Synergistic retrievals of leaf area index and soil moisture from Sentinel-1 and Sentinel-2

Article

Published Version

Open access

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).
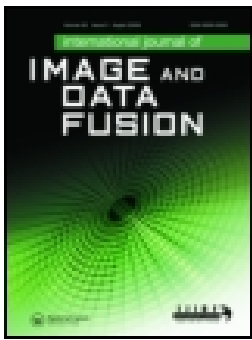
To link to this article DOI: http://dx.doi.org/10.1080/19479832.2022.2149629

Publisher: Taylor & Francis

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Synergistic retrievals of leaf area index and soil moisture from Sentinel-1 and Sentinel-2

T. Quaife, E. M. Pinnington, P. Marzahn, T. Kaminski, M. Vossbeck, J. Timmermans, C. Isola, B. Rommen & A. Loew

Published online: 01 Dec 2022.

Submit your article to this journal ⬚

Article views: 100

View related articles ⬚

View Crossmark data ⬚

Taylor & Francis
Taylor & Francis Group

Synergistic retrievals of leaf area index and soil moisture from Sentinel-1 and Sentinel-2

T. Quaife [a], E. M. Pinnington[a], P. Marzahn[b,c], T. Kaminski[d], M. Vossbeck[d], J. Timmermans[e,f,g], C. Isola[h], B. Rommen[h] and A. Loew[b]

aNational Centre for Earth Observation, University of Reading, Reading, UK; bDepartment of Geography, Ludwig-Maximilians-Universität München, Munich, Germany; cGeodesy and Geoinformatics, University of Rostock, Rostock, Germany; dThe Inversion Lab, Hamburg, Germany; eVirtual Laboratory & Innovations Centre, Lifewatch ERIC, Science, Amsterdam, The Netherlands; fInstitute for Biodiversity and Ecosystem Dynamics, Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands; gInstitute of Environmental Sciences, Leiden University, Leiden, The Netherlands; hEuropean Space Research and Technology Centre, Noordwijk, The Netherlands

ABSTRACT

Joint retrieval of vegetation status from synthetic aperture radar (SAR) and optical data holds much promise due to the complimentary of the information in the two wavelength domains. SAR penetrates the canopy and includes information about the water status of the soil and vegetation, whereas optical data contains information about the amount and health of leaves. However, due to inherent complexities of combining these data sources there has been relatively little progress in joint retrieval of information over vegetation canopies. In this study, data from Sentinel–1 and Sentinel–2 were used to invert coupled radiative transfer models to provide synergistic retrievals of leaf area index and soil moisture. Results for leaf area are excellent and enhanced by the use of both data sources (RSME is always less than 0.5 and has a correlation of better than 0.95 when using both together), but results for soil moisture are mixed with joint retrievals generally showing the lowest RMSE but underestimating the variability of the field data. Examples of such synergistic retrieval of plant properties from optical and SAR data using physically based radiative transfer models are uncommon in the literature, but these results highlight the potential for this approach.

1. Introduction

With the launch of the Sentinel–1 (Torres *et al.* 2012) and Sentinel–2 (Drusch *et al.* 2012) satellites, a new era in Earth Observation has started that allows for the development of novel approaches that use these observations at high temporal frequencies and fine spatial resolution with complementary information from the optical and microwave spectral domains. One important application for these data is the monitoring of crops (Torres *et al.* 2012).

CONTACT T. Quaife ✉ t.l.quaife@reading.ac.uk 🖃 National Centre for Earth Observation, University of Reading, Reading RG6 6ET, UK

Optical remote sensing of crops and crop yield has a long history, stretching back as nearly as far as the advent of satellite remote sensing itself (e.g. Idso *et al.* (1977), Tucker (1980)). Most early examples use vegetation indices (primarily NDVI – the Normalised Difference Vegetation Index) from AVHRR or Landsat data and calibrate empirical relationships with variables such as yield or leaf area. Indeed, even today the majority of studies using optical remote sensing data to monitor crops continue to focus on the use of vegetation indices. For example, NDVI for yield determination in soybean (Liu and Kogan 2002), the Perpendicular Vegetation Index (PVI) for determining absorbed radiation in cotton and wheat crops (Wiegand *et al.* 1986) and the Transformed Soil–Adjusted Vegetation Index (TSAVI) for determining green crop area in wheat (Broge and Mortensen 2002). Viña *et al.* (2011) compared a large number of vegetation indices to estimate the leaf area index of maize and soybean and found that those designed to be sensitive to chlorophyll concentration (e.g. the MERIS Terrestrial Chlorophyll Index, MTCI) gave the best predictions. Other vegetation indices analysed by Viña *et al.* (2011) were NDVI, the Enhanced Vegetation Index (EVI), the simple ratio (SR), the Green Atmospherically Resistant Vegetation Index (GARVI) and Wide-Dynamic Range Vegetation Index (WDRVI).

The advantage of these and similar techniques is that they are relatively straight forward to apply and computationally cheap to implement. In principle, they give reasonable results for the areas and cover types over which they were calibrated. However, their major shortcoming is that with little or no underlying physics they are not generalisable beyond the areas which they were calibrated for and applications on an appreciable scale are unlikely to yield accurate results. In particular, such techniques are not amenable to combining data from different sensors and different domains of the electromagnetic spectrum. To combine data from optical sensors such as Sentinel-2 with SAR sensors such as Sentinel-1 some form of physically based model is required.

Examples of crop remote sensing using physical models of radiative transfer (RT) in the optical domain are far less numerous than those using empirical techniques. One approach that has been used by a number of authors is to use a vegetation canopy radiative transfer model to calibrate the vegetation indices. Féret *et al.* (2011) used the PROSPECT-5 model to develop relationships between spectral indices and leaf pigment concentrations using partial least squares at the leaf level. The study included maize and soybean crops as well as a number of tree species. They showed that it is possible to construct simple polynomial relationships that describe a large amount of variability in the data. However, it is likely to be more complex when moving to a full canopy due to the influence of soil optical properties and the physical structure of the canopy. Punalekar *et al.* (2018) inverted the PROSAIL model against Sentinel–2A data to determine the leaf area index for several areas of pasture land in the UK, and from that inferred biomass. They used a look–up table based approach and demonstrated that it produced more accurate estimates than an empirical relationship calibrated against NDVI. They also noted that for operational management of pasture in the UK the frequent impact of cloud on the Sentinel–2 data would likely limit its usefulness. This is something that can be directly addressed using Sentinel–1 data.

The retrieval of biogeophysical parameters over agricultural landscapes also has a long tradition in microwave remote sensing. The general challenge is the separation of the different contributions of surface properties to the Synthetic Aperture Radar (SAR) signal.

These typically comprise components such as soil moisture, surface roughness, vegetation water content and vegetation structural effects and need to be disentangled during a retrieval process.

Unlike in the optical domain there has been a greater focus on physical modelling for SAR applications. Attempts to describe the backscattering from vegetation covered areas have been made since the late 1970s. They have evolved from the simple 'cloud' model of Attema and Ulaby (1978) to multilayered, multi-constituent models like the MIchigan MIcrowave Canopy Scattering Model (MIMICS) proposed by Ulaby and Elachi (1990) or the radiative transfer model of Karam *et al*. (1992). More complex radiative transfer (RT) models have been developed to take into account the 3-dimensional canopy structure (e.g. Floury, 1999; Martinez *et al*. (2000); Disney and Lewis (2003); Lewis *et al*. (2003); Bracaglia (1995)). Some examples of empirical indices from SAR data also exist, for example Kim and van Zyl (2001) proposed the Radar Vegetation Index (RVI), which is sensitive to the fresh biomass and vegetation water content (Kim and Won 2003). Mattia *et al*. (2003) used ENVISAT ASAR data for the retrieval of fresh biomass and LAI values over wheat fields by means of the HH/VV ratio and Satalino *et al*. (2006) used the same approach for the retrieval of LAI and compared results with LAI derived from in-situ data. They concluded that LAI can be retrieved with the same accuracy as with optical (MERIS) data. McNairn *et al*. (2012) used Radarsat2 data for the retrieval of LAI data and concluded that, similar to Mattia *et al*. (2003), the co–pol ratio as well as the HV/VV ratio and HV backscatter is a good estimator of LAI. However, with the availability of fully polarimetric data, they suggest the usage of H from the H, A, alpha decomposition as well as the volume component from the Freeman–Durden decomposition for the retrieval of LAI. Paloscia *et al*. (2012) used dual frequency X- and L-Band data for the retrieval of plant moisture content. Using an empirical model combining L-Band HH-pol and X-Band VV-pol backscatter values in dependency of the growth structure of the plants, they inverted plant moisture content within the expected range and accuracy. Usually vegetation biomass and LAI over agricultural fields is estimated from SAR data at high frequencies. However, from the SIR-X/C as well as multi-frequency airborne campaigns it has been observed that L-band backscattering exhibits a high sensitivity to vegetation biomass of crops characterised by large leaves (e.g. corn and sunflowers); whereas higher frequencies (C and X bands) showed good agreement with the development of plants with narrow leaves (e.g. wheat) (Ferrazzoli and Guerriero 1996, Mattia *et al*. 2003).

The advent of Sentinel–1 and Sentinel–2 in orbit at the same time provides an opportunity to develop new retrieval techniques to exploit the synergies between the microwave (specifically SAR) and optical domains. Because of a historical lack of contemporaneous observations in these domains the combination of SAR and optical instruments for crop monitoring is less common than the use of either one on its own but the potential of this combination to provide information to crop models is significant as the two types of observations contain very different and complementary information. Prévot *et al*. (2003) present one of the few examples; they used PROSAIL and the zero order SAR model of Attema and Ulaby (1978) to assimilate both optical and SAR data into the STICS crop model. A small number of parameters were retrieved: the sowing date, duration of maximum LAI, the field capacity of the first soil layer and parameters that control stem density and evaporation. Interestingly, the authors note that whilst the optical data improved the model outputs the SAR data did not improve the model performance.

They suggest that the main reason for this is that the relatively low volumes of SAR data available made it difficult to characterise the soil moisture correctly. There is clearly the potential to improve on this situation: with the advent of Sentinel–1 the amount of SAR data available will be much greater and newer data assimilation techniques are capable of dealing with a larger number of parameters.

Other examples using both SAR and optical data tend to involve some level of empirical calibration, which limits the generality of the solution. Hosseini *et al.* (2019), for example, used a neural network to estimate the biomass of maize from SAR (RADARSAT–1) and optical data (RAPIDEYE). In a first step the authors calibrated a water cloud model (for the SAR data) and several vegetation indices (for the optical data) against biomass observations. The neural network was then used to predict the optical biomass estimates using the SAR-based estimates as inputs. This approach improved the overall prediction of the biomass estimates compared to the individually calibrated models. It is very likely, however, that the results of this study will be site dependant because of the individual model calibrations.

This paper examines synergistic retrievals from Sentinel–1 and Sentinel–2 data by inverting coupled optical and microwave radiative transfer models. The retrieval of the actual state of the land surface is typically an undetermined problem; the number of observables is typically much smaller than the number of unknown parameters in both the optical and SAR domain. However, the SAR and optical data contain complementary information that can be exploited by advanced retrieval algorithms which helps to alleviate this problem.

2. Data and methods

2.1. Satellite data

Sentinel–1 and Sentinel–2 data were downloaded from the Sentinel data hub API (https://www.sentinel-hub.com/) for the sites described in Section 2.2. For Sentinel–2, Level–2a surface reflectance data were acquired, resampled to a common spatial resolution and co–registered with the Sentinel–1 data using the ESA SNAP toolbox. The surface reflectance data, as opposed to, say, a Level–1 product, is motivated by the need to eliminate the influence of the atmosphere on the data, which will introduce error into the retrieval algorithm. Cloudy pixels were eliminated based on the Level–2a internal cloud mask (European Space Agency, E. 2015). No further pre–processing was carried out on the Sentinel–2 data.

Additional pre-processing steps for the original single look complex (SLC), Level–1 Sentinel–1 data, all using the SNAP toolbox, included thermal noise removal (to remove thermal noise in the data), radiometric correction (to correct the backscatter values for the local terrain), speckle filtering using a combined spatial and multi-temporal Lee filter (to reduce the impact of speckle), final sigma $_0$ conversion and backscatter normalisation to account for changes in incidence angles across the swath (Weiß *et al.* 2020). The objective of applying these steps is to produce backscatter values that are consistent with those predicted by the RT model and hence reduce errors in the retrieved soil moisture and LAI. Following preliminary experiments (results not shown), the input to the retrieval scheme was selected to be the HV polarisation from Sentinel–1 data and for Sentinel–2 bands 4

(665 nm), 5 (705 nm), 6 (740 nm), 7 (783 nm) and 8 (842 nm) were selected, where the wavelength indicated represents the centre of the band–pass function. No increase skill was observed from including additional optical bands or SAR polarisations.

2.2. Field data

Field data were acquired as part of the Munich North Isar (MNI) campaign which takes place in farmland to the north of the city of Munich (Germany). The MNI fields used in this study are labelled 508, which is growing wheat, 515 which is growing maize and 542 which is growing triticale (a hybrid of wheat and rye). The campaign started end of March 2017 and continued until harvesting of the corresponding crops. Measurements of leaf area index (LAI) were acquired by sampling across the fields and then averaging to provide a per–field LAI. Soil moisture measurements were carried out at the sampling locations indicated in Figure 1. The sampling points within each field are referred to in this manuscript as 'low', 'mid' and 'high' representing their positions from South to North within the field. The figure shows the fields using the NDVI from Sentinel–2 to highlight the vegetation, and calculated as $(B7 - B4)/(B7 + B4)$, where $B4$ and $B7$ are bands 4 and 7 described in the previous section. LAI was measured using a Li–Corr LAI2000 instrument. Each sampling location is equipped with a data logger (Decagon EM50) and probes that measure the soil moisture dynamics (Decagon 5TM) in 5 cm, 10 cm and 30 cm depths with two repetitions each.

2.3. Radiative transfer models

2.3.1. Optical radiative transfer

The semi–discrete model of Gobron *et al.* (1997) represents canopy reflectance via the addition of three terms:

$$\rho_{canopy} = \rho^1(z_0, \Omega, \Omega_0) + \rho^0(z_0, \Omega, \Omega_0) + \rho^M(z_0, \mu, \mu_0) \quad (1)$$

where ρ_{canopy} is the canopy reflectance, ρ^1 is the reflectance due to photons that have had a single interaction with the canopy, ρ^0 is the reflectance due to photons that have only interacted with the soil and ρ^M is the reflectance due to photons that have had multiple interactions within the canopy, possibly including the soil. A hotspot term is included in ρ^1 and ρ^0 to account for the enhanced probability of a photon exiting the canopy if it leaves along the same or similar path via which it entered. Leaf optical properties are prescribed using the PROSPECT model (Jacquemoud and Baret 1990) and soil optical properties via the model of Price (1990). PROSPECT parameters were held constant at typical values for healthy cereal crops. In addition, a very simple model of the impact of soil moisture on the first component of the Price reflectance spectra (Price 1990) was included such that

$$\rho_{s1}^* = \rho_{s1}(1 - \alpha\theta_m), \quad (2)$$

where ρ_{s1} is the first orthogonal component reflectance spectra from the Price model, θ_m is the soil moisture (m^3/m^3), ρ_{s1}^* is the modified reflectance spectra and α is a term that weights the influence of the soil moisture on the reflectance. The effect of this modification is a first-order approximation to make the soil darker when it is wetter.
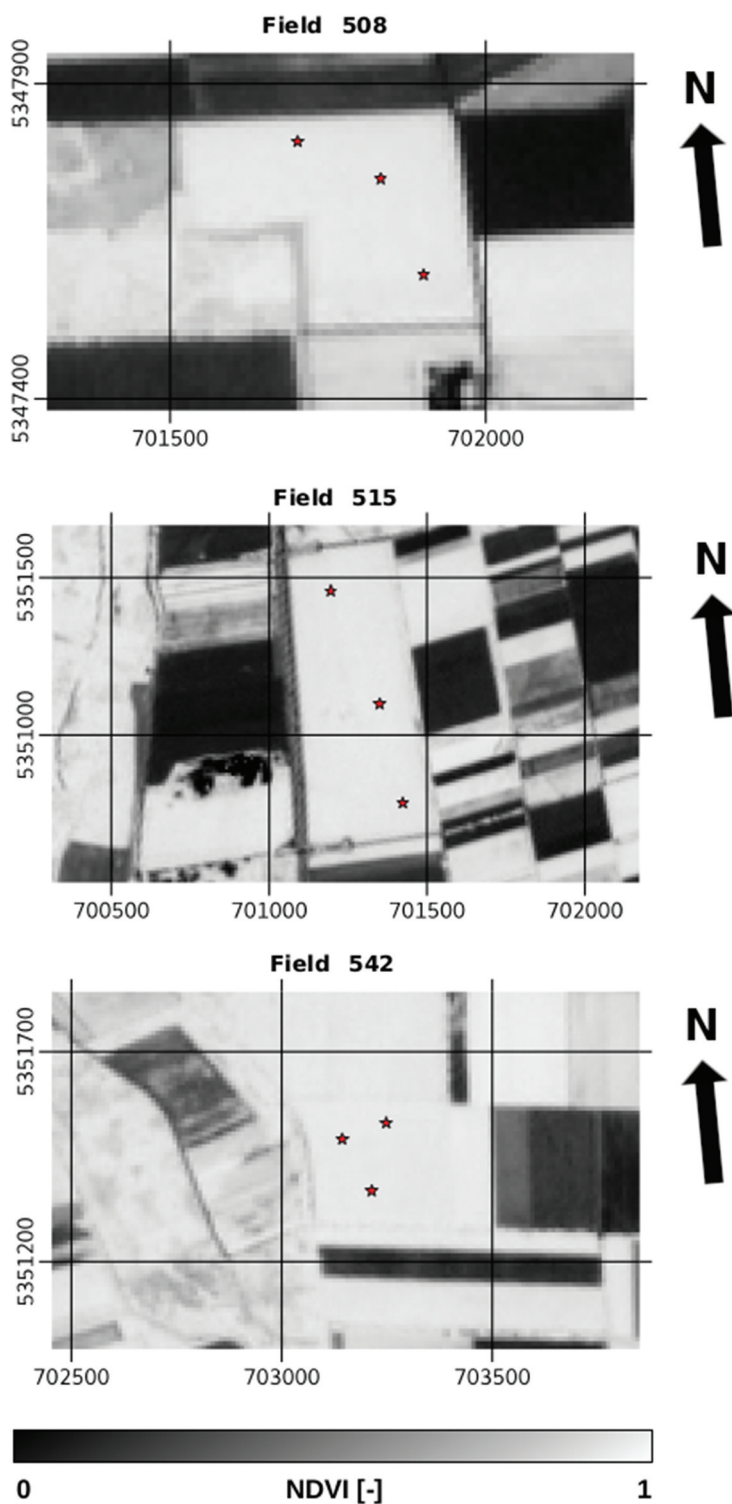
Figure 1. Sentinel–2 NDVI images of the three fields used in this study. From left to right they are growing wheat, maize and triticale. The images were acquired on the 26th of June 2017. The grid lines mark out the Eastings and Northings of the UTM grid with units of metres.

To convert from the top–of–canopy reflectance spectra computed by the semi–discrete model to the wave bands of Sentinel–2 the model outputs were convolved with the spectral response functions of the instrument.

2.3.2. Microwave radiative transfer

For the microwave domain, the semi–empirical single scattering radiative transfer model developed by De Roo $et\ al.$ (2001) was used to model the backscatter values of Sentinel–1. The total single-scattering pq–polarised backscattering coefficient (m^2/m^2) is expressed by the following equation:

$$\sigma^0_{pq} = \sigma^0_{g_{pq}} + \sigma^0_{c_{pq}} + \sigma^0_{gcg_{pq}} + \sigma^0_{cg_{pq}} \tag{3}$$

where each component represents a different scattering mechanism. For a particular polarisation configuration, these components are: $\sigma^0_{g_{pq}}$, the direct backscatter contribution of the underlying soil surface (including two-way attenuation by the canopy); $\sigma^0_{c_{pq}}$, direct backscatter contribution from the canopy; $\sigma^0_{gcg_{pq}}$, ground–canopy–ground scattering contribution and; $\sigma^0_{cg_{pq}}$, the combined ground–canopy and canopy–ground scattering contribution. For the soil backscattering term, $\sigma^0_{g_{pq}}$, the model of Oh $et\ al.$ (1992) is used. The other components are modelled as a uniform water cloud following Attema and Ulaby (1978).

2.4. Retrieval algorithm

To link a satellite observation to the state of the land surface, s_i, at acquisition time t_i a forward model or observation operator (Kaminski and Mathieu 2017) is needed that simulates the satellite observations given s_i. In the current study, the observation operators for Sentinel–1 and Sentinel–2 are provided by the models described in sections 2.3.2 and 2.3.1 respectively. The observation operators are denoted by $H_b(s_i, x_b, g_b)$ and $H_o(s_i, x_o, g_o)$, using the index b for backscatter and o for optical. The vectors x_b and x_o denote uncertain parameters in the formulation of the respective observation operators while g_b and g_o are known variables, for example, sun–sensor geometry. The retrieval is formulated by concatenating the unknown sequence of states s_i at all times t_i into one long vector s, the state trajectory. A dynamic model is also introduced that will help constrain the solution. If the model was perfect, M would fulfill the equation

$$0 = M(s) - s \tag{4}$$

The above four pieces of information are represented by a probability density function (PDF), with respective means b (for backscatter), o (for optical), p (prior), and m (for the deviation of $M(s)$ from s) and respective covariance matrices $C(b)$, $C(o)$, $C(p)$, and $C(m)$ which quantify the respective uncertainties. C_b and C_o include the uncertainty that arises from errors in the respective observation operators that cannot be corrected for by perfect state and parameter values.

For a compact notation all unknown quantities are assembled, that is, the trajectory s and the parameters in the observation operators x_b and x_o, into one long vector $\tilde{x}$ called

the control vector. The retrieval is then achieved by finding an $\tilde{x}$ that minimises the cost function

$$J(\tilde{x}) = J_b(\tilde{x}) + J_o(\tilde{x}) + J_p(\tilde{x}) + J_m(\tilde{x}), \tag{5}$$

which is the sum of four contributions, one dedicated to each piece of information:

$$J_b(\tilde{x}) = \frac{1}{2} \sum_{i=1,n_b} (H_b(\tilde{s}_i, \tilde{x}_b, g_b) - b_i)^T C(b)^{-1} (H_b(\tilde{s}_i, \tilde{x}_b, g_b) - b_i), \tag{6}$$

where n_b is the number of S1 acquisitions;

$$J_o(\tilde{x}) = \frac{1}{2} \sum_{i=1,n_o} (H_o(\tilde{s}_i, \tilde{x}_o, g_o) - o_i)^T C(o)^{-1} (H_o(\tilde{s}_i, \tilde{x}_o, g_o) - o_i), \tag{7}$$

where n_o is the number of S2 acquisitions;

$$J_p(\tilde{x}) = \frac{1}{2} (\tilde{x} - p)^T C(p)^{-1} (\tilde{x} - p); \tag{8}$$

$$J_m(\tilde{x}) = \frac{1}{2} (M(s) - s)^T C(m)^{-1} (M(s) - s). \tag{9}$$

In this implementation the dynamical model is kept both simple and generic, and, as with Lewis *et al.* (2012), the solution is allowed to infer a state vector that deviates from the model prediction. Such an approach is often called 'weak–constraint variational' (Zupanski 1997) in contrast to variational approaches that exclude deviations from the model trajectory, that is, where the model equations act as a strong constraint on the minimisation problem. The advantage of the weak-constraint approach (in particular with high uncertainty $C(m)$) is that it leaves high flexibility to fit the Sentinel observations. In other words, the retrieved trajectory will be primarily determined by observations rather than by the dynamical model – the output will be primarily an satellite data product rather than a model output constrained by observations.

The form of the model used here follows that of Lewis *et al.* (2012), who use a simple linear (more precisely a simple *affine*) model:

$$M(s) = As + b \;, \tag{10}$$

where the model parameters (i.e. the matrix A and the vector b) can be specified to suit a particular problem. In the following retrievals the matrix A of Equation 10 is populated such that a given component variable v_i of the state vector s_i at a given point in time i will be simulated as

$$v_i = a_i v_{i-1} + b_i \tag{11}$$

with $a_i = 1$ and $b_i = 0$. This equates to a simple zero-order model ('today is like tomorrow'). The consequence of using a model formulation like this is that it imposes a degree of temporal smoothness on the results, but also allows the information from each observation to influence the retrievals at other time steps. This is analogous to the temporal constraints on linear BRDF inversion described by Quaife and Lewis (2010).

The minimisation of Eq. 5 is performed iteratively and relies on the capability to evaluate J and its gradient. The Automatic Differentiation tool TAPENADE (Hascoët and

Pascual 2013) was applied to generate code that efficiently evaluates the gradient of *J*, the so-called adjoint of *J*. Figure 2 illustrates the algorithm described in this section.

3. Results and discussion

3.1. Retrievals

For each of the evaluation points described in 2.2 retrievals of leaf area index and soil moisture using Sentinel–1 and Sentinel–2 individually and together were compared to the field data. Figure 3 shows the results for each of these combinations for the mid point of field 508. The general pattern here is indicative of the results for each of the evaluation locations and the full results are summarised using Taylor diagrams described later in this section. Sentinel–1 only retrievals tend to capture the mean soil moisture reasonably well but the variability, especially later in the growing season, is not well correlated with the field observations. This is likely due to the poor representation of LAI in the retrieval, which is consequently very poorly correlated with the field observations and is biased low for most of the growing season. The retrieval algorithm is adjusting the soil moisture to be too high to compensate for the low LAI. This is perhaps unsurprising and suggests a level of equifinality between the two variables in the SAR signal. The Sentinel–2 only retrievals show a much stronger correlation with LAI, representing the phenology of the crops well, with a small bias in the order of 1.0 LAI units. The soil moisture appears to represent the mean well but this is not borne out across all of the data points (see Figure 6) where in general the RMSE and correlation of the Sentinel–2 only soil moisture results is poor. For the joint retrieval the LAI matches very well. The influence of Sentinel-2 on the retrievals pulls the Sentinel–1 results into line via the dynamic model that is implemented in the retrieval algorithm. In turn this allows the soil moisture to take on a more realistic temporal evolution, albeit somewhat biased high and not exhibiting sufficient variability in this example.

Figures 4 and 5 show the joint retrieval results for each of the evaluation points for leaf area index and soil moisture respectively. For LAI the results are consistently good, in particular for fields 508 and 542. The phenological profile is well captured despite the fact that this is somewhat different between the fields. Field 542, for example, which is growing triticale starts its greening–up phase weeks later than the wheat growing in field 508 and this is well captured by the satellite data. Field 515 shows a positive bias in the retrieval results, which then appears to become unbiased towards the end of the season. The same is observed in the 'mid' point of field 542. A consistent pattern observed for the wheat field is that the ground observed LAI appears to slow its decline towards the end of the season, whereas the retrieved LAI continues to reduce. This could be due to the effective LAI of the retrieval being reduced to compensate for a yellowing of the wheat foliage as the plant becomes senescent, whereas the optical measurement technique used to estimate the LAI relies only on the amount of light intercepted and is hence not sensitive to changes in the leaf colour. However, this pattern is not observed in the maize or triticale measurements and both of these species yellow as they pass maturity in much the same way as wheat.

The joint Sentinel–1 plus Sentinel–2 retrievals of soil moisture for each point are shown in Figure 5. For field 508 the retrievals capture the magnitude, with a small positive bias,
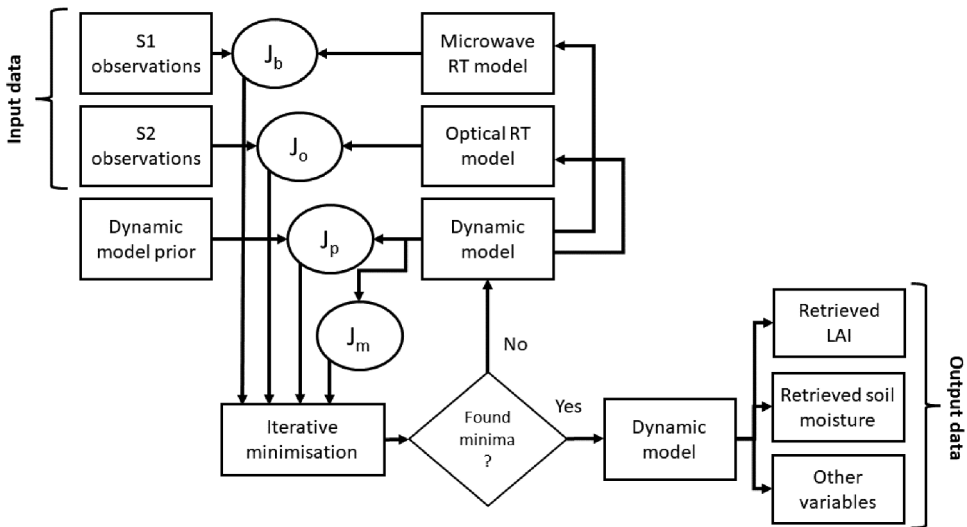
Figure 2. Flow chart outlining the algorithm and data described in Section 2. The box labelled 'other variables' is intended to indicate that, although we have not done so here, the system can be used to retrieve any of the variables that can be computed from output of the dynamical and RT models used.

and some of the early season dynamics. For field 542 the ground observed soil moisture is quite variable and the retrievals generally don't capture that variability, except for the 'low' validation point. As with field 508 there is a small positive bias. For field 515 there exists a significant positive bias in the retrievals.

Statistical summaries of the results for all nine points and the three possible combinations of sensors entering the retrievals are shown in Figure 6 for LAI and Figure 7 for soil moisture using Taylor diagrams (Taylor, 2001). The Taylor diagram allows for easy comparison of multiple experiments. The relative standard deviation of the retrievals to the ground observations is on both the x–axis and y–axis and increases radially from the origin. Correlation is shown on the polar axis such that the smaller the angle between a point and the horizontal axis the more correlated it is with the observations. Because there is an algebraic relationship between the relative standard deviation, correlation and RMSE it is also possible to plot the RMSE on the Taylor diagram; these are the radial contours that emerge from the 1.0 point of the x–axis. For LAI the Taylor diagram shows a clear picture. The Sentinel–1 only retrievals have very poor reproduction of the variability in the ground–truth data and the RMSE is > 0.75 for all experiments. The Sentinel–2 only retrievals over–estimate the variability in the observations, but tend to have a correlation better the 0.8 and an RMSE > 0.5. For the joint retrievals the results are better than the individual cases for all points with RSME always less than 0.5 and a correlation of better than 0.95. The Taylor diagram for the soil moisture is less clear and, by reference to Figure 5, much more dependent on the crop type. The joint retrievals tend to have a lower RMSE and higher correlation than the retrievals from either Sentinel on its own but this is not always the case. Especially for field 515, a high RMSE can be observed resulting from a visible bias in the retrievals, while for field 508, this bias is marginal resulting in small RMSE values. In addition, the joint retrievals also appear to
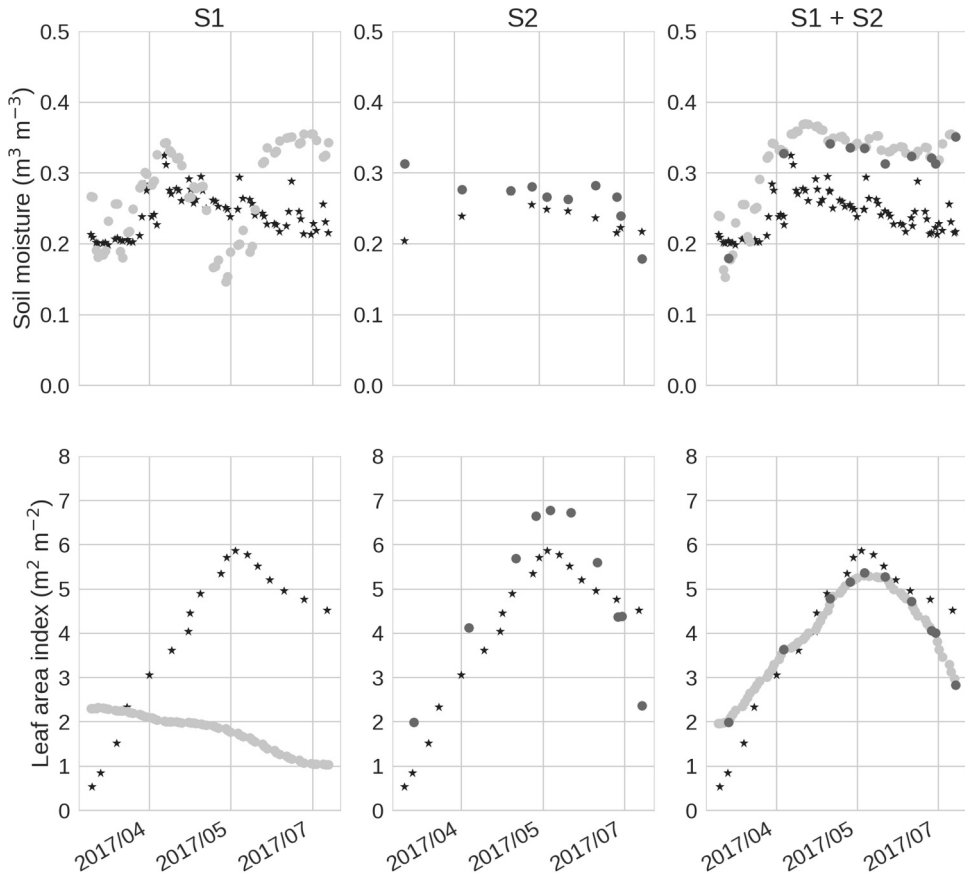
Figure 3. Comparisons of the retrievals for Sentinel–1 (left hand column) and Sentinel–2 (middle column) individually and jointly (right hand column) for field 508, 'mid' sampling point. Stars indicate field observations. Dark points retrievals at Sentinel–2 acquisition times and light grey points retrievals at Sentinel–1 acquisition times.

suppress the variability quite significantly in most cases, whereas the individual retrievals especially tend to contain too much variability.

The overall number of validation points in this study is relatively small: nine points in total, evenly split across three different crop types. This represents, in part, the complexities of undertaking such fieldwork. To build further confidence in the retrieval technique presented here would require expanding the experiments to a wider range of crop and soil types with a larger number of samples for each. The technique is likely to work with similar levels of accuracy for other grass crops as these most closely satisfy the underlying assumptions of the radiative transfer models used (i.e. the vegetation canopy can be represented as a homogeneous turbid medium) but it is unknown how the technique will work for other types of crops (for example, root vegetables such as potato). In addition, the soil type is similar in each of the fields studied here and so sampling across a wider range of soils will help strengthen understanding of the performance of the retrievals. However, given the less conclusive soil moisture results shown in this study, it is difficult to anticipate how the retrievals will be affected by different soil types.
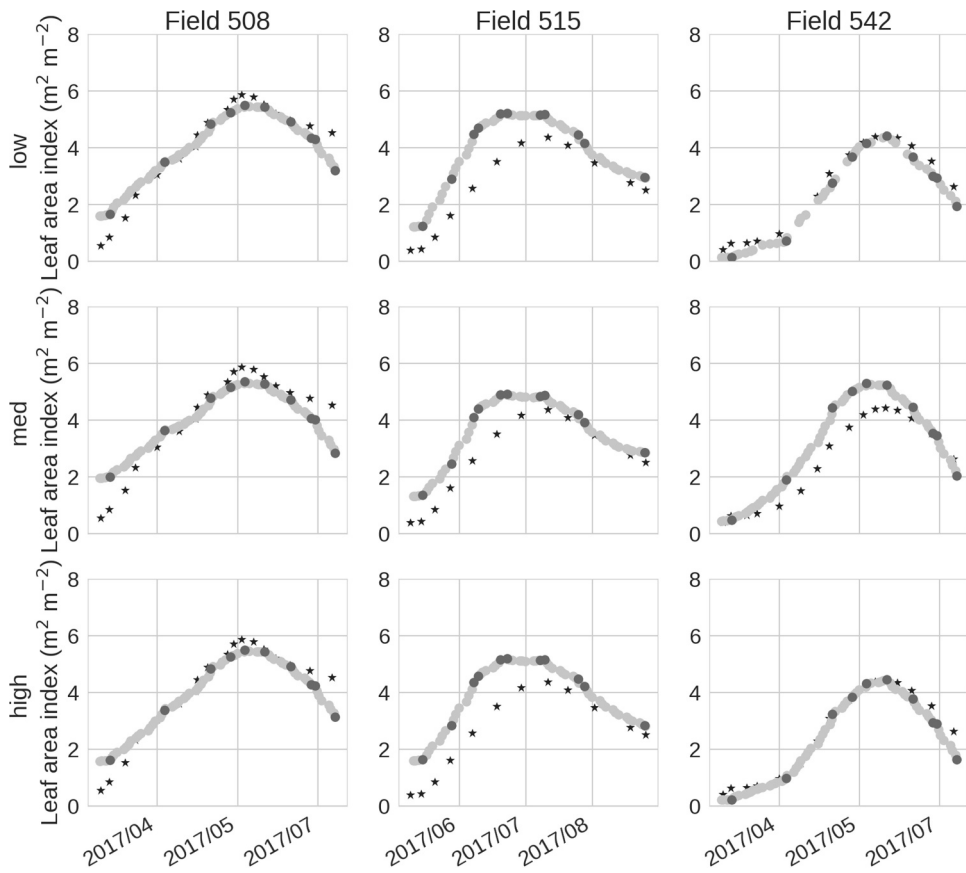
Figure 4. LAI retrievals using all Sentinel–1 and Sentinel–2 data from the 2017 growing season for each field (columns) and each sampling point (rows). Stars indicate field observations. Dark points retrievals at Sentinel–2 acquisition times and light grey points retrievals at Sentinel–1 acquisition times.

The within–field sampling strategy was determined prior to the current study and consequently the available LAI data were not acquired at the points at which the soil moisture probes are placed and at which the satellite data retrievals were carried out. Instead, field averages have been used. Given the high quality of the results for LAI this does not appear to have induced any problems with the results; for the field studied the averaged LAI likely provides a good approximation of the LAI at each of the individual points. If it were possible to have individual LAI time series at each of the sampling points this may allow further nuance in the retrievals to be undertood, such as the extent to which they can capture within–field LAI variability.

3.2. Future work

One aspect of the current study that can clearly be improved upon is the inclusion of other variables from the respective radiative transfer models in the retrieval. This was not attempted during this study as the complexity of the overall system was already high and
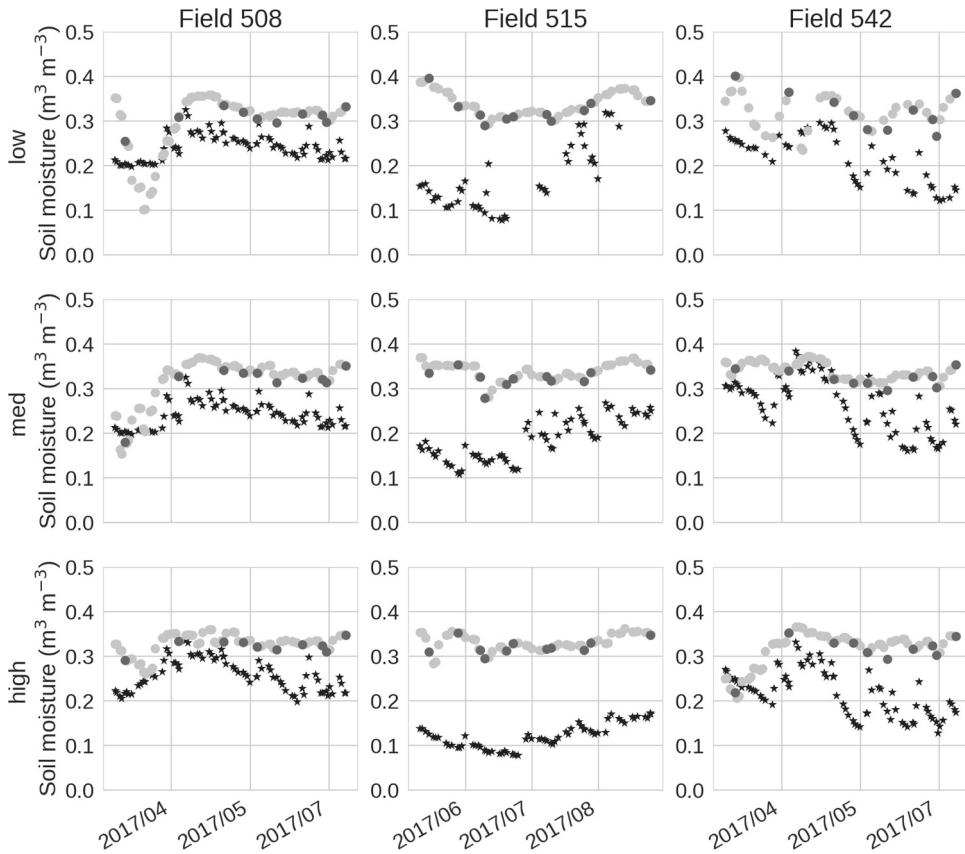
Figure 5. Soil moisture retrievals using all Sentinel–1 and Sentinel–2 data from the 2017 growing season for each field (columns) and each sampling point (rows). Stars indicate field observations. Dark points retrievals at Sentinel–2 acquisition times and light grey points retrievals at Sentinel–1 acquisition times.

the objective was to gain confidence in the LAI and soil moisture retrievals. Arguably the next most important variable to consider would be leaf water content as this has an impact on both the microwave signal and the short–wave infra red channels of Sentinel–2. This should be the next step forward in similar studies looking to exploit synergies between these two Sentinel missions. Including leaf pigments such as chlorophyll should also be a priority and may help to improve retrievals as plants become senescent.

Prior and data uncertainties required as inputs to the retrieval algorithm described in Section 2.4 in this study were assigned on the basis of expert elicitation, which is common practice in Bayesian inference, but more analytical methods could be employed for this purpose. Specifically, understanding the errors in the Sentinel–2 Level–2a surface reflectance and the microwave and optical radiative transfer models used as observation operators here have the potential to improve the retrievals. In addition, being able to objectively specify the off–diagonal elements in the covariance matrices is known to improve the performance of such algorithms (Pinnington *et al.* 2016).
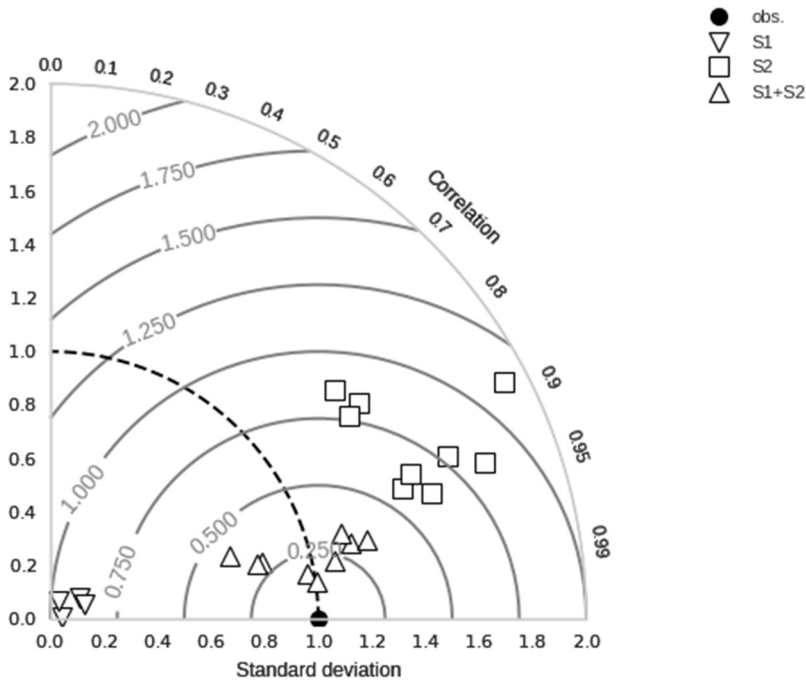
Figure 6. Taylor diagram showing summary results for LAI for each experiment. Distance from the origin indicates the level of variability compared to observations, with the dark dashed line representing the variability in the observations. The angle subtended with the horizontal axis represent correlation with the observations and the radial contours extending from the black dot represent relative RMSE. The closer a point to the black dot (which represent the observations) the better the retrieval has performed.

The particular choice of Sentinel–1 polarisation (HV) and Sentinel–2 band combination (4 through 8) employed was based on preliminary experiments that suggested these worked well with the choice of radiative transfer models. It is not necessarily the case that the combination used here will be the best in every situation or even optimal for this particular experiment. For Sentinel–2 for example, including all bands may lead to over confidence in the results and biasing the retrievals towards the optical data unless errors are correctly specified in the matrix $C(o)$. It is likely that this will need to include off–diagonal elements as well to represent correlated uncertainties in bands that are close to each other spectrally. In the examples in this paper $C(o)$ was set as a diagonal matrix. If other parameters had been targetted for retrieval it is possible additional skill could have been gained from other bands.

The software used for the retrievals in this study was designed to allow flexibility but not computational efficiency. Ultimately, time series retrieval is likely too slow to be used across entire images. Consequently, some further work is required to make this tool more efficient. Analysis of the computational performance of the retrievals (not shown here) revealed a large amount of time being used in the optical radiative transfer model, despite the fact that there are relatively far fewer data points collected by Sentinel–2. The bulk of this is being used to perform multiple scattering calculations (ρ^{M} in Equation 1). A future
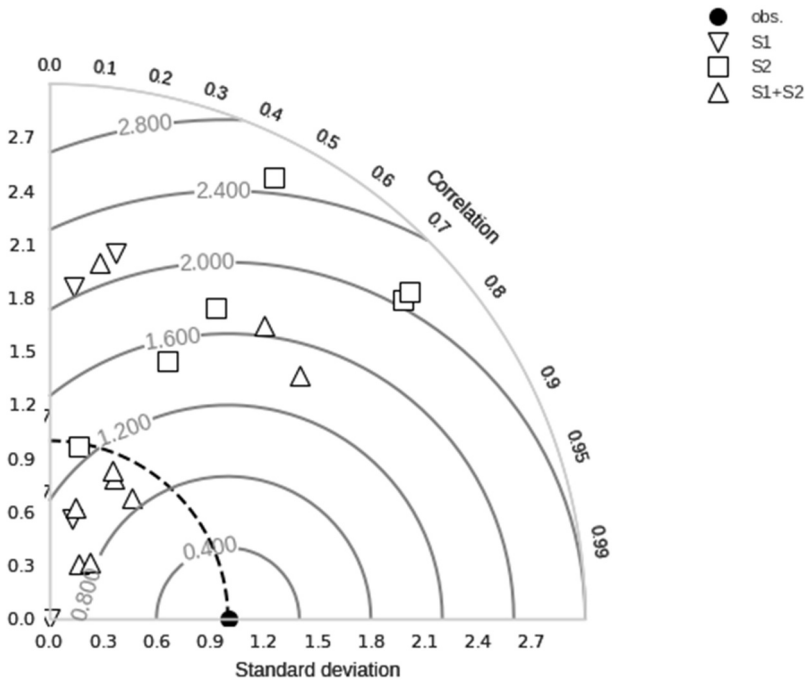
Figure 7. As Figure 6 but for soil moisture.

evolution of the system should probably include a modified version of the semi–discrete model with an approximation to the multiple scattering term.

4. Conclusions

The study has demonstrated a joint retrieval algorithm for LAI and soil moisture from Sentinel–1 and Sentinel–2 data using physically based radiative transfer models. The inversion scheme uses adjoint versions of the computer code of these models to perform the cost–function minimisation and the values of the retrieved variables are constrained temporally using a dynamic model that prescribes the degree of smoothness in their evolution. By comparison against field observations from agricultural sites in Germany it was shown that the joint retrieval improves estimates of leaf area index in every case compared to individual retrievals, and soil moisture in the majority of cases. There are very few examples of joint retrievals using physically based radiative transfer in the optical and SAR domains and none, to our knowledge, using Sentinel–1 and Sentinel–2. The results presented here show promise for future developments in this field. More generally this approach paves the way for combining a much wider range of observation types in retrieval systems or terrestrial data assimilation systems that aim at providing a consistent view on the terrestrial state that is informed by observations. For example, using a range of microwave frequencies, optical and thermal data. In addition, the matrix-based dynamic model could be informed by process-based models to provide prior information on likely magnitudes and timing of events, for example, under what conditions vegetation is likely to start greening–up.

Acknowledgments

Disclosure statement

Funding

ORCID

T. Quaife　http://orcid.org/0000-0001-6896-4613

References

Attema, E. and Ulaby, F., 1978. Vegetation modeled as a water cloud.*Radio. Science*, 13, 357–364.

Bracaglia, M., 1995. A fully polarimetric multiple scattering model for crops. *Remote Sensing of Environment*, 54 (3), 170–179. doi:10.1016/0034-4257(95)00151-4

Broge, N.H. and Mortensen, J.V., 2002. Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data. *Remote sensing of environment*, 81 (1), 45–57.

De Roo, R.D., *et al.* 2001. A semi-empirical backscattering model at l-band and c-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 39 (4), 864–872. doi:10.1109/36.917912

Disney, M.I., Saich, P., and Lewis, P., 2003. Modelling the radiometric response of a dynamic, 3d structural model of scots pine in the optical and microwave domains.In:*Proc. IEEE Int. Geos. RS Symposium*, Toulouse, France. vol. 6, 3537–3539.

Drusch, M., *et al.*, 2012. Sentinel-2: esa's optical high-resolution mission for gmes operational services. *Remote Sensing of Environment*, 120, 25–36. doi:10.1016/j.rse.2011.11.026

European Space Agency, E., 2015. *Sentinel-2 user handbook*. European Space Agency. https://sentinels.copernicus.eu/web/sentinel/user-guides/document-library/-/asset_publisher/xlslt4309D5h/content/sentinel-2-user-handbook

Féret, J.B., *et al.*, 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling.*Remote sensing of. environment*, 115 (10), 2742–2750.

Ferrazzoli, P. and Guerriero, L., 1996. Passive microwave remote sensing of forests: a model investigation. *IEEE Transactions on Geoscience and Remote Sensing*, 34 (2), 433–443. doi:10.1109/36.485121

Floury, N., 1999. Modélisation radars des couverts forestiers: Application à la télédétection. PhD Thesis, Paris Diderot University, Paris.

Gobron, N., *et al.* 1997. A semidiscrete model for the scattering of light by vegetation. *Journal of Geophysical Research: Atmospheres*, 102 (D8), 9431–9446. doi:10.1029/96JD04013

Hascoët, L. and Pascual, V., 2013. The tapenade automatic differentiation tool: principles, model, and specification. *ACM Transactions On Mathematical Software*, 39(3), 1–43, 10.1145/2450153.2450158, Available from

Hosseini, M., *et al*., 2019. Synthetic aperture radar and optical satellite data for estimating the biomass of corn. *International Journal of Applied Earth Observation and Geoinformation*, 83, 101933. doi:10.1016/j.jag.2019.101933

Idso, S.B., Jackson, R.D., and Reginato, R.J., 1977. Remote-sensing of crop yields. *Science*, 196 (4285), 19–25. doi:10.1126/science.196.4285.19

Jacquemoud, S. and Baret, F., 1990. Prospect: a model of leaf optical properties spectra. *Remote Sensing of Environment*, 34 (2), 75–91. doi:10.1016/0034-4257(90)90100-Z

Kaminski, T. and Mathieu, -P.-P., 2017. Reviews and syntheses: flying the satellite into your model: on the role of observation operators in constraining models of the earth system and the carbon cycle. *Biogeosciences*, 14 (9), 2343–2357. Available from http://www.biogeosciences.net/14/2343/2017/

Karam, M.A., *et al*., 1992. A microwave scattering model for layered vegetation. *IEEE Trans. Geosci. Remote Sensing*, 30 (4), 767–784. doi:10.1109/36.158872

Kim, Y. and van Zyl, J., 2001. Comparison of forest parameter estimation techniques using sar data. In:*Proc. IEEE 2001 International Geoscience and Remote Sensing Symposium IGARSS '01*, 9–13 July, Sydney, Australia. vol. 3, 1395–1397.

Kim, S.W. and Won, J.S., 2003. Measurements of soil compaction rate by using jers-1 sar and a prediction model. *Geoscience and remote sensing. IEEE Transactions on*, 41 (11), 2683–2686.

Lewis, P., *et al*., 2003. Modelling the radiometric response of a dynamic, 3d model of wheat in the optical and microwave domains.*In:Proc. IEEE Int. Geos. RS Symposium*, 21-25 July 2003, Toulouse, France.

Lewis, P., *et al*., 2012. An earth observation land data assimilation system (eo-ldas. *Remote Sensing of Environment*, 120, 219–235. The Sentinel Missions – New Opportunities for Science: http://www.sciencedirect.com/science/article/pii/S0034425712000788

Liu, W. and Kogan, F., 2002. Monitoring Brazilian soybean production using NOAA/AVHRR based vegetation condition indices. *International Journal of Remote Sensing*, 23 (6), 1161–1179. doi:10.1080/01431160110076126

Martinez, J.M., *et al*., 2000. Measurements and modelling of vertical backscatter distribution in forest canopy. *IEEE Trans. Geosci. Remote Sensing*, 38 (2), 710–719. doi:10.1109/36.842000

Mattia, F., *et al*., 2003. A comparison between soil roughness statistics used in surface scattering models derived from mechanical and laser profilers. *Geoscience and Remote Sensing. IEEE Transactions on*, 41 (7), 1659–1671.

McNairn, H., *et al*., 2012. Establishing crop productivity using radarsat-2.*ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIX-B8,283–287. https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B8/283/2012/

Oh, Y., Sarabandi, K., and Ulaby, F.T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *Geoscience and Remote Sensing, IEEE Transactions on*, 30 (2), 370–381. doi:10.1109/36.134086

Paloscia, S., Pettinato, S., and Santi, E., 2012. Combining l and x band sar data for estimating biomass and soil moisture of agriculturalfields. *European Journal of Remote Sensing*, 45 (1), 99–109. doi:10.5721/EuJRS20124510

Pinnington, E.M., *et al*., 2016. Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using four-dimensional variational data assimilation. *Agricultural and Forest Meteorology*, 228, 299–314. doi:10.1016/j.agrformet.2016.07.006

Prévot, L., *et al*., 2003. Assimilating optical and radar data into the stics crop model for wheat. *Agronomie*, 23 (4), 297–303. doi:10.1051/agro:2003003

Price, J.C., 1990. On the information content of soil reflectance spectra. *Remote Sensing of Environment*, 33 (2), 113–121. doi:10.1016/0034-4257(90)90037-M

Punalekar, S.M., *et al*., 2018. Application of sentinel-2a data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment*, 218, 207–220. doi:10.1016/j.rse.2018.09.028

Quaife, T. and Lewis, P., 2010. Temporal constraints on linear brdf model parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 48 (5), 2445–2450. doi:10.1109/TGRS.2009.2038901

Satalino, G., Dente, L., and Mattia, F., 2006. Integration of meris and asar data for lai estimation of wheatfields.In:*Proc. IEEE International Conference on Geoscience and Remote Sensing Symposium IGARSS 2006*, July 31 Aug. 4, Denver, Colorado. 2255–2258.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research Atmospheres*, 106 (D7), 7183–7192.

Torres, R., *et al.*, 2012. GMES Sentinel-1 mission. *Remote sensing of environment*, 120, 9–24.

Tucker, C.J., 1980. A critical review of remote sensing and other methods for non-destructive estimation of standing crop biomass. *Grass and Forage Science*, 35 (3), 177–182.

Ulaby, F T. and Elachi, C., 1990. *Radar polarimetry for geoscience appications*. Norwood, MA, USA: Artech House, Inc.

Viña, A., *et al.*, 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 115 (12), 3468–3478. doi:10.1016/j.rse.2011.08.010

Weiß, T., *et al.*, 2020. Evaluation of different radiative transfer models for microwave backscatter estimation of wheatfields. *Remote Sensing*, 12 (3037), 3037. doi:10.3390/rs12183037

Wiegand, C.L., Richardson, A.J., and Nixon, P.R., 1986. Spectral components analysis: a bridge between spectral observations and agrometeorological crop models. *IEEE Transactions on Geoscience and Remote Sensing*, GE–24 (1), 83–89. doi:10.1109/TGRS.1986.289688

Zupanski, D., 1997. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Weather Rev*, 125 (9), 2274–2292. doi:10.1175/1520-0493(1997)125<2274:AGWCAT>2.0.CO;2