaches that use currently available landscape covariates derived from digital elevation models and multispectral remote sensing images (see Grunwald (2009) for a review). However, the published papers specifically considering subsurface soil properties are in the minority in the DSM literature, and the papers considering both the soil properties targeted in this study and Mediterranean or arid regions are even more scarce. The studies all reported weak performance (Ciampalini et al., 2012a; Gallichand and Marcotte, 1993). In particular, Ciampalini et al. dealt with a 2822-km^2 study region that included the one considered in this study. They obtained estimated confidence intervals that were 63% higher on average than those derived from the results of Table 3 and, furthermore, significantly underestimated the uncertainty. The hyperspectral images then clearly added value to the soil surface information. Classical DSM approaches are less limited in space due to the costs of airborne hyperspectral data. It can be expected, however, that these costs will significantly decrease thanks to the launching of several hyperspectral satellites by the end of this decade. PRISMA (PRecursore IperSpettrale della Missione Applicativa) is an Italian hyperspectral sensor expected to be launched in 2013 (spatial resolution: 30 m) (Giampaolo et al., 2008), EnMap (*En*vironmental *M*apping and *A*nalysis *P*rogram, http://www.enmap.org/) is a German hyperspectral sensor expected to be launched in 2014 (spatial resolution: 30 m) (Stuffler et al., 2007), and HYPXIM is a French hyperspectral sensor expected to be launched in 2018–2020 (spatial resolution: 15 m) (Michel et al., 2010).

Our approach can also be compared with using proximal sensing approaches that dealt specifically with the mapping of subsurface soil properties. Geophysical methods such as electro-magnetic induction (EMI) were used to estimate the textural properties of sub-surface soil layers (Triantafilis and Lesch, 2005; Abdu et al., 2008). In particular, Triantafilis and Lesch estimated the bulk soil average clay content to a depth of 7 m with good accuracy (RMSE < 60 g/kg) from a large set of field EMI measurements that were used as inputs of geostatistical

procedures. Although those results clearly outperformed those obtained in this paper, much field work was required for proxy EMI measurements, and the resolution in depth was much coarser than our specification.

Finally, using hyperspectral data in association with legacy measured soil profiles can be considered in the future as a possible compromise between classical DSM approaches and proximal soil sensing for mapping the subsurface soil properties at moderate costs.

5.2. On surface–subsurface functions

The surface–subsurface functions that we proposed in this paper can be considered part of a new type of pedotransfer functions (Bouma, 1989), i.e., predictive functions of certain soil properties from other more easily measured properties. We demonstrated that these functions could overcome the depth limitations of remote sensing soil characterizations of soils that more or less hamper the use of remote sensing in digital soil mapping, regardless of the sensors used. The case study we addressed was favorable for applying such functions because it mainly included recently formed soils. Therefore, any impact of long-term weathering and geochemical processes, whose variations in space could have made surface–subsurface relationships between soil properties much more complex to depict, were removed. This was however less true for CEC whose variations at the soil surfaces were also driven by agricultural practices, which explained the smaller performances of the surface–subsurface functions for this property. Additionally, the correct set of legacy measured soil profiles must be carefully selected to calibrate the surface–subsurface functions. This set must include a sufficient number of profiles while avoiding outliers. Finally, the development of efficient surface–subsurface functions for use in synergy with remote sensing is an open research subject that may constitute a new opportunity to valorize the local, national and continental legacy soil databases. This advancement will not be possible without involving the considerable amount of pedological knowledge on soil variations and processes that has been accumulated for many years.

5.3. Adding field data

The performance of our method was clearly hampered by the soil database used for calibrating the surface–subsurface functions. First, the profile spacing was too large to capture with great detail the variations in soil properties with depth across the study area. Second, there were not enough profiles with bare soil surface hyperspectral measurements to build a co-regionalization model that would have permitted a more efficient prediction through co-kriging (Ciampalini et al., 2012b). These problems are linked to the use of legacy data, which does not enable controlling the spatial sampling of field data. To overcome this problem would require adding field data following a spatial sampling oriented by hyperspectral maps of surface soil properties, e.g., using latin hypercube sampling (Minasny and McBratney, 2006). However, classical profiles from wet laboratory analyses are still extremely expensive. Replacing them with cheaper techniques, such as soil spectroscopy (Viscarra Rossel et al., 2006) or proximal soil sensing (Viscarra Rossel et al., 2011), should be explored.

6. Conclusions

In this paper, we presented an approach that enabled the estimation of subsurface soil properties from surface properties by merging data from legacy measured profiles, hyperspectral imagery and landscape covariates. We investigated surface–subsurface functions, including linear models, Random Forests variables, and landscape covariates for mapping soil properties. The error propagation was addressed using a Monte Carlo approach to estimate the mapping uncertainties. The main outcomes of this study are as follows:

- By providing unprecedented insight into the variation patterns of some surface soil properties, Vis–Nir hyperspectral imagery provides potentially valuable information for mapping soil properties at various depth intervals.
- The extension to greater depths can be achieved by surface–subsurface functions calibrated from a set of legacy measured soil profiles retrieved from local or regional soil databases. Future work should be conducted to optimize such functions in various pedological contexts.
- New field samplings with low cost soil sensing techniques would significantly improve the present performances that are limited by only using legacy data.

Acknowledgments

The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National de Cartographie et de Télédétection, Tunisia) for providing the AISA-Dual images for this study. This hyperspectral data acquisition was granted by IRD, INRA and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01). We are also indebted to Yves Blanca (IRD-UMR LISAH Montpellier) for the field work.

References

Abdu, H., Robinson, D.A., Seyfried, M., Jones, S.B., 2008. Geophysical imaging of watershed subsurface patterns and prediction of soil texture and water holding capacity. Water Resources Research 44.
Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use and Management 25, 124–127.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45.
Boettinger, J., Ramsey, R.D., Bodily, J.M., 2008. Landsat spectral data for digital soil mapping. In: Hartemink, A., McBratney, A.B., Mendonca Santos, L. (Eds.), Digital Soil Mapping with Limited Data, pp. 193–199.
Bouma, J., 1989. Using soil survey data for quantitative land evaluation. Advances in Soil Science 177–213.
Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.
Breiman, L., Friedmann, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth & Brooks.
Ciampalini, R., Lagacherie, P., Hamrouni, H., 2012a. Documenting GlobalSoilMap.net grid cells from legacy measured soil profile and global available covariates in Northern Tunisia. Digital Soil Assessments and Beyond. CRC Press, pp. 439–444.
Ciampalini, R., Lagacherie, P., Monestiez, P., Walker, E., Gomez, C., 2012b. Co-kriging of soil properties with Vis–NIR hyperspectral covariates in the Cap Bon region (Tunisia). Digital Soil Assessments and Beyond. CRC Press, pp. 393–398.
Ciampalini, R., Lagacherie, P., Gomez, C., Grünberger, O., Hamrouni, M.H., Mekki, I., Richard, A., 2013. Detecting, correcting and interpreting the biases of measured soil profile data: a case study in the Cap Bon Region (Tunisia). Geoderma 68–76.
Gallichand, J., Marcotte, D., 1993. Mapping clay content for subsurface drainage in the Nile Delta. Geoderma 58, 165–179.
Giampaolo, P., Cisbani, A., De Cosmo, V., Galeazzi, C., Labate, D., Melozzi, M., 2008. Hyperspectral instruments for Earth observation. International Conference on Space Optics, October 14–17.Toulouse, France.
GlobalSoilMap.net, 2011. Specifications GlobalSoilMap.net Products, Version 2.0. 48 (http://www.globalsoilmap.net/system/files/GlobalSoilMap.net%20specifications_V2_March2011.pdf).
Gomez, C., Lagacherie, P., Bacha, S., 2012a. Using Vis–NIR hyperspectral data to map topsoil properties over bare soils in the Cap-Bon Region, Tunisia. In: Minasny, B., Malone, B., McBratney, A.B. (Eds.), Digital Soil Assessment and Beyond. Taylor & Francis, London, pp. 387–392.
Gomez, C., Lagacherie, P., Coulouma, G., 2012b. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. Geoderma 189–190, 176–185.
Gray, J.M., Humphreys, G.S., Deckers, J.A., 2011. Distribution patterns of World Reference Base soil groups relative to soil forming factors groups relative to soil forming factors. Geoderma 373–383.
Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — digital soil mapping using Random Forests analysis. Geoderma 146, 102–113.
Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152, 195–207.
Holderbach, L., 1964. Exposé des méthodes d'analyses de sol et d'eau en usage à la subdivision d'études de pédologie et d'hydrogéologie. Secrétariat d'état à l'Agriculture, République de Tunisie 95.
IAO, 2002. Land resources of the Oued Lebna catchment (Tunisia). 20th Course Professional Master: Remote Sensing and Natural Resources Evaluation.Istituto Agronomico per l'Oltremare, Florence, Italie pp. 132.

Lagacherie, P., 2008. Digital soil mapping: a state of the art. p. 3–14. In: Hartemink, A.E., McBratney, A.B., Mendonca Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science.

Lagacherie, P., McBratney, A.B., 2006. Chapter 1 Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. Developments in Soil Science 3–22.

Liang, S., 1997. An investigation of remotely-sensed soil depth in the optical region. International Journal of Remote Sensing 18, 3395–3408.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. (R News 2, 18–22Available at http://cran.r-project.org/doc/Rnews/).

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109, 41–73.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Michel, S., Lefevre-Fonollosa, M.-J., Hosford, S., 2010. HYPXIM — a hyperspectral satellite defined for science, security and defence users. Hyperspectral Workshop 2010.Frascati, Italy 17–19 (March 2010).

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences 32, 1378–1388.

Mulder, V.L., De Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — a review. Geoderma 162, 1–19.

Owe, M., Van de Griend, A., 1998. Comparison of soil moisture penetration depths for several bare soils at two microwave frequencies and implications for remote sensing. Water Resources Research 34, 2319–2327.

Rossel, R.A.V., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Advances in Agronomy. , 113. Elsevier.

Rossiter, D.G., 2004. Digital soil resource inventories: status and prospects. Soil Use and Management 20, 296–301.

Schwanghart, W., Jarmer, T., 2011. Linking spatial patterns of soil organic carbon to topography — a case study from south-eastern Spain. Geomorphology 126, 252–263.

Stuffler, T., Kaufmann, H., Hofer, S., Förster, K.-P., Schreier, G., Müller, A., Eckardt, A., Bach, H., Penne, B., Benz, U., Haydn, R., 2007. The EnMAP hyperspectral imager — an advanced optical payload for future applications in Earth observation programmes. Acta Astronautica 61 (1–6), 115–120.

Triantafilis, J., Lesch, S.M., 2005. Mapping clay content variation using electromagnetic induction techniques. Computers and Electronics in Agriculture 46, 203–237.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma 183–184, 124–142.