

A data-driven approach to estimate leaf area index for Landsat images over the contiguous US



Yanghui Kang^{a,b,c,*}, Mutlu Ozdogan^{b,d}, Feng Gao^c, Martha C. Anderson^c, William A. White^c, Yun Yang^c, Yang Yang^c, Tyler A. Erickson^e

^a Department of Geography, University of Wisconsin-Madison, 550 N. Park St., Madison, WI 53706, USA

^b Nelson Institute Center for Sustainability and the Global Environment, University of Wisconsin-Madison, 1710 University Avenue, Madison, WI 53726, USA

^c Hydrology and Remote Sensing Laboratory, USDA ARS, Beltsville, MD 20705, USA

^d Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA

^e Google Inc., Mountain View, CA 94043, USA

ARTICLE INFO

Editor: Jing M. Chen

Keywords:

Leaf area index

Landsat

MODIS

Machine learning

Google earth engine

ABSTRACT

Leaf Area Index (LAI) is a fundamental vegetation biophysical variable serving as an essential input to many land surface and atmospheric models. Long-term LAI maps are typically generated with satellite images at moderate spatial resolution (0.25 to 1 km), such as those from the Moderate Resolution Imaging Spectroradiometer (MODIS). While useful for regional-scale land surface modeling, these moderate resolution products often cannot resolve spatial heterogeneity important for many agricultural and hydrological applications. This paper proposes an approach to map LAI at 30-m resolution based on Landsat images for the Contiguous US (CONUS) consistent with the MODIS product, aimed at multi-scale modeling applications. The algorithm was driven by 1.6 million spatially homogeneous samples derived from MODIS LAI and Landsat surface reflectance products from 2006 to 2018. Based on these samples, we trained separate random forest models to estimate LAI from Landsat surface reflectance for eight biomes of the National Land Cover Database (NLCD). A balanced sample design regarding the saturation status of MODIS LAI and a machine-learning-based noise detection technique were introduced to mitigate the trade-off in estimation accuracy between medium LAI (e.g., 3 to 4, unsaturated) and high LAI (e.g., 4–6, saturated).

This approach was evaluated using ground measurements from 19 National Ecological Observatory Network (NEON) sites and eight independent sites from other sources. These sites comprise a representative sample of forests, grasslands, shrublands, and croplands across the US. For NEON sites, the LAI estimates show an overall Root Mean Squared Error (RMSE) of 0.8 with r^2 of 0.88. For the eight independent sites, the Landsat LAI algorithm achieves RMSE between 0.52 and 0.91. The uncertainty in Landsat estimated LAI varies across biomes and locations. The proposed algorithm was implemented on the Google Earth Engine platform, allowing for the rapid generation of long-term high-resolution LAI records from the 1980s using Landsat images (code is available at <https://github.com/yanghuikang/Landsat-LAI>). Our findings also highlight the importance of sample balance on regression-based modeling in remote sensing applications.

1. Introduction

Leaf Area Index (LAI) is an essential vegetation structural variable driving water and carbon fluxes, energy exchanges, and net primary productivity in global ecosystems (Baret et al., 2013; Fang et al., 2019). Long-term LAI records, established by moderate-resolution satellite observations (250–1000 m), have been used to understand phenology (Valderrama-Landeros et al., 2016; Verger et al., 2016), ecosystem

productivity (Alton, 2016; Zhou et al., 2017), and global vegetation change (Alton, 2018; Chen et al., 2019; Morton et al., 2014). However, many hydrological and agricultural applications require the ability to resolve spatial heterogeneity that exists in many natural and managed ecosystems (Anderson et al., 2016; Kang and Özdogan, 2019). Hence, data from decametric resolution (10–100 m) sensing systems, such as the Landsat and Sentinel missions, are more desirable (Kang et al., 2016; Malenovský et al., 2012; Meyer et al., 2019; Wang et al., 2019).

* Corresponding author at: Department of Geography, University of Wisconsin-Madison, 550 N. Park St, Madison, WI 53706, USA.

E-mail address: kangyanghui@gmail.com (Y. Kang).

Many studies have used Landsat or Sentinel images to estimate LAI (Fernandes et al., 2014b; Ganguly et al., 2012; Meyer et al., 2019; Mousivand et al., 2015; Pfeifer et al., 2012; Verrelst et al., 2015b; Wang et al., 2019; Weiss and Baret, 2016). Empirical relationships can be established between satellite-derived indicators and ground measured LAI (Gonsamo and Chen, 2014; Houborg and McCabe, 2018a; Kang et al., 2016; Turner et al., 1999; Wang et al., 2019). Commonly used indicators include vegetation indices (VIs), which are arithmetic transformations of multiple spectral bands to maximize the sensitivity to vegetative components (Curran and Steven, 1983; Frampton et al., 2013; Viña et al., 2011; Wang et al., 2005). With sufficient ground samples, this approach proves the most accurate in relatively small areas (Walther et al., 2004) and has also been used to validate LAI retrievals at coarser resolution (Fang et al., 2019; Fernandes et al., 2003; Garrigues et al., 2008; Xu et al., 2018). However, at regional to global scales, ground observations are only sparsely available, and the underlying LAI-VI relationship can be confounded by external factors, such as soil background, leaf chlorophyll content, canopy structure, crop type, and sun-surface-sensor geometries (Fang et al., 2019; Levitan et al., 2019). Thus, it is impractical to apply empirical functions to large areas without significant bias in underrepresented areas or vegetation types (Kang et al., 2016; Levitan et al., 2019).

Another approach is to estimate LAI by solving a radiative transfer (RT) model with search algorithms or Look Up Tables (LUT), where in-situ LAI measurements are not needed (Ganguly et al., 2012; Goel and Thompson, 1984; Houborg and Boegh, 2008; Houborg and McCabe, 2018a). The RT approach has been used to produce global moderate-resolution LAI products, such as from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Myneni et al., 2002). Ganguly et al. (2012) formalized an LAI estimation algorithm for Landsat based on RT modeling and LUT. Instead of using LUT, a hybrid approach employs machine learning to map the relationships between LAI and satellite observations based on RT model simulations (Verrelst et al., 2015a). For example, the CYCLOPES (Baret et al., 2007) LAI product is generated using neural networks trained on PROSPECT (Jacquemoud and Baret, 1990) and SAIL (Verhoef, 1984) model (PROSAIL) simulations. The Sentinel-2 Level 2 LAI product is built in a similar manner using neural networks and PROSAIL model outputs (Fernandes et al., 2014b; Weiss and Baret, 2016). Yet, challenges remain for the RT methods to effectively regularize the ill-posed model inversion without site-specific information about canopy structure, biochemical properties, and soil background (Combal et al., 2003; Fang et al., 2019; Houborg et al., 2015; Houborg and McCabe, 2018a).

A data-fusion-based approach has also been proposed to leverage existing moderate-resolution LAI products derived from RT inversion and use machine learning models to associate surface reflectance from other satellites with existing LAI records. Both GEOF (Baret et al., 2016; Baret et al., 2013) and GLASS (Xiao et al., 2016) global moderate-resolution products were derived from an integration of MODIS (Myneni et al., 2002; K. Yan et al., 2016a) and CYCLOPES (Baret et al., 2007) products to exploit the merits of both products and reduce uncertainty. For decametric resolution sensors, Gao et al. (2012) introduced an algorithm to generate Landsat-based LAI using homogeneous MODIS LAI data. This algorithm identifies MODIS pixels where Landsat surface reflectance is relatively homogenous to mitigate the scale effect. Once a homogenous sample set is established, a tree-based machine learning model (Cubist) is trained to predict LAI from Landsat images (Gao et al., 2014; Gao et al., 2012). Another method to associate surface reflectance at decametric resolution with LAI at moderate resolution is through pixel unmixing. For example, Zhou et al. (2018) disaggregated MODIS LAI to Landsat pixel levels and established machine learning models based on surface reflectance and unmixed LAI at 30 m.

The data fusion approach has two notable merits. First, like the RT-based methods, it does not require ground measured LAI samples and could potentially generalize towards large areas. Second, this approach, when applied to decametric sensors (e.g., Landsat, Sentinel-2) using

moderate-resolution LAI products (e.g., MODIS, VIIRS, PROBA-V), provides a high degree of consistency across spatial scales, which is essential for land surface and hydrological models that operate at multiple scales. For example, the Atmosphere-Land Exchange Inverse (ALEXI) model uses MODIS LAI for energy partitioning and evapotranspiration (ET) estimation at 4-km resolution (Anderson et al., 2007a, 2007b), while a disaggregation approach (DisALEXI) employs Landsat-based LAI to disseminate ALEXI ET field to higher spatial resolutions (Anderson et al., 2018; Anderson et al., 2011; Norman et al., 2003; Yang et al., 2018a). The consistency of LAI estimation at fine and coarse grids is thus critical to such applications. Recently, DisALEXI, among other major satellite-based ET models, has been implemented by the OpenET project (<https://etdata.org/>) for 30-m ET mapping over the contiguous US (CONUS). Thus, a MODIS-consistent LAI algorithm for Landsat that is easily scalable towards large areas is highly desirable.

The approach proposed by Gao et al. (Gao et al., 2012) for decametric sensors has so far only been applied and evaluated locally (Gao et al., 2014; Gao et al., 2012; Sun et al., 2017; Zhou et al., 2018). In these applications, homogeneous or unmixed training samples were generated within a single Landsat footprint using samples from one to more years. This local sampling scheme is subject to spatial and temporal discontinuity and low computational efficiency when applied for wall-to-wall LAI mapping over large areas. With the help of machine learning and cloud-computing techniques, Gao et al. (2012)'s approach can be potentially generalized to continental and global scales. Another issue with previous applications using MODIS is that only LAI values produced by the main RT algorithm without saturation were considered (Gao et al., 2012; Houborg et al., 2016; Zhou et al., 2018). The MODIS LAI algorithm flags retrievals as "saturation" when the surface reflectance falls in a predefined saturation domain (Knyazikhin et al., 1998). The saturation domains describe conditions for dense canopies, where the surface reflectance becomes insensitive to LAI (Gower et al., 1999; Kang et al., 2016). As most high LAI values (greater than $4 \text{ m}^2/\text{m}^2$) are derived in the saturation domain, eliminating saturated LAI from the training set would significantly underestimate LAI in dense canopies (Houborg et al., 2016). Since the LAI retrieved with saturation has higher uncertainty, more in-depth analysis is required to determine the optimal balance between saturated and unsaturated LAI training samples.

In this study, we aim to develop a generalized approach to estimate LAI based on Landsat 5, 7, 8 images over CONUS. Landsat provides spatial details essential for many agricultural, hydrological, and climate adaptation applications. A high degree of consistency with the MODIS LAI also supports downstream applications across spatial scales. To achieve these goals, we first generated 1.6 million spatially homogeneous samples from MODIS LAI and Landsat surface reflectance images between 2006 and 2018 over CONUS. These samples were stratified by biome and Landsat sensor and quality controlled to remove noise and maintain sample balance. The resulting 1.6 million high-quality samples were used to train random forest models for each biome, which can be applied to estimate LAI from Landsat images on Google Earth Engine. This approach improved Gao et al. (2012)'s method with spatial generalization over CONUS and advanced sampling design. The Landsat-based LAI estimates were inter-compared to MODIS LAI and evaluated using ground-measured LAI for major biomes in CONUS. The Landsat LAI algorithm and the training dataset are publicly available (<https://github.com/yanghuikang/Landsat-LAI>).

2. Data and methodology

2.1. Overview of the MODIS LAI algorithm

We selected the MODIS LAI product as the basis since it is based on physical RT modeling and has reached Validation Stage 2 – "product accuracy is assessed over a substantial set of locations and time periods", defined by the Committee on Earth Observation Satellites (CEOS)

Working Group on Calibration and Validation (Brown et al., 2020a; Fernandes et al., 2014a; K. Yan et al., 2016b). More importantly, MODIS LAI is widely used in state-of-the-art land surface models, and a consistent Landsat LAI approach will be beneficial for many related applications (Lawrence and Chase, 2007; Ryu et al., 2011).

The MODIS Version 6 LAI and Fraction of Photosynthetically Active Radiation (FPAR) product (MCD15A3H) is a 4-day composite product at 500-m spatial resolution (K. Yan et al., 2016a). The MODIS LAI algorithm includes a set of three-dimensional (3D) RT models with biome-specific soil and canopy properties (1D RT models are used in some biomes). The algorithm uses eight biomes to represent global vegetation patterns: grasses and cereal crops, shrublands, broadleaf croplands, savannas, evergreen broadleaf forests, deciduous broadleaf forests, evergreen needleleaf forests, and deciduous needleleaf forests. RT model inversion is enabled by biome-specific LUTs, which store simulated bi-directional reflectance factors (BRFs) and corresponding soil/canopy patterns which include LAI. When the main algorithm fails to find a solution, a backup algorithm estimates LAI based on biome-specific relationships between LAI and the Normalized Difference Vegetation Index (NDVI). The MODIS LAI product flags the estimation method: Main (RT) model with no saturation, Main (RT) model with saturation, back-up method due to bad geometry, and back-up method due to other problems. Any retrievals that fall in a biome-specific saturation domain are flagged as “saturation” (Knyazikhin et al., 1998). The saturation domains are defined by saturation curves between the dispersion (i.e., standard deviation) and the retrieved values of the LAI with specified uncertainties.

2.2. NLCD land cover and MODIS LAI biome matching

The National Land Cover Database (NLCD) products (L. Yang et al., 2018b) were used to retrieve 30-m biome types. The NLCD products

were produced at 30-m resolution and two- to three-year intervals between 1992 and 2016. The average overall accuracy ranges between 71% to 97% (L. Yang et al., 2018b). There are major differences between NLCD land cover classification schema and biomes used by the MODIS LAI product, as the latter primarily considers canopy structures. To evaluate the discrepancy, we created a random sample of 50,000 pixels from the 2011 MODIS Land Cover (MCD12Q1.006) map in CONUS and compared the land cover categories between MODIS and the 2011 NLCD map. The NLCD map was aggregated to MODIS grids and pixels where the dominating class cover less than 90% of the pixel area were masked out. Based on these samples, we mapped the relationship between NLCD and MODIS land cover types using a Sankey diagram (Fig. 1). Results show that the land cover types of NLCD and MODIS do not present a one-to-one mapping. The NLCD Mixed Forests type primarily belongs to MODIS's Deciduous Broadleaf Forest (DBF) biome. The NLCD Woody Wetlands class corresponds to MODIS DBF, Evergreen Broadleaf Forest (EBF), and Savannas. The NLCD Emergent Herbaceous Wetlands class contains MODIS Savannas and Grasslands. About 40% of the NLCD Shrubs are classified as Grasslands by MODIS. A large portion of NLCD Croplands, mainly cereal crops, is grouped into Grasslands in MODIS. Based on these mappings, we defined eight biomes from NLCD land cover types to maintain consistency with MODIS (Table 1). The only deviation of our definitions from the NLCD scheme is the combination of Pasture and Grassland into a single class.

2.3. LAI sample generation

2.3.1. Overview

A high-quality sample set was generated from MODIS LAI (500 m) and corresponding Landsat surface reflectance (aggregated from 30 m to 500 m) across CONUS from 2006 to 2018 using Google Earth Engine. This dataset consisted of spatially homogeneous MODIS pixels randomly

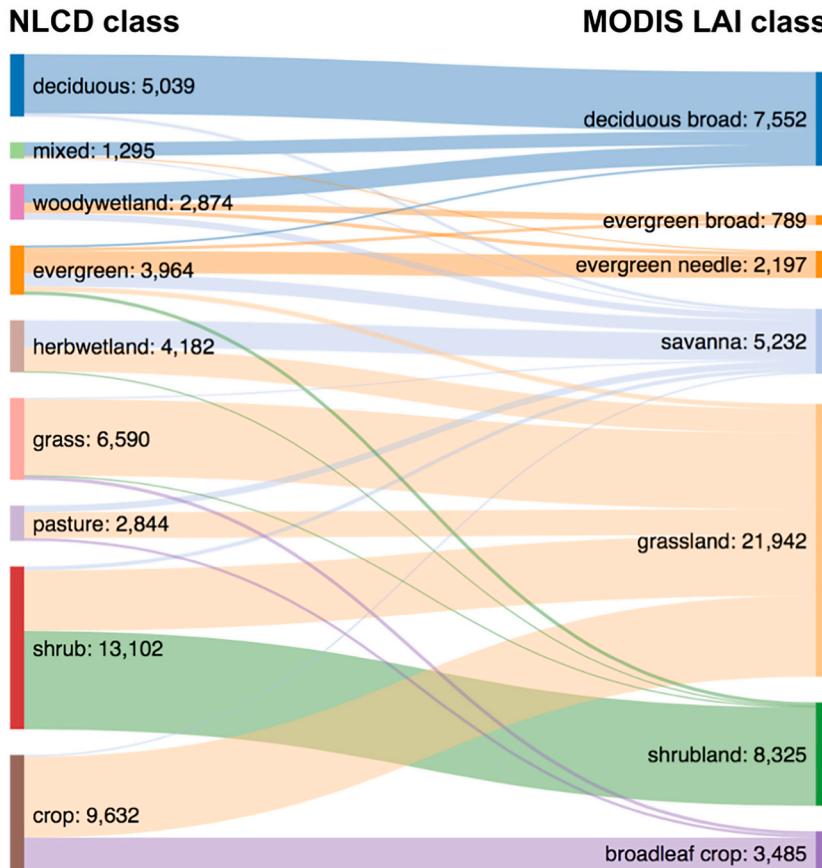


Fig. 1. Sankey diagram showing the mapping between land cover types from NLCD (left) and MODIS LAI Biome classifications (right) over CONUS. The underlying data contains 50,000 random pixels from MODIS 2011 land cover map over CONUS. The majority NLCD class was extracted for these pixels. Numbers indicate the sample count in the class. Widths of the flows are linearly proportional to sample counts. Abbreviates for NLCD land cover types: “deciduous” = Deciduous Forest (NLCD 41); “mixed” = Mixed Forest (NLCD 43); “evergreen” = Evergreen Forest (NLCD 42); “woodywetland” = Woody Wetlands (NLCD 90); “herbwetland” = Emergent Herbaceous Wetlands (NLCD 95); “grass” = Grassland/Herbaceous (NLCD 71); “pasture” = Pasture/Hay (NLCD 81); “shrub” = Shrub/Scrub (NLCD 52); “crop” = Cultivated Crop (NLCD 82). Abbreviates for MODIS biomes: “deciduous broad” = Deciduous Broadleaf Forest (MODIS 6); “evergreen broad” = Evergreen Broadleaf Forest (MODIS 5); “evergreen needle” = Evergreen Needleleaf Forest (MODIS 7) = “savanna”: Savannas (MODIS 4); “grassland” = Grasslands (MODIS 1); “shrubland” = Shrublands (MODIS 2); “broadleaf crop” = Broadleaf Croplands (MODIS 3).

Table 1

Biome classification scheme based on NLCD for Landsat LAI estimation.

Biome	NLCD Code	MODIS Biome
Deciduous Forest	41	Deciduous Broadleaf Forest (DBF)
Evergreen Forest	42	Evergreen Needleleaf Forest (ENF), Savannas
Mixed Forest	43	DBF, ENF
Shrub	52	Grasslands, Shrublands
Grassland	71, 81	Grasslands
Cropland	82	Grasslands, Broadleaf Croplands
Woody wetland	90	DBF, Evergreen Broadleaf Forest (EBF), ENF, Savannas
Herbaceous wetland	95	Savannas, Grasslands

selected across CONUS and stratified by Landsat sensor, biome, and saturation status. Fig. 2 shows the workflow of the sample generation and LAI mapping processes. The sample generation involves five steps: 1) generation of spatially representative MODIS sample locations based on land cover homogeneity, hereafter referred to as “spatial samples”; 2) generation of spatially pure MODIS LAI samples with corresponding Landsat surface reflectance from spatial samples; 3) sample cleaning and outlier removal; 4) sample balancing; and 5) machine-learning-based screening (hereafter referred to as saturation screening) (Fig. 2). The spatial distribution of final training samples is shown in Fig. A1.

2.3.2. Spatial sample generation

Spatial samples were generated from the MODIS land cover map and the NLCD land cover map for 2011 (Fig. 2). We randomly selected approximately 100,000 MODIS pixels (500-m) within which more than

90% of the 30-m NLCD pixels belong to a single biome. The MODIS land cover type should match that of the majority NLCD biome based on Table 1. Since the NLCD 2016 product (L. Yang et al., 2018b) was not yet available by the time we generated these samples, we retained pixels in which the majority NLCD land cover did not change across the NLCD 2001, 2006, and 2011 maps. Note that the sample size (100,000), which is significantly larger than that used in previous data-fusion-based global LAI products, was chosen to provide spatial representativeness over CONUS.

2.3.3. Pure LAI sample generation

For each spatial sample, we extracted high-quality MODIS LAI retrievals (those from the main algorithm) and spatially aggregated Landsat surface reflectance values from imagery collected over 2006 to 2018 (Fig. 2). Landsat images were masked to exclude clouds, cloud shadows, and water and reprojected to match MODIS’s projection. These LAI-reflectance samples were further screened for spatial homogeneity. We followed the procedure described in Gao et al. (2012) to determine the spatial purity of MODIS pixels. Within each MODIS LAI pixel, we aggregated Landsat surface reflectance of six bands (blue, green, red, near-infrared (NIR), short-wave-infrared (SWIR1), SWIR2) by mean (μ) and standard deviation (σ). The coefficient of variation (CV) was calculated as the ratio of μ and σ , and pixels were considered to be homogeneous if their average CV across six bands was smaller/lower than 0.15. Note that we used MODIS LAI generated by the main algorithm both with or without saturation. Including saturated LAI values ensures that our samples cover high LAI values. Each pure LAI sample contains a MODIS LAI value and the mean Landsat surface reflectance of each band. NDVI and Normalized Difference Water Index (NDWI) were also computed for each sample.

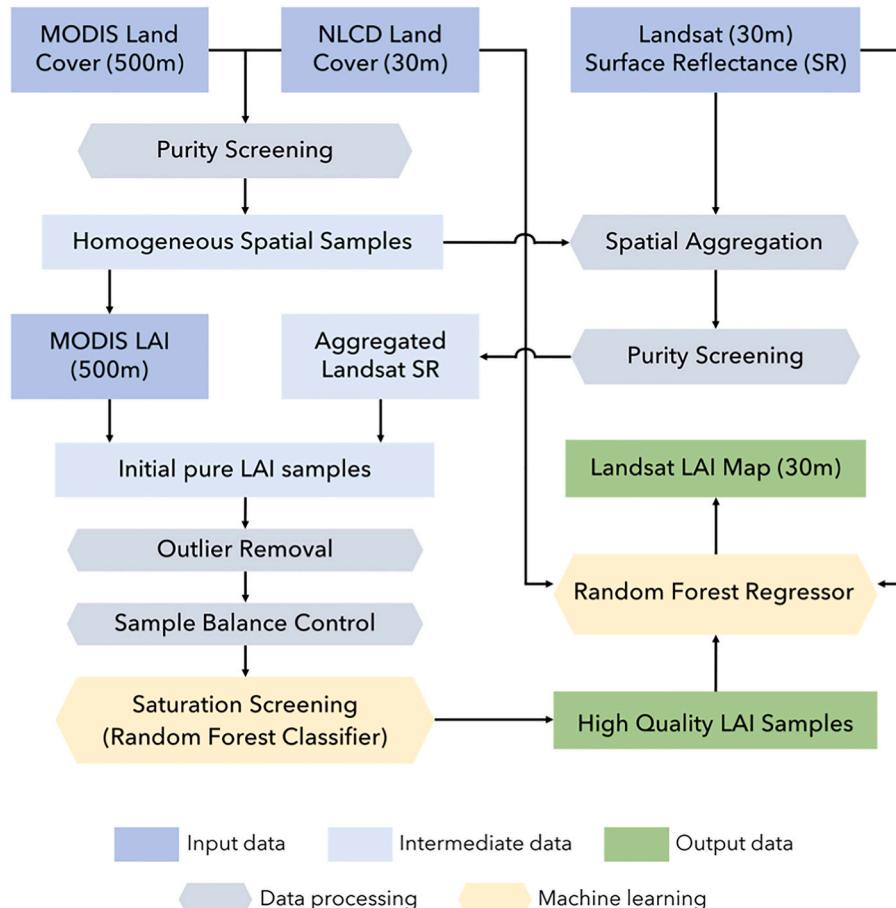


Fig. 2. Workflow of the LAI sample generation and Landsat LAI mapping processes.

2.3.4. Outlier removal

The homogeneous samples were screened to exclude invalid NDVI values (i.e., less than 0 or greater than 1). Then, we used the relationship between NDVI and LAI to remove outliers that fall outside of the normal range of NDVI for different LAI values (Kang et al., 2016). LAI values were binned into 0.2 m²/m² intervals. Within each bin, samples were removed if their NDVI values fell outside of the 1.5 interquartile range (IQR) from the first and third quartile. IQR is the difference between the first and third quartile. Here we used NDVI rather than individual bands (e.g. red and NIR) because of its strong relationship with LAI.

2.3.5. Sample balancing

We adopted a two-step approach to control sample balance among different biomes and among different LAI values (Fig. 2). First, we down-sampled LAI samples to reduce redundancy in overly populated biomes or LAI value ranges. The samples' initial distribution tended to skew towards more spatially homogeneous biomes (e.g., shrub) and low LAI values. For biomes with highly skewed LAI distribution, we computed the sample density in the LAI-NIR-Red space and down-sampled densely populated areas weighted by reversed squared density. As a result, sample distributions became more uniform.

The second step used learning curves to determine the optimal size of saturated and unsaturated LAI samples for high overall accuracy. A learning curve describes the change of model accuracy for varying training sample sizes. We constructed learning curves by changing the size of saturated or unsaturated samples using a 10-fold cross-validation scheme. In the 10-fold cross-validation, the entire sample set was randomly divided into 10 equal-sized splits. In each round, nine splits were used as training samples (i.e. training set), while the other one was for testing (i.e. testing set). Saturated and unsaturated samples were randomly drawn from the training set and used to train models. Models were then applied to the test set and the Mean Absolute Error (MAE) was computed. We controlled the saturated or unsaturated sample size to range from 100 to 20,000. This process was repeated 10 times so that each split served as the testing set once. The average MAE for each training sample size across 10 splits was recorded. Results showed that the optimal overall score was achieved with an equal number of saturated and unsaturated samples (see Section 3.1.2). Therefore, for each Landsat sensor and each biome, we subsampled 40,000 unsaturated and 40,000 saturated samples using the down-sampling process described in the first step. Note that saturated (high) LAI values are rare for shrub-lands, thus we did not include any saturated values for this biome.

2.3.6. Machine-learning-based sample screening

The MODIS LAI algorithm determines saturation based on the relationship between the LAI retrieval and its dispersion (i.e. standard deviation) from all candidate values. Saturated LAI values are typically above four with high uncertainties, while unsaturated LAI is below four with narrower dispersion. We found that some saturated and unsaturated MODIS LAI samples share similar spectral signatures from Landsat but contrasting LAI values, which undermines the skill of machine learning models (see details in Section 3.1.1). This is mainly because of differences between MODIS and Landsat sensing/ processing systems, such as projected instantaneous field of view (PIFOV), radiometric configuration, viewing geometries, atmospheric correction, geolocation errors, temporal compositing, etc. We designed a screening procedure (hereafter referred to as saturation screening) to identify and remove confounding unsaturated and saturated samples that share similar spectral signatures from the training dataset (Fig. 2).

The saturation screening process used the random forest classifier to identify saturated and unsaturated samples based on surface reflectance (green, red, NIR, SWIR), VIs (NDVI, NDWI), and solar illumination angles (Fig. 2). We built separate models for each biome and each Landsat sensor. Based on 10-fold cross-validation, the misclassified samples were removed from the dataset. The effect of saturation screening was analyzed in terms of cross-validation accuracy. Based on initial

assessments, the random forest classifier requires a balanced sample set to achieve optimal accuracy. Therefore, we performed the saturation screening after balancing the samples.

2.4. Machine learning algorithm

Given differences in radiative transfer mechanisms between biomes and discrepancies in spectral characteristics between Landsat sensors, we built random forest models for each biome (Table 1) and sensor (TM, ETM+, OLI) combination on Google Earth Engine (Fig. 2). Building separate models allows us to explicitly use the land cover information and incorporate a much larger training sample set (~1.6 million), as Google Earth Engine has a memory limit that constrains model complexity. Model inputs include surface reflectance in green, red, NIR, and SWIR1 bands, NDVI, NDWI (Gao, 1996), geographic coordinates (longitude, latitude) at pixel center, as well as solar zenith and azimuth angles at the scene center. Vegetation indices were incorporated to improve model performance due to their strong relationships with LAI. Spatial coordinates were used to incorporate spatial autocorrelation (You et al., 2017). The solar illumination geometry acknowledges the Bidirectional Reflectance Distribution Function (BRDF) effect on LAI retrieval. The sensor view geometry and per-pixel solar illumination angles are not available in the Landsat surface reflectance collection on Google Earth Engine, and thus were not included here.

In the exploratory analysis, we evaluated the effect of hyperparameters (number of trees, minimum leaf population, and number of variables per split) on model accuracy using 5-fold cross-validation. We found only marginal improvements in model accuracy as the model complexity increases. Interestingly, complex models with smaller leaves, more variables per split, and more trees could cause "speckles" or "salt and pepper" patterns on the LAI map, likely due to over-fitting. As a result, we used a hundred regression trees, set the minimum leaf population at 50 and number of variables per split at five, as a compromise between accuracy and spatial continuity. We evaluated the model performance by 10-fold cross-validation and used Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), normalized RMSE (nRMSE) and bias as the error metrics. Normalized RMSE was computed as the ratio between RMSE and the mean reference MODIS LAI. Bias was the mean difference between prediction and reference LAI. The final sample set was stored as a Google Earth Engine asset. To generate an LAI map for a Landsat image, the algorithm draws corresponding samples according to the NLCD land cover map, trains biome and sensor-specific random forest regressors, and applies the model to the Landsat image to estimate LAI (Fig. 2). We flag LAI estimates whose input surface reflectance data fall outside of training samples' convex hull, which is subject to extrapolation. We followed the approach used by the Sentinel-2 Level 2 LAI algorithm to create the convex hull (Weiss and Baret, 2016).

2.5. Comparison with MODIS LAI

To assess the consistency between MODIS and aggregated Landsat LAI and evaluate the impact of MODIS scaling errors, we compared LAI estimation based on Landsat against MODIS LAI using an independent sample set with different levels of spatial heterogeneity. This set was created by first sampling 5000 pixels from the MODIS 2011 land cover map for each NLCD biome and then populating these samples with MODIS LAI and Landsat LAI estimation from 2008 to 2017. In total, there were around 610,000 LAI samples. Within each sample, Landsat LAI estimates at 30-m resolution were aggregated by the mean to 500-m MODIS grids. The spatial distribution of these samples is shown in Fig. A2.

Since the MODIS LAI algorithm assumes that pixels are "pure" consisting of a single land cover type, the MODIS LAI is subject to scaling errors due to spatial heterogeneity within its 500 m pixels (Chen, 1999; Wu and Li, 2009; K. Yan et al., 2016a). This scale effect consequently

impacts the consistency between MODIS and aggregated Landsat LAI in heterogeneous areas. Therefore, we analyzed the difference between MODIS and aggregated Landsat LAI regarding different levels of spatial heterogeneity. First, we used information entropy to quantify spatial heterogeneity caused by mixed biomes. Information entropy was computed based on the percentage of area for each NLCD biome within a MODIS pixel following the equation below:

$$H = - \sum_{i=1}^n P(x_i) \log P(x_i) \quad \text{Eq. (1)}$$

$P(x_i)$ denotes the area proportion of biome i and H is the information entropy; n is the total number of 30-m NLCD pixels within a MODIS 500-m pixel. A uniform pixel with a single NLCD class will have an information entropy value of zero. Higher entropy values indicate higher degrees of spatial heterogeneity. Second, for pixels where 90% of the area was covered by a single biome (pure-biome pixels), we used the average CV of Landsat surface reflectance to indicate spatial heterogeneity caused by varied vegetation conditions. For these samples, we also computed the average Landsat surface reflectance within MODIS pixels and applied biome-specific random forest models to produce another set of Landsat LAI estimation (LAI aggregated on SR). This analysis was expected to further reduce the scale effect by aggregating the input data rather than the output. The difference between MODIS and Landsat LAI was quantified by MAE, RMSE, bias, and coefficient of determination (r^2).

2.6. Direct validation using *in-situ* measurements

2.6.1. In-situ data

We validated the Landsat LAI algorithm using two sets of ground measurements. The first set includes measurements from 19 National Ecological Observatory Network (NEON) sites (Table 2). In these sites digital hemispherical photography (DHP) data were regularly collected using the sampling protocol defined by NEON (Meier et al., 2018). The raw DHP data was processed and released by the Copernicus Ground Based Observations for Validation (GBOV) (L. Brown et al., 2020a). The second set consists of ground measurements obtained from eight research stations or Ameriflux sites (Table 3). Each study site adopted a different measuring and sampling approach. The two datasets together covered forest, grass, shrub, and crop biomes.

Table 2

Summary of NEON sites used for ground validation (processed by the GBOV project).

Site	Code	Location	Latitude, Longitude	Biome	Date	No. of samples
Bartlett Experimental Forest	BART	New Hampshire, US	44.03639, -71.2873	Deciduous Forest, Mixed Forest	2014–2018	93
Blandy Experimental Farm	BLAN	Virginia, US	39.0603,-78.0716	Deciduous Forest	2016–2018	31
Central Plains Experimental Range	CPER	Colorado, US	40.8155,-104.7460	Grassland	2014–2018	221
Disney Wilderness Preserve	DSNY	Florida, US	28.1250,-81.4362	Grassland	2014–2018	91
Harvard Forest	HARV	Massachusetts, US	42.5369,-72.1727	Evergreen Forest, Mixed Forest	2014–2018	132
Jones Ecological Research Center	JERC	Georgia, US	31.1948,-84.4686	Evergreen Forest	2014–2018	150
Jornada	JORN	New Mexico, US	32.5907,-106.843	Shrub	2015–2018	168
Moab	MOAB	Utah, US	38.2483,-109.388	Shrub	2015–2018	153
Niwot Ridge Mountain Research Station	NIWO	Colorado, US	40.0543,-105.582	Grassland	2015–2018	53
Onaqui	ONAQ	Utah, US	40.1776,-112.452	Shrub	2014–2018	174
Oak Ridge	ORNL	Tennessee, US	35.9641,-84.2826	Deciduous Forest	2015–2018	84
Ordway-Swisher Biological Station	OSBS	Florida, US	29.6893,-81.9934	Evergreen Forest	2013–2018	91
Smithsonian Conservation Biology Institute	SCBI	Virginia, US	38.8929,-78.1395	Deciduous Forest	2015–2018	39
Smithsonian Environmental Research Center	SERC	Maryland, US	38.8901,-76.56	Deciduous Forest	2015–2018	41
Steigerwaldt Land Services	STEI	Wisconsin, US	45.5089,-89.5864	Deciduous Forest	2016–2018	61
North Sterling	STER	Colorado, US	40.4619,-103.029	Cropland	2014–2018	126
Talladega National Forest	TALL	Alabama, US	32.9505,-87.3933	Evergreen Forest	2015–2018	48
UNDERC	UNDE	Michigan, US	46.2339,-89.5373	Deciduous Forest, Mixed Forest, Woody Wetland	2015–2018	53
Woodworth	WOOD	North Dakota, US	47.1282,-99.2414	Grassland	2014–2018	159

2.6.1.1. NEON-GBOV. We used ground measurements from 19 NEON sites (Table 2). In each site, DHP images were collected at multiple plots every two weeks during the growing season. The size of these plots ranges between 20 m × 20 m and 40 m × 40 m. Within each plot, photos were taken at 12 points located at cardinal directions. The raw photos were processed by the GBOV project and converted to Plant Area Index (PAI), as a proxy of LAI. PAI accounts for both leaves and woody elements, and thus is greater than LAI. GBOV reports PAI and effective PAI retrieved using two methods: Miller's (1967) and Warren-Wilson's (1963). Foliage clumping is an important factor that causes indirect optical instruments to substantially underestimate LAI and/or PAI (Yan et al., 2019). In this dataset, clumping effect was corrected using the approach proposed by (Lang and Yueqin, 1986) to estimate PAI from effective PAI. We screened the DHP data using quality flags provided by GBOV and used PAI values combining both upward-facing and downward-facing photos from Warren-Wilson's approach (L. Brown et al., 2020a). Detailed information about the ground measurement can be found from NEON (Meier et al., 2018). Detailed information about GBOV processing and spatial distribution of the NEON-GBOV sites can be found from Brown et al. (2020) and the GBOV website (<https://gbov.acri.fr>).

2.6.1.2. Other sites. We used ground measurements of LAI/PAI at three flux towers: EMS (Ameriflux Site US-Ha1), HEM (Ameriflux Site US-Ha2), and LPH (Ameriflux Site US-LPH) from the Harvard Forest ecological research site (Table 3). The EMS tower is located in a deciduous broadleaf forest. PAI was measured by LAI-2000 (Li-cor, Lincoln, NE, USA) in 20-m diameter plots along ten transects from the tower. Within each plot, five subsamples were taken and averaged. The HEM tower is in an old-growth hemlock forest. PAI was measured in 12 plots along transects from the tower. Each plot has a diameter of six meters. The LPH tower is in a deciduous broadleaf forest. PAI was measured in 36 plots surrounding the tower, each with a diameter of six meters. The LAI data for these sites were downloaded from the Harvard Forest Data Archive (Munger and Wofsy, 2018; Orwig, 2019).

We obtained ground LAI measurements from five Ameriflux sites: US-Shd, US-Var, US-Ne1, US-Ne2, and US-Ne3 (Table 3). US-Shd (Verma, 2001) and US-Var (Ma et al., 2020; Ma et al., 2007) are located in grassland. In US-Shd, destructive samples were collected in nine 33.3 cm × 33.3 cm square plots randomly selected in the tower footprint every two weeks during the growing season from 1997 to

Table 3

Summary of eight independent study sites used for ground validation.

Source	Site	Location	Latitude, Longitude	Biome	Date	Measurement Method	LAI definition	Clumping Correction	No. of samples	Note
Harvard Forest	HARV-EMS	Massachusetts, US	42.5377,-72.1714	Deciduous Forest	2006–2018	LAI-2000	Project Area Index (PAI)	Partial ^a	140	
	HARV-HEM	Massachusetts, US	42.5394,-72.1778	Evergreen Forest	2012–2016	LAI-2000	PAI	Partial	11	
	HARV-LPH	Massachusetts, US	42.5419,-72.1849	Deciduous Forest	2008–2010	LAI-2000	PAI	Partial	32	
	US-Shd	Oklahoma, US	36.9600,-96.6831	Grassland	1997–2000	Destructive	LAI	Yes	41	
Ameriflux	US-Var	California, US	36.9333,-96.6833	Grassland	2002–2013	Destructive	LAI	Yes	70	
	US-Ne1	Nebraska, US	41.1650,-96.4766	Cropland	2001–2012	Destructive	LAI	Yes	68	Continuous irrigated maize field
	US-Ne2	Nebraska, US	41.1649,-96.4701	Cropland	2001–2012	Destructive	LAI	Yes	76	Irrigated maize and soybean rotation field
	US-Ne3	Nebraska, US	41.1797,-96.4396	Cropland	2001–2012	Destructive	LAI	Yes	67	Rainfed maize and soybean rotation field
SMEX02-IA	SMEX02-IA	Iowa, US	42.0491,-93.6072	Cropland	2002	LAI-2000	PAI	Partial	257	Rainfed maize and soybean fields
GRAPEX	Tower 1	California, US	38.2893,-121.1182	Cropland	2013–2018	LAI-2200	PAI	Partial	10	Vineyard
	Tower 2	California, US	38.2805,-121.1181	Cropland	2013–2018	LAI-2200	PAI	Partial	9	Vineyard

^a LAI-2000 and LAI-2200 partially account for clumping effects at the scale larger than the sensor field-of-view. An independent procedure is often required to compute clumping index in order to fully correct clumping effects for these instruments (Fang et al., 2019).

2001. LAI measured in the nine plots were averaged. In US-Var, destructive samples were collected in three 20 cm × 20 cm square plots within the tower footprint and the averaged LAI was reported. US-Ne1, US-Ne2, US-Ne3 are collectively referred to as the Mead site (Suyker, 2001a, 2001b, 2001c). US-Ne1 is an irrigated maize field, US-Ne2 is an irrigated maize and soybean rotation field, and US-Ne3 is a rainfed maize and soybean rotation field. LAI was measured by destructive sampling within the fields at weekly time intervals between 2001 and 2012.

We also obtained ground measurements from the SMEX02 (Soil Moisture Experiment of 2002) experiment. During the experiment, PAI was measured using LAI-2000 for 21 corn and 10 soybean fields within the Walnut Creek Watershed in Iowa between June 15 and July 8, 2002. Within each field, three sampling locations were selected representing high, low, and average vegetation cover respectively. Each sampling plot was 12-m long and 10-row across. Row spacing was 0.76 for corn and 0.25–0.38 for soybean fields. Detailed sampling and measurement methods can be found in Anderson et al. (2004).

The last set of measurements came from the Grape Remote sensing Atmospheric Profiling and Evapotranspiration eXperiment (GRAPEX) project (Kustas et al., 2018). PAI was measured using LAI2200C in two vineyards. In each field, data were collected in 25 grids which include five cross-row transects each spanning five rows. The average row spacing was 3.35 m, and the spacing between each transect was about 10 m. The 25 PAI measurements were averaged for each field. We used the PAI that included both vines and inter-row cover crops whenever available. PAI measurements in GRAPEX represent structured crop types. Detailed descriptions of the measurement have been provided in Sun et al. (2017) and White et al. (2018).

Note that instruments like LAI2000 and LAI2200 only partially consider foliage clumping at scales larger than the field-of-view of the instrument sensor leading to underestimation of PAI (Chen, 1996; Ryu et al., 2010) (Table 3). For forests and fruit trees, the underestimation might be partly offset by the inclusion of woody elements (Fang et al., 2012b; G. Yan et al., 2016), leaving the estimated PAI closer to true LAI. Generally, field LAI measurements are subject to uncertainties due to

LAI definition, measurement methods, inclusion of woody elements, and clumping effects (Fang et al., 2019). Thus, measurement protocols and potential sources of uncertainty need to be considered in each site for the interpretation of validation results.

2.6.2. Spatiotemporal matchup and error metrics

At Harvard forest sites, since sampling plots were too small compared to Landsat pixel scale, we aggregated the plot measurements of PAI within 200 m from towers and compare them to the mean Landsat values aggregated within a 13-pixel by 13-pixel window around each tower. For other sites, we extracted the average Landsat LAI in a window of three-by-three pixels centered at the measurement location to minimize registration errors. Each location was manually checked using high-resolution satellite images on Google Earth for the spatial homogeneity. Individual locations that contain a large proportion of non-vegetative areas were removed. For Ameriflux sites, the three-by-three Landsat pixel windows were centered at the tower location, assuming that ground measurements were representative of tower footprints. At GRAPEX sites, Landsat extraction windows were centered at the centroid of 25 measuring grids. Since these fields are relatively homogeneous, validation results do not show substantial differences when varying the window size from one pixel to five pixels. For SMEX-02, we extracted Landsat LAI for each sampling plot within the fields.

Ground measurements are matched to estimated LAI from Landsat image acquired at the day of measurement when possible. When there was no coincident Landsat overpass around the measurement time, we performed simple linear interpolation using the closest Landsat LAI observations before and after the measurement. If the gap between two Landsat observations was greater than 20 days, we discarded the sample.

Following CEOS recommendations (Fernandes et al., 2014a), we reported the total uncertainty, bias, and precision for validation results. The total uncertainty includes systematic and random errors. Bias is the expected difference between estimation and reference representing the systematic error. Precision is the dispersion of estimation from the expected value representing the random error. The total uncertainty was

measured by RMSE and Mean Squared Error (MSE) (Eq.(2)). MSE can be decomposed into squared bias and variance of the difference between estimation and reference (Eq.(2,3,4)). Here variance can be used to approximate precision.

$$MSE = E[(\hat{\theta} - \theta)^2] = bias^2 + Var \quad \text{Eq.(2)}$$

$$bias = E(\hat{\theta} - \theta) \quad \text{Eq.(3)}$$

$$Var(\hat{\theta} - \theta) = E\{[(\hat{\theta} - \theta) - E(\hat{\theta} - \theta)]^2\} \quad \text{Eq.(4)}$$

Moreover, since errors in satellite-derived LAI often depend on the LAI value, we fitted Thiel-Sen regression relationships between estimated and reference LAI and used the residuals after fitting as an approximation of precision following CEOS's recommendations

(Fernandes et al., 2014a). Additionally, we calculated the Uncertainty Agreement Ratio (UAR) as the percentage of LAI estimates falling within the maximum of 1 unit or 20% of reference LAI (L. Brown et al., 2020a).

3. Results

3.1. Sample generation

3.1.1. Comparison of saturated and unsaturated samples

The LAI distribution, dispersion, and spectral signature of saturated and unsaturated samples shows substantial distinctions. Fig. 3a illustrates the saturation domain for deciduous forest, cropland, and wetland samples. All saturated MODIS LAI samples fall along a curve defined by the dispersion of candidate LAI values against the retrieved LAI from the MODIS algorithm. The spread around the curve indicates the

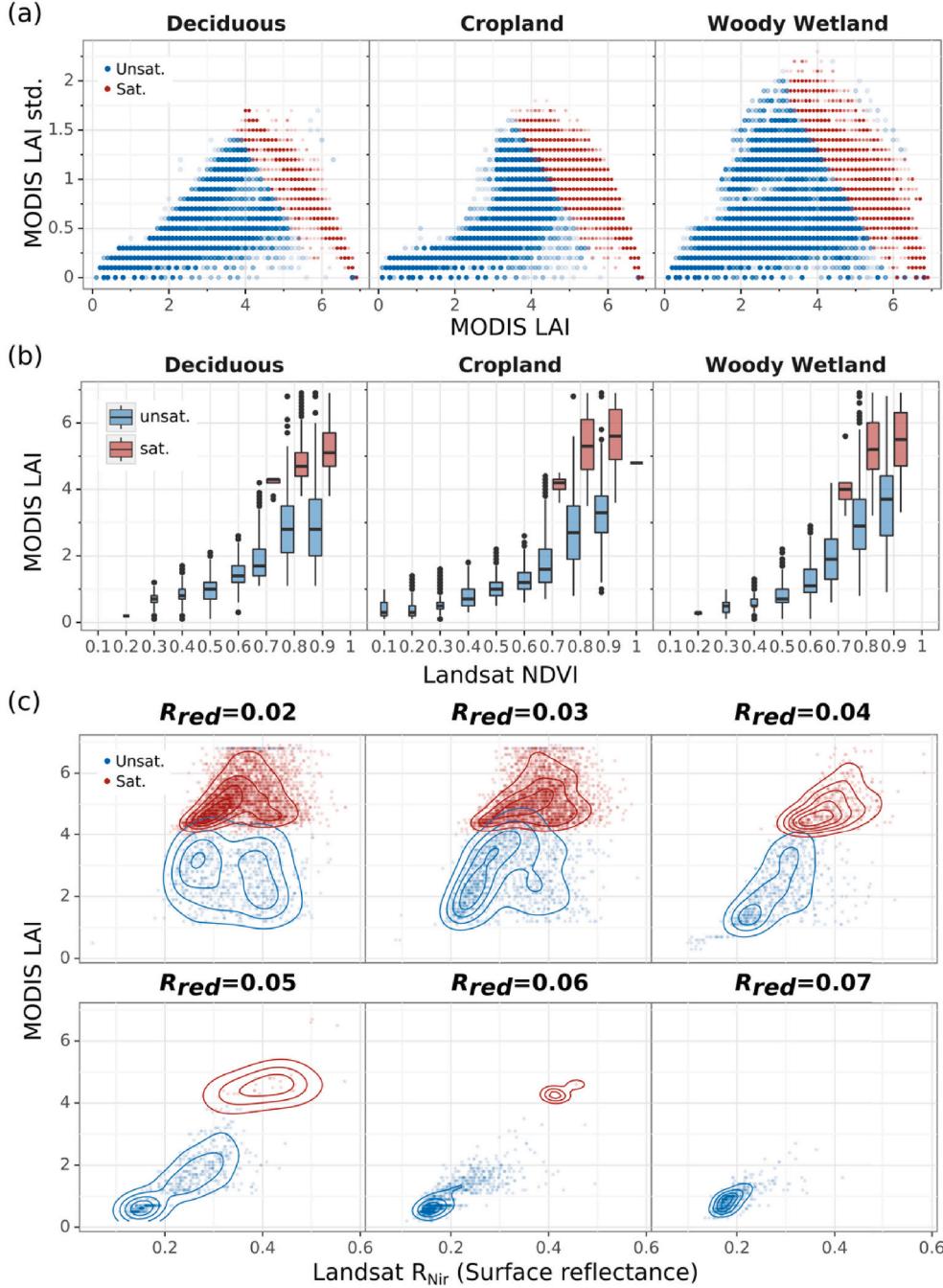


Fig. 3. Spectral signatures of saturated and unsaturated LAI samples for selected biomes. a) The saturation domain defined by the MODIS LAI retrieval and its dispersion. Points for unsaturated samples (blue) are slightly large than those for unsaturated samples for better visualization of overlaps. b) Comparison of saturated and unsaturated LAI values against Landsat NDVI. NDVI is binned by a width of 0.1. c) Relationship between LAI and NIR reflectance (Landsat) within different red reflectance ranges. Results are shown only for deciduous forest samples. Panel titles indicate the reflectance value in red band. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

uncertainty. Saturated and unsaturated samples are not distinguishable by LAI alone. There exists a substantial overlap between saturated and unsaturated LAI samples from 4 to 6 m^2/m^2 .

Fig. 3b shows the relationship between MODIS LAI and Landsat NDVI for saturated and unsaturated samples. All saturated samples possess high NDVI (above 0.65) and high LAI values (above 3.5), while unsaturated samples occupy a full NDVI range with low to medium LAI (below 3). Saturated and unsaturated LAI show different relationships with NDVI. **Fig. 3c** further illustrates the spectral overlap between saturated and unsaturated samples. With the same red and NIR surface reflectance, saturated and unsaturated samples show distinctive LAI values. The relationships between LAI and NIR are more diverse for low red reflectance. Note that only red and NIR bands are used in the MODIS algorithm to estimate LAI.

3.1.2. Effects of the balance between saturated and unsaturated samples

The relative proportion of saturated and unsaturated training samples has a substantial impact on the overall estimation accuracy and that of saturated and unsaturated LAI (**Fig. 4**). In all biomes, the estimation error of saturated samples decreases as more such samples are incorporated in the training set (**Fig. 4a**). The errors sharply reduce as the saturated sample size increases from 100 to 1000 and gradually stabilizes after the sample size reaches 10,000 to 15,000. Meanwhile, adding saturated samples to the training set adversely affects the accuracy of unsaturated samples due to the spectral overlap. In the overlapped area (e.g., $\text{NDVI} > 0.7$ as shown in **Fig. 3b**), the presence of saturated training samples (high LAI) increases the estimated LAI and results in an overestimation for unsaturated samples. Nevertheless, the negative effect on the accuracy of unsaturated samples is not as substantial as the improvement for saturated samples for most biomes. Therefore, the overall estimation error continues to decrease while the size of unsaturated samples increases.

The analysis of the unsaturated sample size shows similar results (**Fig. 4b**). Increasing the proportion of unsaturated sample improves the estimation accuracy of unsaturated LAI considerably while reducing the accuracy of saturated LAI. The overall error drops as unsaturated samples grow in number and stabilizes after 10,000 to 15,000 samples. The unsaturated error is initially higher than the saturated error when a minimum number of unsaturated samples are involved in training, but as more such samples are incorporated, the unsaturated error would eventually drop below the saturated error.

The optimal size of saturated and unsaturated samples for the overall estimation is found to be around 1:1 (**Fig. 4c**). Under a fixed total sample size, when the ratio between saturated and unsaturated sample size changes from 1/10 to 10, saturated and unsaturated samples experience a monotonic decrease and increase of estimation error respectively (**Fig. 4c**). The overall error generally presents a bell-shaped curve with the lowest value found when the sample is balanced between saturated and unsaturated LAI.

3.1.3. Saturation screening

To reduce the confounding effect of spectral overlap, we used a random forest classifier to identify saturated and unsaturated samples as labeled by the MODIS algorithm. Samples not correctly classified were removed from the training dataset as they were considered to be spectrally indistinguishable. The overall cross-validation accuracy ranges from 86% to 94%, depending on the biome type (**Table 4**). 10 to 20% of unsaturated samples and four to 10% of saturated samples are mislabeled (**Fig. 5**).

The potential effect of misclassified samples is illustrated by the relationship between LAI and NDVI (**Fig. 6**). The misclassified unsaturated samples (False Positive) share similar spectral signals as the saturated samples (True Positive) but with lower LAI. For example, for deciduous forests, the NDVI of correctly classified saturated samples is above 0.5, and the LAI ranges between 4 and 6 m^2/m^2 . Misclassified unsaturated samples have a similar NDVI range, but the LAI is between 1

and 4 m^2/m^2 . Therefore, when a new example comes with similar spectral signatures, the model is likely to consider both saturated samples and unsaturated samples, and underestimate LAI. The misclassified saturated samples (False Negative) have slightly lower LAI than most saturated samples, and their NDVI is in between that of unsaturated and saturated samples. The presence of misclassified samples could lead to an overestimation of unsaturated LAI and underestimation of saturated LAI.

3.1.4. Final sample statistics

The final training sample set includes 1.6 million LAI samples with 60,000 to 75,000 samples per biome and Landsat sensor. Most biomes have a bimodal LAI distribution corresponding to unsaturated and saturated samples (**Fig. 7**). We initially generated over 3.6 million samples. Then, 40,000 outliers were removed, the sample balancing process removed about 1.4 million redundant samples, and the saturation screening process discarded another 20,000. The final sample size was determined according to the sample balance, sample size analyses, and the resource limit of Google Earth Engine.

3.2. Cross-validation of machine learning models

The performance of machine learning models was evaluated using 10-fold cross-validation. The MAE based on clean samples ranges from 0.1 to 0.5 m^2/m^2 , RMSE ranges between 0.21 and 0.69 m^2/m^2 , and nRMSE between 0.13 and 0.25, depending on the biome type (**Table A1**, **Fig. 8a**). Shrubs have the lowest error ($0.1 \text{ m}^2/\text{m}^2$) among all biomes, as most of the LAI is below 2 m^2/m^2 . The MAE for forests, grasslands, and emergent herbaceous wetlands are between 0.3 and 0.4 m^2/m^2 , and the MAE for cropland and woody wetlands are around 0.5 m^2/m^2 .

The saturation screening process reduces the cross-validation error substantially, especially in terms of bias (**Fig. 8b**). For unsaturated samples, the reduction in MAE is around 0.2 m^2/m^2 and up to 0.4 m^2/m^2 (mixed forest). For saturated samples, the MAE drops by about 0.07 m^2/m^2 . The improvement in accuracy is mainly attributed to the reduced bias (**Fig. 8b**). Prior to the saturation screening, misclassified samples had larger errors and biases than the correctly classified samples (**Fig. 8c**). After eliminating these samples, the bias is substantially reduced towards zero. The correctly classified samples, i.e., those remain in the clean training set, also see a large reduction in bias and MAE after screening (**Fig. 8c**).

After removing the misclassified samples, the scatter between predicted and observed (MODIS) LAI follows the one-to-one line in the unsaturated domain. However, the model cannot capture the full variability of saturated LAI, leading to an overestimation for LAI ranging from 4 m^2/m^2 to 5 m^2/m^2 and an underestimation above 5 m^2/m^2 (**Fig. 8d**). This is primarily because surface reflectance has a low sensitivity to the changes of LAI in the saturation domain. For cropland and grassland, saturated MODIS LAI samples have a range of 4 m^2/m^2 to 7 m^2/m^2 , while the predicted LAI only ranges from 5 m^2/m^2 to 6 m^2/m^2 . For evergreen forest and mixed forest, the sensitivity is higher, resulting in a slightly smaller bias.

3.3. Comparison to MODIS LAI

3.3.1. General comparison results

When Landsat-based LAI estimates are aggregated and compared to MODIS LAI using an independent sample set, overall MAE ranges between 0.11 m^2/m^2 and 0.82 m^2/m^2 across eight biomes (**Fig. 9a**). The unsaturated MAE is between 0.10 m^2/m^2 and 0.84 m^2/m^2 , while the saturated MAE is higher (0.63 m^2/m^2 –1.15 m^2/m^2) (**Fig. 9a**). Generally, there is an overestimation of unsaturated LAI and underestimation of saturated LAI (**Fig. 9b**, **Fig. A3**). Removing misclassified samples increased the positive bias of unsaturated LAI slightly, but more considerably reduced the negative bias in saturated samples, especially for forests and woody wetlands (**Fig. 9b**). The effect of the sample size of

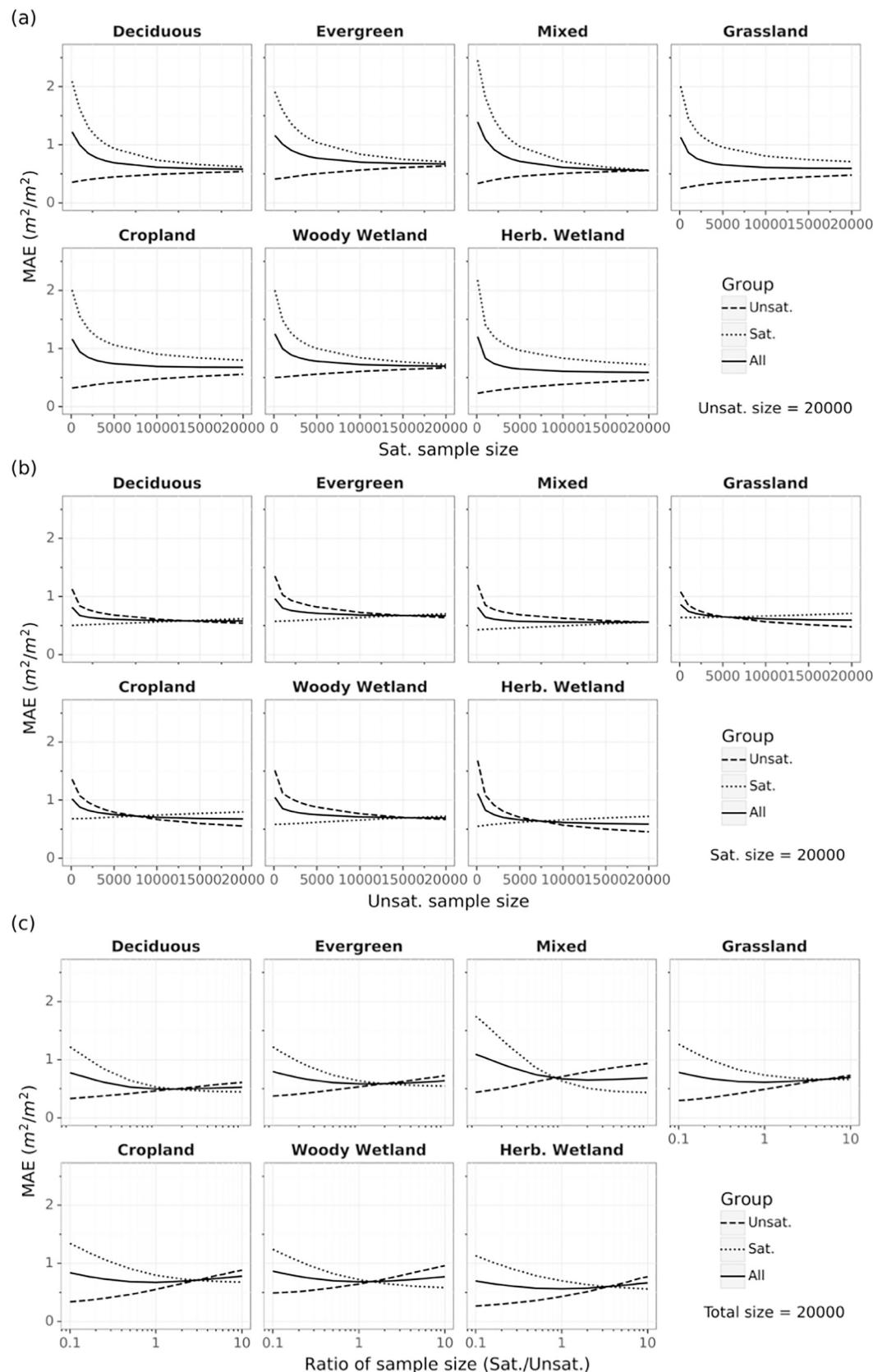


Fig. 4. The impact of training sample balance on LAI estimation accuracy. a) Cross-validation MAE for unsaturated (Unsat.), saturated (Sat.), and combined samples (All.) as a function of saturated sample size in the training set (unsaturated training sample size fixed at 20,000). b) Cross-validation MAE as a function of unsaturated sample size (saturated sample size fixed at 20,000). c) Cross-validation MAE as a function of the ratio between saturated and unsaturated samples in the training set (total sample size fixed at 20,000).

Table 4

Saturation classification accuracy based on 10-fold cross-validation (Landsat 8).

Biome	Accuracy	F1	Precision ^a	Recall ^b
Deciduous	0.89	0.89	0.83	0.96
Evergreen	0.88	0.87	0.82	0.93
Mixed	0.90	0.92	0.87	0.98
Grassland	0.88	0.85	0.81	0.90
Cropland	0.86	0.87	0.81	0.93
Woody Wetland	0.86	0.88	0.84	0.92
Herbaceous Wetland	0.94	0.92	0.91	0.93

^a Precision is defined as the proportion of correct identifications out of all predicted saturated samples.

^b Recall is defined as the proportion of correct identifications out of all actual saturated samples.

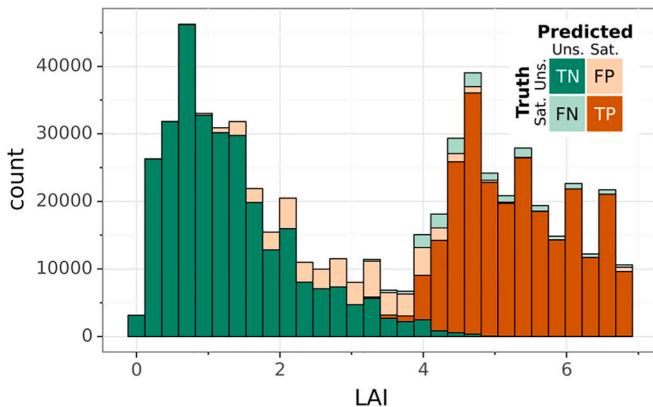


Fig. 5. Distribution of LAI samples based on saturation classification results. Uns.: Unsaturated; Sat.: Saturated; TN: true negative – correctly classified unsaturated samples; TP: true positive – correctly classified; FP: false positive – unsaturated samples but misclassified to saturated; FN: false negative – saturated samples but misclassified to unsaturated. Bars of different categories are stacked.

saturated LAI is also illustrated in Fig. 9c. As the size of saturated samples in the training set increases, the MAE for saturated LAI decreases, while that of unsaturated LAI increases. For most biomes changes in unsaturated LAI error are subtle, but for woody wetland, the increase of unsaturated LAI error is substantial.

3.3.2. Analysis of the scale effect

The difference between aggregated Landsat LAI and MODIS LAI is partly caused by the scale effect, due to the heterogeneity within MODIS pixels and the nonlinearity of the LAI algorithm (Chen, 1999; Wu and Li, 2009). As the heterogeneity of the MODIS pixel increases, indicated by both the mixture of different biomes (information entropy) and CV of Landsat surface reflectance, MAE and bias errors between MODIS and aggregated Landsat LAI increase for both saturated and unsaturated samples (Fig. 10a-d, Table A2). This suggests that the scale effect of MODIS LAI algorithm increase with spatial heterogeneity. The geolocation errors in both MODIS and Landsat images might also contribute to this trend. Yan et al. (2016a) showed that the relationship between MODIS LAI and surface reflectance is a concave function before saturation for deciduous broadleaf forests. In this case, the averaged LAI across Landsat pixels tends to be higher than that of MODIS, which uses surface reflectance at coarse resolution.

To further understand the scale effect, we produced another set of Landsat LAI estimation, by applying random forest models to the aggregated Landsat surface reflectance within MODIS pixels that were covered predominately by a single biome. Results show that the bias error decreases for both unsaturated and saturated LAI in heterogeneous samples, while the impact on homogeneous samples ($CV < 0.2$) is relatively small (Fig. 10c-f). The MAE decreases in unsaturated samples with slight inflation for saturation samples, indicating an increase in variance (random) error. For all biomes, the bias error in both saturated and unsaturated LAI sees a substantial decline (Fig. 10e-f). For example, the positive bias in $LAI < 3$ is reduced for cropland and grassland when aggregating happens at the surface reflectance level (Fig. A3 and A4).

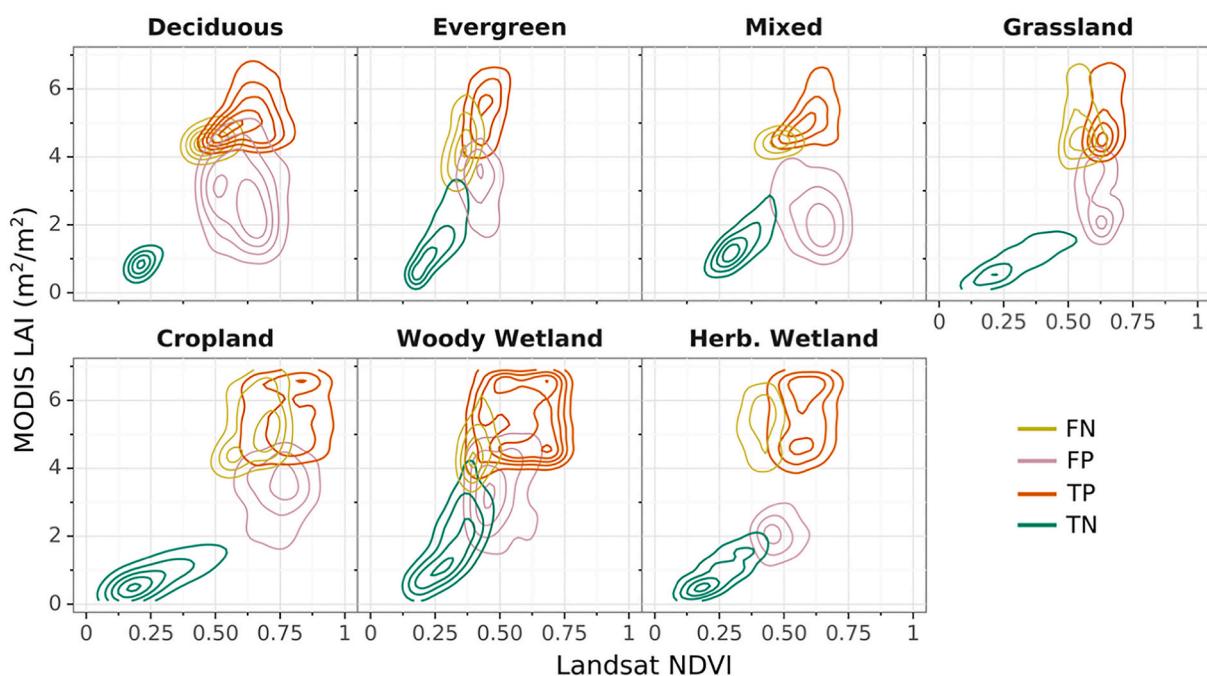


Fig. 6. 2D density contours of the relationship between MODIS LAI and Landsat NDVI for different saturation classification results. TN: true negative – correctly classified unsaturated samples; TP: true positive – correctly classified saturated samples; FP: false positive – unsaturated samples misclassified to saturated; FN: false negative – saturated samples misclassified to unsaturated.

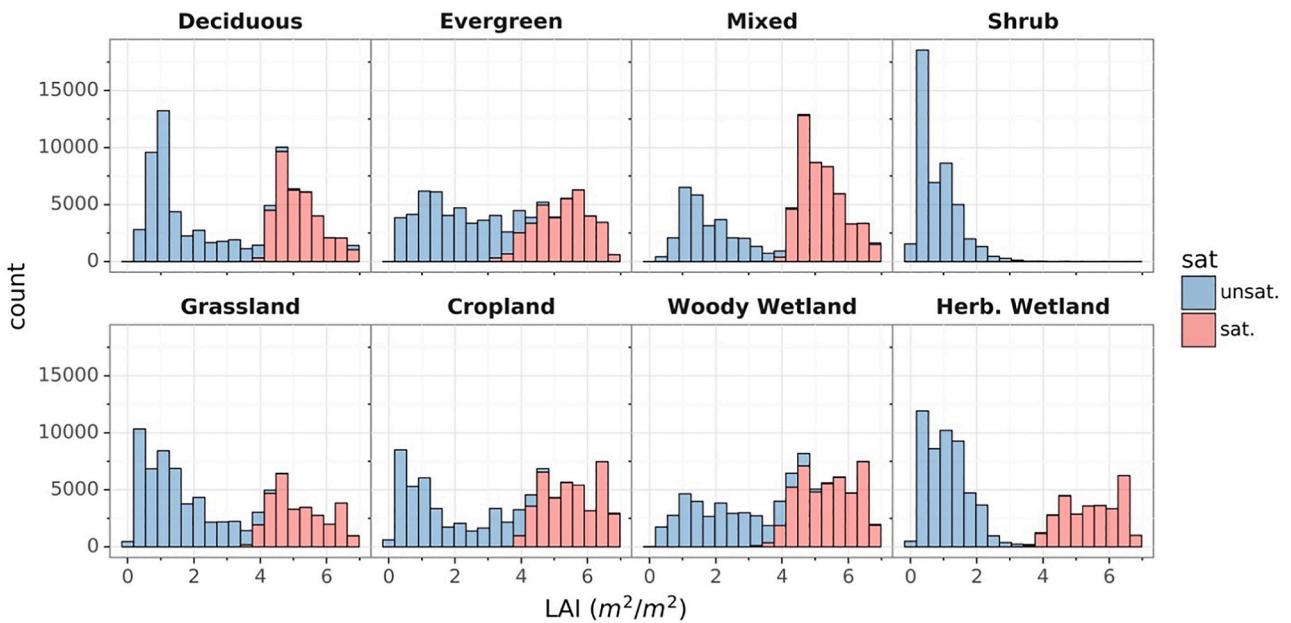


Fig. 7. LAI distributions of unsaturated and saturated samples for eight biomes (Landsat 8). The distributions for Landsat 5 and 7 are similar. Bars of different colors are stacked.

3.4. Validation with ground measurements

3.4.1. NEON-GBOV

A total of 1968 ground PAI measurements from the NEON-GBOV dataset were used for direct validation. The overall RMSE is $0.8 \text{ m}^2/\text{m}^2$, MAE is $0.61 \text{ m}^2/\text{m}^2$, and r^2 is 0.88 (Table 5). 85% of samples meet the accuracy requirement of 1 LAI/PAI unit or 20%. Overall, the total uncertainty is dominated by random error (variance = 0.71) with a positive bias $\sim 0.37 \text{ m}^2/\text{m}^2$. The uncertainty and its components vary by biome type and site (Fig. 11, Fig. A5). Deciduous forest and mixed forest have an RMSE of 0.86 and 0.96 respectively with small negative biases, and the UAR is 84% and 78% respectively. Evergreen forests have the highest RMSE of 1.16, mostly driven by a large positive bias (~ 0.78), with only 63% of the samples meeting the accuracy requirement. Similarly, Brown et al. (2020a) reported an overestimation of evergreen forest LAI by MODIS for NEON sites, where in situ reference measurements were upscaled through transfer functions to match MODIS spatial resolution. Notably, when Miller's method was used to estimate PAI from DHP photos, the bias for evergreen forests decreases by half to 0.37, and the RMSE reduces to 0.88. The RMSE of shrub, grassland, and cropland is 0.31, 0.81, and 0.55 respectively, driven by a positive bias for low PAI. These biomes also achieve high UAR (100% for shrubs, 84% for grasslands, and 98% for croplands). The RMSE for woody wetland is 0.48 with a negligibly small bias and 93% UAR, but the result might not be representative due to the small sample size – only 14. Additionally, we found that differences between estimated and reference PAI increase with the temporal gap between two consecutive Landsat observations. The overall RMSE for estimations with a temporal gap of less than 10 days, between 10 days and 20 days, and above 20 days (less than 50 days) is 0.78, 0.83, and 0.85 respectively. If the closest satellite estimation was used without interpolation, the corresponding RMSE is slightly higher (0.79, 0.84, 0.98 respectively).

The error between estimated LAI and reference PAI displays clear patterns as a function of LAI for different biomes (Fig. 12). For deciduous forests, residuals of low (<2) and high (>4) LAI are mostly below 1 LAI unit without significant bias. But medium LAI values between 2 and 4 show a negative bias, although the sample size is too small (<20) to draw sound conclusions. For mixed forests, residuals for LAI >4 are mostly below 1, but there is a negative bias in medium LAI. The positive

bias for evergreen forests increases with LAI. As a result, the trend of residuals changes considerably after removing the linear trend (compare column 3 and column 4). A similar trend is also found in grasslands over low to medium LAI. Finally, cropland samples have a positive bias in low LAI, while there is no sample for medium to high LAI.

Additionally, we found that the balanced sample design and saturation screening process has considerably reduced the negative bias for deciduous forests, mixed forests, and woody wetlands (Fig. A6). Saturation screening led to a considerable reduction in bias (Fig. A6b). The negative bias in these biomes decreases as more saturated samples are included in the training set (Fig. A6c). For evergreen forests, the positive bias is enlarged due to the inclusion of saturated samples and removal of confounding samples. The sample balancing and screening process have a negligible effect for grasslands and croplands because the small LAI values are not affected by saturated samples.

3.4.2. Other sites

The validation based on the compiled dataset was conducted and analyzed separately by site, because each site used a different measuring protocol (Fig. 13). The RMSE ranges from 0.52 to $0.91 \text{ m}^2/\text{m}^2$, and the MAE is between 0.43 and $0.69 \text{ m}^2/\text{m}^2$ (Table 6). For forests (HARV-EMS, HARV-HEM, HARV-LPH), there is an overestimation for high LAI. At the GRAPEX site, we did not find any substantial bias in the Landsat LAI estimation. At the Mead site, there is an underestimation for high LAI and overestimation for low LAI values. The underestimation may be partly related to the saturation issue of surface reflectance for dense vegetation. The same bias has been reported in previous studies over the same sites (Houborg et al., 2016; Houborg et al., 2015). The overestimation of LAI values between 0 and $1 \text{ m}^2/\text{m}^2$ may be due to undersampling of similar values in the training set and effects of soil background. A similar pattern is observed in the SMEX02-IA site. The overestimation of medium LAI observations in SMEX02-IA might be partly attributed to clumping effects. For US-Shd, there is not any significant bias, except for a slight overestimation for very low LAI (below $0.5 \text{ m}^2/\text{m}^2$). For US-Var, the estimated LAI is higher than observations between 2 and $3 \text{ m}^2/\text{m}^2$. Note that in situ measurements contain uncertainties from measuring instruments, sampling protocol, as well as temporal and spatial match-up with satellite observations. Therefore, the validation results need to be interpreted with caution, and careful consideration

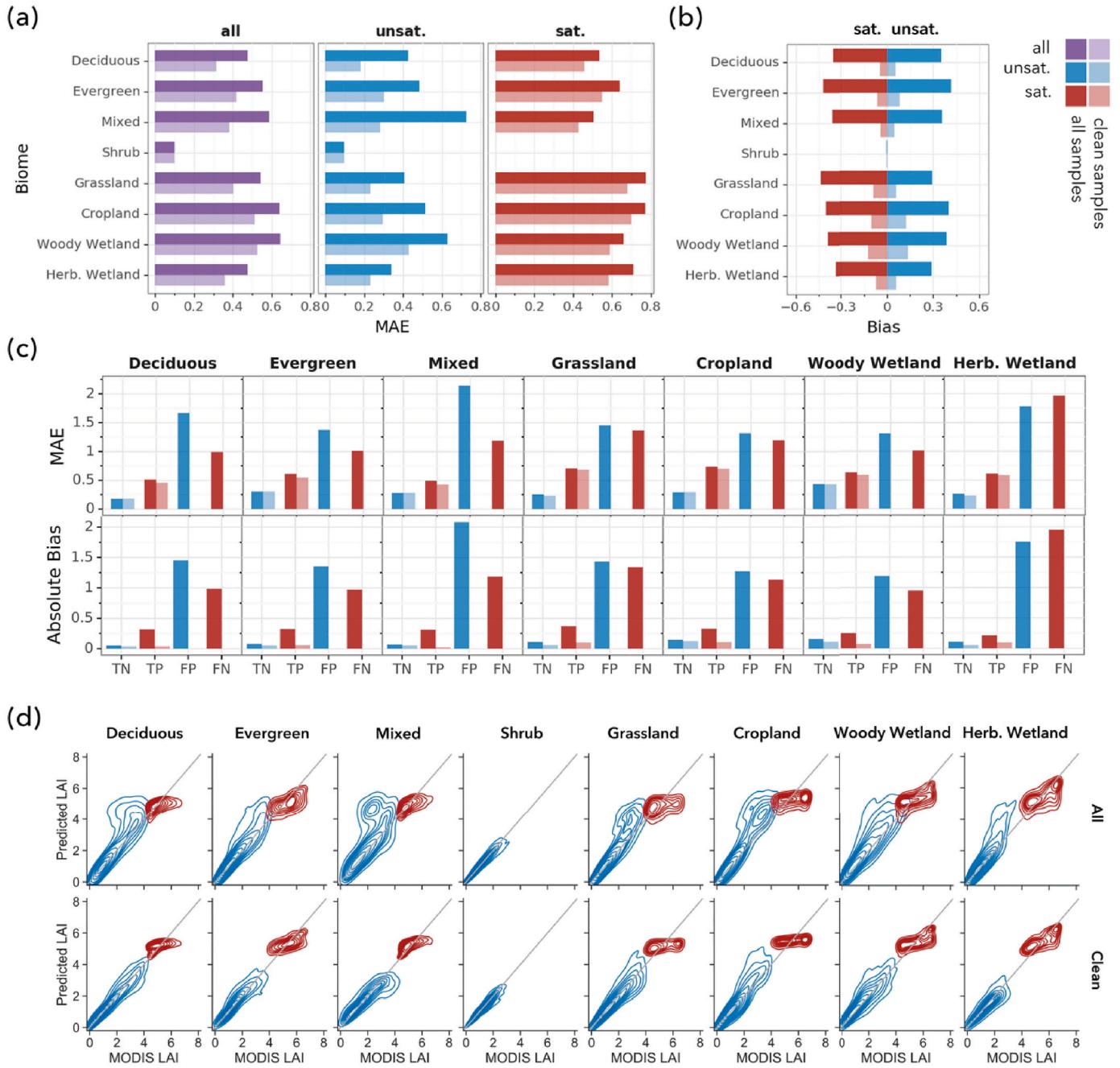


Fig. 8. 10-fold Cross-validation results for eight biomes before (all samples) and after (clean samples) saturation screening. a) Mean Absolute Error (MAE) of LAI for saturated, unsaturated, and combined samples. b) Bias of LAI for saturated, unsaturated, and combined samples. c) MAE and absolute bias of different types of samples (labeled by the results of saturation classification). TN: true negative – correctly classified unsaturated samples; TP: true positive – correctly classified; FP: false positive – unsaturated samples but misclassified to saturated; FN: false negative – saturated samples but misclassified to unsaturated. d) 2D density contour of Landsat predicted vs. MODIS LAI before (All) and after (Clean) saturation screening.

should be given to specific measurement methods at each site.

Fig. 14 shows the temporal evolution of Landsat estimated LAI, ground measured LAI/PAI, and MODIS LAI (500-m resolution). In most sites, the seasonal variation of Landsat LAI agrees well with measurements. In contrast, there are large discrepancies between ground observations and MODIS LAI due to the scale difference, especially for heterogeneous landscapes such as croplands. For example, in GRAPEX and Mead, MODIS LAI is lower than observations at the peak vegetative stage since MODIS operates at a much coarser resolution than the scale of the fields and thus includes information from non-agriculture or non-vegetation surfaces. But with a higher resolution, Landsat LAI captures

the seasonal patterns better. In HARV-HEM, both Landsat LAI and ground measurements show a smaller seasonal variation than that of the deciduous forests sites (HARV-EMS, HARV-LPS), since the evergreen forests retain some leaves during fall and winter. But the MODIS LAI is close to zero in winter, likely because the coarse MODIS pixel also contains patches of deciduous forests.

A comparison between LAI maps derived from MODIS and Landsat indicates the general consistency between the two and the ability of Landsat images to resolve high spatial variabilities in both agricultural and natural ecosystems (Fig. 15). In Fig. 15a, MODIS and Landsat LAI images captured in early Spring, mid-Spring, and mid-Summer depict

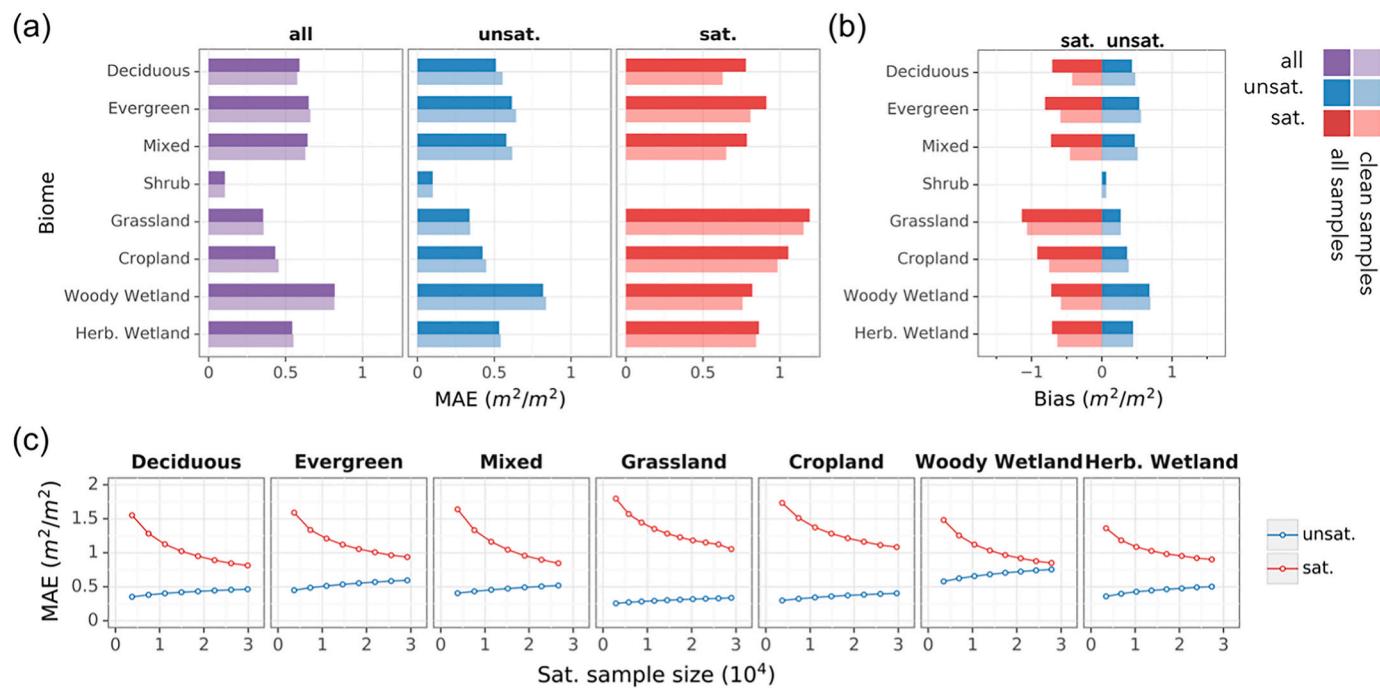


Fig. 9. Effects of sample balance and saturation screening on the comparison between aggregated Landsat LAI and MODIS LAI over CONUS. (a) MAE by biome for unsaturated (unsat.), saturated (sat.), and combined (all) samples. (a) bias by biome for unsaturated (unsat.) and saturated (sat.) samples. (c) Effect of saturated training sample size on the estimation error of saturated and unsaturated LAI.

seasonal changes of agricultural, natural grassland, and forests in parts of the California Central Valley. The spatial patterns and temporal trends of MODIS and Landsat agree well, but Landsat provides more explicit spatial information. In Harvard forests, Landsat LAI maps demonstrate differences in the temporal evolutions of different forest types, which cannot be revealed by MODIS (Fig. 15b). Similarly, in the Mead site, differences in phenological cycles of the three fields (Ne1, Ne2, and Ne3) are reconstructed by the Landsat LAI images but not from the MODIS images (Fig. 15c).

4. Discussion

4.1. Considerations in sample and model design

In this paper, we carefully analyzed the spectral overlap between saturated and unsaturated LAI, which led to a trade-off in the random forest model accuracy of unsaturated LAI ($2\text{--}4 m^2/m^2$) and saturated LAI ($4\text{--}6 m^2/m^2$). Two strategies were proposed to mitigate this effect including 1) adopting a balanced sample design regarding saturation status and 2) using supervised machine learning to remove the most confounding samples. Both strategies were proven effective from the intercomparison with MODIS LAI and the direct validation with ground measurements from NEON sites. Sample imbalance is a prominent problem identified in the machine learning literature, which causes bias towards the prevalent sample category (Krawczyk, 2016; Maalouf and Siddiqi, 2014). While this problem has been studied in the context of classification by the remote sensing community (e.g., Waldner et al., 2019), not much attention has been paid to regression models for estimating ecological and environmental variables. To this end, findings in this paper provide valuable clues on how a skewed dataset could affect the estimation bias and highlight the importance of sample balance for regression problems. Future work could address the effect of balance in a more general context including all predictor variables (surface reflectance, VIs, and geometry) and LAI value ranges. The impact of the size and representativeness of spatial samples should also be assessed.

The proposed approach may benefit from improvements in both

spatial matchup and incorporation of pixel-based illumination and viewing angles. The ground PIFOV of off-nadir MODIS observations is greater than 500 m, thus aggregating Landsat LAI pixels within one MODIS pixel could lead to geolocation errors. Moreover, additional registration errors could arise from the projection difference between MODIS and Landsat. Future studies should improve spatial matchup either by excluding MODIS observations with large viewing angles or using a larger spatial window to determine homogeneity and extract samples. Note that since MODIS LAI is a temporal composite product, it is challenging to locate the actual input image and the viewing angle for MODIS LAI retrievals. Another potential improvement is to incorporate solar illumination and sensor viewing angles at pixel levels to account for the BRDF effect. Currently, we used scene-based solar angles since sensor viewing angles and pixel-based values are not available on Google Earth Engine. Our analysis showed that model performance considerably improved even when only scene-average solar angles were used. Therefore, pixel-based angles will likely further improve the model. Additionally, terrain effects should also be considered as they affect both illumination-viewing geometries and PIFOV (Hao et al., 2018; Wen et al., 2018). Moreover, the impact of using geographic coordinates as predictor variables also needs careful assessment to determine whether and how spatial autocorrelation can be incorporated.

The approach described here provides Landsat-based LAI that is highly consistent with MODIS, especially in homogeneous areas. Since the MODIS LAI algorithm is not scale-invariant, the 500 m resolution MODIS products suffer from scale effect for heterogeneous areas (K. Yan et al., 2016a). Therefore, the degree of consistency between Landsat and MODIS was found to decrease as the spatial heterogeneity increases (Fig. 10). As shown by our analysis, MAE between MODIS and aggregated Landsat LAI increases by 0.4 due to a mixture of land cover (Fig. 10a) and by 0.6 due to vegetation heterogeneity within a single biome (Fig. 10b), indicating a substantial scale effect in MODIS LAI, often overlooked by downstream applications. Thus, future efforts are required to quantify this scale effect and understand its impact on land surface and hydrological modeling.

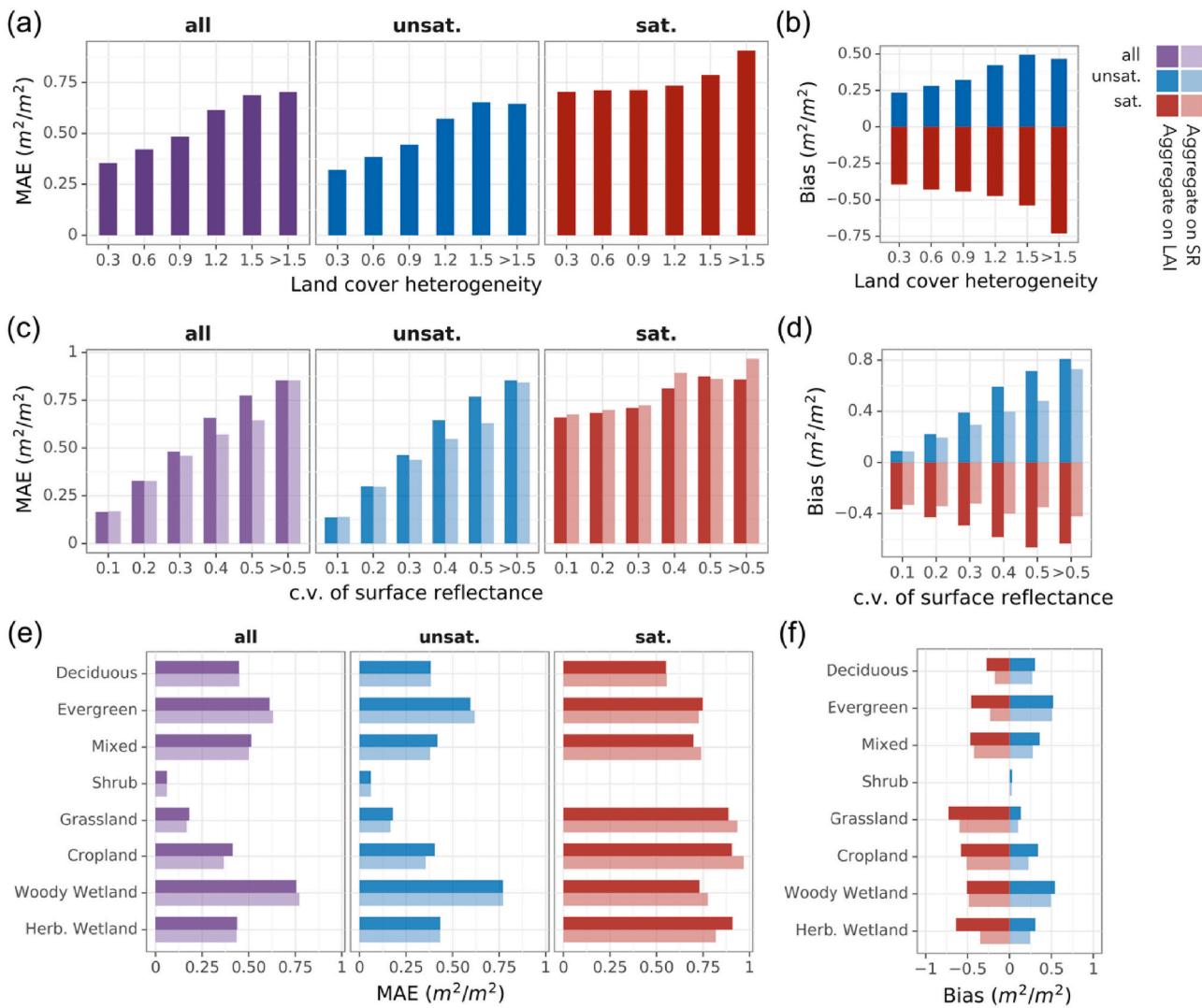


Fig. 10. The impact of pixel homogeneity and scale effects on the comparison between aggregated Landsat LAI and MODIS LAI over CONUS. a) and b) show MAE and bias for samples with different levels of land cover heterogeneity as indicated by the information entropy of NLCD biome within a MODIS pixel. In c) – f), the results only include samples where 90% of the area is covered by a single biome. Bold colors show results from aggregating Landsat LAI, and subtle colors show results from aggregating Landsat SR, which is converted to LAI. c) and d) show MAE and bias for samples with different levels of spatial homogeneity under various spatial aggregation schemes. CV (coefficient of variation) is computed using Landsat surface reflectance and used as an indicator of spatial homogeneity. e) and f) show MAE and bias of saturated and unsaturated samples for different biomes.

Table 5

Summary of direct validation errors by biome for the NEON-GBOV dataset. Variance is a proxy of precision defined as the statistical variance of the difference between estimated LAI and reference PAI. UAR (Uncertainty Agreement Ratio) represents the percentage of samples whose estimation accuracy conforms to the user's requirement – the maximum of 1 unit or 20% of LAI.

Biome	Count	PAI range	RMSE	MAE	Bias	Variance	r^2	UAR
Deciduous Forest	334	0.6–8.2	0.86	0.62	-0.18	0.84	0.73	84%
Evergreen Forest	332	0.5–5.5	1.16	0.92	0.78	0.86	0.75	63%
Mixed Forest	143	1.7–6.5	0.96	0.72	-0.29	0.92	0.55	78%
Shrub	495	0–0.2	0.31	0.30	0.30	0.10	0.37	100%
Grassland	524	0–2.1	0.81	0.70	0.70	0.41	0.65	84%
Cropland	126	0–2.5	0.55	0.50	0.50	0.24	0.84	98%
Woody Wetland	14	2.1–5.6	0.48	0.35	-0.07	0.48	0.87	93%
Total	1968	0–8.2	0.80	0.61	0.37	0.71	0.88	85%

4.2. Reliance on land cover map

Our LAI algorithm uses a high-resolution land-cover product (e.g., the NLCD dataset) to determine the biome type for each pixel and apply the appropriate machine learning model. The relationship between LAI

and surface reflectance typically depends on the biome type. Moreover, building separate models for different biomes allows us to involve more training samples and overcome the memory limit of Google Earth Engine.

However, relying on a land cover product prevents the application of

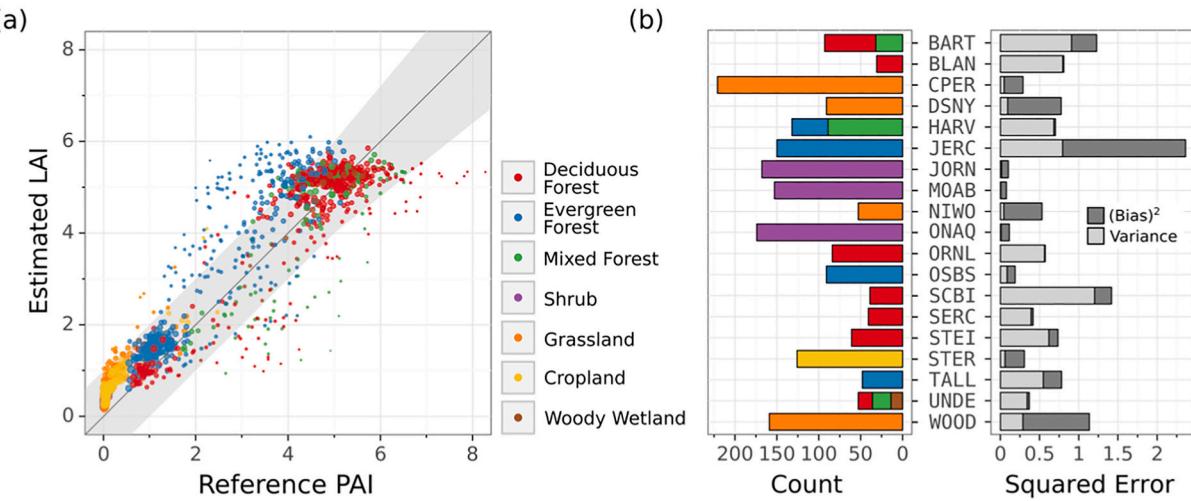


Fig. 11. Comparison between Landsat estimated LAI and reference PAI from the NEON-GBOV dataset. (a) Scatter plot of estimated LAI and reference PAI by biome. Points are scaled by density. The gray shaded area defines uncertainty requirements. (b) Sample counts by biome (left) and relative contributions of systematic (bias squared) and random (variance) components of MSE (right) for each NEON site.

our approach to areas where such data does not exist or is not updated frequently. Moreover, misclassification errors in NLCD could lead to estimation bias in LAI (Fang et al., 2019). Fang et al. (2013) found that the misclassification between similar canopies like crops, grass, and shrub generally does not lead to large errors in MODIS LAI. But the confusion between woody and herbaceous vegetation leads to a considerable bias. In our ground validation analysis, the US-Shd site (grass) is misclassified as cropland, but the estimated LAI does not change much whether the true label or the misclassified label was used. In HARV-EMS and HARV-LPH (deciduous forests), we found that some pixels were misclassified as mixed forest, but this again does not have a considerable impact on the LAI estimation, since the confusion happens within similar biomes. Generally, the NLCD products achieve high accuracy in classes like deciduous forests, evergreen forests, grasslands, and croplands (80% to 90%), but the classification of mixed forests and wetlands is less reliable (L. Yang et al., 2018b). Besides NLCD, the recently published Land Change Monitoring, Assessment, and Projection (LCMAP) product can be an alternative source of land cover information, which is updated yearly (J. F. Brown et al., 2020b). At a global scale, several datasets are available at 10 to 30 m resolution, such as GlobeLand30 (Chen et al., 2014) and FROM-GLC (Gong et al., 2019; Gong et al., 2013). However, global land cover classification at fine resolutions is still challenging, and often these datasets are not updated frequently.

Unlike MODIS, other global moderate-resolution LAI products, such as GEOV (Baret et al., 2016; Baret et al., 2013) and GLASS (Xiao et al., 2016), do not require prior knowledge on land cover. The Sentinel 2 SL2P algorithm is also independent of any land cover information (Weiss and Baret, 2016), but notably, its performance strongly depends on biome type due to the generalized algorithm assumption (Djamai et al., 2019). Future work should investigate the impact of misclassification on LAI estimation and develop global LAI algorithms using widely sampled training sets without the need for land cover maps. Also noteworthy is that the original algorithm proposed by Gao et al. (2012) does not rely on land cover maps, suggesting the potential of machine learning to learn such information together with LAI. To this end, deep neural networks should be explored for their superior abilities in learning spatiotemporal patterns (Tong et al., 2020; Zhong et al., 2019).

4.3. Considerations on ground validation

Quantitative characterization of the uncertainty associated with satellite-derived LAI products is essential for downstream applications.

LAI retrievals from decametric sensors (e.g., Landsat, SPOT, and Sentinel-2) are usually validated through direct field-to-satellite comparison at site levels. In this work, Landsat estimated LAI was compared to ground measurements in 19 NEON sites and eight independent sites from various sources. The validation result depends on biome and site. Notably, there are appreciable biases in Landsat LAI estimates. There is an overestimation for low LAI values ($LAI < 1$) for shrubs, grasslands, and croplands, although the extent of these errors is small. There is a positive bias for evergreen forest over medium to high LAI based on the NEON dataset. It is worth noting that systematic and random errors occur in LAI estimated from both satellite and ground observations. Sources of uncertainties in satellite estimation include noises in input data, model assumptions, mixed pixels, spatial registration, and errors in land cover classification (Fang et al., 2019; Fernandes et al., 2014a). For our approach, algorithmic assumptions of MODIS LAI, temporal compositing of the MODIS product, mechanical differences between Landsat and MODIS sensing systems, and the temporal interpolation of the Landsat LAI all introduce uncertainties. Uncertainties in field LAI measurements could result from inversion models, clumping effects, the inclusion of understory vegetation, as well as human errors (Leblanc and Fournier, 2014; Zou et al., 2020). Future studies are required to understand sources of bias and random errors in both satellite estimation and field measurements and provide comprehensive assessments of these differences.

It is also important to note several challenges in the systematic validation of satellite LAI products, especially at decametric resolutions. First, validation activities have been limited by the availability of field data. Most data were collected in agricultural fields and forests, but data for savanna and wetlands are minimal. Second, the quality of many field measurements of LAI has not been standardized. Notably, several international initiatives have ongoing efforts in the standardization of ground measurements and validation of Earth observation data products. CEOS defines best practice protocols to validate global LAI products (Fernandes et al., 2014a). GBOV aims to provide high-quality in situ reference data with wide spatial and temporal coverage and the NEON dataset used in this study is among the first sets of data from GBOV (L. Brown et al., 2020a). Finally, ground measurements are not always concurrent with the satellite overpass. Temporal compositing or smoothing of satellite data is sometimes desired to increase the availability of usable ground data and reduce noise associated with satellite signals (Houborg et al., 2016; Houborg and McCabe, 2018b).

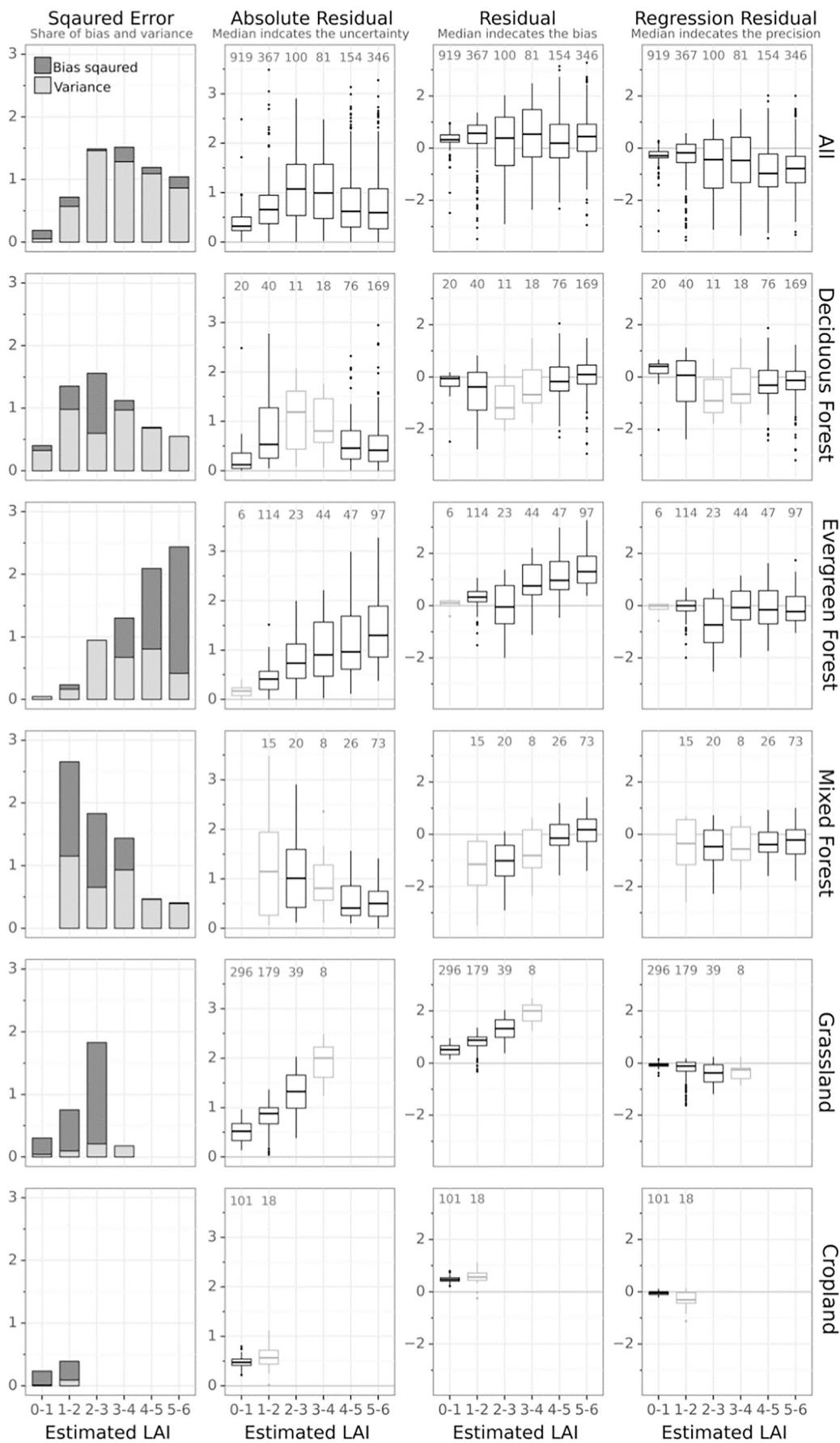


Fig. 12. Uncertainty statistics as a function of estimated LAI by comparing Landsat estimated LAI to reference PAI of the NEON-GBOV dataset. The first column shows mean squared error (MAE) and the relative contribution from the systematic (bias squared) and random (variance) components. The second to fourth columns show boxplots of absolute residuals, residuals, and regression residuals as a function of estimated LAI respectively. In a boxplot, the middle bar presents the median, the box covers 25% and 75% percentiles, and the difference between the two is called the interquartile range (IQR). The whiskers extend from the box to $1.5 \times \text{IQR}$, and data beyond the whiskers are outliers and drawn individually. The median or middle bars of the absolute residual, residual, and regression residual indicate total uncertainty (column 2), bias (column 3), and precision (column 4) respectively. The numbers on top are the count of samples in each LAI value range. Boxes with less than 20 samples (i.e., not statistically significant) are indicated with light gray color. Groups with less than five samples are not shown. The first row shows a summary of all the samples, and the rest shows data for deciduous forests, evergreen forests, mixed forests, grasslands, and croplands. The shrub is not included because all reference samples are below 0.5 LAI. Woody wetland is not shown due to the small sample size.

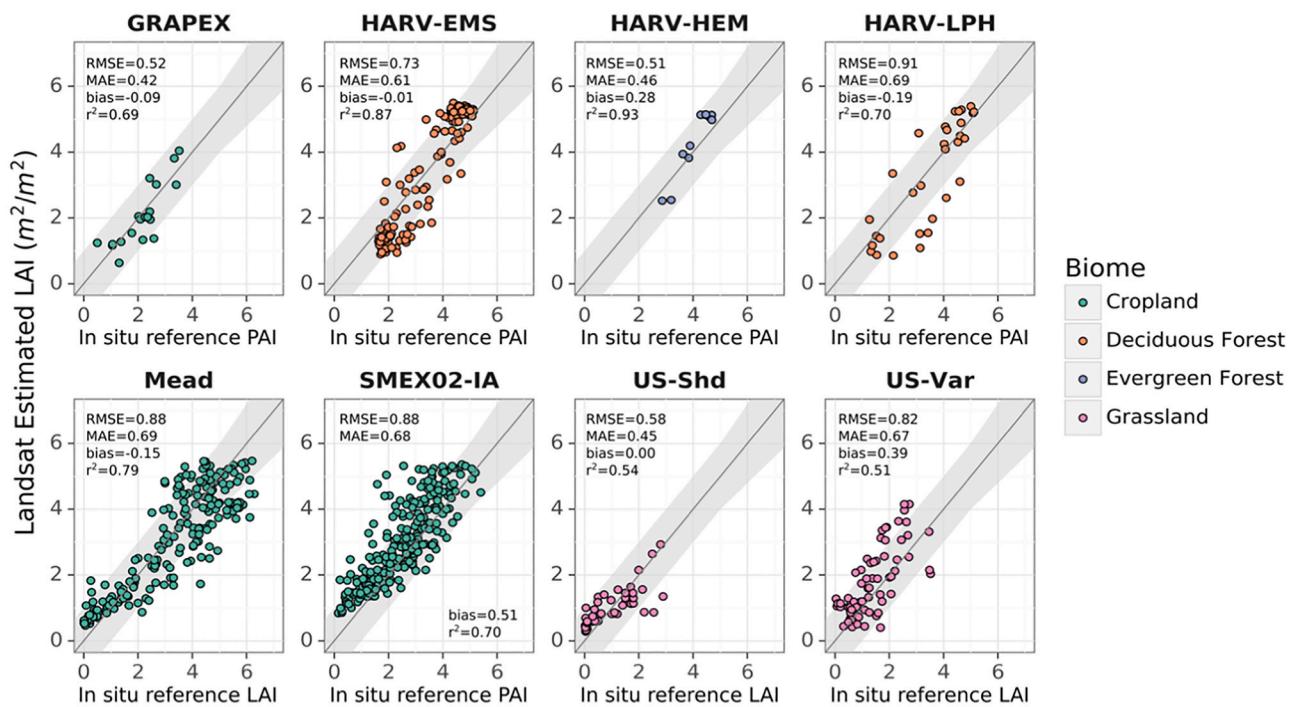


Fig. 13. Comparison of Landsat estimated LAI and ground measurements for eight independent validation sites. The gray shaded area defines uncertainty requirements.

Table 6

Summary of direct validation errors for the eight independent sites. Variance is a proxy of precision defined as the statistical variance of the difference between estimated LAI and reference LAI or PAI. Please refer to Table 3 for the specific LAI definition of each site. UAR (Uncertainty Agreement Ratio) represents the percentage of samples whose estimation accuracy conforms the user's requirement – the maximum of 1 unit or 20% of LAI.

Biome	Count	LAI/PAI range	RMSE	MAE	Bias	Variance	r ²	UAR
GRAPEX	19	0.5–3.5	0.52	0.42	-0.09	0.84	0.69	95%
HARV-EMS	140	1.6–5.1	0.73	0.61	-0.01	0.86	0.87	84%
HARV-HEM	11	2.9–4.7	0.51	0.46	0.28	0.92	0.93	100%
HARV-LPH	32	1.3–5.1	0.91	0.69	-0.19	0.10	0.70	72%
Mead	211	0–6.3	0.88	0.69	-0.15	0.41	0.79	72%
SMEX02-IA	257	0.2–5.4	0.88	0.68	0.51	0.24	0.78	78%
US-Shd	41	0–2.9	0.58	0.45	0.00	0.48	0.93	93%
US-Var	70	0–3.5	0.82	0.67	0.39	0.03	0.73	73%

4.4. Insights for building decametric satellite LAI products

The development of global decametric-resolution LAI products is highly desirable for various applications. Compared to empirical LAI-VI relationships, approaches based on RT modeling are more pertinent for this task without the need for ground samples, but challenges remain in recalibration requirements, computational efficiency, and reflectance saturation. Ganguly et al. (2012) developed a prototype Landsat LAI algorithm based on the MODIS RT model and noted the challenge to accurately estimate model parameters for different Landsat sensors. The hybrid approach employed by the Sentinel-2 SL2P algorithm provides a computationally efficient way to solve RT models with machine learning algorithms (Weiss and Baret, 2016). But since RT modeling is sensitive to the underlying surface reflectance product, both LUT and training datasets used by the hybrid approach need to be re-built for new sensors. This study provides an alternative based on existing moderate-resolution global LAI products. The training samples can be automatically regenerated and processed to adapt the models for new Landsat and MODIS collections using Google Earth Engine. The consistency with MODIS LAI supports land surface and hydrological modeling cross spatial scales. The data-fusion scheme provides further flexibilities in synthesizing various satellite products. The framework is adaptable to

other moderate-resolution LAI products (e.g., Copernicus Global Land Service (CGLS) 300 m LAI and Visible Infrared Radiometer Suite (VIIRS) 500 m LAI) and other decametric multispectral sensors such as Sentinel-2. Using ensembles of moderate resolution LAI products as the backbone could also potentially reduce biases and improve the LAI estimation precision.

Theoretical limitations in RT modeling should also be noted. RT models employ various assumptions to generalize complex vegetation conditions. Therefore, uncertainties especially in the form of bias arises when real conditions violate assumptions. For example, previous studies found that the SL2P algorithm underestimates LAI in biomes like forests and row crops, where the turbid medium assumption of the SAIL RT model does not hold (Brown et al., 2019; Djamai et al., 2019). The MODIS algorithm, which our approach is based on, uses biome-specific 3D RT models and hence is less susceptible to such issues. Nevertheless, estimation bias is still noted for several biomes from our validation results. Moreover, the saturation of surface reflectance in dense vegetation is an intrinsic issue for multispectral remote sensing techniques. Further improvements in LAI estimation will likely benefit from incorporating information on canopy structure from active remote sensing such as LiDAR (Tang et al., 2015) and synthetic aperture radar (Wang et al., 2019).

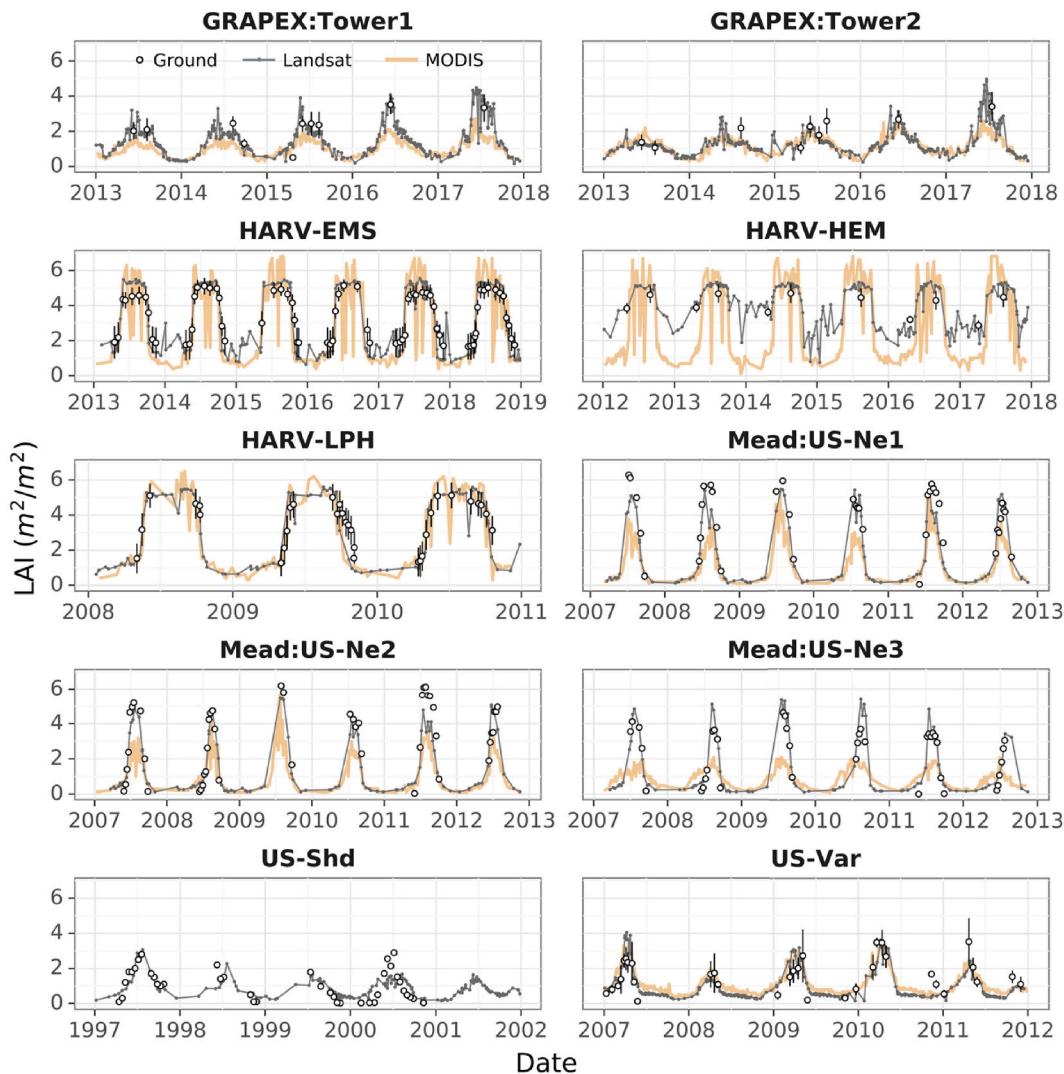


Fig. 14. Time series of Landsat estimated LAI, MODIS LAI, and ground measured LAI/PAI for GRAPEX, Harvard Forest (HARV), Mead, US-Shd, and US-Var. The standard deviations of subplot measurements in the ground measured LAI/PAI were plotted as error bars whenever available. For clear visualization, only selected date ranges were shown for HARV-EMS, Mead, and US-Var. Note that ground measurements in GRAPEX and Harvard Forest are PAI, and those for other sites are LAI.

Additionally, the temporal observation frequency remains a major limit of Landsat-based products due to satellite revisit cycle and cloud coverage. A previous study found that the global average probability of acquiring one cloud-free observation per month on land locations is 0.55 combining both Landsat-5 and Landsat-7 images during a 36-month period centered at 2010 (Kovalskyy and Roy, 2013). The limitation in temporal resolution might be mitigated by combining Sentinel-2, which would improve the temporal frequency of cloud-free observations to 8.4 days on average globally (Claverie et al., 2018), or integrating moderate-resolution images through image fusion techniques (Gao et al., 2015; Houborg et al., 2016).

Future efforts are also required to quantify product uncertainty/precision. Global LAI products often provide theoretical uncertainty estimation from the RT model, which might be challenging for non-parametric machine learning models. A Monte Carlo approach might be explored to approximate prediction uncertainty for random forest (Coulston et al., 2016; Menth and Hooker, 2016). Bayes-based models such as Gaussian processes and Bayesian neural networks could also be explored to derive uncertainty (Verrelst et al., 2013). Intercomparison with other LAI products is also highly desired to understand product precision (Fang et al., 2012a).

5. Conclusions

In this study, we developed and evaluated an LAI mapping algorithm for CONUS based on Landsat images. The algorithm is driven by 1.6 million spatially homogeneous LAI samples derived from MODIS LAI and Landsat surface reflectance. We built random forest models to estimate LAI for eight biomes and three Landsat sensors on Google Earth Engine. The cross-validation MAE ranges from 0.1 to $0.5 \text{ m}^2/\text{m}^2$. The Landsat LAI was aggregated and compared to independent MODIS LAI samples for both homogeneous and heterogeneous areas, and the MAE is between 0.11 and $0.82 \text{ m}^2/\text{m}^2$, despite the scale effect of MODIS LAI. We validated the Landsat LAI algorithm using a large set of ground measurements over CONUS. The RMSE ranges between 0.31 and $1.16 \text{ m}^2/\text{m}^2$ ($\text{MAE} \sim 0.30$ to $0.92 \text{ m}^2/\text{m}^2$) depending on biome and site.

Another contribution of this study is the in-depth analysis of the saturation issue and sample balance. We observe spectral overlaps between unsaturated and saturated MODIS LAI. The overlap issue introduces a tradeoff between the accuracy of regular retrievals in the medium LAI range (3 to $4 \text{ m}^2/\text{m}^2$) and that of the saturated estimates for high LAI values (4 to $6 \text{ m}^2/\text{m}^2$). A balanced sample design regarding saturation status is optimal to achieve the best overall accuracy. A novel approach based on supervised machine learning removes the most

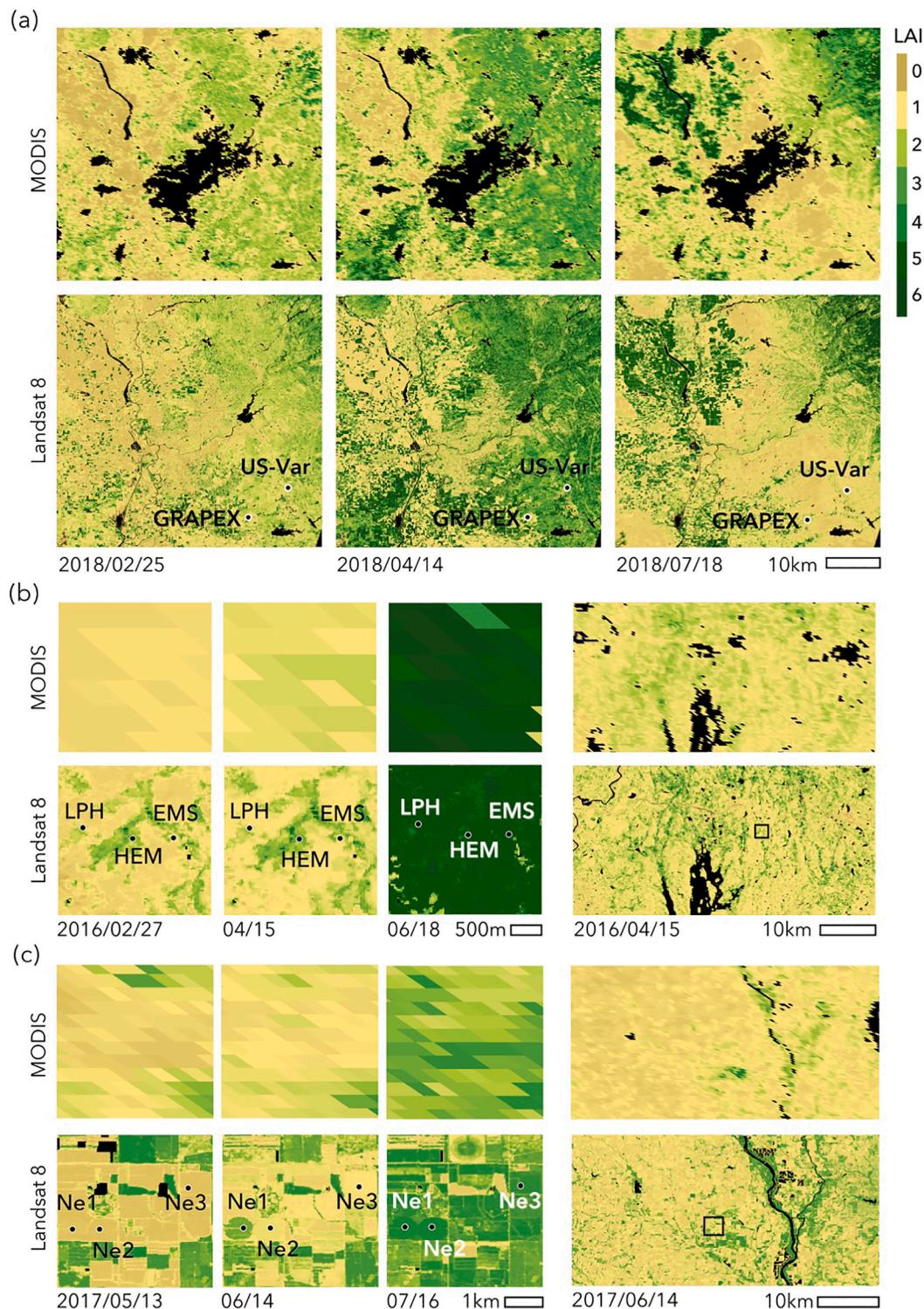


Fig. 15. Comparison of Landsat and MODIS LAI maps at California Central Valley (a), Harvard Forest (b), and Mead (c) sites. (a) LAI maps derived from MODIS and Landsat 8 in parts of California central valley surrounding Sacramento, CA at three timestamps in 2018. This area includes both GRAPEX and US-Var. The large no-data area (i.e., black pixels) in MODIS maps is the City of Sacramento, as MODIS does not produce LAI estimations for urban areas. The Landsat algorithm produces LAI for urban and other non-vegetative land cover types using training samples of all biomes. (b) Zoom-in views (left) and landscape views (right) of Landsat and MODIS LAI maps for Harvard forests at three timestamps in 2016. (c) Zoom-in views (left) and landscape views (right) of Landsat and MODIS LAI maps for the Mead Site at three timestamps in 2017. Black boxes in the landscape view of Landsat maps define the extent of zoom-in views.

confounding samples from the training set and improves the model performance, especially for high LAI values.

The proposed approach allows for generating long-term high-resolution LAI records beyond the MODIS program (since 1999), thanks to

the legacy of the Landsat mission (back to the 1980s). The improved resolution is essential to resolve high spatial heterogeneity in both natural and human-intervened landscapes. The algorithm is powered by cloud computing techniques of the Google Earth Engine platform. Our

findings also highlight the need for a deeper understanding of sample balance in regression-based applications for the retrieval of environmental and ecological variables using satellite images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by the S.D. Bechtel, Jr. Foundation; the Gordon and Betty Moore Foundation; the Walton Family Foundation; the Lyda Hill Philanthropies; the Water Funder Initiative; and the OpenET project. This research was supported in part by an appointment

to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by ORAU under DOE contract number DE-SC0014664. Y. K. also acknowledge partial funding from the NASA Space and Earth Science Fellowship (NNX12AN59H). We would like to thank the staff of the Viticulture Research Division of E. & J. Gallo Winery for the collection and processing of field data during GRAPEX Intensive Observation Periods. We are grateful to Andy Suyker, William Munger, Mark Vanscoy, Audrey Barker Plotkin, Luke Brown, and Christophe Ereboung for providing detailed information regarding ground LAI measurements. We would also like extend our thanks to the editors and five anonymous reviewers for their constructive comments which has substantially strengthened the manuscript.

Appendix A. Supplementary figures and tables

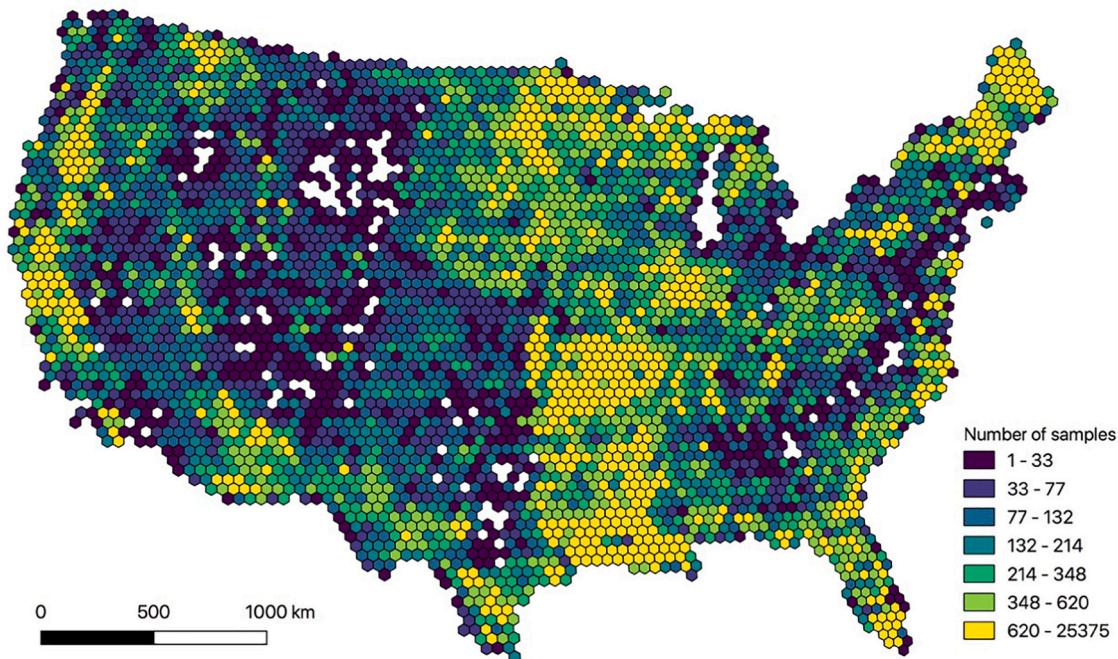


Fig. A1. Spatial distribution of training samples (1.6 million) over CONUS. The color of hexagon grids represents the number of samples within the area. The hexagon grids were generated using H3 (A Hexagonal Hierarchical Geospatial Indexing System) (<https://github.com/uber/h3>).

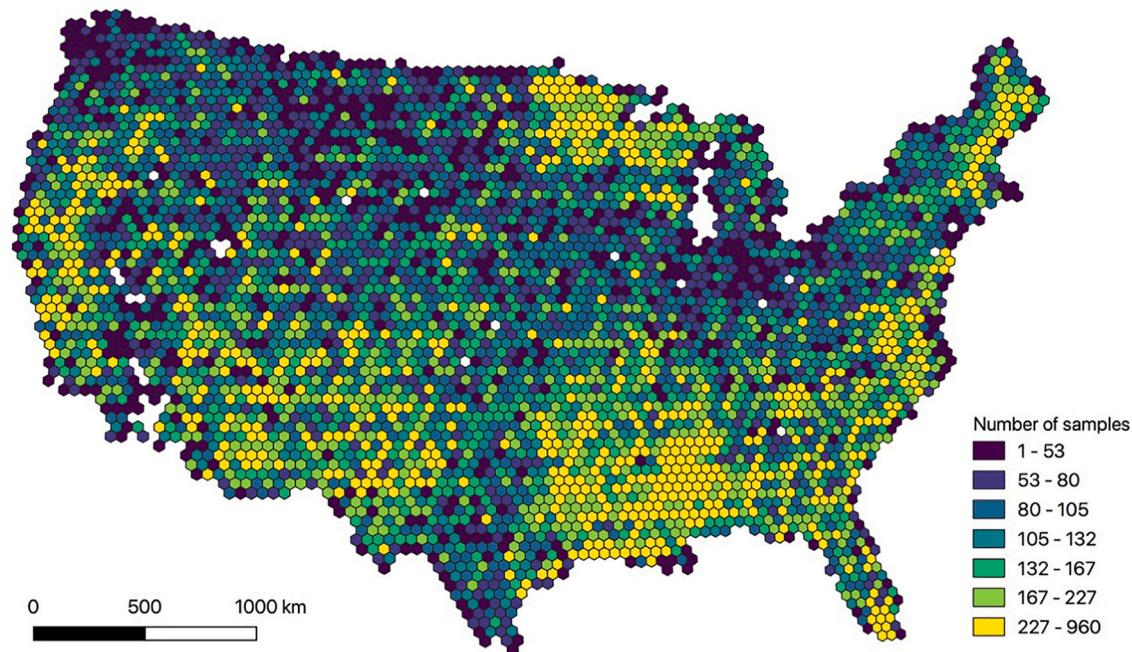


Fig. A2. Spatial distribution of samples (610,000) for intercomparison between Landsat and MODIS. The color of hexagon grids represents the number of samples within the area. The hexagon grids were generated using H3 (A Hexagonal Hierarchical Geospatial Indexing System) (<https://github.com/uber/h3>).

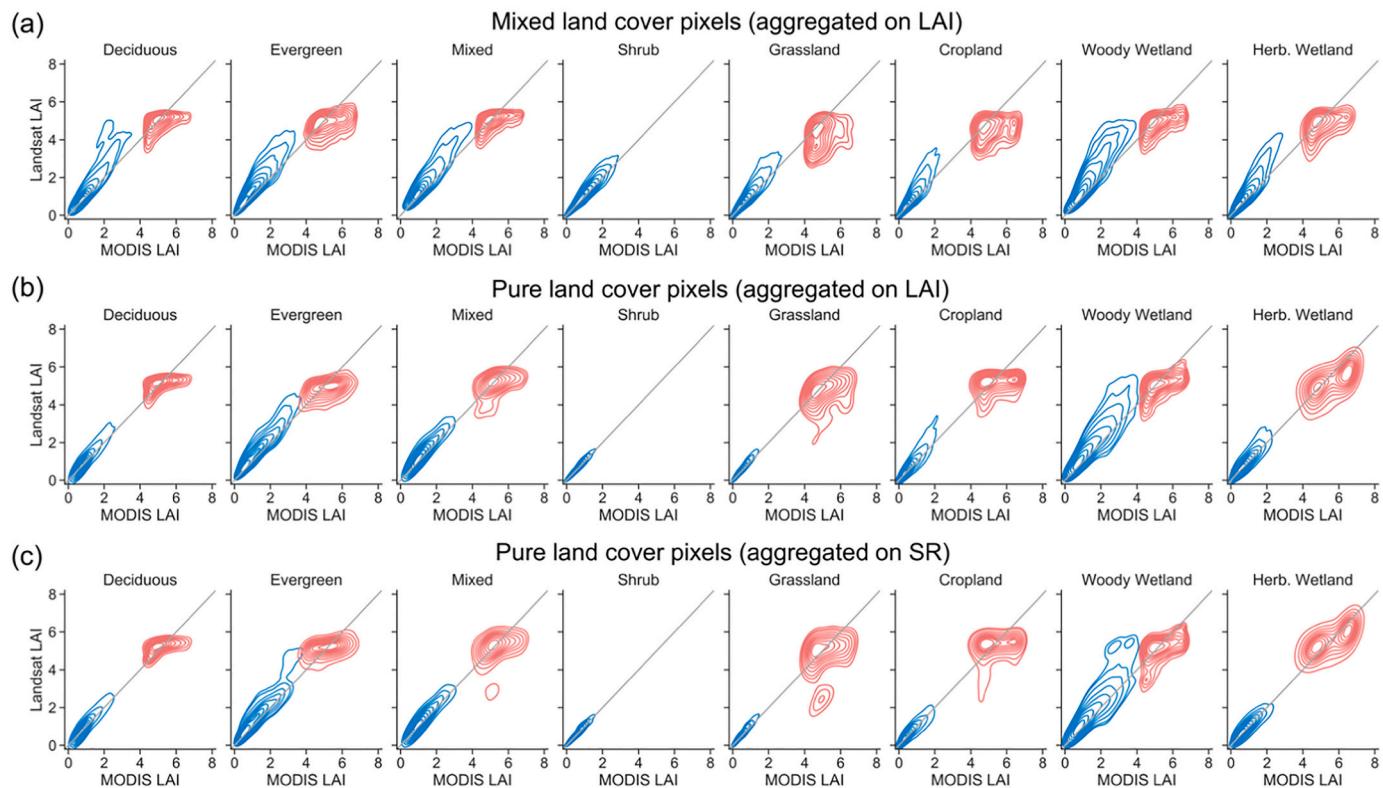


Fig. A3. 2D density contours between aggregated Landsat LAI and MODIS LAI for (a) pixels with mixed biome types, (b-c) pixels with 90% of the area covered by one biome. In (b), Landsat LAI was produced by averaging LAI values. In (c) Landsat LAI was computed based on averaged surface reflectance (SR) values using machine learning models. Blue indicates unsaturated samples, and red indicates saturated samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

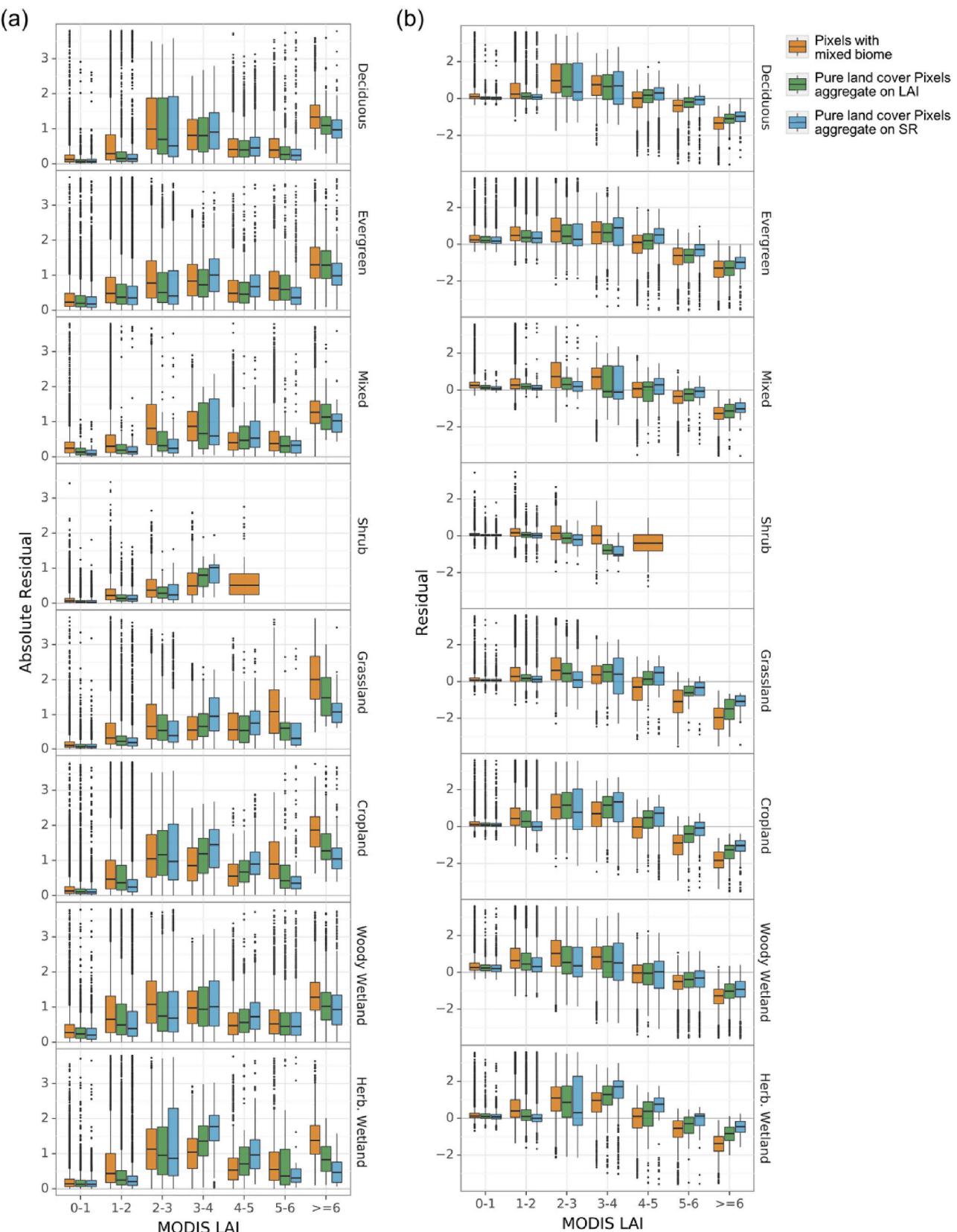


Fig. A4. Boxplots of absolute residuals (a) and residuals (b) between Landsat and MODIS LAI as a function of MODIS LAI. In a boxplot, the middle bar presents the median, the box covers the 25% and 75% percentiles, and the difference between the two is called the interquartile range (IQR). Whiskers extend from the box to $1.5 \times \text{IQR}$, and data beyond the whiskers are outliers and drawn individually. Boxes with less than 20 samples are not shown.

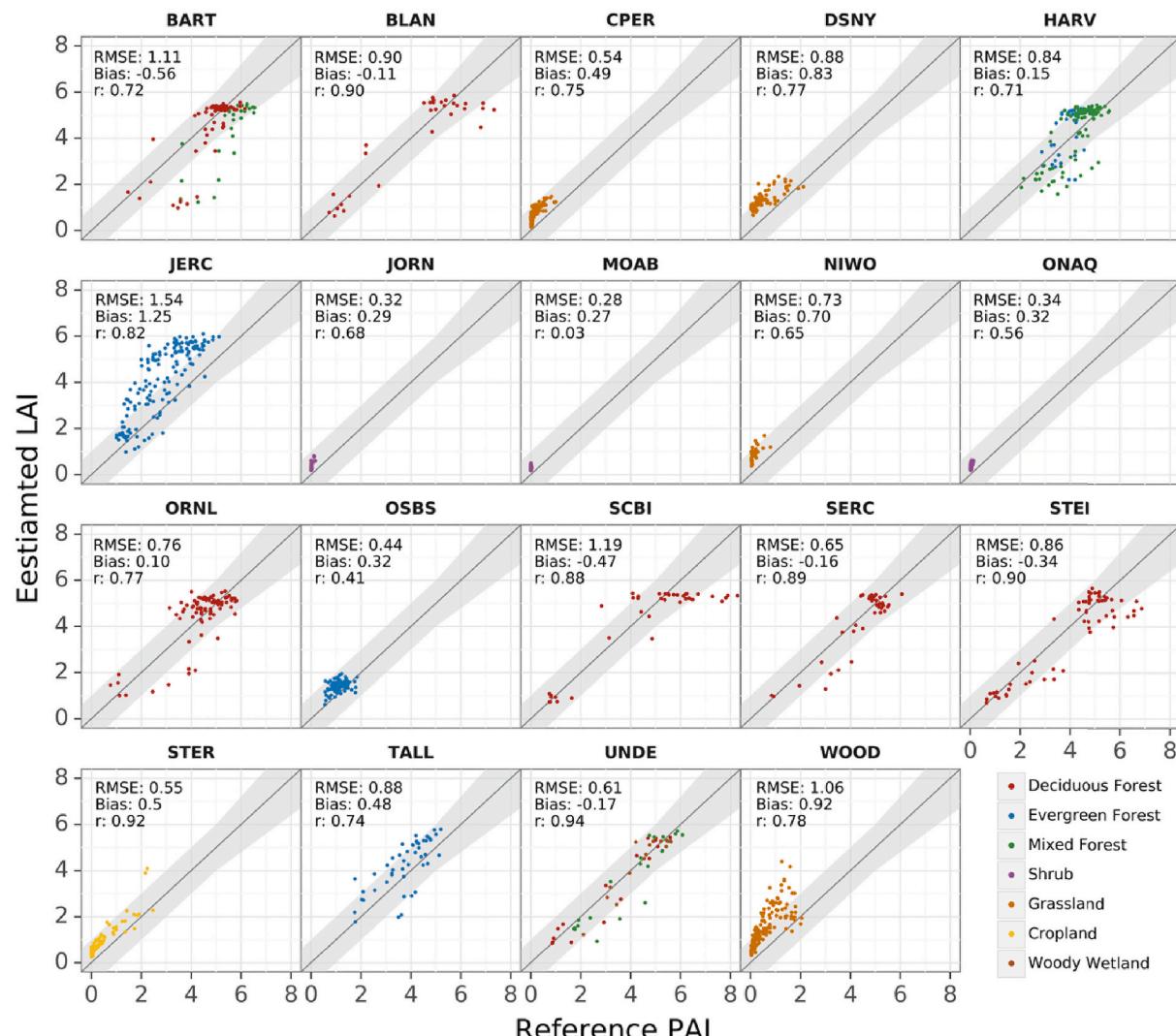


Fig. A5. Comparison between estimated LAI and reference PAI in 19 NEON sites. Gray lines represent a 1:1 relationship, and the shaded area indicates the uncertainty requirement.

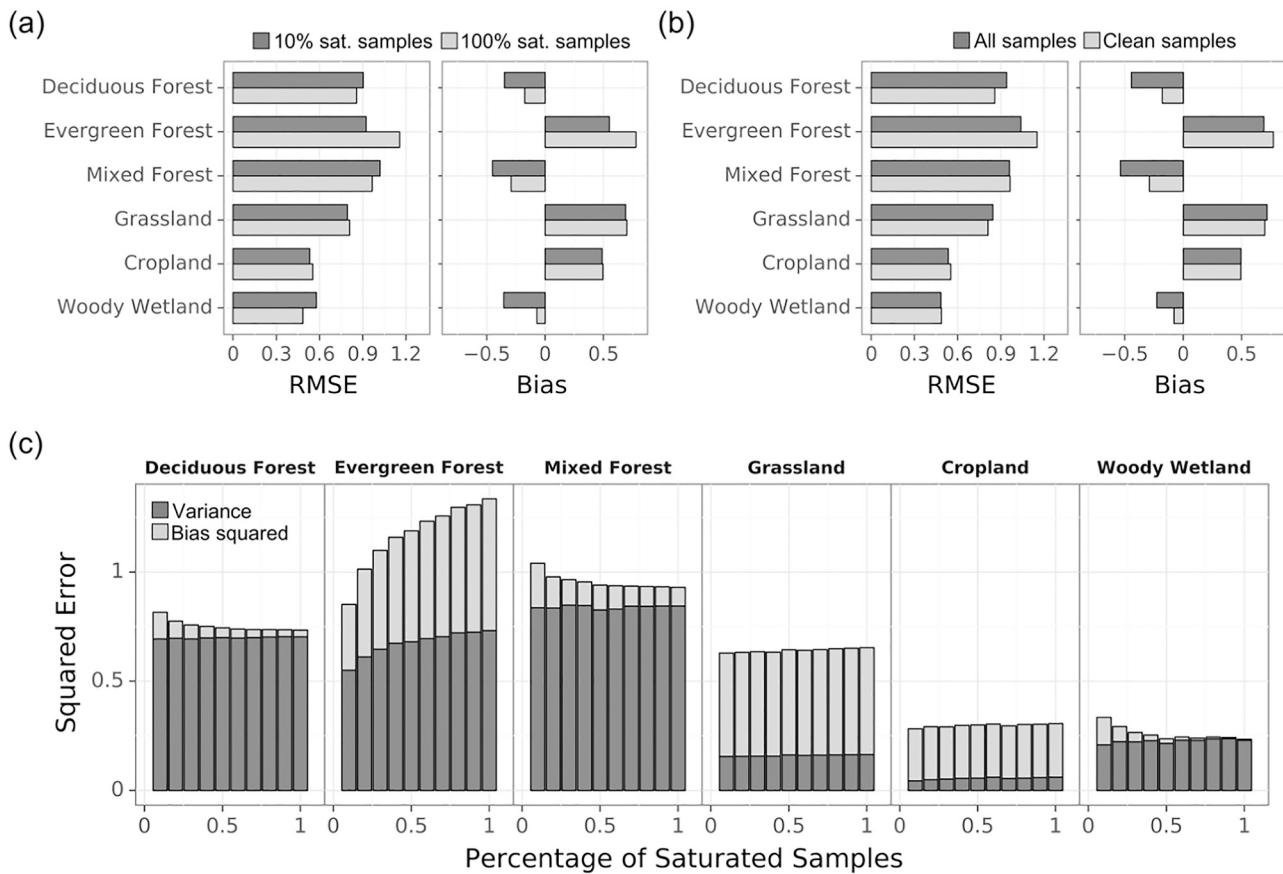


Fig. A6. Effect of saturated sample size and saturation screening on ground validation results for NEON sites. (a) RMSE and bias in LAI estimation from models using different numbers of saturated LAI samples in the training set. Dark gray bars correspond to a scenario where only 10% of all saturated LAI samples were used. Light gray bars correspond to a scenario where all the saturated samples were involved meaning a 1:1 ratio between saturated and unsaturated samples. Both scenarios used all unsaturated LAI samples. (b) RMSE and bias by biome before (All) and after (clean) saturation screening of the training dataset. (c) Impacts of the number of saturated samples used in the training set on the bias (bias squared) and precision (variance) error. The x axis shows the percentage of all available saturated samples used in the training set. All unsaturated samples were used in all cases.

Table A1
Summary of cross validation errors by biome (Landsat 8).

Biome	MAE	RMSE	nRMSE	bias	r^2
Deciduous	0.32	0.49	0.15	0.00	0.95
Evergreen	0.42	0.58	0.17	0.00	0.91
Mixed	0.38	0.53	0.13	0.00	0.91
Shrub	0.10	0.21	0.25	-0.00	0.89
Grassland	0.40	0.61	0.22	0.00	0.92
Cropland	0.51	0.71	0.20	0.00	0.91
Woody Wetland	0.53	0.69	0.17	-0.00	0.87
Herbaceous Wetland	0.36	0.55	0.21	0.00	0.94

Table A2
Intercomparison between MODIS LAI and aggregated Landsat LAI.

Biome	Mixed-biome Pixels						-	Pure-biome Pixels					
	MAE	RMSE	nRMSE	bias	variance	r^2		MAE	RMSE	nRMSE	bias	variance	r^2
Deciduous	0.59	0.92	0.37	0.22	0.80	0.78		0.45	0.82	0.29	0.09	0.81	0.85
Evergreen	0.68	0.95	0.47	0.43	0.71	0.71		0.60	0.88	0.44	0.41	0.78	0.71
Mixed	0.63	0.91	0.33	0.21	0.79	0.77		0.48	0.81	0.30	0.12	0.80	0.83
Shrub	0.14	0.25	0.43	0.09	0.06	0.87		0.06	0.10	0.30	0.03	0.09	0.87
Grassland	0.41	0.71	0.57	0.28	0.42	0.73		0.16	0.30	0.46	0.12	0.28	0.87
Cropland	0.47	0.78	0.70	0.38	0.46	0.74		0.41	0.73	0.74	0.32	0.66	0.79
Woody Wetland	0.83	1.12	0.43	0.40	1.09	0.67		0.74	1.04	0.33	0.15	1.03	0.72
Herbaceous Wetland	0.58	0.89	0.64	0.44	0.61	0.70		0.42	0.73	0.62	0.27	0.68	0.71

References

- Alton, P.B., 2016. The sensitivity of models of gross primary productivity to meteorological and leaf area forcing: a comparison between a Penman-Monteith ecophysiological approach and the MODIS light-use efficiency algorithm. *Agric. For. Meteorol.* <https://doi.org/10.1016/j.agrformet.2015.11.010>.
- Alton, P.B., 2018. Decadal trends in photosynthetic capacity and leaf area index inferred from satellite remote sensing for global vegetation types. *Agric. For. Meteorol.* <https://doi.org/10.1016/j.agrformet.2017.11.020>.
- Anderson, M.C., Neale, C.M.U., Li, F., Norman, J.M., Kustas, W.P., Jayanthi, H., Chavez, J., 2004. Upscaling ground observations of vegetation water content, canopy height, and leaf area index during SMEX02 using aircraft and Landsat imagery. *Remote Sens. Environ.* 92, 447–464. <https://doi.org/10.1016/j.rse.2004.03.019>.
- Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007a. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation. *J. Geophys. Res. Atmos.* 112, 1–13. <https://doi.org/10.1029/2006JD007506>.
- Anderson, M.C., Norman, J.M., Mecikalski, J.R., Otkin, J.A., Kustas, W.P., 2007b. A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 2. Surface moisture climatology. *J. Geophys. Res. Atmos.* 112, 1–13. <https://doi.org/10.1029/2006JD007507>.
- Anderson, M., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., D'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. *Hydrol. Earth Syst. Sci.* 15, 223–239. <https://doi.org/10.5194/hess-15-223-2011>.
- Anderson, M.C., Zolin, C.A., Sentelhas, P.C., Hain, C.R., Semmens, K., Tugrul Yilmaz, M., Gao, F., Otkin, J.A., Tetraut, R., 2016. The evaporative stress index as an indicator of agricultural drought in Brazil: an assessment based on crop yield impacts. *Remote Sens. Environ.* <https://doi.org/10.1016/j.rse.2015.11.034>.
- Anderson, M., Gao, F., Knipper, K., Hain, C., Dulaney, W., Baldocchi, D., Eichelmann, E., Hemes, K., Yang, Y., Medellin-Azuara, J., Kustas, W., 2018. Field-scale assessment of land and water use change over the California delta using remote sensing. *Remote Sens.* 10 <https://doi.org/10.3390/rs10060889>.
- Ma, Siyan, Xu, Liukang, Verfaillie, Joseph, Baldocchi, Dennis, 2020. AmeriFlux US-Var Vaira Ranch- Ione, Ver. 15-5. AmeriFlux AMP, (Dataset). <https://doi.org/10.17190/AMF/1245984>.
- Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. *Remote Sens. Environ.* 110, 275–286. <https://doi.org/10.1016/j.rse.2007.02.018>.
- Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholczyk, P., Smets, B., 2013. GEOV1: LAI and fAPAR essential climate variables and fCOVER global time series capitalizing over existing products. Part1: principles of development and production. *Remote Sens. Environ.* 137, 299–309. <https://doi.org/10.1016/j.rse.2012.12.027>.
- Baret, F., Weiss, M., Verge, A., Smets, B., 2016. ATBD for LAI, fAPAR and fCover from PROBA-V products at 300M resolution (GEOV3).
- Brown, L., Ogutu, B.O., Dash, J., 2019. Estimating forest leaf area index and canopy chlorophyll content with Sentinel-2: an evaluation of two hybrid retrieval algorithms. *Remote Sens.* 11 <https://doi.org/10.3390/rs11151752>.
- Brown, L., Meier, C., Morris, H., Pastor-Guzman, J., Bai, G., Lerebours, C., Gobron, N., Lanconelli, C., Clerici, M., Dash, J., 2020a. Evaluation of global leaf area index and fraction of absorbed photosynthetically active radiation products over North America using Copernicus ground based observations for validation data. *Remote Sens. Environ.* 247, 111935. <https://doi.org/10.1016/j.rse.2020.111935>.
- Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelman, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R., Rover, J., 2020b. Lessons learned implementing an operational continuous United States national land change monitoring capability: the land change monitoring, assessment, and projection (LCMAP) approach. *Remote Sens. Environ.* 238, 111356. <https://doi.org/10.1016/j.rse.2019.111356>.
- Chen, J.M., 1996. Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agric. For. Meteorol.* 80, 135–163. [https://doi.org/10.1016/0168-1923\(95\)02291-0](https://doi.org/10.1016/0168-1923(95)02291-0).
- Chen, J.M., 1999. Spatial scaling of a remotely sensed surface parameter by contexture. *Remote Sens. Environ.* 69, 30–42. [https://doi.org/10.1016/S0034-4257\(99\)00006-1](https://doi.org/10.1016/S0034-4257(99)00006-1).
- Chen, Jun, Chen, Jin, Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2014. Global land cover mapping at 30m resolution: a POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* <https://doi.org/10.1016/j.isprsjprs.2014.09.002>.
- Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R.K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R.R., Myneni, R.B., 2019. China and India lead in greening of the world through land-use management. *Nat. Sustain.* <https://doi.org/10.1038/s41893-019-0220-7>.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161. <https://doi.org/10.1016/j.rse.2018.09.002>.
- Combal, B., Baret, F., Weiss, M., Trubuil, A., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* 84, 1–15.
- Coulston, J.W., Blinn, C.E., Thomas, V.A., Wynne, R.H., 2016. Approximating prediction uncertainty for random forest regression models. *Photogramm. Eng. Remote. Sens.* 82, 189–197. <https://doi.org/10.14358/PERS.82.3.189>.
- Curran, P., Steven, M., 1983. Multispectral remote sensing for the estimation of green leaf area index. *Philos. Trans. R. Soc. London. Ser. A, Math. Phys. Sci.* 309, 257–270.
- Djamai, N., Fernandes, R., Weiss, M., McNairn, H., Goita, K., 2019. Validation of the sentinel simplified level 2 product prototype processor (SL2P) for mapping cropland biophysical variables using Sentinel-2/MSI and Landsat-8/OLI data. *Remote Sens. Environ.* 225, 416–430. <https://doi.org/10.1016/j.rse.2019.03.020>.
- Fang, H., Wei, S., Jiang, C., Scipal, K., 2012a. Theoretical uncertainty analysis of global MODIS, CYCLOPES, and GLOBCARBON LAI products using a triple collocation method. *Remote Sens. Environ.* 124, 610–621. <https://doi.org/10.1016/j.rse.2012.06.013>.
- Fang, H., Wei, S., Liang, S., 2012b. Validation of MODIS and CYCLOPES LAI products using global field measurement data. *Remote Sens. Environ.* 119, 43–54. <https://doi.org/10.1016/j.rse.2011.12.006>.
- Fang, H., Li, W., Myneni, R.B., 2013. The impact of potential land cover misclassification on modis leaf area index (LAI) estimation: a statistical perspective. *Remote Sens.* 5, 830–844. <https://doi.org/10.3390/rs5020830>.
- Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Rev. Geophys.* <https://doi.org/10.1029/2018RG000608>, 2018RG000608.
- Fernandes, R., Butson, C., Leblanc, S., Latifovic, R., 2003. Landsat-5 TM and Landsat-7 ETM+ based accuracy assessment of leaf area index products for Canada derived from SPOT-4 VEGETATION data. *Can. J. Remote. Sens.* 29, 241–258. <https://doi.org/10.5589/m02-092>.
- Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., Widlowski, J.L., 2014a. Global Leaf Area Index Product Validation Good Practices. Version 2.0. In G. Schaepman-Strub, M. Román, & J. Nickeson (Eds.), Best Practice for Satellite-Derived Land Product Validation (p. 76). Land Product Validation Subgroup (WGCV/CEOS). <https://doi.org/10.5067/doc/ceoswgcv/lpv/lai.002>.
- Fernandes, R., Weiss, M., Camacho, F., Berthelot, B., Baret, F., Duca, R., 2014b. Development and assessment of leaf area index algorithms for the Sentinel-2 multispectral imager. *Int. Geosci. Remote Sens. Symp.* 3922–3925. <https://doi.org/10.1109/IGARSS.2014.6947342>.
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>.
- Ganguly, S., Nemani, R.R., Zhang, G., Hashimoto, H., Milesi, C., Michaelis, A., Wang, W., Votava, P., Samanta, A., Melton, F., Dungan, J.L., Vermote, E., Gao, F., Knyazikhin, Y., Myneni, R.B., 2012. Generating global leaf area index from Landsat: algorithm formulation and demonstration. *Remote Sens. Environ.* 122, 185–202. <https://doi.org/10.1016/j.rse.2011.10.032>.
- Gao, B.C., 1996. NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).
- Gao, F., Anderson, M.C., Kustas, W.P., Wang, Y., 2012. Simple method for retrieving leaf area index from Landsat using MODIS leaf area index products as reference. *J. Appl. Remote. Sens.* 6, 063554. <https://doi.org/10.1117/1.JRS.6.063554>.
- Gao, F., Anderson, M.C., Kustas, W.P., Wang, Y., 2014. Retrieving leaf area index from Landsat using MODIS LAI products and field measurements. *IEEE Geosci. Remote Sens. Lett.* 11, 773–777. <https://doi.org/10.1109/LGRS.2013.2278782>.
- Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. *IEEE Geosci. Remote Sens. Mag.* <https://doi.org/10.1109/MGRS.2015.2434351>.
- Garrigues, S., Shabanov, N.V., Swanson, K., Morissette, J.T., Baret, F., Myneni, R.B., 2008. Intercomparison and sensitivity analysis of leaf area index retrievals from LAI-2000, AccuPAR, and digital hemispherical photography over croplands. *Agric. For. Meteorol.* 148, 1193–1209. <https://doi.org/10.1016/j.agrformet.2008.02.014>.
- Goel, N.S., Thompson, R.L., 1984. Inversion of vegetation canopy reflectance models for estimating agronomic variables. IV. Total inversion of the SAIL model. *Remote Sens. Environ.* 15, 237–253. [https://doi.org/10.1016/0034-4257\(84\)90034-8](https://doi.org/10.1016/0034-4257(84)90034-8).
- Gong, P., Wang, J., Yu, Le, Zhao, Yongchao, Zhao, Yuan Yuan, Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Xu, F., Wu, L., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, Han, Zhu, P., Zhao, Z., Zhang, Haiying, Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, Liang, Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, Jin, Chen, Jun, 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* 34, 2607–2654. <https://doi.org/10.1080/01431161.2012.748992>.
- Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, Wenyu, Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, Weijia, Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., Song, L., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci. Bull.* 64, 370–373. <https://doi.org/10.1016/j.scib.2019.03.002>.
- Gonsamo, A., Chen, J.M., 2014. Continuous observation of leaf area index at Fluxnet-Canada sites. *Agric. For. Meteorol.* 189–190, 168–174. <https://doi.org/10.1016/j.agrformet.2014.01.016>.
- Gower, S., Kucharik, C., Norman, J., 1999. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sens. Environ.* 70, 29–51.

- Hao, D., Wen, J., Xiao, Q., Wu, S., Lin, X., You, D., Tang, Y., 2018. Modeling anisotropic reflectance over composite sloping terrain. *IEEE Trans. Geosci. Remote Sens.* 56, 3903–3923. <https://doi.org/10.1109/TGRS.2018.2816015>.
- Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. *Remote Sens. Environ.* 112, 186–202. <https://doi.org/10.1016/j.rse.2007.04.012>.
- Houborg, R., McCabe, M.F., 2018a. A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. *ISPRS J. Photogramm. Remote Sens.* 135, 173–188. <https://doi.org/10.1016/j.isprsjprs.2017.10.004>.
- Houborg, R., McCabe, M.F., 2018b. A Cubesat enabled spatio-temporal enhancement method (CESTEM) utilizing planet, Landsat and MODIS data. *Remote Sens. Environ.* 209, 211–226. <https://doi.org/10.1016/j.rse.2018.02.067>.
- Houborg, R., McCabe, M., Cescatti, A., Gao, F., Schull, M., Gitelson, A., 2015. Joint leaf chlorophyll content and leaf area index retrieval from Landsat data using a regularized model inversion system (REGFLC). *Remote Sens. Environ.* 159, 203–221. <https://doi.org/10.1016/j.rse.2014.12.008>.
- Houborg, R., McCabe, M.F., Gao, F., 2016. A Spatio-temporal enhancement method for medium resolution LAI (STEM-LAI). *Int. J. Appl. Earth Obs. Geoinf.* 47, 15–29. <https://doi.org/10.1016/j.jag.2015.11.013>.
- Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. *Remote Sens. Environ.* 34, 75–91. [https://doi.org/10.1016/0034-4257\(90\)90100-Z](https://doi.org/10.1016/0034-4257(90)90100-Z).
- Kang, Y., Özdogan, M., 2019. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sens. Environ.* 228, 144–163. <https://doi.org/10.1016/j.rse.2019.04.005>.
- Kang, Y., Özdogan, M., Zipper, S.C., Román, M.O., Walker, J., Hong, S.Y., Marshall, M., Magliulo, V., Moreno, J., Alonso, L., Miyata, A., Kimball, B., Loheide, S.P., 2016. How universal is the relationship between remotely sensed vegetation indices and crop leaf area index? A global assessment. *Remote Sens.* 8, 597. <https://doi.org/10.3390/rs8070597>.
- Knyazikhin, Y., Martonchik, J.V., Diner, D.J., Myneni, R.B., Verstraete, M., Pinty, B., Gobron, N., 1998. Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active. *J. Geophys. Res.* 103, 32239. <https://doi.org/10.1029/98JD02461>.
- Kovalskyy, V., Roy, D.P., 2013. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30m Landsat data product generation. *Remote Sens. Environ.* 130, 280–293. <https://doi.org/10.1016/j.rse.2012.12.003>.
- Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. *Prog. Artif. Intell.* 5, 221–232. <https://doi.org/10.1007/s13748-016-0094-0>.
- Kustas, W.P., Anderson, M.C., Alfieri, J.G., Knipppen, K., Torres-Rua, A., Parry, C.K., Nieto, H., Agam, N., White, W.A., Gao, F., McKee, L., Prueger, J.H., Hippo, L.E., Los, S., Alsina, M.M., Sanchez, L., Sams, B., Dokoozlian, N., McKee, M., Jones, S., Yang, Y., Wilson, T.G., Lei, F., McElrone, A., Heitman, J.L., Howard, A.M., Post, K., Melton, F., Hain, C., 2018. The grape remote sensing atmospheric profile and evapotranspiration experiment. *Bull. Am. Meteorol. Soc.* 99, 1791–1812. <https://doi.org/10.1175/BAMS-D-16-0244.1>.
- Lang, A.R.G., Yueqin, X., 1986. Estimation of leaf area index from transmission of direct sunlight in discontinuous canopies. *Agric. For. Meteorol.* 37, 229–243. [https://doi.org/10.1016/0168-1923\(86\)90033-X](https://doi.org/10.1016/0168-1923(86)90033-X).
- Lawrence, P.J., Chase, T.N., 2007. Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *J. Geophys. Res. Biogeosci.* 112 <https://doi.org/10.1029/2006JG000168>.
- Leblanc, S.G., Fournier, R.A., 2014. Hemispherical photography simulations with an architectural model to assess retrieval of leaf area index. *Agric. For. Meteorol.* 194, 64–76. <https://doi.org/10.1016/j.agrformat.2014.03.016>.
- Levitin, N., Kang, Y., Özdogan, M., Magliulo, V., Castillo, P., Moshashay, F., Gross, B., 2019. Evaluation of the uncertainty in satellite-based crop state variable retrievals due to site and growth stage specific factors and their potential in coupling with crop growth models. *Remote Sens.* 11, 1928. <https://doi.org/10.3390/rs11161928>.
- Ma, S., Baldocchi, D.D., Xu, L., Hehn, T., 2007. Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California. *Agric. For. Meteorol.* <https://doi.org/10.1016/j.agrformat.2007.07.008>.
- Maalouf, M., Siddiqi, M., 2014. Weighted logistic regression for large-scale imbalanced and rare events data. *Knowledge-Based Syst.* 59, 142–148. <https://doi.org/10.1016/j.knosys.2014.01.012>.
- Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sens. Environ.* 120, 91–101. <https://doi.org/10.1016/j.rse.2011.09.026>.
- Meier, C., Everhart, J., Jones, K., 2018. TOS Protocol and Procedure: Measurement of Leaf Area Index.
- Mentch, L., Hooker, G., 2016. Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *J. Mach. Learn. Res.* 17, 1–41.
- Meyer, L.H., Heurich, M., Beudert, B., Premier, J., Pfugmacher, D., 2019. Comparison of Landsat-8 and Sentinel-2 data for estimation of leaf area index in temperate forests. *Remote Sens.* 11, 1–16. <https://doi.org/10.3390/rs11101160>.
- Miller, J.B., 1967. A formula for average foliage density. *Aust. J. Bot.* 15, 141–144. <https://doi.org/10.1071/BT9670141>.
- Morton, D.C., Nagol, J., Carabajal, C.C., Rosette, J., Palace, M., Cook, B.D., Vermote, E.F., Harding, D.J., North, P.R.J., 2014. Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature* 506, 221–224. <https://doi.org/10.1038/nature13006>.
- Mousivand, A., Menenti, M., Gorte, B., Verhoef, W., 2015. Multi-temporal, multi-sensor retrieval of terrestrial vegetation properties from spectral-directional radiometric data. *Remote Sens. Environ.* 158, 311–330. <https://doi.org/10.1016/j.rse.2014.10.030>.
- Munger, W., Wofsy, S., 2018. Biomass inventories at Harvard Forest EMS tower since 1993. *Harvard Forest Data Archive: HF069*. <http://pasta.lternet.edu/package/doipl/knb-lter-hfr/69/31>.
- Myneni, R., Hoffman, S., Knyazikhin, Y., Privette, J., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G., Lotsch, A., Friedl, M., Morisette, J., Votava, P., Nemani, R., Running, S., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* 83, 214–231. [https://doi.org/10.1016/S0034-4257\(02\)00074-3](https://doi.org/10.1016/S0034-4257(02)00074-3).
- Norman, J.M., Anderson, M.C., Kustas, W.P., French, A.N., Mecikalski, J., Torn, R., Diak, G.R., Schmugge, T.J., Tanner, B.C.W., 2003. Remote sensing of surface energy fluxes at 101-m pixel resolutions. *Water Resour. Res.* 39 <https://doi.org/10.1029/2002WR001775>.
- Orwig, D.H., 2019. Leaf Area Index at Harvard Forest HEM and LPH Towers since 1998. *Harvard Forest Data Archive: HF150*. <http://pasta.lternet.edu/package/doipl/knb-lter-hfr/150/19>.
- Pfeifer, M., Gonsamo, A., Disney, M., Pellikka, P., Marchant, R., 2012. Leaf area index for biomes of the eastern Arc Mountains: Landsat and SPOT observations along precipitation and altitude gradients. *Remote Sens. Environ.* 118, 103–115. <https://doi.org/10.1016/j.rse.2011.11.009>.
- Ryu, Y., Nilson, T., Kobayashi, H., Sonnentag, O., Law, B.E., Baldocchi, D.D., 2010. On the correct estimation of effective leaf area index: does it reveal information on clumping effects? *Agric. For. Meteorol.* 150, 463–472. <https://doi.org/10.1016/j.agrformat.2010.01.009>.
- Ryu, Y., Baldocchi, D.D., Kobayashi, H., Van Ingen, C., Li, J., Black, T.A., Beringer, J., Van Gorsel, E., Knoll, A., Law, B.E., Roupsard, O., 2011. Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. *Glob. Biogeochem. Cycles* 25, 1–24. <https://doi.org/10.1029/2011GB004053>.
- Sun, L., Gao, F., Anderson, M.C., Kustas, W.P., Alsina, M.M., Sanchez, L., Sams, B., McKee, L., Dulaney, W., White, W.A., Alfieri, J.G., Prueger, J.H., Melton, F., Post, K., 2017. Daily mapping of 30 m LAI and NDVI for grape yield prediction in California vineyards. *Remote Sens.* 9 <https://doi.org/10.3390/rs9040317>.
- Suyker, A., 2001a. AmeriFlux US-Ne1 Mead - irrigated continuous maize site. <https://doi.org/10.17190/AMF/1246084>.
- Suyker, A., 2001b. AmeriFlux US-Ne2 Mead - irrigated maize-soybean rotation site. <https://doi.org/10.17190/AMF/1246085>.
- Suyker, A., 2001c. AmeriFlux US-Ne3 Mead - rainfed maize-soybean rotation site. <https://doi.org/10.17190/AMF/1246086>.
- Tang, H., Ganguly, S., Zhang, G., Hofton, M.A., Nelson, R.F., Dubayah, R., 2015. Characterizing leaf area index (LAI) and vertical foliage profile (VFP) over the United States. *Biogeosci. Discuss.* 12, 13675–13710. <https://doi.org/10.5194/bgd-12-13675-2015>.
- Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* 237, 111322. <https://doi.org/10.1016/j.rse.2019.111322>.
- Turner, D., Cohen, W., Kennedy, R.E., Fassnacht, K.S., Briggs, J.M., 1999. Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sens. Environ.* 70, 52–68. [https://doi.org/10.1016/S0034-4257\(99\)00057-7](https://doi.org/10.1016/S0034-4257(99)00057-7).
- Valderrama-Landers, L.H., España-Boquera, M.L., Baret, F., 2016. Deforestation in Michoacan, Mexico, from CYCLOPES-LAI time series (2000–2006). *IEEE J. Sel. Top. Appl. Earth Obs. Remot. Sens.* <https://doi.org/10.1109/JSTARS.2016.2597742>.
- Verger, A., Filella, I., Baret, F., Peñuelas, J., 2016. Vegetation baseline phenology from kilometric global LAI satellite products. *Remote Sens. Environ.* 178, 1–14. <https://doi.org/10.1016/j.rse.2016.02.057>.
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.* 16, 125–141.
- Verma, S., 2001. AmeriFlux US-Shd Shidler- Oklahoma. <https://doi.org/10.17190/AMF/1246095>.
- Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS J. Photogramm. Remote Sens.* 86, 157–167. <https://doi.org/10.1016/j.isprsjprs.2013.09.012>.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015a. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - a review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G.P.W., Camps-Valls, G., Moreno, J., 2015b. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods - a comparison. *ISPRS J. Photogramm. Remote Sens.* 108, 260–272. <https://doi.org/10.1016/j.isprsjprs.2015.04.013>.
- Viña, A., Gitelson, A.A., Ngwu-Robertson, A.L., Peng, Y., 2011. Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sens. Environ.* 115, 3468–3478. <https://doi.org/10.1016/j.rse.2011.08.010>.
- Waldner, F., Chen, Y., Lawes, R., Hochman, Z., 2019. Needle in a haystack : mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sens. Environ.* 233, 111375. <https://doi.org/10.1016/j.rse.2019.111375>.
- Walther, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., Liang, S., 2004. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sens. Environ.* 92, 465–474. <https://doi.org/10.1016/j.rse.2004.06.003>.

- Wang, Q., Adiku, S., Tenhunen, J., Granier, A., 2005. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sens. Environ.* 94, 244–255. <https://doi.org/10.1016/j.rse.2004.10.006>.
- Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS J. Photogramm. Remote Sens.* 154, 189–201. <https://doi.org/10.1016/j.isprsjprs.2019.06.007>.
- Weiss, M., Baret, F., 2016. S2ToolBox Level 2 products: LAI, FAPAR, FCOVER - Version 1.1. *Sentin. ToolBox Level2 Prod.*
- Wen, J., Liu, Qiang, Xiao, Q., Liu, Qinhuo, You, D., Hao, D., Wu, S., Lin, X., 2018. Characterizing land surface anisotropic reflectance over rugged terrain: a review of concepts and recent developments. *Remote Sens.* 10. <https://doi.org/10.3390/rs10030370>.
- White, W.A., Alsina, M.M., Nieto, H., McKee, L.G., Gao, F., Kustas, W.P., 2018. Determining a robust indirect measurement of leaf area index in California vineyards for validating remote sensing-based retrievals. *Irrig. Sci.* 0, 0. <https://doi.org/10.1007/s00271-018-0614-8>.
- Wilson, J.W., 1963. Estimation of foliage denseness and foliage angle by inclined point quadrats. *Aust. J. Bot.* 11, 95–105. <https://doi.org/10.1071/BT9630095>.
- Wu, H., Li, Z.L., 2009. Scale issues in remote sensing: a review on analysis, processing and modeling. *Sensors* 9, 1768–1793. <https://doi.org/10.3390/s90301768>.
- Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., Song, J., 2016. Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. *IEEE Trans. Geosci. Remote Sens.* 54, 5301–5318. <https://doi.org/10.1109/TGRS.2016.2560522>.
- Xu, B., Li, J., Park, T., Liu, Q., Zeng, Y., Yin, G., Zhao, J., Fan, W., Yang, L., Knyazikhin, Y., Myneni, R.B., 2018. An integrated method for validating long-term leaf area index products using global networks of site-based measurements. *Remote Sens. Environ.* 209, 134–151. <https://doi.org/10.1016/j.rse.2018.02.049>.
- Yan, G., Hu, R., Wang, Y., Ren, H., Song, W., Qi, J., Chen, L., 2016. Scale effect in indirect measurement of leaf area index. *IEEE Trans. Geosci. Remote Sens.* <https://doi.org/10.1109/TGRS.2016.2519098>.
- Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2016a. Evaluation of MODIS LAI/FPAR product collection 6. Part 1: consistency and improvements. *Remote Sens.* 8, 1–16. <https://doi.org/10.3390/rs8050359>.
- Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R.R., Knyazikhin, Y., Myneni, R.B., 2016b. Evaluation of MODIS LAI/FPAR product collection 6. Part 2: validation and intercomparison. *Remote Sens.* 8, 460. <https://doi.org/10.3390/rs8060460>.
- Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., Zhang, W., 2019. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agric. For. Meteorol.* 265, 390–411. <https://doi.org/10.1016/j.agrformet.2018.11.033>.
- Yang, Y., Anderson, M.C., Gao, F., Wardlow, B., Hain, C.R., Otkin, J.A., Alfieri, J., Yang, Yun, Sun, L., Dulaney, W., 2018a. Field-scale mapping of evaporative stress indicators of crop yield: an application over Mead, NE, USA. *Remote Sens. Environ.* 210, 387–402. <https://doi.org/10.1016/j.rse.2018.02.020>.
- Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Bender, S.M., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Granneman, B., Liknes, G.C., Rigge, M., Xian, G., 2018b. A new generation of the United States National Land Cover Database: requirements, research priorities, design, and implementation strategies. *ISPRS J. Photogramm. Remote Sens.* 146, 108–123. <https://doi.org/10.1016/j.isprsjprs.2018.09.006>.
- You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. In: 2017 Association for the Advancement of Artificial Intelligence, pp. 4559–4565. <https://doi.org/10.5555/3298023.3298229>.
- Zhong, L., Hu, L., Zhou, H., Tao, X., 2019. Deep learning based winter wheat mapping using statistical data as ground references in Kansas and northern Texas, US. *Remote Sens. Environ.* 233, 111411. <https://doi.org/10.1016/j.rse.2019.111411>.
- Zhou, Y., Hilker, T., Ju, W., Coops, N.C., Black, T.A., Chen, J.M., Wu, X., 2017. Modeling gross primary production for sunlit and shaded canopies across an evergreen and a deciduous site in Canada. *IEEE Trans. Geosci. Remote Sens.* <https://doi.org/10.1109/TGRS.2016.2615102>.
- Zhou, J., Zhang, S., Yang, H., Xiao, Z., Gao, F., 2018. The retrieval of 30-m resolution LAI from landsat data by combining MODIS products. *Remote Sens.* 10, 1–22. <https://doi.org/10.3390/rs10081187>.
- Zou, J., Hou, W., Chen, L., Wang, Q., Zhong, P., Zuo, Y., Luo, S., Leng, P., 2020. Evaluating the impact of sampling schemes on leaf area index measurements from digital hemispherical photography in Larix principis-rupprechtii forest plots. *For. Ecosyst.* 7 <https://doi.org/10.1186/s40663-020-00262-z>.