A river runs through it: Robust automated mapping of riparian woodlands and land surface phenology across dryland regions

Conor A. McMahon [a,*], Dar A. Roberts [a,c], John C. Stella [b], Anna T. Trugman [a,c], Michael B. Singer [c,d,e], Kelly K. Caylor [a,c]

[a] Department of Geography, University of California, Santa Barbara, CA 93106, United States of America
[b] Department of Sustainable Resources Management, State University of New York College of Environmental Science and Forestry, Syracuse, NY 13210, United States of America
[c] Earth Research Institute, University of California, Santa Barbara, CA 93106, United States of America
[d] School of Earth and Environmental Sciences, Cardiff University, Cardiff CF10 3AT, United Kingdom
[e] Water Research Institute, Cardiff University, Cardiff CF10 3AX, United Kingdom

ARTICLE INFO

ABSTRACT

Riparian woodlands in drylands are critically important to human society, global biodiversity, and regional water and energy budgets. These sensitive ecosystems have experienced substantial degradation over the last several decades from climatic change and direct human activity. Nevertheless, quantifying long-term change in dryland riparian woodlands remains a major challenge, and much uncertainty exists in their remaining extent, historical breadth, and likely future trajectories. Dryland landscapes show large, fine-scale spatial heterogeneity in seasonal greenness patterns, driven in part by spatial variation in water availability. Riparian woodlands occur where water is concentrated in the landscape, either as aboveground streamflow or subsurface groundwater. In arid and semi-arid climates, this renders them phenologically distinctive from upland ecosystems. However, despite their importance and distinctiveness, there are currently no automated methods for delineating dryland riparian woodlands across regional extents in the cloud. Here we designed and implemented a cloud-based algorithm to retrieve dryland land surface phenology patterns from multispectral satellite imagery and conducted sensitivity analyses using real and simulated data to demonstrate that the approach is robust for MODIS, Sentinel-2, and Landsat over realistic ranges of noise and cloud cover. We then designed a series of random forest vegetation classifiers that integrate phenological and spectral information, vegetative structure from LiDAR, and topography from LiDAR or the Shuttle Radar Topography Mission. We implemented classifiers for three local study sites and then generalized our model to run regionally across the southwestern United States, with balanced accuracy for the riparian woodland class ranging from 94.5% to 97.5% when validated with local to regional datasets. Generally, phenological information proved more important than any other data source for mapping riparian woodlands, which showed more stability in interannual phenology than did upland vegetation types. To our knowledge, ours is the first regional, annual, automatically-generated and updated approach for mapping dryland riparian woodlands in the southwestern United States, paving the way for improved modeling and management efforts on watershed to regional scales. We also provide one of the first operational, exclusively cloud-based methods to extract dryland land surface phenology patterns using Landsat, Sentinel-2, MODIS, or other sensors, providing a framework for future studies investigating other aspects of long-term or spatial variation in dryland vegetative seasonality across the globe.

1. Introduction

Drylands are defined as regions with relatively high evaporative demand in comparison to water supply. For example, the United Nations defines drylands as regions where the aridity index, or the ratio of precipitation to evaporative demand, is <0.65 (Safriel et al., 2005). In drylands, riparian woodlands cover a small percentage of land area (Sands, 1980; Katibah, 1984; Swift, 1984; Salo et al., 2016) but are

* Corresponding author.
E-mail address: conormcmahon@ucsb.edu (C.A. McMahon).

responsible for a disproportionately large fraction of landscape-scale primary productivity and evapotranspiration (Jenerette et al., 2009; Swetnam et al., 2017; Ma et al., 2018; Doody et al., 2023). Riparian woodlands also support numerous ecosystem services including water filtration, nutrient cycling, local microclimate regulation, and recreation, and they provide habitat to many threatened and endangered species (Ballard et al., 2004; Ranalli and Macalady, 2010; Gunawardena et al., 2017; Palmer and Ruhi, 2019; Nagler et al., 2020; Nagler et al., 2021). At the same time, these ecosystems are vulnerable to climate change, groundwater extraction and damming, invasive species, and land-use change (Rood et al., 2003; Jenerette et al., 2009; Stella et al., 2013; Ma et al., 2018; Albano et al., 2020; Mayes et al., 2020; Nagler et al., 2020; Nagler et al., 2021; Nagler, 2022b). Dryland riparian woodlands have already declined to a small fraction of their historical extent (Katibah, 1984; Swift, 1984; Salo et al., 2016; Albano et al., 2020; Nagler et al., 2020; Nagler et al., 2021). Mapping their remaining area and their spatial distribution through time is important for conservation but is challenging because riparian woodlands are complex and diverse, temporally dynamic (with regular flood disturbance), and narrow in spatial extent.

Although considerable amounts of time and funding are spent managing riparian woodlands and their associated fauna (including numerous threatened and endangered species), most riparian woodlands are mapped manually through some combination of field work and/or interpretation of aerial imagery. For example, more than a dozen rivers in the Lower Colorado River Basin in Arizona have had riparian extents manually digitized (Nagler, 2022b). Manual approaches can provide species-level information, high accuracy, and spatial precision. These maps can be used alongside time series data for monitoring ecosystem process, including evapotranspiration change. However, manual delineation is extremely laborious, which limits coverage and typically precludes regular updates or incorporation of change in woodland extent over time. Substantial effort has been made to automate riparian woodland delineation using manually delineated datasets as training alongside various remotely sensed data (Rusnák et al., 2022), including multispectral satellite imagery (Townsend and Walsh, 2001; Dennison et al., 2009; Johansen et al., 2010; Jia et al., 2020; Melichar et al., 2023; Rabanaque et al., 2021), fusion of aerial and multispectral imagery (Nagler et al., 2022a), hyperspectral aerial imagery (Godfrey et al., 2023), digital elevation models and LiDAR (Johansen et al., 2010; Salo et al., 2016; Rabanaque et al., 2021; Godfrey et al., 2023), unmanned aerial systems (Dunford et al., 2009), and collaborative frameworks involving a combination of remote sensing data sources and expert stakeholders (Doody et al., 2017). In application, these efforts are limited mainly by the spatial and temporal availability of LiDAR, hyperspectral imagery, and drone surveys, or by the need to continually re-train multispectral satellite classifiers in new regions with different vegetation types. No general approach currently exists to automatically and annually map dryland riparian woodlands across regional or greater extents using widespread and frequently updated imagery (Rusnák et al., 2022).

Across global drylands, water availability is a major seasonal constraint on primary productivity (Huxman et al., 2004; Jenerette et al., 2009; Ma et al., 2018). Because of this, vegetative leaf phenology and the timing of precipitation are also related, particularly in upland settings (Smith et al., 2019; Albano et al., 2020; QGIS Geographic Information System, 2021; Warter et al., 2021, 2023; Currier and Sala, 2022; Melichar et al., 2023). Some upland plant species exhibit drought avoidance strategies with perennial summer-deciduous or short-lived annual life histories, matching their growth seasons to water availability (Huxman et al., 2004). Other dryland trees and shrubs are evergreen and resist water scarcity via strong stomatal regulation, highly resistant xylem, or deep roots to access persistent water in deeper soil layers (Cooper, 1922; Stromberg, 1993; Smith et al., 2019; Manning et al., 2020; Maestre et al., 2021).

By contrast, well-watered streams usually have shallow

groundwater, which is constantly available to phreatophytic plants. This encourages establishment of trees adapted primarily for light and temperature limitation, rather than water limitation (Doody et al., 2023). In drylands, dense woodlands of winter-deciduous broadleaved trees – with short-lived leaves, low water use efficiency, and vulnerable xylem, but fast growth and high efficiency in light use – are therefore primarily limited to riparian environments and other groundwater-dependent ecosystems (Rood et al., 2003; Stromberg and Merritt, 2015; Manning et al., 2020; Melichar et al., 2023). The tree species characteristic of this habitat type are both vital for a suite of threatened and endangered species (Ballard et al., 2004; Hatten et al., 2010; Johnson et al., 2017) and are especially vulnerable to loss of streamflow or groundwater under human mediated hydrologic and climatic change (Rood et al., 2003; Sands, 1980; Stromberg, 1993; Stromberg and Merritt, 2015; Albano et al., 2020; Nagler et al., 2020; Nagler et al., 2021; Manning et al., 2020; Mayes et al., 2020; QGIS Geographic Information System, 2021; QGIS Geographic Information System, 2021; Melichar et al., 2023).

Multispectral satellite imagery can be used to estimate phenology patterns on landscape scales, allowing identification of plants by phenological type (Jönsson and Eklundh, 2004; Verbesselt et al., 2010; Smith et al., 2019; Nietupski et al., 2021; Melichar et al., 2023). Most existing approaches involve collecting all imagery from one or several years, evaluating a vegetation index such as the normalized difference vegetation index (NDVI; Rouse et al., 1973), filtering to remove cloud-contaminated values, and then fitting a parametric function to the remaining data. Popular functions include double logistic (Nietupski et al., 2021) and piecewise polynomial or Fourier (harmonic) fits (Verbesselt et al., 2010; Brooks et al., 2012; Xin et al., 2013; Zhou et al., 2023). Most phenology retrieval algorithms are designed to run offline on downloaded data, rather than in the cloud, which limits their applicability over large areas (e.g. Melichar et al., 2023). Additionally, many authors have focused on the Moderate Resolution Imaging Spectroradiometer (MODIS) or other sensors with very frequent return intervals but coarse spatial resolution (Jönsson and Eklundh, 2004; Verbesselt et al., 2010; Kong et al., 2019; QGIS Geographic Information System, 2021), which is problematic in dryland riparian environments where the entire woodland corridor is usually narrower than a single 250 m MODIS pixel. Some authors have attempted to circumvent this problem by fusing data from multiple sensors with different imaging properties, such as Landsat and MODIS (Gao et al., 2006; Dennison et al., 2009; Doody et al., 2017; QGIS Geographic Information System, 2021). However, algorithmic complexity and spatial assumptions in such models limit their rapid extensibility over broad areas, particularly on an annual basis. Others aggregate seasonal data from across multiple consecutive years (Melichar et al., 2023) or aggregate over space to the reach or regional level (Nagler et al., 2020; Nagler et al., 2021), which increases the effective number of samples but reduces ability to detect change, disturbance, or spatial patterns.

In this paper, we introduce a rapidly-scalable, cloud-based local regression algorithm to retrieve pixel-scale annual greenness phenology patterns from multispectral satellite imagery with 10 to 30 m spatial resolution, focusing on Landsat and Sentinel-2. We conduct sensitivity analyses with simulated data representing these and several other satellite constellations to investigate the effects of sensor noise, cloud fraction, and return interval on our phenology retrievals. We also compare our Landsat- or Sentinel-2-derived phenology estimates to estimates based on MODIS, which has a coarser spatial resolution but much faster return interval.

We then build random forest vegetation classification models to map vegetation types in drylands. We integrate phenology, summer spectral characteristics, canopy height from LiDAR, and topographic information, and focus on mapping the most groundwater-dependent riparian woodlands of phreatophytic, winter-deciduous trees. We initially trained and tested models locally at three focal sites and then built a regional classifier trained in California and tested across the

southwestern United States. Finally, we used random forest variable importance (VIP) to interrogate the relative value of different input variables for mapping vegetation types in this context. Our approach represents the first regional method for mapping riparian woodland extent and phenology in the cloud, and our phenology and classification approaches are straightforwardly extensible and could be applied to global dryland regions and a diverse suite of sensors, including data sources other than greenness.

2. Methods

We developed an approach for mapping seasonal change in plant greenness on landscape scales using multispectral satellite imagery, temporal and spectral cloud filters, and moving-window polynomial regression (Fig. 1). We validated our phenology retrievals using inter-sensor comparisons and simulation-based sensitivity tests. We then used random forest classifiers and phenology combined with other information to map land cover classes at focal test sites and regionally across the southwestern United States.

Because mapping seasonality requires repeated measurements through time, satellite imaging return interval is critically important. Simultaneously, because riparian woodlands are typically narrow in spatial extent, spatial resolution is also very important to our application. Unfortunately, these two factors usually trade off in satellite datasets due to technical limitations in sensor and orbit design. For reference, we list the properties of several satellites discussed in this paper in Table 1.

We built our classifiers on satellite imagery from either the Landsat (Wulder et al., 2019) or Sentinel-2 (Copernicus Sentinel Data, 2022) datasets, which have relatively high spatial resolution (30 and 10 m, respectively) but relatively infrequent return intervals (16 and 5 days) – a challenge that has limited their use in other studies on land surface phenology. Landsat has a longer data record, with continuous observation by a series of similar satellites since 1984, while the first Sentinel-2

satellite was launched in 2015, and the launch of a second satellite in 2018 halved the sampling interval (from 10 to 5 days). Landsat thus has the advantage in data record length, while Sentinel-2 has a relatively short record but generates more scenes each year with higher spatial resolution. Because the Landsat program has included a series of satellites partly overlapping in temporal coverage, there are periods within its four-decade record when two or even three satellites were simultaneously collecting imagery – for example, Landsat 8 and 9 are both operational at the time of writing. Combining data across multiple Landsat satellites during periods of overlap can decrease the imaging return interval, providing two or even three times as much imagery.

In the absence of on-the-ground phenology data, we also provided a cross-sensor comparison for our Landsat and Sentinel-2 models against phenology curves derived from MODIS, which has daily imagery but a much coarser spatial resolution (250 m), rendering it inappropriate for all but the broadest dryland riparian woodlands. Finally, in simulation-based sensitivity tests we included a fourth satellite constellation, PlanetScope, which has near daily imagery at very high spatial resolution (nominally 3 m). In contrast to Landsat, Sentinel-2, and MODIS, PlanetScope is not globally available for free within Earth Engine, so we included only simulated data from that constellation for comparison.

2.1. Site descriptions

We focused on three study areas, each centered on a United States military base: Vandenberg Space Force Base (VSFB) in Santa Barbara County, California, including the Santa Ynez River and portions of the Santa Ynez Mountains; Marine Corps Base Camp Pendleton (MCBCP) in San Diego County, California, including the Santa Margarita River; and Fort Huachuca Army Installation (FH) in Cochise County, Arizona, including the Huachuca Mountains and the nearby San Pedro River. Our study sites span a range of climates and include both large lowland rivers and small montane tributaries, as well as a diversity of upland vegetation types (Fig. 2).
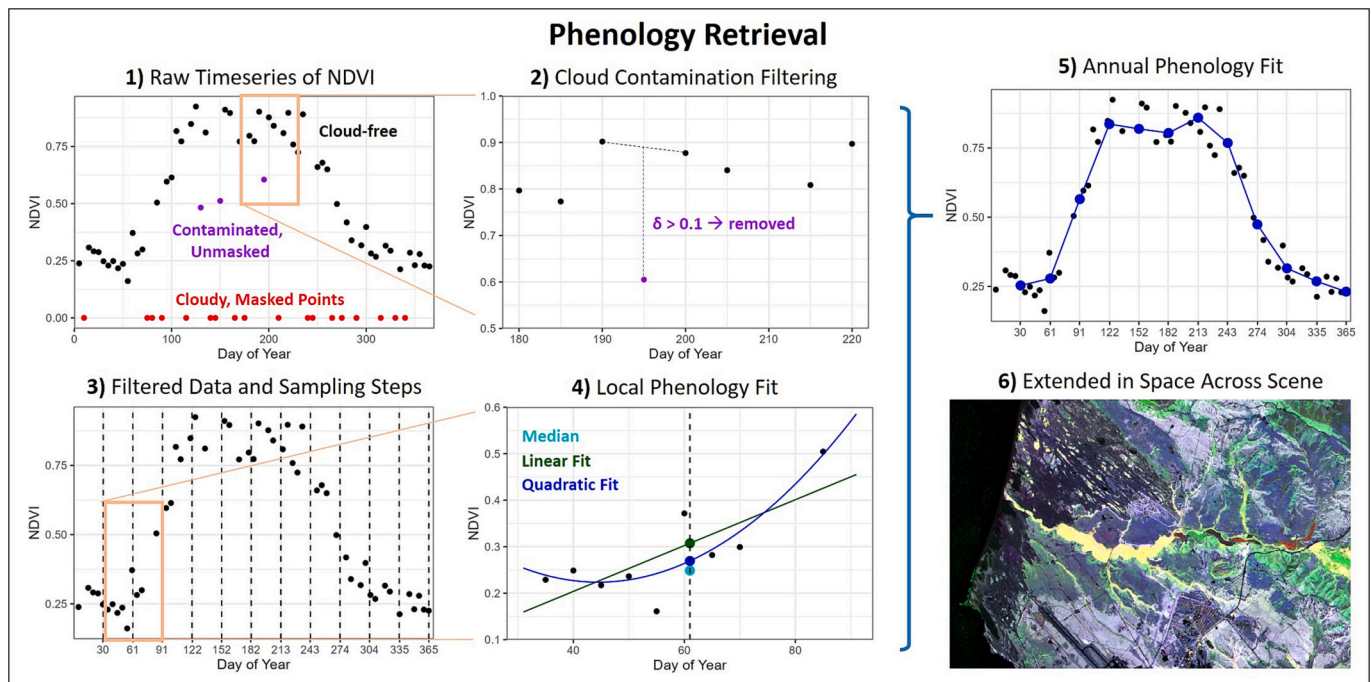

Fig. 1. Overall framework for phenology retrieval. In sub-plots 1 and 3, orange windows indicate zoomed-in extents for sub-plots 2 and 4. 1) Surface reflectance satellite imagery is collected and converted to a vegetation index (here the Normalized Difference Vegetation Index, NDVI); 2) Temporal filters are applied to remove pixels contaminated by clouds and other atmospheric effects; 3) Sampling steps are fixed at regular intervals in time (here, monthly); 4) Phenology values are fit using a quadratic, linear, or median value within time windows around each target timestamp; 5) Values are fit at each timestamp to produce a regularly-spaced seasonal NDVI curve; 6) The process is repeated for each pixel to produce spatial maps of seasonality.

Table 1

List of multispectral satellite platforms used in this paper. PlanetScope is generated by a large constellation of small satellites which together produce variable imaging timelines over both space and time. In some cases, PlanetScope imagery are available on a daily basis, but in our simulations we treat PlanetScope as having a regular 2-day return interval for simplicity.

	Period of Record		Data Properties		Use in This Paper		
	Start	End	Res. (m)	P (days)	Sensitivity Tests	Cross-validation	Veg. Classifiers
Landsat 5 (TM)	1984	2013	30	16	x		x
Landsat 8 (OLI)	2013	–	30	16	x		x
Landsat 8 + 9 (OLI)	2022	–	30	8	x	x	x
Sentinel-2a	2015	–	10	10			
Sentinel-2a + 2b	2018	–	10	5	x	x	x
MODIS	1999	–	250	1	x	x	
PlanetScope	2014	–	3	2*	x		

Both VSFB and MCBCP in coastal Southern California have Mediterranean climates characterized by cool, wet winters and warm, dry summers, with MCBCP being slightly warmer and receiving slightly less overall precipitation. By contrast, FH is located in an interior desert with a monsoonal climate, cold winters, hot summers, and rain primarily from intense storms in late summer, with smaller rain events in winter which often do not lead to surface flows (Sabathier et al., 2021, 2023). Temperature and precipitation across these dryland sites therefore differ sharply in both magnitude and seasonal timing (Fig. 2), driving differences in timing of vegetative phenology for upland plant species, which acquire most of their water from rainfall and shallow soil layers. By contrast, the phenological patterns of groundwater-supported riparian woodlands are much less variable across this region, typically featuring low greenness in the winter and consistently high greenness through the entire warm period (winter-deciduous trees).

2.2. Land cover classes

We used satellite observations to map land cover classes, including both functional vegetation types and abiotic surfaces. In each classifier, we sought to include as many surface classes as could be reasonably discriminated from one another, given the limits of both training data and salience among classes. For local classifiers, we were able to discriminate and map many fine class designations because the local contexts included relatively little within-class variability and fewer phenologically similar classes. When we generalized our three local models to a single regional model, we were also forced to generalize and simplify the upland class list to account for the diversity of vegetative types that occur on regional scales. This was also necessary because of limitations in the regional training data, which could not be crosswalked to some of the finer upland class types that were straightforward to delineate manually for local classifiers (Section 2.6). Classes are listed in Table 2, and additional detail on each is provided in Table S1 in the supplement.

In all cases, including both local and regional models, we focused primarily on obligately riparian woodlands of deciduous riparian woodlands (DRW): shallow-rooted phreatophytic trees dependent on constant access to shallow groundwater during the summer growing period (e.g. *Salix, Populus, Acer, Alnus, Platanus, Fraxinus,* etc.). We also mapped trees in evergreen woodlands (EW), which are often facultatively riparian in our region, particularly at higher elevation (Sands, 1980), and can also contribute to some of the ecosystem services provided by DRW. We included the option to combine riparian EW and DRW into an overall riparian woodland class (RW). When doing so, we used relative elevation (rooting height above the nearest thalweg point) to distinguish between evergreen trees with or without access to stream-associated groundwater, including evergreen trees as 'riparian woodland' only when their relative elevation with respect to the stream channel was <10 m. In our region, DRW trees are typically very shallow-rooted (e.g. 0.5 to 2.5 m) and only grow on sites with shallow groundwater, typically close to the channel. In contrast, many of our evergreen tree species can root more deeply, up to 10 m, and therefore may still access riparian groundwater when further from the channel (Cooper, 1922; Stromberg, 1993; Canadell et al., 1996).

In each classifier we also included some non-riparian vegetation types. In local classifiers we mapped a variety of natural upland
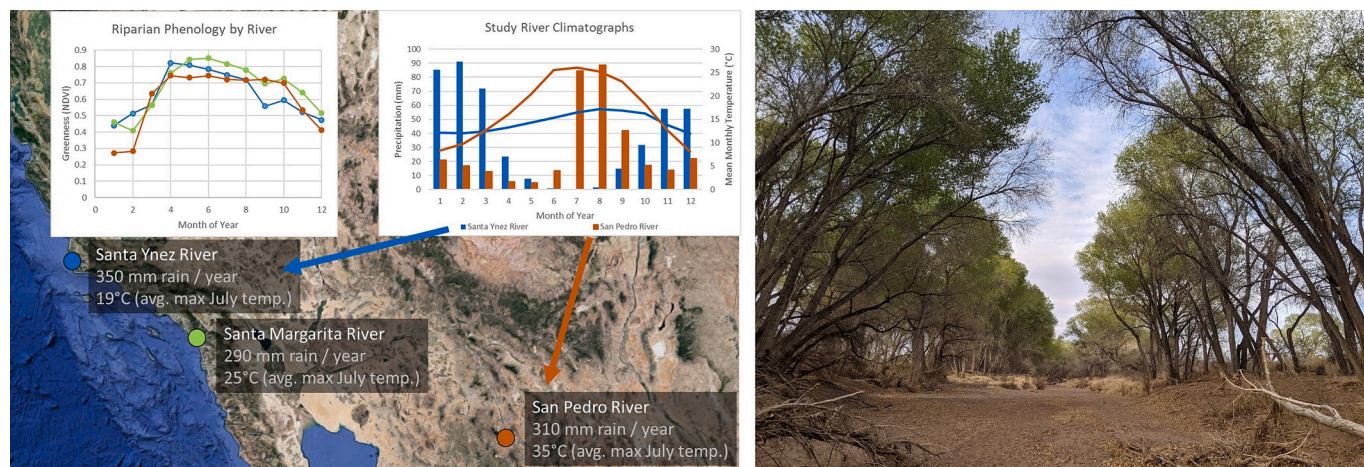

Fig. 2. Geographic, climatic, and habitat context for this study. At left: focal sites from the American Southwest with differences in climate (climatograph inset; lines are mean temperature and bars are precipitation) but similar riparian woodland composition and seasonality (phenology inset). Colors in the inset graphs correspond to the three focal sites – blue for VSFB, green for MCBCP, and orange for FH. At right: representative example of Fremont cottonwood (*Populus fremontii*) gallery woodland along an intermittent reach of the San Pedro River near FH. Climatographs are based on areal averages of long-term normal PRISM data across each study extent (PRISM Climate Group, 2014). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Land cover classes (and associated acronyms) used in various classifiers. All models included DRW, which was the focus of this study, alongside at least one evergreen woodland class and various non-woodland classes. The geographic contexts in which each class was used are indicated with an X for the following focal sites: Fort Huachuca Army Installation (FH); Marine Corps Base Camp Pendleton (MCBCP); Vandenberg Space Force Base (VSFB); and the regional southwestern United States (R). Class descriptions are provided in the supplement (Table S1). Example phenology curves are provided in Figs. 4, 5, S1, S4, and S5.

	FH	MCBCP	VSFB	R	
DRW	X	X	X	X	Deciduous Riparian Woodland
EW	X			X	Evergreen Woodland
RW				X	Riparian Woodland
RS		X			Riparian Scrub
EOW		X	X		Evergreen Oak Woodland
ENW			X		Evergreen Non-native Woodland
MB	X				Mesquite Bosque
CH		X	X		Chaparral
SS		X	X		Sage Scrub
DS	X				Desert Shrubland
GR	X	X	X		Grassland
WET		X			Wetland
NV	X	X	X		Non-Vegetated
CR	X		X	X	Crops
UR	X	X	X		Urban
W		X	X		Water
ONV				X	Other Natural Vegetation

vegetation types, as well as crops, urban areas, bare sediment or rock, and open water. In our regional classifier we lumped most upland vegetation, resulting in four classes that retain functional meaning and phenological salience across the southwestern United States: DRW, EW, other natural vegetation (ONV), and crops (CR), again with an option to combine DRW and near-stream EW as a single RW class. In each case, we retained as many upland classes as could be reasonably discriminated from one another using the data sources available. This reduces confusion by improving representation of rare classes, and allows determination of which upland classes are most frequently confused with riparian woodland.

2.3. Greenness phenology and surface reflectance retrieval

To facilitate regional mapping of greenness phenology and land cover classes, we implemented a generalized phenology retrieval algorithm to estimate continuous seasonal NDVI curves using the Google Earth Engine platform (Gorelick et al., 2017). We used this algorithm to produce separate seasonal greenness time series for each pixel and year (Fig. 1).

First, for each year and pixel, we aggregate all surface reflectance data from a given multispectral satellite constellation (for example, Landsat, Sentinel-2, or MODIS), removing any data covered by the cloud mask for that imagery. Despite the use of existing spectral cloud filters distributed with each product, some pixels affected by clouds, cloud shadows, or other localized atmospheric contaminants were not reliably removed (Fig. 1). To account for this, we applied an additional temporal filter, which compares each pixel value to the values immediately before and after in the time series and masks pixels when the NDVI of the target image was >0.1 below the prediction of a linear regression in time between neighbor values. Clouds and cloud shadows tend to have much lower greenness than healthy vegetation. Consequently, pixels that are even partly obscured by cloud cover show a strong reduction in greenness vs. the previous and following scenes, even when the overall mixed pixel spectrum is still sufficiently plant-like to be missed by the spectral cloud filter. We applied this temporal cloud filter twice in series to remove cloudy pixels even when multiple consecutive images were contaminated in this way.

Next, we applied an adaptive local regression to smooth the data and

estimate NDVI values at arbitrary points in time. Our approach needed to be flexible across a greatly varying range of imaging frequencies, both because of variation in return interval across sensors and because even within a given data archive, cloud impacts lead to strong variation in clear scene count between years, locations, and even seasons within a year. We used a local regression around each time step to allow each modeled NDVI value to be estimated independently across both space and season, substantially increasing the capacity for parallelization in Google Earth Engine. We used an adaptive regression order (a quadratic, linear, or median fit) because higher-order fits prevent clipping of local maxima and minima within the timeseries, but the large variation in the number of cloud-free scenes for individual fitting windows causes variation in the efficacy of higher-order fits. A regression will only smooth noise when the number of data points used in the fit is meaningfully higher than the order of the regression – otherwise, the fit will respond strongly to noise and provide inaccurate estimates – so we included linear or median fits for instances with few cloud-free scenes in the sample window. The algorithm we developed is as follows (and is illustrated visually in Fig. 1):

1. Specify the number of annual time steps to use (e.g. 12 for monthly estimates or 52 for weekly estimates of greenness, etc.) and evenly sample time steps through the year. Resampling from scene dates to fixed time steps allows the same set of temporal predictors to be used across variation in cloud cover and the edges of image tiles.
2. Aggregate all cloud-free NDVI data within a 60-day window centered on each time step. For example, imagery with no cloud cover and a 1-day return interval (MODIS) will have about 60 scenes in each window, while imagery with 40% cloud cover and a 16-day return interval (Landsat) will have only about 2 scenes.
3. Calculate an estimated NDVI value for each time step, using the following decision framework:
   a. First, find the median NDVI value of samples in the 60-day window, and also apply both a quadratic regression and a linear regression in the window to estimate the NDVI value at the sampling time.
   b. For sampling windows with $n \geq 6$ scenes, if the quadratic regression predicts a physically reasonable NDVI value (-1.0 to 1.0), and the absolute value of the z-score of the prediction with respect to the distribution of values in the sample window is <1.5 standard deviations (i.e., not a statistical outlier), use the quadratic prediction.
   c. OTHERWISE, for sampling windows with $n \geq 3$ scenes, if the linear regression predicted a physically reasonable NDVI value (-1.0 to 1.0), and the absolute value of the z-score of the prediction with respect to the distribution of values in the time window is <1.5 standard deviations, use the linear prediction.
   d. OTHERWISE, use the median NDVI value of the neighborhood.
4. Finally, after each time step is filled with an NDVI estimate in this way, use linear interpolation to fill any gaps (which only occur when no cloud-free observations occurred in the 60-day sampling window).

We also considered other vegetation indices (VIs) as alternatives to NDVI, including the Soil Adjusted Vegetation Index (SAVI – Huete, 1988) and Enhanced Vegetation Index (Huete et al., 2002). In comparison to NDVI, SAVI partially addresses soil background effects for areas with low plant cover (such as drylands) (Qi et al., 1994; Huete, 1988). Similarly, EVI partially corrects for the atmosphere and soil background; it also continues to respond to increases in leaf area index (LAI) over very dense canopies, after NDVI has saturated (Huete et al., 2002).

However, because both SAVI and EVI introduce terms which are not simple band ratios (the L term for soil brightness), they are both sensitive to overall changes in illumination from shadowing or topographic effects (Matsushita et al., 2007; Galvao et al., 2011; Liao et al., 2015; Chen et al., 2020; Yin et al., 2022). This is problematic over montane

and valley sites with substantial variation in slope and aspect, particularly for phenology retrieval, as the solar angle changes seasonally. In contrast, NDVI is less sensitive to topographic effects (Chen et al., 2020). Additionally, while EVI has greater sensitivity than NDVI at high LAI values, NDVI has greater sensitivity at low LAI (Huete et al., 2002), and dryland landscapes develop LAI mostly on the lower end of this sensitivity-response continuum (Stromberg et al., 1993). Soil background effects may lead to problematic variation in annual minimum NDVI for sites with the same vegetative cover but different soils. Analogously, saturation can lead to similar peak NDVI values at sites with differing maximum LAI. However, because our classifier is based on phenology, peak and minimum values are less impactful than the timing of shifts in greenness, which are captured very similarly by all three vegetation indices, albeit sometimes with slight differences in timing of senescence (Fig. S1, Huete et al., 2002; de Silveira et al., 2007; Wardlow et al., 2007). As a result, other studies comparing phenological vegetation classifiers show negligible or inconsistent differences in accuracy when the VI is changed between NDVI and EVI (Huete et al., 2002; de Silveira et al., 2007; Wardlow et al., 2007; Wardlow and Egbert, 2010; Halabuk et al., 2015; Yan et al., 2015; Melichar et al., 2023). Finally, EVI requires inclusion of a blue band in addition to the red and NIR bands in SAVI and NDVI. Some sensors (e.g. AVHRR) do not include a blue band; others (e.g. MODIS) include a blue band with coarser spatial resolution than the red and NIR bands. This means that for many potential applications, including portions of our analysis, NDVI or SAVI can be produced at finer spatial resolution than EVI, which is important in narrow riparian areas.

We performed comparisons using NDVI, SAVI, or EVI. We found that spatial patterns in seasonality were consistent across the three indices (Fig. S1). Additionally, pixel-wise correlation between monthly values of these indices was high (0.97 median correlation for NDVI vs. SAVI and 0.96 for NDVI vs. EVI). We found no difference in riparian class accuracy when switching between indices, and slight reductions in overall class accuracy for both SAVI and EVI in comparison to NDVI (Fig. S2). Consequently, we proceed with NDVI to minimize topographic effects in montane catchments and maximize spatial resolution. However, we have also included EVI and SAVI as outputs in our Earth Engine library for other users to apply in other contexts.

To validate the consistency of our phenology algorithm across

sensors and differences in cloud cover, we used it to extract weekly NDVI from 2022 imagery in the Sentinel-2, Landsat, and MODIS archives around San Antonio Creek on VSFB (Fig. 3). We selected this site because it is one of our broader and more contiguous riparian areas, facilitating accurate riparian phenology retrievals even when using MODIS. We extracted phenology samples for *Salix lasiolepis* riparian woodland (DRW), maritime chaparral (CH), annual grassland (GR), and bare beach sand (NV), totaling about 0.6 km^2 of area, and directly compared mean phenology curves for each class across sensors (Fig. S3). We also aggregated the Sentinel-2 and Landsat imagery up to mean values at the 250 m spatial scale of MODIS and directly compared weekly greenness values from each sensor pair for individual MODIS pixels (Fig. S4).

We then applied our phenology algorithm to extract seasonality at each focal site (FH, MCBCP, and VSFB) and at the sample points used to train and validate the regional-scale classifier (Section 2.6). Although our algorithm allows phenology curves to be sampled at arbitrary time intervals, we built all classifiers using monthly greenness fits, using imagery from the same year as the training / validation data.

Finally, we also generated mean summer spectra images ("leaf-on" conditions for DRW) in the blue, green, red, NIR, SWIR1, and SWIR2 bands for each site using Earth Engine. We calculated the mean surface reflectance across all cloud-free imagery (either Landsat or Sentinel-2 depending on the classifier, see section 2.7) between June 1 and August 31 in each test year for each band. For Sentinel-2, we disaggregated the two SWIR bands from their native 20 m resolution to match the 10 m resolution used for all other bands.

2.4. Phenology retrieval – Sensitivity simulations

We developed simulations to test our ability to accurately retrieve land surface phenology through partial cloud cover, sensor noise, and variation in atmospheric correction quality. In doing so, we sought to determine where in space and time each of our focal sensors could be relied upon to retrieve surface phenology, given regional variation in cloud cover. We assumed the surface was covered by a combination of leafy vegetation and soil, and used a double-logistic function to estimate leaf fractional cover as a function of day-of-year, varying from 0 to 100% leaf cover, with parameter definitions given in Table 3:
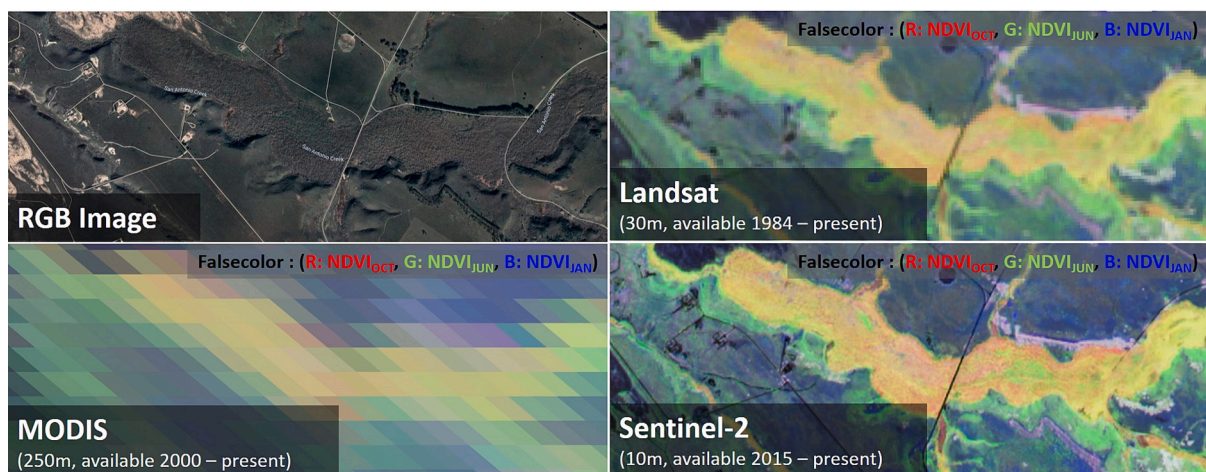

Fig. 3. Example phenology imagery produced for San Antonio Creek on VSFB, comparing results when using different sensor platforms. Clockwise from the upper left: Google Earth truecolor satellite imagery; phenology falsecolor from Landsat 8; phenology falsecolor from Sentinel-2a + 2b; phenology falsecolor from MODIS. Colors in phenology images are displayed with October, June, and January NDVI loaded as RGB. Based on these colour loadings, riparian woodland dominated by *Salix lasiolepis* (DRW) displays as yellow-orange because it is winter-deciduous and has high greenness in October and June but not in January. By contrast, surrounding uplands that are covered mostly by grasses and forbs (GR) are blue because they are senesced in June and October. Evergreen *Eucalyptus globulus* trees (ENW or EW) on the north bank of the stream are white. Google satellite background: Imagery ©2023 Airbus, Imagery ©2023 Airbus, CNES / Airbus, Landsat / Copernicus, Maxar Technologies, USDA/FPAC/GEO, Map data ©2023. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Definitions of parameters and range of values used in phenology simulations.

	Range of Values	Description
F_{leaf}	0.0 to 1.0	Driven by double-logistic function
f	0.00, 0.05, 0.10, … 0.85, 0.90, 0.95	Fraction of cloudy days
T	1, 2, 5, 8, 16, 30	Sensor return interval
SNR	2, 5, 10, 20, 100	Signal-to-noise ratio
σ_{fixed}	0.01, 0.02, 0.05, 0.10	Fixed noise term
n	10, 20, 30, 40, 50, 60	Width of sampling window for local regression
k_{spring}	2, 5, 10, 20, 30	Exponential rate of greenness change in spring
k_{fall}	2, 5, 10, 20, 30	Exponential rate of greenness decline in fall
t_{spring}	90	Timing of spring inflection point (day of year)
t_{fall}	150, 180, 210, 240, 270, 300	Timing of fall inflection point (day of year)

$$F_{leaf} = \frac{1}{1 + e^{-k_{spring}\left(t - t_{spring}\right)}} - \frac{1}{1 + e^{-k_{fall}\left(t - t_{fall}\right)}} \tag{1}$$

We chose fixed spectral endmembers for pure vegetation with red reflectance at 0.05 and near infrared (NIR) reflectance at 0.5, and soil with red reflectance at 0.24 and NIR reflectance at 0.4. At each time sample, we estimated overall reflectance as a linear combination of the endmember reflectance, weighted by fractional cover:

$$R_{red} = 0.24 \bullet \left(1 - F_{leaf}\right) + 0.05 \bullet F_{leaf} \tag{2}$$

$$R_{NIR} = 0.40 \bullet \left(1 - F_{leaf}\right) + 0.50 \bullet F_{leaf} \tag{3}$$

Varying k_{spring}, k_{fall}, t_{spring}, and t_{fall} across the values shown in Table 3 resulted in 150 different underlying seasonal phenology curves. We sampled these curves at 52 weekly time steps to build a reference time series of 'true' seasonal NDVI.

Next, we simulated sensor retrieval of NDVI by sampling R_{red} and R_{NIR} at a separate fixed interval (sensor return interval, T) and adding random Gaussian noise $N(0, \sigma)$ with mean 0 and standard deviation σ to each reflectance value. We broke σ into two noise terms: a term based on the signal-to-noise-ratio (SNR) proportional to the reflectance, which we varied across simulations to model variation in radiometric sensor quality, and a constant noise term for errors in atmospheric correction and surface reflectance retrieval:

$$\sigma = \sigma_{fixed} + SNR \bullet \frac{R}{2}$$

$$R_{noisy} = R + N(0, \sigma) \tag{3}$$

We then calculated NDVI at each sensor return date based on the two noise-contaminated reflectance values. The range of values we used for the fixed and signal-dependent noise terms were based on Okin and Gu (2015) who used varying SNR from 15 to 355 for various Landsat sensors and bands, and who report RMSE values between known surface reflectance and simulated reflectance retrievals following atmospheric correction which vary from 0.0269 for Landsat OLI to 0.0656 for Landsat 5 TM.

In each test, we assumed a variable fraction, f, of cloudy days, and randomly removed that fraction of sample points from the vector of simulated sensor values for each test case, resulting in final simulated time series with 365(1-*f*)/T samples, each point representing a single distinct cloud-free image acquisition. For example, a Landsat-like 16 days for T provides 22 samples in the year, and with f = 0.5 for cloud cover half of those were randomly removed, leaving 11 cloud-free observations. Finally, we applied our moving-window polynomial regression algorithm to retrieve weekly greenness estimates based on the simulated sensor values, using a time window radius of N days.

All permutations of the parameter values in Table 3 were generated,

Table 4

Variables used for each random forest classifier. Summer spectral values (6) and monthly phenology values (12) were extracted from either Sentinel-2 (S2) or Landsat (L). Reflectance values were cloud-free averages from June 1 to August 31. Topographic and hydrologic variables (4) were computed using either SRTM or LiDAR – these include elevation (DEM), terrain slope, and vertical and horizontal distance to the nearest stream flowline (respectively Rel. Elev. and Creek Dist.). Canopy height (CHM) was only included in models which included LiDAR variables.

Site		Huachuca - Local Classifier				Pendleton - Local Classifier				Vandenberg - Local Classifier					Regional
Model		All	No Li.	Sp.	Ph.	Li.	All	No Li.	Sp.	Ph.	Li.	SRTM	2013	2015	2017	2019	
Spectra	Blue	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
	Green	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
	Red	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
	NIR	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
	SWIR1	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
	SWIR2	S2	S2	S2			S2	S2	S2			SRTM	L	L	L	L	L or S2
Seasonality	Jan NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Feb NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Mar NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Apr NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	May NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Jun NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Jul NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Aug NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Sep NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Oct NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Nov NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
	Dec NDVI	S2	S2		S2		S2	S2		S2		SRTM	L	L	L	L	L or S2
3D Structure	CHM	LiDAR				LiDAR	LiDAR				LiDAR						
	DEM	LiDAR	SRTM			LiDAR	LiDAR	SRTM			LiDAR	SRTM	SRTM	SRTM	SRTM	SRTM	SRTM
	Slope	LiDAR	SRTM			LiDAR	LiDAR	SRTM			LiDAR	SRTM	SRTM	SRTM	SRTM	SRTM	SRTM
	Rel. Elev.	LiDAR	SRTM			LiDAR	LiDAR	SRTM			LiDAR	SRTM	SRTM	SRTM	SRTM	SRTM	SRTM
	Creek Dist.	LiDAR	SRTM			LiDAR	LiDAR	SRTM			LiDAR	SRTM	SRTM	SRTM	SRTM	SRTM	SRTM

resulting in 432,000 simulated phenology retrievals. For each test, we computed the R^2 and root mean square error between underlying phenology and the retrievals from our curve-fitting algorithm.

2.5. Vegetation structure and topography

We mapped topography using LiDAR or the Shuttle Radar Topography Mission (SRTM - Farr and Kobrick, 2000), and vegetation structure using LiDAR. When working with LiDAR, we first extracted ground returns by selecting local minima among last-returns and filtering for points with locally-smooth surface normals (<0.5 rad average deviation with respect to neighboring surface angles within 1 m). We extracted vegetation as points that were >0.5 m higher than a Delaunay triangulated irregular network fit to the ground returns. We rasterized the terrain point cloud to a digital elevation model (DEM) and digital surface model by taking the minimum and maximum values within 10 m pixels (the native resolution of our Sentinel-2 NDVI imagery), and then computed the local terrain slope (θ). Maximum vegetation height in each pixel (CHM) was determined as the difference between the digital surface model and digital elevation model. This simple difference will overestimate vegetation height for sloped terrain due to differences in the maximum vegetation height and minimum ground height at opposite edges of a pixel, so CHM values were normalized to account for terrain:

$$CHM = CHM_0 - tan\,(\theta) \bullet W_{pixel} \qquad (3)$$

Where W_{pixel} is the pixel width (10 m) and CHM_0 is the uncorrected canopy height. Finally, we evaluated relative elevation as the vertical distance between each terrain point and the nearest stream flowline mapped in the National Hydrography Plus dataset (USGS, 2019). The source files, example scripts, and more details on the methods used to run these analyses are available in the Dirt or Leaf (McMahon, 2022a) and LiDAR Raster Stats (McMahon, 2022b) C++ packages. LiDAR for the two military bases was provided confidentially by the bases and is not publicly available; however, an additional new LIDAR survey was flown along the San Pedro River (McMahon, 2022), which is now available online.

LiDAR was processed in this way for FH and MCBCP, but not for VSFB, where LiDAR was not available (Table 4). However, topographic information was also generated for all sites using SRTM, either to replace LiDAR in comparative tests for FH and MCBCP (Tables 3, S2), or as the sole source of terrain information at VSFB and the regional classifier. While SRTM is not able to map vegetation structure, it requires much less preprocessing, consumes far smaller data volumes, and is available globally. When using SRTM for locally-trained models, we repeated the above analysis, but instead of processing point cloud data we worked directly from the 10 m DEM provided by SRTM, and we did not include structural information on vegetation.

For our regional model, we relied upon existing topographic information in the Google Earth Engine data catalog, again including the SRTM digital elevation model and terrain slope, as well as an existing global map of relative elevation at 30 m (Donchyts et al., 2016), disaggregated to the 10 m scale of Sentinel-2 for any analyses using that satellite constellation.

2.6. Extraction of training points

For each of our locally-trained models, we used the QGIS and QField apps (QGIS 2021) to manually delineate polygons representing examples of each vegetation type. We labeled polygons from each class using field visits, high-resolution aerial imagery, and LiDAR for vegetative structure and topography. For each test site, we assigned half of the available polygons in each class to training and half to validation, to ensure an independent validation (individual stands of vegetation were never represented in both training and validation datasets). We then randomly sampled pixels without replacement from each polygon until each class had the same number of points in the training and validation sets (equal to one half as many points as the rarest class, for a class-balanced test). We sampled the phenology, summer surface reflectance, SRTM, and LiDAR rasters for each pixel and aggregated the result into a dataframe. This produced approximately 500 total samples for each class in each test site for both the training and validation datasets.

We based the training data for our regional model on CALVEG (McClellan 2004), which provides maps of existing vegetation across California, classified with the National Vegetation Classification System (NVCS). Since our analysis was focused on drylands, we removed mesic ecoregions by subsetting the dataset to include only the following ecoregions: the South Coast, South Interior, Central Coast, Central Valley, South Sierran, and Great Basin. We aggregated all NVCS divisions representing four target classes: DRW, EW, ONV, and Crops (Table 2). We sampled 3000 points for training and validation from DRW and 1000 each for the other classes (Fig. S5). We stratified our sampling within each class by NVCS division to ensure representation of the diversity of subtypes within each vegetative class. For example, NVCS division D193, "Vancouverian Flooded & Swamp Forest Division", and D013, "Western North American Interior Flooded Forest Division", were given equal representation in the DRW sample sets. Within each division, we randomly sampled polygons without replacement, with sampling probability weighted by polygon area, then buffered polygons inward by 15 m and sampled a point in the buffered polygon. This ensured that a 30 m Landsat pixel centered at that point would not overlap other polygons, and that no polygon would contribute more than one point. We used Google Earth Engine to extract Landsat phenology, summer surface reflectance, SRTM topographic information, and latitude at each training or validation point as described in Sections 2.3 and 2.4 using data from the same year as the source products for each CALVEG polygon.

Some CALVEG polygons are mapped over sparse vegetation, which covers a small fraction of the polygon's overall area – for example, portions of a stand mapped in CALVEG as 'evergreen woodland' might have <50% overall tree cover, with the rest taken up by herbaceous plants or bare soil. To account for this, we subset the training data to reject points where the target class was a minority of the actual overall cover. For DRW, we removed points where the June–August average NDVI was lower than 0.5 or where the increase in average NDVI for summer (June–August) over winter (January–February) was <0.1. For EW, we removed points where June–August NDVI was lower than 0.5 or where the standard deviation in monthly NDVI across the year was >10% of the mean annual NDVI.

2.7. Model training and validation

We trained 18 random forest classifiers using different combinations of data sources and training sites (Table 4). For each classifier, we trained a random forest model using a combination of structural or topographic information, phenology data, and training / validation polygons, with different datasets used in each classifier (Table 4). Whenever used, individual band values for summer surface reflectance and individual monthly greenness values from the phenology curve were treated as separate predictors in the random forest model. Topographic and structural models were also included as individual predictors.

At both FH and MCBCP, we used Sentinel-2 for optical imagery, and built models on the following data sources (number of predictors for each category in parentheses):

1. Phenology (12) + Summer Spectra (6) + LiDAR (5)
2. Phenology (12) + Summer Spectra (6) + SRTM (4)
3. Phenology alone (12)
4. Summer Spectra alone (6)
5. LiDAR alone (5)
6. SRTM alone (4)

LiDAR was not available at VSFB, so we built models using SRTM for topographic information and Landsat for phenology and summer spectra. However, to test the models' performance over time as phenology varied between years, we tested four different years of imagery from VSFB with divergent water availability: 2013, 2015, 2017, and 2019. Annual rainfall at the nearby Lompoc rain gauge (Santa Barbara County Flood Control District, 2023) was 184, 203, 563, and 519 mm during these four rain years, compared to a long-term average of 367 mm. We trained four different classifiers on imagery from each of the four years, then used each classifier to predict validation datasets in each year, resulting in 16 total validation assessments (e.g. trained in 2013 + validated in 2013, trained in 2013 + validated in 2015, etc.).

Our regional model was built using SRTM (4 predictors), latitude (1 predictor), summer surface reflectance (6 predictors), and phenology (12 predictors). We validated the regional model both by comparing it directly to an independent validation dataset extracted from CALVEG

Table 5

Summary of results for all classifiers. Each row represents a different classifier test. Lo and Re designate locally vs. regionally trained models. All scales are in meters (either 10 m or 30 m, depending on which optical sensor was used). FH is Fort Huachuca, MCBCP is Marine Corps Base Camp Pendleton, VSFB is Vandenberg Space Force Base, and SBC is Santa Barbara Creeks. Struct., Sp., and Ph. respectively refer to the data source for canopy or terrain structure, average summer reflectance, or vegetative phenology. S2 is Sentinel-2 and L is Landsat. Sens., Spec., and BA are respectively sensitivity, specificity, and balanced accuracy for the riparian class (the Target column denotes the riparian class used – either DRW or RW, depending on the test). Acc. and Kappa are the overall accuracy and kappa statistic for the classifier across all classes. All accuracy statistics are reported as percentages (%). Colors are scaled from 70% (red) to 100% (blue).

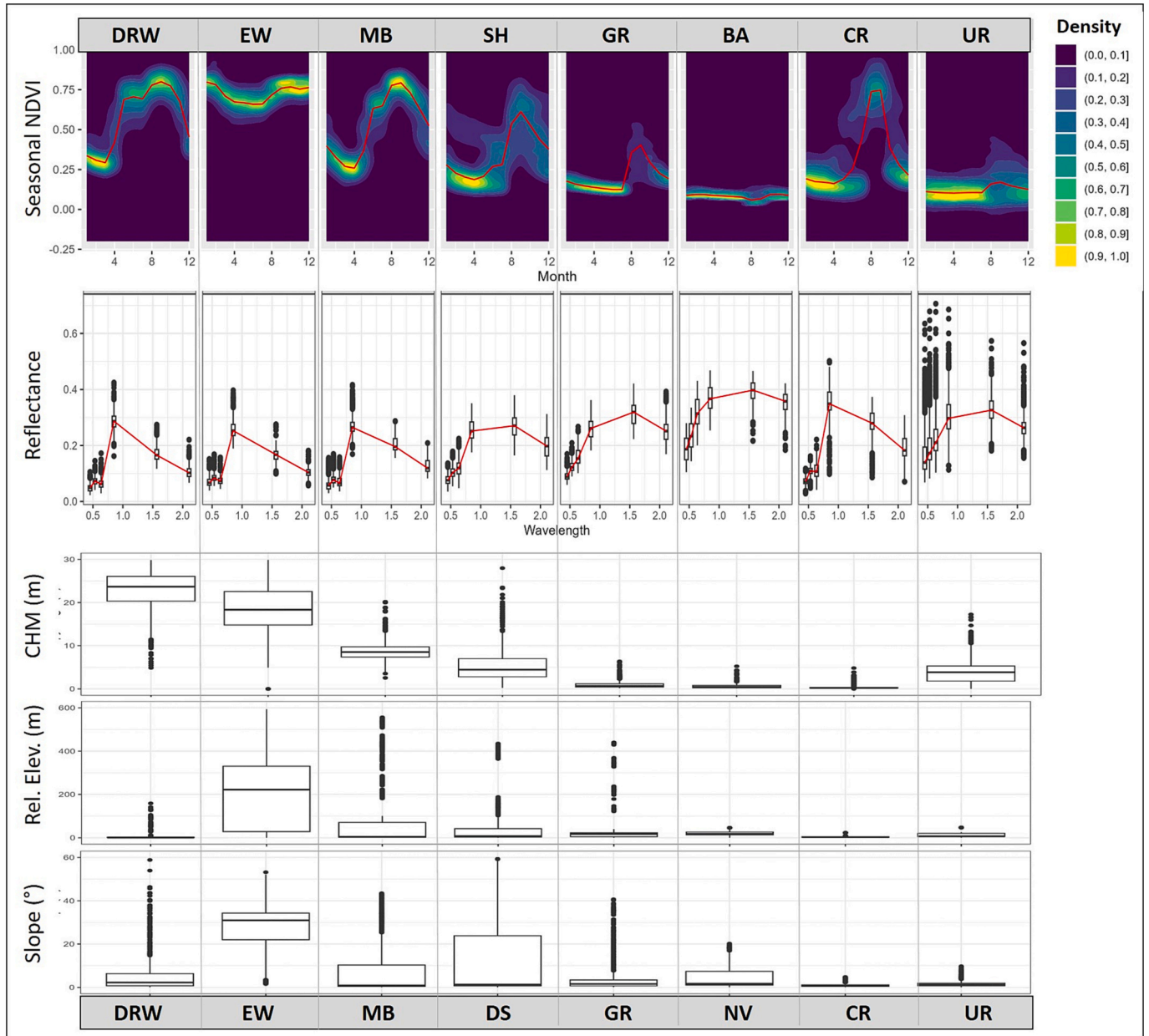| Context				Data				Riparian			Overall	
Extent	Scale (m)	Site	Year	Target	St.	Sp.	Ph.	Sens.	Spec.	BA	Acc.	Kappa
Lo	10	FH	2021	DRW	LiDAR	S2	S2	95.4	99.3	95.6	99.3	0.97
Lo	10	FH	2021	DRW	SRTM	S2	S2	92.7	98.4	95.5	86.2	0.84
Lo	10	FH	2021	DRW	---	S2	---	80.7	95.2	88.0	67.1	0.62
Lo	10	FH	2021	DRW	---	---	S2	90.4	96.9	93.6	80.7	0.78
Lo	10	FH	2021	DRW	LiDAR	---	---	96.5	99.9	98.2	70.1	0.66
Lo	10	FH	2021	DRW	SRTM	---	---	47.9	97.8	72.8	47.8	0.40
Lo	10	MCBCP	2021	DRW	LiDAR	S2	S2	96.8	99.7	98.2	84.6	0.83
Lo	10	MCBCP	2021	DRW	SRTM	S2	S2	95.0	99.4	97.2	79.5	0.77
Lo	10	MCBCP	2021	DRW	---	S2	---	78.0	95.3	86.7	63.4	0.59
Lo	10	MCBCP	2021	DRW	---	---	S2	94.4	99.5	97.0	71.6	0.68
Lo	10	MCBCP	2021	DRW	LiDAR	---	---	74.8	94.8	84.8	69.1	0.66
Lo	10	MCBCP	2021	DRW	SRTM	---	---	43.4	88.4	65.9	53.8	0.49
Lo	30	VSFB	2013	DRW	SRTM	L	L	98.6	99.6	99.1	91.9	0.91
Lo	30	VSFB	2015	DRW	SRTM	L	L	98.4	99.4	98.9	87.4	0.86
Lo	30	VSFB	2017	DRW	SRTM	L	L	97.4	99.5	98.4	90.3	0.89
Lo	30	VSFB	2019	DRW	SRTM	L	L	95.1	99.6	97.4	89.6	0.88
Re	10	FH	2015	DRW	SRTM	S2	S2	100.0	94.3	97.2	96.8	0.97
Re	10	MCBCP	2016	RW	SRTM	S2	S2	97.8	91.2	94.5	97.0	0.86
Re	10	SBC	2021	RW	SRTM	S2	S2	96.6	94.6	95.6	95.8	0.91
Re	30	CLVG	2019	DRW	SRTM	L	L	95.8	100.0	97.9	92.5	0.89
Re	30	CLVG	2019	DRW	SRTM	L	L	96.1	99.7	97.9	97.8	0.96
Re	30	CLVG	2019	RW	SRTM	L	L	96.8	99.0	97.9	97.1	0.94

Fig. 4. Seasonal NDVI phenology, average summer spectra, canopy height (CHM), relative elevation, and terrain slope for training data extracted at Fort Huachuca and the San Pedro River, separated by surface class. Key to classes: deciduous riparian woodland (DRW), evergreen woodland (EW), mesquite bosque (MB), desert shrubland (DS), grassland (GR), non-vegetated terrain (NV), cropland (CR), and urban (UR). For more detail on classes see Table S1.

and by predicting classes for manually-delineated vegetation polygons from each of our study sites. When training the model and when validating with CALVEG, we used Landsat for surface reflectance and phenology data, because some CALVEG polygons predate the launch of Sentinel-2. When validating with our local polygons we used Sentinel-2 imagery to predict class values because of its finer spatial resolution and to provide an additional cross-sensor comparison.

For each classifier, we trained a random forest model using the training datasets, with 50 classification trees and a bag fraction of 0.5. We report relative variable importance (VIP) for each classifier with each combination of variables (Table S2, S3). We then predicted classes at the independent validation pixels and report error matrices (Table S3 – S6). Finally, we calculated overall accuracy and Cohen's kappa statistic (Cohen, 1960) across all classes, as well as sensitivity, specificity, positive prediction rate (PPV), negative prediction rate (NPV), and balanced accuracy (BA) for each individual class (Table 5):

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

$$NPV = \frac{TN}{TN + FN} \tag{7}$$

$$BA = \frac{Sensitivity + Specificity}{2} \tag{8}$$

Where TP is the number of true positive cases, TN is true negatives, FP is false positives, and FN is false negatives.

3. Results

3.1. Phenology and terrain retrievals

Phenology trends differed between classes (Figs. 4, 5, S1, S4, S4), particularly for DRW, which was usually the only class with low greenness in winter and high greenness through the entire warm season. In Arizona, xeroriparian mesquite bosque (MB) also had a winter-deciduous phenology, but it greened up later in spring than DRW and had a much lower canopy height (Fig. 4). Some evergreen classes like evergreen oak woodland (EOW) and chaparral (CH) also had similar phenology but were separable by LiDAR or SRTM metrics (Fig. S6).

Classes showed some variability in magnitude of greenness and timing of phenological events between years (Fig. 5). However, broad phenological patterns still separated most classes – especially for DRW, which had lower interannual variability in phenology than the other vegetative classes. For DRW in dry vs. wet years, onset of spring greenness and fall senescence both occurred slightly earlier, and peak greenness values during summer were slightly lower (by about 0.025 NDVI). The largest differences in wet vs. dry years were for the annual grassland class (GR), which senesced much earlier in the two dry years compared to the two wet years. The evergreen classes also showed some reduction in late-season greenness during dry years relative to wet years.

Phenology retrievals were consistent when the algorithm was applied separately to data from different sensors, building confidence in the robustness of the method. At coastal sites around San Antonio Creek on VSFB, estimates of mean weekly phenology in 2022 were similar between MODIS, Landsat, and Sentinel-2 for DRW, chaparral, and grassland (Fig. S3). Phenology retrievals from Landsat and Sentinel-2
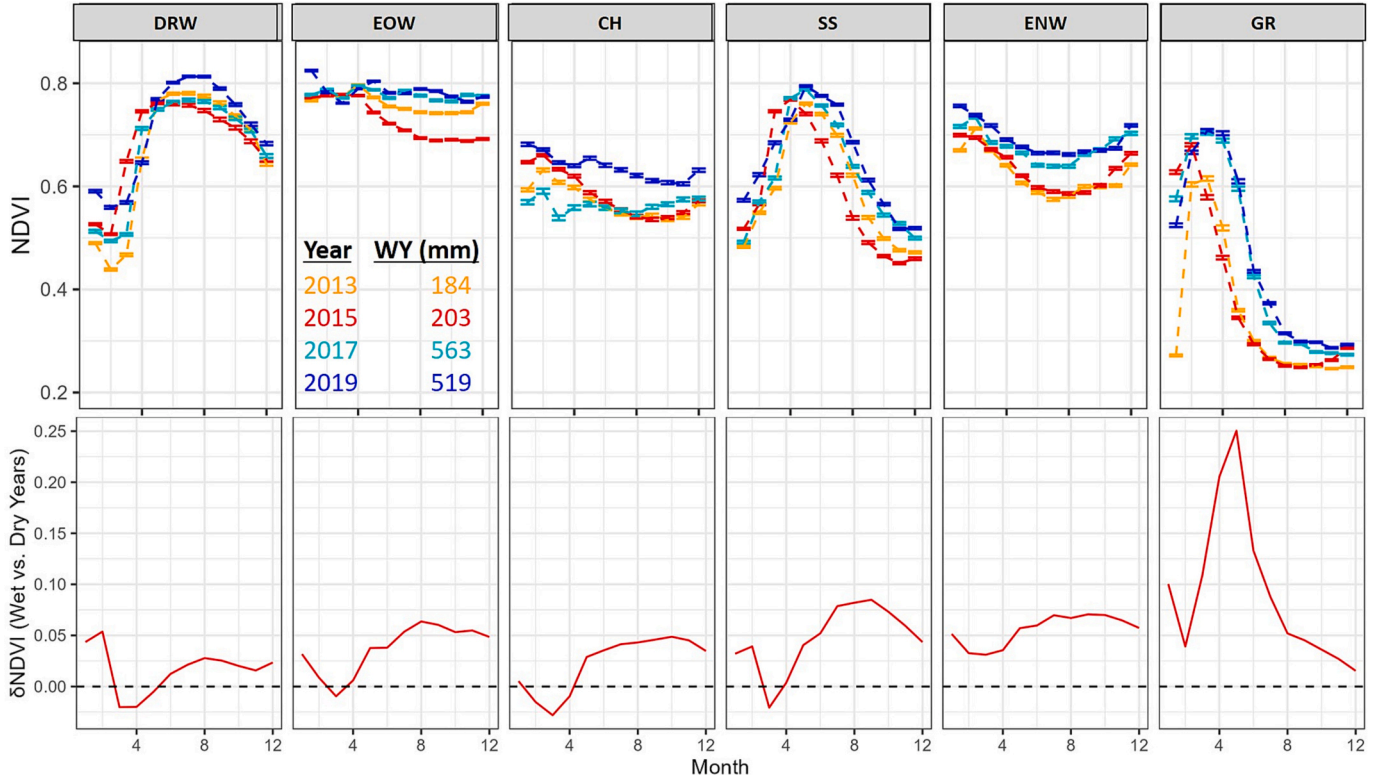

Fig. 5. Variation in Landsat-derived phenology at VSFB across four years, separated by vegetation type (Table 2). In upper plots, lines show the mean monthly NDVI across all training points in each class, with bars denoting one standard error around the mean. Line colour shows the year; total water year precipitation in mm (during the previous wet season) is given for each year in an inset. Lower plots show the average increase in monthly greenness in wet (2017, 2019) minus dry (2013, 2015) years. Classes are deciduous riparian woodland (DRW), evergreen oak woodland (EOW), chaparral (CH), sage scrub (SS), evergreen non-native woodland (ENW), and grassland (GR) – for more information on classes see Tables 2 and S1.
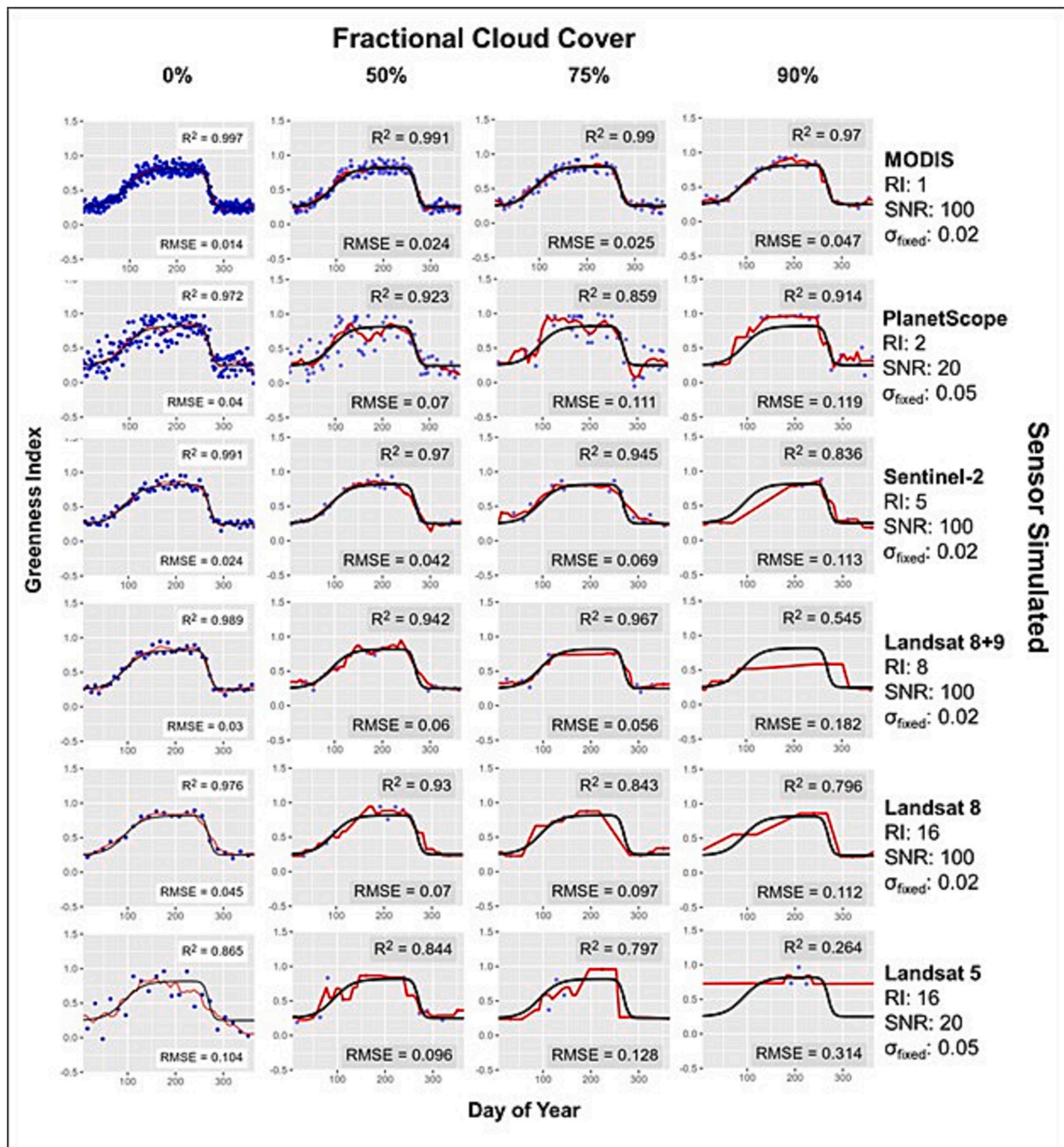
Fig. 6. Examples of phenology retrievals for simulated noisy and cloud-contaminated data. In each plot, the black curve represents the simulated "true" phenology curve of the surface. Blue dots represent satellite samples injected with noise and subsampled due to cloud cover. The red line shows the phenology curve retrieved from the noisy and cloud-contaminated data; good model fits are those where the red and black curves are similar. For each plot, the R^2 and root mean square error between the retrieved (red) and true (black) lines are given in the top right and bottom right corners, respectively. Columns show variation in the fraction of scenes obscured by clouds, from 0% to 90% of all scenes. Rows correspond to tests with varying noise (SNR, σ_{fixed}) and imaging return interval (RI). Each row is intended to qualitatively correspond to an actual satellite platform. From top to bottom: MODIS; PlanetScope; Sentinel-2a + Sentinel-2b; Landsat 8 + Landsat 9; Landsat 8 alone; and Landsat 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

had greater cross-sensor correspondence than MODIS had with either other sensor – for example, the Pearson correlation coefficient between weekly Landsat and Sentinel-2 estimates of greenness was 0.966, 0.985, and 0.987 for DRW, chaparral, and grassland, respectively. The analogous values for comparisons between MODIS and Landsat or Sentinel-2 ranged from 0.907 to 0.982. Root mean square differences in paired weekly estimates of NDVI varied from 0.028 to 0.080, depending on the sensor pair and surface class. After aggregating Landsat and Sentinel-2

to the resolution of MODIS, pairs of weekly NDVI estimates at individual pixels were also highly repeatable across sensors (Fig. S4), with r between Landsat and Sentinel-2 of 0.99.

At FH and MCBCP, we used linear regressions to compare LiDAR vs. SRTM estimates of absolute elevation, relative elevation, and terrain slope. At the relatively coarse spatial resolution of 10 m, we found that the two methods produced similar terrain metrics. At training points on FH, the Pearson correlation coefficient between LiDAR vs. SRTM
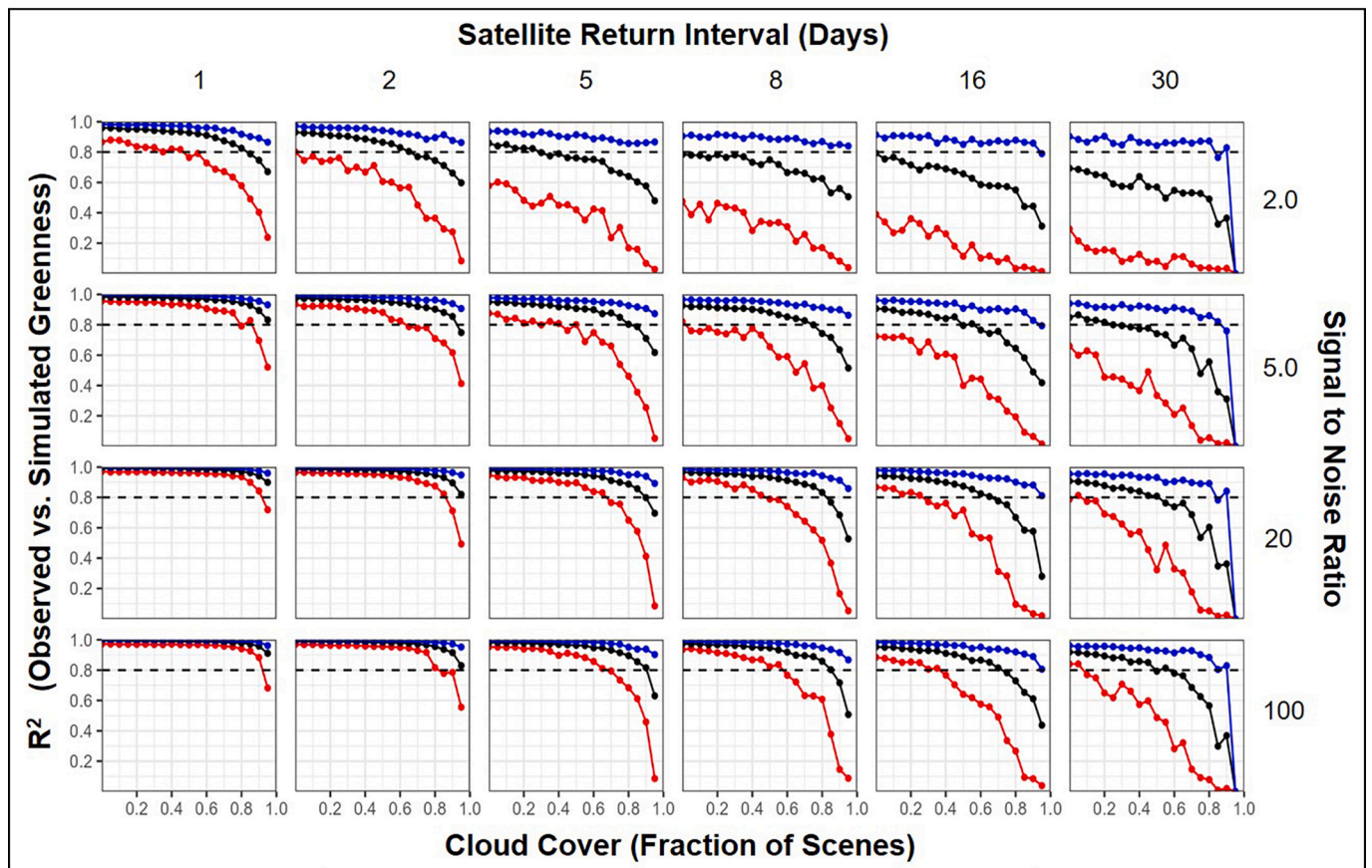
Fig. 7. Comparison of R² values between simulated "true" greenness and greenness retrieved from the phenology estimator after injection of cloud cover and noise, with $n = 30$ and $\sigma_{fixed} = 0.02$ (Table 3). Black lines are the median R² value across all simulations for each test case. Blue and red lines are respectively the upper and lower 5th percentiles of R² values for all tests. Multiple iterations were tested at each combination of imaging parameters because the shape of the phenology curve was varied even as imaging parameters were held constant. Plots are split by return interval (horizontal, from 1 to 30 days return interval) and the signal to noise ratio (vertical, factors of 2 to 100 times as much signal information content as noise). A horizontal dotted line marks R² = 0.8. For help interpreting the quality of a phenology retrieval based on its R² value, see example plots in Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

estimates of absolute elevation and relative elevation were 0.999 and 0.990, respectively. The relationship was weaker for terrain slope ($r = 0.933$), implying a greater sensitivity to fine-scale variation in terrain not captured by the 30 m SRTM product. The analogous r values at MCBCP were 0.999 for absolute elevation, 0.973 for relative elevation, and 0.865 for terrain slope.

3.2. Phenology retrieval simulation

We compared simulated phenology retrievals from a fixed underlying phenology curve and variable sensor parameters (Fig. 6), and also across all combinations of sensor parameters and phenology curves (Fig. 7). Errors are stochastic, and the importance of cloud contamination varies depending on when it occurs during the growing season; for this reason, some retrievals with lower cloud cover failed even as some retrievals with higher cloud cover succeeded (e.g. the bottom row of Fig. 6). Overall patterns of relative performance emerge when comparing across all tests with each parameter set (Fig. 7). The local polynomial regression consistently retrieved the underlying phenology when cloud cover was low even in the presence of substantial noise, but retrievals failed increasingly often with higher cloud cover, or with moderate cloud cover and very high noise (Figs. 6, 7).

There were strong tradeoffs in performance between return interval, cloud cover, and noise. For example, simulations with a 2-day return interval modeled the underlying phenology fairly well even with 90% cloud cover, but with a Landsat-like 16-day return interval the same

amount of cloud cover was much more impactful. Increases in noise reduced the cloud cover thresholds at which phenology retrieval became unreliable. However, within the currently realistic range of values for return interval and noise in surface reflectance data, return interval and cloud cover were the most important predictors of phenology retrieval efficacy.

3.3. Classification results

When comparing classifier effectiveness at FH and MCBP with different input parameters (Tables 3, S2), we found that accuracy was highest at both test sites when all data sources were included (LiDAR, summer reflectance, and phenology). When we excluded canopy height and derived terrain metrics from SRTM instead of LiDAR, we had lower overall classification accuracy (86.2% vs. 99.3% at FH and 79.5% vs. 84.6% at MCBCP). However, the balanced class accuracy for the riparian woodland class changed much less when LiDAR was replaced with SRTM (95.5% vs. 95.6% at FH and 97.2% vs. 98.2% at MCBCP).

When comparing classes derived from a single data type, the best results in both overall accuracy and balanced DRW accuracy occurred when using phenology data, and the worst results were found when only using SRTM. Results were intermediate and similar when using only summer reflectance or LiDAR.

When all data types were used, the most important variables at FH (Table S2) were canopy height (VIP = 100), January greenness (41.3), February greenness (40.7), horizontal distance to the nearest stream

(37.2), April greenness (35.7), and elevation (34.2). When using SRTM instead of LiDAR, the most important variables were again February greenness, April greenness, horizontal stream distance, January greenness, and elevation (although importance and order changed slightly). The most important spectral reflectance bands were generally the SWIR wavelengths, plus green for the test with all data except LiDAR. The most important phenology bands were in winter and early spring (December to April). Most class error was urban pixels misclassified as GR, or MB misclassified as DS and vice versa (Table S3). The highest balanced class accuracies were for agriculture (99.4%), deciduous riparian woodland (97.4%), and evergreen woodland (96.4%), and the lowest were for urban areas (86.0%) and shrubland (88.8%).

At MCBCP, when using all data together, the top four and the sixth most important variables (Table S2) were structural or topographic: elevation (VIP = 100), canopy height (47.3), slope (45.9), relative elevation (26.5), and stream distance (20.3). The next most important variables were all phenology – from most to least important these were November, August, September, October, January, and February greenness. When LiDAR was replaced with SRTM, the importance list changed substantially, and the most important variables were elevation (100), NIR reflectance (30.7), November greenness (28.1), September greenness (24.5), July greenness (21.7), and horizontal stream distance (19.7). The most important spectral reflectance band was usually NIR, and red was consistently the least important. The most important phenology bands were in late winter (January to March) and in late summer or fall (September to November). Most class error was between urban and non-vegetated bare soil pixels, or sage scrub misclassified as riparian scrub, evergreen oak woodland, or chaparral (Table S4). The highest balanced class accuracies were for water (100%), deciduous riparian woodland (98.2%), and grassland (96.5%), and the lowest were for non-vegetated bare soil patches (68.2%, again because of confusion with urban pixels), and sage scrub (73.4%).

At VSFB, we used the same set of input data (SRTM, surface reflectance, and phenology) to classify four years with different precipitation (Tables S2, S5). When classifiers were trained and tested in the same year, overall accuracy and riparian balanced class accuracy were fairly consistent across years (respectively ranging from 87.4% to 91.9% and from 97.4% to 99.1%). The most important metrics on average were elevation (average VIP = 100), green reflectance (64.8), terrain slope (61.6), SWIR1 reflectance (55.7), NIR reflectance (54.3), March greenness (51.2), and February greenness (42.9). Generally, the most important phenological data were from late winter to early spring, and late summer to early fall. NIR and SWIR1 reflectance both had substantially higher importance during the two dry vs. the two wet years (average VIP of 62.5 vs. 46.1 for NIR and 70.9 vs. 40.5 for SWIR1 in dry and wet years, respectively). Class confusion on VSFB was qualitatively similar to MCBCP, which also has a Mediterranean climate (Table S5). Most confusion was between upland classes, particularly those which are evergreen and show little greenness change through the year.

To test the stability of phenology-based classifications through time, we also trained and tested classifiers on imagery from different years (Table S6). Differences in training vs. validation year affected classifier performance by a modest amount. When the model was trained and tested in different years, overall class accuracy declined from an average of 89.8% to 83.9%, while balanced riparian class accuracy declined from 98.5% to 96.3% (both $p < 0.05$, two-sample *t*-test). The decline in accuracy was greater when the data were trained in wet and tested in dry years, or vice versa, and this difference was more pronounced for upland than for riparian classes.

3.4. Regional riparian mapping

In addition to the accuracy statistics presented in Table 5, variable importance and confusion matrices for our regional classifier are provided in the supplement (Table S7). Example maps are included for focal sites (Fig. 9). We initially performed site-level validation of the regional

model on the San Pedro River, testing our ability to discriminate DRW from other vegetative classes using Sentinel-2 and SRTM. With the coarser class designations in the regional model, cottonwood-willow and mesquite were lumped into a single DRW class. Given that constraint, the model performed with high overall accuracy (96.8%) and was able to map the combined deciduous riparian class with 97.2% balanced class accuracy. DRW pixels that were mislabeled as upland included 7.1% of all cottonwood pixels and 4.3% of all mesquite pixels.

We next tested the Google Earth Engine model on Marine Corps Base Camp Pendleton, and this time evaluated its ability to map any type of riparian woodland (RW, evergreen or deciduous) vs. other types of vegetation. Here, the overall accuracy was again high at 97.0%. The riparian woodland specificity was lower at 91.2%, with some non-woodland area lumped into the riparian woodland output. Most of the upland vegetation that was confused for riparian woodland at this site was chaparral (83.6% of all such errors).

Our final site-level validation focused on small creeks in the Santa Ynez Mountains just outside the boundaries of Vandenberg Space Force Base, again using Sentinel-2 imagery. In this context, we first tested the model's performance at mapping any type of riparian woodland (RW, evergreen or deciduous) vs. non-woodland or non-riparian vegetation types. For that test, the balanced riparian accuracy was 95.6%. However, in these narrow mountain streams there was substantial confusion between deciduous and evergreen canopies *within* the riparian woodland class. Most forest canopies that were manually labeled deciduous were classified as evergreen (62%), whereas fewer were classified as deciduous (36%). Nevertheless, deciduous areas classified as evergreen still had higher summer than winter greenness based on the phenology algorithm (mean NDVI increase of 0.186), whereas truly evergreen canopies showed no change or declines in greenness during summer (mean change of −0.1); this difference rendered the two canopy types consistently separable with manual inspection of phenology imagery.

Additional validation using the regional CALVEG dataset and Landsat imagery produced accuracy ranging from 92.5 to 97.9% (Tables 4, S7). When evaluated at the highest level of taxonomic precision computed by the regional model (to classes of DRW, EW, ONV, or CR) the overall accuracy was 91.0%, and the riparian sensitivity and specificity were respectively 99.3% and 93.7%. The lower overall accuracy compared to riparian accuracy reflects the fact that for this test, most confusion was between non-riparian classes (ONV and CR). When mapping DRW vs. all other vegetation types the overall accuracy was higher (97.1%), and riparian user's and producer's accuracy remained high. The same trend was repeated with the validation for any type of riparian woodland (RW vs. Other, 97.5% overall accuracy).

4. Discussion

4.1. Phenology retrievals

We developed a cloud-based utility to retrieve land surface phenology which is generalizable across sensors and can be applied rapidly across broad spatial extents. The approach allows direct comparison across data sources with substantial variation in sampling density. This enables inter-comparison of seasonal trends across different satellite archives, different geographical areas on Earth, and/or analysis of interannual or seasonal time series at the same location. Advantages of the approach include its computational efficiency and its large capacity for parallelization, which enables rapid scaling in space and time. We are not aware of other operational, cloud-based phenology retrieval algorithms which can be applied over large extents (e.g. > 10,000 km^2), in short time periods (e.g. < 30 min) at arbitrary dryland locations or for arbitrary multispectral sensors, including Landsat and Sentinel-2 at their native resolutions.

We tried to compare our phenology retrievals to another open-source, cross-sensor platform which was recently released on Earth Engine – HANTS-GEE (Zhou et al., 2023). HANTS-GEE has a user

interface which facilitates point-based estimates of phenology using a Fourier transform method, and also allows extraction of phenology images similar to those which we produced. The authors used their algorithm to create a global dataset of MODIS land surface phenology at 5 km resolution, and also presented a small set of point-sampled phenology curves from Landsat and Sentinel-2, but they did not publish information on the computational speed or the maximal extent over which the algorithm could be run to reconstruct images at fine spatial resolution, which was our focus. We tried to use HANTS-GEE to extract 10 m monthly phenology data from Sentinel-2 over an area of approximately 12,600 km^2 over Santa Barbara County (corresponding to our VSFB site). Unfortunately, we were not able to do so – the program failed during four consecutive runs, timing out after >900 min in each case, as it exceeded its computational resources and was canceled by the Earth Engine server. By contrast, our algorithm extracted phenology from the same area, sensor, and spatial resolution on the first attempt in 18 min of real-world runtime (93,503 Earth Engine Compute Units). In comparison to other existing methods, we believe that our approach represents a substantial step forward in computational power for retrieval of fine-scale phenology patterns across broad spatial extents.

Our phenology retrievals were consistent when compared across satellites with different spatial and temporal resolution (Figs. 3, S1, S2). Because of differences in overpass timing, imagery from each sensor was occluded by clouds on different dates. The consistent phenology retrievals across sensors despite variable cloud impacts suggests that the algorithm is robust to moderate cloud cover and differences in sensor design and imaging geometry. Simulations showed that cloud cover, return interval, and measurement or algorithmic error are each important constraints on effective phenology retrieval. However, within the range of current realistic parameters for satellite surface reflectance, cloud cover and return interval dominated and showed strong tradeoffs (Fig. 7), while random algorithmic and radiometric noise was partially smoothed out.

Return interval and sensor noise are fixed for a given satellite, while cloud cover varies regionally and temporally. Our simulation results help to constrain the maximum level of cloud cover through which a given satellite can reliably retrieve surface phenology. Generally, for the ideal case with low cloud cover and rapid return intervals, all tests showed good correspondence between the underlying and retrieved phenology curves. As either of these parameters degraded, the phenology retrievals became less reliable (with lower average R^2 and greater variability in R^2 between modeled and underlying phenology curves). Standards for acceptable accuracy should be adopted to determine when and where phenology retrievals can be relied upon. In Fig. 7 we illustrate multiple quantiles for each test case to help identify the worst conditions at which these phenology retrievals will meet the user's judgment and needs. For example, to preserve 95% of pixels with $R^2 > 0.8$, a sensor like Landsat 8 OLI (with low noise and a 16-day return interval) could not tolerate more than approximately 30% overall cloud cover. However, jointly using two Landsat satellites with overlapping coverage (e.g. Landsat 8 and 9) decreases the return interval to 8 days, as does working in the overlap zone between two adjacent Landsat tiles; either approach would allow the model to perform well with up to approximately 50% cloud cover. Sentinel-2, with a 5-day return interval, could reliably retrieve phenology curves even at 60% cloud cover. And finally, a system with higher radiometric noise and algorithmic uncertainty but near-daily imagery, like the PlanetScope constellation, could provide estimates of phenology at >80% cloud cover.

The simulation results are especially important for users considering extension of our methods at mesic sites with greater cloud cover. We focused on drylands, which typically have relatively low cloud cover. For example, in 2021 and 2022, approximately 27% and 28% of Landsat pixels were occluded by clouds at our San Pedro River test site in inland Arizona, which permitted reliable phenology retrievals even with a single Landsat satellite (Fig. 7). Although differences in water availability, temperature, and elevation could drive differences in cloud

cover, we checked for and did not observe any difference in cloud cover between riparian areas and adjacent uplands. However, fog may be a concern for some coastal dryland sites. The Santa Ynez River runs 145 km from inland Santa Barbara County to the Pacific Ocean. At its mouth, clouds and coastal fog in 2021 and 2022 affected 37% and 38% of observations (Fig. S8). By contrast, the city of Lompoc (14 km inland along the same river) is less affected by marine fog and had only 15% and 22% cloud cover in the same two years. In this setting, a single Landsat satellite might struggle to map phenology immediately along the coast but would be sufficient a short distance inland, while Sentinel-2 or two Landsat satellites would be sufficient even at the coast.

Cloud cover can be seasonally autocorrelated. For example, in coastal Southern California, clouds are most common during the months of June and July, while in Southern Arizona monsoon rainstorms bring the greatest cloud cover in August and September. Clouds are most likely to affect phenology retrievals during times of greenness shift, rather than during periods when little change is occurring (whether that be in the leaf-off or leaf-on condition). Consequently, users interested in the effects of clouds on phenology retrieval should consider the cloud cover during periods of leaf flush and senescence. For deciduous riparian woodlands, that means that cloud cover during the early spring and late fall is most critical (e.g. April–May and October–November). As a result, in our focal region, the cloudiest seasons are not aligned with the primary periods of change in riparian leaf cover, reducing the importance of temporal autocorrelation in cloud cover. However, other users are advised to consider the interaction between cloud seasonality and observed phenology carefully. We suggest that users interested in extending our phenology algorithm to new regions use our simulations to determine the required temporal interval to achieve adequate sampling. Users should conservatively reference cloud fractions which are typical of the cloudiest periods coincident with phenological shifts in their region.

4.2. Land surface classification

One advantage of random forest models is the ability to directly report the relative importance of each variable in differentiating between classes (Table S2, S7). Higher importance is afforded to variables which contribute more to the separation of land cover classes. In our analysis, relative variable importance differed across sites and test cases. Vegetation height from LiDAR had high variable importance when it was included, but riparian class accuracy decreased only slightly when topographic information was sourced from SRTM instead and canopy height was excluded (Figs. S6, S8), which is encouraging because SRTM is available globally and LiDAR is not. In all tests except at VSFB, including the regional model, phenological variables had higher importance than summer spectra, and tests using only phenological information outperformed tests using only summer spectra – especially for the DRW class – showing that phenological data have substantial value for mapping vegetative types in drylands. However, this trend in relative importance may not generalize to non-dryland ecosystems, where leaf phenology may vary less across functional vegetation types – for example, in mapping tropical forest in Puerto Rico, Martinuzzi et al. (2012) saw declines in accuracy when replacing LiDAR with SRTM, and did not see an improvement when using multi-temporal Landsat imagery (two vs. one scene). In our analysis on drylands, accuracy was highest when topographic, phenological, and reflectance information was included together, indicating that the three data types are complementary.

For the local models, variable importance was greatest for phenology bands corresponding to periods with rapid shifts in landscape-scale greenness – for example, late winter and late summer at the two sites with Mediterranean climates, and late winter through the monsoon at FH (Table S4). By contrast, the importance of phenology bands was lower during periods with steady and uniformly high or low greenness across vegetation types (including portions of the winter and summer).

Seasonal timing differs regionally and can also shift between years, particularly for upland vegetation in drylands, which often has phenology tied directly to precipitation. At VSFB, summer NIR and SWIR reflectance had higher importance in particularly dry years. These two bands are respectively important in mapping healthy green vegetation and in discriminating dead or senesced vegetation from soils, which may be especially useful under drought conditions. We also saw clear differences in seasonality between dry and wet years, with lower peak greenness and earlier onset of late-season senescence for most vegetation types during dry years (Fig. 5). Deciduous riparian woodlands and some other vegetative types showed differences in the timing of spring greenup between dry and wet years, which could be related to the covariation between temperature and precipitation in winter and spring. Spring leaf out date in many deciduous tree species varies in response partly to winter temperatures, so warm winters under drought could lead to earlier leaf out. Additionally, we observed differences in phenological timing across ecoregions – for example, riparian woodlands in interior desert sites greened up later, senesced earlier, and fell to lower minimum winter greenness values in comparison to coastal ecosystems (Fig. S7). However, broad patterns still separated vegetation types across ecoregions, and our phenology-based classifiers were repeatable and consistently effective even when trained in one year or location and tested in another (Table S6). This was especially true for DRW, which showed less interannual change in phenological timing than the other vegetation types we classified. In most models, error matrices showed that DRW class accuracy was the highest or near highest among all classes, demonstrating the phenological distinctiveness and consistency of this land cover type in dryland environments (Figs. S3 – S5, S7). We suspect that this difference results from the reduced dependence of riparian woodlands on rainfall, because in contrast to most of the landscape, riparian trees are able to access stream-associated groundwater even during the dry season. Year-round access to water means these plants show seasonal cycles constrained primarily by temperature and photoperiod, which are less variable between years than precipitation.

In comparison to previous automated riparian mapping efforts, our classifiers had similar or favorable performance. For example, Salo et al. (2016) used only topographic information to map montane riparian zones, and for all Strahler stream orders they had a maximum kappa statistic across several methods of 0.38, which is similar to our kappa values using only SRTM (0.40 at FH and 0.49 at MCBCP) but much lower than the models using all available data (0.87 and 0.83, respectively). Other authors using multispectral satellite imagery, sometimes in combination with LiDAR, achieved classification accuracies ranging from 87% to 97%, which are similar to the results we report here (Townsend and Walsh, 2001; Jia et al., 2020; Rabanaque et al., 2021). Melichar et al. (2023) also used Landsat-based phenology metrics to map vegetation types in the southwestern United States. They did so offline, and aggregated samples across 8 consecutive years to increase the effective sampling frequency. In comparison to our annualized phenology classifier, their approach was effective for mapping long-term vegetation cover, but is less able to resolve interannual change in cover. This may be important for riparian woodlands, which are frequently disturbed by floods. They achieved similar class accuracy for some riparian vegetation types (F-score of 0.91 for "North American Warm Desert Riparian Mesquite Bosque") but lower accuracy for other, rarer and more sparsely-distributed riparian classes (F-score of 0.45 for "North American Warm Desert Lower Montane Riparian Woodland and Shrubland"). Our work demonstrates the regional scalability of annual phenology-based classifiers, while their broad focus across plant functional types demonstrates the utility and potential for extension of cloud-based phenology methods to map other, non-riparian vegetation types throughout dryland regions.

With local classifiers, we were able to map a broad diversity of classes, including more detailed vegetation classes than were used in the more general regional model. In particular, local classifiers permitted

more subtypes of riparian vegetation to be differentiated, including cottonwood-willow gallery woodland, mesquite bosque, riparian scrub, and wetlands. Some widespread species like the Fremont cottonwood, *Populus fremontii*, differ in phenological timing by as much as several months across their range. To account for seasonal variability resulting from regional differences in climate and photoperiod, we included latitude and elevation in the regional model, which are linked to photoperiod and climate. However, local classifiers can refine the results for focal sites by entirely excluding this range-wide variability.

Local classifiers are generally not extensible outside of the region where they were trained, however, and the laborious process of retraining the classifier for a new site makes them of limited use to application-oriented users such as land managers and conservation planners, who may not have the time or expertise to build a new model specifically for their focal site. In contrast, the regional classifier included fewer vegetative classes, but was still able to map them with accuracy of 92.5% or greater even when tested on regions not included in the training set (e.g. trained in California and tested in Arizona). A major advantage of the regional model is that it can be automatically updated and applied in new areas, which could substantially lower the barrier to use for a greater set of geospatial information users.

The regional model was effective at mapping pure stands of DRW along lowland rivers, and at mapping RW of any type, despite months of variation in the timing of spring and fall across our study area (Fig. S5). However, in complex mixed forests it struggled to separate small, isolated patches of deciduous woodland embedded in larger evergreen forests. This is noteworthy because in drylands, lower-order montane streams often feature a very narrow band of deciduous woodland along the channel surrounded by a larger woodland of evergreen broadleaved and coniferous trees; in many cases, evergreen trees can even occur in the midstory growing under a canopy of deciduous trees (Sands, 1980). In such contexts, our regional model often labeled the ribbon of deciduous woodland along the channel as evergreen woodland along with the surrounding matrix. However, consistent differences in phenology still separated isolated deciduous or mixed areas from purely evergreen stands, and our locally-trained models were able to separate those forest types (e.g. along Huachuca Creek in Fig. 8). Therefore, apart from locally-trained models, another potential solution for handling complex woodlands with sub-pixel mixing among classes is to rely directly on the continuous phenological information itself instead of using discrete class labels – mapping phenometrics such as start and end of season and magnitude of greenness change. Alternatively, fractional cover by deciduous or evergreen species could be mapped at the sub-pixel scale by applying a method like spectral mixture analysis to the phenology data, although this may require design of spectral libraries with temporally varying endmembers (Dudley et al., 2015). These extensions may be especially useful when using spatially coarser sensors like Landsat in environments with particularly narrow riparian areas, like small montane streams. Finally, users could focus on datasets with even finer spatial resolution, such as PlanetScope imagery.

4.3. Contributions and continuing work

By implementing a rapid, open-source, entirely cloud-based phenology retrieval algorithm for Landsat and Sentinel-2, we hope to substantially lower the barrier for ongoing research on phenological patterns, particularly in dryland environments. The climatic, hydrologic, and photoperiod drivers of variability and long-term change in plant phenology have been much studied in deciduous woodlands in mesic environments (Morisette et al., 2009; Piao et al., 2019; Zohner et al., 2023). By contrast, phenology patterns in drylands are relatively poorly characterized because dryland ecosystems tend to have both very fine-scale spatial heterogeneity and extremely flashy, irregular seasonal patterns which are tied to rainfall and streamflow events (Broich et al., 2014; Smith et al., 2019); the former problem is particularly salient in riparian environments. We recommend ongoing work to investigate
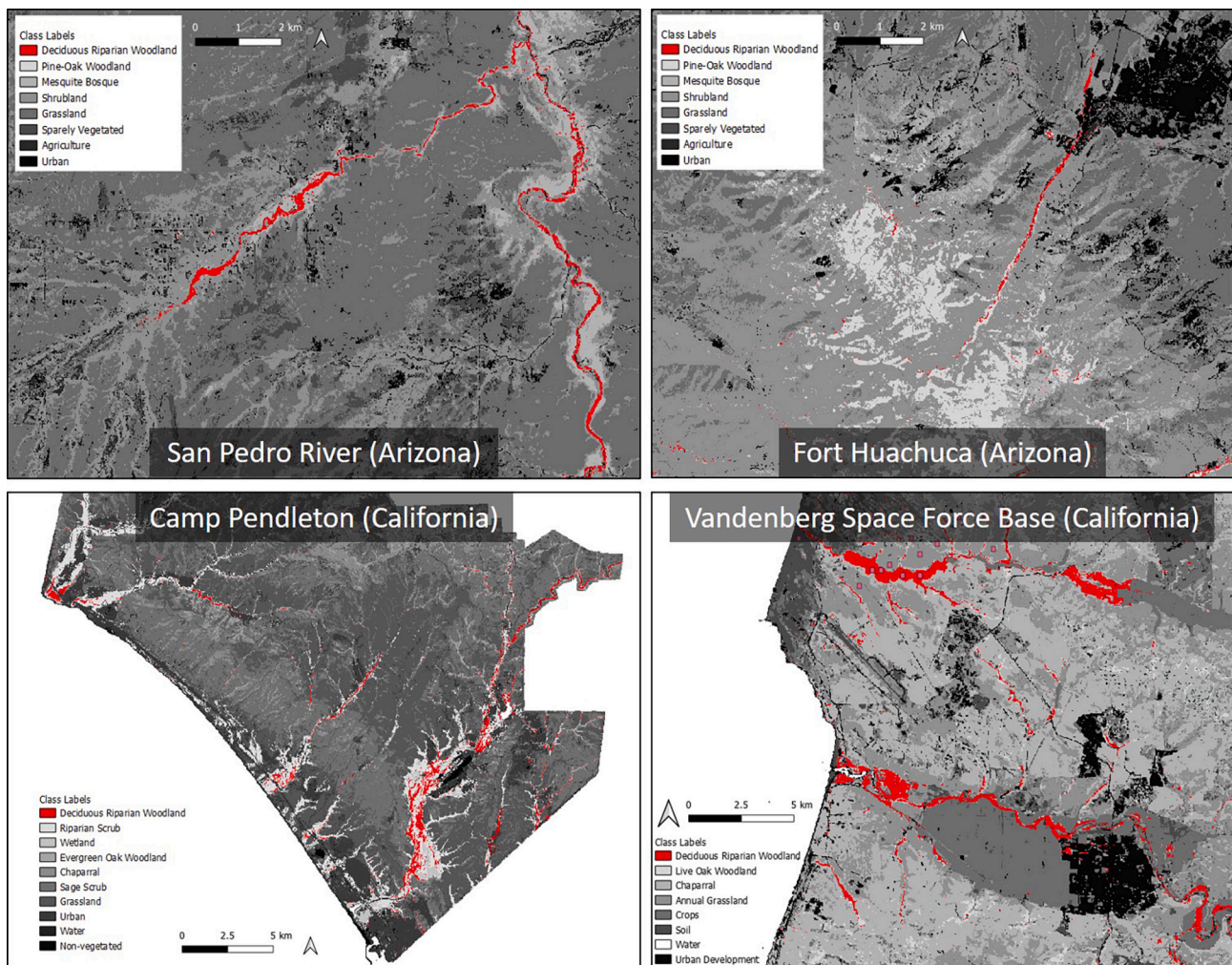
Fig. 8. Mapped vegetation types from local classifiers applied to study sites. Clockwise from upper left: the confluence of the San Pedro and Babocomari Rivers near Fort Huachuca; Huachuca Creek on Fort Huachuca; San Antonio Creek and the Santa Ynez River flowing through Vandenberg Space Force Base; the Santa Margarita River flowing across Marine Corps Base Camp Pendleton. In each set of images, the riparian woodland class is highlighted in red, and other vegetation types are grayscale. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phenological patterns using our method, particularly for vegetation types (such as riparian woodland) which are typically too narrow to effectively model using MODIS.

Our phenology algorithm could also be extended to operate on products other than NDVI. These could include raw surface reflectance values, other spectral indices, fractional cover from spectral mixture analysis, land surface temperature, or any other temporally continuous, numeric product which varies on seasonal timescales and which is mapped on daily to weekly intervals by satellites. Additionally, other functions could be explored for use in the regression, including harmonic fits, which have been used elsewhere for phenology estimates (e. g. Verbesselt et al., 2010; Brooks et al., 2012; Qinchuan, 2013; Zhou et al., 2023). The algorithm could also be extended to directly estimate phenometrics like seasonal start and end dates and annual minima and maxima. As noted above, these could then be used to provide finer detail on complex, mixed habitats such as montane streams, where individual pixels may include multiple different phenological types.

As new vegetation and satellite products become available, other opportunities will emerge for extending our work, as well. VegCAMP is a new field-based vegetation mapping protocol which could be used to augment or replace the older CALVEG data we used here, particularly once current issues are resolved related to lack of unified classification across surveys. Our maps could also be compared to other existing riparian woodland maps which have been generated for specific regional

contexts (e.g. those mapped on the Little Colorado River - Nagler et al., 2022a). Our classifiers which use SRTM or MODIS could be extended to use newer products from platforms like TerraSAR-X and VIIRS. Finally, we believe that over the next decade, the upcoming Surface Biology and Geology Mission and other spaceborne hyperspectral sensors will provide particularly key opportunities for continued application of our phenology retrieval methods.

Our simulations comparing the influence of sensor parameters and cloud cover on phenology retrieval could also be expanded in future work. We assumed that cloud cover was distributed evenly through the year, but researchers working in a particular area may benefit from using temporally autocorrelated cloud distributions which match local meteorological patterns. The simulations could also be extended to feature more robust representations of atmospheric effects and the noise they engender, including via use of the MODTRAN radiative transfer software to separately estimate noise levels in each band (Berk et al., 2014).

Our maps of riparian woodland extent could support diverse lines of scientific inquiry and land or watershed management on local to regional scales. Much recent attention has focused on the remote sensing data record as a tool to demonstrate the adverse effects of climate change, damming, and groundwater extraction on riparian woodlands and their associated fauna (Mayes et al., 2020; Kibler et al., 2021). Our maps complement this existing literature both by allowing future
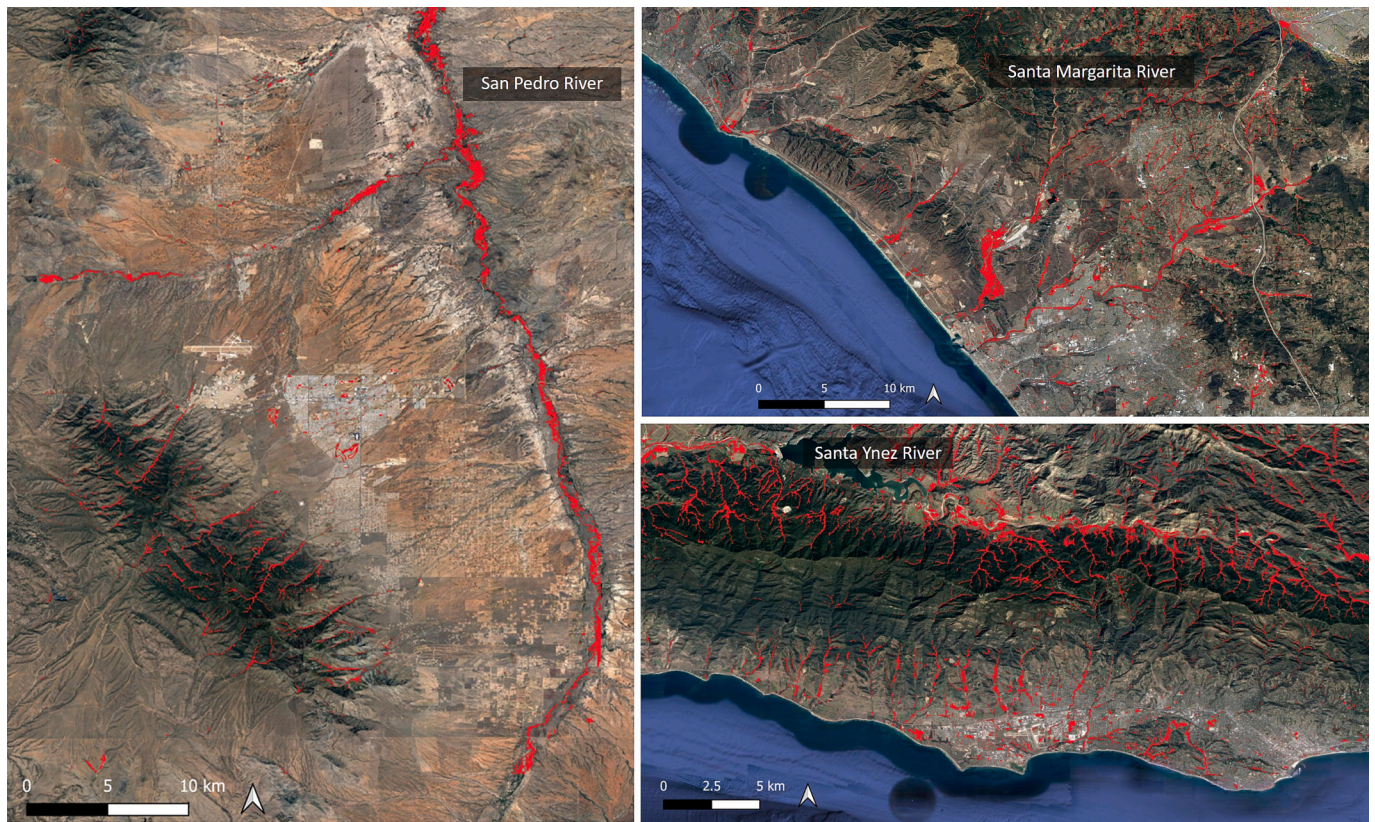
Fig. 9. Examples of riparian woodland maps for each study region. All images show Google truecolor satellite imagery, with areas classified as riparian woodland in the regional Google Earth Engine model overlain in red. The left image tile shows southwestern Cochise County, Arizona, including the San Pedro and Babocomari Rivers, and small streams throughout the Huachuca Mountains. The upper right tile shows Marine Corps Base Camp Pendleton in San Diego County, California, including the Santa Margarita River. The lower right tile shows southern Santa Barbara County, California, with a focus on the Santa Ynez River in the north and small montane creeks throughout the Santa Ynez Mountains to the south. Background: Imagery ©2023 Landsat / Copernicus, Imagery ©2023 TerraMetrics, Map Data ©2023 Google, INEGI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

analyses to focus on particular vegetative types (e.g. DRW vs. other vegetation, such as riparian scrub, wetlands, and bare riverwash, each of which responds to hydroclimatic drivers and disturbance in different ways) and by allowing change in riparian woodland extent to be directly tracked over time. Despite their small spatial area, riparian woodlands in drylands also feature prominently in landscape water and carbon budgets (Jenerette et al., 2009; Swetnam et al., 2017; Ma et al., 2018), so improved maps are also likely to improve the characterization of land cover and vegetation dynamics in watershed- to regional-scale climate and hydrological models. Finally, dryland riparian woodlands are critical habitat for a long list of threatened and endangered species (), and simple estimates of the mean and variability in NDVI are currently utilized by management professionals to map priority habitat for certain listed species (Hatten et al., 2010; Johnson et al., 2017). Our maps present a powerful new tool to inform and improve such habitat-mapping efforts on a regional scale.

5. Conclusion

We developed a new tool for estimating seasonal greenness patterns in vegetation from multispectral satellite imagery, which is relatively insensitive to sensor type and temporal sampling interval, and can be rapidly applied in the cloud to retrieve phenological patterns over large areas at fine spatial scale. We tested the effects of cloud cover by interrogating both real and simulated data and conclude that our approach is applicable to Landsat or Sentinel-2 data in most dryland regions, where cloud cover is relatively low, allowing access to 40 years of phenology patterns at 30 m scale, or 6 years at 10 m scale. We used phenology data, summer surface reflectance, and topography to map

riparian woodlands at several focal sites, and then generalized the approach to a regional model which we trained in California and tested in both California and Arizona. We provide rapidly extensible, cloud-based methods for retrieving phenology patterns and mapping deciduous riparian woodlands in drylands using Landsat and Sentinel-2. We believe these methods improve substantially upon contemporary work in terms of processing speed and the scope over which they can be applied without exhausting computational resources. Phenological timing is critical in deciduous riparian woodlands for migratory wildlife and hydrological budgets, and to ensure alignment of vegetative cycles with water availability for primary productivity, seed dispersal, and establishment of new woodlands. Dryland riparian woodlands have already declined to a small fraction of their historic area, and they continue to face threats from ongoing climate change, groundwater extraction, and river management, so mapping long-term changes in their health and extent is now more important than ever (Katibah, 1984; Swift, 1984; Rood et al., 2003; Jenerette et al., 2009; Stella et al., 2013; Salo et al., 2016; Ma et al., 2018; Albano et al., 2020; Nagler et al., 2020; Nagler et al., 2021; Mayes et al., 2020). We believe that our maps and phenology algorithms provide powerful new tools for both academic study and management of riparian woodlands across global drylands.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2024.114056.

Environmental Research and Development Program. Co-author is an editorial board member for RSE and previously served as an associated editor - Dar Roberts.

CRediT authorship contribution statement

Conor A. McMahon: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft. **Dar A. Roberts:** Funding acquisition, Methodology, Resources, Software, Supervision, Writing – review & editing. **John C. Stella:** Funding acquisition, Supervision, Writing – review & editing. **Anna T. Trugman:** Supervision, Writing – review & editing. **Michael B. Singer:** Funding acquisition, Supervision, Writing – review & editing. **Kelly K. Caylor:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Data availability

Data will be made available on request.

References

Albano, C.M., McGwire, K.C., Hausner, M.B., McEvoy, D.J., Morton, C.G., Huntington, J. L., 2020. Drought sensitivity and trends of riparian vegetation vigor in Nevada, USA (1985-2018). Remote Sens. 12 (9), 1362.

Ballard, G., Burnett, R., Burton, D., Chrisney, A., Comrack, L., Elliott, G., Gardali, T., Geupel, G., Heath, S., Humple, D., Kus, B., Lynes, M., Pitkin, M., Pomara, L., Scoggin, S., Small, S. Stralberg, D., Toniolo, V., 2004. The riparian bird conservation plan: a strategy for reversing the decline of riparian associated birds in California. California Partners in Flight.

Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., van den Bosch, J., 2014. MODTRAN6: a major upgrade of the MODTRAN radiative transfer code. proc. SPIE 9088, algorithms and Technologies for Multispectral, hyperspectral, and Ultraspectral imagery XX 90880H. https://doi.org/10.1117/12.2050433.

Broich, M., Huete, A., Tulbure, M.G., Ma, X., Xin, Q., Paget, M., Restrepo-Coupe, N., Davies, K., Devadas, R., Held, A., 2014. Land surface phenological response to decadal climate variability across Australia using satellite remote sensing. Biogeosciences 11, 5181–5198. https://doi.org/10.5194/bg-11-5181-2014.

Brooks, E.B., Thomas, V.A., Wynne, R.H., Coulston, J.W., 2012. Fitting the multitemporal curve: a Fourier series approach to the missing data problem in remote sensing analysis. IEEE Trans. Geosci. Remote Sens. 50 (9) https://doi.org/10.1109/TGRS.2012.2183137.

Canadell, J., Jackson, R.B., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.D., 1996. Maximum Rooting Depth of Vegetation Types at the Global Scale.

Chen, R., Yin, G., Liu, G., Li, J., Verger, A., 2020. Evaluation and normalization of topographic effects on vegetation indices. Remote Sens. 12 (14), 2290. https://doi.org/10.3390/rs12142290.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20 (1), 37–46. https://doi.org/10.1177/001316446002000104.

Cooper, W.S., 1922. The broad-sclerophyll vegetation of California: an ecological study of the chaparral and its related communities. Carnegie Institution of Washington 89–92.

Copernicus Sentinel Data, 2022. Retrieved from ASF DAAC during December, 2022, processed by ESA.

Currier, C., Sala, O., 2022. Precipitation versus temperature as phenology controls in drylands. Ecology 103 (11). https://doi.org/10.1002/ecy.3793.

de Silveira, E.M.O., De Carvalho, L.M.T., Junior, F.W.A., De Mello, J.M., 2007. The Assessment of Vegetation Seasonal Dynamics Using Multitemporal NDVI and EVI Images Derived from MODIS. In: 2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images. IEEE, Leuven, Belgium, pp. 1–5. https://doi.org/10.1109/MULTITEMP.2007.4293049.

Dennison, P.E., Nagler, P.L., Hultine, K.R., Glenn, E.P., Ehleringer, J.R., 2009. Remote monitoring of tamarisk defoliation and evapotranspiration following saltcedar leaf beetle attack. Remote Sens. Environ. 133 (7), 1462–1472. https://doi.org/10.1016/j.rse.2008.05.022.

Donchyts, G., Winsemius, H., Schellekens, J., Erickson, T., Gao, H., Savenije, H., van de Giesen, N., 2016. Global 30m height above the nearest drainage. European Geosci. Union 2016. https://doi.org/10.13140/RG.2.1.3956.8880.

Doody, T.M., Barron, O.V., Dowsley, K., Emelyanova, I., Fawcett, J., Overton, I.C., Pritchard, J.L., Van Dijk, A.I., Warren, G., 2017. Continental mapping of groundwater dependent ecosystems: A methodological framework to integrate diverse data and expert opinion. J. Hydrology: Regional Studies 10, 61–81. https://doi.org/10.1016/j.ejrh.2017.01.003.

Doody, T.M., Gao, S., Vervoort, W., Pritchard, J., Davies, M., Nolan, M., Nagler, P.L., 2023. A river basin spatial model to quantitatively advance understanding of riverine tree response dynamics to water availability and hydrological management. J. Environ. Manag. 332, 117393 https://doi.org/10.1016/j.jenvman.2023.117393.

Dudley, K.L., Dennison, P.E., Roth, K.L., Roberts, D.A., Coates, A.R., 2015. A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients. Remote Sens. Environ. 167, 121–134.

Dunford, R., Michel, K., Gagnage, M., Piégay, H., Trémelo, M.L., 2009. Potential and constraints of unmanned aerial vehicle technology for the characterization of Mediterranean riparian forest. Int. J. Remote Sens. 30, 4915–4935.

Farr, T.G., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. EOS Trans. AGU 583.

Galvao, L.S., dos Santos, J.R., Roberts, D.A., Breunig, F.M., Toomey, M., de Moura, Y.M., 2011. On intra-annual EVI variability in the dry season of tropical forest: a case study with MODIS and hyperspectral data. Remote Sens. Environ. 115, 2350–2359.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44 (8) https://doi.org/10.1109/TGRS.2006.872081.

Godfrey, J., Lejot, J., Demarchi, L., Bizzi, S., Michel, K., Piégay, H., 2023. Combining hyperspectral, LiDAR, and forestry data to characterize riparian forests along age and hydrological gradients. Remote Sens. 15 (1) https://doi.org/10.3390/rs15010017.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Gunawardena, K., Wells, M., Kershaw, T., 2017. Utilising green and bluespace to mitigate urban heat island intensity. Sci. Total Environ. 584-585, 1040–1055.

Halabuk, A., Mojses, M., Halabuk, M., David, S., 2015. Towards detection of cutting in Hay meadows by using of NDVI and EVI time series. Remote Sens. 7 (5), 6107–6132. https://doi.org/10.3390/rs70506107.

Hatten, J.R., Paxton, E.H., Sogge, M.K., 2010. Modeling the dynamic habitat and breeding population of the southwestern willow flycatcher. Ecol. Model. 221 (13–14), 1674–1686. https://doi.org/10.1016/j.ecolmodel.2010.03.026.

Huete, A.R (August 1988). A soil-adjusted vegetation index (SAVI), Remote Sens. Environ., 25 (3): 295–309. doi:https://doi.org/10.1016/0034-4257(88)90106-X.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Huxman, T.E., Snyder, K.A., Tissue, D., Leffler, A.J., Ogle, K., Pockman, W.T., Sandquist, D.R., Potts, D.L., Schwinning, S., 2004. Precipitation pulses and carbon fluxes in semiarid and arid ecosystems. Oecologia 141, 254–268.

Jenerette, G.D., Scott, R.L., Barron-Gafford, G.A., Huxman, T.E., 2009. Gross primary production variability associated with meteorology, physiology, leaf area, and water supply in contrasting woodland and grassland semiarid riparian ecosystems. J. Geophys. Res. 114 (G4), G04010.

Jia, M., Mao, D., Wang, Z., Ren, C., Zhu, Q., Li, X., Zhang, Y., 2020. Tracking long-term floodplain wetland changes: A case study in the China side of the Amur River basin. Int. J. Appl. Earth Obs. Geoinf. 92 https://doi.org/10.1016/j.jag.2020.102185.

Johansen, K., Phinn, S., Witte, C., 2010. Mapping of riparian zone attributes using discrete return LiDAR, QuickBird and SPOT-5 imagery: assessing accuracy and costs. Remote Sens. Environ. 114, 2679–2691. https://doi.org/10.1016/j.rse.2010.06.004.

Johnson, M.J., Hatten, J.R., Holmes, J.A., Shafroth, P.B., 2017. Identifying western yellow-billed cuckoo breeding habitat with a dual modelling approach. Ecol. Model. 347, 50–62. https://doi.org/10.1016/j.ecolmodel.2016.12.010.

Jönsson, P., Eklundh, L., 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30 (8), 833–845. https://doi.org/10.1016/j.cageo.2004.05.006.

Katibah, E.F., 1984. A brief history of riparian forests in the central valley of California. In: Warner, R.E., Hendrix, K.M. (Eds.), California Riparian Systems. University of California Press, pp. 23–29.

Kibler, C.L., Schmidt, E.C., Roberts, D.A., Stella, J.C., Kui, L., Lambert, A.M., Singer, M. B., 2021. A brown wave of riparian woodland mortality following groundwater declines during the 2012-2019 California drought. Environ. Res. Lett. 16 https://doi.org/10.1088/1748-9326/ac1377.

Kong, D., Zhang, Y., Gu, X., Wang, D., 2019. A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine. ISPRS Journal of Photogrammetry and Remote Sensing 155, 13–24. https://doi.org/10.1016/j.isprsjprs.2019.06.014.

Liao, Z., He, B., Quan, X., 2015. Modified enhanced vegetation index for reducing topographic effects. J. Appl. Remote. Sens. 9 (1) https://doi.org/10.1117/1.JRS.9.096068.

Ma, X., Feng, Q., Su, Y., Yu, T., Deo, R.C., 2018. Characteristics of ecosystem water use efficiency in a desert riparian forest. Environ. Earth Sci. 77 (10), 358.

Maestre, F.T., Benito, B.M., Berdugo, M., Concostrina-Zubiri, L., Delgado-Baquerizo, M., Eldridge, D.J., Guirado, E., Gross, N., Kéfi, S., Le Bagousse-Pinguet, Y., Ochoa-Hueso, R., Soliveres, S., 2021. Biogeography of global drylands. New Phytol. 231 (2), 540–558. https://doi.org/10.1111/nph.17395.

Manning, A., Julian, J.P., Doyle, M., W., 2020. Riparian vegetation as an indicator of stream channel presence and connectivity in arid environments. J. Arid Environ. 178 https://doi.org/10.1016/j.jaridenv.2020.104167.

Martinuzzi, S., Gould, W.A., Vierling, L.A., Hudak, A.T., Nelson, R.F., Evans, J.S., 2012. Quantifying tropical dry forest type and succession: substantial improvement with LiDAR. Biotropica 45 (2), 135–146. https://doi.org/10.1111/j.1744-7429.2012.00904.x.

Matsushita, B., Yang, W., Chen, J., Onda, Y., Qiu, G., 2007. Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: A case study in high-density cypress Forest. Sensors 7 (11), 2636–2651. https://doi.org/10.3390/s7112636.

Mayes, M., Caylor, K.K., Singer, M.B., Stella, J.C., Roberts, D., Nagler, P., 2020. Climate Sensitivity of Water Use by Riparian Woodlands at Landscape Scales. Hydrological Processes, October, pp. 4884–4903.

McMahon, C. (2022). Lidar Survey of the San Pedro River, AZ 2021. National Center for Airborne Laser Mapping (NCALM). Distributed by OpenTopography. doi: 10.5069/G98050T2.

McClellan, 2004. Existing vegetation – CALVEG, Pacific southwest region, CA: USDA-Forest Service. EvegTile03B_99_04_v2.

McMahon, Conor, 2022a. Dirt or Leaf. v.vol. 1.0.0, Zenodo. https://doi.org/10.5281/zenodo.7388625.

McMahon, Conor, 2022b. LiDAR Raster Stats. v.vol. 1.0.0, Zenodo. https://doi.org/10.5281/zenodo.7388629.

Melichar, M., Didan, K., Barreto-Muñoz, A., Duberstein, J.N., Hernández, E.J., Crimmins, T., Li, Haiquan, Traphagen, M., Thomas, K.A., Nagler, P.L., 2023. Random Forest classification of multitemporal Landsat 8 spectral data and phenology metrics for land cover mapping in the Sonoran and Mojave deserts. Remote Sens. 15 (1266) https://doi.org/10.3390/rs15051266.

Morisette, J.T., Richardson, A.D., Knapp, A.K., Fisher, J.I., Graham, E.A., Abatzoglou, J., Wilson, B.E., Breshears, D.D., Henebry, G.M., Hanes, J.M., Liang, L., 2009. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Front. Ecol. Environ. 7 (5), 231–284. https://doi.org/10.1890/070217.

Nagler, P.L., 2022b. Riparian Vegetation Trends in Evapotranspiration. U.S. DOI Remote Sensing Activities, 2022 Riparian Vegetation Trends In Evapotranspiration | Land Imaging Report Site (usgs.gov).

Nagler, P.L., Barreto-Muñoz, A., Chavoshi Borujeni, S., Jarchow, C.J., Gómez-Sapiens, M. M., Nouri, H., Herrmann, S.M., Didan, K., 2020. Ecohydrological responses to surface flow across borders: two decades of changes in vegetation greenness and water use in the riparian corridor of the Colorado River delta. Hydrol. Process. 34 (25), 4851–4883.

Nagler, P.L., Barreto-Muñoz, A., Chavoshi Borujeni, S., Nouri, H., Jarchow, C.J., Didan, K., 2021. Riparian area changes in greenness and water use on the lower Colorado river in the USA from 2000 to 2020. Remote Sens. 13 (7), 1332.

Zohner, C.M., Mirzagholi, L., Renner, S.S., Mo, L., Rebindaine, D., Bucher, R., Palouš, D., Vitasse, Y., Fu, Y.H., Stocker, B.D., Crowther, T.W., 2023. Effect of climate warming on the timing of autumn leaf senescence reverses after the summer solstice. Science 381 (6653). https://doi.org/10.1126/science.adf5098.

Nagler, P.L., Barreto-Muñoz, A., Sall, I., Lurtz, M.R., Didan, K., 2022a. Riparian plant evapotranspiration and consumptive use for selected areas of the little Colorado River watershed on the Navajo nation. Remote Sens. 15 (1), 52.

Nietupski, T.C., Kennedy, R.E., Temesgen, H., Kerns, B.K., 2021. Spatiotemporal image fusion in Google earth engine for annual estimates of land surface phenology in a heterogenous landscape. Int. J. Appl. Earth Obs. Geoinf. 99 https://doi.org/10.1016/j.jag.2021.102323.

Okin, G.S., Gu, J., 2015. The impact of atmospheric conditions and instrument noise on atmospheric correction and spectral mixture analysis of multispectral imagery. Remote Sens. Environ. 164 https://doi.org/10.1016/j.rse.2015.03.032.

Palmer, M., Ruhi, A., 2019. Linkages between flow regime, biota, and ecosystem processes: implications for river restoration. Science 365 (6459).

Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, Xu, Shen, M., Zhu, X., 2019. Plant phenology and global climate change: current progresses and challenges. Glob. Chang. Biol. https://doi.org/10.1111/gcb.14619.

QGIS.org, 2021. QGIS Geographic Information System. QGIS Association. http://www.qgis.org/.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.

Rabanaque, M.P., Martínez-Fernández, V., Calle, Mikel, Benito, G., 2021. Basin-wide hydromorphological analysis of ephemeral streams using machine learning algorithms. Earth Surf. Process. Landf. 47 (1), 328–344. https://doi.org/10.1002/esp.5250.

Ranalli, A.J., Macalady, D.L., 2010. The importance of the riparian zone and instream processes in nitrate attenuation in undisturbed and agricultural watersheds – A review of the scientific literature. J. Hydrol. 389 (3–4), 406–415.

Rood, S.B., Braatne, J.H., Hughes, F.M.R., 2003. Ecophysiology of riparian cottonwoods: stream flow dependency, water relations and restoration. Tree Physiol. 23 (16), 1113–1124.

Rouse, J.W., Hass, R.H., Schell, J., Deering, D., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Res. Technol. Satellite (ERTS) Symp. 1, 309–317.

Rusnák, M., Goga, T., Michaleje, L., Šulc Michalková, M., Máčka, Z., Bertalan, L., Kidová, A., 2022. Remote sensing of riparian ecosystems. Remote Sens. 14, 2645. https://doi.org/10.3390/rs14112645.

Sabathier, R., Singer, M.B., Stella, J.C., Roberts, D.A., Caylor, K.K., 2021. Vegetation responses to climatic and geologic controls on water availability in southeastern Arizona. Environ. Res. Lett. 16 (6), 064029 https://doi.org/10.1088/1748-9326/abfe8c.

Sabathier, R., Singer, M.B., Stella, J.C., Roberts, D.A., Caylor, K.K., Jaeger, K.L., Olden, J. D., 2023. High resolution spatiotemporal patterns of flow at the landscape scale in

montane non-perennial streams. River Res. Appl. 39 (2), 225–240. https://onlinelibrary.wiley.com/abs/10.1002/rra.4076.

Safriel, U., Adeel, Z., Niemeijer, D., Puigdefabregas, J., White, R., Lal, R., Winslow, M., Ziedler, J., Prince, S., Archer, E., King, C., Shapiro, B., Wessels, K., Nielsen, T., Portnov, B., Reshef, I., Thonell, J., Lachman, E., McNab, D., 2005. Dryland Systems. In: Millenium Ecosystem Assessment. Island Press, Washington DC, Ecosystems and Human Well Being.

Salo, J.A., Theobald, D.M., Brown, T.C., 2016. Evaluation of methods for delineating riparian zones in a semi-arid montane watershed. J. Am. Water Resour. Assoc. 52, 632–647.

Sands, A., 1980. Riparian Forests in California: Their Ecology and Conservation: A Symposium, UCANR Publications, pp. 2–4.

Santa Barbara County Flood Control District, 2023. Monthly and Yearly Rainfall Record. retrieved on May 9, 2023 from. https://files.countyofsb.org/pwd/hydrology/historic%20data/rainfall/439mdd.pdf.

Smith, W.K., Dannenberg, M.P., Yan, D., Herrmann, S., Barnes, M.L., Barron-Gafford, G. A., Biederman, J.A., Ferrenberg, S., Fox, A.M., Hudson, A., Knowles, J.F., MacBean, N., Moore, D.J.P., Nagler, P.L., Reed, S.C., Rutherford, W.A., Scott, R.L., Wang, X., Yang, J., 2019. Remote sensing of dryland ecosystem structure and function: progress, challenges, and opportunities. Remote Sens. Environ. 233 https://doi.org/10.1016/j.rse.2019.111401.

Stella, J.C., Rodríguez-González, P.M., Dufour, S., Bendix, J., 2013. Riparian vegetation research in Mediterranean-climate regions: common patterns, ecological processes, and considerations for management. Hydrobiologia 719, 291–315.

Stromberg, J.C., 1993. Root patterns and hydrogeomorphic niches of riparian plants in the American southwest. J. Arid Environ. 94, 1–9.

Stromberg, J.C., Merritt, D.M., 2015. Riparian plant guilds of ephemeral, intermittent, and perennial rivers. Freshw. Biol. 61 (8), 1259–1275. https://doi.org/10.1111/fwb.12686.

Stromberg, J.C., Wilkins, S.D., Tress, J.A., 1993. Vegetation-hydrology models: implications for management of Prosopis velutina (velvet mesquite). Riparian Ecosystems 3 (2), 307–314. https://doi.org/10.2307/1941833.

Swetnam, T.L., Brooks, P.D., Barnard, H.R., Harpold, A.A., Gallo, E.L., 2017. Topographically driven differences in energy and water constrain climatic control on forest carbon sequestration. Ecosphere 8, 4.

Swift, B.L., 1984. Status of riparian ecosystems in the United States. J. Am. Water Resour. Assoc. 20 (2), 223–228.

Townsend, P.A., Walsh, S.J., 2001. Remote sensing of forested wetlands: application of multitemporal and multispectral satellite imagery to determine plant community composition and structure in southeastern USA. Plant Ecol. 157 https://doi.org/10.1023/A:1013999513172.

United States Geological Survey (USGS), 2019. National Hydrography Plus (NHDPlus). Retrieved from. https://www.usgs.gov/national-hydrography/access-national-hydrography-products.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114 (1), 106–115. https://doi.org/10.1016/j.rse.2009.08.014.

Wardlow, B.D., Egbert, S.L.A., 2010. Comparison of MODIS 250-m EVI and NDVI data for crop mapping: A case study for Southwest Kansas. Int. J. Remote Sens. 31 (3), 805–830. https://doi.org/10.1080/01431160902897858.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. central Great Plains. Remote Sens. Environ. 108 (3), 290–310. https://doi.org/10.1016/j.rse.2006.11.021.

Warter, M.M., Singer, M.B., Cuthbert, M.O., Roberts, D., Caylor, K.K., Sabathier, R., Stella, J., 2021. Drought onset and propagation into soil moisture and grassland vegetation responses during the 2012–2019 major drought in Southern California. Hydrol. Earth Syst. Sci. 25, 3713–3729.

Warter, M.M., Singer, M.B., Cuthbert, M.O., Roberts, D., Caylor, K.K., Sabathier, R., Stella, J., 2023. Modeling seasonal vegetation phenology from hydroclimatic drivers for contrasting plant functional groups within drylands of the southwestern USA. Environ. Res.: Ecol. 2 (2).

Wulder, M.A., Loveland, T.R., Roy, D.P., Crawford, C.J., Masek, J.G., Woodcock, C.E., Allen, R.G., Anderson, M.C., Belward, A.S., Cohen, W.B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J.D., Hostert, P., Hughes, M.J., Huntington, J., Johnson, D.M., Kennedy, R., Kilic, A., Li, Z., Lymburner, L., McCorkel, J., Pahlevan, N., Scambos, T.A., Schaaf, C., Schott, J.R., Sheng, Y., Storey, J., Vermote, E., Vogelmann, J., White, J.C., Wynne, R.H., Zhu, Z., 2019. Current status of Landsat program, science, and applications. Remote Sens. Environ. 225, 127–147.

Xin, Q., Olofsson, P., Zhu, Zhe, Tan, B., Woodcock, C.E., 2013. Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data. Remote Sens. Environ. 135, 234–247. https://doi.org/10.1016/j.rse.2013.04.002.

Yan, E., Wang, G., Lin, H., Xia, C., Sun, H., 2015. Phenology-based classification of vegetation cover types in Northeast China using MODIS NDVI and EVI time series. Int. J. Remote Sens. 36 (2), 489–512. https://doi.org/10.1080/01431161.2014.999167.

Yin, H., Tan, B., Frantz, D., Radeloff, V.C., 2022. Integrated topographic corrections improve Forest mapping using Landsat imagery. Int. J. Appl. Earth Obs. Geoinf. 108, 102716 https://doi.org/10.1016/j.jag.2022.102716.

Zhou, J., Menenti, M., Jia, L., Gao, B., Zhao, F., Cui, Y., Xiong, X., Liu, X., Li, D., 2023. A scalable software package for time series reconstruction of remote sensing datasets on the Google earth engine platform. Intern. J. Digital Earth 16 (1), 988–1007. https://doi.org/10.1080/17538947.2023.2192004.