Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research 62 (2018) 1654–1674

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Simulating arbitrary hyperspectral bandsets from multispectral observations via a generic Earth Observation-Land Data Assimilation System (EO-LDAS)

M. Chernetskiy [a,*], N. Gobron [b], J. Gómez-Dans [a,c], O. Morgan [b], M. Disney [a,c], P. Lewis [a,c], C. Schmullius [d]

[a] University College London (UCL), United Kingdom
[b] European Commission Joint Research Centre, Ispra, VA, Italy
[c] NERC National Centre for Earth Observation (NCEO), United Kingdom
[d] Friedrich Schiller University, Institute of Geography – Department for Earth Observation, Jena, Germany

Received 16 February 2018; received in revised form 3 July 2018; accepted 15 July 2018
Available online 23 July 2018

Abstract

This paper presents results of using multi-sensor and multi-angular constraints in the generic Earth Observation-Land Data Assimilation System (EO-LDAS) for reproducing arbitrary bandsets of hyperspectral reflectance at the top-of-canopy (TOC) level by merging observations from multispectral sensors with different spectral characteristics. This is demonstrated by combining Multi-angle Imaging Spectroradiometer (MISR) and Landsat Enhanced Thematic Mapper Plus (ETM+) data to simulate the Compact High Resolution Imaging Spectrometer CHRIS/PROBA hyperspectral signal over an agricultural test site, in Barrax, Spain. However, the method can be more generally applied to any combination of spectral data, providing a tool for merging EO data to any arbitrary hyperspectral bandset.

Comparisons are presented using both synthetic and observed MISR and Landsat data, and retrieving surface biophysical properties. We find that when using simulated MISR and Landsat data, the CHRIS/PROBA hyperspectral signal is reproduced with RMSE 0.0001–0.04. LAI is retrieved with r^2 from 0.97 to 0.99 and RMSE of from 0.21 to 0.38. The results based on observed MISR and Landsat data have lower performances, with RMSE for the reproduced CHRIS/PROBA hyperspectral signal varying from 0.007 to 0.2. LAI is retrievedwith r^2 from 0.7 to 0.9 and RMSE from 0.7 to 1.4. We found that for the data considered here the main spectral variations in the visible and near infrared regions can be described by a limited number of parameters (3–4) that can be estimated from multispectral information. Results show that the method can be used to simulate arbitrary bandsets, which will be of importance to any application which requires combining new and existing streams of new EO data in the optical domain, particularly intercalibration of EO satellites in order to get continuous time series of surface reflectance, across programmes and sensors of different designs.

Keywords: MISR; Landsat; CHRIS/PROBA; EO-LDAS; Semi-discrete radiative transfer model; Barrax

1. Introduction

An understanding of surface reflectance over the solar reflective domain (with wavelengths from 400 to 2500 nm) is important in order to monitor the land surface with spaceborne passive optical sensors. Typically, these sensors acquire data in a limited set of bands (e.g. multispectral sensors, such as Landsat (6 bands), Moderate Resolution Imaging Spectroradiometer (MODIS) (7 bands) or Sentinel-2/MSI (12 bands). By contrast, the Compact High Resolution Imaging Spectrometer (CHRIS)

* Corresponding author.
E-mail address: m.chernetskiy@ucl.ac.uk (M. Chernetskiy).

https://doi.org/10.1016/j.asr.2018.07.015
0273-1177/

instrument on board Proba-1 (Barnsley et al., 2004) collects data with a higher spectral resolution (62 bands of 1.3–12 nm width), typically referred to as a hyperspectral sensor. Although hyperspectral data are presented as airborne sensors measurements for decades, they are not so common on the space borne platforms. We note that a number of space missions are expected to be launched in the next few years with the remit of acquiring hyperspectral data. Among these missions are the Environmental Mapping and Analysis Program (EnMAP), Hyperspectral PRecursor of the Application Mission (PRISMA) and Hyperspectral Infrared Imager (HyspIRI) (Guanter et al., 2015; Candela et al., 2016; Lee et al., 2015).

Hyperspectral data are routinely collected from airborne sensors (e.g. Asner et al., 2016), on the ground and have been used for validation and calibration of space borne sensors (Gupta et al., 1998, Baccini et al., 2007, Hay et al., 1997, 2001). Hyperspectral sensors are also being mounted on automated acquisition platforms in flux towers (Porcar-Castell et al., 2015). There has been an increased interest in hyperspectral observations as a way to characterise leaf traits such as specific leaf area or leaf nitrogen content (Roelofsen, 2014, Musavi, 2016).

A typical application of Earth Observation (EO) data in the optical domain over the land surface is the retrieval of biophysical parameters, often carried out through the inversion of a radiative transfer (RT) model. The resulting derived parameters are usually 'validated' by comparisons with ground based measurements (Baret et al., 2006) and/or comparisons among different products (Disney et al., 2016). These validation methods have a number of shortcomings, as detailed for leaf area index (LAI) in e.g. (Disney et al., 2016), which mentions that a number of incompatible assumptions can be made when gathering "ground truth" data and the retrieval scheme chosen (or among different products). The issue of the scale of the measurements is also important (Pfeifer et al., 2012; Widlowski et al., 2005). Validation of other parameters that describe leaf optical properties is also fraught with complications due to similar reasons. An additional, independent, test of an inversion scheme is that the results obtained from the inversion ought to allow one to predict observations from a different sensor, with arbitrary angular and spectral properties. In this respect, hyperspectral sensors present a spectrally comprehensive dataset to compare against.

Data assimilation (DA) schemes, such as the Earth Observation Land Data Assimilation System (EO-LDAS) of Lewis et al. (2012) and Gomez-Dans et al. (2016) produce inferences of land surface parameters based on EO data combined with a number of *a priori* additional constraints. The EO-LDAS approach maps land surface parameters (such as LAI, or leaf and soil optical properties) to surface directional reflectance by means of an RT model scheme. Thus, if the land surface parameters are known, the RT model can be used to predict observations from another sensor, with different acquisition geometries, spectral characteristics, etc. Additionally, the ability to produce a complete time series of parameters allows the prediction of observations when no other sensor data is available.

An example of this approach is provided in Verhoef and Bach (2003), where estimates of LAI, fraction of brown leaves and soil moisture derived from inverting Landsat data, were used to forward model HyMap imaging spectrometer observations. Verhoef and Bach (2007) concentrate on simulation of Top Of Atmosphere (TOA) multi-angular hyperspectral signal by coupling soil-leaf-canopy and atmosphere RT models. In order to validate the RT model results, they simulate hyperspectral signal of CHRIS/PROBA at the bottom-of-atmosphere (BOA) level for bare soil, maize, dense and sparse forest. Comparison of results between real and simulated CHRIS/Proba measurements show RMSE from 0.011 to 0.027.

The inverse problem is known to be ill-posed (Kimes et al., 2000), in practice meaning that there may be infinite solutions that fit the observations equally well. One way around this is to use prior constraints (Combal et al., 2003), complemented by regularisation approaches (Lewis et al., 2012) or by models of the parameter evolution (Koetz et al., 2005; Quaife et al., 2007; Gomez-Dans et al., 2016). Adding more (independent) observations is also an obvious way to add more constraints to the problem. In all these cases, the original set of observations are being complemented by extra information that restricts the solution space.

Several studies have shown that multi-angular information can improve retrieval of land parameters such as LAI. For instance, Knyazikhin et al. (1998) describe the algorithm for synergistic retrieval of LAI and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) from MODIS and MISR measurements. Gobron et al., (2000) demonstrate that using multiple observational angles from MISR reduces the number of solutions when inverting a canopy RT model. Other studies have demonstrated that MISR can also provide information on structure and heterogeneity of vegetation (Pinty et al., 2002; Gobron et al., 2000; Widlowski et al., 2004).

There are currently in excess of 100 EO sensors acquiring data over the land surface. Products that combine observations from different sensors are still relatively rare, as the different spectral, spatial, angular and temporal characteristics of the data, as well as artefacts introduced by parts of the processing, result in a challenging problem. The identification of so-called essential climate variables (ECVs) (Hollman et al., 2013, Bojinksy et al., 2014) is providing a push towards datasets of scientific parameters that use observations from all available satellites, resulting in consistent, uncertainty-quantified, long term records.

In this paper, we aim to advance the development of multi-sensor products by demonstrating a method for using data from a sensor with relatively few spectral bands can be used to predict data from a hyperspectral sensor, via a DA approach. To achieve this we use data from the

MISR (Diner et al., 1998) and Landsat ETM+ sensors to predict bottom-of-atmosphere (BOA) hyperspectral reflectance from the CHRIS/PROBA sensor.

The main assumption we make here is that the major spectral variations in the visible-near infrared region (400–1000 nm, broadly) are controlled by a small number of mechanisms that can be parameterised by a few magnitudes (basis functions) provided by multispectral information. For example Price (1994; 1998) showed that vegetation reflectance exhibits a strong degree of redundancy i.e. much of the variation is concentrated in a few key features. Price (1990) examined the reflectance spectra of 500 different soils at field scale, and showed via principal component analysis that 99.6% of the variance could be explained by just 5 principal components. The main reason of this redundancy is the broad spectral representation of a limited number of underlying physical parameters. More recently, Verrelst et al., (2016) carried out a sensitivity analysis of the PROSAIL model (Jacquemoud et al., 2009) and reported that the main driving factors are LAI, Leaf Angular Distribution (LAD) and soil coefficients, which contribute up to 90% of whole spectral variability. Another example of this issue of redundancy is given by Mousivand et al., (2015) who found that the main driving factors for the Soil-Leaf-Canopy (SLC) model (Verhoef and Bach 2003) and CHRIS-Proba are crown coverage, LAI, leaf inclination distribution function, soil moisture, chlorophyll content of green leaves and fraction of brown leaves. Under this assumption, these few parameters may be retrieved from a limited spectral sampling, and then used to forward model observations from other sensors with different spectral properties. This in turn allows sensors with different spectral properties to be compared directly, or more importantly, merged in a common spectral framework.

In the first part of the study, we test the assumptions above with a theoretical experiment, in which a synthetic observational scenario is used to assess the limitations of the proposed approach without considering complicating factors such as residual effects from atmospheric correction, gridding or sensor calibration. In the second part we apply the same approach to the real multispectral data of MISR and Landsat, both separately and then combined.

2. Methods

Fig. 1 shows the steps of the study. In the "Synthetic data" section, we generate multispectral Landsat and MISR data, then we use the EO-LDAS approach to invert the semi-discrete RT model of Gobron et al. (1997) and retrieve the state parameters. Then we use forward model to simulate hyperspectral and multi-angular observations from CHRIS-PROBA. In the "Real data" section, we apply the same steps to the real multispectral data. For "Comparison", we use field measurements collected during the SPectrabARrax Campaign (SPARC) 2004 (Gandia et al., 2004).

2.1. Study site

The test site of this study is an agricultural site located near Barrax in Spain (Fig. 2).

Barrax is an European Space Agency (ESA) test site where many field campaigns have been carried out over

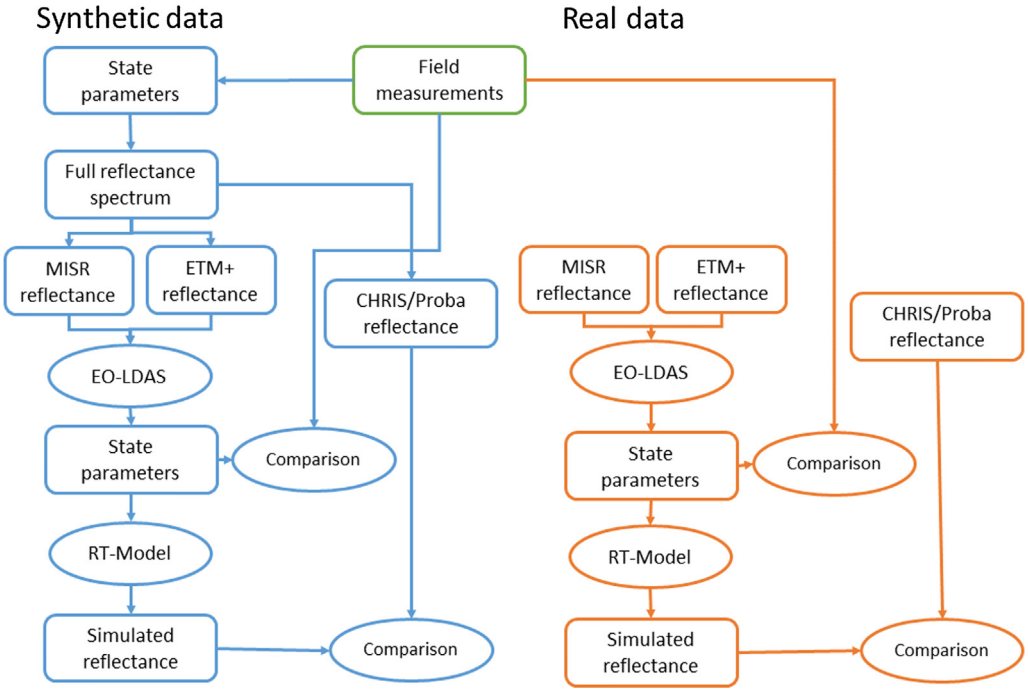

Fig. 1. Simulating synthetic and real CHRIS/PROBA spectra.

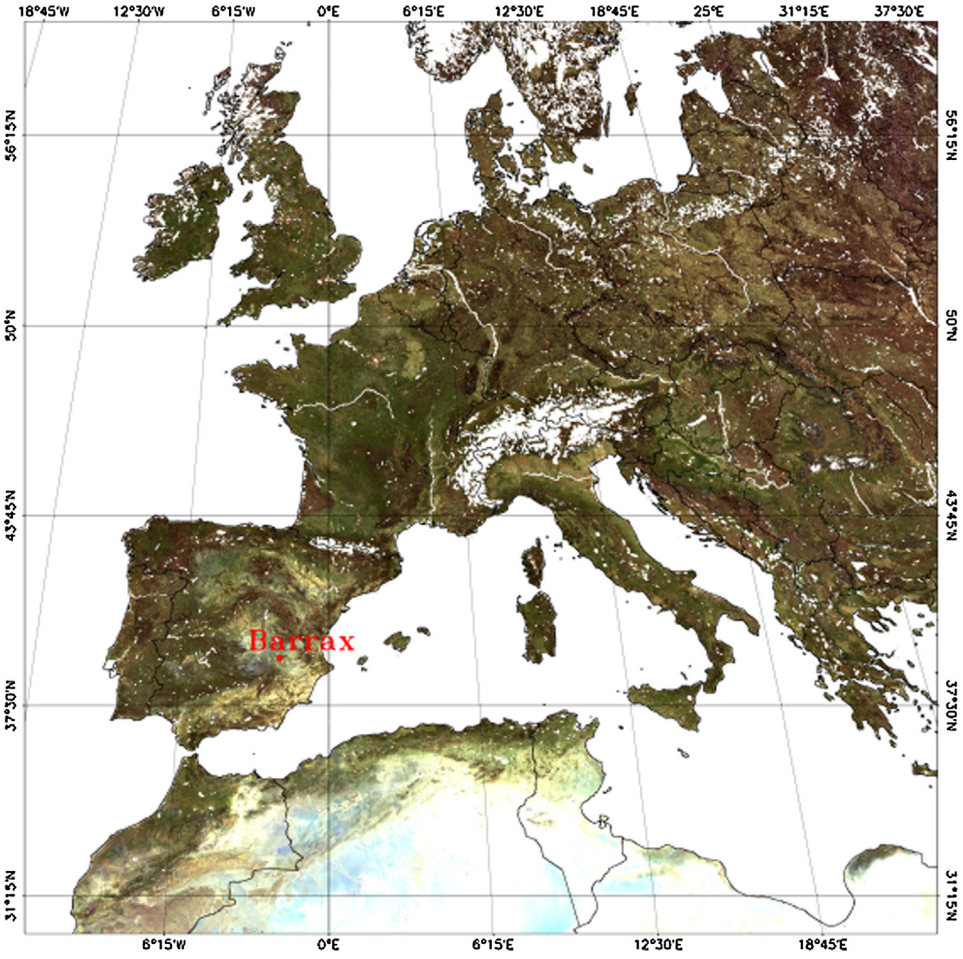

Fig. 2. The Barrax test site on the satellite map of Europe.

the years 2003, 2004, 2005 and 2009. The activities of the SPARC 2004 (Gandia et al., 2004) ground-based campaign included the retrieval of biophysical parameters, such as LAI, chlorophyll concentration, leaf equivalent water thickness and leaf dry matter. Ground spectral measurements were taken with an Analytical Spectral Devices (ASD) FieldSpec Pro FR Spectroradiometer (ASD Inc.), including bare soil and vegetation of some fields. The crop types across the site include alfalfa, sugar beet, corn, garlic, onion, potato, sunflower and vineyard, with a total of 18 fields being measured. Biophysical parameters which are relevant for this work are presented in Table 1.

2.2. Remote sensing data

Data from the CHRIS/Proba sensor have been collected contemporary with the SPARC ground measurements campaign. In mode 1, the data provide 62 spectral bands in the visible and near-infrared region, from 411 to 997 nm with a spatial resolution at nadir of 34 m and five view zenith angles: ±55°, ±36° and 0° (nominally). Spectral resolution varies from 1.3 nm at the 410 nm band to 12 nm at the 1050 nm band (Barnsley et al., 2004). Fig. 3 illustrates the illumination (stars coloured in orange) and viewing geometry (green dots) of CHRIS/Proba data acquired over the Barrax site on 16th July 2004.

MISR is currently the only operational multi-angular optical sensor that obtains information globally and simultaneously at multiple (nine) view angles. We use the MISR top-of-canopy (TOC) reflectance at 275 m spatial resolution produced by the sharpening method of Verstraete et al. (2012) for the 16th July 2004. The data comprise four spectral bands at 446 nm, 558 nm, 672 nm and 867 nm with nine cameras at nadir view, ±26.1°, ±45.6°, ±60.0°, and ±70.5°. Acquisition geometry for all cameras is also represented in Fig. 3 in blue dots. MISR per band uncertainties for real data are specified as 0.05 for all four bands. (Chernetskiy et al., 2017).

The Landsat ETM+ data were acquired on the 18th July 2004 with 30 m spatial resolution and 6 spectral bands. The viewing and illumination geometry are also displayed in Fig. 3. Atmospheric correction was performed using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) software (Masek et al., 2012). The default LEDAPS routine was used. All input information required for atmospheric correction was taken from Landsat metadata; ancillary data of (Total Ozone Mapping Spectrometer) TOMS and National Centres for Environmental

Table 1
Parameters from the SPARC database.

Field number	Field code	Field code	LAI (m^2/m^2)	Cab ($\mu g/cm^2$)	Leaf water (Cw) (g/m^2)	Dry matter (Cdm) (g/m^2)
1	Alfalfa	A1	3.73	50.17	36.0	144.6
2	Alfalfa	A2	3.36	48.7	41.55	119.3
3	Sugar Beet	B1	4.48	49.05	64.5	424.1
4	Corn	C1	1.69	49.33	76.2	148.2
5	Corn	C9	2.92	52.94	76.0	148.0
6	Corn	C10	2.57	52.16	76.0	148.0
7	Garlic	G1	0.63	51.71	100.4	654.8
8	Garlic	G1	0.63	48.22	100.0	655.0
9	Garlic	G1	0.63	43.6	100.0	655.0
10	Garlic	G1	0.63	39.02	100.0	655.0
11	Onion	On3	0.95	31.26	71.3	499.4
12	Potato	P2	3.96	37.98	40.7	270.8
13	Potato	P3	4.03	34.85	41.0	271.0
14	Sunflower	SF1	0.5	43.35	87.5	368.6
15	Sunflower	SF1	0.5	44.1	87.0	369.0
16	Sunflower	SF1	0.5	39.76	87.0	369.0
17	Sunflower	SF3	0.66	43.71	69.6	404.9
18	Vineyard	V1	2.51	34.58	91.0	202.3

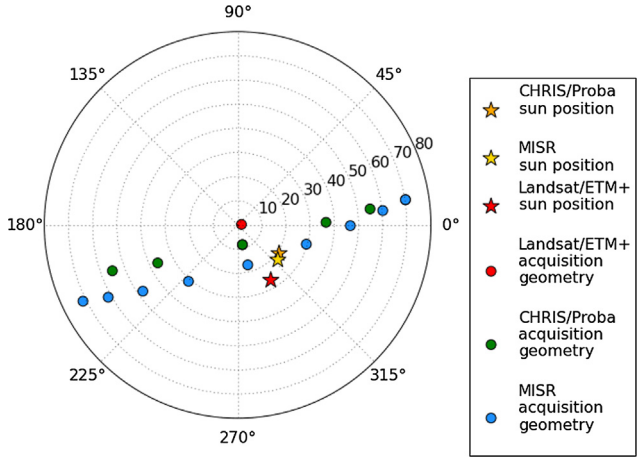

Fig. 3. Acquisition geometry in polar coordinates of the MISR, Landsat ETM+ and CHRIS/Proba data.

Prediction (NCEP) Reanalysis. Fig. 4 shows true colour composites of the three types of satellite observations. We assume that a Landsat pixel corresponds to centre of MISR pixel where we have highest response of Point Spread Function (PSF). Landsat per band uncertainties for real data are assumed to be normal with zero mean and standard deviations of0.0202, 0.0246, 0.0250, 0.0214, 0.0395 and 0.0272 for bands 1–5, 7 respectively (Maiersperger et al., 2013). Real per band uncertainties for MISR and Landsat are not provided and their estimation is a complicated task, which is outside of this study.

Fig. 4 shows that the image of Landsat-7 ETM+ was acquired with Scan Line Corrector (SLC) turned off. This means that we have vertical 'stripes' where data are lost. However apart from some loss of data the SLC-off imagery has the same radiometric and geometric quality as imagery collected before SLC failure. We can see that most of the field measurements are within regions where data are available. The two fields where data are lost are Alfalfa (A2) and Onion (On3) are excluded from the analysis of the real Landsat data.

2.3. Methods

EO-LDAS (Lewis et al., 2012) is a generic variational data assimilation system designed to infer land surface properties (and associated uncertainties) from heterogeneous combinations of observations (interpreted through observation operators, typically RT models) and an array of prior constraints. The DA system (Gomez-Dans and Lewis 2012) defines a cost function, which is then minimised by a gradient descent approach. This cost function describes the model fit to observations, as well as penalties from departing from prior values:

$$J(x) = J_{obs}(x) + J_{prior}(x) \quad (1)$$

where x is the state vector: each element contains the value of a land surface parameter. In this study, x contains ten parameters, eight vegetation and two soil parameters. Various parameter transformations are used to quasi-linearise the RT model (Weiss et al., 2000). $J_{obs}(x)$ is the fit to the observations and $J_{prior}(x)$ is the *a priori* parameter distribution.

The observational operator embedded in $J_{obs}(x)$ provides a mapping from land surface parameters to directional surface reflectance using the semi-discrete radiative transfer canopy model of Gobron et al. (1997), coupled to the leaf optical PROpertiesSPECTra (PROSPECT) of Jacquemoud and Baret (1990) and the spectral soil model of Price (1990) for the albedo of background soil. We use

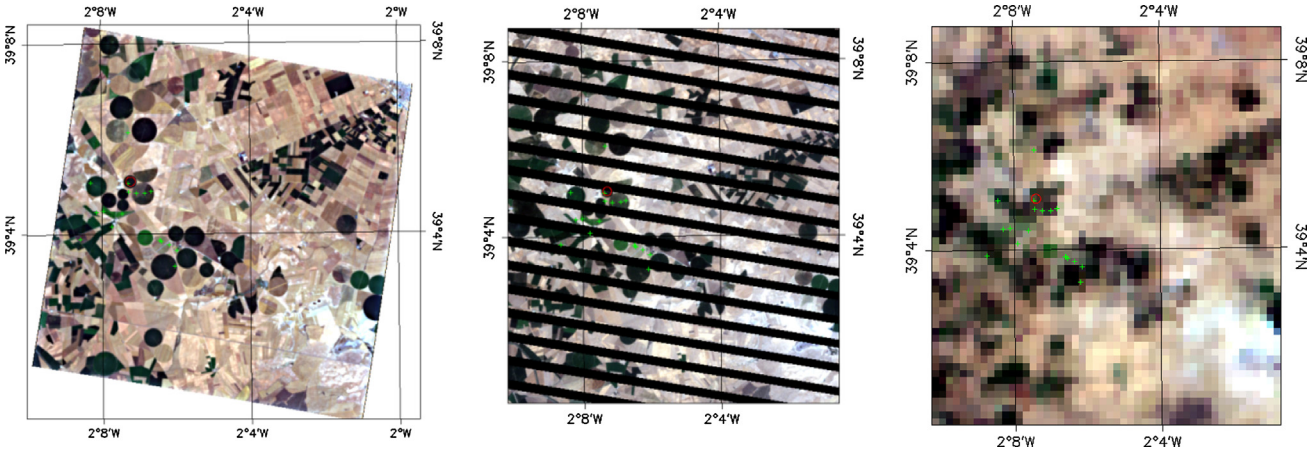

Fig. 4. True colour Images of CHRIS/Proba (left hand side panel), Landsat ETM+ (middle panel) and MISR at 275 m (right hand side panel). Green cross indicate the 18 fields listed in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

modification of the PROSPECT-5 (Lewis et al., 2008, Féret et al., 2008). The soil model is described as:

$$R_0 = s_1\varphi_1 + s_2\varphi_2 \quad (2)$$

where R_0 is soil background reflectance, φ_1 and φ_2 are basis functions (principle components) of Price, s_1 and s_2 are soil model coefficients (Price 1990).

The prior component $J_{prior}(x)$ contains the prior parameter distribution. The prior probability distribution function (pdf) in EO-LDAS is Gaussian (so defined by a mean vector and a covariance matrix) and it encodes the belief on the distribution of the parameter values before the observations are taken into account. In this study, we assume the prior distribution to have means and standard deviations given by the values shown in Table 2. There are 8 state parameters for the canopy and leaf model and two for the soil model. Note that the Leaf Angle Distribution (LAD) is fixed to one of the five Bunnik archetype expressions (Bunnik 1978). Prior information on parameters is taken from (Lewis et al., 2012). LAI and height are set to values which are reasonable for an agricultural field. The prior covariance matrix is assumed to be diagonal.

The aim of the exponential transformation of some parameters from Table 2 is to quasi-linearize the RT model, so that the assumption that the posterior PDF is Gaussian is better met.

The assumptions in EO-LDAS result in the posterior pdf of x being given by a multivariate normal. The mean vector is given by the minimum of $J(x)$ (effectively, minus the log posterior), and the posterior uncertainty is given by the inverse of the Hessian at the minimum, using a linear approximation of the Hessian (Lewis et al., 2012).

One important issue is to find an initial estimate of the state parameters because it determines starting values – the first guess for an inversion process. A good first guess reduces the chance of being trapped in local minima of the decision space and typically results in faster convergence of the gradient descent procedure. A sampling design of 10 starting points over parameter space was used. We assume a uniform distribution and use a Latin Hypercube Sampling (LHS) scheme (Sacks et al., 1989).

Table 2
The bounds of state variables and prior information.

Parameter	Transformation	Lower limit	Upper limit	Prior mean	Prior sd.
LAI [m^2/m^2]	$e^{(-x/2)}$	0.08	0.95	0.37	1.0
Canopy height (xh) [m]	–	0.1	5.0	1	3
The leaf radius (xr) [m]	–	0.001	0.3	0.15	0.15
Concentration of chlorophyll a + b (C_{ab}), [$\mu g/cm^2$]	$e^{(-x/100)}$	0.5	0.75	0.6	0.4
Proportion of senescent material (scen)	–	0.001	1.0	0.5	0.5
Equivalent leaf water (C_w) [cm]	$e^{(-x*50)}$	0.03	0.999	0.6	0.5
Dry matter (C_{dm}), [$\mu g/cm^2$]	$e^{(-x*100)}$	0.13	0.99	0.37	0.7
Number of leaf layers (N)	–	1	2.5	1.5	1.5
Soil PC1 (s_1)	–	−20.0	−1.0	−12.21	1.33
Soil PC2 (s_2)	–	−1.0	2.0	0.32	2.14
Leaf angle distribution (LAD)	–	Planophile – 1 Erectophile – 2 Plagiophile – 3 Extremophile – 4 Uniform – 5			

The optimisation that resulted in the lowest cost was chosen. The main advantage of using a LHS over other methods (e.g. random sampling) is that it will produce a space-filling sampling design for a given number of samples and dimensions.

3. Results

3.1. Soil model

Many fields of the Barrax region have relatively low LAI in range from 0.5 to 1.0, meaning the soil reflectance will have a strong influence on the overall reflectance signal and hence that incorrect soil description can be a source of uncertainties (Table 1). However, availability of the ASD soil measurements gives the possibility of finding more precise φ_1 and φ_2 functions (Eq. (2)). We apply Principal Component Analysis (PCA) to 62 ASD SPARC spectral measurements of bare soil and find that first two principal components (PC) explain more than 97% of all variability (Fig. 5).

We use these PCs as φ_1 and φ_2 vectors of the Price soil model. Then we invert this soil model against available ASD measurements and find s_1 and s_2 coefficients. For this we solve a linear system:

$$\bar{s} = (AC^{-1}A^T)^{-1}AC^{-1}\bar{\rho} \tag{3}$$

where;

A is a matrix of phi functions:

$$\begin{aligned} A = &[[\varphi_1(t_1,\lambda_1),\ldots,\varphi_1(t_1,\lambda_{2101}),\ldots,\varphi_1(t_n,\lambda_{2101})],\\ &[\varphi_2(t_1,\lambda_1),\ldots,\varphi_2(t_1,\lambda_{2101}),\ldots,\varphi_2(t_n,\lambda_{2101})]] \end{aligned} \tag{4}$$

C is a covariance matrix; ρ is a vector of reflectance:

$$\rho = [\rho(t_1\lambda_1),\ldots,\rho(t_1\lambda_{2101}),\ldots,\rho(t_n\lambda_1),\ldots,\rho(t_n\lambda_{2101})] \tag{5}$$

We find that $\mu_{s1} = -12.21$, $\sigma_{s1} = 1.33$ and $\mu_{s2} = 0.32$, $\sigma_{s2} = 2.14$. The soil model defined by the resulting coefficients is shown in Fig. 6. We can see that in general the model follows the ASD measurements except some for some small differences in the NIR region, but still remaining well within the range of uncertainties. This model is a more precise description of the Barrax soil and allows for the use of the resulting μ and σ values in the EO-LDAS prior term.

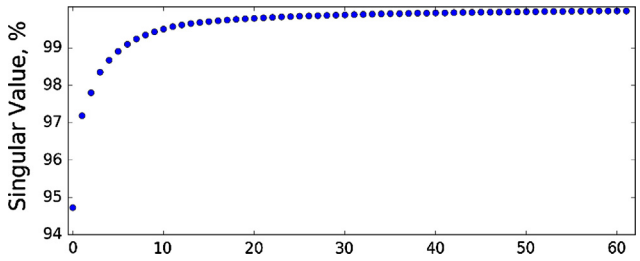

Fig. 5. Results of PCA analysis of 62 ASD measurements.

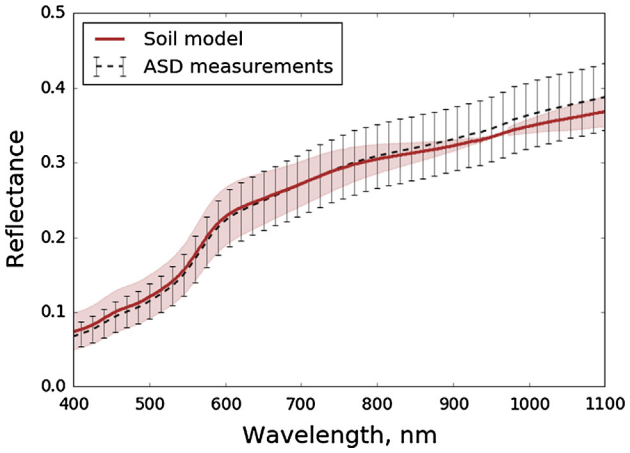

Fig. 6. Soil model with associated uncertainties in 1σ (brown line). Mean and standard deviation (1σ) of bare soil ASD measurements (dotted line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Reproducing real spectral information

Here, we test the potential of the semi-discrete RT model to reproduce spectral characteristics of the Barrax vegetation. Values of LAI, C_{ab}, C_w and C_{dm} are known from the SPARC database (Table 1), s_1 and s_2 were found in the previous section. This means that we know six parameters, which according to, e.g. Mousivand et al., 2015, are most influential. The unknown values of xh, xr, scen and N can be found by inversion of available CHRIS/Proba data. In order to simplify and speed-up the process we generate a look-up-table (LUT) with 4000 entries by means of LHS filled with the values of xh, xr, scen and N. The resulting spectra obtained by minimization of the LUT are shown in Figs. 7 and 8. In order to check ability of the model to reproduce the bidirectional reflectance distribution function (BRDF) in this particular situation, we modelled reflectance for all five CHRIS/Proba 'fly-by' positions. Table 3 shows RMSE in range from 0.004 to 0.061 for all 'fly-by' positions. We can see that slightly higher minimum RMSE (0.009–0.01) is for ±55° positions but highest maximum RMSE (0.061) is for nadir view. Fig. 7 shows that the best results are for the fields with lowest LAI, demonstrating the accuracy of the chosen soil model. The highest RMSE is 0.061 for the potato field P2 which is responsible for highest RMSE for all viewing directions (Table 3). So maximum values of RMSE mostly due to discrepancies in some fields which can not be parameterised with given model assumptions (e.g. continuous canopy) and spatial resolution. If we accept the correctness of the field measurements these discrepancies can potentially be explained by structural effects at the canopy (clumping) and/or leaf level, that cannot be simulated by the available model assumptions. This example also demonstrates that the soil model correctly represents

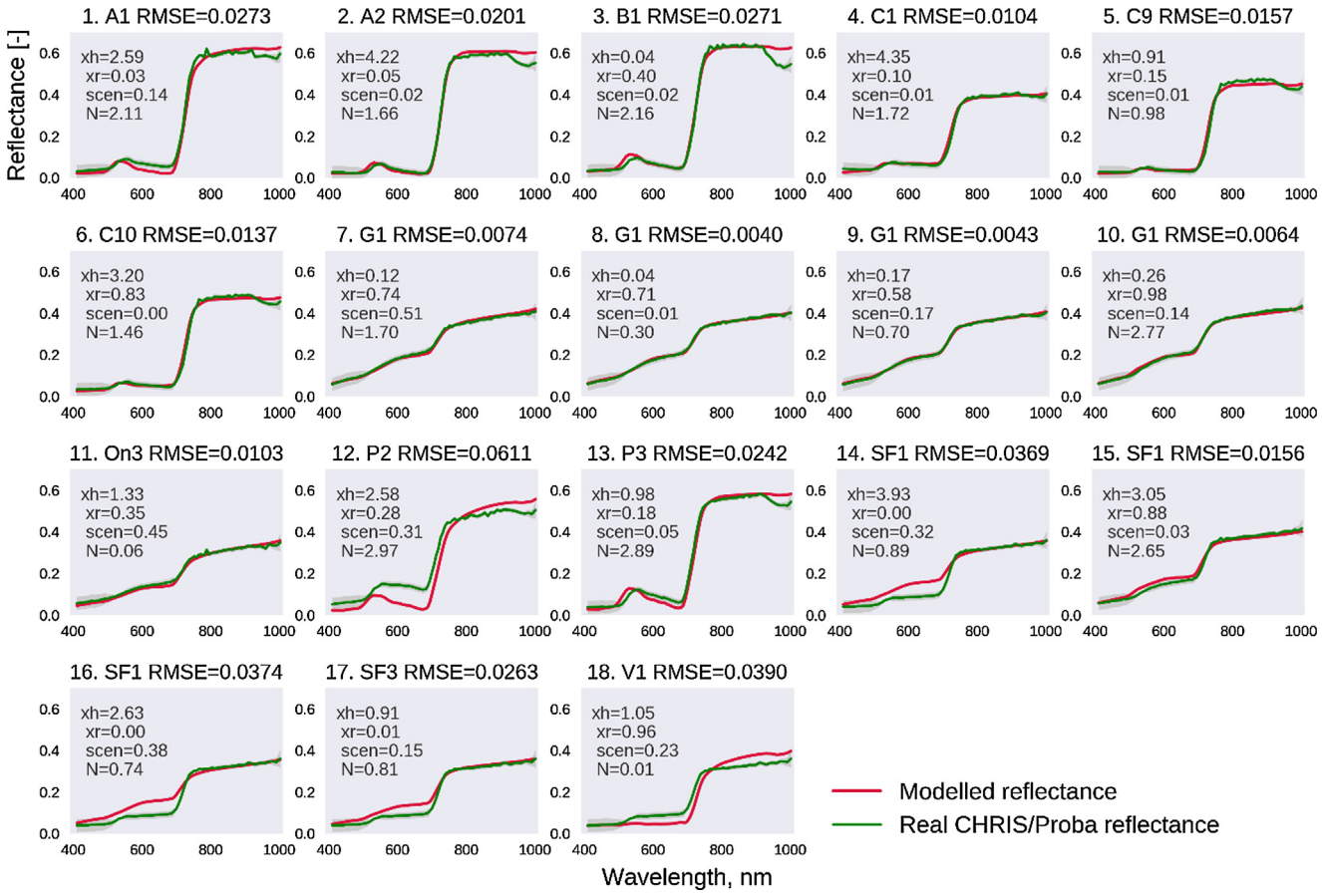

Fig. 7. Real and modelled CHRIS/Proba spectra for the nadir view. LAI, C_{ab}, C_w and $C_{dm}$are fixed to the field measured values. Each panel corresponds to each field.

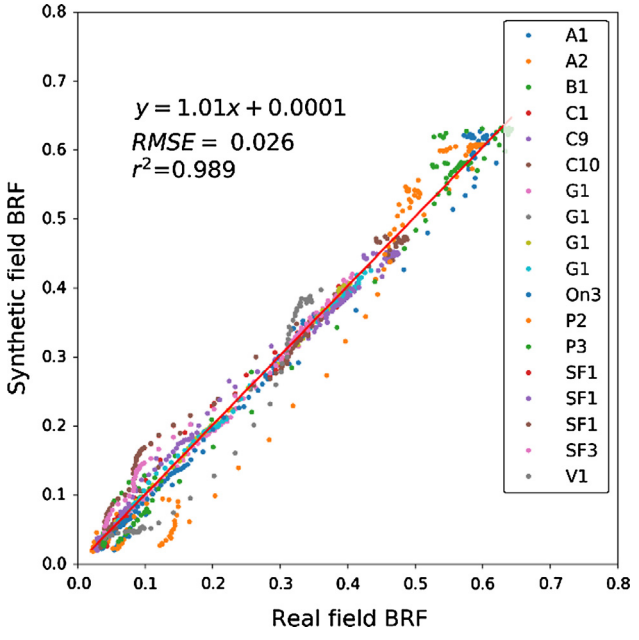

Fig. 8. Scatterplot of real and modelled CHRIS/Proba spectra (400–2500 nm) for the nadir view. LAI, C_{ab}, C_w and $C_{dm}$are fixed to the field measured values. Each field has its own colour code. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Minimum and maximum RMSE for fit between real and modelled CHRIS/Proba. LAI, C_{ab}, C_w and C_{dm} are fixed to the ground truth values.

'fly-by' position	0°	−36°	+36°	+55°	−55°
min RMSE	0.004	0.005	0.005	0.010	0.009
max RMSE	0.061	0.048	0.055	0.059	0.053
mean RMSE	0.022	0.023	0.024	0.024	0.027
std. RMSE	0.015	0.015	0.013	0.013	0.014

background signal and can reproduce BRDF. Results of this section demonstrate that with given ground truth measurements, it is possible to find a set of parameters that allow reproducing hyperspectral data over the Barrax fields in all CHRIS/Proba viewing directions.

Figs. 7 and 8 show that some fields exhibit some difference between real and modelled spectra in green and NIR regions. The most likely reason of this is discontinuous canopies which cannot be detected by given spatial resolution and model assumptions. This exercise potentially can be improved by including field measurements of some parameters which are unavailable in this study such as N and by using different model assumption such as clumping, etc.

3.3. Synthetic data

3.3.1. Generation of the synthetic datasets

Synthetic data represent an 'idealised' test case and allow checking potential of the EO-LDAS parameter retrievals and subsequent reproduction of hyperspectral information.

We simulate the Bidirectional Reflectance Factors (BRFs) of CHRIS/Proba, MISR and Landsat (ETM+) over the 18 SPARC field measurement points. The synthetic data are based on the SPARC database (ESA, 2004) with given information on LAI, C_{ab}, C_w and C_{dm} (Table 1). All other parameters were taken from Lewis et al. (2012) (Table 2). Since each MISR $275 \times 275\ m^2$ pixel contains approximately 81 Landsat pixels we generate 18 9x9 pixel fields with 10 layers of parameters each and add correlated Gaussian noise but keeping the central pixel with information from Tables 1 and 2. Then each field was averaged to a single value and the number of spectral bands was reduced to those of CHRIS, Landsat and MISR. Finally, we added random noise, which we define as 0.1 times the uncertainty of the real data i.e. [0.00202, 0.00246, 0.00250, 0.00214, 0.00395 and 0.00272] (Section 2.2). These numbers are reasonable because they give some noise but keep it low because the aim of the synthetic experiment is to test the best possible (idealised) conditions for retrieval. Increasing of noise can lead to decreasing sensitivity to some of parameters and we will not see potential possibility of retrieval. Input per band uncertainties depend on random noise and are in the range from 0.0002 to 0.002 for Landsat; and from 0.001 to 0.04 for MISR. The resulting synthetic data are similar but not necessarily identical to real spectral information because we have five parameters, which are not available from the SPARC database (Table 4).

3.3.2. Retrieval of biophysical parameters

In this work, we simulate a hyperspectral signal by forward run of the RT model with input parameters retrieved by the model inversion. So it is implied that in order to restore the hyperspectral signal we have to precisely retrieve one or several state parameters.

We solve the problem only for LAI, Cab, Cw and soil coefficients. All other parameters are fixed to values from table 2. Some of the fixed parameters such as N have influence on leaf spectra from 400 to 2500. However, this contribution has been ignored.

One of the goals of this work is to test the influence of the MISR cameras combinations on the retrieval of parameters and on the subsequent simulation of hyperspectral signal. We do the retrievals by increasing the number of cameras from one to nine i.e. An camera on the first step, An-Af on the second, An-Af-Aa on the third, etc. This means that at first, we have the nadir-only camera, and then the range of View Zenith Angles (VZAs) is increasing with increasing number of cameras. We assume that bands and cameras are independent and treat each camera as additional J_{obs} term in equation 1.

Fig. 9 displays the results of retrieval of LAI, C_{ab} and C_w by inversion of MISR and Landsat data separately and then MISR + Landsat together. LAI is retrieved with r^2 from 0.97 to 0.99 and RMSE from 0.21 to 0.38 for all considered cases.

The results from LAI retrievals using Landsat only data are similar to the MISR only results. The RMSE in the retrievals essentially spans the entire dynamic range of the field values.

Table 4
Parameters for synthetic data.

Field N	Field	Canopy height (xh)	Leaf radius (xr)	Proportion of senescent material (scen)	Number of leaf layers (N)	Leaf Angle Distribution (LAD)
1	A1	5	0.01	0	1	1
2	A2	5	0.01	0	1	1
3	B1	5	0.01	0	1	5
4	C1	5	0.01	0	1	2
5	C9	5	0.01	0	1	2
6	C10	5	0.01	0	1	2
7	G1	5	0.01	0	1	2
8	G1	5	0.01	0	1	2
9	G1	5	0.01	0	1	2
10	G1	5	0.01	0	1	2
11	On3	5	0.01	0	1	2
12	P2	5	0.01	0	1	5
13	P3	5	0.01	0	1	5
14	SF1	5	0.01	0	1	1
15	SF1	5	0.01	0	1	1
16	SF1	5	0.01	0	1	1
17	SF3	5	0.01	0	1	1
18	V1	5	0.01	0	1	1

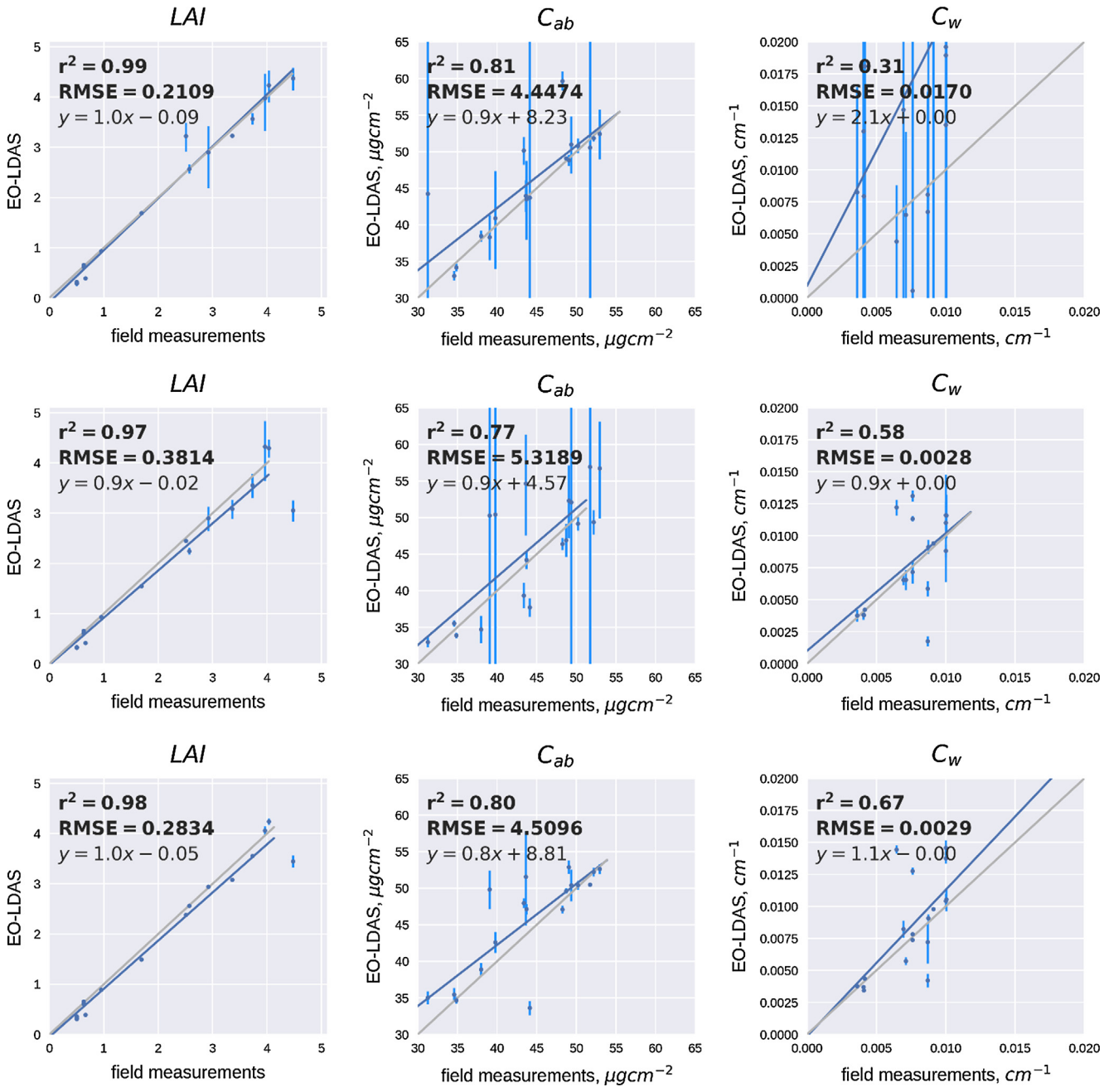

Fig. 9. Comparison of field measurements and EO-LDAS retrievals of LAI, Cab and Cw with synthetic data. MISR only (top panels), Landsat only (middle panels) and MISR + Landsat (bottom panels). MISR nadir camera (An).

Combining both sensors for LAI results in an inference of LAI that, compared with LAI measurements, is similar to the MISR only case.

Retrieval of C_{ab} is similar for all sensors with lowest r^2 and highest RMSE in case of Landsat-only results which is reflected in the highest RMSE and uncertainties. Results of C_w show worst performance for MISR-only, because MISR does not have sensitivity to water content, as spectral bands sensitive to water absorption are not available. Best C_w results are achieved for Landsat-only and the combination of MISR and Landsat, with nearly the same RMSE and better r^2 for the latter. This might be due to the system reacting to a better inference of LAI, which in turn limits the possible values of C_w, resulting in a better estimate of a parameter to which one of the sensors has no sensitivity, as explained above.

Uncertainties are high for C_w in MISR-only case and for C_{ab} in Landsat-only case. However, with the combination of two sensors they are lower for all three parameters.

Results of the retrieval of parameters demonstrate precise retrieval of LAI in all three cases and reasonable retrieval of C_{ab} and C_w in case of combination of two sensors.

These results were acquired using nadir-only MISR-camera. Next results show what happens if the number of the cameras is increased.

Fig. 10 shows the performance of the retrieval with increasing numbers of MISR cameras. We can see that in case of LAI and C_w there is no distinct linear trend with increasing number of cameras. The exception is nadir-only LAI where RMSE is higher. Lower RMSE for C_w in case of Landsat and MISR + Landsat expected because of sensitivity of Landsat to this parameter. The parameter which is affected by increasing of number of cameras is C_{ab}. We can see decreasing of RMSE for retrieval where from four to nine MISR cameras were used.

However, Fig. 11 shows decreasing uncertainties with increasing number of cameras for LAI and Cab. Also it shows that combination of sensors have lower uncertainty in case of all three parameters.

LAI and Cab retrievals show that the mean *a posteriori* does not change significantly, but that as more MISR cameras are considered, the uncertainty decreases. This is an automatic consequence of the set-up of the problem, and assuming uncertainty in the bands is independent.

These results are not unexpected: the increase of more observations from MISR cameras results in a shrinking of the posterior uncertainty, as expected. This is further enhanced by adding the extra Landsat data. Generally speaking, the Landsat inversions are closer to the true values, again as expected due to the lower uncertainty in the measurements.

A similar pattern is also present with leaf chlorophyll concentration, although in this case, the *a posteriori* mean is less stable. This is expected: the assumption of statistical independence in the observations produces a reduction of the posterior variance in both cases. The sensitivity of the data to LAI is higher than to chlorophyll, as well as its distribution having a higher dynamic range, which accounts for the higher stability of the LAI solution relative to the Cab inversion.

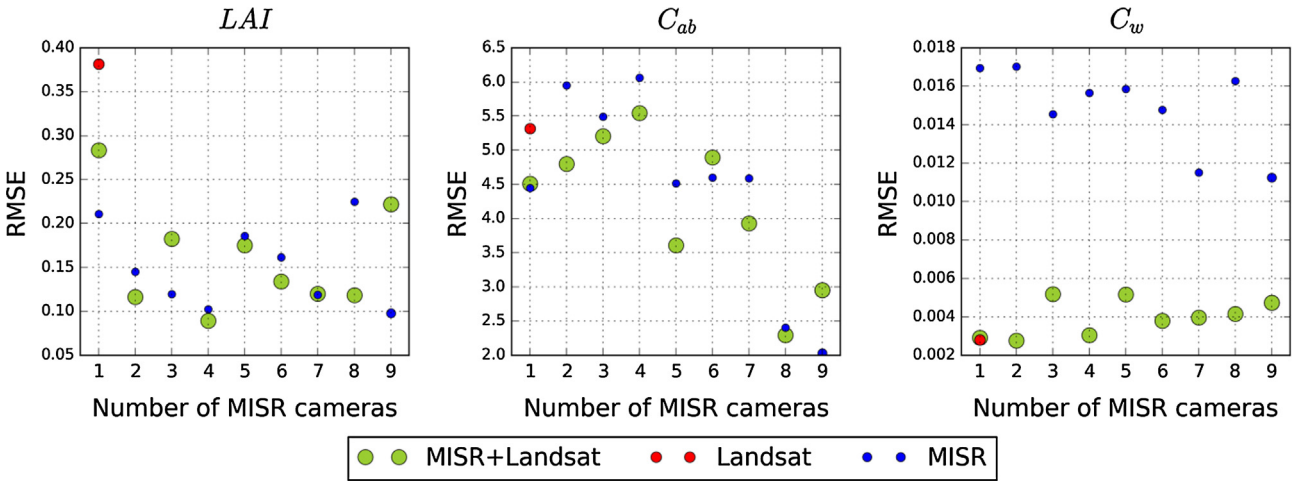

Fig. 10. RMSE of LAI (left hand side panel), Cab (middle panel) and Cw (right hand side panel) for increasing number of MISR cameras.

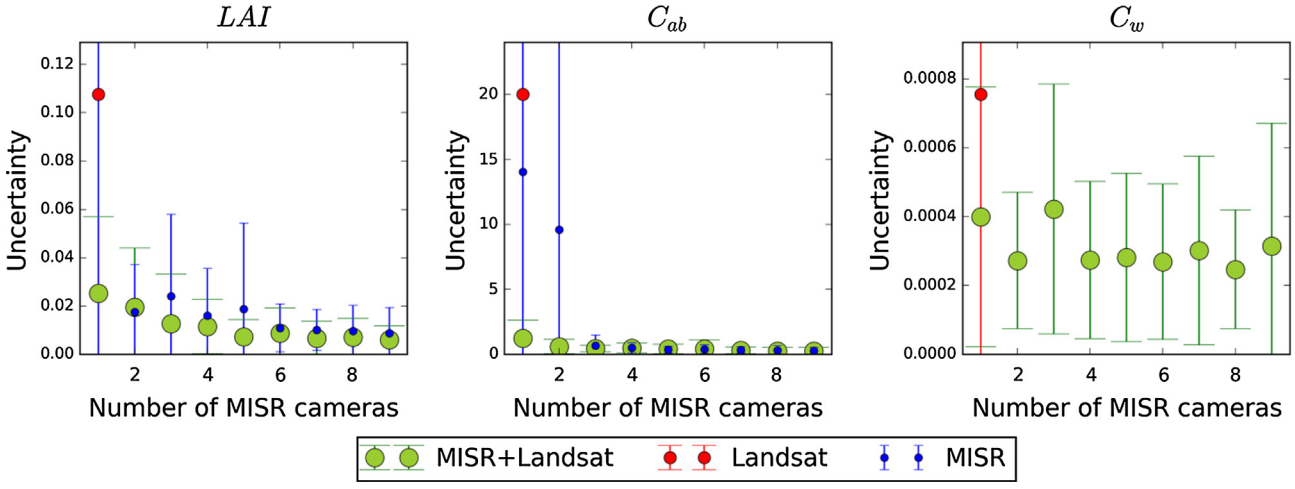

Fig. 11. Averaged uncertainty of LAI (left hand side panel), Cab (middle panel) and Cw (right hand side panel) for increasing number of MISR cameras.

3.3.3. Cross entropy

One of the results of the Section 3.3.2 is that with increasing number of MISR cameras posterior uncertainties are decreasing. The possible reason for this is increasing amount of information. In this section, we make an experiment to find out what is the main reason for decreasing of uncertainties: number of viewing directions or number of observations.

We generate seven synthetic nadir MISR observations with different random noise. For the same state parameters (Table 2) we generate 7 multiangular observations. Then we estimate the state parameters and the cross entropy is calculated as

$$H_e = \frac{1}{2}\ln(det(C_{prior}H_{hess})) \quad (6)$$

where H_e - cross entropy, C_{prior} – prior covariance, H_{hess} – Hessian matrix.

Hessian is a matrix of partial second derivatives, calculated at the minimum of the solution space or maximum *a posteriori* (MAP) point. The Hessian matrix describes the radius of curvature around MAP point, and in the case of Gaussian distributions, it is equivalent to the inverse *a posteriori* covariance matrix, and as such encodes the *a poteriori* uncertainty (e.g. variance/covariance structure) in the retrieved parameters.

Under the assumption of a Gaussian *a priori* pdf for the parameters, Cprior encodes the *a priori* covariance between parameters. In this contribution, we assume Cprior to be diagonal, and thus only define *a priori* variances and not covariances. The cross-entropy He quantifies the "information gain" of the system going from prior to posterior distribution, or in other words, it quantifies how the system interprets the evidence from the observations in terms of parameters.

Fig. 12 demonstrates that with increasing number of observations, the amount of information is increasing. However, it is nearly the same for both nadir-only data and multiangular data until the third camera, and then fluctuates from fourth to seventh cameras.

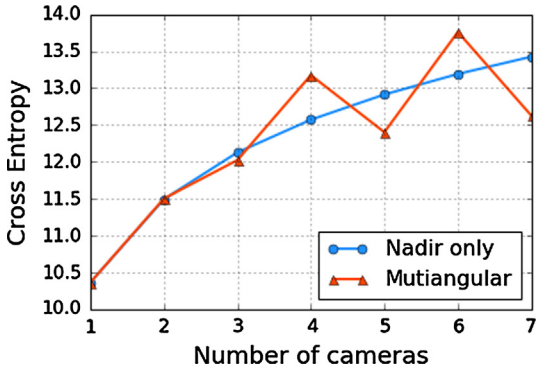

Fig. 12. Changing cross entropy with increasing number of the MISR cameras: 7 nadir only cameras, 7 multiangular observations from An to C MISR cameras.

Note that in nadir only case in the sequence of adding of cameras the cross entropy is increasing but values of the retrieved parameters remain the same.

This example shows that, at least in case of this angular/spectral configuration, increasing number of observations has nearly the same effect as increasing number of different viewing directions. This is in agreement with results of the previous section where we can see that increasing number of MISR cameras leads to a decrease of posterior parameter uncertainties. As above, this is a consequence of the assumptions of independent bands. In practice, this is an important issue, as when EO data has been atmospherically corrected, it is likely that strong correlations in the uncertainty appear in different bands.

3.3.4. Simulation of the CHRIS/PROBA signal

This section investigates possibility to restore hyperspectral signal based on the results from Section 3.3.2.

CHRIS/Proba reflectance is retrieved by forward run of the RT model for each combination of the MISR cameras and for each CHRIS/Proba 'fly-by' positions.

Fig. 13 shows one example of direct comparison between synthetic CHRIS/PROBA and retrieved reflectance spectra for Alfalfa crop field for the nadir camera (An) when using only An camera (left hand side panel) and nine cameras (right side panel). We can see that there is no distinct RMSE trend with increasing number of cameras which is in line with results obtained for retrieval of the state parameters and cross-entropy.

Figs. 14 and 15 show that with given set of retrieved parameters, synthetic CHRIS/PROBA spectra can be simulated very close to unity with both Landsat and MISR + Landsat.

Fig. 16 summarizes the comparison of the 'true' and retrieved hyperspectral data over the 18 fields with synthetic MISR data only. Here we show the nine cameras case because one of the results of Section 3.3.2 is that using nine cameras provides lower uncertainties. In the left-hand column, the results of the simulation of CHRIS/Proba camera VZA 8.4° are shown in terms of RMSE. The right panel corresponds to VZA −55.2°. The lines correspond to the 18 Barrax fields. We show result only for two CHRIS/Proba 'fly-by' positions because no big difference was found for other three directions. Note that the Z scale is logarithmic.

Fig. 16 shows low squared difference for all fields with slightly higher values in NIR region for fields 1–4, 6, 9–11 and 18. However taking into account that highest squared difference varies from 10^{-3} to 10^{-4} the error is very low. RMSE is in the range from 0.0002 to 0.04.

Fig. 17 displays the Landsat only case. Landsat/ETM+ has three additional bands in the SWIR region which are outside the CHRIS/Proba spectral range but due to better constraining of retrieval of parameters there is a noticeable improvement for fields 7–9 and 15–18. RMSE varies in the range from 0.001 to 0.01.

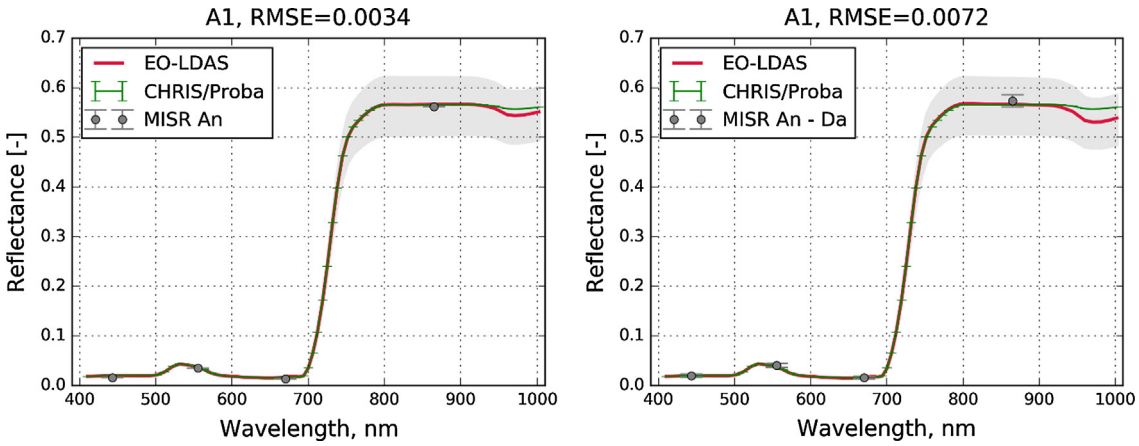

Fig. 13. Comparison of CHRIS synthetic (green line) for nadir view and MISR-only retrieved reflectance spectra (red line) over the Alfalfa field. Left hand side panel is for An (nadir) camera only, right hand side panel for nine cameras. Grey surface indicates the posterior uncertainties in 95% credible interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

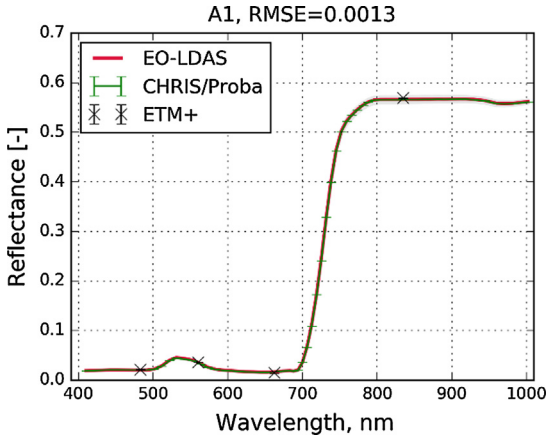

Fig. 14. Comparison of CHRIS synthetic (green line) and Landsat-only retrieved reflectance spectra (red line) over the Alfalfa field. Grey surface indicates the posterior uncertainties in 95% credible interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The MISR + Landsat case and has slightly lower RMSE range for all simulated CHRIS/Proba cameras: from 0.0001 to 0.01. However, in general this result is very similar to Landsat-only. This tells that Landsat data dominates the retrieval process and that even if spectral information content of Landsat is not enough for precise retrieval of all parameters (Figs. 7–9), it is enough for precise simulation of a hyperspectral sensor.

3.4. Real data

Real data experiment in general follows the same steps as synthetic experiment. I.e. we use CHRIS/PROBA as a benchmark for simulated hyperspectral data which are based on forward run of the model using parameters obtained by MISR and Landsat data inversion. Inversion is done by MISR-only data, Landsat-only and with merged MISR + Landsat. We use soil model described in Section 3.1 as a prior.

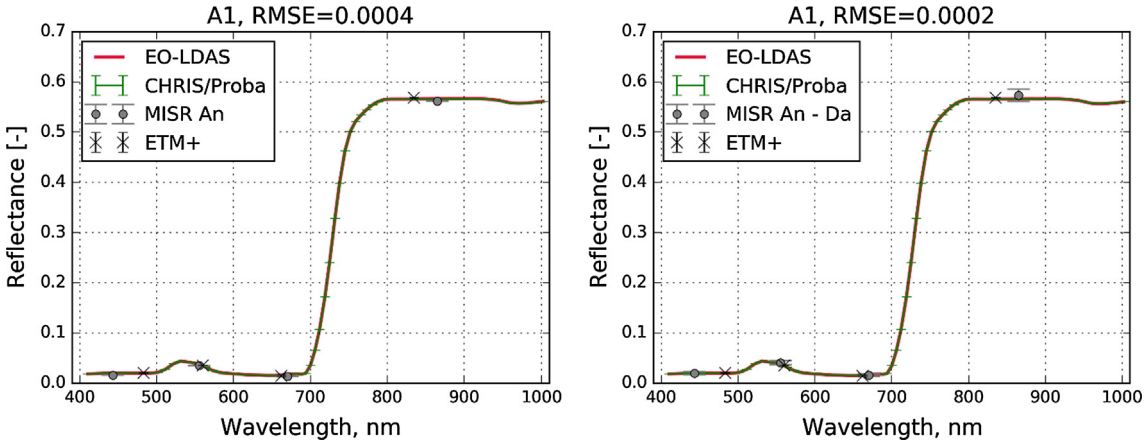

Fig. 15. Comparison of CHRIS synthetic (green line) for nadir view and MISR + Landsat retrieved reflectance spectra (red line) over the Alfalfa field. Left hand side panel is for An (nadir) camera only, right hand side panel for nine cameras. Grey surface indicates the posterior uncertainties in 95% credible interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

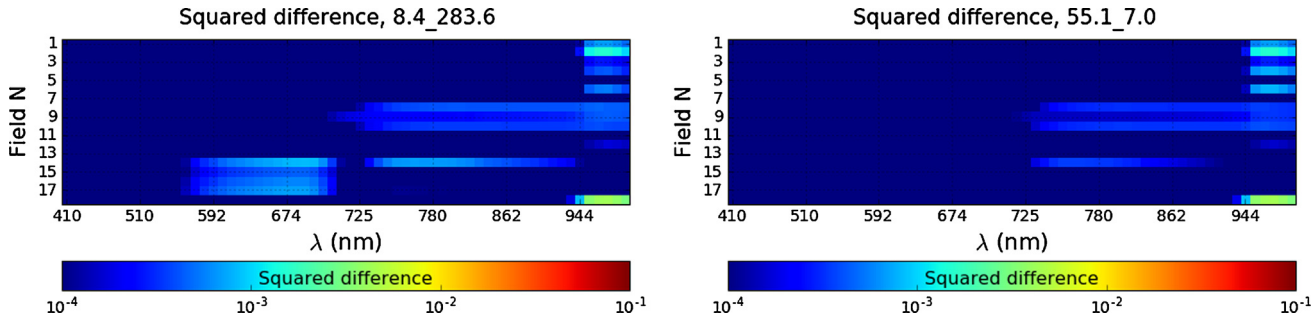

Fig. 16. Squared difference between synthetic CHRIS-Proba and MISR-only (nine cameras) retrieval over 18 fields for CHRIS-Proba camera VZA 8.4° (left hand side panel) and VZA −55.2° (right hand side panel).

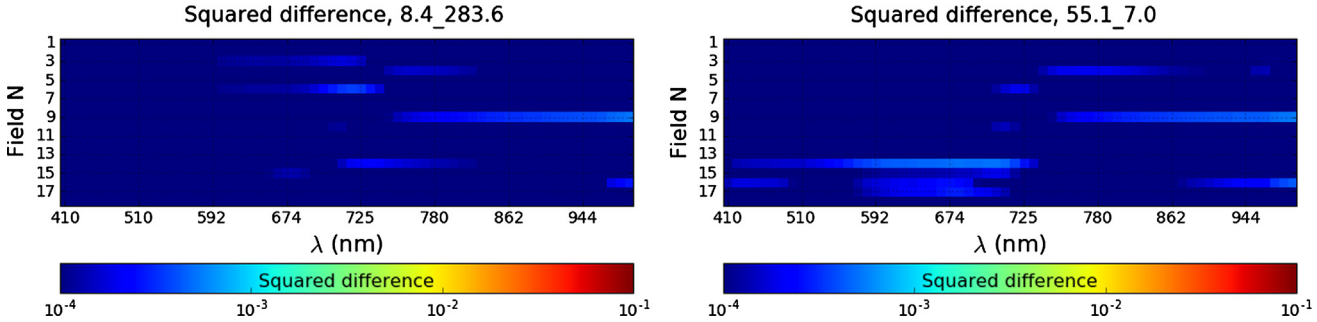

Fig. 17. Squared difference between synthetic CHRIS-Proba and Landsat-only retrieval over 18 fields for CHRIS-Proba camera VZA 8.4° (left hand side panel) and VZA −55.2° (right hand side panel).

3.4.1. Retrieval of biophysical parameters

Fig. 18 shows that retrieval of LAI based on the MISR-only information underestimates higher LAI values, resulting in bias, lower r^2 (0.69) and high RMSE (1.42). In case of Landsat and MISR + Landsat LAI retrieval has better results with slightly lower RMSE and uncertainties for the blended sensors. Retrieval of C_{ab} and C_w does not have meaningful accuracy and mostly show results of the prior minimization. We can see it because many of the C_{ab} values for Landsat and MISR + Landsat cases equal to 29 μg/cm^2.

Fig. 18 shows that only retrieval of LAI is acceptable because for C_{ab} and C_w r^2 is low and RMSE is high. When we compare the LAI retrieved from Landsat data alone and then MISR + Landsat data, with the "true" LAI the correlation is high and the slope is close to unity. The RMSE (Landsat, Landsat + MISR) is around 0.8 LAI units or lower, which is in line with a number of global products (Garrigues et al., 2008). In Landsat-only case, the results of LAI show that the retrieved LAI tends to overestimate the true LAI for high LAI values, which is explained by the saturation of reflectance for high LAI, and thus is a reduced sensitivity effect. The retrieval of C_{ab} and C_w, suggesting a low sensitivity of the real observations to these parameters. Discrepancy between results of this section and synthetic data section is due to idealized synthetic datasets which are aimed to show potential retrieval of parameters.

3.4.2. Simulation of the CHRIS/Proba signal

Results of the Section 3.4.1 show that among three retrieved parameters (LAI, C_w and C_{ab}) only the retrieved LAI error is less than 20%. This means that with given data we know 'correct' values only for three from ten parameters (LAI, s_1 and s_2). The goal of this section is to test whether it is possible to simulate CHRIS/Proba signal if the only known parameters are LAI and the soil coefficients. If so, this would demonstrate the power and flexibility of the reconstruction process.

Fig. 19 illustrates simulation of hyperspectral information using MISR-only data over the Alfalfa A1 field. RMSE between actual and reconstructed CHRIS/Proba spectra ranges from 0.0920 to 0.0985. There is no distinct trend of RMSE/uncertainty with varying the number of MISR cameras.

Unlike the synthetic simulation there is a relatively high discrepancy between simulated and real spectra in the NIR region. The reason for this maybe the much lower spatial resolution of MISR that results in spectrally mixed pixels. This probably means that the procedure with Eq. (2) cannot completely separate the spectra. However, this was expected since Eq. (2) is only a first guess assumption.

Fig. 20 shows results for Landsat-only case. RMSE is lower mainly due to better correspondence in NIR region. This can be explained by higher spatial resolution and better spectral coverage. However, some discrepancy in NIR and red regions is still noticeable.

The next figure illustrates combination of real Landsat and MISR data. Fig. 21 shows that when a combination of the two sensors is used, MISR pulls NIR part of the curve down resulting in lower RMSE. The more cameras we add the more deviation from the actual CHRIS/Proba. However, we can notice better correspondence in visible part. The possible reason for higher error in NIR part is unknown real per band uncertainties.

Fig. 22 summarizes results based on the real MISR-only data over 18 fields. Results for fields with higher LAI (Table 1) exhibit higher squared difference especially in the NIR region. Meanwhile RMSE for fields with lower LAI has lower values. Low squared difference is especially noticeable for the garlic fields 7–9. This result is expected since garlic fields have LAI of 0.63 leading to a very strong signal from soil, which is well described (Section 3.1). Difference between simulations of different CHRIS/Proba cameras is minimal except higher difference is red region for the garlic fields. RMSE is in the range from 0.007 to 0.2 where the lowest RMSE belongs to field seven (garlic).

Fig. 23 displays results of hyperspectral simulations using Landsat-only data. We can see that in general, values of squared difference are lower for all fields, corresponding to the difference between Figs. 19 and 20. However, the difference is slightly higher in NIR region in case of low LAI (fields 7–9). RMSE is in the range from 0.01 to 0.1.

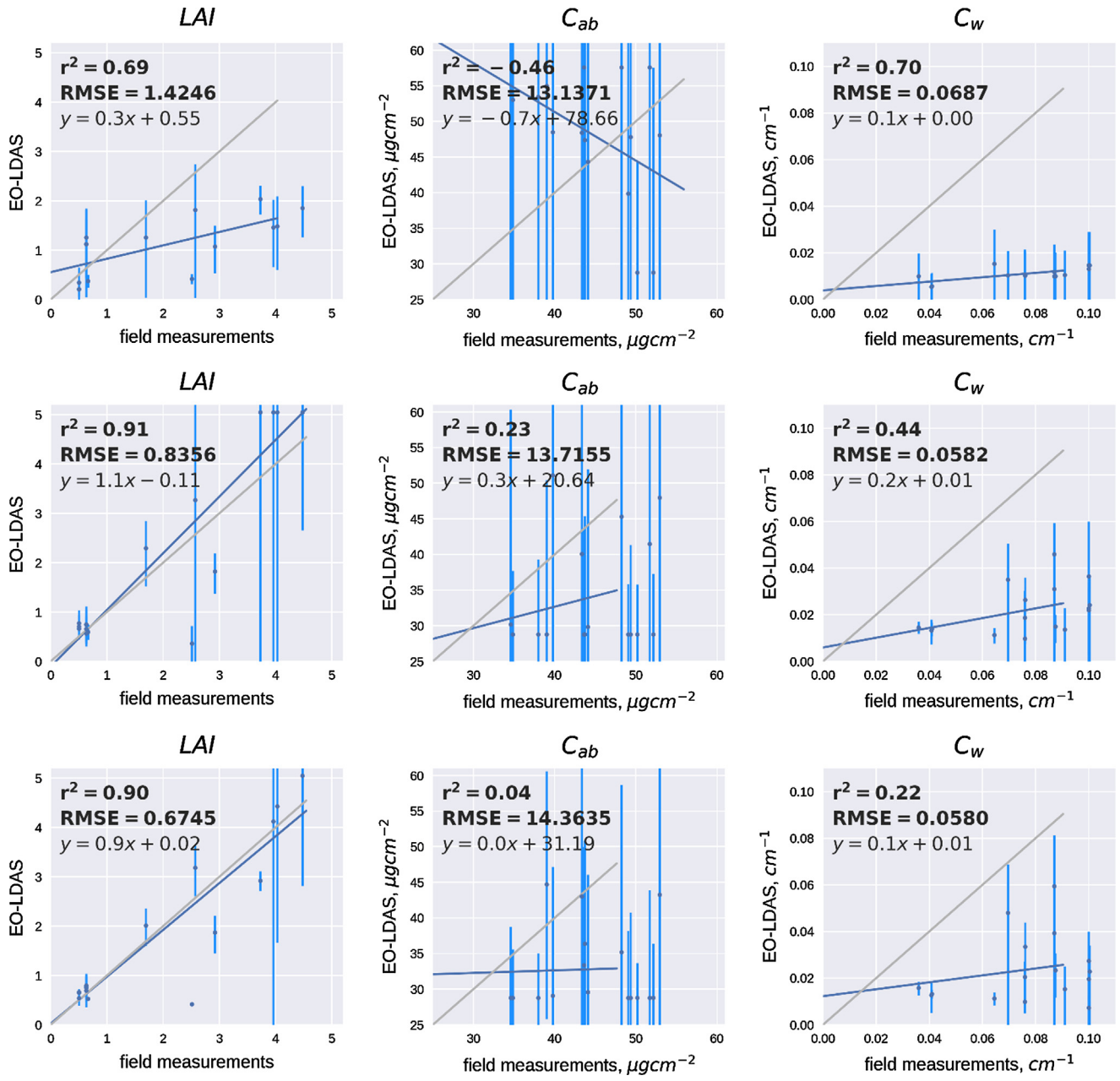

Fig. 18. Comparison of field measurements and EO-LDAS retrievals of LAI, C_{ab} and C_w with real data. MISR only (top panels), Landsat only (middle panels) and MISR + Landsat (bottom panels). MISR nadir camera (An).

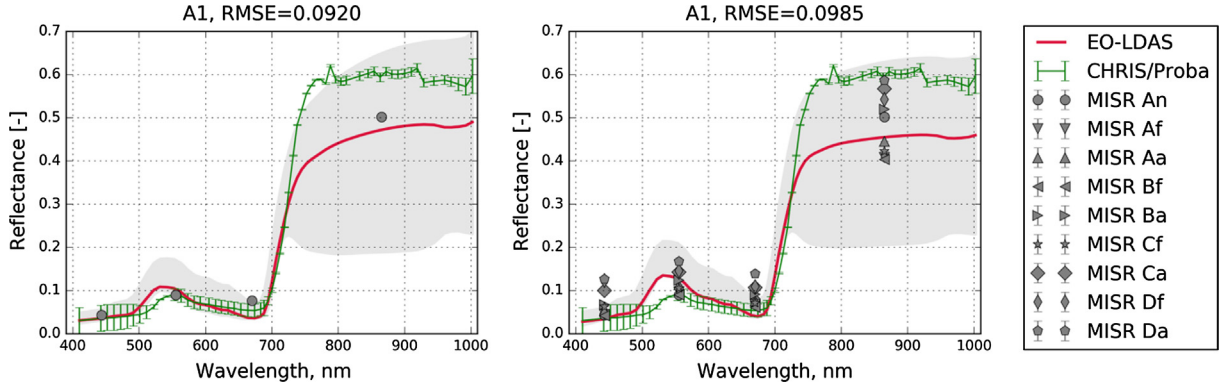

Fig. 19. Comparison of CHRIS actual (green line) for NADIR view and retrieved reflectance spectra (red line) over the Alfalfa field. Retrieval was done using actual MISR data only by increasing the number of cameras from one to nine. Left hand side panel is for An (nadir) camera only, right hand side panel for nine cameras. Grey surface indicates the posterior uncertainties. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

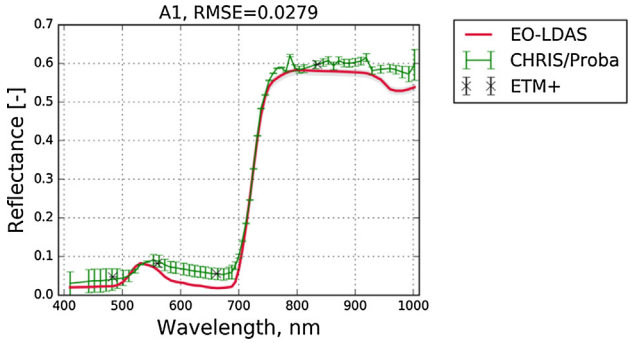

Fig. 20. Comparison of CHRIS actual (green line) for NADIR view and retrieved reflectance spectra (red line) over the Alfalfa field. Retrieval was done using actual Landsat-only data. Grey surface indicates the posterior uncertainties. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 24 shows squared difference for MISR + Landsat. These results correspond to Fig. 21 i.e. small difference from the Landsat-only results. However, unlike the difference between Figs. 20 and 21 there is no higher difference in NIR region except for field A1. RMSE is in the range from 0.02 to 0.1. This tells that in blending the two sensors and using nine MISR cameras, Landsat plays the leading role. However, with adding more cameras results can deviate from Landsat-only decisions according to the results shown on Fig. 21. One other thing to note is that for both Figs. 23 and 24 the differences for CHRIS/Proba VZA 55.2° have slightly higher values in red region for garlic fields.

Note that the fields 2 and 11 on Figs. 23 and 24 have no data because they correspond to SLC-OFF pixels.

3.5. On uncertainty estimation

Realistic uncertainty is one of the key factors to combining observations from different sensors. We can do an experiment assuming that uncertainties of the each MISR band is nine times higher than Landsat ones. This setup makes a balance between two sensors when we use all nine MISR cameras. Results are displayed on Fig. 25. We can see that if we have only one MISR camera the result is almost the same as for Landsat-only case (Fig. 20) but with

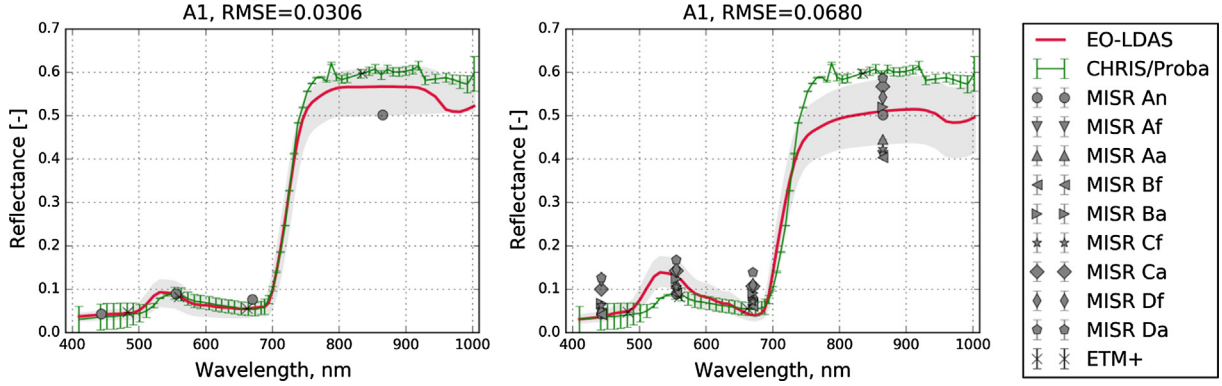

Fig. 21. Comparison of CHRIS actual (green line) for NADIR view and retrieved reflectance spectra (red line) over the Alfalfa field. Retrieval was done using actual MISR data and Landsat by increasing the number of cameras from one to nine. Left panel is for An (nadir) camera only, right panel for nine cameras. Grey surface indicates the posterior uncertainties. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

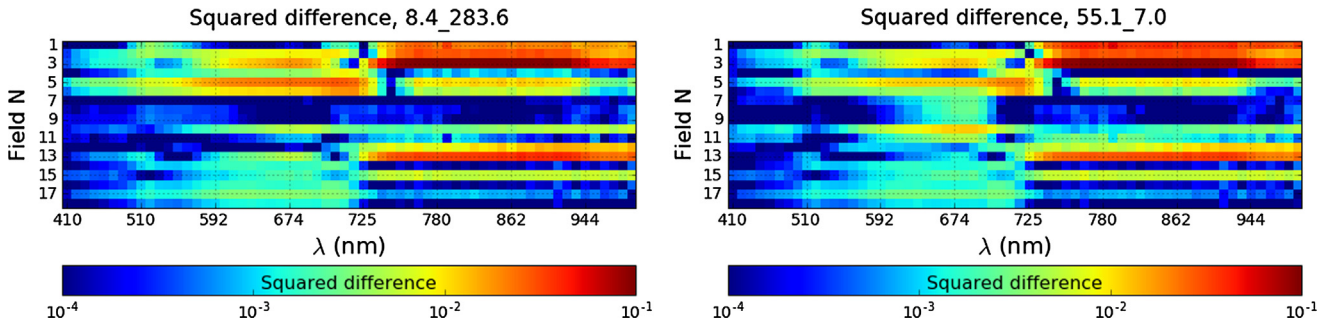

Fig. 22. Squared difference between real CHRIS-Proba and MISR-only (nine cameras) retrieval over 18 fields for CHRIS-Proba camera VZA 8.4° (left hand side panel) and VZA −55.2° (right hand side).

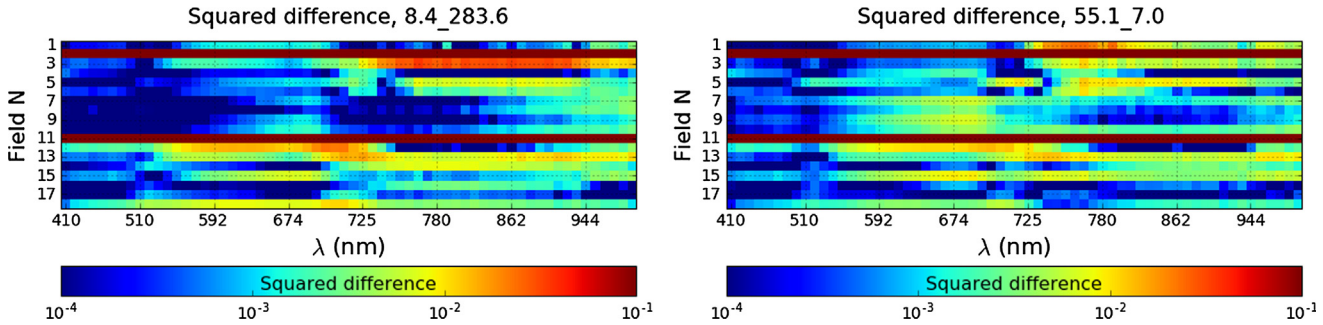

Fig. 23. Squared difference between real CHRIS-Proba and Landsat-only retrieval over 18 fields for CHRIS-Proba camera VZA 8.4° (left hand side panel) and VZA −55.2° (right hand side).

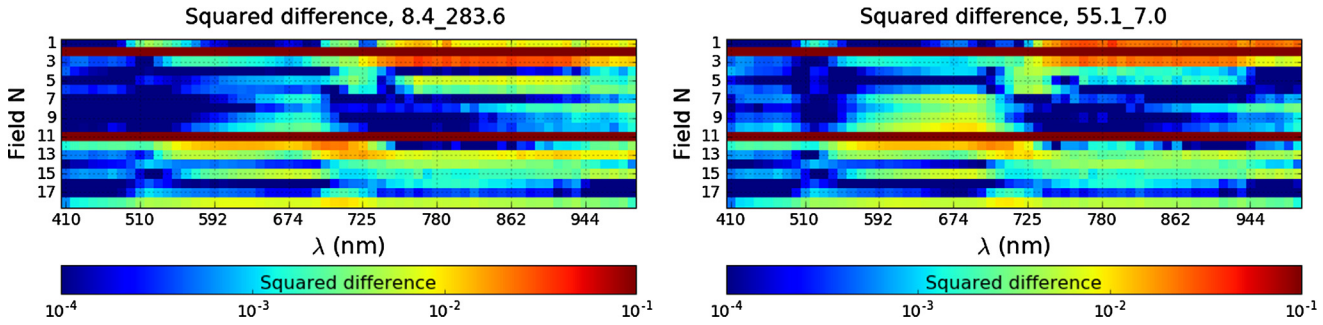

Fig. 24. Squared difference between real CHRIS-Proba and MISR + Landsat (nine cameras) retrieval over 18 fields for CHRIS-Proba camera VZA 8.4° (left hand side panel) and VZA −55.2° (right hand side).

five cameras RMSE is lower than on both Figs. 20 and 21. Due to the fact that uncertainties of each MISR band are much lower than Landsat ones, adding more than one MISR cameras does not pull simulated Proba-curve down in NIR region like on Fig. 25.

3.6. An experiment with fixed LAI

We can make an experiment when we fix LAI to the ground truth value and s_1/s_2 to their prior values. Then we use LHS to generate 100 random LUTs with seven unknown parameters with 7000 entries in each table. We then find best fit to the real CHRIS/Proba data for each of these tables. Results show low RMSE in the range between 0.004 and 0.02. Here we show an example of Alfalfa and garlic fields (Fig. 26 and Table 5 and 6). Tables 5 and 6 show that a broad variation in land surface parameters still results in a prediction very much in agreement with the observations. We can see it as a depiction of ill-posedness, where there is no unique solution, but rather a wide range of values fit the observations equally well. However, taking into account that LAI and soil parameters are fixed we can say that this is in line with Mousivand et al., 2015 and shows that in case of this data the number of the main controlling factors are even less and they are LAI and s_1/s_2.

Fig. 26 displays spectra with minimum and maximum RMSE for the Alfalfa and Garlic fields. We can see that

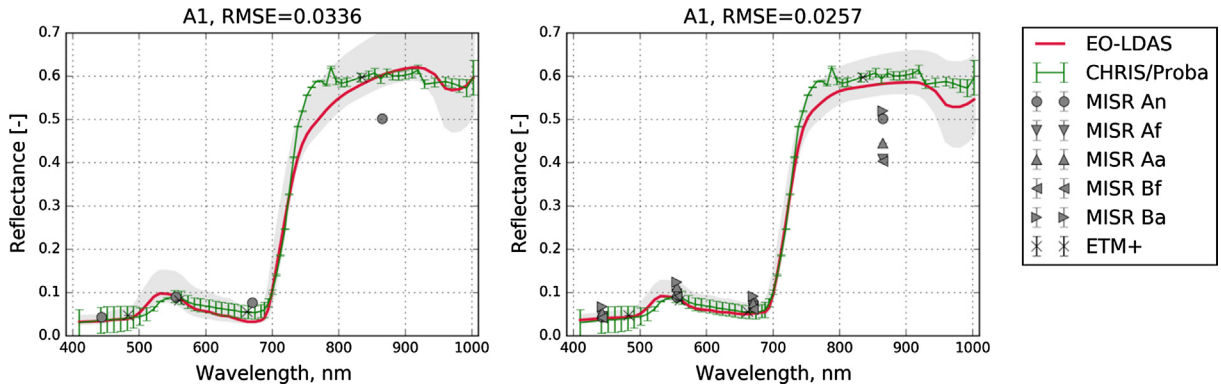

Fig. 25. Comparison of CHRIS real (green line) for nadir view and MISR + Landsat retrieved reflectance spectra (red line) over the Alfalfa field. Left hand side panel is for An (nadir) camera only, right hand side panel for nine cameras. Grey surface indicates the posterior uncertainties in 95% credible interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

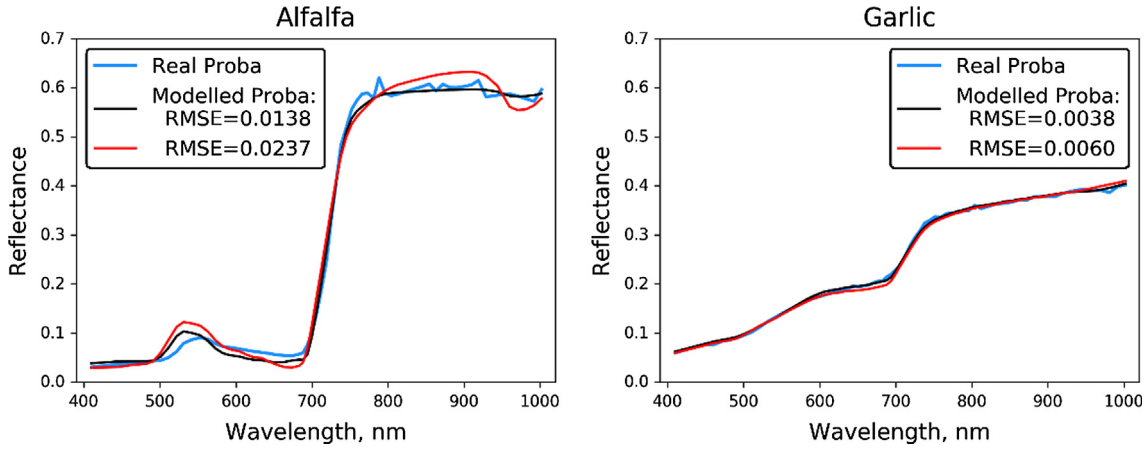

Fig. 26. Simulation of CHRIS/Proba spectra with fixed LAI and s1/s2 to the 'true' values. Left hand side panel – Alfalfa A1 field, right hand side panel – Garlic G1 field. Black lines represent spectra with minimum RMSE, red lines – maximum RMSE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Min and max values for fit with fixed LAI and soil model for the Alfalfa field.

	xhc (m)	rpl (m)	xkab ()	scen	xkw	xkm	xleafn
Min value	0.011	0.038	25.486	0.003	0.000	0.001	0.619
Max value	4.843	0.999	118.157	0.303	0.055	0.024	2.999

Table 6
Min and max values for fit with fixed LAI and soil model for the Garlic field.

	xhc (m)	rpl (m)	xkab ()	scen	xkw	xkm	xleafn
Min value	0.000	0.101	21.724	0.003	0.000	0.000	0.038
Max value	2.131	0.999	84.602	0.683	0.059	0.020	2.946

even in the case of maximum RMSE (0.018), the resulting simulated spectral curve is quite similar to real CHRIS/Proba data.

4. Discussion and conclusion

We propose a way of using the EO-LDAS data assimilation system to reproduce hyperspectral information (effectively arbitrary band combinations) with multi-sensor and multi-angular constraints, using a restricted number of known biophysical parameters derived from by multi-spectral sensors. We have demonstrated the performance of EO-LDAS on MISR and Landsat sensors but the method can be extended to other optical multispectral sensors. We have shown that EO-LDAS provides an optimal way of combining spectral and angular

observations for a two-sensor combination. It was found that increased amount of multiangular observations (provided by MISR) only helps to improve accuracy (RMSE) in the case of the synthetic simulations, and only then for C_{ab}. Calculation of the cross entropy for increasing numbers of the MISR cameras does not show that multiangular observations always provide more information than the same number of mono-angular nadir observations (Fig. 12). This suggests that, at least in case of this experimental setup, the number of observations is of greater significance than different viewing angle positions.

One important note is that per-band uncertainties for both MISR and Landsat surface directional reflectances are not provided. Estimation of these uncertainties will require propagation of the sensor thermal noise, as well as atmospheric composition, through the atmospheric correction component, not a trivial task, and outside the scope of this study.

However, it should also be kept in mind that these are not "real" uncertainties, but is an approximation to the linear part of the Hessian, with higher order terms ignored (Lewis et al., 2012). Additionally, it is assumed that the posterior is Gaussian, and that there is no model error. It can be seen as estimation of steepness of curvature in the decision space, which measures how many possible solutions we have at the end of optimization process. It is not always related to "real" uncertainty, which is very hard to estimate. For example, the process of optimization can be trapped in local minima and in this case posterior uncertainty reflects the situation around this local minimum. However, in many cases posterior uncertainty does reflect the estimation of accuracy, as shown for example by Thacker (1989).

From the above, it is clear that an accurate description of the statistical properties of the inputs (here, surface directional reflectance), is required to produce an accurate estimate of the posterior distribution. It is also important to consider potential biases between observations from different sensors (although the biases can be considered part of the observational uncertainty).

The example from Section 3.5 is an attempt to show that realistic prior uncertainties is an essential factor in merging different sources of information and potentially better results can be achieved if EO data providers can include information about uncertainties to reflectance products.

The real data results suggest that proper estimation of LAI and soil parameters could be enough for simulation of hyperspectral signal between 400 and 1000 nm with RMSE ≤ 0.03. This result is in line with Verrelst et al. (2016) and Mousivand et al. (2015), who both show that LAI has a stronger influence in all bands than most of other parameters. Indeed, the example from the Section 3.6 supports the hypothesis that most of variations of hyperspectral data in visible and infrared ranges are controlled by just a few mechanisms. In case of the data of this work it is mainly LAI and soil parameters.

These comparisons show that EO-LDAS can be used for simulation of hyperspectral data by observations (retrievals) from a multispectral sensor or by a combination of multispectral sensors. In addition, results show that accuracy of retrieved LAI and accuracy of simulated hyperspectral information are consistent. Finally, results demonstrate consistent and physically based technique for merging information from two or more multispectral sensors.

The simulation of one sensor signal by another sensor, which we have demonstrate here, is likely to be useful for the intercalibration of small EO satellites. 92 small satellites were launched in 2013, more than 140 in 2014 and more than 500 expected in 2015–2019 (Messier, 2015). Many of these satellites are designed for EO and can provide quite high spatial resolution data from 10 to 30 m. Large numbers of these satellites can provide both high revisiting rate and fine resolution. However, to combine information from all these sensors they have to be properly intercalibrated. The method we present here is a possibility to intercalibrate small EO satellites in order to get continuous time series of surface reflectance, across programmes and sensors of different designs.

An important issue, which we do not consider here, is how to propagate *a priori* uncertainties through the preprocessing chain: i.e. calibration, georeferencing, atmospheric correction, dependency on spatial resolution etc. So further efforts are required in this area because this is likely to provide a better balance between the cost function terms: models, radiometric information and prior information and can significantly improve results.

Acknowledgements

Authors gratefully acknowledges financial support from the BACI project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640176. We gratefully acknowledge the GIONET project, funded by the European Commission, Marie Curie Programme Initial Training Network, Grant Agreement number PITN-GA-2010- 264509. We also acknowledge MULTIPLY project (Project No. 687320). We acknowledge 40th COSPAR organizing committee for travel financial support; U.S. Geological Survey (USGS) and National Center for Earth Resource Observations and Science (EROS) for providing Landsat data through the Earth Explorer system (http://earthexplorer.usgs.gov/). JRC authors thank Linda Hunt and Michel Verstraete for their technical support on MISR HR processing chain. This research would be impossible without work of JPL/NASA who created and launched into space MISR instrument.

J Gomez-Dans acknowledges support from the NERC NCEO, European Space Agency support under Contract 4000112388/14/I-NB, and the European Union's Horizon 2020 research and innovation programme under grant

agreement No 687320 for the EU H2020 MULTIPLY project.

M. Disney acknowledges support from the NERC NCEO, NERC Standard Grants NE/N00373X/1 and NE/ P011780/1 and the European Union's Horizon 2020 research and innovation programme under grant agreement No 640176 for the EU H2020 BACI project.

References

Asner, G.P., Townsend, P., Martin, R.E., Chadwick, K. D., 2016. Forest biophysical and biochemical properties from hyperspectral and LiDAR remote sensing. In: Land Resources Monitoring, Modeling, and Mapping with Remote Sensing, pp. 429–448.

Baccini, A., Friedl, M.A., Woodcock, C.E., Zhu, Z., 2007. Scaling field data to calibrate and validate moderate spatial resolution remote sensing models. Photogramm. Eng. Remote Sens. 73 (8), 945–954.

Baret, F., Morissette, J.T., Fernandes, R.A., Champeaux, J.L., Myneni, R.B., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., Nickeson, J.E., 2006. Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: proposition of the CEOS-BELMANIP. IEEE Trans. Geosci. Remote Sens. 44 (7), 1794–1803.

Barnsley, M.J., Settle, J.J., Cutter, M., Lobb, D., Teston, F., 2004. The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. IEEE Trans. Geosci. Remote Sens. 42 (7), 1512–1520.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bull. Amer. Meteor. Soc. 95, 1431–1443.

Bunnik, N.J.J., 1978. The multispectral reflectance of shortwave radiation of agricultural crops in relation with their morphological and optical properties, technical report, MededelingenLandbouwhogeschool, Wageningen, Netherlands.

Candela, L., Formaro, R., Guarini, R., Loizzo, R., Longo, F., Varacalli, G., 2016. The PRISMA mission. In: Proceedings of the IEEE IGARSS (International Geoscience and Remote Sensing Symposium) Conference, Beijing, China, July 10–15, 2016.

Garrigues, S., Lacaze, R., Baret, F., Morisette, J.T., Weiss, M., Nickeson, J.E., Fernandes, R., Plummer, S., Shabanov, N.V., Myneni, R.B., Knyazikhin, Y., Yang, W., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. J. Geophys. Res. Biogeosciences 113 (G2), 2156–2202.

Chernetskiy, M., Gomez-Dans, J., Gobron, N., Morgan, O., Lewis, P., Truckenbrodt, S., Schmullius, C., 2017. Estimation of FAPAR over croplands using MISR data and the earth observation land data assimilation system (EO-LDAS). Remote Sens. 9 (7), 656.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. Remote Sens. Environ 84 (1), 1–15.

Diner, D.J., Beckert, J.C., Reilly, T.H., Bruegge, C.J., Conel, J.E., Kahn, R.A., Martonchik, J.V., Ackerman, T.P., Davies, R., Gerstl, S.A.W., Gordon, H.R., Muller, J.-P., Myneni, R.B., Sellers, P.J., Pinty, B., Verstraete, M.M., 1998. Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. IEEE Trans. Geosci. Remote Sens. 36 (4), 1072–1087.

Disney, M., Muller, J.-P., Kharbouche, S., Kaminski, T., Voßbeck, M., Lewis, P., et al., 2016. A new global fAPAR and LAI dataset derived from optimal albedo estimates: comparison with MODIS products. Remote Sens. 8, 275.

ESA: SPARC Data Acquisition Report., 2004. Contract no: 18307/04/NL/FF, University of Valencia [Online], Available: http://earth.esa.int/campaigns/.

Feret, J.B., Francois, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., et al., 2008. (2008) PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112 (6), 3030–3043.

Gandia, S., Fernández, G., García, J.C., Moreno, J. 2004. Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign. In: Proc. 2nd CHRIS/Proba Workshop, ESA/ESRIN, Frascati, Italy.

Gobron, N., Pinty, B., Verstraete, M.M., Govaerts, Y., 1997. A semi-discrete model for the scattering of light by vegetation. J. Geophys. Res. 102, 9431–9446.

Gobron, N., Pinty, B., Verstraete, M.M., Martonchik, J.V., Knyazikhin, Y., Diner, D.J., 2000. Potential of multiangular spectral measurements to characterize land surfaces: conceptual approach and exploratory application. J. Geophys. Res. 105 (D13), 17539.

Gomez-Dans, J., Lewis, P., 2012. Earth Observation Land Data Assimilation System (EO-LDAS), Available online: https://github.com/jgomezdans/eoldas.

Gomez-Dans, J., Lewis, P., Disney, M., 2015. Efficient emulation of radiative transfer codes using gaussian processes and application to land surface parameter inferences. Remote Sens. 8 (2), 119.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 1998. The EnMAPSpaceborne imaging spectroscopy mission for earth observation. Remote Sens. 7 (7), 8830–8857.

Gupta, R.K., Prasad, S., Krishna Rao, P.V., 1998. Evaluation of spatial upscaling algorithms for different land cover types. Adv. Space Res. 22 (5), 625–628.

Hay, G.J., Niemann, K., Goodenough, D.G., 1997. Spatial thresholds, image-objects, and upscaling: a multiscale evaluation. Remote Sens. Environ. 62, 1–19.

Hay, G.J., Marceau, D.J., Dub, P., 2001. A multiscale framework for landscape analysis: object-specific analysis and upscaling. Landscape Ecol. 16, 471–490.

Hollmann, R., Merchant, C.J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R., Holzer-Popp, T., Paul, F., Sandven, S., Sathyendranath, S., van Roozendael, M., Wagner, W., 2013. The ESA climate change initiative: satellite data records for essential climate variables. Bull. Amer. Meteor. Soc. 94, 1541–1552.

Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. Remote Sens. Environ. 34 (2), 75–91.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., Ustin, S.L., 2009. PROSPECT+SAIL models: a review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66.

Kimes, D.S., Knyazikhin, Y., Privette, J.L., Abuelgasim, A.A., Gao, F., 2000. Inversion methods for physically-based models. Remote Sens. Rev. 18, 381–439.

Knyazikhin, Y., Martonchik, J.V., Myneni, R.B., Diner, D.J., Running, S. W., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. J. Geophys. Res. 103 (D24), 32257.

Koetz, B., Baret, F., Poilvé, H., Hill, J., 2005. Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. Remote Sens. Environ. 95, 115–124.

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D. J., Middleton, E.M., 2015. An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. Remote Sens. Environ. 167, 6–19.

Lewis, P., Gómez-Dans, J., Kaminski, T., Settle, J., Quaife, T., Gobron, N., Styles, J., Berger, M., 2012. An earth observation land data assimilation system (EO-LDAS). Remote Sens. Environ. 120, 219–235.

Maiersperger, T.K., Scaramuzza, P.L., Leigh, L., Shrestha, S., Gallo, K. P., Jenkerson, C.B., Dwyer, J.L., 2013. Characterizing LEDAPS surface reflectance products by comparisons with AERONET, field spectrometer, and MODIS data. Remote Sens. Environ. 136, 1–13.

Masek, J.G., Vermote, E.F., Saleous, N., Wolfe, R., Hall, F.G., Huemmrich, F., Gao, F., Kutler, J. and Lim, T.K., 2012. LEDAPS Landsat Calibration, Reflectance, Atmospheric Correction Preprocessing Code. Model product. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. http://dx.doi.or.

Messier, D. Euroconsult Sees Large Market for Smallsats,Parabolic Arc. Retrieved 8 March 2015.

Mousivand, A., Menenti, M., Gorte, B., Verhoef, W., 2015. Multitemporal, multi-sensor retrieval of terrestrial vegetation properties from spectral–directional radiometric data. Remote Sens. Environ. 158, 311–330.

Musavi, T., Migliavacca, M., van de Weg, M.J., Kattge, J., Wohlfahrt, G., van Bodegom, P.M., Reichstein, M., Bahn, M., Carrara, A., Domingues, T.F., Gavazzi, M., Gianelle, D., Gimeno, C., Granier, A., Gruening, C., Havránková, K., Herbst, M., Hrynkiw, C., Kalhori, A., Kaminski, T., Klumpp, K., Kolari, P., Longdoz, B., Minerbi, S., Montagnani, L., Moors, E., Oechel, W.C., Reich, P.B., Rohatyn, S., Rossi, A., Rotenberg, E., Varlagin, A., Wilkinson, M., Wirth, C., Mahecha, M.D., 2016. Potential and limitations of inferring ecosystem photosynthetic capacity from leaf functional traits. Ecol. Evol. 6, 7352–7366.

Pinty, B., Widlowski, J., Gobron, N., Verstraete, M.M., Diner, D.J., Member, A., 2002. Uniqueness of multiangular measurements — Part I : An Indicator of subpixel surface heterogeneity from MISR. IEEE Trans. Geosci. Remote Sens. 40 (7), 1560–1573.

Porcar-Castell, A., Mac Arthur, A., Rossini, M., Eklundh, L., Pacheco-Labrador, J., Anderson, K., et al., 2015. EUROSPEC: at the interface between remote-sensing and ecosystem CO_2 flux measurements in Europe. Biogeosciences 12, 6103–6124.

Pfeifer, M., Disney, M., Quaife, T., Marchant, R., 2012. Terrestrial ecosystems from space: a review of earth observation products for macroecology applications. Glob. Ecol. Biogeogr. 21, 603–624.

Price, J.C., 1990. On the information content of soil reflectance spectra. Remote Sens. Environ. 113–121.

Price, J.C., 1994. Band selection procedure for multispectral scanners. Appl. Opt. 33 (15), 3281–3288.

Price, J.C., 1998. An approach for analysis of reflectance spectra. Remote Sens. Environ. 64 (3), 316–330.

Roelofsen, H.D., van Bodegom, P.M., Kooistra, L., Witte, J.-P.M., 2014. Predicting leaf traits of herbaceous species from their spectral characteristics. Ecol. Evol. 4 (6), 706–719.

Quaife T, Lewis P, Disney MI, De Kauwe M, Williams M, Law B, et al., 2007. Assimilating reflectance data into an ecosystem model to improve estimates of terrestrial carbon flux.

Thacker, W.C., 1989. The role of the Hessian matrix in fitting models to measurements. Geophys. Res. 94 (C5), 6177–6196.

Sacks, J., Welch, W.J., Mitchell, T.J., Wynn, H.P., 1989. Design and analysis of computer experiments. Stat. Sci. 4, 409–423.

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 Sens. Environ. 109 (2), 166–182.

Verhoef, W., Bach, H., 2003. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. Remote Sens. Environ. 87 (1), 23–41.

Verrelst, J., Sabater, N., Rivera, J.P., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., Moreno, J., 2016. Emulation of leaf, canopy and atmosphere radiative transfer models for fast global sensitivity analysis. Remote Sens. 8 (8), 673.

Verstraete, M.M., Hunt, L.A., Scholes, R.J., Clerici, M., Pinty, B., Nelson, D.L., 2012. Generating 275-m resolution land surface products from the multi-angle imaging spectroradiometer data. IEEE Trans. Geosci. Remote Sens. 50 (10), 3980–3990.

Weiss, M., Baret, F., Myneni, R.B., Pragnère, A., Knyazikhin, Y., 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. Agronomie 20 (1), 3–22.

Widlowski, J.-L., Pinty, B., Gobron, N., Verstraete, M.M., Diner, D.J., Davis, A.B., 2004. Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. Clim. Change 67, 403–415.

Widlowski, J.L., Pinty, B., Lavergne, T., Verstraete, M.M., Gobron, N., 2005. Using 1-D models to interpret the reflectance anisotropy of 3-D canopy targets: issues and caveats. IEEE Trans. Geosci. Remote Sens. 43, 2008–2017.