Landsat and Sentinel 2 and fAPAR Satellite Records Based on the MODIS MOD15 Algorithm Using Brute Force Transfer Learning

Introduction

The Global Climate Observing System (GCOS) requires daily global leaf area index (LAI) and fraction of absorbed photosynthetically active radiation (fAPAR) products at 250m resolution (Table 1, GCOS, 2016). Only the Landsat 5,7 and 8 missions and Sentinel-2a and Sentinel-2b missions offer systematic global acquisitions of satellite data records (SDRs) potentially suitable for mapping LAI and fAPAR at ≤50m resolution (European Space Agency, 2012; U.S. Geological Survey 2013).

The Simplified Level 2 Prototype Processor (SL2P) is currently the only free and open algorithm that can derive LAI and fAPAR from imagery acquired from all of these missions (Weiss and Baret, 2016). SL2P has been implemented in Google Earth Engine to facilitate global application (Fernandes et al., 2020). Validation of SL2P indicates that it’s thematic uncertainty generally meets GCOS threshold thematic uncertainty requirements over crops but not over heterogeneous landscapes such as treed vegetation (Djamai, xx; Brown, xx; Brown xx; xx ). Even if an improved version of SL2P, or another new globally applicable algorithm, could be implemented for Landsat and S2 SDRs the resulting products would still require extensive validation to quantify their thematic performance. Ideally, the products should also be consistent with those from other sensors with SDRs that satisfy the daily revisit requirement (e.g. 500m resolution MODISA3H products from MODIS on Terra and Aqua, Myneni et al. 2021)

Table 1. GCOS fAPAR and LAI mapping requirements for modelling applications (GCOS, 2016).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product | Definition | Frequency | Resolution | Thematic Uncertainty | Stability |
| Fraction of Absorbed Photosynthetically Active Radiation | FAPAR is the fraction of the incoming radiation in the 400-700 nm band absorbed by the leaves, which can be used for photosynthesis. ] | Daily | 50m | Goal: max(10%; 0.05)  Threshold:  max() | max(3%; 0.02) |
| Leaf Area Index | One half of the total green leaf area per unit horizontal ground surface area, | Daily | 50m | Goal: 15%  Threshold:  max(0.75,20%) | max(10%; 0.25) |

One solution for the production of ≤50m resolution LAI and fAPAR products consistent with MOD15A3H is to apply a version of the Moderate Resolution Imaging Spectrometer (MODIS) MOD15A1 LAI and fAPAR algorithm (Knyazhikin et al., 1999), currently used for MOD15A3H products, to Landsat and S2 SDRs. A version for Landsat 5 Thematic Mapper (TM ) and Landsat 7 Enhanced Thematic Mapper Plut (ETM+) SDRs had already been developed and shown to generally satisfy GCOS resolution and uncertainty requirements for pixels with homogenous land cover and vegetation structure and correct biome classification (Ganguly, et a.l, 2008a; Ganguly et al. 2008b). However, this version has limitations:

1. there are currently no plans to derive free and open global products or to offer the code for this task (S. Ganguly personal communication);
2. it was implemented for TM and ETM+ surface bi-directional reflectance (BRF) inputs from the LEDAPS processor (Masek et al, 2006) and would need recalibation to account for changes in the calibration of these sensors since 2008 or to Landsat 8 Operational Line Imager (OLI) or Sentinel 2 Multispectral Istrument SDR;
3. it deviates from MOD15A1 in that it uses a shortwave infrared (SWIR) band in addition to a red and near-infrared input band that may decease uncertainty but also consistency;
4. it was implemented in a high performance compute environment that is not open for use. The goal of this study is to develop, implement and validate an algorithm that closely approximates MOD15v6 for Landsat or S2 SDRs and that is capable of systematic processing of regional (>10,000km2) products in a free and open manner.

MOD15v6 is a deterministic algorithm that, given one of xx biome types (Table 2), relates input MODIS red and near infrared bi-directional reflectance, view and solar zenith angle and their relative azimuth angle to the expected value and standard error of either LAI or fAPAR. MOD15A1 is implemented using a cascade of look-up tables (LUTs) populated with radiative transfer (RT) model simulations of reflectance and fAPAR based on prior distributions for biome specific RT model inputs and sensor specific observation geometry.

Transfer learning is a machine learning approach that uses knowledge gained while solving one problem and applies it to solving a different but related problem (Goodfellow et al., 2015). In this context, the problem at hand is mapping Landsat or S2 SDRs to LAI and fAPAR estimates consistent with MOD15A3H and meeing GCOS requirements and the related problem is the MOD15v3 algorithm that maps MODIS inputs to MOD15A3H product outputs.

Transfer learning has been used to approximate the MOD15A1 algorithm for application to SDRs from other imagers (Verger, xx; Campos-Taberner et al. xx; Kang, xx). These studies have used non-linear regression to estimate the mapping given the calibration database in the name of computational efficiency (Verger, xx; Kang, xx). There are three major limitations with this approach: i) a full recalibration must generally be performed to incorporate additional samples, ii) regression assumes conditional residuals are independent and identically distributed while the highly non-linear nature of the relationship between input reflectance and variables such as LAI indicate this may not be satisfied, iii) it is difficult to arrive at unbiased estimates of LAI and fAPAR and their standard error in the presence of measurement error in both inputs and outputs, as is the case when using noisy Landsat or S2 SDRs and MOD15A3H product estimates.

A brute force approach to transfer learning is proposed to address the limitations of regression approaches. Biome specific LUTs for the expected value and standard error of either LAI and fAPAR are populated using a large calibration database of spatially and temporally matched upscaled target clear sky Landsat or S2 surface reflectance inputs and 500m resolution Terra MODIS MOD15A3H outputs. The LUT is then applied to full resolution imagery corresponding to the inputs.

The purpose of this study is to develop and document an open source LUT based transfer learning algorithm to apply an equivalent of the MDO15A1 LAI and fAPAR algorithm to ETM+ imagery. A secondary purpose is to validate derived Landsat and S2 products from this algorithm using in-situ reference data and to quantiy consistency by intercomparison with MOD15A3H product at 500m resolution.

We hypothesize that brute force transfer learning using paired 500m resolution ETM+ inputs and MOD15A3H1 outputs will be sufficient to populate a LUT to provide retrievals for all valid clear sky Landsat or S2 measurements over vegetation. This will be tested by determining the frequency of unmatched inputs as a function of calibration database size and biome.

We hypothesize that the difference between 500m resolution Landsat or S2 products and MOD15A3H products will be substantially less than the MDO15A3H standard error This will be tested by quantifying the ratio of the root mean square difference (RMSD) between products to the standard error of the MOD15A3H estimates as a function of biome type and ETM+ variable estimated.

We hypothesize that the bias and uncertainty of Landsat and S2 products when compared to in-situ sites will be similar to the same statistics reported in the literature for other direct implementations of MOD15A1 (Ganguly et al. 2008b) or transfer learning implementations (Kang et al., 2020). Here similarity will be quantified as the ratio of the difference over similar validation sites compared to the typical standard error of the Landsat or S2 estimates for a given biome and estimated variable.

Theory

For a given biome (Table 2), MOD15A1 is a deterministic mapping between six inputs (MODIS Band 1 red BRF, MODIS Band 2 NIR BRF, view zenith and azimuth angle, and solar zenith and azimuth angle) and the expected value of LAI or fAPAR and its associated standard error. The mapping is not performed by direct look-up but by on demand computation using matches between observed inputs and a calibration database of paired inputs and outputs. The calibration database corresponds to radiative transfer (RT) model simulations of MODIS inputs and outputs based on a prior distribution of canopy and soil parameters and MODIS aquisition geometry. Measurement error is not added to entries in the calibration. The matching returns the expected value and standard error of LAI and fAPAR of all calibration database simulations for the biome within a specified range of absolute Band 1 and Band 2 BRF (Table 2). In this sense MOD15A1 is actually a functional mapping as for any measured input there is only one expected value and associated standard error for Lai and for fAPAR.

There are two distinct approaches to transferring a deterministic functional relationship such as MOD15A1 to a target Landsat or S2 imager: i) apply the composite mapping of the the MOD15A1 and the mapping between Landsat and S2 sensors measurements and MOD15A1 inputs (e.g. Rochdi et al. xx) , ii) ) recalibrate the simulations in the calibration database for the spectral and acquisition geometry of the target imager (e.g. Campos-Tabener et al. xx, Kang et al. xx). We do not adopt the composite mapping approach for two reasons. Firstly, the relationship between the target sensor and MOD15A1 inputs is not functional in that, for the same acquisition geometery, there can be many possible MODIS Band 1 and MODIS Band 2 BRFs for an observed target sensor BRF in similar red and NIR bands due to differences in the spectral response function. One could in principle include additional target sensor input bands to predict the MOD15A1 input bands but this relationship would depend itself on canopy and soil properties (Trichtechnko et al. xx) leading to a level of circularity and possibly biased estimates. Secondly, even in the relationship could be defined with uncertainty much less than the MOD15A1 assumed input uncertainties the propagation of these uncertainties into the estimated standard error would require additional computation that may be challenging to apply for free and open product generation on demand. Instead, we chose to follow the established approach of recalibrating a new database using paired inputs from the target sensor selected and scaled to match corresponding MOD15A1 outputs.

Previous studies have relied on regression estimators to predict LAI or fAPAR based on a calibration database of matching paired inputs from a target fine resolution Landsat sensor and outputs from MOD15A3H or associated products. Regression estimators offer a finite and modest computation demand and can also provide associated prediction standard errors. However, regression makes two fundamental assumptions that may not apply to this transfer learning problem.

1. Firstly, the vast majority of regression methods assume that measurement error of inputs is significantly less than the expected residuals for a perfect model (Draper and Smith, xx). Consistent, i.e. those whose expected value converges as the calibration sample size increases, univartiate regression estimators are available (Fernandes and Leblanc, xx) but there is no generaly analytical solution for consistency or unbiased estimates with multivariate regression. Ad hoc approaches, such as calibration using data with added measurement error (Verger et al., xx) have been used but require significantly greater computation during calibration and do not also provide unbiased estimates of prediction standard error.

ii) regression assumes that conditional residuals are independent and identically distributed (iid). This is typically not the case in the so called saturation zone of the MDO15A1 algorithm where the scater of LAI or fAPAR for matching entries in a calibration database is typically skewed positive and can have a coefficient of variation exceeding one. Indeed, the cross-validation uncetrainty of a regression tree approach based on Landsat OLI Band xx (red) and Band xx (NIR) inputs increased by ~xx% when comparing retrievals in the saturation zone versus the unsaturated zone. While it may be possible to reduce this effect using additional spectral bands as input predictors this has yet to be demonstrated and this would further increase the possibility of increased uncertanity due to measurement errors of the additional bands.

While the impact of measurement error and iid residuals may not always be an issue for the current transfer learninig problem it is reasonable to wonder if there is a transfer learning solution that does not suffer from these problems. We suggest that a LUT coded using a perfect hash function based on a large number of calibration samples mined from matching target sensor and MOD15A3H measurements will address the issues identified while provding a solution with fixed and modest ccomputations with known probability. A binary search of a hash tablerequires at worst log2(n)+1 comparisons . A perfect hash ensures that there is a unique code for each calibration sample so there is no chance of conflicts in LUT addressing. Here we use a code based on simply concatenating scaled and quantized versions of the input and output pair of each calibration sample.

Both computational aspects such as LUT size, performance and calibration time and thematic aspects related to the additional uncertainty due to the transfer learning algorithm depend on the LUT quantization and measurement error of the input and outputs of the calibration samples. These two factors are related in that the impact of input and output quantization on uncertainty will be small where measurement error is large. MOD15A1 scaled outputs are integers ranging from 0 to 100 (Table 3) so an output quantization with two decimal digits will result in no loss in precision assuming a value of 100 is not observed; and typically a value of 100 is associated with a large standard error due to saturation so replacing it with a value of 99 sufficies. The impact of input quantization is more challenging to quantify as its relationship to output uncertainty is non-linear. It can be predicted by noting that the effect of input quantization and measurement error of the input are independent. In this case the the percentage of additional error (Ea) due to a ratio of input quantization to input measurement error (R) is give by:

Ea = 1 – (1^2+Eq^2)^0.5 Equation 1

Equation 1 implies an input quantization error of 0.45 the measurement error will only increase the input error by 10% relative to the error without quantization and, in the case of a locally linear relationship between input and output values, an output error increase by 10%. Of course, the relationship between input and output can be highly non-linear in the saturation zone. In this case, a non-linear quantization can be applied (e.g. for low red reflectances) and the actual error can be estimated by comparing estimates with different levels of quantization.

The MDO15A1 algorithm uses 6 inputs per biome type (Table 4). While it is not a constraint, if that same 6 similar inputs were used by target sensors, a modest input quantization to integers in [0,99] for each input gives a maximum possible bioe specific LUT size of 10^8 elements. Assuming the LUT is sorted a binary search requires <log2(N)+1 operations to find a match. This in turn implies 8 scaling and quantization and <27 look-up operations to find a match in the LUT given a measurement. Both the LUT size and the computation demands for look-up are comparable to regression approaches used for similar transfer learning problems. The challenge of populating the LUT while minizing measurement error of inputs is addressed in Section xx. However, it is possible that not all possible LUT elements can be calibrated using the sampling approach we implement. It is still possible to first model the likelihood of failed match by hold out sampling and to secondly offer the user the option to perform targeted additional sampling in circumstances where failed matches exceed this likelihood. The latter option being facilitated by the use of a perfect hash since one can simply add to a sorted LUT list between the two closest matches.

Notwithstanding sufficient precision in LUT entries, a match between a measurement from a target sensor and a LUT entry will still result in additional uncertainty due to the fact that the target sensor inputs are not identical to MOD15A1. In this case the LUT will no longer be a functional mapping since there can be multiple observed MOD15A1 outputs for a given target sensor input observed during calibration. This is conceptually the same , but independent, type of error as the uncertainty of MOD15A1 inputs (Table 2) . It follows that this additional uncertainty can be quantified by simply returning all matching retrievals. However, to reduce computation and complexity we propose to simply reduce the relational mapping LUT to a functional mapping LUT using the median LAI of all matches with the same input as the new expected value and using the Euclidean sum of the average standard error of these matches and the standard deviation of the LAI across these matches as the new standard error estimate. The advantage of this approach is that the standard deviation of LAI across macthes is an estimate the additional uncertainty due to the target sensor inputs not perfectly matching MOD15A1 inputs.

Table 2. MOD15A1 biome types and associated absolute uncertainty (% BRF) of input MODIS Band 1 and Band 2 BRF to the MOD16v6 algorithm.

|  |  |  |
| --- | --- | --- |
| **Biome** | **Uncertainty Level** | |
|  | **% BRF Band 1** | **% BRF Band 2** |
| Grasses/Cereal crops | 20 | 5 |
| Shrubs | 20 | 5 |
| Broadleaf crops | 20 | 5 |
| Savanna | 20 | 5 |
| Evergreen Broadleaf forest | 30 | 15 |
| Deciduous Broadleaf forest | 30 | 15 |
| Evergreen Needleleaf forest | 30 | 15 |
| Deciduous Needleleaf forest | 30 | 15 |

Table 3. Output specification of MOD15A1 algorithm as implemented for MOD15A3H products.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **MOD15A3H**  **Band**  **Name** | **Description** | **Units** | **Data Type** | **Fill Value** | **Valid Range** | **Scale Factor** |
| Fpar\_500m | Fraction of Photosynthetically Active Radiation | % | 8-bit unsigned integer | 249-255 | 0 to 100 | 0.01 |
| Lai\_500m | Leaf Area Index | m²/m² | 8-bit unsigned integer | 249-255 | 0 to 100 | 0.1 |
| FparLai\_QC | Quality for  FPAR and LAI | Class Flag | 8-bit unsigned integer | 255 | 0 to 254 | N/A |
| FparExtra\_QC | Extra detail Quality for  FPAR and LAI | Class Flag | 8-bit unsigned integer | 255 | 0 to 254 | N/A |
| FparStdDev\_500m | Standard deviation of  FPAR | % | 8-bit unsigned integer | 248-255 | 0 to 100 | 0.01 |
| LaiStdDev\_500m | Standard deviation of LAI | m²/m² | 8-bit unsigned integer | 248-255 | 0 to 100 | 0.1 |

Table 4. Input specification of MOD15A1 algorithm as implemented for MOD15A3H products.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **MOD09GA**  **Band**  **Name** | **Description** | **Units** | **Data Type** | **Fill Value** | **Valid Range** | **Scale Factor** |
| sur\_refl\_b01\_1 | Surface Reflectance Band 1 | N/A | 16-bit signed integer | -28672 | -100 to 16000 | 0.0001 |
| sur\_refl\_b02\_1 | Surface Reflectance Band 2 | N/A | 16-bit signed integer | -28672 | -100 to 16000 | 0.0001 |
| SolarZenith | solar zenith angle | ° | 16-bit signed integer | 0 | 0 to 18000 | 0.01 |
| SensorZenith | view zenith angle | ° | 16-bit signed integer | 0 | 0 to 18000 | 0.01 |
| SolarAzimuth | solar azimuth angle | ° | 16-bit signed integer | 0 | -18000 to 18000 | 0.01 |
| SensorAzimuth | View azimuth angle | ° | 16-bit signed integer | 0 | -18000 to 18000 | 0.01 |

Data Sets

Google Earth Engine datasets are summarized in Table 5 together with links to substantial documentation of algorithms and products. As such, only the aspects of these data relevant to the transfer learning algorithms or generated products are discussed here. Input data layers to transfer learning algorithms are documented in the methods section.

Table 5. List of data sets used.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product ID | Description | Date range | GEE Link | Reference |
| MCD15A3H | Combined 4-day 500m fAPAR and LAI | 2002-07-04, 2021-08-29 | https://developers.google.com/earth-engine/datasets/catalog/MODIS\_006\_MCD15A3H | Myneni et al. 2021. |
| MCD12Q1 | 500m resolution global land cover types at yearly intervals | 2001-01-01, 2019-01-01 | https://developers.google.com/earth-engine/datasets/catalog/MODIS\_006\_MCD12Q1 | Friedl and Sulla-Menashe, 2019. |
| MOD09GA | MODIS bands 1-7 BRF at 500m with 1km acquisition geometry | 2000-02-24, 2021-09-07 | https://developers.google.com/earth-engine/datasets/catalog/MODIS\_006\_MOD09GA | Vermote and Wolfe, 2021. |
| MSI | Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | 2017-03-28, 2021-09-09 | https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS\_S2\_SR#description | ESA, 2015. |
| OLI | USGS Landsat 8 Level 2, Collection 2, Tier 1, surface BRF | 2013-04-11, 2021-08-29 | https://developers.google.com/earth-engine/datasets/catalog/LANDSAT\_LC08\_C02\_T1\_L2 | USGS, 2020a. |
| ETM+ | USGS Landsat 7 Level 2, Collection 2, Tier 1, surface BRF | 1999-01-01, 2021-08-14 | https://developers.google.com/earth-engine/datasets/catalog/LANDSAT\_LE07\_C02\_T1\_L2 | USGS, 2020b. |
| TM | USGS Landsat 5 Level 2, Collection 2, Tier 1, surface BRF | 1984-01-01, 2012-05-05 | https://developers.google.com/earth-engine/datasets/catalog/LANDSAT\_LT05\_C02\_T1\_L2 | USGS, 2020c. |
| NALCMS | North American land cover 30m, 5yr | 2000-2020 | http://www.cec.org/north-american-land-change-monitoring-system/ | Latifovic et al., 2012. |
| Corine | Europe land cover, 100m, 6yr | 2000-2018 | https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS\_CORINE\_V20\_100m?hl=en | Feranec et al. , 2016. |
| CGLS | Global land cover, 100m, circa 2015-1019 | 2015 | https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS\_Landcover\_100m\_Proba-V-C3\_Global?hl=en | Buchhorn et al., 2020. |
| MOD44W | Land/water mask, 250m, 1yr | 2000-01-01, 2015-01-01 | https://developers.google.com/earth-engine/datasets/catalog/MODIS\_006\_MOD44W?hl=en | Carrol et al., 2017. |
| NEON-Insitu | In-situ LAI and fAPAR measurements over USA from the NEON network. | 2016-02-01, 2021-07-01 |  | Meier et al., xx |
| CCRS-Insitu | In-situ LAI and fAPAR measurements over Canada from the CCRS network. | 2018-06-01-2019-09-01 |  | Fernandes et al., xx |

MCD15A3H

MCD15A3H are four day maximum fAPAR global 500m LAI and fAPAR composites produced from MODIS on Terra abd Aqua using MOD15v6. The product includes additional per pixel information of relevance to our study including: associated standard errors of estimates, the sensor platform, if the retrieval corresponded to the saturated or unsaturated region of the mapping, if the main ratiative transfer algorithm was applied. Here, only retrievals from the main radiative transfer algorithm are used and the saturated and unsaturated retrievals are used separately.

MCD12Q1

MCD12Q1 is an annual 500m global land cover map that includes the MCD15A3H biome type. This map is used to calibrate biome specific retrieval algorithms. The geometry is identical to MCD15A3H products.

MOD09GA

MOD09GA is a single day composite of surface BRF for MODIS bands 1-7 and associated acquisition geometry at 1km resolution. This product is used to filter MCD15A34H retrievals to those that match Landsat or SDR acquisition dates and have a limited range of view zenith angle. The geometry is identical to MCD15A3H products.

MOD44W

MOD44W is a monthly land/water mask derived using 250m resolution MODIS imagery using a pixel based supervised classification algorithm (xx). The mask is used to screen out MCD15A3H retrievals that may potentially contain sub-pixel water not represented in the annual MCD12Q1 maps. The geometry is identical to MCD15A3H products when pixels are aggregated to corresponding 500m resolution grids.

MSI

MSI is a Level 2 Geolocated (L2G) surface BRF from the Multispectral Instruments of either the Sentinel 2a or 2b imagers. Each product consists of 15 bands at a spatial resolution of either 10m, 20m or 60m together with acquisition geometry, a simplified land cover label, and clear sky flags.

The radiometric uncertainty is

The geometry uncertainty is

The inputs used for LUT calibration and product retrieval are given in Table xx.

OLI

OLI is a L2G surface BRF from the Landsat 8 Operational Land Imager. Each product consists of eight 30m bands spanning visible, NIR and SWIR and two 100m thermal infrared bands, together with solar geometry and spacer across track roll angle at the scene centre, a land/water label, and clear sky flags.

The radiometric uncertainty is

The geometry uncertainty is

The inputs used for LUT calibration and product retrieval are given in Table xx.

ETM+

ETM+ is a L2G surface BRF from the Landsat 7 Enhanced Thematic Mapper Plus imager. Each product consists of seven 30m bands spanning visible, NIR and SWIR, one 60m thermal infrared band, together with solar geometry and sensor attitude at the scene centre, a land/water label, and clear sky flags. Products after May 31, 2003 suffer from a scan line corrector failure artifact corresponding to missing pixels for alternate scans within each image.

The radiometric uncertainty is

The geometry uncertainty is

The inputs used for LUT calibration and product retrieval are given in Table xx.

TM

ETM+ is a L2G surface BRF from the Landsat 5 Thematic Mapper imager. Each product consists of six 30m bands spanning visible, NIR and SWIR, one 60m thermal infrared band, together with solar geometry and sensor attitude at the scene centre, a land/water label, and clear sky flags.

The radiometric uncertainty is

The geometry uncertainty is

The inputs used for LUT calibration and product retrieval are given in Table xx.

## MOD44W Terra Land Water Mask Derived From MODIS and SRTM Yearly Global 250m

This mask corresponds to an annual land or water classification based on a decision tree classified using calibration based on the SRTM Water Body Data. The mask is included in MOD09GQ and used to specify areas not mapped in the 250m LAI and fAPAR products.

## 250m Resolution Biome Type

Three land cover products were used to specify the biome type for 250m LAI and fAPAR products using legend conversions (Table 9)

Table 9. Legend conversion between available land cover products and MODIS bomes.

|  |  |  |  |
| --- | --- | --- | --- |
| **Biome** | **Input Land Cover Legend Classes** | | |
|  | **NALCMS** | **CORINE** | **CGLS** |
| Grasses/Cereal crops |  |  |  |
| Shrubs |  |  |  |
| Broadleaf crops |  |  |  |
| Savanna |  |  |  |
| Evergreen Broadleaf forest |  |  |  |
| Deciduous Broadleaf forest |  |  |  |
| Evergreen Needleleaf forest |  |  |  |
| Deciduous Needleleaf forest |  |  |  |

### North American Land Change Monotoring System (NALCMS) 30m Land Cover

The North American Land Cover Monitoring System 30m land cover maps correspond to conequitive 5 year intervals from 2000 to 2020. These maps provide consistent change updating with an overall accuracy of >xx% for the simplified legend used to map MODIS biome types (Table 8).

### Corine Land Cover

The Corine 100m land cover corresponds to 6 year intervals from 2000 to present for Europe . These maps provide consistent change updating with an overall accuracy of >xx% for the simplified legend used to map MODIS biome types (Table 8).

### Copernicus Global Land Service (CGLS) Land Cover

The CGLS 100m land cover corresponds to a circa 2016-2020 global land cover. The map has an accuracy of <xx% for the simplified leged used to map MODIS biome types (Table 8). The map is not optimal for retrievals outside this period but there is no other consistent land cover with all MODIS biomes at <=250m resolution outside of North America and Europe available on GEE at present.

Methods

LUT Calibration

The LUT was calibrated using spatially and temporally matching target sensor inputs (Table xx, Table xx, Table xx or Table xx) and MOD15A1 outputs (Table xx).

Matching was performed on a target sensor granule basis since their revist frequency was far less than MOD15A1.

The MOD15A1 product containing the granule and the product for prior and following 4d periods was extracted from archive.

The MOD15A1 product was masked for all pixels where the main RT algorithm was not successful for the three extracted dates.

To minimize biome misclassification the MOD15A1 product was masked for all pixels where the biome changed in any one pixel in a 3x3 neighbourhood between the current and adjanced years in the MCD12 product.

Pixels where the MOD15A1 product was derived for the same date as the target sensor were found where possible. Due to the lack of a pixel specific production time in MOD15A1 this was performed by finding all MOD15A1 terra retrievals with only one corresponding MOD09GQ retrieval in the 4d interval. This was not exhaustive but at least found pixels where a temporal match is assumed.

For all other pixels, to minize temporal mismatch, the MOD15A1 product was masked for all pixels where for any one pixel in a 3x3 neighbourhood where the maximum change in LAI exceeded twice the average standard error of the three retrievals.

A layer indicating valid clear sku target pixels over land was produced and aggregated to all unmasked MOD15A1 pixels using a mean value as an indicator of the percentage valid target input pixels. MOD15A1 pixels where the 3x3 neighbihood had less than xx% valid target pixels were masked.

Inputs from target sensor pixels were aggregated using their mean to matching unmasked MOD15A1 pixels.

For Landsat sensors, the view zenith angle was estimated assuming the sensor track fell along the central axis of the rhomboid defined by the scene corner points in meter coordinates under local UTM projection and that the east and west edges of the rhomboid corresponded to the nominal flat earth scan angle of the sensor (Figure xx). This approximation may result in errors in areas in topographic relief or near the poles. The latter are not a concern due to the lack of vegegation but the former could result in some uncertanities in view zenith angle. The view azimuth angle was not available for Landsat sensors and was defaulted as 90 degrees to ensure it could not be interpreted as a principle plane observation which would be rare for Landsat sensor orbits.

For S2 imagers, the mean incidence view azimuth and zenith angle for the Band 4 (red) detectors was used for the granule.

The 3x3 pixel average of aggregated target inputs and corresponding MOD15A1 outputs were added to the unsaturated calibration database unless one of the input MOD15A1 outputs were flagged saturated, in which case they were added to the saturated calibration database.

Each database was coded into a perfect hash table for the relational mapping using two decimal digits per input or output as indicated in Table xx and Table xx and the example in Table xx.

A second hash table for a functional mapping was derived from the relational hash table but summarizing the median LAI or fAPAR for unique input codes together with a new standard error corresponding to the Euclidean sum of the mean standard error of duplicates and the standard deviation of the duplicates. A second table with the standard deviation of the duplicates following the same order as the entries of the functional hash table was produced to provide a diagnostic for the increase in uncertainty due to the transfer learning process.

Results

Discussion

Conclusions

References

U.S. Geologogical Survey, 2013. Landsat – A global imaging mission. U.S. Department of the Interior, Fact Sheet 2012-3072, https://pubs.usgs.gov/fs/2012/3072/fs2012-3072.pdf accessed on September 7, 2021.

European Space Agency, 2012. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services, Editor K.Fletcher, ESA SP-1322/2, https://sentinel.esa.int/documents/247904/349490/S2\_SP-1322\_2.pdf accessed on September 7, 2021.

Myneni, R., Knyazikhin, Y., Park, T., 2021. MODIS/Terra+Aqua Leaf Area Index/FPAR 4-Day L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-09-09 from https://doi.org/10.5067/MODIS/MCD15A3H.061

Ganguly, S., Samanta, A., Schull, M.A., Shabanov, N.V., Milesi, C., Nemani, R.R., Knyazikhin,

Y. and Myneni, R.B., 2008b. Generating vegetation leaf area index Earth system data record

from multiple sensors. Part 2: Implementation, analysis and validation. Remote Sensing of

Environment, 112(12), pp.4318-4332.

Ganguly, S., Schull, M.A., Samanta, A., Shabanov, N.V., Milesi, C., Nemani, R.R., Knyazikhin,

Y. and Myneni, R.B., 2008a. Generating vegetation leaf area index earth system data record

from multiple sensors. Part 1: Theory. Remote Sensing of Environment, 112(12), pp.4333-

4343.

Goodfellow, I., Benigo, Y., and Courville, A., 2015. Deep Learning. MIT Press, Cambride, 777 pp.

Masek, J. G., Vermote, E. F., Saleous,N. E.,Wolfe, R.,Hall, F.G.,Huemmrich, K. F., et al. (2006).

A Landsat surface reflectance dataset for North America, 1990–2000. IEEE Geoscience

and Remote Sensing Letters, 3(1), 68–72, doi:10.1109/LGRS.2005.857030

USGS, 2020a. Landsat 8 Collection 2 (C2) Level 2 Science Product (L2SP) Guide. US Department of Interior, LSDS-1619 Version 2.0 ,36pp. [https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1619\_Landsat8-C2-L2-ScienceProductGuide-v2.pdf accessed on September 10](https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1619_Landsat8-C2-L2-ScienceProductGuide-v2.pdf%20accessed%20on%20September%2010), 2021.

USGS, 2020b. Landsat 4-7 Collection 2 (C2) Level 2 Science Product (L2SP) Guide, LSDS-1618 Version 3.0 ,36pp.  [https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1618\_Landsat-4-7\_C2-L2-ScienceProductGuide-v3.pdfaccessed on September 10](https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1619_Landsat8-C2-L2-ScienceProductGuide-v2.pdf%20accessed%20on%20September%2010), 2021.

ESA, 2015. Sentinel 2 User Handbook. European Space Agency, Issue 1 Rev 2, 64pp. [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\_User\_Handbook accessed on September 10](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook%20accessed%20on%20September%2010), 2021.

Vermote, E., Wolfe, R., 2021. MODIS/Terra Surface Reflectance Daily L2G Global 250m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2021-09-09 from <https://doi.org/10.5067/MODIS/MOD09GQ.061>.

Campos-Taberner, M.; Moreno-Martínez, Á.; García-Haro, F.J.; Camps-Valls, G.; Robinson, N.P.; Kattge, J.; Running, S.W. Global Estimation of Biophysical Variables from Google Earth Engine Platform. Remote Sens. **2018**, 10, 1167. https://doi.org/10.3390/rs10081167