Remote Sensing of Environment 128 (2013) 197–211

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Integration of in situ measured soil status and remotely sensed hyperspectral data to improve plant production system monitoring: Concept, perspectives and limitations

Laurent Tits [a,*], Ben Somers [b], Jan Stuckens [c], Jamshid Farifteh [a], Pol Coppin [a]

[a] Dept. of Biosystems, M3-BIORES, Katholieke Universiteit Leuven, W. de Croylaan 34, BE-3001 Leuven, Belgium
[b] Flemish Institute for Technological Research (VITO), Centre for Remote Sensing and Earth Observation Processes (TAP), Boeretang 200, BE-2400 Mol, Belgium
[c] GeoID, Researchpark Haasrode, Interleuvenlaan 62, 3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 25 June 2012
Received in revised form 10 October 2012
Accepted 11 October 2012
Available online 8 November 2012

Keywords:
Radiative transfer
Look-up-table
Spectral Mixture Analysis
Chlorophyll
Water
Leaf Area Index
Citrus
Soil moisture
Vegetation index
Agriculture

ABSTRACT

A common problem in agricultural remote sensing is the sub-pixel spectral contribution of background soils, weeds and shadows which impedes the effectiveness of spectral vegetation indices to monitor site-specific variations in crop condition. To address this mixture problem, the present study combines in situ measured soil status and remotely sensed hyperspectral data in an alternative spectral unmixing algorithm. The model driven approach, referred to as Soil Modeling Mixture Analysis (SMMA), combines a general soil reflectance model and a modified spectral mixture model providing as such the opportunity to simultaneously extract the sub-pixel cover fractions and spectral characteristics of crops. The robustness of the approach was extensively tested using ray-tracer data (PBRT) from a virtual orchard, and results showed an improved monitoring of the crop's chlorophyll, water content and Leaf Area Index (LAI). A significant increase in the R^2 between vegetation indices and the biophysical parameters was observed when index values were calculated from the *pure* vegetation signals as extracted by SMMA as opposed to index values calculated from the original (mixed) image pixels (GM1: $\Delta R^2 = 0.19$; MDWI: $\Delta R^2 = 0.38$; sLAIDI: $\Delta R^2 = 0.14$).

1. Introduction

During the last 50 years, agricultural production strategies have changed dramatically, mostly due to economic decisions to reduce the inputs and maximize the profits, and due to environmental guidelines for a more efficient and safer use of chemicals (Pinter et al., 2003). Perhaps the most significant change is the shift towards precision, or site-specific, crop management. Within-field variability is taken into account to optimize the management practices. This requires accurate and consistent information on soil and plant conditions across the farm, and at temporal and spatial scales that match rapidly evolving capabilities to vary cultural procedures, irrigations, and agrochemical inputs (Dorigo et al., 2007; Pinter et al., 2003). An emerging technology in precision agriculture to provide this information is hyperspectral remote sensing (Dorigo et al., 2007; Zarco-Tejada et al., 2004). Hyperspectral data contains valuable information with respect to the physical and chemical properties of surface targets. Moreover, when hyperspectral remote sensing is applied with a high temporal resolution – by sensors on-board of satellites – intensive monitoring of biophysical and biochemical crop characteristics during growth can become a reality, through for instance the use of vegetation indices (Zarco-Tejada et al., 2004) or through radiative transfer model inversions (Jacquemoud et al., 2009). Above ground biomass (Clevers et al., 2007), canopy chlorophyll content (Zarco-Tejada et al., 2004), leaf equivalent water thickness (Eitel et al., 2006), Leaf Area Index (LAI, Delalieux et al., 2008) and leaf nutrient status (Hansen & Schjoerring, 2003) are only a few of the biophysical attributes that have successfully been monitored using features derived from hyperspectral observations.

Aforementioned examples reveal the huge potential of hyperspectral satellite observations for plant production system management. This is also reflected in the mission statements of the upcoming hyperspectral satellites. For the Environmental Mapping and Analysis Program (EnMAP), the main objective is "to investigate a wide range of ecosystem parameters encompassing agriculture, forestry, soil and geological environments, coastal zones and inland waters" (Environmental Mapping & Analysis Program, 2012). The Canadian Hyperspectral Environment and Resources Observer (HERO) states it as "provide the Hyperspectral user community with high quality information on the surface of the Earth, specifically the plants and materials that cover it and their changes with time" (MDA corporation, 2012).

Yet, present-day input parameters provided by hyperspectral satellite observations are not adequate enough. Spectral reflectance distortions caused by the atmosphere and the considerable orbital height of satellites (roughly ranging between 650 and 800 km) drastically impact image interpretation (Brown, 1992). Most of these

* Corresponding author. Tel.: +32-16-328146.
E-mail address: laurent.tits@biw.kuleuven.be (L. Tits).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.rse.2012.10.006

distortions can be corrected for by geometric, radiometric and atmospheric algorithms (Biliouris et al., 2009; Itten & Meyer, 1993). Yet, a persisting issue is the mixture problem associated with the composite nature of pixels (Adams et al., 1993). Because of the large observation height, in combination with the high spectral resolution of hyperspectral satellites, the size of an image pixel is considerable. All current (i.e. Hyperion) and upcoming (e.g. HERO, EnMAP, TAIKI HSC-III) hyperspectral satellites have a spatial resolution of 30 m. It exceeds in many cases the size of the objects of interest. Consequently, the reflectance signal of a pixel is the integrated result of spectral contributions of different subpixel elements building up a pixel. This constrains the accuracy of spectral analysis based on vegetation index development or other feature extraction techniques (Roberts et al., 1993). The mixture problem is also aggravated in agricultural image scenes where mixed pixels prevail because of the discontinuous open canopies typical for most (perennial) cropping systems (Peddle & Smith, 2005). In Fig. 1, this is demonstrated by showing a high resolution aerial image of a citrus orchard. A 30 by 30 m pixel grid is superimposed to illustrate that the reflectance of most satellite pixels in an agricultural image scene cannot be simply interpreted in terms of crop properties only. For most pixels, subpixel mixing of soils and/or shadows occurs.

In more traditional plant system monitoring studies using (spaceborne) remote sensing, the mixing problem is often ignored or rather superficially dealt with. The majority of available unmixing algorithms are focused on roughly estimating the proportional ground cover of the vegetation class in a mixed pixel (e.g. Quarmby et al., 1992; Asner & Heidebrecht, 1992; Lobell & Asner, 2004; Fitzgerald et al., 2005; Peddle & Smith, 2005; Somers et al., 2009a). This technique, hereafter referred to as Area Unmixing (AU), is popular for the rapid, early and low-cost assessment of crop area statistics from multi-temporal and -spectral low (spatial) resolution imagery (Verbeiren et al., 2008), but the technique is clearly unable to extract spectrally 'pure' vegetation characteristics, uncontaminated by pixel components, such as soil and shadow.

Several authors dealt with this problem partially by adjusting existing vegetation indices, in particular the Normalized Difference VI (Tucker, 1979) and the Simple Ratio (Jordan, 1969) indices, to make them more robust for soil background effects. The basic assumption of these soil-adjusted vegetation indices is that soils are characterized by a unique linear relationship between NIR (700–1350 nm) and VIS (400–700 nm) reflectance, i.e., the soil line. Huete (1988) adapted the formula of NDVI by including the coefficients of the soil line in the Soil-Adjusted Vegetation Index (SAVI). The transformed (TSAVI; Baret & Guyot, 1991), modified (MSAVI; Qi et al., 1994) and optimized SAVI (OSAVI; Rondeaux et al., 1996) are all variants of the traditional SAVI. Despite these efforts, the success of the soil-adjusted indices is limited because the soil line is not as generic as assumed while the technique is mainly restricted to corrections in the VIS–NIR spectral domain (Delalieux et al., 2008; Rondeaux et al., 1996).

An accurate monitoring method for critical crop production parameters, however, requires the removal of undesired spectral background effects from mixed image pixels. Consequently, a more generic approach to reduce subpixel background effects is needed allowing the accurate and site-specific monitoring of plant production systems. Tits et al. (in press) proposed a Signal Unmixing (SU) methodology to extract the 'pure' vegetation signal from a mixed pixel signal consisting of soils and vegetation. Using an extensive spectral library for each endmember, a Multiple Endmember Spectral Mixture Analysis (MESMA, Roberts et al., 1998) approach was used to evaluate different endmember combinations, selecting the endmember combination with the lowest modeling error as the spectral signatures of the components in the pixel. Two major limitations of the MESMA methodology are (i) the requirement of large libraries to encompass the spectral variability that can be expected to occur in the field, and (ii) ill-posedness effects resulting in multiple endmember combinations that produce the same mixed spectrum (Tits et al., in press). However, results showed that the performance of the SU model improved significantly with increasing knowledge on the soil endmember, as it minimizes the two difficulties described above.

In this study we hypothesize that the integration of in situ measured soil status data and remotely sensed hyperspectral data can provide the needed information to estimate the spectral signature of the soil endmember, so that the subpixel vegetation reflectance signature can be extracted from the mixed image pixels. Previous studies have already demonstrated the added value of combining remote sensing and in situ data inputs for improved image interpretation. Examples for more general applications are the vicarious calibration of sensors (Dinguirard & Slater, 1999) and the atmospheric correction of satellite images using the invariant-object method (Liang et al., 2001). In situ data is also used in combination with remotely sensed spectral data in applications such as climatology (Reynolds et al., 2002), sediment transportation (Ouillon et al., 2004) and agriculture (Dzikiti et al., 2010). Both Kerr and Ostrovsky (2003) and Zaks and Kucharik (2011) state that a combination of field measurements and remote sensing is needed to solve the problem of scale mismatch between field data and most remote sensing data sources. The same principle is used in this study to introduce a SU technique to remove soil background and shadow effects from optical satellite images. The proposed concept consists of the assimilation of soil reflectance models (Müller & Décamps, 2001; Somers et al., 2009c, 2010) and Spectral Mixture Analysis (SMA; Adams et al.,

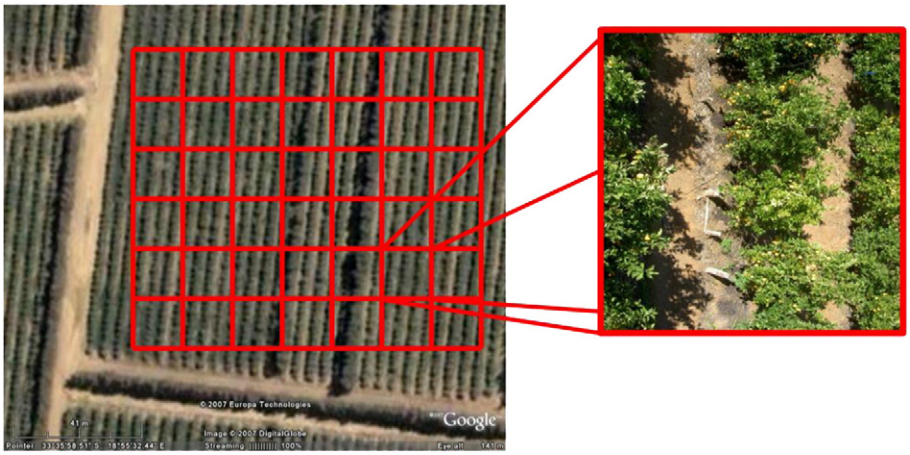

Fig. 1. Example of a 30 by 30 m grid over a high resolution image, combined with a detail of a pixel to illustrate the mixed nature of the pixel. Adapted from Somers (2009d).

1993), to extract pure vegetation spectra from mixed image pixels (Tits et al., in press). The SU technique combines in situ measured soil moisture content (SMC) and optical spectral satellite data in an adapted SMA algorithm to reduce spectral background effects in vegetation indices. The specific goals of this research are threefold and can be summarized as:

i. Evaluate the potential of SMMA to reduce background effects from vegetation indices
ii. Evaluate the added value of SMMA in the monitoring of plant production system state parameters (e.g. canopy water and chlorophyll content, Leaf Area Index).
iii. Evaluate the sensitivity of the proposed technique on the accuracy of the in situ measured SMC

Overall, the corollary of this work is to provide a better definition and quantification of spectral inputs for plant production system models. Yet, the main objective of this study is a 'proof of concept' of SU rather than the provision of a fully operational technique. The SU technique is evaluated using simulated hyperspectral imagery, based on measurements of spectral, dendrometric and orchard-specific properties of citrus orchards in the West-Cape Province of South Africa.

2. Soil Modeling Mixture Analysis

2.1. Concept

To derive accurate information on the spatial and temporal dynamics of the condition of crops from hyperspectral data, it is required that the mixed pixel spectral signatures are unmixed in crop related and non-crop information. We hypothesize that data fusion of in situ measured soil moisture content (SMC) and remotely sensed hyperspectral data can reduce mixing problems and as such lead to an increased accuracy in extracting biophysical and structural crop characteristics from hyperspectral satellite imagery. The concept is illustrated in Fig. 2. A soil moisture sensor network combined with some interpolation model can provide a surface soil moisture content map. The combined analysis of the soil moisture map and the hyperspectral data cube allows reducing background contamination in the spectral data, providing a hyperspectral crop reflectance map from which the spectral influences of soil and shadows are filtered. From the crop reflectance map, providing pure spectral information on the crop condition, spectral indicators can then be derived delivering a crop status map. The crop status map can on its turn be used to steer precision farming activities.

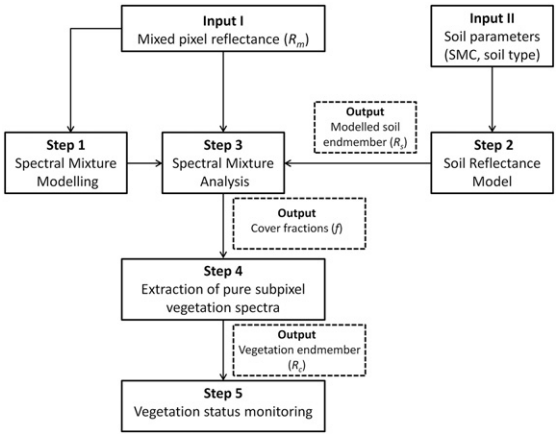

Fig. 2. Schematic overview of the proposed SMMA approach to extract the pure vegetation spectrum from tree–soil mixtures for an improved monitoring of the vegetation status.

The next section describes in detail the theoretical background of the proposed data assimilation approach, referred to as Soil Modeling Mixture Analysis (SMMA).

2.2. Theoretical background

The following paragraphs discuss each step of the SMMA protocol in detail.

Step 1 Spectral Mixture Modeling — A mixed satellite signal (R_m) is traditionally modeled as a combination of pure spectral signatures of its constituent components (i.e., endmembers) weighted by their subpixel fractional cover (Adams et al., 1993). In vegetated areas most of the photons will scatter on multiple components, e.g. tree(s) and soil, before being measured by the satellite sensor resulting in a nonlinear mixing process (Borel & Gerstl, 1994; Somers et al., 2009a). Borel and Gerstl (1994) were one of the first to incorporate these multiple scattering events occurring in plant-soil mixtures into the SMA model. Using a radiosity method, they illustrated that by including additional endmembers, each accounting for a characteristic interaction among ground objects, nonlinear spectral mixing could be accurately and realistically modeled. Somers et al. (2009a) optimized this nonlinear spectral mixing model for the citrus orchards of the study area. The obtained nonlinear mixture model described in Eq. (1) is implemented and used as the basis of the SMMA approach:

$$R_m = R_c \cdot f_c + R_s \cdot f_s + R_{sh} \cdot f_{sh} + R_c \cdot R_s \cdot f_{in} + \varepsilon \quad (1)$$

where R_c, R_s and R_{sh} are the reflectance spectra (R), and f_c, f_s, and f_{sh} the subpixel contribution factors (f) of the crop, soil, and shadow endmembers, respectively. f_{in} accounts for the amount of spectral interaction between the vegetation and soil component. The portion of the spectrum that remains unexplained is the residual term ε. As a simplification, shadows were considered as uniformly black so that R_{sh} can be modeled as photometric shade (i.e., 0% reflectance over all the wavelengths; Roberts et al., 1998).

Step 2 Soil reflectance modeling — The major part of variation in soil reflectance can be assigned to changes in soil type (texture class) and soil moisture content (SMC) (Müller & Décamps, 2001; Somers et al., 2010). Over the past decades a number of quantitative soil moisture reflectance models have been developed (e.g., Liu et al., 2002; Lobell & Asner, 2002; Somers et al., 2010). In 2002, Lobell and Asner (2002) tried to explain the change in soil reflectance with varying SMC using physically interpretable parameters. Based on empirical relationships they suggested the following exponential model:

$$R_{SMC,\lambda} = F_\lambda \cdot R_{dry,\lambda} + (1 - f_\lambda) \cdot R_{dry,\lambda} \cdot e^{(-a\lambda \cdot SMC)} \quad \text{with} \quad f_\lambda = \frac{R_{sat,\lambda}}{R_{dry,\lambda}}. \quad (2)$$

In Eq. (2) $R_{SMC,\lambda}$ is the modeled reflectance of the soil at waveband λ for a specific SMC; $R_{dry,\lambda}$ is the dry reflectance of the soil at waveband λ; and $R_{sat,\lambda}$ is the saturated reflectance at waveband λ which corresponds to the minimal reflectance measured at that band. a_λ is an attenuation factor accounting for the rate by which the reflectance changes with increasing SMC. With spatially explicit SMC data (i.e. soil moisture content map) as an input, these models can be applied to account for the variations in soil reflectance (R_s).

Step 3 Spectral Mixture Analysis — A commonly used approach to estimate subpixel cover fraction (f) maps is Spectral Mixture Analysis (SMA; Adams et al., 1993). SMA is directly driven by the physically explicit mixture model in Eq. (1). Once the endmembers (or target components, i.e. vegetation, soil and shadow) and their spectral signatures are known, and when the number of endmembers is less than the number of spectral bands, the system of equations in (1) is over-determined and may be inverted uniquely using techniques to solve for the fractions with minimal additional error, ε, in the equations. Different factors influence the accuracy of the subpixel cover

fractions estimated by SMA, such as ambient (e.g. inaccurate atmospheric correction), instrumental (e.g. insufficient signal-to-noise ratio), and model structure related noise (e.g. nonlinear mixing) (Bateson et al., 2000). The most profound source of error in SMA, however, is the lack of ability to account for sufficient temporal and spatial spectral variability, as one spectral signature per component present in the scene is used by the traditional SMA model (Somers et al., 2011). To circumvent the endmember variability problem the current study implements the Multiple Endmember Spectral Mixture Analysis technique as proposed by Roberts et al. (1998). Rather than using a rigid set of endmembers, MESMA allows endmembers to vary on a per pixel basis, permitting multiple endmembers for each component and as such countering the fixed-endmember restriction. Spectral libraries are needed to incorporate each plausible endmember condition, and in an iterative procedure different endmember combinations, selected from the spectral libraries, are used to decompose each pixel. The model with the lowest reconstruction error between the modeled and the measured mixed signal is assigned to the pixel, using the corresponding estimated cover fractions in the next step. MESMA has been used in a wide range of applications, and despite the potential benefits, there are a number of methodological and practical limitations. The accuracy of MESMA is determined by the adequacy of the spectral library, and a trade-off exists between keeping the inclusiveness of the library and the computational efficiency (Okin et al., 2001). A broad variation in spectral signatures may further lead to more than one possible combination of the pure spectra resulting in the same mixture spectrum, causing the so called ill-posedness effects (Tits et al., in press).

Step 4 Extraction of pure subpixel vegetation spectra — When a soil reflectance model (providing R_s; Step 2) is coupled to the SMA approach (Step 3), cover fraction estimates can be provided for each of the endmembers (f_s, f_c, f_{sh}, f_{in}). Consequently, the pure crop spectrum, R_c, is the only remaining unknown in the mixture model and can easily be calculated from Eq. (1) providing for each pixel a corrected satellite signal, free from soil background and shadowing effects:

$$R_c = \frac{R_m - R_s \cdot f_s - R_{sh} \cdot f_{sh}}{f_c + R_s f_{in}} + \varepsilon. \quad (3)$$

Step 5 — Vegetation status monitoring — Vegetation indices can now be derived for the extracted crop reflectance spectrum R_c (Step 4). The indices are free from background effects and hence allow to accurately monitor biophysical and biochemical crop parameters at subpixel level.

3. Materials and methods

3.1. Synthetic hyperspectral satellite data

Synthetic hyperspectral satellite data were generated from a ray tracing experiment in a fully calibrated virtual citrus orchard (Stuckens et al., 2009). This type of data allows the simulation of the complexity of real hyperspectral spaceborne data, implicitly incorporating effects such as multiple photon scattering and shadowing and shading effects. Meanwhile, exact ground reference data is available for an objective and extensive evaluation of the algorithm (Pharr & Humphreys, 2004). In the virtual orchards, 3D geometries of trees were built using the triangular mesh algorithm described in Weber and Penn (1995). The spectral interactions with the virtual objects (leaves, branches, stems, soil) and the atmosphere were modeled realistically using bidirectional reflectance distribution function models and sky maps, respectively (Stuckens et al., 2009). An example of the virtual orchard, for different levels of detail, is given in Fig. 3. The Physically Based Ray-Tracer (PBRT) model has been integrated in the web-based RAMI Online Model Checker (ROMC) service (Widlowski et al., 2008).

Spectral reference data for citrus canopies, leaves and soils were measured using a full-range (350–2500 nm) Analytic Spectral Devices (ASD) Fieldspec JR spectroradiometer with a 25° foreoptic. The measurements were performed in the same orchard in South Africa used to calibrate the virtual orchard, as described above. Spectra of different soil types were measured (i.e. Haplic Arenosol, Cambisol, Albic Leptic Luvisol and Haplic Acrisols), over a wide range of SMC. Leaf chlorophyll and water content were derived from the measured leaf spectra through inversion of the PROSPECT model (Jacquemoud & Baret, 1990). The extracted leaf chlorophyll and water content were modified, and the new reflectance coefficients were recalculated with the PROSPECT model (Stuckens et al., 2009). Thus extra variability in the biophysical parameters and the spectral data was created. The chlorophyll and water content of the leaves, both modified and extracted from the spectral measurements are shown in Table 1.

The orchard block had a row spacing of 4 m and a tree spacing of 2 m. The row azimuth was 7.3°, and the average tree height 3 m. To test the performance of the soil reflectance model for different soil types, a Haplic Arenosol and an Albic Leptic Luvisol soil (FAO, 1998), both typical for commercial citrus orchards in the Western Cape Province in South Africa, were used in the simulations (Somers et al., 2010). The gravimetric water content of the soil surface ranged between 0% and 50%. For a detailed description of the design, modalities and application of the virtual orchard model the reader is redirected to Stuckens et al. (2009). For more details on the field campaign, see Somers et al. (2009a, 2010). The image scene consists of four orchard blocks, each with specific plant properties: one block without any stress (i.e. with the reference leaf and tree parameters as shown in Table 1), and one block for each of the modified tree and leaf parameters as described in Table 1. Each of the four blocks consists of the two soil types as described above, and each soil type is separated in dry (gravimetric SMC < 15%) and wet soils (gravimetric SMC > 15%). As a result, there are four different soil sub-blocks in each of the four orchard blocks. The scene was generated such that the variability in soil and canopy properties is more pronounced between blocks than within blocks. The spatial variability in soil and canopy properties of the generated scene is shown in Fig. 4.

The hyperspectral image scene that was generated contained 30 by 30 pixels (spectral range: 400–2400 nm, spectral resolution: 10 nm, pixel size: 30 × 30 m). The sensor specifications were set to approximate those of the currently available hyperspectral satellite sensors. The scene was rendered with a solar elevation of 79.2° and azimuth of 339.6°, corresponding to the local noon in December (the summer season in the Southern hemisphere).

For the creation of a spectral library for the tree endmember needed both in the MESMA and the SMMA approach, pure spectral signatures of trees were extracted from virtual orchard scenes, encompassing the spectral variability that is present in the different orchard blocks. As illumination conditions and orchard structure were identical for the pure tree signatures and the created hyperspectral image scene, and the variability encountered in the image is represented in the library, the signatures in the tree-library could be considered as representative for the trees in the image.

3.2. Experimental methodology

The aim of this study was to increase the accuracy of the biophysical plant parameters derived from a hyperspectral image by extracting the pure canopy signature from the mixed hyperspectral signal using a combination of a soil reflectance model and a SU model, as described in Section 2.2. The performance of this SMMA model was compared to the traditional MESMA approach, as described by Tits et al. (in press). This MESMA approach, without any prior information or in situ data, used a spectral library for the soil endmember. This library consisted

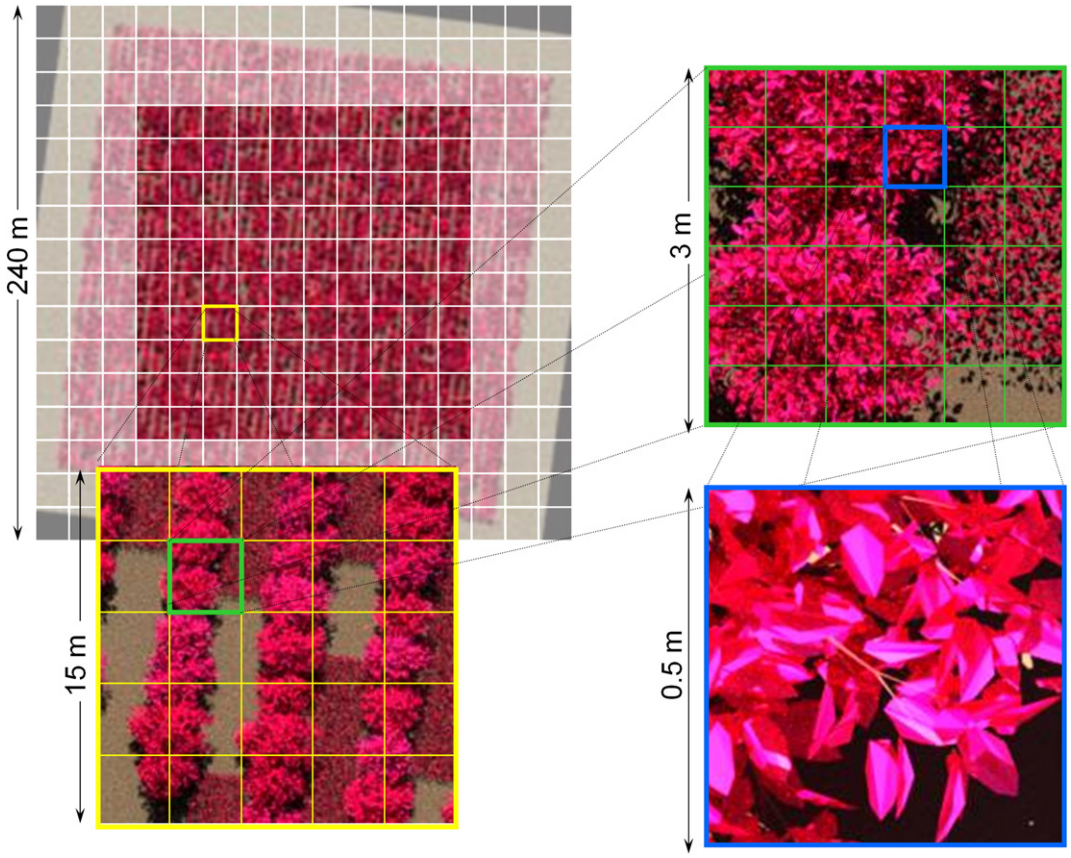

Fig. 3. Color infrared image of the virtual citrus orchard. Different levels of detail are highlighted. From Stuckens et al. (2008).

of field measurements of four different soil types encountered in the same study area as the in situ measurements, (i.e. Haplic Arenosol, Cambisol, Albic Leptic Luvisol and Haplic Acrisols) and different SMCs, encompassing the full range found in the virtual orchard.

For a proper evaluation of the SMMA model, a sensitivity analysis was performed regarding the influence of the accuracy of the modeled soil signal on the extracted vegetation signal and consequently on the derived biophysical parameters. First, the reference soil signal was used as an input in the SMMA model to demonstrate the potential of the SMMA approach. The second part of the sensitivity analysis encompassed an increasing error on the SMC to study the robustness of the SMMA approach.

For a comprehensive evaluation of the performance of the two different SU models and the sensitivity analysis of the SMMA approach, all the different steps (Fig. 2) were evaluated separately, as described below. As such, the influence of each of the separate steps on the accuracy of the model could be evaluated. The only exception was for Step 1, as explained below.

Step 1 Spectral Mixture Modeling — As described by Somers et al. (2009a), nonlinear mixing effects occur in citrus orchards, with the tree–soil interaction being the most important multiple scattering interaction (Eq. 1). The validity of this nonlinear model to account for the multiple scattering effects is tested by comparing its performance with that of the linear model (Eq. 4)

$$R_m = R_c \cdot f_c + R_s \cdot f_s + R_{sh} \cdot f_{sh} + \varepsilon \tag{4}$$

where $R_{c,}$ R_s and R_{sh} are the reflectance spectra (R), and f_c, f_s, and f_{sh} the subpixel contribution factors (f) of the crop, soil, and shadow endmembers, respectively. The portion of the spectrum that remains unexplained is the residual term ε.

In order to minimize the effect of other sources of error, both the reference vegetation and soil signal were used as inputs in the unmixing model. This allowed to evaluate the performance of the mixture model itself, as other sources of error such as endmember variability were minimized.

The performance of the linear and nonlinear mixture model was evaluated by comparing the Relative Root Mean Square Error (RRMSE, Eq. 5) per wavelength between the reference and the extracted vegetation signal. In Eq. (5), a_i denotes the actual or reference reflectance, while e_i the extracted reflectance for each (i) of the n observations. $\bar{a}$ is the average reference reflectance for a specific wavelength over all observations.

$$\mathrm{RRMSE} = \sqrt{\frac{\sum_{i=1}^{n}(a_i - e_i)^2}{n}} \times \frac{1}{a^2} \tag{5}$$

Step 2 Soil reflectance modeling — The parameters needed in the SRM (i.e. $R_{dry,\lambda}$, $R_{sat,\lambda}$ and a_λ, see Section 2.2, Step 2) for the modeling of the spectral signature of the soil endmember were extracted from Somers et al. (2010), where a full calibration of the SRM was

Table 1
Modifications in the tree and leaf parameters to create variation in the scenarios. Mean and standard deviations of the reference values are given as well.

	Reference value	Modifications
LAI	6.3 ± 2.1	56% of reference trees
Chlorophyll (a + b) content of tree leaves (μg/cm^2)	45.01 ± 9.41	50% of reference leaves
Water content of tree leaves (g/cm^2)	0.02 ± 0.0008	70% of reference leaves

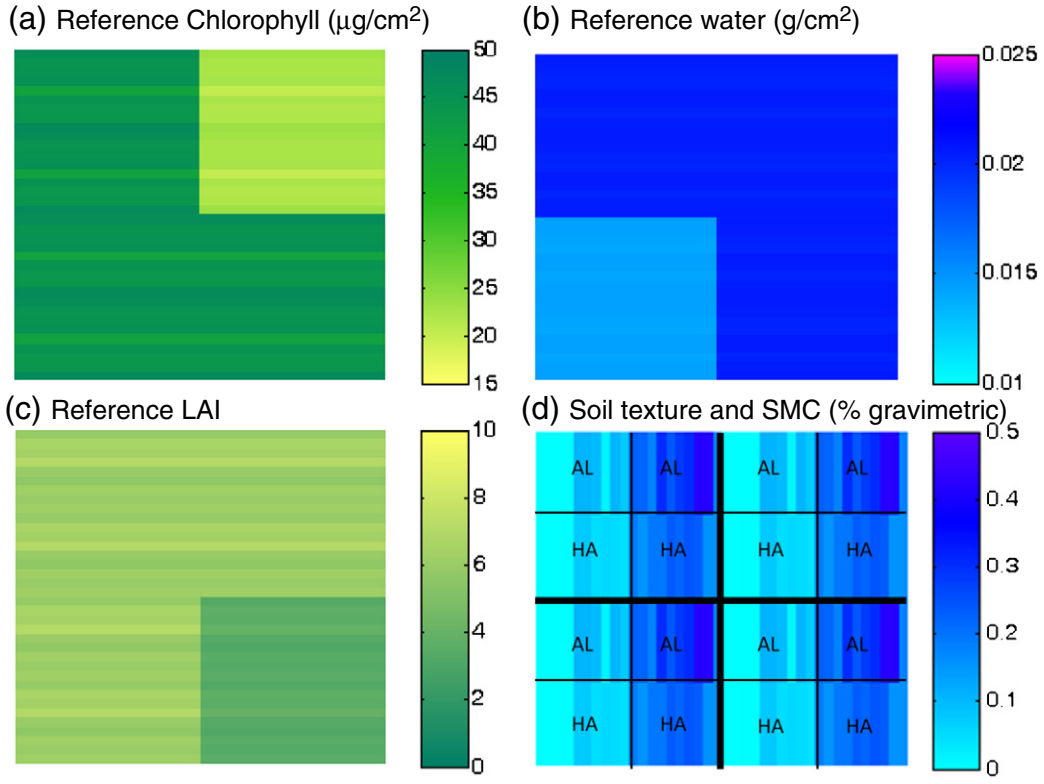

Fig. 4. The per pixel variability of the rendered scenes of the virtual orchard, shown for (a) the chlorophyll content, (b) water content, (c) LAI of the canopies, and (d) the surface gravimetric soil moisture content and the soil type, with Albic Leptic Luvisol (*AL*) and Haplic Arenosol (*HA*).

performed for different soil types, including the Haplic Arenosol and Albic Leptic Luvisol used in this study.

The performance of the soil reflectance model (Eq. 2) was evaluated using the RRMSE per wavelength between the modeled and the reference spectral signature of the soil endmember, for the two different soil types to show the error introduced by the soil reflectance model. The accuracy of the modeled soil was also compared with the soil signal selected from the spectral library with the MESMA model.

Step 3 Spectral Mixture Analysis — a nonlinear mixture model was used to describe the nature of the mixtures, with the addition of an interaction term to account for multiple scattering effects between the tree and soil endmember. As a result, the estimated f_c, f_s, f_{sh} and f_{in} in Eq. (1) are not the physical cover fractions of the different components in the pixel, but the spectral contribution of each endmember to the measured mixed signal. For the proper evaluation of the different SU models, these estimated contribution factors must therefore first be converted to the physical cover fractions. According to the methodology described in Somers et al. (2009a), the contribution factor of the interaction term f_{in} is proportionally redistributed among the contribution factors f_c and f_s of the tree and soil endmember respectively, as shown in Eq. (6) for f_c, converting from contribution factor f_c to cover fraction f'_c.

$$f'_C = f_c + f_{in}\frac{f_c}{f_c + f_s} \tag{6}$$

The resulting estimated cover fraction of the tree endmember was compared with the actual tree cover in the pixel, using the coefficient of determination (R^2), described in Eq. (7).

$$R^2 = 1 - \frac{\sum_{i=1}^{t}(a_i - e_i)^2}{\sum_{i=1}^{t}(a_i - \bar{a})^2} \tag{7}$$

With t as the number of observations and a_i and e_i the actual and estimated values for observation i, respectively. $\bar{a}$ is the average of the a_i values.

Step 4: *Extraction of pure subpixel vegetation spectra* — The performance of the different unmixing models for the extraction of the pure canopy signature was evaluated using the RRMSE (Eq. 5) per wavelength between the extracted vegetation signal and the reference pure canopy signal.

Step 5: *Vegetation status monitoring* — To estimate the three crop health parameters (i.e. chlorophyll content, leaf water content and LAI), different VIs were calculated from the pure vegetation signal extracted from the mixed signals using the methods described above, and correlated with the parameter of interest.

The GM1 (Gitelson & Merzlyak, 1996) index was used to estimate the chlorophyll content of the canopy. Although GM1 was developed for leaf spectral measurements, it was successfully tested by Zarco-Tejada et al. (2004) on tree canopies in Olive orchards and by Delalieux et al. (2008) in Citrus and Peach orchards.

$$\mathrm{GM1} = \mathrm{R}_{750}/\mathrm{R}_{550} \tag{8}$$

To estimate the water content of the canopies, the Maximum Difference Water Index (MDWI) (Eitel et al., 2006) was used, a water index developed to estimate the water content at canopy level. It has recently been tested successfully by Dzikiti et al. (2010) on citrus trees in the same study area described in Section 3.1.

$$\mathrm{MDWI} = (R_{\mathrm{max1500-1750}} - R_{\mathrm{min1500-1750}})/(R_{\mathrm{max1500-1750}} + R_{\mathrm{min1500-1750}}) \tag{9}$$

Most LAI-related indices, like the Normalized Difference Vegetation Index (NDVI), are susceptible not only to changes in LAI but also to changes in chlorophyll content. Moreover, NDVI is not sensitive to changes at high LAI values (i.e., the NDVI values tend to saturate at

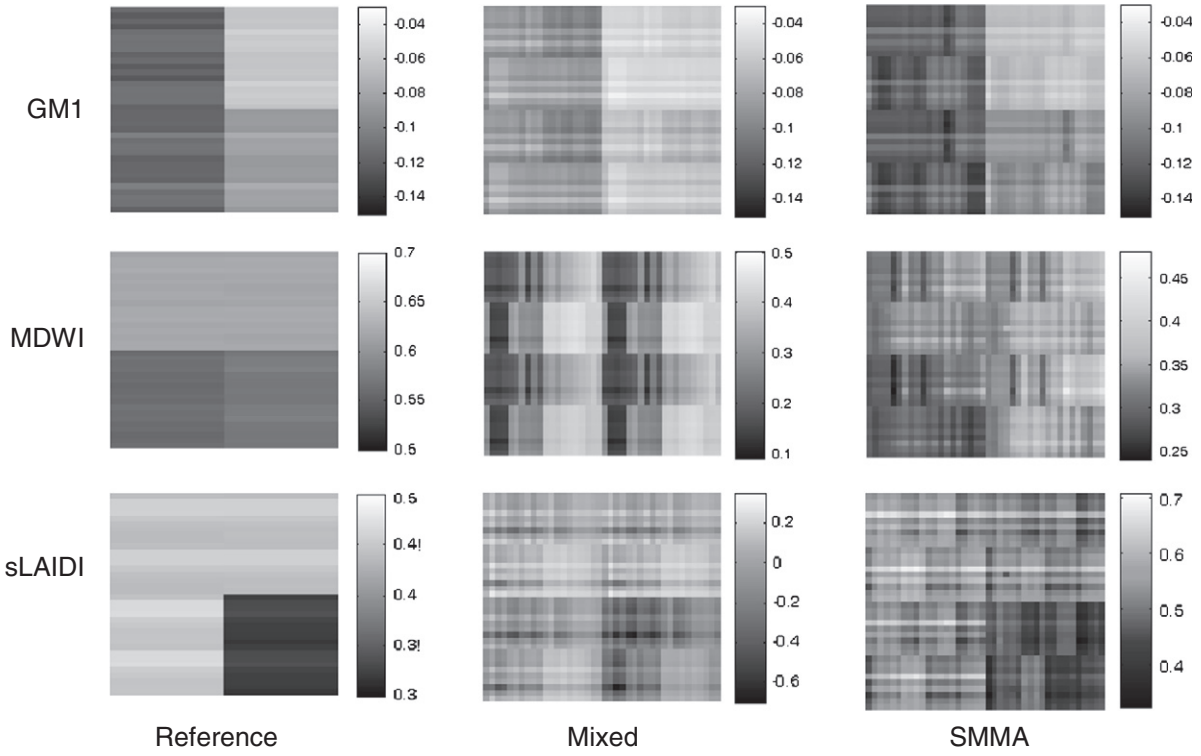

Fig. 5. VI maps of the GM1, MDWI and the sLAIDI, calculated with the pure canopy signals (*Reference*), the mixed pixel signals (*Mixed*), and with the vegetation signals extracted with the SMMA model (*SMMA*).

high LAI and chlorophyll values) (Delalieux et al., 2008; Gao, 1996). To circumvent this problem, Delalieux et al. (2008) proposed the standardized LAI Determining Index (sLAIDI):

$$\text{sLAIDI} = S \cdot \left(\frac{R_{1050} - R_{1250}}{R_{1050} + R_{1250}}\right) \cdot R_{1555} \tag{10}$$

s is a scaling factor to rescale the index between zero and one, and is equal to 40. The index was successfully tested to estimate LAI of 'Midknight Valencia' Citrus (*Citrus sinensis* L.) trees, in the same study area as described in Section 3.1. Results clearly demonstrated that sLAIDI was both independent of chlorophyll content and did not tend to saturate, even at LAI values of 10.

The ability of the extracted vegetation signal to predict the biophysical properties of the vegetation was tested by calculating the R^2 (Eq. 5) between the VI derived from the extracted spectral vegetation signature and the biophysical parameter.

4. Results

The relevance of SMMA as a critical part of the hyperspectral processing chain is highlighted in Fig. 5, showing the VI maps calculated from the reference pure vegetation spectra (*Reference*, left), the mixed pixel spectra (*Mixed*, middle), and the pure vegetation spectra extracted from the SMMA algorithm (*SMMA*, right).

The maps in Fig. 5 clearly illustrate an increased resemblance between the SMMA extracted VI values and the reference values, opposed to the mixed VI values for both GM1 and MDWI. For both the GM1 and the MDWI map, however, the influence of the LAI on both VIs is clearly visible, while sLAIDI is unaffected by either the water or the chlorophyll content of the canopy.

In the following paragraphs, a more detailed evaluation for each of the different steps of the SMMA model is given.

Step 1 Spectral Mixture Modeling — The performance of the linear (Eq. 4) and the nonlinear model (Eq. 1) to extract the pure canopy signature using both the reference vegetation signal and the reference soil signal as inputs, is shown in Fig. 6. The nonlinear mixture model clearly outperforms the linear model with an average RRMSE of 0.05 compared to 0.74 respectively. The RRMSE of the vegetation signal extracted with the nonlinear model is generally very low, i.e. between 0.0015 and 0.03, except for the blue bands between 400 and 520 nm and the edges of the SWIR2 region (1950–2400 nm) where the RRMSE can increase to 0.26.

Step 2 Soil reflectance modeling — The accuracy of the modeled soil signals is shown in Fig. 7(a), for the two different soil types, i.e. Haplic Arenosol (*Arenosol*) and Albic Leptic Luvisol (*Luvisol*). An additional subdivision was made between wet and dry soil, with a SMC of 15% as boundary. Overall, the Arenosols had the lowest modeling error, with a mean RRMSE of 0.14 compared to 0.25 for the Luvisols. In addition, the dry soils were better modeled compared to the wet soils, especially in the SWIR region of the spectrum.

The accuracy of the spectral signatures of the soil endmember used in the two different SU models described in Section 3.2 is shown in Fig. 7(b). The soil signature modeled in the SMMA approach clearly has a higher accuracy than the soil signal selected from the spectral library with the MESMA model, with a mean RRMSE value of 0.18 (*SMMA*) compared to 0.58 (*MESMA*).

The influence of erroneous SMC values on the modeled soil signal is shown in Fig. 8. The error induced on the SMC ranged between 0 and 100%, and an increasing RRMSE could be observed with an increasing error in the SMC. The largest errors could be observed in the VIS and SWIR2 region, although only a small increase in RRMSE of the modeled soil signal could be observed for an error of 25%, with an average increase in RRMSE of 0.03. Larger errors on the SMC, however, resulted in relatively larger errors on the modeled soil signal.

Step 3: *Spectral Mixture Analysis* — The results of the subpixel cover fractions estimated with the SMMA approach for the different soil types, are given in Table 2, together with the overall accuracies for SMMA and MESMA. The cover fractions were best extracted with

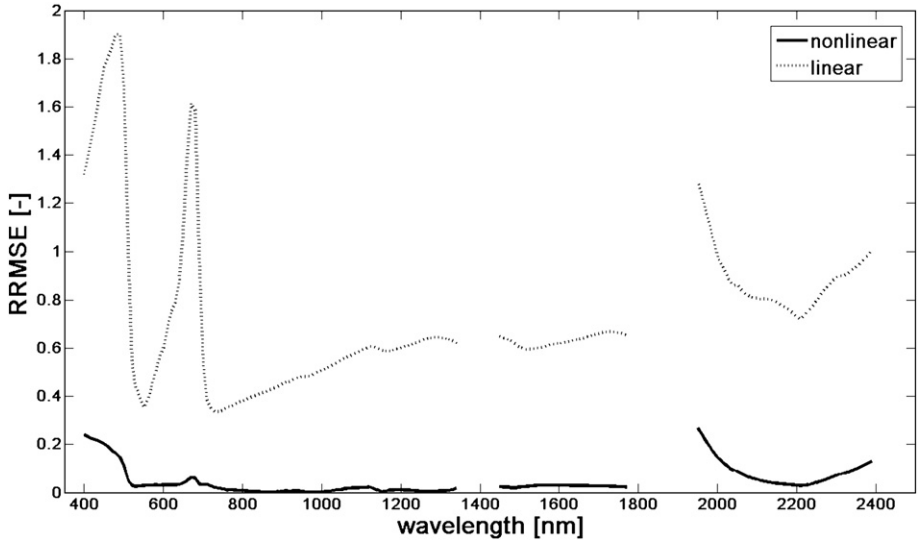

Fig. 6. RRMSE per wavelength between the extracted and the reference canopy signature using both a linear and a nonlinear model, with the reference vegetation and soil signal as inputs.

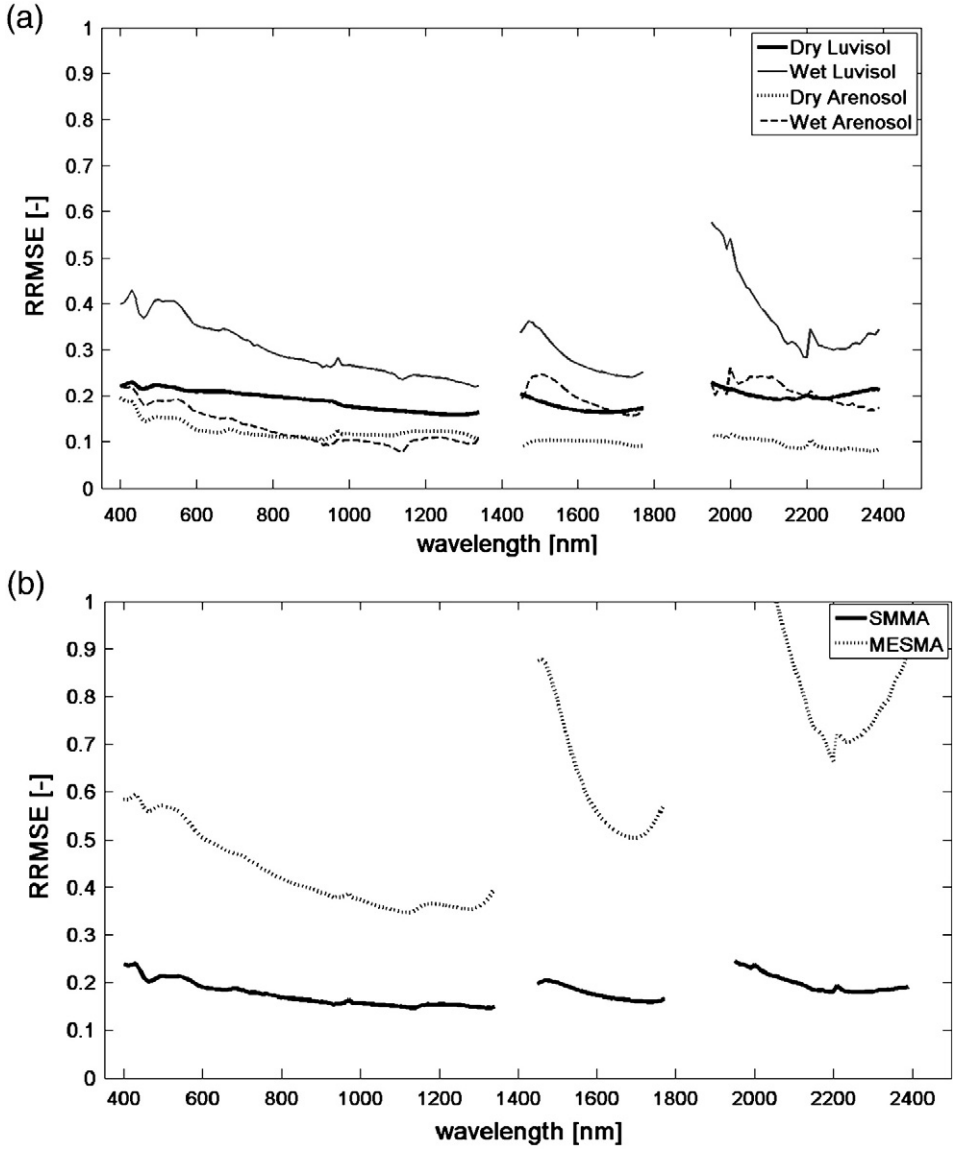

Fig. 7. Accuracy of the soil signals used as inputs in the different SU models, with (a) the modeled soil signals (SMMA), shown per soil type, with an additional separation between dry (SMC<15%) and wet (SMC>15%) soils, and (b) the overall error on the modeled soil signals (*SMMA*) and the soil signals selected from the spectral library with the MESMA model (*MESMA*).

the SMMA model for the orchard scenes with an Albic Leptic Luvisol soil, while for the orchard scenes with Haplic Arenosol soils SMMA had a lower accuracy. The overall accuracy of SMMA, on the other hand, was lower than for the MESMA model, with a $\Delta R^2 = 0.07$.

The influence of erroneous SMC values on the estimation of the cover fractions is shown in Fig. 9. The accuracy of the SU model with the reference soil signature per pixel as input is also shown, with an R^2 value of 0.70. When using the modeled soil signature as an input in the SMMA model instead the R^2 decreased to 0.43, which slowly decreased further to $R^2 = 0.32$ for an error on the SMC of 40%, after which the R^2 dropped to 0.19 for an error on the SMC of 50%. For errors on the SMC larger than 50%, the R^2 remained around 0.13.

Step 4: *Extraction of pure subpixel vegetation spectra* — the accuracy of the canopy signal extracted with the SMMA model, for the two different soil types, is given in Fig. 10(a). Overall, the accuracies for the two soil types are very similar, with an exception in the VIS and the SWIR2. For example, for the green peak at 550 nm, the dry and wet Luvisols have a RRMSE of 0.17 and 0.14 respectively, while the dry Arenosols have a RRMSE of respectively 0.20 and 0.15.

The overall accuracy of the SMMA model and the MESMA model for the extraction of the pure canopy signature is shown in Fig. 10(b). Similar results were obtained in the VIS region of the spectrum, with only a small difference in RRMSE of 0.06 at 550 nm. In the NIR and the SWIR region, however, SMMA outperforms the MESMA model, with an average difference in RRMSE of 0.10 and 0.11 in the NIR and SWIR1 respectively, and a maximum difference of 0.12 in the SWIR2.

The influence of erroneous SMC values on the extraction of the pure canopy signal, is shown in Fig. 11. The accuracy of the extracted canopy signal using the reference soil signal as an input is shown as well, with RRMSE values around 0.08 in the green bands and the NIR. Using the soil reflectance model, however, the RRMSE at 550 nm increased from 0.08 to 0.17, while no significant increase was observed for the NIR. In the SWIR1 region even a decrease in RRMSE could be observed, e.g. from 0.11 to 0.08 at 1550 nm.

Increasing the error on the SMC highlighted these errors on the extracted canopy signals, with an increase in RRMSE in the VIS and the edges of the SWIR2 region. In the NIR and SWIR1, however, accuracies remained relatively constant.

Step 5: *Vegetation status monitoring* — The pure crop spectra extracted in Step 4 (previous paragraph) were used to derive the LAI, chlorophyll and water content of the canopy, using the VIs described in Section 3.2. In Table 3 the R^2 is given between the calculated VIs and the parameter of interest. VIs were calculated on the reference pure canopy signals, the mixed pixel signals, and the canopy signals as extracted by the SMMA and the MESMA approach.

Table 2
The coefficient of determination (R^2) for the estimated cover fractions with the SMMA approach for the two soil types, with an additional separation between dry (SMC<15%) and wet (SMC>15%) soils. The overall accuracy of the cover fractions estimated with the SMMA model (*SMMA*) is also given, together with the accuracy of the MESMA model (*MESMA*).

	Dry Luvisol	Wet Luvisol	Dry Arenosol	Wet Arenosol	SMMA	MESMA
R^2	0.57	0.44	0.44	0.41	0.43	0.50

For the VIs extracted from the reference signatures of the canopies, high accuracies were obtained for the three VIs. These accuracies dropped when the mixed pixel signals were used for VI calculation, with even no correlation between the water content of the trees and the MDWI. The accuracies could be improved when working with the canopy signals extracted with one of the SU models.

For the SMMA approach, results are given for the different soil types separately. For the pixels containing dry Luvisols, high accuracies compared to the other soil types were obtained for GM1, while accuracies were low for the MDWI and mediocre for sLAIDI. The opposite was true for the wet Arenosols. The dry Arenosols, on the other hand, had a relatively high accuracy for all three VIs, while the wet Luvisols had a low accuracy for all VIs.

Yet, the SMMA approach showed overall higher accuracies (*SMMA* in Table 3) compared to the VIs calculated from the mixed signals, with an increase in R^2 of 0.19, 0.38 and 0.14 for GM1, MDWI and sLAIDI respectively. These results clearly demonstrate the added value of the SMMA approach for the monitoring of citrus orchards with hyperspectral imagery. Although MESMA obtained similar results for GM1, the accuracies were lower for MDWI and sLAIDI compared to the SMMA approach. The accuracy of the MESMA approach for sLAIDI was even lower than the accuracies obtained with the mixed pixel signals.

The influence of erroneous SMC values on the correlation of the VIs with the biophysical parameters is shown in Fig. 12. The correlation with the VIs calculated from the extracted vegetation using the reference soil signal as an input in the SU model is shown as well (*Reference soil*). The GM1 VI was very sensitive to the accuracy of the soil signal used as input in the SMMA, as the R^2 decreased by 0.28 when using the modeled soil signal instead of the reference soil, while the R^2 decreased further with even small errors on the SMC. No correlation remained ($R^2 = 0.02$) for an error of 40% on the SMC.

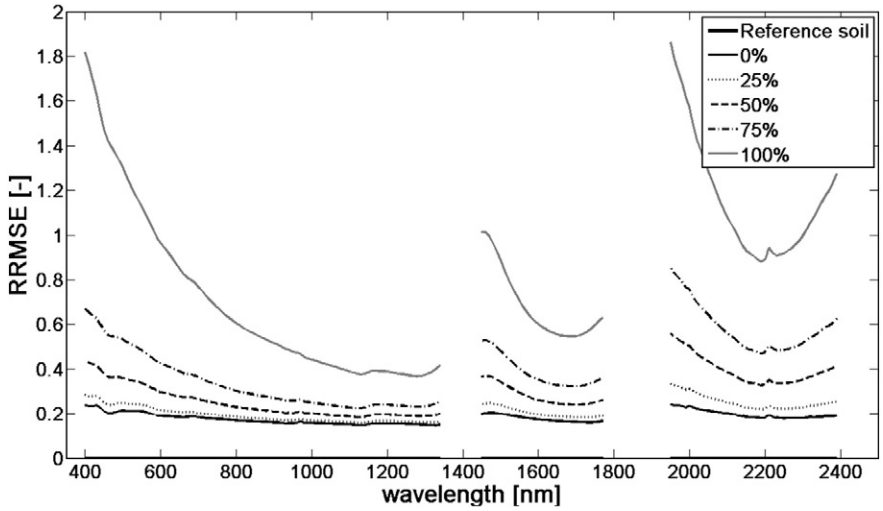

Fig. 8. Sensitivity analysis on the accuracy of the SMC on the modeled soil signal. The error in terms of percentage on the SMC is increased from 0 (*0%*) to 100 (*100%*).

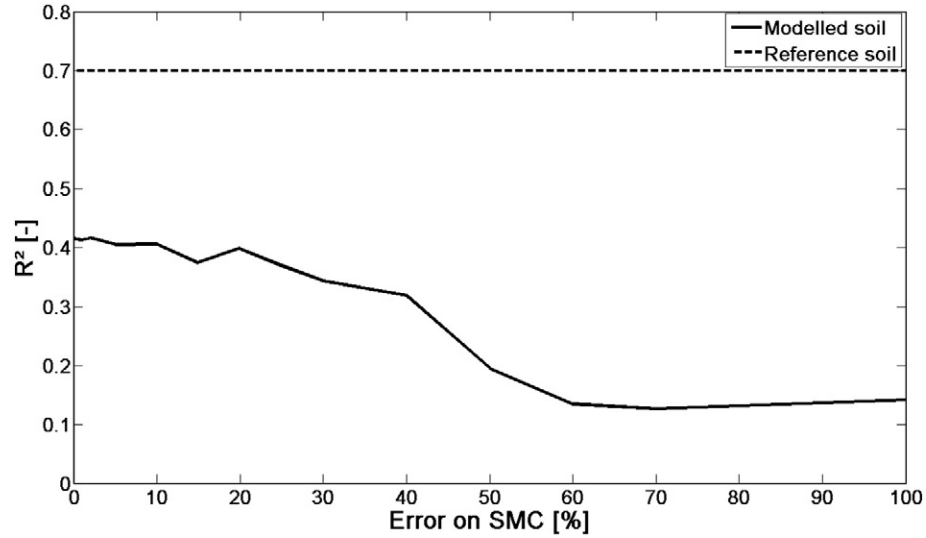

Fig. 9. Sensitivity analysis on the accuracy of the SMC on the estimated cover fractions (*Modeled soil*). The accuracy of the estimated cover fractions using the reference soil signal is shown as well (*Reference soil*).

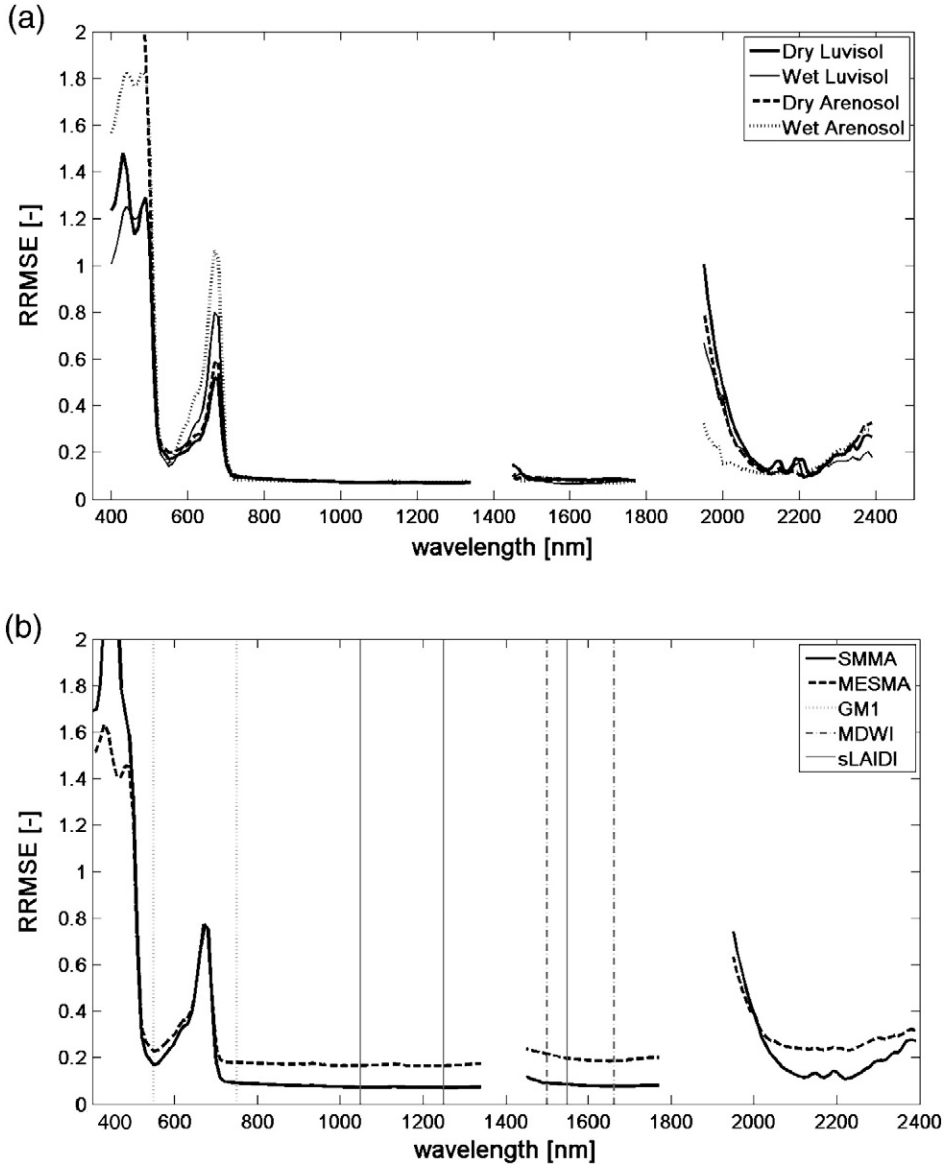

Fig. 10. Accuracy of the extracted canopy signals from (a) the SMMA model per soil type, with an additional separation between dry (SMC<15%) and wet (SMC>15%) soils, and (b) the two different SU models, i.e. the SMMA and the MESMA model. The wavebands used to calculate the different VIs are highlighted with gray vertical lines (*GM1*, *MDWI*, *sLAIDI*).

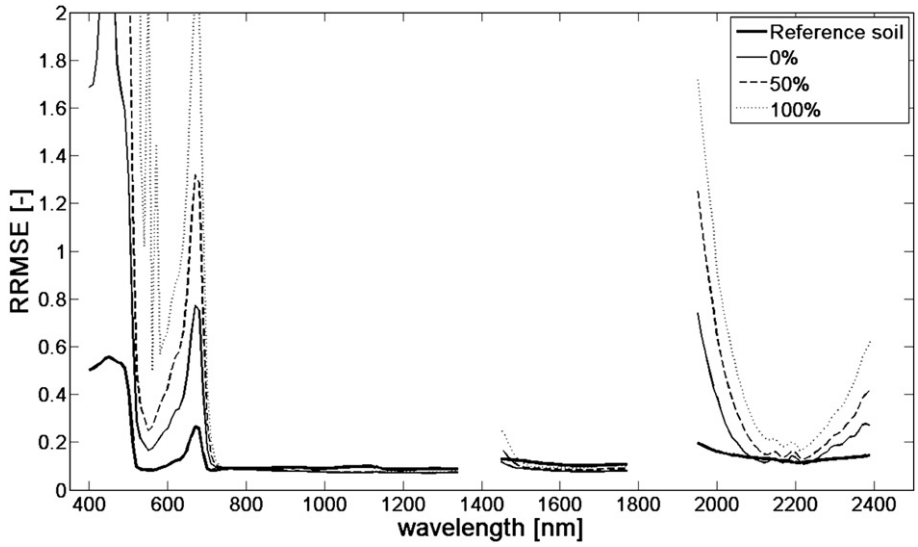

Fig. 11. Sensitivity analysis on the accuracy of the SMC on the extracted pure canopy signal. The error in terms of percentage on the SMC is increased from zero (*0%*) to 100 (*100%*). The accuracy of the extracted canopy signal using the reference soil signal is shown as well (*Reference soil*).

MDWI and sLAIDI were more robust for errors on the modeled soil signal, as the decrease in R^2 was limited to 0.16 and 0.08 respectively, while the decrease in correlations due to increasing errors on the SMC followed similar trends for both VI. For errors up to 30%, only a slight decrease in R^2 was observed.

5. Discussion

5.1. Signal Unmixing

The added value of applying a SU model for extracting biophysical parameters from hyperspectral imagery is clearly demonstrated in Fig. 5. The maps of the VI calculated with the SMMA model show the spatial variability of the biophysical parameters more pronounced than the maps based on the mixed pixel signals. This is also confirmed by the correlations shown in Table 3. The mixed pixel signals significantly decreased the correlation with the parameters of interest, with even no correlation between MDWI and the water content of the canopies. By comparing Figs. 4 and 5, it is clear that the values of MDWI calculated from the mixed pixel signals are mainly determined by the water content of the soils, as MDWI follows the same pattern as the SMC in Fig. 4(d). The SMMA approach could mitigate the influence of the SMC on the MDWI, although some interference still occurs. This is also demonstrated by comparing the modeling error of the soil endmember in Fig. 7(a) and the correlation of the MDWI in Table 3. The soils with the lowest modeling errors are the Arenosols, and the influence of SMC is thus better mitigated on Arenosols, resulting in the highest correlation for the MDWI in Table 3. These results show the added value of a spectral unmixing model to reduce the effects of the background components on the spectral signature of the crops, resulting in an increased accuracy of the estimated biophysical parameters.

It must be noted that the different VIs used in this study to provide an estimation of the biophysical parameters, may not be the most optimal VI available. In Fig. 5, the influence of the LAI on the GM1 and the MDWI is clearly visible, hampering the accurate extraction of chlorophyll and water content. The focus of this research, however, was the extraction of the crop spectral signature from a mixed pixel, and how this signal can be used to estimate the parameters of interest with an increased accuracy. How this step is performed, either with VI (and which VI are optimal) or with more model based approaches such as radiative transfer model inversion, is a topic on which more research is needed.

In the following paragraphs, a more comprehensive evaluation is given of the two different SU models, as well as a more in-depth discussion of the performance of the SMMA model.

5.2. SMMA vs. MESMA

Although both SU models improve the accuracies of the estimated biophysical parameters, SMMA outperforms MESMA for nearly all the different steps in the SU approach shown in Fig. 2.

While MESMA iteratively selects a soil signal from a spectral library, SMMA models the reflectance of the soil endmember using the SMC and soil type as input. The mean RRMSE of the modeled soil signal was 0.18 compared to 0.58 for the soil endmember selected with MESMA. A more accurate representation of the soil endmember was thus used as an input in Eq. (3) for SMMA, resulting in a better extraction of the pure canopy signal, with a mean ΔRRMSE of 0.09. The high RRMSE values in Fig. 10 in the blue and red bands, and at the edges of the SWIR2 region, are due to the very low reflectance values of the canopy signals at these wavebands, and small deviations in reflectance will result in large relative errors.

Table 3
R^2 between the parameter of interest (i.e. chlorophyll content, water content and LAI) and the VI calculated based on either the pure crop signature (*Ground truth*), the mixed pixel signals (*Mixed*) or the vegetation signal extracted with the SMMA approach for the two soil types, with an additional separation between dry (SMC<15%) and wet (SMC>15%) soils. The overall correlation estimated with the SMMA model (*SMMA*) is also given, together with the accuracy of the MESMA model (*MESMA*).

	Ground truth	Mixed	Dry Luvisol	Wet Luvisol	Dry Arenosol	Wet Arenosol	SMMA	MESMA
Chlorophyll ~ GM1	0.82	0.34	0.61	0.54	0.73	0.44	0.54	0.51
Water ~ MDWI	0.63	0.01	0.33	0.42	0.50	0.68	0.39	0.22
LAI ~ sLAIDI	0.95	0.35	0.52	0.46	0.55	0.60	0.49	0.27

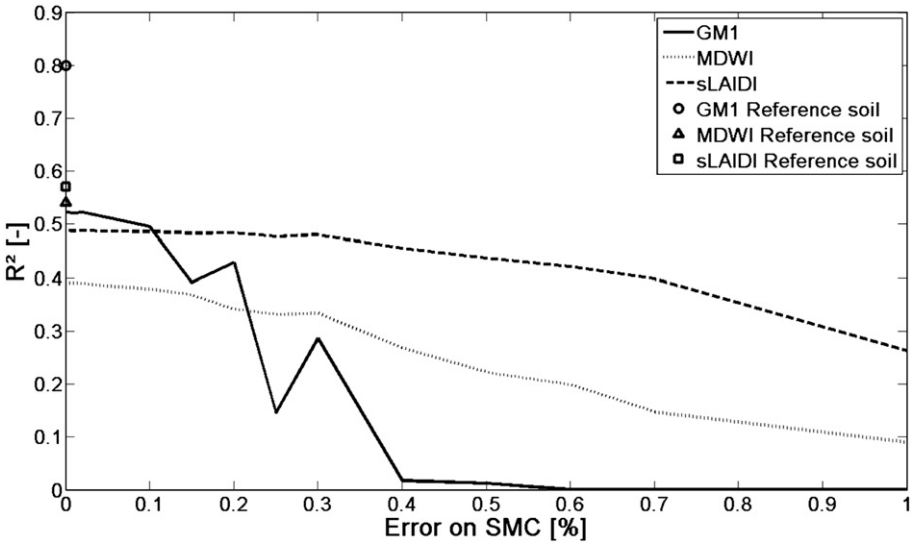

Fig. 12. Sensitivity analysis on the accuracy of the SMC on the correlation between the parameter of interest (i.e. chlorophyll content, water content and LAI) and the VI calculated from the pure canopy signal extracted with the SMMA model. The error in terms of percentage on the SMC is increased from zero (*0%*) to 100 (*100%*). Correlations for the canopy signals extracted using the reference soil signal are shown as well (*Reference soil*).

The more accurate extraction of the pure canopy signal by SMMA resulted in a more accurate estimation of the biophysical parameters, as shown in Table 3. In Fig. 10(b), the accuracy of the extracted canopy signals is shown, with the vertical lines depicting the wavebands used for the calculation of the three VIs. For GM1, the difference between the two SU models was small ($\Delta R^2 = 0.03$), as the accuracy of the extracted canopy signal at 550 nm, used for the computation of GM1 (Eq. 8) was similar for the two SU models (ΔRRMSE = 0.06). For both MDWI and sLAIDI, SMMA outperformed MESMA, with a $\Delta R^2 = 0.17$ and $\Delta R^2 = 0.22$ respectively, which is consistent with the lower error on the extracted canopy signals in the used wavebands (Fig. 10(b)).

One of the factors influencing the performance of the MESMA model, is the ill-posedness effect, i.e. a broad variation in endmembers in the spectral libraries may lead to more than one possible combination of the pure spectra resulting in the same mixture spectrum (Tits et al., in press). The endmember combination with the lowest modeling error is thus not necessarily the best representation of the components present in the mixed pixel. By using a soil reflectance model to represent the soil endmember this ill-posedness effect is significantly reduced. The results obtained in this study are in line with the conclusions of Tits et al. (in press), who demonstrated better accuracies for the MESMA model with increasing information on the soil endmember. The SMMA approach thus provides a means, based on in situ measurements, to use the information of the soil endmember (i.e. SMC and soil type) to reduce the ill-posedness effects hampering the MESMA model.

An additional advantage of providing an estimation of the soil endmember is the reduction of the number of iterations that are needed in the unmixing process. When working with a library for the soil endmember, more iterations are needed to evaluate different tree–soil endmember combinations. By providing a soil endmember estimation, a better accuracy can be obtained with less iterations, as was demonstrated by Tits et al. (in press). As a result, SMMA will not only increase the accuracy of the SU model, but will also reduce its computation time.

It must be noted, however, that although a better representation of the soil endmember resulted in a better extraction of the pure canopy signal, the same trend could not be observed for the estimated cover fractions in step 3, as shown in Table 2. MESMA obtained better results, with a $\Delta R^2 = 0.08$ compared to SMMA. The same trend was observed by Tits et al. (in press), where an improvement in the extraction of the canopy signal did not result in a better estimation of the cover fractions. However, the focus of the proposed SMMA approach is on the extraction of the pure canopy signal, and is not preferable for an accurate extraction of the cover fractions of the different endmembers. More accurate and easier to implement alternatives are available to take endmember variability into account in an unmixing model for the estimation of the cover fractions, of which an overview is given in Somers et al. (2011).

5.3. Soil reflectance modeling

The reflectance of the soil endmember was modeled using the soil reflectance model (Eq. 2) based on the soil texture and the SMC. Two different soil types were tested (i.e. Haplic Arenosol and an Albic Leptic Luvisol), and an additional separation was made between dry and wet soils.

The Arenosols could be modeled best, with a mean RRMSE of 0.14 compared to 0.25 for the Luvisols, which is in line with the findings of Somers et al. (2010), where the soil reflectance model was tested for six different soil types.

The SMC also had an influence on the accuracy of the modeled soil signal, with higher modeling accuracies for the dry soils. Some nuance is needed, however, as the reflectance values of dry soils are higher, and the same absolute error will result in smaller relative errors compared to wet soils. For a more complete comparison, the absolute error (RMSE) per wavelength is given in Fig. 13. Again, Arenosols outperform the Luvisols. The dry soils, however, have a larger absolute modeling error compared to the wet soils, while having a lower relative error. This needs to be taken into account in the further discussion, as the spectral unmixing algorithm used in step 3 (Fig. 2) is based on a Least-Squares optimization, and absolute errors might be more important.

Despite the differences in modeling error between the different soil types, similar accuracies were observed for the extracted canopy signals (Fig. 10) in the NIR and SWIR1. The indifference of the NIR regarding the accuracy of the modeled soil signal could be due to the uneven distribution of the differences in reflectance between the VIS and the NIR/SWIR to the SMA output. The SMA algorithm (step 3, Fig. 2) uses a Least-Squares estimate between modeled and measured mixed pixel, and this value is dominantly determined by the NIR and SWIR in vegetative systems (Somers et al., 2009b; Tits et al., in press). Small errors in the modeled soil signal will thus result

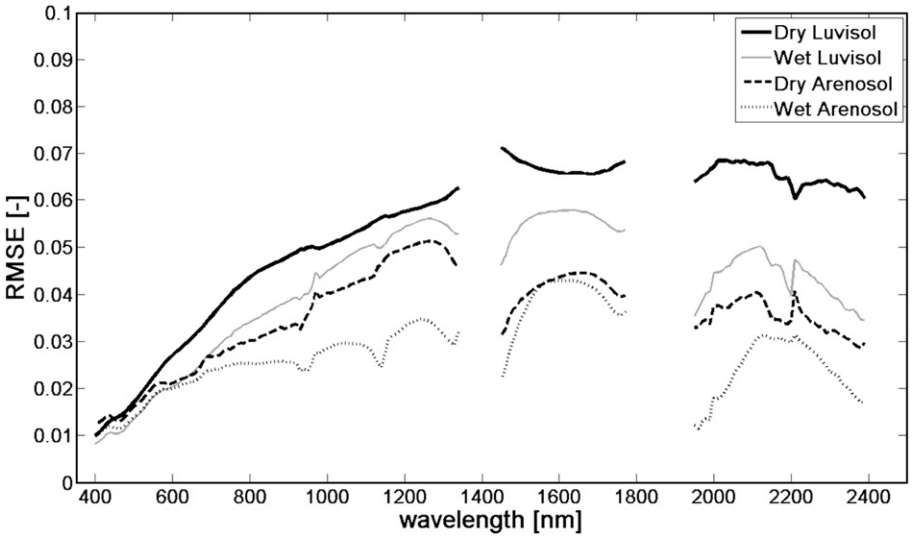

Fig. 13. Accuracy of the modeled soil signals, shown per soil type, with an additional separation between dry (SMC<15%) and wet (SMC>15%) soils.

in relatively larger errors in the VIS domain for the extracted vegetation signal. If the accuracy of the extracted canopy signal in the VIS is of importance, one must take this uneven contribution into account. Tits et al. (in press) proposed a segmentation technique where the three spectral regions (i.e. VIS, NIR, SWIR) are unmixed separately, as such removing the dominance of the NIR region in the Least-Squares optimization, and increasing the accuracy of the extracted canopy signal in the VIS. Other methods may include a weighted unmixing approach (Somers et al., 2009b) or spectral normalization (Wu, 2004).

The same trend could be observed in the sensitivity analysis on the accuracy of the SMC. In the sensitivity analysis, an increasing error on the SMC increased the error of the modeled soil signal (Fig. 8), which on its turn decreased the accuracy of the extracted canopy signal in the VIS and SWIR2, but not in the NIR region of the spectrum (Fig. 11). This indifference of the NIR and SWIR1 on the accuracy of the SMC resulted in stable results for MDWI and sLAIDI, as these VIs only use wavebands from these regions (Fig. 12). GM1, on the other hand, which uses the 550 nm waveband, was very sensitive to the accuracy of the SMC.

However, even if the soil signal is exactly known, the vegetation signal cannot perfectly be extracted from the mixed pixels. In Fig. 6, the RRMSE of the extracted vegetation signal is shown if both the reference vegetation and soil signal are used as input signals in the SMA model. This may be due to the residual term ε in Eq. (1). This residual term is the proportion of the mixed signal that cannot be modeled, and includes factors such as nonlinear interactions that have not been accounted for (Somers et al., 2009a), adjacency effects (Richter et al., 2006), and shadow that does not have a zero percent reflectance (Fitzgerald et al., 2005). To increase the potential accuracy of the SMMA approach, more research is needed on these topics.

In Fig. 11, however, the RRMSE of the extracted vegetation signal is shown when only the correct soil signal is used as an input, together with a library for the tree endmember. The resulting increase in RRMSE between Figs. 6 and 11 is due to the iterative methodology of the approach. As the spectral library for the tree endmember does not contain all possible spectral signatures, but rather encompasses the range that may be encountered in the image, small errors in the estimated cover fractions will occur, which will be further translated into the extracted vegetation signal. More research is needed on the creation of libraries which are large enough to include the spectral variability needed for an accurate unmixing, while still being manageable regarding computation time (Okin et al., 2001).

6. Conclusions and future perspectives

This study focused on the composite spectral nature of image pixels and the adverse consequences for the site-specific monitoring of agricultural attributes in citrus orchards. The key challenge was to develop a methodology to eliminate or minimize undesired background effects in vegetation indices. A Signal Unmixing (SU) technique was presented and successfully tested on simulated hyperspectral satellite imagery. The combination of in situ measured soil moisture content (SMC) data and remotely sensed hyperspectral data in an adapted Multiple Endmember Spectral Mixture Analysis (MESMA) technique provided the opportunity to extract the pure and complete hyperspectral vegetation spectrum from mixed image pixels. The extracted vegetation spectrum was implemented to reduce background contamination in vegetation index calculations. This new approach of SU opens opportunities for accurate crop status monitoring at the subpixel level. The combination of in situ data and remotely sensed hyperspectral data could as such address a critical missing link in the hyperspectral processing chain for precision farming applications, which confirms the research hypothesis. The SU technique is a good example of the added value of in situ (i.e., SMC) and hyperspectral data fusion leading to an increased accuracy in extracting biophysical and structural crop characteristics from hyperspectral satellite imagery.

It should be highlighted, however, that this work should be considered as a proof of concept study. Although the potential of SU has been demonstrated, a number of potential improvements and residual challenges have been revealed during this research and should be addressed prior to the operational implementation of the approach.

Next steps towards the operational implementation of SU include testing its effectiveness to account for multiple soil types and the co-registration of the different datasets. Apart from sandy soils, clayey and silty soils should be included in the analysis, as to test the transferability of the method for different plant production systems. In addition, the mutual effects of soil type, moisture content, changing view and illumination conditions and soil undulation or surface roughness effects should be addressed. Future research should continue on the work of Verhoef and Bach (2007) who were among the first to tackle this issue. Also the co-registration between the in situ collected soil status information and the remotely sensed hyperspectral image cubes needs proper attention when considering the operational implementation of the approach.

Probably the most important drawback of SU is the poor capacity to properly account for the effects of weeds. The approach allows extracting the pure vegetation signal but does not allow to discriminate between weed and crop spectra. The extracted signal actually still is a mixture of weeds and crops, and this again reduces the effectiveness of spectral indices to monitor crop related information. A detailed study on the magnitude and nature of the effect of weeds on SU is required while intensive research on the spectral properties of weeds and crops needs to be performed. Novel insights and innovative concepts/solutions are necessary. Current research is evaluating the usefulness of multi-angular and/or multi-temporal spectral information to separate both vegetation types (Somers & Asner, in press; Stuckens et al., 2010).

Acknowledgements

This work was supported by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The work of B. Somers was supported by the Belgian Science Policy Office in the frame of the Stereo II programme — Project VEGEMIX (SR/67/146).

References

Adams, J. B., Smith, M. O., & Gillespie, A. R. (1993). Imaging spectroscopy: Interpretation based on spectral mixture analysis. In C. M. Pieters, & P. Englert (Eds.), *Remote geochemical analysis: Elemental and mineralogical composition 7* (pp. 145–166). New York: Cambridge University Press.

Asner, G. P., & Heidebrecht, K. B. (1992). Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: Comparing multispectral and hyperspectral observations. *International Journal of Remote Sensing*, *23*, 3939–3958.

Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indexes for LAI and APAR assessment. *Remote Sensing of Environment*, *35*, 161–173.

Bateson, C. A., Asner, G. P., & Wessman, C. A. (2000). Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, *38*, 1083–1094.

Biliouris, D., Van der Zande, D., Verstraeten, W. W., Stuckens, J., Muys, B., Dutré, Ph, et al. (2009). RPV model parameters based on hyperspectral bidirectional reflectance measurements of *Fagus sylvatica* L. leaves. *Remote Sensing*, *1*, 92–106.

Borel, C., & Gerstl, S. (1994). Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sensing of Environment*, *47*, 403–416.

Brown, L. (1992). A survey of image registration techniques. *Computing Surveys*, *24*, 325–376.

Clevers, J. G. P. W., van der Heijden, G. W. A. M., Verzakov, S., & Schaepman, M. E. (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering and Remote Sensing*, *73*, 1141–1148.

Delalieux, S., Somers, B., Hereijgers, S., Verstraeten, W. W., Keulemans, W., & Coppin, P. (2008). A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. *Remote Sensing of Environment*, *112*, 3762–3772.

Dinguirard, M., & Slater, P. N. (1999). Calibration of space-multispectral imaging sensors: A review. *Remote Sensing of Environment*, *68*, 194–205.

Dorigo, W., Zurita-Milla, R., de Wit, A., Brazile, J., Singh, R., & Schaepman, M. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, *9*, 165–193.

Dzikiti, S., Verreynne, J. S., Stuckens, J., Strever, A., Verstraeten, W. W., Swennen, R., et al. (2010). Determining the water status of Satsuma mandarin trees [*Citrus unshiu* Marcovitch] using spectral indices and by combining hyperspectral and physiological data. *Agricultural And Forest Meteorology*, *150*, 369–379.

Eitel, J., Gessler, P., Smith, A., & Robberecht, R. (2006). Suitability of existing and novel spectral indices to remotely detect water stress in Populus spp. *Forest Ecology and Management*, *229*, 170–182.

Environmental Mapping and Analysis Program (2012). Available online: www.enmap.org. Accessed last on June 20th 2012

FAO (1998). FAO, world reference base for soil resources. *World soil resource report no 84*. Rome: Food and Agriculture Organization of the United Nations.

Fitzgerald, G. J., Pinter, P. J., Hunsaker, D. J., & Clarke, T. R. (2005). Multiple shadow fractions in spectral mixture analysis of a cotton canopy. *Remote Sensing of Environment*, *97*, 526–539.

Gao, B. (1996). NDWI — A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*, 257–266.

Gitelson, A., & Merzlyak, M. (1996). Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *Journal of Plant Physiology*, *148*, 494–500.

Hansen, P., & Schjoerring, J. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, *86*, 542–553.

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*, 295–309.

Itten, K. I., & Meyer, P. (1993). Geometric and radiometric correction of TM data of mountainous forested areas. *IEEE Transactions on Geoscience and Remote Sensing*, *31*, 764–770.

Jacquemoud, S., & Baret, F. (1990). Prospect—A model of leaf optical properties spectra. *Remote Sensing of Environment*, *34*, 75–91.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., et al. (2009). PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, *113*, S56–S66.

Jordan, C. (1969). Derivation of Leaf-Area Index from quality of light on forest floor. *Ecology*, *50*, 663–666.

Kerr, J., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends In Ecology & Evolution*, *18*, 299–305.

Liang, S., Fang, H., & Chen, M. (2001). Atmospheric correction of Landsat ETM + land surface imagery — Part I: Methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 2490–2498.

Liu, W., Baret, F., Gu, X., Tong, Q., Zheng, L., & Zhang, B. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, *81*, 238–246.

Lobell, D., & Asner, G. (2002). Moisture effects on soil reflectance. *Soil Science Society of America Journal*, *66*, 722–727.

Lobell, D., & Asner, G. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, *93*, 412–422.

MDA corporation (2012) [available online: sm.mdacorporation.com]. Accessed last on June 20th 2012.

Müller, E., & Décamps, H. (2001). Modeling soil moisture-reflectance. *Remote Sensing of Environment*, *76*, 173–180.

Okin, G. S., Roberts, D. A., Murray, B., Curkendall, D. W., & Okin, W. J. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment*, *77*, 212–225.

Ouillon, S., Douillet, P., & Andrefouet, S. (2004). Coupling satellite data with in situ measurements and numerical modeling to study fine suspended-sediment transport: A study for the lagoon of New Caledonia. *Coral Reefs*, *23*, 109–122.

Peddle, D. R., & Smith, A. M. (2005). Spectral mixture analysis of agricultural crops: Endmember validation and biophysical estimation in potato plots. *International Journal of Remote Sensing*, *26*, 4959–4979.

Pharr, M., & Humphreys, G. (2004). Physically based rendering: From theory to implementation. San Francisco, USA: Morgan Kaufmann Publishers.

Pinter, P., Hatfield, J., Schepers, J., Barnes, E., Moran, M., Daughtry, C., et al. (2003). Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, *69*, 647–664.

Qi, J., Chehbouni, A., Huete, A., Kerr, Y., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*, 119–126.

Quarmby, N. A., Townshend, J. R., Settle, J. J., White, K. H., Milnes, M., Hindle, T. L., et al. (1992). Linear mixture modelling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, *13*, 415–425.

Reynolds, R., Rayner, N., Smith, T., Stokes, D., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, *15*, 1609–1625.

Richter, R., Bachmann, M., Dorigo, W., & Muller, A. (2006). Influence of the adjacency effect on reflectance measurements. *IEEE Transactions on Geoscience and Remote Sensing Letter*, *3*, 565–569.

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. O. (1998). Mapping Chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, *65*, 267–279.

Roberts, D. A., Smith, M. O., & Adams, J. B. (1993). Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, *44*, 255–269.

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*, 95–107.

Somers, B. (2009d). Hyperspectral Unmixing for Plant Production Monitoring. PhD Disertation, Arenberg Doctoral School, Katholieke Universiteit Leuven, pp. 208.

Somers, B., Asner, G. P. in press. Invasive species mapping in Hawaiian rainforests using multi-temporal Hyperion spaceborne imaging spectroscopy. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, (DOI http://dx.doi.org/10.1109/JSTARS.2012.2203796).

Somers, B., Asner, G. P., Tits, L., & Coppin, P. (2011). Endmember variability in spectral mixture analysis: A review. *Remote Sensing of Environment*, *115*, 1603–1616.

Somers, B., Cools, K., Delalieux, S., Stuckens, J., der Zande, D. V., Verstraeten, W., et al. (2009). Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards. *Remote Sensing of Environment*, *113*, 1183–1193.

Somers, B., Delalieux, S., Stuckens, J., Verstraeten, W. W., & Coppin, P. (2009). A weighted linear spectral mixture analysis approach to address endmember variability in agricultural production systems. *International Journal of Remote Sensing*, *30*, 139–147.

Somers, B., Delalieux, S., Verstraeten, W. W., & Coppin, P. (2009). A conceptual framework for the simultaneous extraction of subpixel spatial extent and spectral characteristics of crops. *Photogrammetric Engineering and Remote Sensing*, *75*, 57–68.

Somers, B., Gysels, V., Verstraeten, W. W., Delalieux, S., & Coppin, P. (2010). Modelling moisture-induced soil reflectance changes in cultivated sandy soils: A case study in citrus orchards. *European Journal of Soil Science*, *61*, 1091–1105.

Stuckens, J., Somers, B., Verstraeten, W., & Coppin, P. (2008). Simulation of physiology induced hyperspectral changes in a virtual citrus orchard as detected by medium and high resolution remote sensing platforms, Geophysical Research Abstracts. EGU General Assembly. Vienna, Austria, April 13–18, 2008 European Geosciences Union.

Stuckens, J., Somers, B., Albrigo, G. L., Dzikiti, S., Verstraeten, W. W., Swennen, R., et al. (2010). Off-nadir viewing for reducing spectral mixture issues in citrus orchards. *Photogrammetric Engineering and Remote Sensing*, *76*, 1261–1274.

Stuckens, J., Somers, B., Delalieux, S., Verstraeten, W. W., & Coppin, P. (2009). The impact of common assumptions on canopy radiative transfer simulations: A case study in citrus orchards. *Journal of Qualitative Spectroscopy & Radiative Transfer*, *110*, 1–21.

Tits, L., Somers, B., & Coppin, P. (2012). The potential and limitations of a clustering approach for the improved efficiency of Multiple Endmember Spectral Mixture Analysis in plant production system monitoring. *IEEE Transactions on Geoscience and Remote Sensing*, *50*, 2273–2286.

Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*, 127–150.

Verbeiren, S., Eerens, H., Piccard, I., Bauwens, I., & Van Orshoven, J. (2008). Sub-pixel classification of SPOT-VEGETATION time series for the assessment of regional crop areas in Belgium. *International Journal of Applied Earth Observation and Geoinformation*, *10*, 486–497.

Verhoef, W., & Bach, H. (2007). Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sensing of Environment*, *109*, 166–182.

Weber, J., & Penn, J. (1995). Creation and rendering of realistic trees. *Proceedings of the 22nd annual conference on computer graphics and interactive techniques* (pp. 119–128).

Widlowski, J. -L., Robustelli, M., Disney, M., Gastellu-Etchegorry, J. -P., Lavergne, T., Lewis, P., et al. (2008). The RAMI on-line model checker (ROMC): A web-based benchmarking facility for canopy reflectance models. *Remote Sensing of Environment*, *112*, 1144–1150.

Wu, C. (2004). Normalized spectral mixture analysis for monitoring urban composition using ETM + imagery. *Remote Sensing of Environment*, *93*, 480–492.

Zaks, D. P. M., & Kucharik, C. J. (2011). Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters*, *6*, 1–10.

Zarco-Tejada, P., Miller, J., Morales, A., Berjon, A., & Aguera, J. (2004). Hyperspectral indices and model simulation for chlorophyll estimations in open canopy tree crops. *Remote Sensing of Environment*, *90*, 463–476.