MOD-LSP v1.0: MODIS-Based Land Surface Parameters for the Variable Infiltration Capacity Model over the Continental US, Mexico, and Southern Canada

Dataset Description, Methods and User Guide

Theodore J. Bohn[1] and Enrique R. Vivoni[1,2]

1. School of Earth and Space Exploration,
Arizona State University, Tempe, AZ 85287

2. School of Sustainable Engineering and the Built Environment,
Arizona State University, Tempe, AZ 85287

April 2, 2019

Table of Contents

1. Introduction

This document describes the MOD-LSP MODIS-based land surface parameters for the Variable Infiltration Capacity (VIC) model (Liang et al. 1994) compatible with release 5.0 and later (Hamman et al. 2018) in image mode. The MOD-LSP spatial domain covers the continental United States, Mexico, and southern Canada. This spatial domain and 0.625° (6 km) grid resolution are compatible with the gridded daily meteorological forcings of (Livneh et al. 2015), which can be disaggregated to hourly time step via the MetSim tool (Bennett et al. 2018) using domain files developed for PITRI precipitation disaggregation (Bohn et al. 2019). These domain files also can accompany the MOD-LSP parameter files as inputs to VIC simulations.

These parameters have two main purposes: (1) to improve upon previous widely-used parameters over the region with updated, higher-resolution land cover maps and spatially explicit observations of surface properties; and (2) to expand from a single parameter set corresponding to one point in time to a series of parameter sets that account for temporal variability at seasonal to decadal scales.

The previous generation of VIC model parameters over the United States (Maurer et al. 2002; Mitchell et al. 2004; Livneh et al. 2013), Mexico (Zhu and Lettenmaier 2007), and North America (Livneh et al. 2015) share several limitations: (1) they were derived from a relatively coarse-resolution land cover map, based on AVHRR imagery acquired in the early 1990s (Hansen et al. 2000); (2) the annual cycle of monthly leaf area index (LAI) values was derived by spatially interpolating a sparse (< 100 points over North America) set of pixels from a single year of an AVHRR-based dataset (Myneni et al. 1997); (3) in all land cover classes except explicit bare soil, the vegetation canopy was assumed to have 100% area coverage (no gaps); (4) each land cover class was assigned a single spatially (and temporally, for most classes) invariant value of albedo obtained from literature; and (5) urban areas were treated as bare soil.

In addition, the prior parameter sets did not account for land cover variability and change, including: (1) long-term land cover conversions in response to climate change, disturbance, and land use; and (2) interannual variations in phenology (plant characteristics such as LAI and albedo) in response to climate fluctuations. Both types of change occur at varying rates on the landscape (e.g., withering and eventual death of trees in response to drought; Breshears et al. 2005). The land cover map in prior VIC parameter sets represented land surface conditions circa 1992-3. Because no subsequent maps with similar methodology were made, the parameters were not suitable for studies of land cover change. In the subsequent two decades, many land cover change datasets, consisting of land cover classifications using the same methodology and spaced several years apart, have been developed at a variety of scales. In particular, the Multi-Resolution Land Characteristics Consortium (MLRC) National Land Cover Database (NLCD; Homer et al. 2015) contains consistent classifications over the United States for years 1992, 2001, 2006, 2011, and 2016; and the Instituto Nacional de Estadística y Geografía (INEGI) Uso del Suelo y Vegetación land cover product (INEGI; INEGI, 2014) contains consistent classifications over Mexico for years 1985, 1993, 2002, 2007, and 2011. Similarly, interannual variability in phenology, as observed by the Moderate Resolution Imaging Spectroradiometer (MODIS), has recently received attention in hydrologic modeling studies (Hogue et al. 2005; Zhang et al. 2013; Tang et al. 2012; Ford and Quiring 2013; Parr et al. 2015; Tesemma et al. 2015; Bohn and Vivoni 2016; Liu et al. 2018).

MOD-LSP was developed to overcome the limitations of prior parameter sets and facilitate studies of land cover change, over the continental United States, Mexico, and southern Canada. To facilitate land cover change analysis, several versions of the MOD-LSP parameters have been generated using harmonized NLCD and INEGI classifications from years 1992/3, 2001/2, and 2011, as well as a land cover classification based on MODIS. In terms of phenology, MOD-LSP introduces the following improvements: (1) the annual cycle of phenology has been derived from 17 years (2000-2016) of 8- and 16-day MODIS products at nearly all of the 500 m pixels in the domain, yielding spatially explicit and statistically representative estimates of phenology for each land cover class in each grid cell; (2) taking advantage of recent VIC model development (Bohn and Vivoni 2016), canopy fraction (f_{canopy}) and albedo have also been estimated from MODIS products in similar fashion to LAI; and (3) values have been provided for urban areas. Alternate versions of the annual cycles of phenology derived from a single year corresponding to the land cover classification have also been generated. Finally, to account for interannual variability in phenology, monthly time series of phenology spanning the period 2000-2016 have been generated as optional vegetative forcings for the VIC model.

2. Methods

2.1. Overview

Processing of the land surface properties centers around spatial averaging of the MODIS phenology observations over the classes of the chosen land cover classification within each output grid cell, yielding a separate spatial average for each class (Fig. 1). Before aggregation could take place, the NLCD and INEGI land cover classifications needed to be harmonized. After aggregation, data gaps were filled via temporal and spatial interpolation. Finally, to be usable in VIC simulations, other non-MODIS land surface properties were copied from prior parameter sets, mapped from old to new land cover classes where necessary.

2.2. Study Domain

The study domain spans the region bounded by 14.5° and 53° N latitude and 125° and 67° W longitude (Fig. 2). This domain (also used in Livneh et al., 2015) is an extension of the CONUS domain, over which several gridded meteorology and hydrology datasets (Maurer et al. 2002; Mitchell et al. 2004; Xia et al. 2012; Livneh et al. 2013) have been constructed, to include the nation of Mexico; hence we will refer to the domain as CONUS_MX hereafter. In those cases for which our analysis omits southern Canada, we will refer to that subset of the domain as USMX.

2.3. VIC Parameters

2.3.1. General Information

Several spatial datasets were used in creating and evaluating the parameter datasets under consideration herein. Details of satellite sensors, spatial resolution, and acquisition dates for these datasets can be found in Table 1. The VIC parameter datasets discussed herein (described in Table 2) include “L2015”, the parameter set used in Livneh et al. (2015), covering the CONUS_MX domain (obtained from http://ciresgroups.colorado.edu/livneh/data/daily-observational-hydrometeorology-data-set-north-american-extent); and “MOD-LSP”, the suite of new parameter sets with phenology based on MODIS observations, covering either CONUS_MX or USMX domains, depending on the underlying land cover classifications. All VIC parameter datasets were created at 0.0625° (6 km) spatial resolution.

2.3.2. Land Cover Classifications

The land cover classification underlying the L2015 dataset was the AVHRR-based University of Maryland (UMD) land cover product (Hansen et al. 2000), modified for the North American Land Data Assimilation project (NLDAS) (Mitchell et al. 2004) to exclude the Open Water, Urban, and Perennial Ice/Snow classes (UMD-NLDAS hereafter). In each grid cell, the Urban and Perennial Ice/Snow classes were replaced by bare soil, and the Open Water class was replaced by all other classes present in the grid cell in proportion to their area fractions.

The MOD-LSP parameter sets are named based on the underlying land cover classification, according to the convention: *lc_source.lc_year*, where *lc_source* indicates the land cover source and methodology and *lc_year* indicates the year or set of years to which the land cover classification pertains. Values of *lc_source* include: MOD_IGBP, based on the IGBP classification of the MOD12Q1.005 land cover product (Friedl et al. 2010), covering the CONUS_MX domain; and NLCD_INEGI, which covers the USMX domain and is based on the National Land Cover Database (NLCD; Homer et al., 2015) over the United States and the Uso del Suelo y Vegetación land cover product of the Instituto Nacional de Estadística y Geografía (INEGI; INEGI, 2014) over Mexico. To combine the NLCD and INEGI products into a single product, the more numerous INEGI land cover classes were mapped to the 16 NLCD 2011 classes, as shown in Table 3. Details of the procedure can be found in the NLCD_INEGI GitHub archive (https://github.com/tbohn/NLCD_INEGI/releases/tag/v1.5) (Bohn 2019a).

Because the MOD12Q1.005 land cover product provided 13 separate annual maps over the period 2001-2013, during which many pixels oscillated back and forth among several classes, each pixel in the MOD_IGBP parameter set was assigned the most frequent land cover class encountered over the 13-year period, yielding a single map that represented the entire period.

This has been denoted by setting *lc_year* to "mode". The NLCD_INEGI datasets used the National Land Cover Database (NLCD) maps from 1992, 2001, and 2011 (Homer et al. 2007; Fry et al. 2009; Homer et al. 2015) over the US and the Uso del Suelo y Vegetación maps from 1993, 2002, and 2011 (INEGI 2014) over Mexico. It should be noted that although both the NLCD and INEGI products were derived from Landsat imagery, they were created with different methodologies: NLCD was classified automatically on a pixel-by-pixel basis, while the INEGI product was delineated manually into polygons of various classes. Furthermore, the NLCD 1992 product used a different method from that used in subsequent years. The MRLC issued a 1992-2001 land cover change retrofit product (Fry et al. 2009) to facilitate comparison between those years. However, the retrofit product only described changes in broad Anderson Level I land cover categories (e.g., "forest" or "shrub/grassland"). We created a version of the NLCD 1992 map that was both consistent with the retrofit and contained information about all classes as follows: where the 1992-2001 retrofit product indicated that land cover change had occurred, classes at Anderson Level II (e.g., "deciduous forest" or "shrubland") were taken from the NLCD 1992 classification (with 1992 class codes mapped to the 2001 legend); otherwise classes were taken from the NLCD 2001 classification (Homer et al. 2007). For the NLCD_INEGI parameter sets, *lc_year* was set to the year of the NLCD product used (1992, 2001, or 2011), with the INEGI products from 1993 and 2002 used in *lc_years* 1992 and 2001. The *lc_year* code was prepended with "s" in some cases to indicate that only "stable" pixels (those that did not change land cover class over the period 2001-2011) were used to compute spatial average values of phenology variables (see below). Details of the procedure can be found in the NLCD_INEGI GitHub archive (https://github.com/tbohn/NLCD_INEGI/releases/tag/v1.5) (Bohn 2019a).

2.3.3. Time-Varying Surface Properties

The VIC model requires specification of a repeating annual cycle of 12 monthly values of the following time-varying surface properties (phenology) for each land cover class of each grid cell: Leaf Area Index (LAI), albedo, roughness length, and displacement height. An additional annual cycle of canopy fraction (f_{canopy}) is optional, and was included in the MOD-LSP parameter sets. For roughness length and displacement height, all VIC parameter datasets used a spatially-invariant repeating annual cycle of monthly values for each land cover class, taken from literature (Nijssen et al. 2001).

The L2015 parameter set contained a unique annual cycle of monthly LAI values for each land cover class in each grid cell, based on the AVHRR-based product of Buermann et al. (2002) (B2002 hereafter). This annual cycle was derived from only a single year of values from 1992-1993, and from a sparse (< 50) set of pixels from each land cover class, which were extended over the entire CONUS_MX domain by spatial interpolation. For albedo, L2015 used a constant, spatially-invariant value for each land cover class, taken from literature (Nijssen et al. 2001). L2015 did not contain values of f_{canopy}, leading the VIC model to use a default value of 1.0 everywhere.

The MOD-LSP parameter sets obtained monthly phenology from MODIS products: 8-day LAI from the MOD15A2H.006 product for the period between 2000-02-18 and 2001-06-26 and the MCD15A2H.006 product for the period 2001-07-04 through the end of 2016 (Myneni et al. 2002); albedo from the “White-Sky Albedo from shortwave broadband” variable in the 1-day MCD43A3.006 product (Schaaf et al. 2002) for the entire period 2000-2016; and f_{canopy} derived from Normalized Difference Vegetation Index (NDVI) of the 16-day MOD13A1.006 product (Huete et al. 2002) for the entire period 2000-2016. For consistency with LAI, only those albedo observations corresponding to the 8-day LAI schedule were used. For those 8-day intervals when the 16-day NDVI data were unavailable, NDVI values were treated as missing data. Phenology variables were aggregated to the 6 km grid by computing a separate spatial average value for each land cover class in the grid cell:

$$\bar{x}(c,t) = \frac{1}{N_l(c)} \sum_{k_l=1}^{N_l(c)} x\big(k_p(k_l), t\big) \quad , \qquad (1)$$

where x is a phenology variable, k_l is an index of the set of $N_l(c)$ land cover pixels of class c within the cell, $k_p(k_l)$ is the index of the MODIS pixel containing land cover pixel k_l, and t is the index of the time dimension (at 8-day intervals). For the MOD_IGBP dataset, land cover pixels were defined on the same grid as phenology pixels (500 m sinusoidal projection), so that each phenology pixel corresponded to exactly one land cover pixel. For the NLCD_INEGI datasets, the smaller size of land cover pixels (30 m) led to multiple land cover pixels of potentially many different classes corresponding to a single phenology pixel and taking on the same value of x.

Pixels that were contaminated by poor retrievals (dead sensor, bad geometry, clouds present) or snow, as indicated by the FparLAI_QC and FparExtra_QC variables included in the MCD15A2H.006 product, were excluded from the spatial average. In some cases, this led to entire grid cells missing data for one or more acquisition dates, particularly for cells north of 50° N latitude in winter (Fig. 3a-c). These missing values were filled by interpolation later in the process flow. Even after filtering out pixels that were flagged for snow, observations immediately before and after the flagged observations often exhibited albedo values much larger than the typical range of variability during the rest of the year, indicating at least partial snow coverage. Therefore, after the aggregation step, we further nulled out all observations for which the albedo exceeded four standard deviations above the long-term mean.

The MCD15A2H.006 product did not provide estimates of LAI over pixels classified in the MOD12Q1.005 product as open water, barren, perennial snow/ice, wetlands, urban, or unclassified. With the exception of urban pixels, LAI was set to "missing" in these cases. For urban pixels, LAI was estimated via an empirical LAI-NDVI relationship estimated by sampling pixels of shrubland, grassland, and forest classes surrounding the metropolitan areas of Phoenix, Los Angeles, San Francisco, Portland, and Seattle:

$$LAI = 8\,(NDVI - NDVI_{min})^2 \quad , \quad (2)$$

where $NDVI_{min}$ = 0.1.

f_{canopy} was derived from NDVI following Bohn and Vivoni (2016):

$$f_{canopy} = [(NDVI - NDVI_{min})/(NDVI_{max} - NDVI_{min})]^2 \quad , \quad (3)$$

where $NDVI_{min}$ = 0.1 and $NDVI_{max}$ = 0.8.

Gaps (8-day intervals missing data) were filled via the following process (Fig. 3): (1) The time series of each grid cell (Fig. 3d) was separated into its climatological mean and standard deviation (Fig. 3e,f) and standardized anomalies (Fig. 3g), following:

$$x'(c,y,d) = \big(x(c,y,d) - \mu_x(c,d)\big)/\sigma_x(c,d) \quad , \quad (4)$$

Where x' = standardized anomaly, c = land cover class index, y = year, d = day of year, x = phenological variable, and μ_x and σ_x = climatological mean and standard deviation of x for class c and day of year d. (2) All values of μ_x and σ_x for which fewer than 5 observations were available on a given day of the year were set to "missing". (3) In the anomaly time series, missing values either on the first or last time steps or for which the nearest valid data points were more than 2 intervals away were set to 0, and remaining gaps in the anomaly time series were linearly interpolated (Fig. 3h). (4) Temporal gaps in μ_x and σ_x were filled by linear interpolation, treating the climatological cycle as periodic (Fig. 3i,j). (5) Anomalies were recombined with μ_x and σ_x to assemble the final gap-filled time series (Fig. 3k). (6) Any remaining gaps (cells for which no observations were available at any time) were filled with spatial interpolation from values from the same land cover class in neighboring cells, using a Gaussian kernel with σ = 1 cell. The final gap-filled data (Fig. 3l-n) are therefore estimates of snow- and cloud-free values. In the case of albedo, it is presumed that a land surface model such as VIC would replace the input albedo with a simulated snow albedo when snow is present.

The requisite repeating annual cycle of monthly phenology was computed from different sets of years for different MOD-LSP parameter sets. For the MOD_IGBP.mode parameter set, the gap-filled 8-day timeseries of MODIS phenology spanning 2000-2016 were aggregated to a single year of climatological mean monthly values. For the NLCD_INEGI.sYYYY (where YYYY = 1992, 2001, or 2011) parameter sets, these climatological mean cycles were computed using only those 30-m land cover pixels that had a stable land cover class between the 2001 and 2011 maps, thereby providing a climatological cycle free of the impacts of land cover conversion. For the NLCD_INEGI.YYYY (where YYYY = 2001 or 2011) parameter sets, the annual cycle was taken only from year YYYY of the gap-filled MODIS observations.

Thus, the difference between simulations using 2011 and 2001 parameter sets would show the full impact of land cover conversion and within-class interannual variability in phenology between the years 2001 and 2011, while the difference between simulations using s2011 and s2001 would show the impact of land cover conversion alone. To assess the impact of interannual variability of phenology alone, an additional set of 17-year monthly time series was prepared for all MOD-LSP parameter sets via aggregating from the gap-filled 8-day records of

LAI, f_{canopy}, and albedo. For the NLCD_INEGI.2001 and NLCD_INEGI.2011 parameter sets, these time series include the impact of land cover conversion on the time series of phenology (e.g., forest-shrub conversion due to wildfire or logging); however the classification of the pixels in these cases will be that of either year 2001 or 2011.

2.3.4. Non-MODIS Surface Properties

All parameter sets obtained grid cell elevations from the USGS GTOPO30 digital elevation model (Gesch et al. 1999). For the MOD-LSP datasets, GTOPO30 was aggregated directly to 0.0625° resolution, but L2015 was sub-sampled from earlier 0.125° (12 km) spatial resolution datasets; thus their land masks differ along the coastlines. All VIC parameter datasets used the soil properties of L2015, which were obtained from the FAO-UNESCO Digital Soil Map of the World (FAO/UNESCO 1998). Similarly, all VIC parameter datasets used the conceptual soil parameter values (e.g., b_{infilt}, *Dsmax*, and soil layer thicknesses) of L2015, which were derived via calibration in previous studies (Maurer et al. 2002; Zhu and Lettenmaier 2007).

For all VIC parameter datasets, time- and space-invariant surface properties for each land cover class (e.g., presence of overstory, stomatal and architectural resistance parameters) were taken from the L2015 dataset, which in turn obtained these values from literature (Nijssen et al. 2001). Because the new parameter datasets used different land cover classification schemes than the L2015 dataset, these time- and space-invariant parameters were mapped to the most appropriate class of the UMD-NLDAS classification as shown in Tables 4 and 5. Because the L2015 dataset did not contain open water, perennial ice/snow, or urban classes, properties for these classes in the MOD_IGBP and NLCD-INEGI datasets were taken from the grassland class of the L2015 dataset. However, the grassland properties only applied to the vegetated portions of these classes, which were prescribed by MODIS observations of LAI and f_{canopy}. Furthermore, the properties of the open water class can be overridden with those of water when the VIC lake model is turned on.

2.4. Evaluation Datasets

To evaluate the values of f_{canopy} in the new parameter datasets, two independent datasets were used: the NLCD US Forest Service Percent Tree Canopy Cartographic product (Coulston et al. 2012) (NLCD-Forest hereafter) and the NLCD Shrubland product (Xian et al. 2015) (NLCD-Shrub hereafter). These datasets were aggregated from 30-m to 0.0625 ° resolution over the NLCD_INEGI.2011 land cover classification in a similar fashion to the MODIS observations.

2.5. Hydrologic Model

The VIC model formulation is described in detail elsewhere (Liang et al. 1994; Hamman et al. 2018). As of release 4.2, the VIC model optionally can read the annual cycle of f_{canopy} from the parameter file, and when $f_{canopy} < 1.0$, soil evaporation (E_{soil}) will be computed in the gaps between plants (Bohn and Vivoni 2016); otherwise, as in prior releases of the model, f_{canopy} is assumed to equal 1.0 for all land cover classes except the explicit bare soil class. As of release 5.0, in image mode, spatially explicit annual cyles of LAI and albedo must be specified in the parameter file. For albedo, the values from the parameter file are used under snow-free conditions, but the simulated snow albedo is used when snow is present. As of release 4.2, the VIC model can optionally read time series of phenology as additional forcing variables, but as such must be specified at the same time step as the meteorological forcings, which are typically hourly (Bohn and Vivoni 2016). Although the VIC model contains a lake component (Bowling and Lettenmaier 2010), activating it requires additional parameters that are not included in the MOD-LSP parameter sets.

2.6. Public Access

All parameter sets discussed herein are stored in NetCDF files, formatted for input to the Variable Infiltration Capacity (VIC) model (Liang et al. 1994), version 5.0 and later (Hamman et al. 2018), in image mode. The parameter sets are available for download at Zenodo (Bohn and Vivoni 2019a). The L2015 parameter set was converted from ascii to VIC-5-compliant NetCDF format by the "tonic" tool (https://github.com/UW-Hydro/tonic/releases/0.2). The MOD-LSP parameter sets were created via Python scripts (that use the xarray package (Hoyer and Hamman 2017)) from the VIC_Landcover_MODIS_NLCD_INEGI project (Bohn 2019b) archived on GitHub (https://github.com/tbohn/VIC_Landcover_MODIS_NLCD_INEGI/releases/tag/v1.4). The NLCD_INEGI harmonized US-Mexico land cover classifications for years 1992/3, 2001/2, and 2011 are available for download at Zenodo (Bohn and Vivoni 2019b). The scripts used to create the NLCD_INEGI harmonized land cover classifications are archived on GitHub (https://github.com/tbohn/NLCD_INEGI/releases/tag/v1.5) (Bohn 2019a).

3. Results and Discussion

3.1. Forest and Shrubland Canopy Fraction Evaluation

The MOD-LSP estimates of f_{canopy} performed better, relative to the NLCD-Shrub and NLCD-Forest products, than L2015 (Table 6; Fig. 4). July average values of f_{canopy} from NLCD_INEGI.s2011 tended to underestimate the NLCD-Shrub values and overestimate NLCD-forest values by factors of 0.70 and 1.24, respectively, yielding RMSE values of 0.09 and 0.21, respectively. However, the L2015 parameter set, which implicitly assumed f_{canopy} = 1.0 in all vegetated classes, grossly overestimated the NLCD-Shrub and NLCD-Forest values by factors of 7.24 and 2.28, and yielded RMSE values of 0.87 and 0.61.

It should be noted that Eqn (3) neglects impacts of non-linearity of spatial aggregation of heterogeneous NDVI and impacts of soil brightness, soil wetness, and canopy shading (Jiang et al. 2006). Thus, there will be local biases due to local soil and vegetation properties; furthermore, some of the seasonal cycle in f_{canopy} estimated via Eqn (3) are due to seasonal changes in solar angle. Thus, this formula could stand to be refined. Nevertheless, Bohn and Vivoni (2016) found that VIC simulations using f_{canopy} values derived from MODIS NDVI via Eqn (3) resulted in substantial improvements in simulated ET at 60+ shrubland and forest AMERIFLUX (Baldocchi et al. 2001) eddy covariance tower sites.

3.2. Urban Phenology Evaluation

A comprehensive evaluation of urban LAI and f_{canopy} from MOD-LSP has not been undertaken. However, over Santa Barbara, CA, July values of LAI and f_{canopy} compare favorably to 71 randomly-sampled field observations (Alonzo et al. 2015). Observed LAI values had a mean of 1.0035 and standard error of 0.1097. The NLCD_INEGI.s2011 LAI values, averaged over all urban classes in the grid cells containing Santa Barbara, had a mean of 0.8843. Alonzo et al. (2015) reported canopy coverage values of 0.20 (from the UFORE model) and 0.254 (estimated from high-resolution imagery by the city of Santa Barbara). The NLCD_INEGI.s2011 July urban f_{canopy} values had a mean of 0.2256.

3.3. Comparison with Prior Parameter Set – Land Cover Distributions

The L2015, MOD_IGBP, and NLCD_INEGI.2011 land cover classifications generally agree on the geographic distributions of the major land cover categories, although there are notable differences (Fig. 5). For the “forest” category (the sum of all forest and wooded grassland or savanna classes), all three land cover datasets show high coverage in eastern Canada and northeastern United states, the southeastern United States, southern Mexico, the mountains west of 100 °W longitude, and the Pacific Northwest coast (Fig. 5a-c). However, L2015 has higher forest coverage around the fringes of these regions than MOD_IGBP, while NLCD_INEGI has lower coverage. This might be due in part to the spatial resolutions of the underlying land cover products (1 km for UMD, 500 m for MOD12Q1.005, and 30 m for NLCD and INEGI). In particular, the NLCD products do not include classes for mixes of trees and grass (e.g., savanna), likely due to the difficulty of identifying such a mixture from an individual 30-m pixel. For shrublands (Fig. 5d-f) and grasslands (Fig. 5g-i), the three land cover classifications agree on the general locations of these classes (arid western North America) but differ on the boundary between shrubland and grassland. L2015 is more similar to NLCD_INEGI in its partitioning between shrublands and grasslands, with shrublands extending into the northwestern United states, while in MOD12Q1.005 shrublands are confined to the southwestern United States and western Mexico. The three products again agree on the general distribution of agriculture (Fig. 5j-l), but MOD_IGBP has the greatest agricultural coverage in most regions of the US and Canada, and NLCD_INEGI has the greatest coverage in Texas and Mexico.

3.4. Comparison with Prior Parameter Set - Phenology

The improvements of MOD-LSP phenology over that of prior studies are evident when comparing maps over North America. For the “mixed forest” class (Fig. 6), the July LAI of L2015 (Fig. 6a) contains regions of homogeneous values with abrupt boundaries, while MOD_IGBP.mode (Fig. 6d) exhibits a more physically reasonable distribution. The L2015 July f_{canopy} (Fig. 6b) is 1.0 everywhere, while that of MOD_IGBP.mode (Fig. 6e) varies from 0.5 to 1.0 and shows a similar spatial distribution to LAI. Similarly, the L2015 July albedo (Fig. 6c) is 0.18 everywhere, while that of MOD_IGBP.mode (Fig. 6f) ranges from 0.10 to 0.18. Similar problems are evident in the “cropland” class of L2015, relative to that of MOD_IGBP.mode (Fig. 7).

MOD-LSP and L2015 parameter sets also differ in total “scene” phenology (the area-weighted average phenology across all land cover classes) (Fig. 8). In January, the “scene” LAI of L2015 (Fig. 8a) is substantially higher than that of MOD_IGBP.mode (Fig. 8g), particularly in the northern forests. This could be due to both (1) L2015 holding LAI constant year-round for evergreen forest and (2) possible residual impacts from clouds and snow that were not completely accounted for in the MOD-LSP processing. However, in July, the LAI of L2015 (Fig. 8d) and MOD_IGBP.mode (Fig. 8j) look more similar. f_{canopy} remains 1.0 year-round in all vegetated cells in L2015 (Fig. 8b,e) but displays a seasonal cycle in most of the domain in MOD_IGBP.mode (Fig. 8h,k). For albedo, spatial variation is evident in L2015, ranging from 0.12 to 0.20, but there is no temporal variation except over croplands (Fig. 8c,f). The MOD_IGBP.mode albedo exhibits much larger spatial variability than L2015, ranging from 0.10 to 0.30 (Fig. 8i,l).

3.5. Impacts on Hydrology

The aforementioned differences in land cover distributions and phenology between the L2015 and MOD-LSP parameter sets (Figs 5-8) impact water and energy flux partitioning in hydrology simulations. Holding f_{canopy} constant at 1.0 (Fig. 9), the differences in mean annual ET (and Q) between simulations using the MOD_IGBP and L2015 parameter sets are positively (negatively) correlated with the difference in mean annual LAI between the two parameter sets. In the warmer and wetter regions of the domain (south and east of the dashed lines), ET and Q exhibit sensitivities to differences in LAI of about +/-200 mm y^{-1} per unit change in LAI.

However, the differences diminish in the cooler and drier portions of the domain (to the north and west of the dashed lines).

Similarly, allowing f_{canopy} to vary (Fig. 10) results in changes of about -/+50 mm y^{-1} per unit change in f_{canopy} in the warmer and wetter portions of the domain, due to reduced canopy evaporation being outweighed by the increased soil evaporation and transpiration from increased throughfall (Fig. 10). Again, sensitivities to changes in f_{canopy} diminish as climate becomes drier, but where annual P falls primarily as snow (e.g., the western mountains), the sensitivities of ET and Q to changes in f_{canopy} change sign, due to reduced snow storage in the forest canopy, where aerodynamic resistance is lower (and turbulent fluxes higher) than at the surface of the ground snow pack. The total difference between MOD_IGBP with varying f_{canopy} and L2015 is dominated by the difference due to LAI, but the f_{canopy} impacts reinforce the LAI impacts in the western mountains.

3.6. Land Cover Change

The NLCD_INEGI products show substantial land cover change between 2001 and 2011 (Fig. 11). Forests experienced substantial losses in the southeastern United States, the Pacific Northwest, and much of Mexico (Fig. 11a,b). Shrublands and grasslands expanded in the southeastern US and Pacific Northwest, but shrank in northern Mexico (Fig. 11g,h). Agriculture expanded substantially in Mexico (Fig. 11m,n) and cities expanded everywhere (Fig. 11s,t). At smaller scales (boxes 1-4 in Fig. 11b,h,n,t), the NCLD_INEGI products show the loss of forest to shrub/grassland in response to drought and fire in the Rocky Mountains (box 1; Breshears et al. 2005); a mix of forest-shrub, crop-forest, and crop-urban conversion around Lake Michigan (box 2); forest-shrub, forest-crop, and shrub-crop conversion in northwestern Mexico due to pasture clearing and agricultural expansion (box 3; Bohn et al. 2018a); and forest-shrub and forest-urban conversion around Atlanta (box 4).

These changes in land cover, coupled with interannual variations in phenology (via the NLCD_INEGI.2001.2001_2001 and NLCD_INEGI.2011.2011_2011 parameter sets), led to changes in hydrologic fluxes (Fig. 12). Forest-shrub conversion in the Rocky Mountains (box 1) and forest-agriculture conversion in northwestern Mexico (box 3) were accompanied by reductions in LAI and f_{canopy}, leading to reduced ET and increased Q. However, crop-urban conversion around Lake Michigan (box 2) and forest-urban conversion around Atlanta (box 4) did not lead to reductions in LAI or f_{canopy} and had minimal impacts on hydrology. Furthermore, interannual fluctuations in LAI and f_{canopy} in eastern Texas (box 5) led to substantial changes in hydrology without being accompanied by substantial land cover change.

4. Parameter Sets, File Names, and Configuration for Running VIC

There are three types of files included in the MOD-LSP project (Table 7): (1) VIC parameter files, which contain land cover class area fractions ("Cv") in all grid cells, spatially-varying annual cycles of monthly phenology variables ("LAI", "fcanopy", "albedo") for all classes in all grid cells, soil properties, and vegetation structural properties that are invariant in time and space; (2) "veg_hist" compressed tar files, each containing 17 yearly files of monthly time-varying phenology variables; and (3) a template ("global_param.template") for the global parameter file, which specifies the names, locations, and contents of all input and output files, and simulation parameters such as model time step, start and end dates, and model physics options.

To create a global parameter file from the template, the user should make a copy of the template and edit it. Lines beginning with “#” are comment statements that are ignored by the VIC model. All words enclosed in “< >” are placeholders that need to be replaced by the desired values (the “< >” must be removed). All options in the global parameter file are explained in the VIC model documentation (https://vic.readthedocs.io) but the most relevant options are explained in Table 8. Notable variables in the parameter and veg_hist files are described in Table 9.

The veg_hist files are optional; if omitted, the VIC model will use the repeating annual cycle of phenology from the parameter file. These files must contain data at the same time step length as the meteorological forcings. To use the veg_hist files requires (1) disaggregating these files of monthly values to an hourly time step (assuming meteorology is hourly) using the scripts *disagg_veghist_monthly2hourly_nc.py* and *wrap_disagg_veghist_monthly2hourly_nc.pl* from the VIC_Landcover_MODIS_NLCD_INEGI GitHub project (https://github.com/tbohn/VIC_Landcover_MODIS_NLCD_INEGI/releases/tag/v1.4) and (2) setting the appropriate options in the global parameter file (Table 8). Instructions for this procedure can be found in the GitHub archive.

References

Alonzo, M., B. Bookhagen, J. P. McFadden, A. Sun, and D. A. Roberts, 2015: Mapping urban forest leaf area index with airborne lidar using penetration metrics and allometry. *Remote Sens. Environ.*, **162**, 141–153.

Baldocchi, D., and Coauthors, 2001: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.*, **82**, 2415–2434.

Bennett, A., J. J. Hamman, B. Nijssen, E. A. Clark, and K. M. Andreadis, 2018: UW-Hydro/MetSim: Version 1.1.0 (version 1.1.0). *Zenodo*, doi:10.5281/zenodo.1256120. http://doi.org/10.5281/zenodo.1256120 (Accessed June 7, 2018).

Bohn, T. J., 2019a: tbohn/AscGridTools: Tools for processing ascii-format ESRI grid files, v1.0 (Version 1.4). *Zenodo*, doi:10.5281/zenodo.2628880. http://doi.org/10.5281/zenodo.2628880 (Accessed April 4, 2019).

——, 2019b: tbohn/VIC_Landcover_MODIS_NLCD_INEGI: Tools to Create MOD-LSP MODIS-Based Parameters for the Variable Infiltration Capacity (VIC) Model over the Continental United States, Mexico, and Southern Canada, v1.4 (Version v1.4). *Zenodo*, doi:10.5281/zenodo.2628942. http://doi.org/10.5281/zenodo.2628942 (Accessed April 4, 2019).

——, and E. R. Vivoni, 2016: Process-based characterization of evapotranspiration sources over the North American monsoon region. *Water Resour. Res.*, **52**, 358–384.

——, and ——, 2019a: MOD-LSP: MODIS-Based Parameters for Variable Infiltration Capacity (VIC) Model over the Continental US, Mexico, and Southern Canada (Version 1.0) [Data set]. *Zenodo*, doi:10.5281/zenodo.2612560. https://zenodo.org/record/2612560.

——, and ——, 2019b: NLCD_INEGI: Harmonized US-Mexico Land Cover Change Dataset, 1992/2001/2011 (Version 1.1) [Data set]. *Zenodo*, doi:10.5281/zenodo.2591501. https://zenodo.org/record/2591501.

——, ——, G. Mascaro, and D. D. White, 2018a: Land and water use changes in the US-Mexico Border Region, 1992-2011. *Environ. Res. Lett.*, **13**, 114005.

——, K. M. Whitney, G. Mascaro, and E. R. Vivoni, 2018b: Parameters for PITRI Precipitation Temporal Disaggregation over continental US, Mexico, and southern Canada, 1981-2013 (Version 1) [Data set]. *Zenodo*, doi:10.5281/zenodo.1402223. http://doi.org/10.5281/zenodo.1402223 (Accessed September 14, 2018).

——, ——, ——, and ——, 2019: A deterministic approach for approximating the diurnal cycle of precipitation for use in large-scale hydrological modeling. *J. Hydrometeorol.*, **20**, 297–317, doi:10.1175/JHM-D-18-0203.1.

Bowling, L. C., and D. P. Lettenmaier, 2010: Modeling the effects of lakes and wetlands on the water balance of Arctic environments. *J. Hydrometeorol.*, **11**, 276–295.

Breshears, D. D., and Coauthors, 2005: Regional vegetation die-off in response to global-change-type drought. *Proc. Natl. Acad. Sci.*, **102**, 15144–15148.

Buermann, W., Y. Wang, J. Dong, L. Zhou, X. Zeng, R. E. Dickinson, C. S. Potter, and R. B. Myneni, 2002: Analysis of a multiyear global vegetation leaf area index data set. *J. Geophys. Res. Atmospheres*, **107**, ACL-14.

Coulston, J. W., G. G. Moisen, B. T. Wilson, M. V. Finco, W. B. Cohen, and C. K. Brewer, 2012: Modeling percent tree canopy cover: a pilot study. *Photogramm. Eng. Remote Sens.*, **78**, 715–727.

FAO/UNESCO, 1998: *Digital Soil Map of the World and Derived Soil Properties*. Food and Agriculture Organization/United Nations Educational, Scientific, and Cultural Organization (FAO/UNESCO), Rome, Rome, http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/ (Accessed January 1, 2018).

Ford, T. W., and S. M. Quiring, 2013: Influence of MODIS-derived dynamic vegetation on VIC-simulated soil moisture in Oklahoma. *J. Hydrometeorol.*, **14**, 1910–1921.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang, 2010: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.*, **114**, 168–182.

Fry, J., M. Coan, C. Homer, D. Meyer, and J. Wickham, 2009: *Completion of the National Land Cover Database (NLCD) 1992-2001 land cover change retrofit product*. US Geological Survey,.

Gesch, D. B., K. L. Verdin, and S. K. Greenlee, 1999: New land surface digital elevation model covers the Earth. *EOS Trans. Am. Geophys. Union*, **80**, 69–70.

Hamman, J. J., B. Nijssen, T. J. Bohn, D. R. Gergel, and Y. Mao, 2018: The Variable Infiltration Capacity Model, Version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geosci. Model Dev.*, **11**, 3481–3496.

Hansen, M., R. DeFries, J. R. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.*, **21**, 1331–1364.

Hogue, T. S., L. Bastidas, H. Gupta, S. Sorooshian, K. Mitchell, and W. Emmerich, 2005: Evaluation and transferability of the Noah land surface model in semiarid environments. *J. Hydrometeorol.*, **6**, 68–84.

Homer, C., and Coauthors, 2007: Completion of the 2001 national land cover database for the counterminous United States. *Photogramm. Eng. Remote Sens.*, **73**, 337.

Homer, C. G., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States - Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.*, **81**, 345–354.

Hoyer, S., and J. Hamman, 2017: xarray: ND labeled Arrays and Datasets in Python. *J. Open Res. Softw.*, **5**.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, 2002: Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.*, **83**, 195–213.

INEGI, 2014: Conjunto de datos vectoriales de Uso del Suelo y Vegetación, Escala 1:250 000, Serie V (Capa Unión). http://www.inegi.org.mx/geo/contenidos/recnat/usosuelo/ (Accessed October 1, 2015).

Jiang, Z., A. R. Huete, J. Chen, Y. Chen, J. Li, G. Yan, and X. Zhang, 2006: Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sens. Environ.*, **101**, 366–378.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. Atmospheres*, **99**, 14415–14428.

Liu, M., J. C. Adam, A. S. Richey, Z. Zhu, and R. B. Myneni, 2018: Factors controlling changes in evapotranspiration, runoff, and soil moisture over the conterminous US: Accounting for vegetation dynamics. *J. Hydrol.*, **565**, 123–137.

Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Clim.*, **26**, 9384–9392.

——, T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950–2013. *Nat. Sci. Data*, **2**, 150042, doi:10.1038/sdata.2015.42.

Maurer, E., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Clim.*, **15**, 3237–3251.

Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *J. Geophys. Res. Atmospheres*, **109**.

Myneni, R., and Coauthors, 2002: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.*, **83**, 214–231.

Myneni, R. B., R. Ramakrishna, R. Nemani, and S. W. Running, 1997: Estimation of global leaf area index and absorbed PAR using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.*, **35**, 1380–1393.

Nijssen, B., R. Schnur, and D. P. Lettenmaier, 2001: Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93. *J. Clim.*, **14**, 1790–1808.

Parr, D., G. Wang, and D. Bjerklie, 2015: Integrating remote sensing data on evapotranspiration and leaf area index with hydrological modeling: Impacts on model performance and future predictions. *J. Hydrometeorol.*, **16**, 2086–2100.

Schaaf, C. B., and Coauthors, 2002: First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.*, **83**, 135–148.

Tang, Q., E. R. Vivoni, F. Muñoz-Arriola, and D. P. Lettenmaier, 2012: Predictability of evapotranspiration patterns using remotely sensed vegetation dynamics during the North American monsoon. *J. Hydrometeorol.*, **13**, 103–121.

Tesemma, Z., Y. Wei, M. Peel, and A. Western, 2015: The effect of year-to-year variability of leaf area index on Variable Infiltration Capacity model performance and simulation of runoff. *Adv. Water Resour.*, **83**, 310–322.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmospheres*, **117**.

Xian, G., C. Homer, M. Rigge, H. Shi, and D. Meyer, 2015: Characterization of shrubland ecosystem components as continuous fields in the northwest United States. *Remote Sens. Environ.*, **168**, 286–300.

Zhang, X., Q. Tang, J. Zheng, and Q. Ge, 2013: Warming/cooling effects of cropland greenness changes during 1982–2006 in the North China Plain. *Environ. Res. Lett.*, **8**, 024038.

Zhu, C., and D. P. Lettenmaier, 2007: Long-term climate and derived surface hydrology and energy flux data for Mexico: 1925–2004. *J. Clim.*, **20**, 1936–1946.

Tables

Table 1. Descriptions of spatial datasets used in this study.

Type	**Abbreviation**	**Dataset**	**Reference**	**Satellite Source**	**Acquisition Date Range**	**Temporal Interval**	**Resolution**
Land Cover Classification	UMD	University of Maryland Land Cover	(Hansen et al. 2000)	AVHRR	1981-1994	Single map	1 km
	MOD12Q1.051	MODIS Collection 5 Global Land Cover	(Friedl et al. 2010)	MODIS	2001-2013	Annual	500 m
	NLCD	National Land Cover Database	(Homer et al. 2007; Fry et al. 2009; Homer et al. 2015)	LANDSAT	1992-2011	Years 1992, 2001, 2006, 2011	30 m
	INEGI	Uso del Suelo y Vegetación	(INEGI 2014)	LANDSAT	1992-2011	Years 1993, 2002, 2007, 2011	n/a (imagery delineated by hand into polygons)
Static Surface Properties	GTOPO30	Global 30 Arc-Second Digital Elevation Model	(Gesch et al. 1999)	n/a	n/a	Single map	30” (1 km)
	NLCD-Forest	National Land Cover Database USFS Tree Cover, Cartographic	(Coulston et al. 2012)	LANDSAT	2008-2009	Single map	30 m
	NLCD-Shrub	National Land Cover Database Shrub Cover	(Xian et al. 2015)	WorldView-2, LANDSAT	2013	Single map	30 m
	FAO	FAO-UNESCO Digital Soil Map of the World	(FAO/UNESCO 1998)	n/a	n/a	Single map	5’ (8 km)
Time-Varying Surface Properties	B2002	Leaf Area Index	(Myneni et al. 1997; Buermann et al. 2002)	AVHRR	1981-1994	10 day	5’ (8 km)
	MCD15A2H.006	Leaf Area Index	(Myneni et al. 2002)	MODIS	2000-2016	8 day	500 m
	MOD13A1.006	Normalized Difference Vegetation Index	(Huete et al. 2002)	MODIS	2000-2016	16 day	500 m

	MCD43A3.006	Albedo	(Schaaf et al. 2002)	MODIS	2000-2016	1 day	500 m

Table 2. Description of various versions of VIC input parameters. All VIC parameters have 0.0625 degree (6 km) spatial resolution.

<table>
<tr><th rowspan="2">Land Cover Code
(lc_source.lc_year)</th><th rowspan="2">Domain</th><th rowspan="2">Soil Properties</th><th colspan="3">Land Cover Classification</th><th colspan="3">Phenology</th><th rowspan="2">Comment</th></tr>
<tr><th>Source</th><th>Temporal Coverage</th><th>Classification Scheme</th><th>Source</th><th>Seasonal cycle</th><th>Time Series</th></tr>
<tr><td>L2015</td><td>CONUS_MX</td><td>FAO</td><td>UMD</td><td>Single snapshot representing 1981-1994</td><td>UMD, with open water and urban replaced by bare soil</td><td>LAI: AVHRR
f_{canopy}: n/a
Albedo: n/a</td><td>Single cycle from 1992-1993</td><td>n/a</td><td>Phenology derived by spatial interpolation of values from < 50 pixels in each class</td></tr>
<tr><td>MOD_IGBP.mode</td><td>CONUS_MX</td><td>FAO</td><td>MODIS MOD12Q1.051</td><td>For each pixel, mode of class over 2001-2013</td><td>IGBP</td><td rowspan="6">LAI: MCD15A2H.006
f_{canopy}: MOD13A1.006
Albedo: MCD43A3.006 (Taken from “White-Sky Albedo for shortwave broadband”)</td><td>Monthly climatology 2000-2016</td><td>Monthly, 2000-2016</td><td></td></tr>
<tr><td>NLCD_INEGI.2011</td><td>USMX</td><td>FAO</td><td rowspan="5">NLCD over US;
INEGI over Mexico</td><td>2011</td><td>NLCD 2011</td><td>2011</td><td>n/a</td><td></td></tr>
<tr><td>NLCD_INEGI.2001</td><td>USMX</td><td>FAO</td><td>2001 over US; 2002 over Mexico</td><td>NLCD 2011</td><td>2001</td><td>n/a</td><td></td></tr>
<tr><td>NLCD_INEGI.s2011</td><td>USMX</td><td>FAO</td><td>2011</td><td>NLCD 2011</td><td rowspan="3">Monthly climatology 2000-2016 from pixels with stable class between 2001 and 2011</td><td>Monthly, 2000-2016</td><td></td></tr>
<tr><td>NLCD_INEGI.s2001</td><td>USMX</td><td>FAO</td><td>2001 over US; 2002 over Mexico</td><td>NLCD 2011</td><td>Monthly, 2000-2016</td><td></td></tr>
<tr><td>NLCD_INEGI.s1992</td><td>USMX</td><td>FAO</td><td>1992 over US; 1993 over Mexico</td><td>NLCD 2011</td><td>n/a</td><td>Over US, NLCD 1992-2001 retrofit change product applied to 2001, preserving relative proportions of shrub and grass relative to their total and of crop and pasture relative to their total</td></tr>
</table>

Table 3. Primary INEGI codes and descriptions, mappings to NLCD classes. Yellow indicates codes present in 1993 only. Green indicates codes not present in 1993.

Code	Description	NLCD ClassName[1]	NLCD ClassNum[1]	Comment
[R]	Area Agricola; Riego suspendido	Cultivated Crops	82	Present in 1993 only
ACUI	Acuicola	Cultivated Crops	82	Not present in 1993
ADV	Desprovisto de veg.	Barren Land	31	Not present in 1993
AH	Asentamientos humanos	Developed, low intensity	22	"Human Settlements"; Not present in 1993
BA	Bosque de oyamel	Needleleaf Forest	42	"Fir forest"
BB	Bosque cedro	Needleleaf Forest	42	"Cedar forest"
BC	Bosque cultivado	Broadleaf Forest	41	"Cultivated forest"; assumed to be orchards
BG	Bosque de galeria	Broadleaf Forest	41	"Riparian forest"
BI	Bosque inducido	Broadleaf Forest	41	Not present in 1993
BJ	Bosque de tascate	Needleleaf Forest	42	"Juniper forest"
BM	Bosque mesofilo de montana	Broadleaf Forest	41	"Cloud forest"; assumed predominantly broadleaf
BP	Bosque de pino	Needleleaf Forest	42	"Pine forest"
BPQ	Bosque de pino-encino	Mixed Forest	43	"Pine-Oak forest"
BQ	Bosque de encino	Broadleaf Forest	41	"Oak forest"
BQP	Bosque de Encino-pino	Mixed Forest	43	"Oak-Pine forest"
BS	Bosque de ayarin	Needleleaf Forest	42	"Spruce forest"
BW	Bosque de bajo abierto	Broadleaf Forest	41	Present in 1993 only
DV	Sin veg. aparente	Barren Land	31	
H2O	Cuerpo de agua	Open Water	11	
HA	humedad	Cultivated Crops	82	Rainfed, Annual
HAP	humedad	Cultivated Crops	82	Rainfed, Annual, permanent
HAS	humedad	Cultivated Crops	82	Rainfed, Annual, semi-permanent
HP	humedad	Cultivated Crops	82	Rainfed, Permanent
HS	humedad	Cultivated Crops	82	Rainfed, Semi-permanent
HSP	humedad	Cultivated Crops	82	Rainfed, Semi-permanent/ permanent
MC	Matorral crasicuale	Shrub/Scrub	52	
MDM	Matorral desertico microfilo	Shrub/Scrub	52	
MDR	Matorral desertico rosetofilo	Shrub/Scrub	52	
MET	Matorral espinoso tamauli-peco	Shrub/Scrub	52	
MJ	Matorral de coniferas	Shrub/Scrub	52	Present in 1993 only
MK	mezquital	Shrub/Scrub	52	
MKE	Mezquital tropical	Shrub/Scrub	52	Not present in 1993
MKX	Mezquital xerofilo	Shrub/Scrub	52	Not present in 1993
ML	chaparral	Shrub/Scrub	52	
MRC	Matorral rosetofilo costero	Shrub/Scrub	52	
MSC	Matorral sarcocuale	Shrub/Scrub	52	

MSCC	Matorral sarco-crasicuale	Shrub/Scrub	52	
MSM	Matorral sub-montano	Shrub/Scrub	52	
MSN	Matorral sarco-crasicuale de neblina	Shrub/Scrub	52	
MST	Matorral subtropical	Shrub/Scrub	52	
MU	huizachal	Shrub/Scrub	52	Present in 1993 only
P/E	Pais extranjero	(nodata)	-1	Outside the boundaries of Mexico
PA	Pastizal - huizachal	Grassland/Herbaceous	71	Present in 1993 only
PC	Pastizal cultivado	Pasture/Hay	81	
PH	Pastizal halofilo	Grassland/Herbaceous	71	Assumed to be natural salt-tolerant grassland
PI	Pastizal inducido	Pasture/Hay	81	
PN	Pastizal natural	Grassland/Herbaceous	71	
PT	Veg. de Peten	Woody Wetland	90	Not present in 1993 (was "manglar")
PY	Pastizal gipsofilo	Grassland/Herbaceous	71	
RA	riego	Cultivated Crops	82	Irrigated, Annual
RAP	riego	Cultivated Crops	82	Irrigated, Annual/ permanent
RAS	riego	Cultivated Crops	82	Irrigated, Annual/ semi-permanent
RP	riego	Cultivated Crops	82	Irrigated, Permanent
RS	riego	Cultivated Crops	82	Irrigated, Semi-permanent
RSP	riego	Cultivated Crops	82	Irrigated, Semi-permanent/ permanent
SAP	Selva alta perenni-folia	Broadleaf Forest	41	
SAQ	Selva alta sub-perenni-folia	Broadleaf Forest	41	
SBC	Selva baja caducifolia	Broadleaf Forest	41	
SBK	Selva baja espinosa	Broadleaf Forest	41	
SBP	Selva baja perreni-folia	Broadleaf Forest	41	
SBQ	Selva baja sub-perreni-folia	Broadleaf Forest	41	
SBQP	Selva baja sub-perreni-folia	Broadleaf Forest	41	
SBS	Selva baja subcaduci-folia	Broadleaf Forest	41	
SG	Selva de galeria	Broadleaf Forest	41	"Riparian forest"
SMC	Selva mediana caduci-folia	Broadleaf Forest	41	
SMP	Selva mediana perenni-folia	Broadleaf Forest	41	
SMQ	Selva mediana sub-caducifolia	Broadleaf Forest	41	
SMS	Selva mediana subcaduci-folia	Broadleaf Forest	41	
TA	Temporal	Cultivated Crops	82	Rainfed, Annual
TAP	Temporal	Cultivated Crops	82	Rainfed. Annual/ permanent
TAS	Temporal	Cultivated Crops	82	Rainfed, Annual/ permanent
TP	Temporal	Cultivated Crops	82	Rainfed, permanent
TS	Temporal	Cultivated Crops	82	Rainfed, permanent

TSP	Temporal	Cultivated Crops	82	Rainfed, permanent
VA	Popal	Emergent Herbaceous Wetland	95	Emergent veg – herbaceous (lily/water hyacinth)
VD	Veg. de desiertos arenosos	Barren Land	31	VD areas tend to be classified as “bare” in MODIS and NLCD classifications
VG	Veg. de galeria	Woody Wetland	90	“Riparian vegetation”; assume that there are shrubs and trees
VH	Veg. halofila	Emergent Herbaceous Wetland	95	Salt marsh (halophyllic veg)
VHH	Veg. halofila hidrofila	Emergent Herbaceous Wetland	95	Not present in 1993
VM	Manglar	Woody Wetland	90	“Mangrove swamp”
VP	Palmar	Broadleaf Forest	41	“Palm”; Present in 1993 only
VPI	Palmar Inducido	Broadleaf Forest	41	“Palm”; Not present in 1993
VPN	Palmar natural	Broadleaf Forest	41	“Palm”; Not present in 1993
VS	Sabana	Broadleaf Forest	41	“savanna”
VSI	Sabanoide	Broadleaf Forest	41	Assumed similar to savanna; Not present in 1993
VT	Tular	Emergent Herbaceous Wetland	95	Cattails, reeds, etc
VU	Veg. de las dunas costeras	Shrub/Scrub	52	
VW	Pradera de alta montana	Grassland/Herbaceous	71	
VY	Veg. gipsofila	Shrub/Scrub	52	
ZU	Zona urbana	Developed, High intensity	24	

[1]The NLCD classes of “Deciduous Forest” and “Evergreen Forest” (codes 41 and 42) were interpreted to correspond to broadleaf and needleleaf forest, respectively, in Mexico. Over the CONUS_MX domain, Deciduous Needleleaf Forest is extremely rare, so that the NLCD Evergreen Forest class essentially denotes Needleleaf Forest. Similarly, in the UMD-NLDAS dataset, the time-invariant structural parameters for Deciduous Broadleaf Forest were the same as those for Evergreen Broadleaf Forest, so that our interpretation did not adversely impact parameter values.

Table 4. Mapping between MOD-IGBP classes and UMD-NLDAS classes from which structural parameters were taken.

MOD-IGBP		UMD-NLDAS	
Code	**Names**	**Code**	**Name**
0	Open Water	10	Grassland
1	Evergreen Needleleaf Forest	1	Evergreen Needleleaf Forest
2	Evergreen Broadleaf Forest	3	Evergreen Broadleaf Forest
3	Deciduous Needleleaf Forest	2	Deciduous Needleleaf Forest
4	Deciduous Broadleaf Forest	4	Deciduous Broadleaf Forest
5	Mixed Forest	5	Mixed Forest
6	Closed Shrubland	8	Closed Shrubland
7	Open Shrubland	9	Open Shrubland
8	Woody Savanna	6	Woodland
9	Savanna	7	Savanna
10	Grassland	10	Grassland
11	Wetland	10	Grassland
12	Cropland	11	Cropland
13	Urban	10	Grassland
14	Crop-Natural Mosaic	11	Cropland
15	Perennial Ice/Snow	10	Grassland
16	Barren	10	Grassland

Table 5. Mapping between NLCD-INEGI classes and UMD-NLDAS classes from which structural parameters were taken.

NLCD-INEGI		UMD-NLDAS	
Code	**Names**	**Code**	**Name**
11	Open Water	10	Grassland
12	Perennial Ice/Snow	10	Grassland
21	Developed, Open Space	10	Grassland
22	Developed, Low Intensity	10	Grassland
23	Developed, Medium Intensity	10	Grassland
24	Developed, High Intensity	10	Grassland
31	Barren Land	10	Grassland
41	Deciduous Forest[1]	4	Deciduous Broadleaf Forest
42	Evergreen Forest[1]	1	Evergreen Needleleaf Forest
43	Mixed Forest	5	Mixed Forest
52	Shrub/Scrub	9	Open Shrubland
71	Grassland/Herbaceous	10	Grassland
81	Pasture and Hay	11	Cropland
82	Cultivated Crops	11	Cropland
90	Woody Wetlands	10	Grassland
95	Emergent Herbaceous Wetlands	10	Grassland

[1]The NLCD classes of "Deciduous Forest" and "Evergreen Forest" (codes 41 and 42) were interpreted to correspond to broadleaf and needleleaf forest, respectively, in Mexico. Over the CONUS_MX domain, Deciduous Needleleaf Forest is extremely rare, so that the NLCD Evergreen Forest class essentially denotes Needleleaf Forest. Similarly, in the UMD-NLDAS dataset, the time-invariant structural parameters for Deciduous Broadleaf Forest were the same as those for Evergreen Broadleaf Forest, so that our interpretation did not adversely impact parameter values.

Table 6. Statistics of L2015 and MOD_LSP (NLCD_INEGI.2011) July f_{canopy} values relative to NLCD-Shrub (Xian et al. 2015) and NLCD-Forest (Coulston et al. 2012) products. The subscript "ref" refers to NLCD-Shrub and NLCD-Forest.

Land Cover Category	**Parameter Set**	**Statistics**			
		μ	**μ / μ_{ref}**	**Pearson r**	**RMSE**
Shrub	L2015	1.0000	7.2368	0.0000	0.8658
	NLCD_INEGI.2011	0.0962	0.6964	0.3643	0.0940
Forest	L2015	1.0000	2.2810	0.0000	0.6120
	NLCD_INEGI.2011	0.5458	1.2449	0.6893	0.2064

Table 7. Files included in this dataset.

Name	**Description**
global_param.template	Template for global parameter file for VIC model

params.*domain*.*lc_source*.*lc_year*.*phen_years*.nc	VIC soil and vegetation parameters. In these files, phenology data (LAI, f_{canopy}, albedo) for each class in each grid cell consist of a single set of 12 climatological monthly mean values. File naming convention is as follows:
	• *domain* = domain name ("CONUS_MX" or "USMX")
	• *lc_source* = land cover source name ("L2015", "MOD_IGBP", or "NLCD_INEGI")
	• *lc_year* = code describing the year(s) of acquisition from which pixels are assigned classes; can be one of ("mode", *year*, or "s"*year*) where:
	    o *mode* = class is the most common class from all maps over the period 2000-2013
	    o *year* = class comes from the land cover map for year *year*
	    o "s"*year* = same as *year*, but the "s" signifies that the phenology data were only aggregated over pixels for which land cover class was stable (did not change) between 2001 and 2011. Pixels for which land cover class changed were assigned phenology values via spatial interpolation from stable neighbors of the class to which the pixel belonged in the given *lc_year*.
	• *phen_years* = years of MODIS products from which phenology values are derived, in the format *startyear_endyear*.

veg_hist.*domain*.*lc_source*.*lc_year*.*phen_years*.monthly.tgz	Gzipped tar file of directory containing yearly files (one file per year) of time-varying monthly phenology. Individual files follow the naming convention: veg_hist.*domain*.*lc_source*.*lc_year*.*phen_years*.monthly.*year*.nc where • *domain*.*lc_source*.*lc_year*.*phen_years* = same naming convention as for the params files • *year* = year of phenology to which this file pertains Note: these files must be converted from monthly values to sub-daily values (same time step as the meteorological forcings) via the following scripts, available at https://github.com/tbohn/VIC_Landcover_MODIS_NLCD_INEGI/releases/tag/v1.4: • disagg_veghist_monthly2hourly_nc.py • wrap_disagg_veghist_monthly2hourly_nc.pl

Table 8. Explanation of placeholders in global_param.template file.

Option	Description	Default
<DOMAIN_FILE>	Path/name of the domain file. The land mask in the domain file determines the spatial domain of the simulation. The spatial dimensions (resolution, numbers of rows and columns, latitudes and longitudes of grid cell centers) must match those in the desired VIC parameter file exactly. Domain files covering the CONUS_MX and USMX domains are available on Zenodo at (https://zenodo.org/record/2564019)(Bohn et al. 2018b).	None.
<PARAM_FILE>	Path/name of the desired parameter file.	None.
<LAI_SRC>, <FCAN_SRC>, <ALB_SRC>	Options to tell the VIC model where to obtain phenology variables from. To use the repeating annual cycles of monthly values stored in the parameter file, these should be set to FROM_VEGPARAM. To use the monthly time series of phenology from the veg_hist files, these should be set to FROM_VEGHIST (see below for instructions on preparing the veg_hist files). Because the L2015 parameter file does not contain f_{canopy} values and has no associated veg_hist file, <FCAN_SRC> must be set to FROM_DEFAULT when using the L2015 parameter file (which instructs VIC to use a value of 1.0 everywhere).	<FCAN_SRC>: FROM_DEFAULT All others: FROM_VEGPARAM
<FORCING_DIR>	Path of the directory containing meteorological forcing files.	None.
<FORCING_PFX>	Prefix of meteorological forcing files. Forcing files are assumed to be named as *forcing_pfx*.YYYY.nc, where YYYY = the year covered by the file.	None.
<VEGHIST_DIR>	Path of the directory containing veg_hist files.	Use repeating cycle from parameter file.
<VEGHIST_PFX>	Prefix of veg_hist files. Veg_hist files are assumed to be named as *veghist_pfx*.YYYY.nc, where YYYY = the year covered by the file.	Use repeating cycle from parameter file.

Table 9. Definitions of selected variables in the params and veg_hist files.

Name	**Definition**
veg_descr	Land cover class name
veg_class	Class ID code
mask	Land mask. The MOD-LSP parameter files have a slightly different mask than the L2015 file, primarily along coastlines, due to the higher-resolution processing of the DEM in the MOD-LSP processing.
run_cell	Mask determining which cells to run in the VIC simulation. Run_cell is currently set to equal the L2015 mask in the MOD_IGBP parameter file, to allow it to be used with the L2015 meteorological forcings. Similarly, the NLCD_INEGI parameter files set run_cell to the intersection of the L2015 and USMX land masks. To use these parameter files with a different set of meteorological forcings, run_cell must be modified to be consistent with the land mask of those forcings.
LAI	Monthly Leaf Area Index
fcanopy	Monthly f_{canopy}
albedo	Monthly albedo

Figures

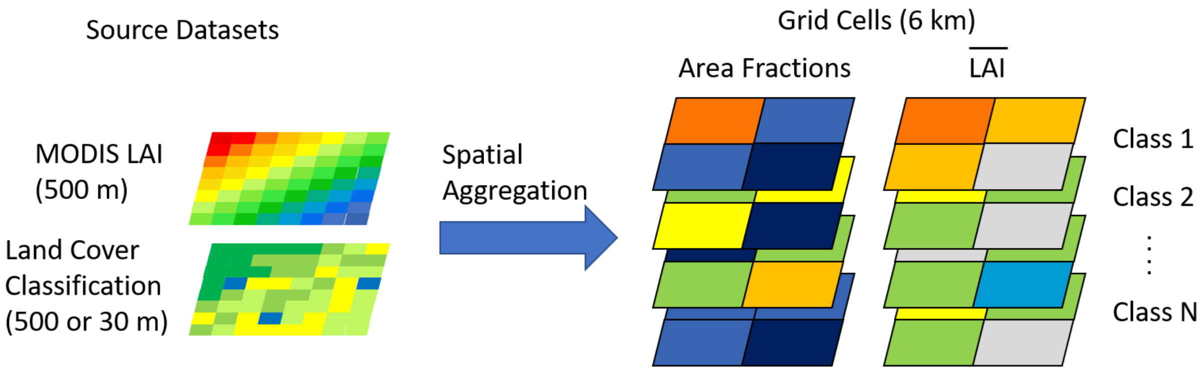

Fig. 1. Example process flow for LAI, showing spatial aggregation of MODIS LAI over a land cover classification to generate spatial average values for each land cover class at 6 km resolution.

Fig. 2. Spatial extents of the USMX (dark gray region) and CONUS_MX (union of dark and light gray regions) domains.

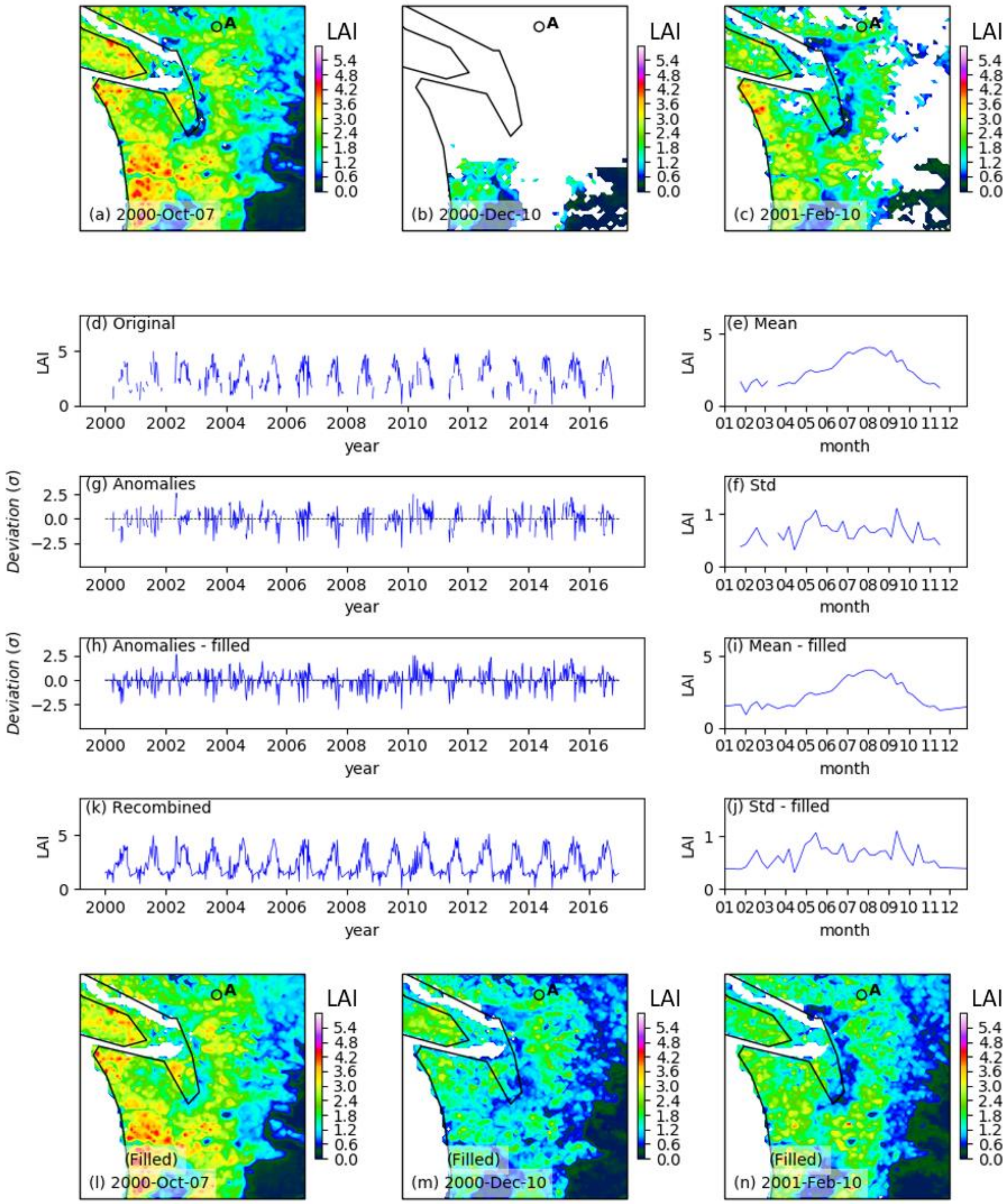

Fig. 3. Demonstration of gap filling process using LAI over the geographic region bounded by 45-50° N latitude and 120-125° W longitude. (a-c) Maps of 8-day MOD15A2H.006 LAI aggregated over land cover classes from the MOD12Q1.051 IGBP land cover classification at 0.0625° resolution, for days 2000-Oct-07, 2000-Dec-10, and 2001-Feb-10. Gaps are present due to low solar angle, snow, and clouds. (d) Time series of LAI for mixed forest, from the grid cell denoted by “A” in panels (a-c). (e-f) Climatological mean and standard deviation of LAI for each

8-day interval over the period 2000-2016. (g) Anomalies of LAI relative to the climatological mean, in units of numbers of standard deviations from the mean. (h-j) Climatological mean, standard deviation, and anomalies after gap filling. (k) Gap-filled time series created by recombining gap-filled mean, standard deviation, and anomalies of panels (h-j). (l-n) Gap-filled versions of the maps in panels (a-c).

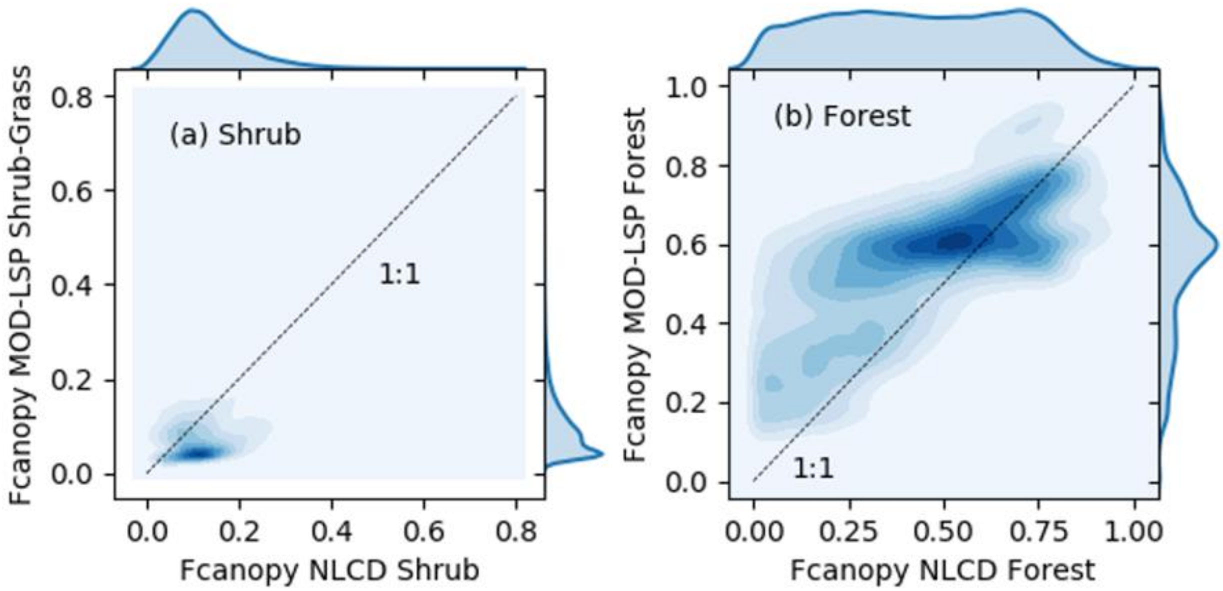

Fig. 4. Comparison of MOD-LSP f_{canopy} from NLCD_INEGI.2011.2000_2016 parameters to NLCD canopy cover products aggregated to 0.0625° (6 km) resolution. (a) MOD-LSP f_{canopy} from shrubland and grassland classes vs. NLCD-Shrub (Xian et al. 2015); (b) MOD-LSP f_{canopy} from all forest classes vs NLCD-Forest product (Coulston et al. 2012).

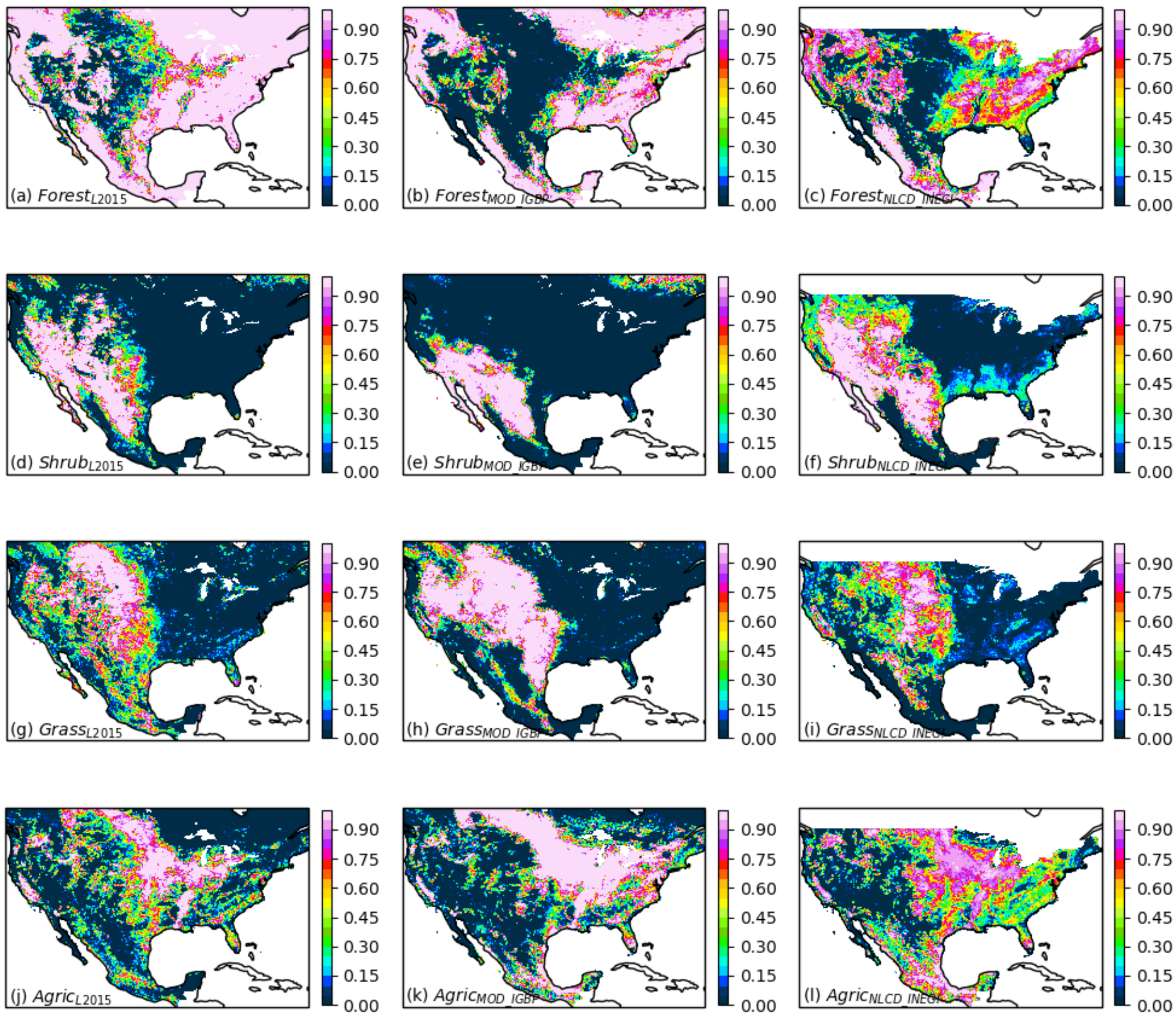

Fig. 5. Comparison of geographic distributions of broad land cover categories in the L2015, MOD_IGBP, and NLCD_INEGI.2011 parameter sets. Maps show the total area coverage fraction of (a-c) all forest classes; (d-f) all shrub classes; (g-i) all grassland classes; (j-l) all agricultural and pastoral classes.

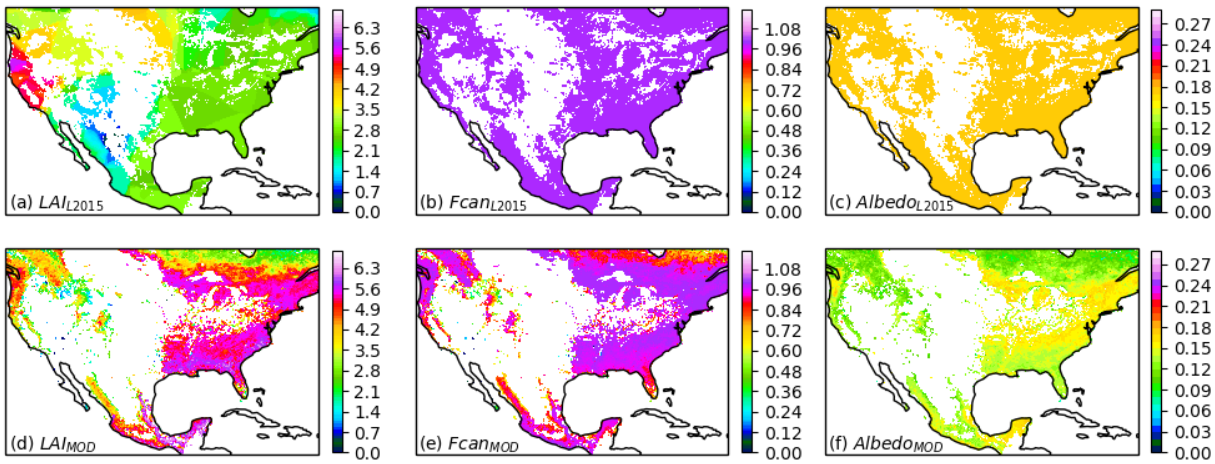

Fig. 6. Comparison of climatological average July phenology between L2015 and MOD-LSP MOD_IGBP datasets for the mixed forest class. (a-c) LAI, f_{canopy}, and albedo from L2015; (d-f) LAI, f_{canopy}, and albedo from MOD-LSP MOD_IGBP.

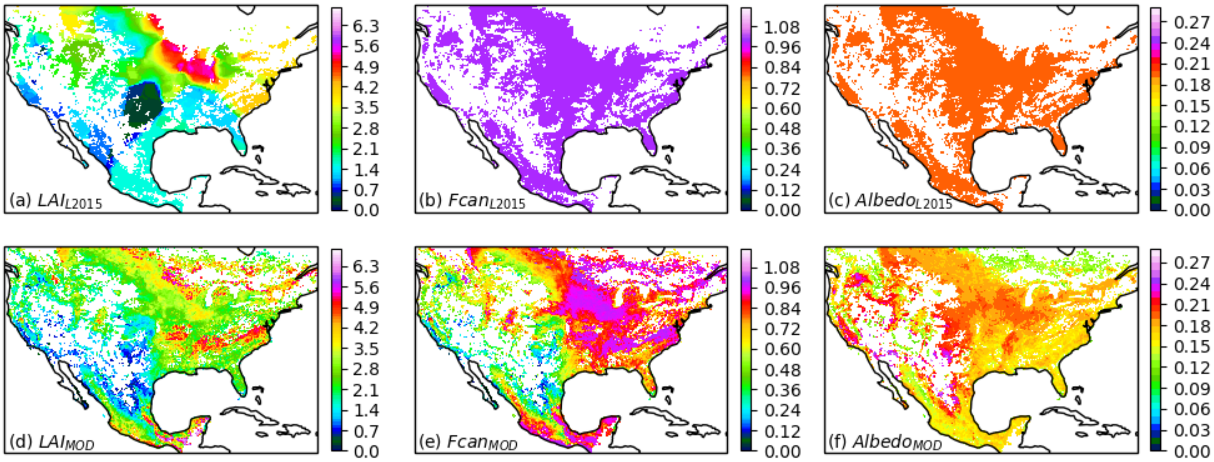

Fig. 7. Comparison of climatological average July phenology between L2015 and MOD-LSP MOD_IGBP datasets for the cropland class. (a-c) LAI, f_{canopy}, and albedo from L2015; (d-f) LAI, f_{canopy}, and albedo from MOD-LSP MOD_IGBP.

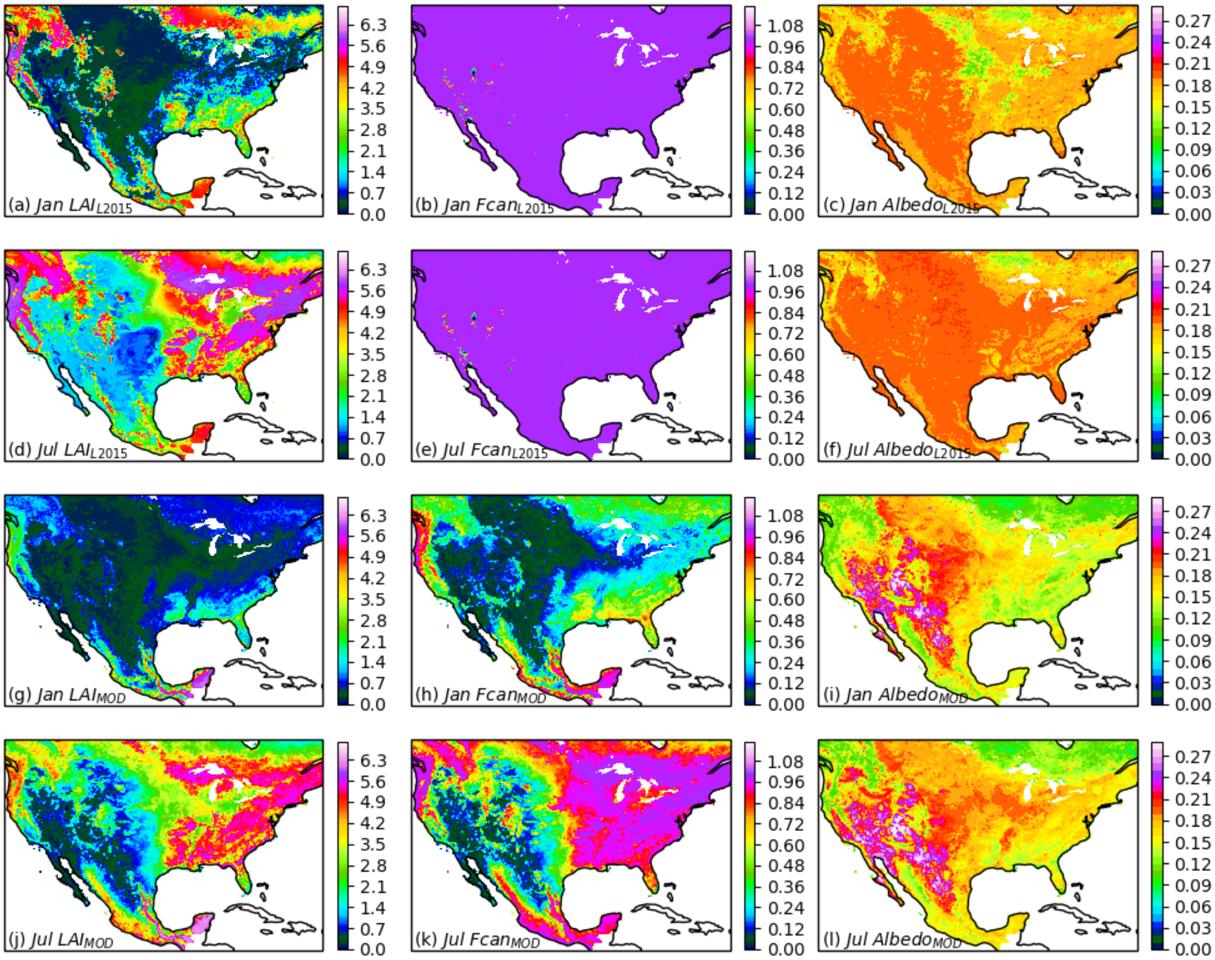

Fig. 8. Comparison of climatological average January and July phenology between L2015 and MOD-LSP MOD_IGBP datasets for the area-weighted average across all land cover classes. (a-f) LAI, f_{canopy}, and albedo from L2015, for (a-c) January and (d-f) July; (g-l) LAI, f_{canopy}, and albedo from MOD-LSP MOD_IGBP, for (g-i) January and (j-l) July.

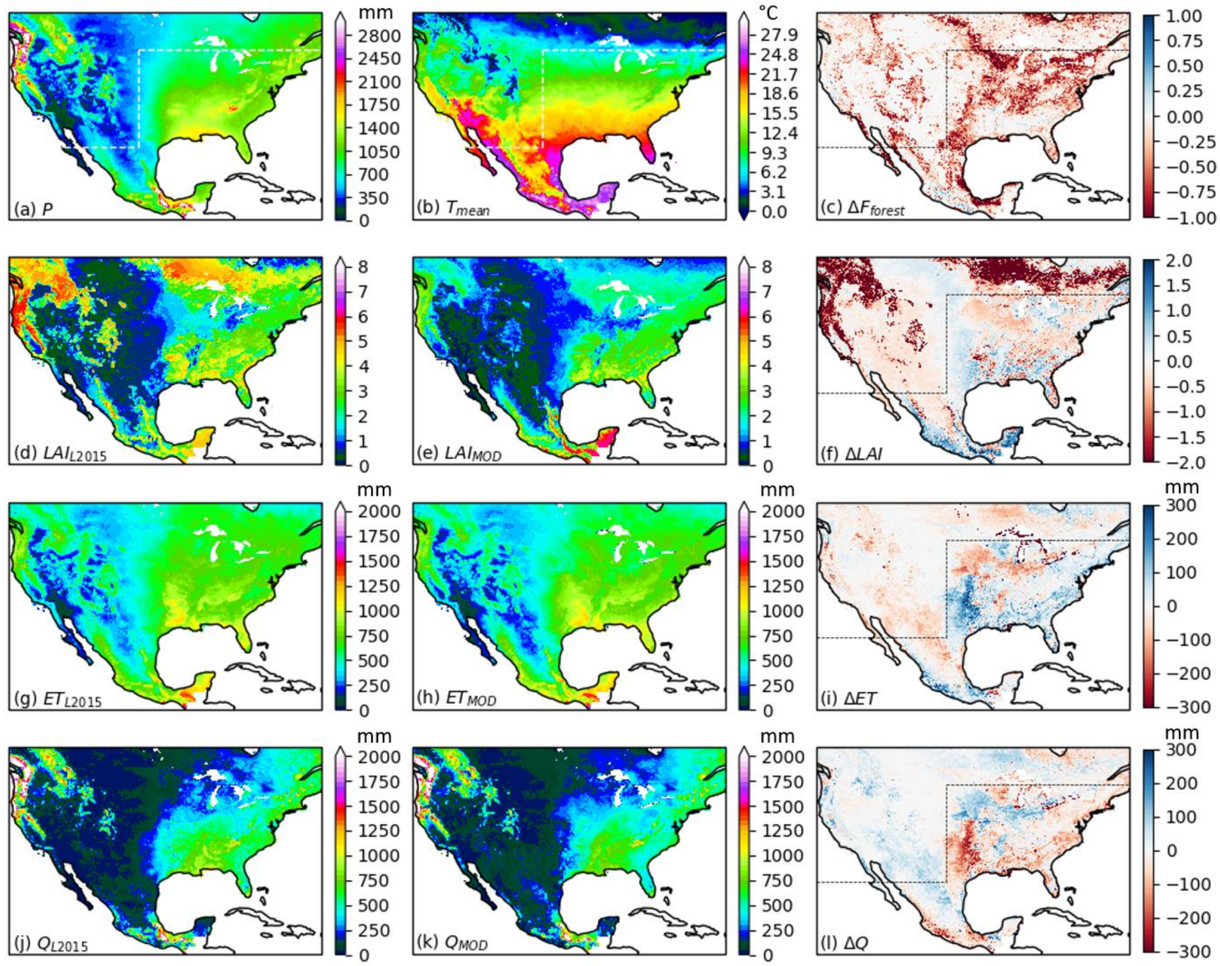

Fig. 9. Comparison of annual hydrologic terms between simulations using the L2015 and MOD-LSP MOD_IGBP datasets over the period 1981-2013 (for f_{canopy} uniformly held to 1.0). (a-b) Mean annual P and T from L2015 meteorological forcings; (c) difference in forest area fraction between L2015 and MOD-LSP MOD_IGBP; (d-f) mean annual LAI from L2015 and MOD_IGBP simulations and their difference; (g-i) mean annual ET from L2015 and MOD_IGBP simulations and their difference; (j-l) mean annual Q from L2015 and MOD_IGBP simulations and their difference. Dashed lines denote boundary between cold/dry climates and warm/wet climates.

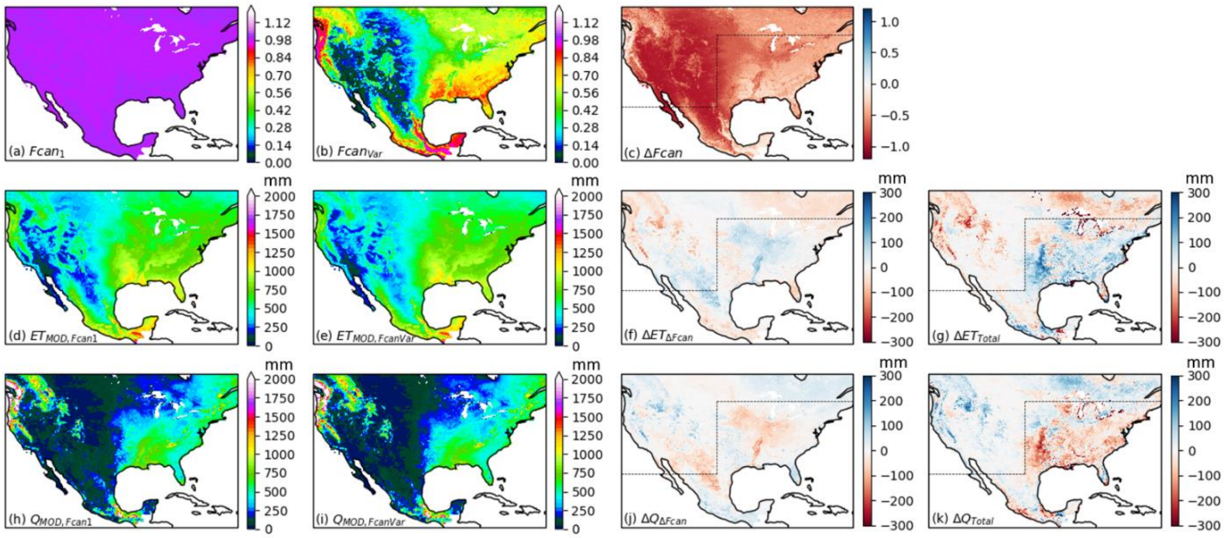

Fig. 10. Comparison of annual hydrologic terms between simulations with uniform and spatially-varying f_{canopy}, using the MOD-LSP MOD_IGBP dataset over the period 1981-2013. (a-c) Uniform f_{canopy}, mean annual spatially-varying f_{canopy}, and their difference; (d-f) mean annual ET from uniform and spatially-varying f_{canopy} simulations and their difference; (h-j) mean annual Q from uniform and spatially-varying f_{canopy} simulations and their difference; (g) and (k) total difference in (g) ET and (k) Q between MOD-IGBP spatially-varying f_{canopy} and L2015 uniform f_{canopy} simulations. Dashed lines denote boundary between cold/dry climates and warm/wet climates.

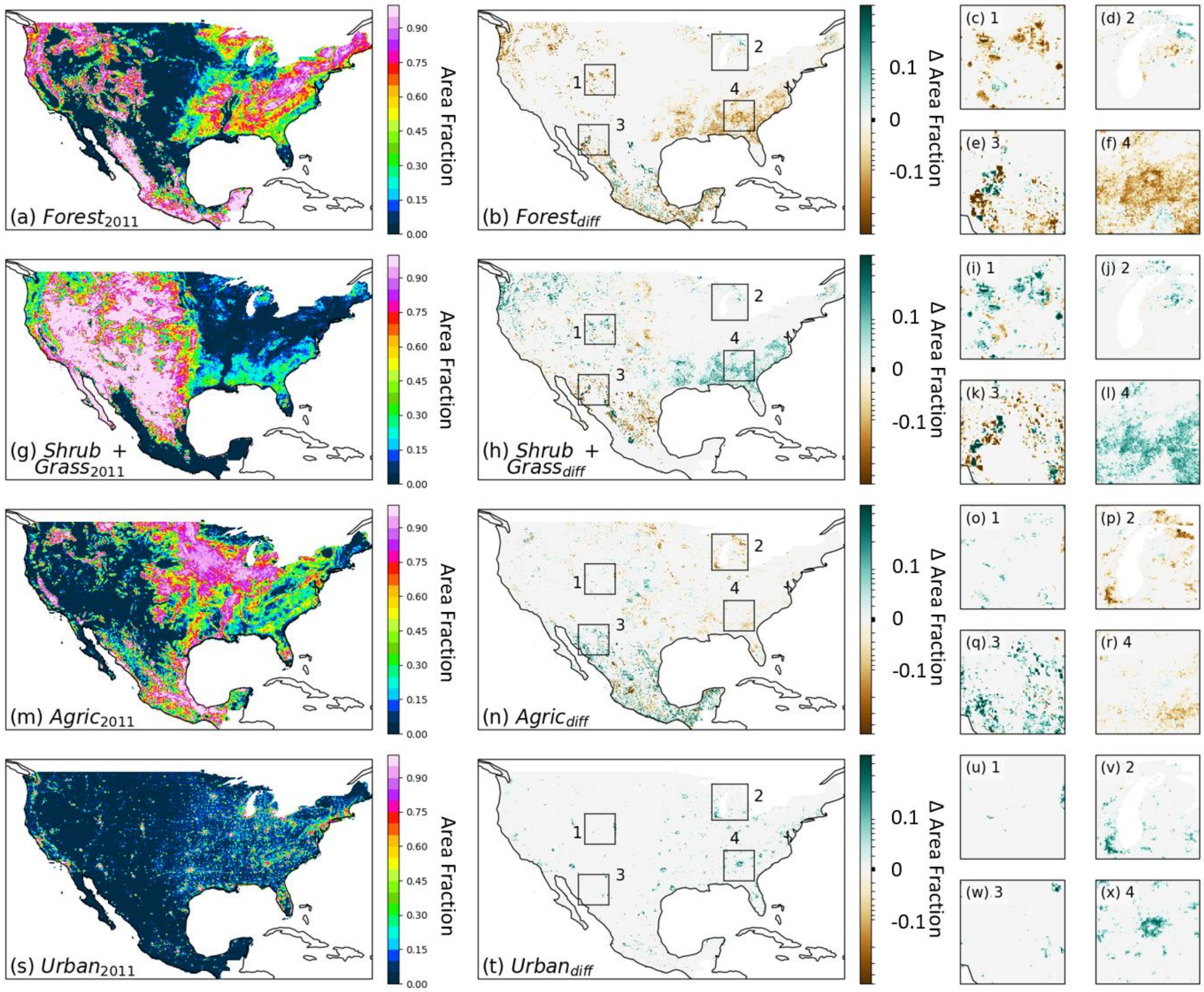

Fig. 11. Changes in the geographic distributions of broad land cover classes between 1992 and 2011 in the NLCD_INEGI parameter sets. Panel (a) Distribution of all forest classes in 2011; (b) change in distribution of forest between 1992 and 2011; (c-f) magnification of boxes 1-4 in panel (b). Panels (g-l) same as (a-f) for combined areas of shrubland and grassland classes. Panels (m-r) same as (a-f) for combined areas of agricultural and pastoral classes. Panels (s-x) same as (a-f) for combined areas of urban classes.

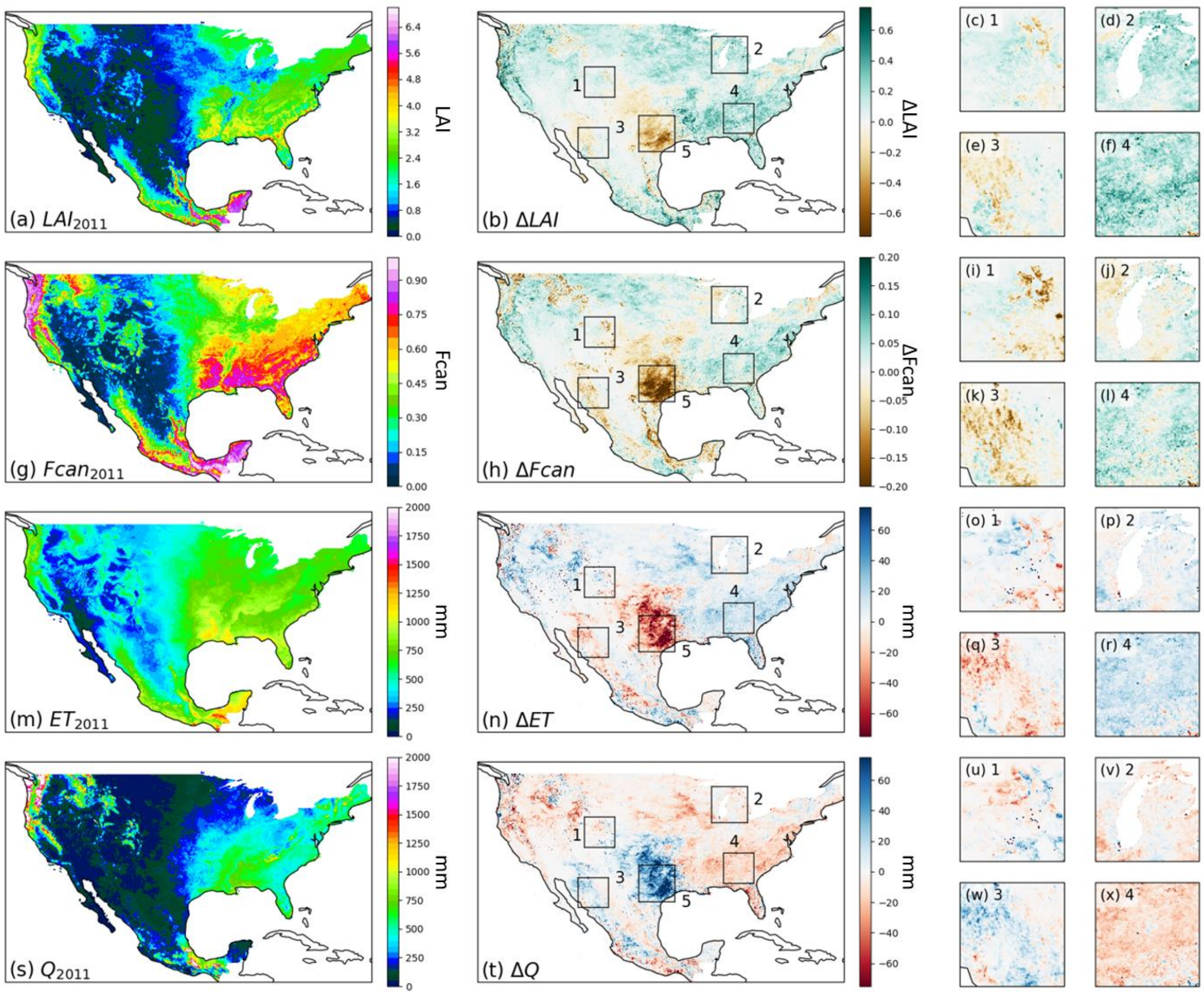

Fig. 12. Changes in the phenology and annual hydrologic fluxes between 2001 and 2011 in the NLCD_INEGI.2001.2001_2001 and NLCD_INEGI.2011.2011_2011 parameter sets. Panel (a) mean annual LAI in 2011; (b) change in mean annual LAI between 2001 and 2011; (c-f) magnification of boxes 1-4 in panel (b). Panels (g-l) same as (a-f) for mean annual f_{canopy}. Panels (m-r) same as (a-f) for mean annual ET. Panels (s-x) same as (a-f) for mean annual Q.