Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index: BACI

Deliverable 2.2: Definition and specification of protocols for merging EO data and specification of “state vector”: time, space, wavelength requirements, ancillary data (elevation, land cover), projections, file formats (Technical requirements document)

Project title:	Detecting changes in essential ecosystem and biodiversity properties- towards a Biosphere Atmosphere Change Index
Project Acronym:	BACI
Grant Agreement number:	640176
Main pillar:	Industrial Leadership
Topic:	EO- 1- 2014: New ideas for Earth-relevant space applications
Start date of the project:	1st April 2015
Duration of the project:	48 months
Dissemination level:	Public
Responsible of the deliverable	Mathias Disney (UCL) Phone: +44(0) 20 7679 0592 Email: mathias.disney@ucl.ac.uk
Contributors	Mat Disney (UCL Geography) Maxim Chernetskiy (UCL Geography) Jose Gomez-Dans (UCL Geography) Christiane Schmullius (FSU Jena) Marcel Urban (FSU Jena) Johannes Balling (FSU Jena)
Date of submission:	30th September 2016

Summary

This deliverable D2.2 provides the definition and specification of protocols for merging EO data and specification of ‘surface state vector’ (SSV): time, space, wavelength requirements, ancillary data, including projections and file formats. This deliverable forms the basis of a technical requirements specification for the production of the state vector.

Aim/Outcome

- Specification of ‘surface state vector’ (SSV) (D2.2)

Acronyms

CEDA	Centre for Environmental Data Archival
CEMS	Climate and Environmental Monitoring from Space
ECV	Essential Climate Variable
EO	Earth Observation
fAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FSU	Friedrich Schiller University
GWS	CEDA/CEMS Group Work Space
JASMIN	Joint Analysis System (Meeting Infrastructure Needs)
LAI	Leaf Area Index
UCL	University College London

Contents

1 Introduction 6
2 Definition Of Ssv Framework, Input, Output, Ancillary Data 6
3 Data-Merging Framework: Summary 6
3.1 Input, Outputs, Ancillary Data 7
3.2 Input, Outputs 8
3.2.1 Inputs 8
3.2.2 Outputs 10
4 Results 11
4.1 Temporally Regularized Modis Data 11
4.2 Temporally Regularized Sentinel Data 13
4.3 Temporally Regularized Estimates Of Biophysical Properties 14
5 Summary 16
6 References 17

List of figures

Figure 1 MODIS reflectance band 2 (NIR), for the Hainich test site (Germany, lat/lon: 51.08, 19.45), characterised as a deciduous broad-leaf forest site, for the period 2001-2015. Grey area is uncertainty in +/- 1 σ. 1225 MODIS tiles............................11
Figure 2 MODIS reflectance band 2 (NIR), for the Somalia test site (Horn of Africa, part of BACI FT, lat/lon: 47.00, 6.00), characterised as desert/savannah, for the period 2001-2015. 3413 MODIS tiles. 12
Figure 3 MODIS reflectance band 2 (NIR), for the Viterbo test site, Italy (BACI FT, lat/lon: 42.38, 12.03), characterised as agricultural test site, for the period 2001-2015. 3678 MODIS tiles. 12
Figure 4 MODIS reflectance band 2 (NIR), for the Wytham Woods test site, UK (lat/lon: (-1.34, 51.78), characterised as deciduous broadleaf forest, for the period 2001-2015. 281 MODIS tiles. 13
Figure 5 Sentinel-1 VH backscatter (linear) for the Viterbo test site, Italy (BACI FT), characterised as agricultural test site, for the period 10/2014 to 10/2015. Shades of grey are 5%, 25% and 75% credible interval of uncertainties. 14
Figure 6 Sentinel-1 VV backscatter (linear) for the Viterbo test site, Italy (BACI FT), characterised as agricultural test site, for the period 10/2014 to 10/2015. Shades of grey are 5%, 25% and 75% credible interval of uncertainties. 14
Figure 7 Sentinel-1 VH backscatter (linear) for the Wytham Woods test site, UK (BACI FT), characterised as deciduous broadleaf, for the period 10/2014 to 10/2015. Shades of grey are 1%, 5% and 25% credible interval of uncertainties. 14
Figure 8 SSV-derived surface biophysical properties, and uncertainties for the Hainich test site 2000-2008...15
Figure 9 SSV-derived LAI and uncertainties for the Wytham Woods test site 2001...16

1 Introduction

This document is intended to fulfil the requirements for **D2.2** i.e. to provide details of the technical requirements for the generation of the EO-derived SSV, via merging of EO data from various streams. The requirements for SSV specification include: time, space and wavelength requirements, ancillary data, projections, file formats.

2 Definition of SSV framework, input, output, ancillary data

The SSV, as defined in the BACI proposal, is the best estimate from EO, representing the state of a point/region on the land surface at a given time as a function of input data i.e. $SSV = f(\rho, \sigma^0(x, y, t, \lambda,) \cdots)$, in particular reflectance ρ, and $\partial\rho/\partial t$ i.e. change in reflectance since the last observation, backscatter σ^0, as well as other potential variables such as LST, vegetation state and ancillary information), with uncertainty. Two key aspects of the approach are that: i) SSV definition should be flexible enough to allow for inclusion of other EO data where useful; and ii) the framework should include assessment of uncertainty. The latter is important in terms of assessing the strength and certainty of change as indicated by the SSV, from the perspective of the observation uncertainties, as well as for incorporating the impact of this uncertainty into the subsequent BACI and downstream analysis.

3 Data-merging framework: summary

The main tool used for data merging for the BACI SSV is the Earth Observation Land Data Assimilation Scheme (EO-LDAS) (Lewis et al, 2012; EO-LDAS, 2013). The code for EO-LDAS is open and available from EO-LDAS (2013).

The EO-LDAS is an optimal estimation data assimilation (DA) framework developed expressly for merging data from different sources with different sampling properties (spectral and temporal particularly). The EO-LDAS is a Bayesian estimator (Enting, 2002), based on the scheme of Tarantola (2015). All observations (including any prior information on the state variables) is represented by a prior probability density function (PDF), which when combined yields a posteriori PDF for the parameters, which is the result/solution of the assimilation problem. Assuming PDFs are Gaussian and the models can be considered linear (or nearly, e.g. via transformation) then the posterior parameter PDF can also be approximated as:

$$\rho(\chi) = exp\big(-J(\chi)\big) \qquad (1)$$

which is the maximum likelihood estimate of the state variables χ. This is the minimum of a cost function which takes the form:

$$J(\chi) = \sum_i J_i(\chi)\ (1) \qquad (2)$$

where $J_i(\chi)$is a cost function expressing a constraint i, a member of some set of constraints.

EO-LDAS has two main components: (i) a set of constraints, expressed via Eq. (1); (ii) an assimilation algorithm, i.e. a way to apply the constraints to achieve the optimal estimate of the state vector. The set of constraints in EO-LDAS involves: (i) an observational constraint $J_{obs}(\chi)$, requiring data (from EO or ground measurements) and a model for translating from state space to observation space (the observation operator); (ii) a dynamic model constraint $J_{model}(\chi)$, conditioning the temporal (and/or spatial) evolution of the state vector; (iii) physical or empirical bounds and/or distribution constraints $J_{prior}(\chi)$ to the state vector elements. Thus, the core of EO-LDAS becomes:

$$J(\chi) = J_{obs}(\chi) + J_{prior}(\chi) + J_{model}(\chi) \qquad (3)$$

Each of these constraints has associated with it an error model represented by a covariance matrix. Note that the symbol χ refers to the set of state variables that we wish to estimate. In EO-LDAS this essentially means a representation of the state at each sample point in time (and/or space) that we consider. So, e.g. estimating Leaf Area Index and leaf Chlorophyll content at one location for every day of the year, would involve a state vector with 365×2 elements. EO-LDAS can also include 'static' state representations affecting one or more of the constraints considered constant in space/time.

3.1 Input, outputs, ancillary data

Following initial testing of the SSV approach using EO-LDAS, for the BACI SSV we limit the **approach from the 1D vegetation RT model inversion case, to the generation of temporally regularized surface reflectance and backscatter products**. The vegetation EO-LDAS application allows for retrieval of surface biophysical properties, based on optimal estimate of biophysical radiative transfer (RT) model parameters. However, following feedback and discussions from other WPs, we note the following:

- The surface state and change, $SSV = f(\rho, \sigma^0(x, y, t, \lambda,) \cdots)$ and ΔSSV are most directly captured by the observations going into the SSV themselves;
- Derived products including estimates of LAI, fAPAR and other surface biophysical properties can be desirable from the perspective of attribution of changes in SSV. However any RT model required to derive these estimates will impose additional constraints and assumptions on the original data, potentially masking or exaggerating changes in the SSV;
- The change analysis to be carried out on the SSV by the ML group in WP5 (and others), will likely be most sensitive to change in an SSV free of additional model assumptions: if a change is real, it should be visible in the observations themselves.

As a result of this, the BACI SSV is proposed to be generated from the best-estimate of observations i.e. temporally regularized, multi-wavelength time series of observations, in general avoiding biophysical RT model derivations of surface parameters. However, we note that in producing this regularized time series SSV, it is also straightforward to produce estimates of surface biophysical properties, should they be desirable, and taking into account the caveats provided above.

For generating the best-estimate of surface state from optical data, we implement the inversion of linearised bidirectional reflectance distribution function (BRDF) Kernel Models by temporal regularization. In general this implementation can be described by the EO-LDAS framework represented by eq. 3. These models assume BRF as linear combination of kernels which are functions of view and sun angles i.e. $\bar{\rho} = f(f_{iso}, f_{vol}, f_{geo})$ where the f_{iso}, f_{vol} and f_{geo} are isotropic, volumetric and geometric kernels representing the respective scattering 'shapes' from the surface (Schaaf et al. 2002). The linear BRDF model approach is important as it allows us to normalise viewing and illumination geometry i.e. placing all observations in a common view, sun angle configuration. This removes any view and sun-angle dependencies from the data, which can provide significant changes over time, and can be mistaken for changes in the surface state. The recent analysis by Morton et al. (2014) for example suggests that the Amazon 'greening' controversy may have been due to residual sun-angle effects in composited data.

The view and sun angle normalized BRDF, is then used, along with the microwave time series, as input for temporal regularization. This provides an optimal time series interpolation, based on the preceding and following observations in the time series, and using the EO-LDAS framework as the temporal smoother. Temporal regularization works not as a post processing filter but as a constraint which reduces the number of possible solutions of an inversion (Lauvernet et al. 2008; Knorr et al. 2010; Quaife and Lewis 2010; Lewis et al. 2012). This is a key technique in the always "ill-posed" problem of inferring parameters from EO. In a perfect or noise-free case, we can use a model which describes the temporal development of a desired parameter (LAI, BRDF kernel etc.). However in a real situation this kind of a model is usually not available. Here, we use a simple 'zero order' process model that assumes "today is the same as tomorrow" i.e. that the temporal development of the parameter of interest is smooth.

Temporal regularization provides optimal interpolation i.e. the best estimate of a continuous surface state observation (in an optimal estimation sense). However temporal regularization requires an estimate of γ, the model error or 'smoothness' parameter, describing the degree to which the regularization process should rely on fitting to previous observations, or the current observation. If is too large, then the time series can be over-smoothed causing smaller magnitude or shorter duration changes to be down-weighted.

Inversion of BRDF models and temporal regularization allows us to obtain a full time series of reflectance, broad band (BB) albedo and the associated posterior uncertainties. Temporal regularization of surface reflectance is an important technique to fill time gaps, normalize BRDF variations i.e. to account for different spectral/angular sampling from different sensors.

The SSV generation as designed is flexible and able to incorporate observations from other sources. However the overhead for this is not the inclusion and combination of the data per se, but the downloading and pre-processing of the data (consistent spatial resolution). By processing on the CEMS system we have minimized this overhead in that these data are archived on the system and so we minimize the movement of large amounts of data.

3.2 Input, outputs

3.2.1 Inputs

- NASA MODIS, across the time-period 2000 – present: global coverage, 7 spectral bands, 1km spatial resolution, ISIN projection;

- NASA MISR, across the time-period 2000 – present: intermittent spatial coverage, 4 spectral bands, 275m spatial resolution, SOM projection. Only considered for regional test cases due to spatial coverage;
- NASA Landsat, 2000 – present (and historical, but using 2000+ for BACI): intermittent spatial coverage, 7 vis-SWIR bands, 30m spatial resolution, UTM projection. Considered for local test cases, due to inconsistent spatial coverage;
- ESA Sentinel-1 microwave backscatter, 2014-present: processed to calibrated backscatter,

To come, if applicable

- Sentinel 1B (2015-): to be processed as for existing S1A data;
- Sentinel 2A (mid-2015- and B 2017): to be considered in the same way as for Landsat i.e. for local test cases where spatial coverage is likely to prove a significant advantage;
- Sentinel 3A (early 2016): to be processed in a similar way as for MODIS i.e. for regional to global cases, particularly where temporal coverage is likely to prove an advantage;
- MODIS LST, 2000-present: applying the temporal regularization framework to provide a gap-filled time series of observations;

Clearly the Sentinel observations will not provide the time series coverage required to identify anything other than much shorter term anomalies. However in conjunction with the existing SSV, this may be sufficient to demonstrate the additional utility of higher spatial resolution and increased temporal coverage provided by both Sentinel 2 (regional) and 3 (global). The design of the SSV is intended to allow these data streams to be incorporated, albeit with the overhead of collecting and archiving the data, which can be the largest single time constraint given the volumes of data required, as described above. To try and mitigate the lack of historical SAR time series, we are now including recently-released ESA ASAR data.

Historical microwave data

In the second half of 2016, the European Space Agency opened the archive of historical microwave data to the public, which allows the registered users to download ENVISAT ASAR C-Band data free of charge.

The data download is done via EOLi (Earth Observation Link), which is an open source application provided by the ESA (https://earth.esa.int/web/guest/eoli). This earth observation catalogue and ordering service allow the registered user to search, browse and download ESA earth observation data as well as satellite data from third party missions (e.g. ALOS). Regular registered users have a download query restriction by 5 satellite scenes per day. It is possible to enlarge the number of downloads up to 30 scenes per day by writing a project proposal (data service request), which have been done for the BACI project.

The processing of the ENVISAT SAR data was done utilizing the SNAP Toolbox (Sentinel Application Platform), which is an open source software, which were developed for pre-processing, visualisation and analysing different sources of earth observation data, e.g. from the Sentinel mission, ENVISAT data as well as third party missions ALOS 1 and ALOS 2.

The processing of the ENVISAT ASAR data* (Image Mode Precision Image) used within BACI is done using the following consecutive processing steps: (1) import and multi-looking, (2) retrieving the parameters from the orbit file, (3) radiometric calibration, (4) Terrain-correction and terrain-flattening, and (5) output.

*Data provided by the European Space Agency (© ESA (2016).

3.2.2 Outputs

SSV, comprising:

- Temporally regularized (optimally smoothed and gap-filled) estimates of surface state comprising reflectance and backscatter time series;
- Uncertainty estimate of resulting regularized surface state;
- Covering the time period:
  - o 2000-present for MODIS as base input, with Landsat and MISR for local areas/periods;
  - o 2014- for Sentinel 1A for BACI FT sites;
  - o 2000- for ENVISAT ASAR back to for selected FT sites;

File format(s), as defined in MS2:

- **netCDF**: all SSV data will be provided as netCDF as agreed with consortium, can be used for point and spatial data, widely-used, handled by most/all main software tools (particularly Python/GDAL);

Ancillary data required **for**, and/or metadata resulting **from,** SSV generation:

- **Observation type**: name and type of observation ('raw', low-level, processed or derived (eg LST);
- **Calibration and units**: SI in the case of lower-level observations; categorical in the case of eg land cover, in which case class definitions will be provided;
- **Origin of observation**: source, citation, contact person/URL/institution;
- **Date, time, location/area**: i.e. lat, lon range and projection if not UTM)
- **Projection**: UTM (zones);
- **Terrain correction (for SAR products)**: specify whether terrain correction has been applied;
- **Climate, land cover, trait information**: The MODIS land cover map will be used if LC is to be used to stratify analysis; non-EO fields such as climate, traits etc fields can be incorporated at the machine learning stage, if they are also in the same spatial resolution (1km) and projection (UTM), netCDF.

Naming conventions:

- Unique to sensor, product, site, date (where version could be within product, and/or defined by a BACI consortium member based on their own processing).

For some datasets (e.g. MODIS particularly) it will be assumed, unless specified otherwise, that ALL datasets compiled under a given **SENSOR_PRODUCT_SITE** are processed in the same way and therefore have the same metadata properties. If so, then a single metadata file will be used; if not this will be reflected in the naming convention e.g. by version, so that a user can track which product has been processed in what way and by whom.

4 Results

4.1 Temporally regularized MODIS data

The following figures show example time series of normalized MODIS during 2000-2015 with uncertainties, for different areas. This is the core of the SSV i.e. temporally regularized, gap-filled estimate of surface state, with associated uncertainty. While focus is on the BACI FT sites, other sites have been selected for test SSV generation where there are sufficient ground data (including field observations of potential surface state changes) that the resulting SSV can be validated as far as possible.

These SSV values will be used as input to the BACI machine learning (ML) approach, which aims to identify change 'hotspots' from EO and ancillary data (climate, traits, function etc).

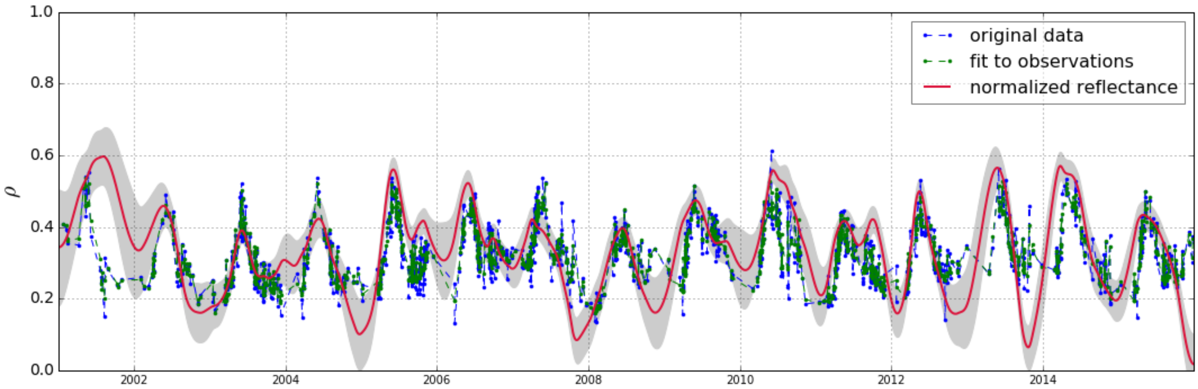

Figure 1 MODIS reflectance band 2 (NIR), for the Hainich test site (Germany, lat/lon: 51.08, 19.45), characterised as a deciduous broad-leaf forest site, for the period 2001-2015. Grey area is uncertainty in +/- 1 σ. 1225 MODIS tiles.

The dark blue symbols are the original MODIS reflectance filtered by standard MODIS quality assurance (QA) flags i.e. cloud level and sensor response. The red line is normalized reflectance. The green line is the EO-LDAS derived fit to observations i.e. reflectance calculated by forward model assuming the same geometry as original data.

At the deciduous broad-leaf site, the seasonal variation is obvious, along with considerable other variation, even in the regularized case. This demonstrates a major advantage of the regularized approach over, say a RT-based model retrieval – where the RT model has to retrieve e.g. LAI on the basis of noisy observations, the resulting LAI can increase and decrese significantly from day-to-day (see Disney et al., 2016 for an example of this problem). This poses problems for models either driven by, or compared with the LAI values, as well as for change detection, which will be masked by the large day-to-day variation. This can arise due to the input data, the model assumptions and also ancillary data used for the RT model retrieval.

Figure 2 MODIS reflectance band 2 (NIR), for the Somalia test site (Horn of Africa, part of BACI FT, lat/lon: 47.00, 6.00), characterised as desert/savannah, for the period 2001-2015. 3413 MODIS tiles.

In this case we see there is almost no discernible seasonal phenology, and a series of rapid intra-annual fluctuations, along with some slightly more significant increases and reductions. These latter might make ideal cases for candidate rapid changes/disturbance. What is most obvious here is the impact of BRDF on magnitude, which has acted to cause the normalized reflectance to be substantially higher than the observed reflectance. In a relatively stable target area such as this, this will not have a large impact on a change detection strategy. However, for subtle changes against a background of perhaps greater seasonal and/or interannual variation, this magnitude variation can be important.

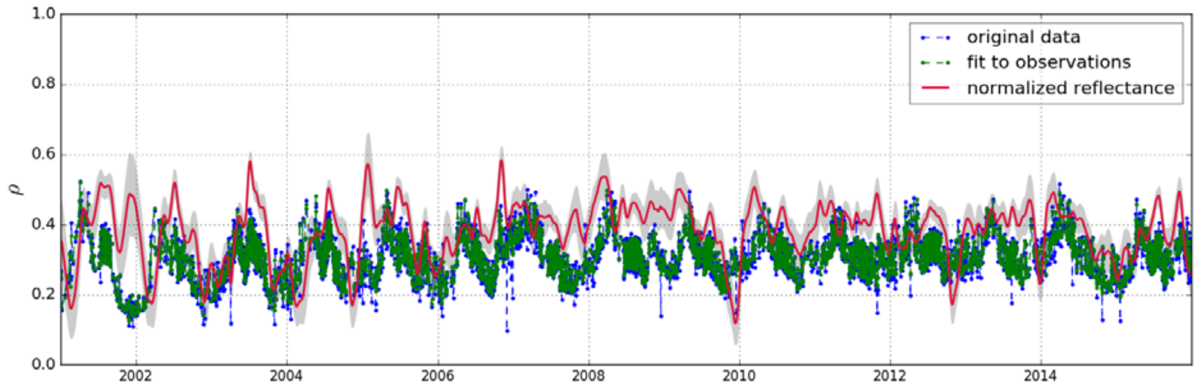

Figure 3 MODIS reflectance band 2 (NIR), for the Viterbo test site, Italy (BACI FT, lat/lon: 42.38, 12.03), characterised as agricultural test site, for the period 2001-2015. 3678 MODIS tiles.

In the Viterbo case, the strong seasonal cycle is again visible, stronger even than for the deciduous woodland site at Hainisch. The difference caused by the normalization is more apparent here.

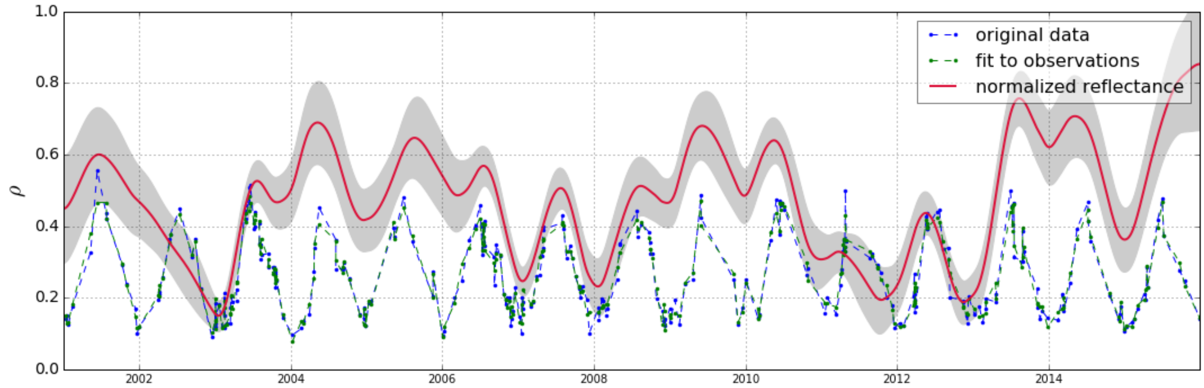

Figure 4 MODIS reflectance band 2 (NIR), for the Wytham Woods test site, UK (lat/lon: (-1.34, 51.78), characterised as deciduous broadleaf forest, for the period 2001-2015. 281 MODIS tiles.

For the Wytham Woods case, the seasonal cycle is visible, but much weaker than the Viterbo and Hainisch sites, despite the deciduous nature of the majority of the woodland. This may in part be due to the varying canopy type (some coniferous, varying understory) and also due to the more temporally sparse nature of the observations. Figures 2 and 4 show that the normalized reflectance has higher values than input data. This is due to difference in view/sun geometry especially during wintertime. We can see it observing that fit-to-observation (green curve) is very close to the input reflectance values. This example demonstrates importance of reflectance normalization.

4.2 Temporally regularized Sentinel data

The examples below show Sentinel-1 data from BACI CEMS archive (where there also ALOS Palsar potentially for some sites). The Sentinel-1 data are aggregated to the MODIS 500m resolution.

The figures below show essentially a single year of Sentinel 1 data over the test sites, and as a result we cannot draw any conclusions about seasonal cycles etc. The regular nature of the observations is apparent however. This will not be the case for optical due to clouds, and this is a key reason for including microwave observations in the SSV. A second key reason is that of course the microwave signal is sensitive to different surface properties than optical observations – surface roughness and moisture primarily. Where we have multiple optical data observations, they will likely be strongly correlated in terms of information content, even if they have different spectral bands. This is because the bands almost always cover the same general regions (visible, NIR, SWIR), even if at different specific band locations. Optical observations are often strongly correlated in time as well: clouds, and sun-synchronous orbits tend to mean that when observations are obscured from one platform/sensor, they are often obscured from others. Higher revisit frequency can reduce this issue, but rarely remove it.

Part of the reason for the high uncertainty values of Sentinel-1 time series is due to aggregation to MODIS resolution. Because of nature of microwave data (coherent interference), the backscatter signal has a broad range of possible values. When spatially aggregated to lower resolution this tends to substantially increase uncertainty. This is the reason why it is so important to use a method which takes uncertainties into account. We can see in Figures 5-7 that uncertainties are higher in between acquisitions i.e. we restored intervals between satellite acquisitions but we trust these intervals less.

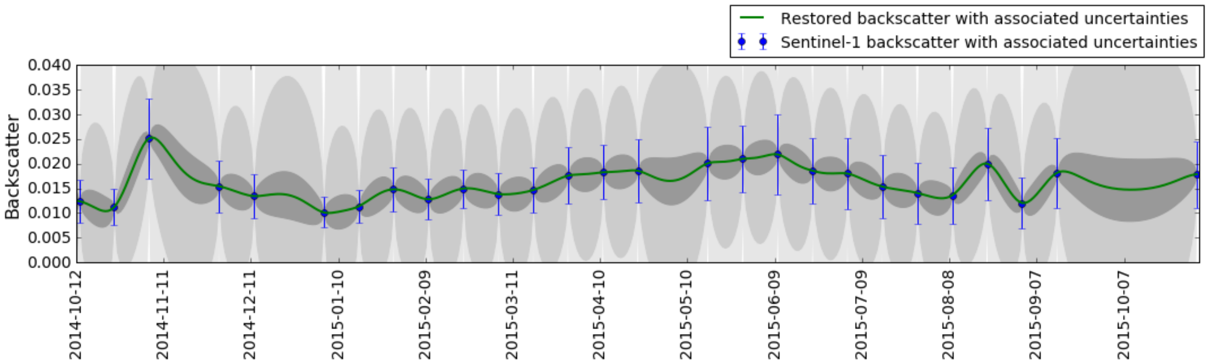

Figure 5 Sentinel-1 VH backscatter (linear) for the Viterbo test site, Italy (BACI FT), characterised as agricultural test site, for the period 10/2014 to 10/2015. Shades of grey are 5%, 25% and 75% credible interval of uncertainties.

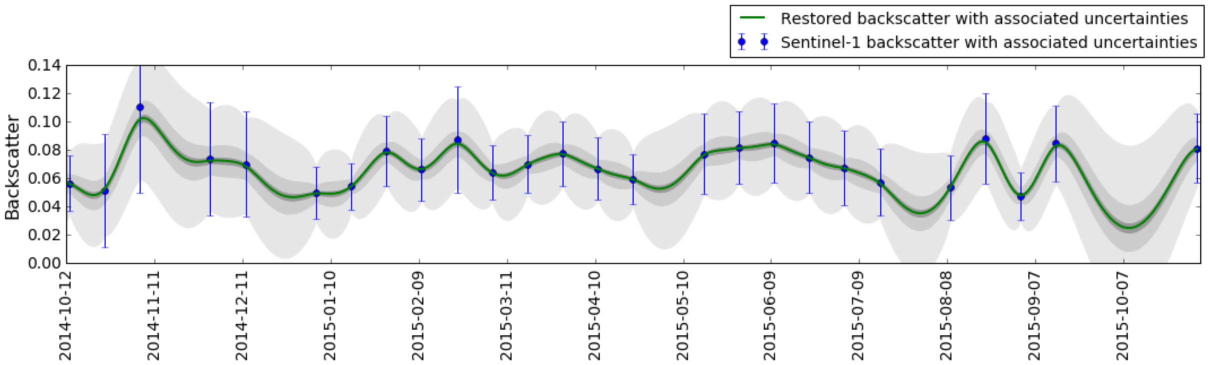

Figure 6 Sentinel-1 VV backscatter (linear) for the Viterbo test site, Italy (BACI FT), characterised as agricultural test site, for the period 10/2014 to 10/2015. Shades of grey are 5%, 25% and 75% credible interval of uncertainties.

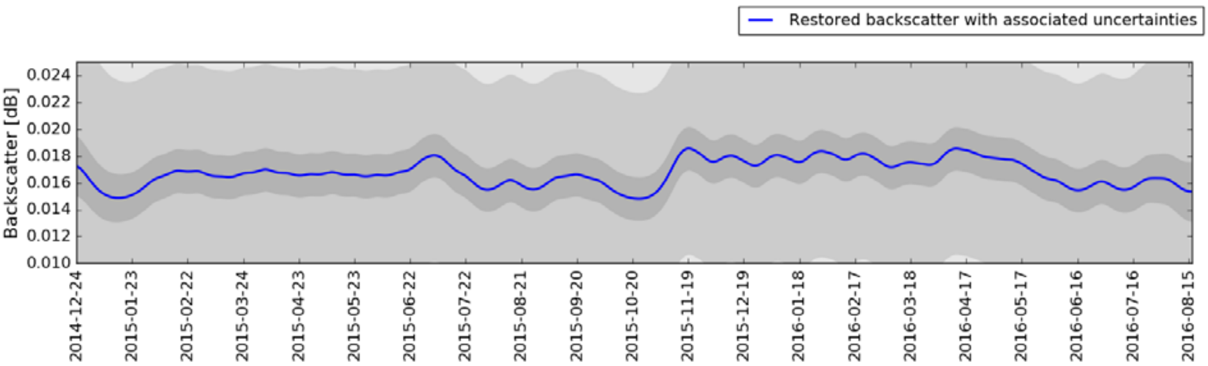

Figure 7 Sentinel-1 VH backscatter (linear) for the Wytham Woods test site, UK (BACI FT), characterised as deciduous broadleaf, for the period 10/2014 to 10/2015. Shades of grey are 1%, 5% and 25% credible interval of uncertainties.

We note that the treatment of optical and microwave data in a common regularization framework in the manner shown above is one aspect of what makes the BACI SSV unique and novel.

4.3 Temporally regularized estimates of biophysical properties

The following figure shows results of the full EO-LDAS inversion of a canopy 1D RT model, resulting in time series of: LAI, leaf chlorophyll content, leaf water content and soil brightness, with associated uncertainties. These results were obtained with the NADIM 1D semidiscrete canopy RT model (Gobron et al., 1997), using dynamic prior information and temporal regularization. The dynamic prior information constraints in this case are the beginning and end of vegetation season where we know that LAI, chlorophyll and leaf water have very small values. Temporal regularization imposes requirement of smoothness.

As discussed above, while these type of parameter retrievals are potentially of interest in terms of attribution (e.g. whether any observed change is due to vegetation or background changes), the imposition of the RT model assumptions (canopy structural, leaf optical properties here) may mask more subtle changes in the underlying SSV signal used to derive the parameters. This is why we propose that using the 'unadulterated' SSV is likely to be better from a change detection perspective.

In each of the examples below, the shaded regions represent 75% and 95% of credible interval of uncertainty in retrieval. Vertical grey lines show MODIS acquisitions, which were used for the retrieval. The variable, often sparse nature of these observations due to cloud, highlight the desirability of combining observations from multiple streams, particularly microwave where possible: even optical data from multiple sources rarely provide much better temporal sampling due to the high temporal correlation between them.

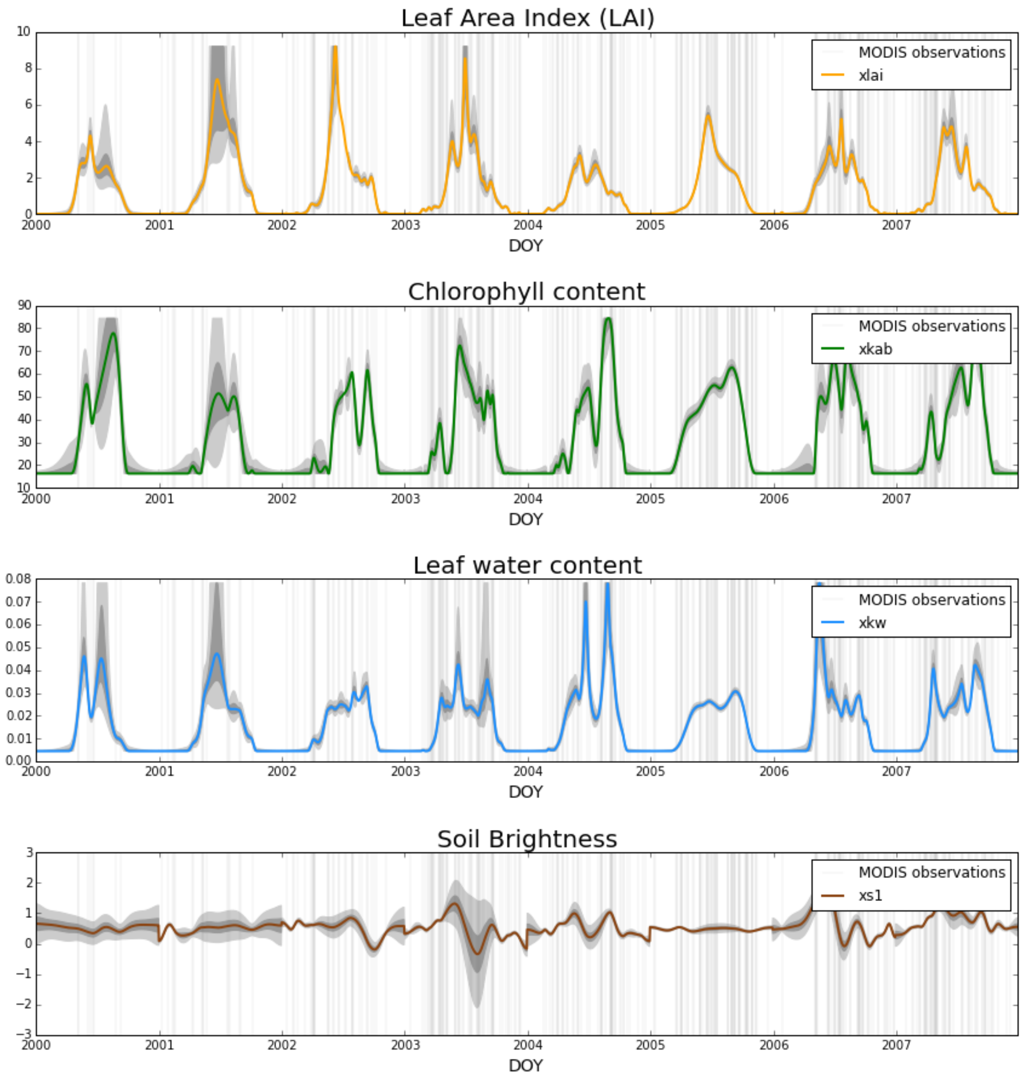

Figure 8 SSV-derived surface biophysical properties, and uncertainties for the Hainich test site 2000-2008.

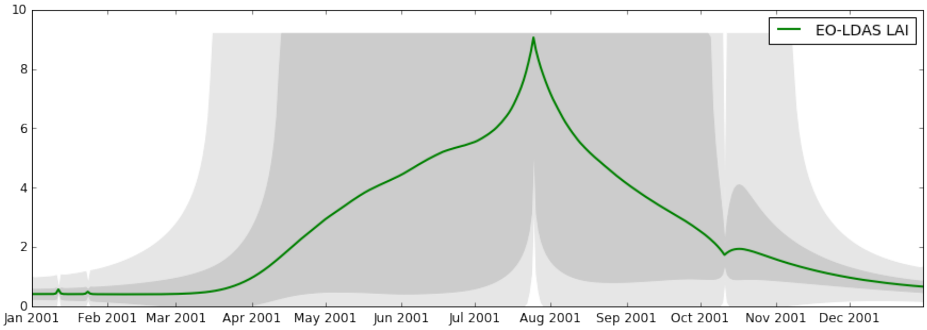

Figure 9 SSV-derived LAI and uncertainties for the Wytham Woods test site 2001.

We see in the figures above, that the uncertainty is apparently much lower than for the original SSV. Whilst this is true in the sense of the parameter uncertainty, of course this is not true necessarily in the sense of a time-series of surface state. This is in part due to the strong constraints used in the parameter retrieval of start and end of growing season.

5 Summary

The above description details the production of the BACI surface state vector (SSV), using temporal regularization within the EO-LDAS framework, to combine observations of the surface state from optical and microwave observations (**D2.2**). The resulting SSV output is an optimal estimate of temporally regularized, spatial time series of the surface, providing a unique and novel time-series vector of observations for input to the BACI ML for change detection and downstream product generation.

In conclusion, the BACI SSV framework:

- combines data across wavelengths, including optical and microwave (with different spatial and temporal properties), into a common observation vector;
- uses these data to generate: optimally smoothed and filtered time series of reflectance, albedo and backscatter, as the core SSV output, with consistent uncertainties (key for use in further quantitative modelling and change detection chain, particularly attribution);
- allows for retrieval of biophysical parameters based on RT model inversion, with quantified uncertainties;
- is flexible, to allow extension to other time-series observations, notably new Sentinel observations (1-3), but also including potentially thermal (e.g. MODIS LST) and historical microwave (ENVISAT ASAR, where available).

The treatment of optical and microwave data in a common regularization framework in the manner demonstrated makes the BACI SSV unique and novel.

6 References

Disney, M. I., Muller, J.-P., Kharbouche, S., Kaminski, T., Vossbeck, M., Lewis, P. and Pinty, B. (2016) A new global fAPAR and LAI dataset derived from optimal albedo estimates: comparison with MODIS products, Remote Sensing, *8*(4), 275; doi:10.3390/rs8040275 (PDF version, 26MB).

EO-LDAS code base: https://github.com/jgomezdans/eoldas.

Gobron N., Pinty B. and Verstraete M. M. (1997) A Semidiscrete Model for the Scattering of Light by Vegetation. *Journal of Geophysical Research - Atmospheres*, 102(D8): 9431-9446. DOI: 10.1029/96JD04013.

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R. and Mathieu, P.-P., (2010) Carbon cycle data assimilation with a generic phenology model. Journal of Geophysical Research, 115(G4), G04017.

Lauvernet, C., Baret, F., Hascoët, L., Buis, S. and Le Dimet, F.-X., (2008) Multitemporal-patch ensemble inversion of coupled surface–atmosphere radiative transfer models for land surface characterization. Remote Sensing of Environment, 112(3), 851–861.

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

Morton, D.C., Nagol, J., Carabajal, C.C., Rosette, J., Palace, M., Cook, B.D., Vermote, E.F., Harding, D.J. and North, P.R., 2014. Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature*, *506*(7487), pp.221-224.

Quaife, T. and Lewis, P. (2010) Temporal constraints on linear BRF model parameters IEEE Transactions on Geoscience and Remote Sensing, doi:10.1109/TGRS.2009.2038901

Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., Muller, J.P. Disney, M. I., Roberts, G., Lewis, P. et al. 2002. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote sensing of Environment*, *83*(1), pp.135-148.