



## Improved mapping of forest type using spectral-temporal Landsat features

Valerie J. Pasquarella<sup>a,b,c,\*</sup>, Christopher E. Holden<sup>a</sup>, Curtis E. Woodcock<sup>a</sup>



<sup>a</sup> Department of Earth and Environment, Boston University, 675 Commonwealth Ave., Boston, MA 02215, United States

<sup>b</sup> Department of Environmental Conservation, University of Massachusetts Amherst, 160 Holdsworth Way, Amherst, MA 01003, United States

<sup>c</sup> Northeast Climate Science Center, University of Massachusetts Amherst, 233 Morrill Science Center, 611 North Pleasant Street, Amherst, MA 01003, United States

### ARTICLE INFO

**Keywords:**

Landsat time series

Forest type

Classification

Spectral-temporal

### ABSTRACT

Multi-spectral imagery from the Landsat family of satellites has been used to map forest properties for decades, but accurate forest type characterizations at a 30-m Landsat resolution have remained an ongoing challenge, especially over large areas. We combined existing Landsat time series algorithms to quantify both harmonic and phenological metrics in a new set of spectral-temporal features that can be produced seamlessly across many Landsat scenes. Harmonic metrics characterize mean annual reflectance and seasonal variability, while phenological metrics quantify the timing of seasonal events. We assessed the performance of spectral-temporal features derived from time series of all available observations (1985–2015) relative to more conventional single date and multi-date inputs. Performance was determined based on agreement with a reference dataset for eight New England forest types at both the pixel and polygon scale. We found that spectral-temporal features consistently and significantly (paired *t*-test,  $p < 0.01$ ) outperformed all feature sets derived from individual images and multi-date combinations in all measures of agreement considered. Harmonic features, such as annual amplitude and model fit error, aid in distinguishing deciduous hardwoods from conifer species, while phenology features, like the timing of autumn onset and growing season length, were useful in separating hardwood classes. This study represents an important step toward large-scale forest type mapping using spectral-temporal Landsat features by providing a quantitative assessment of the advantages of harmonic and phenology features derived from time series of Landsat data as compared with more conventional single-date and multi-date classification inputs.

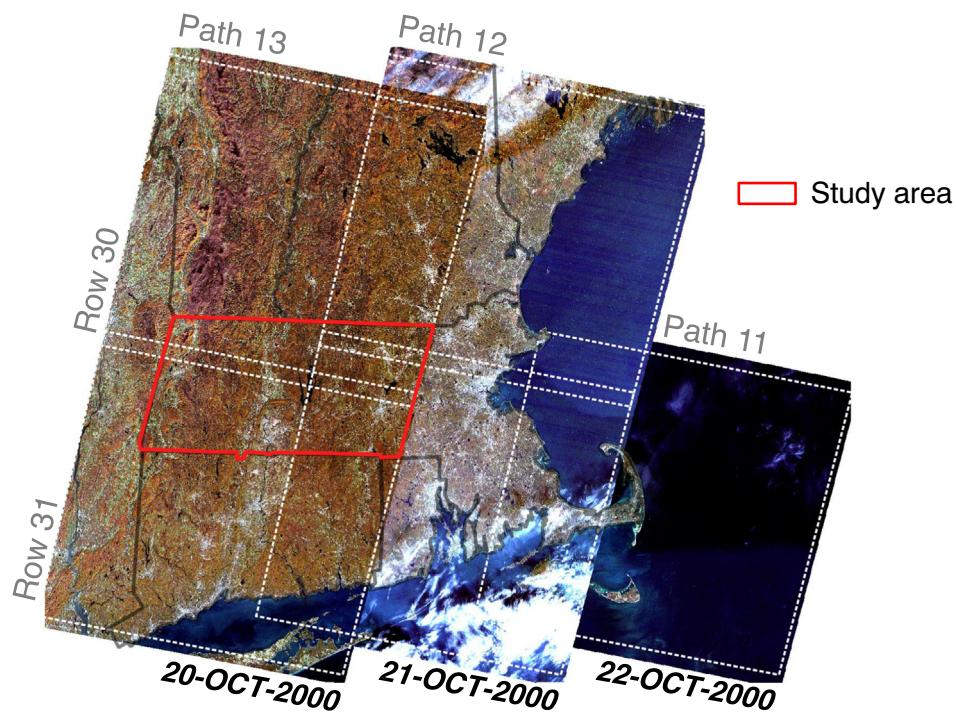
### 1. Introduction

Species composition is a fundamental characteristic of forested ecosystems (Gonçalves, 2017), and maps of forest types with species-level detail are highly desired by researchers, resource managers and policymakers (Iverson et al., 1989; He et al., 1998; Kerr and Ostrovsky, 2003). Since the launch of the first Landsat satellite in 1972, forest resource mapping has been a focus of the Landsat program (Dodge and Bryant, 1976; Iverson et al., 1989; Cohen and Goward, 2004; Wulder et al., 2012). However, prior to the opening of the Landsat archive in 2008 (Woodcock et al., 2008), efforts to map forest types using Landsat data have typically relied on the careful selection of a single image or set of images that maximized spectral differences among tree species during key phenological periods such as leaf-on, leaf-off, spring flush, and autumn senescence (Walsh, 1980; Williams and Nelson, 1986; Wolter et al., 1995; Mickelson et al., 1998; Dymond et al., 2002; Reese et al., 2002; Brown de Colstoun, 2003). While individual images can be used to characterize seasonal variability in spectral reflectance, image

availability and timing has remained an ongoing challenge to using Landsat for large area forest type mapping.

Landsat images are collected once every 8 to 16 days, and each acquisition is subject to contamination by clouds and cloud shadows. This issue is typically overcome by selecting relatively cloud-free images for analysis, or in some cases, through best-available-pixel compositing (Roy et al., 2010; White et al., 2014; Thompson et al., 2015), but when dealing with multi-seasonal image inputs, the timing of observations is critical. Clear images (or pixels, in the case of composites) may not be available for key periods such as spring onset and autumn offset within a given year, and because the timing of these events can vary across years, selecting image dates that best characterize species-specific spectral signatures can be difficult. These challenges are further compounded when working across Landsat orbit paths, which are always offset by at least a day (when there are two sensors in space) and subject to different cloud conditions. Thus, selecting images that are consistent in time and space across Landsat orbit paths is effectively impossible in even moderately cloudy regions,

\* Corresponding author at: Department of Environmental Conservation, University of Massachusetts Amherst, 160 Holdsworth Way, Amherst, MA 01003, United States.  
E-mail addresses: [valpasq@umass.edu](mailto:valpasq@umass.edu) (V.J. Pasquarella), [ceholden@bu.edu](mailto:ceholden@bu.edu) (C.E. Holden), [curtis@bu.edu](mailto:curtis@bu.edu) (C.E. Woodcock).



**Fig. 1.** Example image mosaic for Massachusetts. Images were acquired on three consecutive days in October 2000. Imagery from Path 13 are relatively clear, but imagery from Path 12, acquired one day later, are of much lower quality and show significant cloud contamination. Study area (outlined in red) represents the portion of Massachusetts the lies within Path 13. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

particularly when multiple dates of images from different seasons or years across multiple scenes are required.

Though several large-scale products have been developed for the contiguous United States using multi-date Landsat imagery, including the National Land Cover Dataset (NLCD) and LandFire vegetation layers (Wickham et al., 2014) and tree species range maps (Ellenwood et al., 2015), there is evidence that data from sensors with higher spatial, spectral, and/or temporal resolution offer improvements over Landsat for mapping forest types. For example, Mora et al. (2010) identify tree species based on crown metrics derived from very high spatial resolution (< 1 m) imagery, while Martin et al. (1998) and Plourde et al. (2007) use hyperspectral imagery to characterize forest species composition. Over larger extents, forest type maps have typically been produced using data from coarser resolution sensors with more frequent repeat time such as AVHRR (Zhu and Evans, 1994) and MODIS (Ruefenacht et al., 2008; Wilson et al., 2012), suggesting the importance of the temporal domain for distinguishing among tree species.

The opening of the Landsat archive for free public use in 2008 and subsequent development of algorithms for extracting information from Landsat time series (e.g. Zhu et al., 2012; Melaas et al., 2013; Zhu and Woodcock, 2014; Brooks et al., 2014; DeVries et al., 2015) have created new opportunities to use the Landsat temporal domain to improve classification of forest types. Rather than using a single image or set of images to discriminate among forest types, it is now possible to characterize both spectral and temporal variability in canopy reflectance from dense time series of all available observations (Kennedy et al., 2014; Pasquarella et al., 2016). However, the relative value of spectral-temporal features derived from time series as compared to more conventional single-date and multi-date spectral features for forest type classification has not yet been established.

In this study, we assess the utility of more conventional single-date and multi-date classification inputs relative to that of spectral-temporal features derived from time series of all available Landsat TM/ETM + observations for discriminating among relatively homogenous forest patches in the Northeastern US. We test two key types of spectral-temporal features: *harmonic metrics*, which characterize mean annual reflectance as well as seasonal variability, and *phenological metrics*,

which quantify the timing of seasonal events, such as spring onset, peak growing season, and autumn offset. These features have been developed and applied separately in previous studies (Fisher et al., 2006; Zhu and Woodcock, 2014; Melaas et al., 2013; Melaas et al., 2016), but are combined for the first time in this study for the specific purpose of forest type discrimination. Using agreement with a reference dataset for Massachusetts at both the pixel and polygon scales for validation of results, we test the relative performance of single-date, multi-date, and spectral-temporal inputs. Specifically, we test a series of classifications that use consistent training and testing datasets, but vary the features used as inputs. In addition to overall performance, we also consider class-level agreement and the contributions of various features in discriminating among different forest types. Our goal is not to estimate the accuracy of a forest map or make statistical inference regarding forest area, but rather to determine how agreement with a reference dataset changes as a function of the inputs used. In this context, higher agreement with the reference data is the measure of success used.

## 2. Materials & methods

### 2.1. Study area

Our study area covers the western portion of Massachusetts within Landsat World Reference System 2 (WRS-2) Path/Rows 13/30 and 13/31. Though relatively small in size, Massachusetts a challenging test bed both ecologically and in terms of data processing. Massachusetts lies at the transition between northern and southern temperate forest zones, with latitudinal, elevational and geological gradients influencing variability in forest composition (Westveld, 1956; Hall et al., 2002; Fralish, 2003). This spatial heterogeneity in environmental conditions is further complicated by a long history of human and natural disturbances that have periodically altered the successional state and/or composition of individual forest patches (Bromley, 1935; Foster, 1992; Foster and Motzkin, 1998; Gerhardt and Foster, 2002; Hall et al., 2002). With both local and regional factors determining the age, composition, and structure, forests in this region are particularly challenging to classify at the forest type level.

Five Landsat scenes spanning three WRS-2 Paths and two UTM

zones are required to cover the Massachusetts mainland and islands, and cloud cover can vary significantly from image to image (Fig. 1). Therefore, assembling a mosaic of images with nearly consistent dates from even two adjacent Paths is difficult, if not impossible. We chose to limit our analysis to the western portion of Massachusetts that falls within WRS-2 Path 13. By working within a single Path, we were able to reduce uncertainties that would result from combining images from different dates across Paths.

## 2.2. Forest type reference data

The collection of a reference dataset characterizing tree species composition and/or general forest type would require a substantial field effort that was beyond the scope of this study. Therefore, we chose to rely on an existing forest type dataset produced by the Massachusetts Division of Fisheries and Wildlife (MassWildlife; [MassWildlife Staff, 2013](#)) as reference data. This polygon dataset provides forest community labels for forested stands on individual MassWildlife properties across the state. Polygon labels are based on aerial photo interpretation of common forest types, incorporation of ancillary wetlands data, and ground-truthing by MassWildlife field staff. Aerial photos used for interpretation were collected between 1999 and 2005, and field visits were conducted between 2001 and 2008. Despite the different periods of assessment, it is assumed that this data layer accurately reflects the general forest condition of MassWildlife lands during the first decade of the 21st century ([MassWildlife Staff, 2013](#)).

The original MassWildlife dataset, which is considered public record and can be requested directly from MassWildlife, included 17 forest types, which were defined based on proportions of species at the polygon scale. In an effort to reduce uncertainty associated with mixed types and ensure that polygon labels were representative at the pixel scale, we selected the eight classes defined as > 75% dominated by a single species or assemblage of species for use in our analysis. The class labels and definitions used in this study are shown in Table 1.

To further improve the quality of the reference dataset, reference polygons were filtered based on size and shape. Small polygons tend to be prone to edge effects, while percentage-based class definitions may lead to large numbers of pixels from the non-labeled class in very large polygons. Therefore, polygons smaller than five pixels and polygons larger than 1000 pixels were removed from the analysis. Polygons with a pixel-to-polygon area ratio of > 1.5 (pixel-based area 50% greater than polygon-based area) were also removed, eliminating extremely long, thin polygons, which are most likely artifacts of the hand-digitizing process. To ensure that only undisturbed stands were included, reference data was also filtered to remove pixels identified as having changed prior to the year 2000 by the Continuous Change Detection and Classification (CCDC) algorithm used in the spectral-temporal portion of this analysis (see Section 2.3.2). The final reference dataset used in this study consisted of 2802 polygons covering 145,841 pixels, with an average polygon size of about 5.2 km<sup>2</sup> (58 pixels). The number of polygons and pixels for each forest type is provided in Table 1.

**Table 1**

“Pure” forest classes, extracted from MA Division of Fisheries & Wildlife Land Cover dataset.

Class name	Abb.	Description	Polygons	Pixels
Hardwood swamp	Hsw	Wetland: > 75% hardwood species (single or combined)	496	11,480
Softwood swamp	Ssw	Wetland: > 75% softwood species (single or combined)	311	10,249
Northern Hardwoods	NH	> 75% NH species (i.e. beech, birch, maple, ash, aspen, cherry - single or combined)	889	71,865
Central Hardwoods	CH	> 75% oak-hickory species	548	39,168
White pine	Wp	> 75% softwood (> 75% of softwood component = white pine)	325	14,734
Hemlock/white pine	HeWp	> 75% softwood (> 25% of softwood component = hemlock; 0–75% = white pine)	140	8598
Spruce-Fir	SF	> 75% softwood (> 50% of softwood component = Spruce-Fir)	64	3266
Pitch pine/scrub oak	Pp	> 75% pitch pine overstory; 0–25% oak species	29	2520

## 2.3. Classification inputs

We tested the utility of three types of classification inputs for forest type classification: (1) single-date and multi-date spectral features from individual Landsat images, (2) spectral-temporal features derived from time series of all available Landsat observations, and (3) ancillary datasets, including topography and wetland probability layers. An overview of classification inputs is provided in Fig. 2.

All images used in this analysis were Level 1 precision- and terrain-corrected (L1T) Climate Data Record (CDR) Landsat 4 and Landsat 5 TM, and Landsat 7 ETM+ surface reflectance products with < 80% cloud cover. Landsat CDR products have been orthorectified, radiometrically calibrated, and atmospherically corrected using LEDAPS, and thermal band data have been converted into top of atmosphere brightness temperature ([Masek et al., 2006](#); [Loveland and Dwyer, 2012](#); [Markham and Helder, 2012](#); [U.S. Geologic Survey, 2015](#)). Reflectance data were also transformed into Tasseled Cap Brightness (TCB), Tasseled Cap Greenness (TCG), and Tasseled Cap Wetness (TCW) components using the coefficients provided by [Crist \(1985\)](#). While we acknowledge that many other indices, transforms, and even multi-date differencing approaches could be used to differentiate among forest types (e.g. [Wolter et al., 1995](#); [Maiersperger et al., 2001](#); [Dymond et al., 2002](#)), it was not possible to explore every possible spectral combination for this study; thus, we limited our analysis to the six TM/ETM+ optical bands, the TC transforms of these bands, and brightness temperature data.

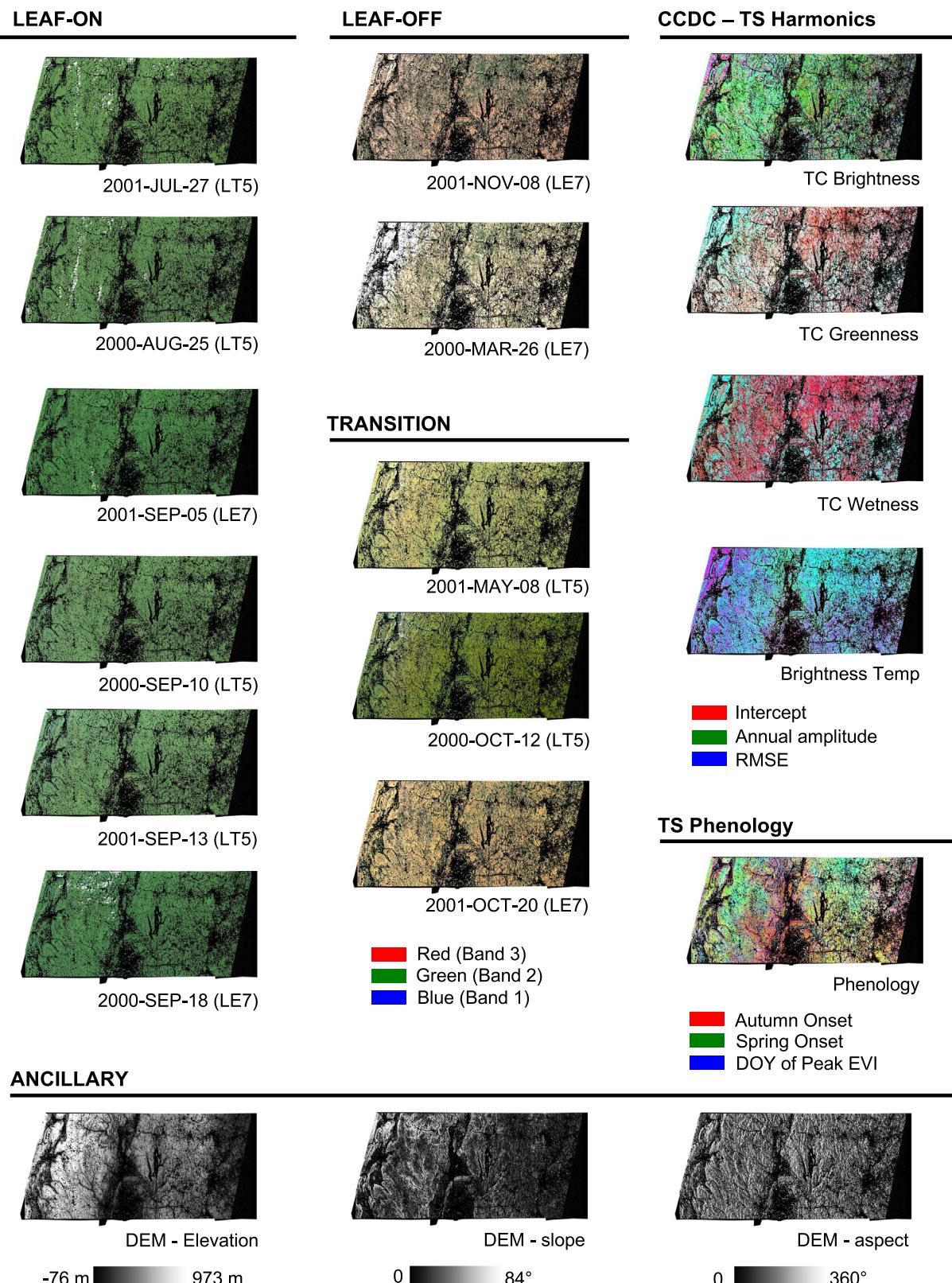
### 2.3.1. Individual images

To approximate more conventional single-date and multi-date classification approaches, we identified all high-quality images from the years 2000 and 2001 to roughly coincide with the timing of the reference data collection while also taking advantage of the SLC-on era of Landsat 7. Images were selected to minimize cloud cover, and the Fmask results from the CDR product were used to remove clouds, cloud shadows, and snow ([Zhu and Woodcock, 2012](#); [Zhu et al., 2015a](#)). The final image dataset included eleven images—six from the growing season/leaf-on period, two from phenological transition periods/shoulder seasons, and two acquired during winter/leaf-off conditions. True color RGB composites of images used in this study are shown in Fig. 2.

### 2.3.2. Spectral-temporal features

Spectral-temporal features were derived from time series of all available observations from 1985 to 2015. We used two different approaches that have been independently developed and validated to extract both harmonic and phenology features from time series of all clear Landsat observations from 1985 to 2015.

A Python implementation of the Continuous Change Detection and Classification (CCDC) algorithm ([Holden et al., 2016](#); described more fully in [Zhu and Woodcock, 2014](#), [Zhu et al., 2015b](#), and [Holden and Woodcock, 2016](#)) was first used to identify stable (non-changing) time series segments and produce harmonic spectral-temporal features. A simple Fourier-style regression model was fit to all available clear



**Fig. 2.** Classification inputs for Western Massachusetts. Non-forested pixels have been masked. Single-date images are shown as true color composites with a consistent color stretch. Spectral-temporal features are also displayed as color composites, with different features assigned to the RGB channels as indicated. Ancillary DEM information is included to show the underlying topographic gradient. Note that individual composites created from spectral-temporal features show a wider range of tones and colors than any individual date of imagery. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

observations for each pixel, resulting in a set of harmonic features that includes an estimated (1) intercept, (2) annual amplitude, and (3) Root Mean Square Error (RMSE) for each spectral band. Because these features correspond to stable time segments, not individual dates of imagery, features can be extracted for any date within the period of the time series. We extracted the harmonic features for the time series segments that intersect July 1, 2000 to coincide with the reference data. We also used CCDC change detection results to remove any pixels that exhibited a time series break indicative of land cover change or forest disturbance prior to July 1, 2000.

In addition to the harmonic CCDC features, we also calculated a series of phenology metrics that quantify forest phenology (Fisher et al., 2006; Melaas et al., 2013). Using a Python implementation of the phenology algorithm initially developed by Melaas et al. (2013) (Holden et al., 2016), Enhanced Vegetation Index (EVI) time series were used to estimate the long-term mean phenology of each pixel. The resulting phenology features include: (1) day of year (DOY) of spring onset, (2) DOY of autumn onset, (3) length of growing season, (4) peak EVI, (5) DOY of peak EVI, and (6) the correlation coefficient of the phenology model.

As illustrated in Fig. 3, harmonic features and phenology features quantify different types of variability in spectral-temporal signatures. The CCDC approach uses the temporal domain to characterize spectral variability, and thus CCDC features are oriented relative to reflectance on the y-axis. In contrast, the phenology algorithm estimates the timing of key phenological events for each pixel. These phenology variables are best thought of as being oriented relative to DOY (time) on the x-axis, with the exception of the peak EVI, which is the EVI value corresponding to the DOY of peak EVI. By using both CCDC and phenology metrics, the annual variability in both the spectral and temporal domains is characterized, providing two major “axes” within each spectral band for discriminating among forest types. When visualized spatially, spectral-temporal feature composites tend to show more pronounced spatial variability within forested areas compared to individual image dates (Fig. 2).

### 2.3.3. Ancillary datasets

Numerous studies have shown that incorporation of ancillary

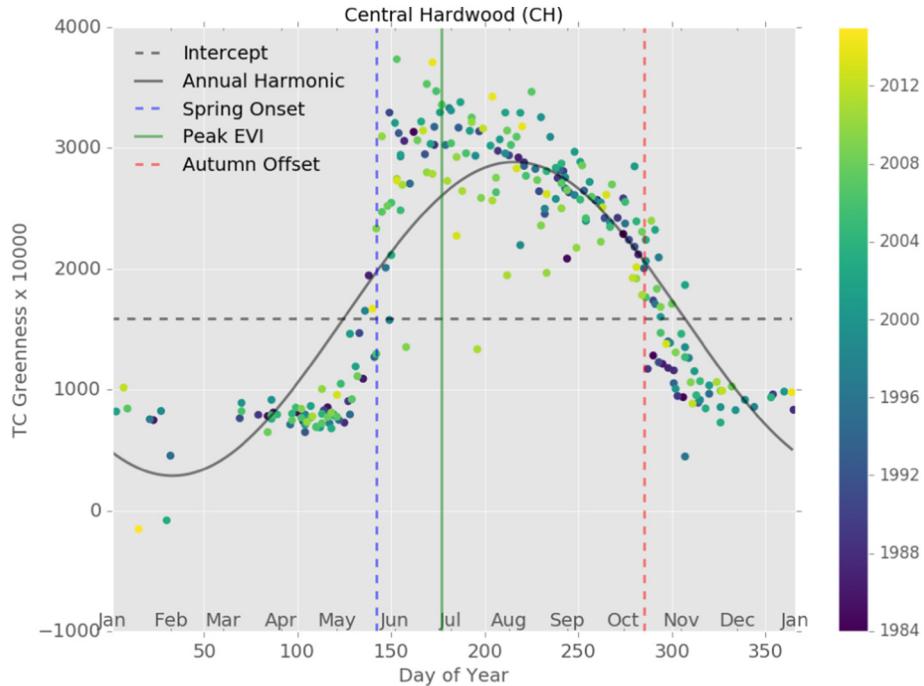
classification features improves mapping of forest communities (Franklin et al., 1986; Sader et al., 1995; He et al., 1998; Khatami et al., 2016), therefore we used two additional datasets to assess how spectral-temporal features perform in combination with ancillary information. The first ancillary dataset is a Digital Elevation Model (DEM), from which we derived slope, aspect, and a transformation of aspect, cos (aspect - 45°), that characterizes the north-south orientation of slopes (Franklin et al., 1986) (Fig. 2). The DEM was obtained from the Massachusetts Office of Geographic Information Systems (MassGIS) data portal.

NOAA's C-CAP National Wetlands Potential dataset was also used to provide the wetland likelihood for each pixel (NOAA, 2014). This nationally standardized, 30 m raster dataset was developed through a modeling process combining multiple GIS and remote sensing data sets including soil characteristics, elevation, existing wetland inventories, hydrographical extents and Landsat 7 imagery. Given the well-recognized challenge of classifying forested wetlands, we anticipated this data layer would improve separability of the hardwood and softwood wetland types from upland forest communities.

### 2.4. Forest type classification

A series of classifications were used to systematically compare the utility of individual image dates, spectral-temporal features, and ancillary data for discriminating among eight relatively homogenous forest types native to New England. Single-date classifications used ten spectral features (TM/ETM+ Bands 1–7, TCB, TCG, TCW) from a single image. Multi-date pairings increased information content by including seasonal variability and doubling the feature set used for classification. The spectral-temporal feature set included intercepts, annual amplitudes, and RMSE for ten bands/indices (TM/ETM+ Bands 1–7, TCB, TCG, TCW) and the six phenology metrics previously described, for a total of 36 spectral-temporal features.

A Random Forest (RF) classifier (Breiman, 2001) was used to assign forest type labels to pixels based on the given combination of feature inputs. The RF family of classifiers have become increasingly common in remote sensing applications due to their flexible, non-parametric nature and ability to limit overfitting (Belgiu and Drăguț,



**Fig. 3.** Harmonic and phenology spectral-temporal features. Harmonic intercept and annual amplitude quantify properties related to reflectance, while phenological features quantify properties related to the timing of seasonal transitions.

2016; Gómez et al., 2016), and have been shown to generally outperform a number of other classifiers across a variety of datasets (Fernández-Delgado et al., 2014). We employed a Python implementation of the RF classifier (Pedregosa et al., 2011), building ensembles of 500 trees, as per the default proposed by Belgiu and Drăguț (2016). To account for the unequal class sizes in our reference dataset, we set class weights to be proportional to class frequencies in the input data, which has been shown to improve RF performance with imbalanced datasets (Chen et al., 2004). We employed a winner-takes-all hard classification approach, where each pixel was assigned to a single forest type for the purposes of assessing agreement, though we also output RF probabilities for individual classes.

### 2.5. Agreement assessment

Agreement between RF results and the MassWildlife forest type reference dataset served as our measure of feature set performance. To generate a mean and standard deviation for agreement scores, thirty RF classifications were generated for each feature set. For each classification, training data were selected a three-fold cross-validation strategy was used where two-thirds of the reference data were used for training and one-third reserved for testing. Given the inherent autocorrelation of pixels derived from the same polygon in the reference dataset, folds were assigned based on polygon ID, ensuring that all pixels from a given polygon were placed in the same fold (Friedl et al., 2000). Fold assignment was determined by a random number generator within each run, but the random seed was kept consistent such that classifications for all feature sets used the same combinations of training and testing data. Because agreement scores can vary substantially depending on the metric used, we assessed the quality of classification results using three commonly used metrics, specifically (1) out-of-bag (OOB) agreement, (2) overall agreement, and (3) area-weighted agreement.

OOB agreement scores, which apply the bagging approach of the RF algorithm to internally assess classifier performance, are intended to provide an estimate of error with little additional computing effort (Breiman, 1996; Belgiu and Drăguț, 2016). Many remote sensing studies have utilized internally produced OOB scores as a measure of classification agreement or accuracy (e.g. Rodriguez-Galiano et al., 2012; DeVries et al., 2016; Thompson et al., 2015), and when training data is comprised of independent samples, OOB estimates can provide a reasonable estimate of agreement. Yet in many reference datasets, like the forest type polygons used here, training data are defined at a patch or polygon scale, and pixels drawn from the same patch or polygon are likely to exhibit a large degree of spatial autocorrelation. The OOB agreement scores are calculated internally for each fold and may include both training and testing pixels from the same polygon. Therefore, the OOB estimates of agreement are included primarily to demonstrate the effect of autocorrelation on agreement scores.

Measures of overall agreement and area-weighted agreement were calculated using independent testing data as part of our three-fold cross validation. To account for patch autocorrelation, we selected training and testing folds based on polygon ID, ensuring that all pixels from a given polygon were assigned to the same fold. In the case of overall agreement, each site (e.g. pixel, polygon) equally contributes to the final measure of agreement, while in the case of area-weighted agreement, the contribution of each site is inversely proportional to the total number of sites in that class, such that each class (rather than each site) contributes equally to the final measure of agreement. Therefore, area-weighted agreement better reflects classification performance for both rare and common classes, and as the number of sites labeled correctly in each class increases, area-weighted agreement will approach overall agreement.

Because of the difference in scale between the reference dataset and classification features/predictions, agreement was assessed at both the pixel and polygon scale. Out-of-bag, overall and area-weighted agreement were calculated at the pixel scale, and overall agreement was

assessed at the polygon scale using a plurality rule. To confirm whether observed differences in agreement across classifications using different feature sets were statistically significant, we used a series of paired t-tests to compare agreement scores. Each individual single-date, multi-date, and spectral-temporal classification score was paired with each of the other classification scores to test for statistically significant differences in mean level of agreement. Paired F-tests revealed that most classifications had statistically significant differences in variance, therefore Welch's t-test, which does not assume equal population variance, was used for all comparisons.

As a final step in our assessment, confusion matrices were generated to examine both the nature and magnitude of disagreement between the reference and predicted labels for individual classes. Mean and standard deviation confusion matrices were calculated for each feature set, and mean confusion matrices were also used to estimate user's and producer's agreement scores.

### 2.6. Feature importance assessment

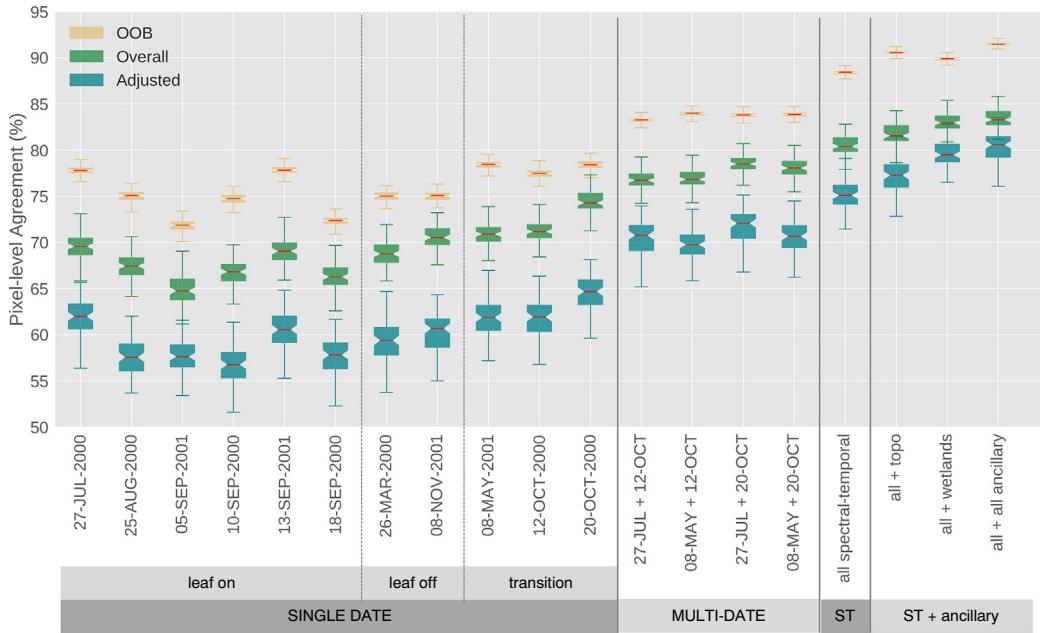
In addition to considering differences in agreement across feature sets, we also sought to determine which spectral-temporal features were most important for classification of different forest types. Feature importance scores were calculated internally by the RF algorithm using the Gini criterion. The Gini importance score describes the relative importance of features and has been widely used as a measure of feature importance in remote sensing. For example, Corcoran et al. (2013) use the Gini criterion to assess the significance of remotely sensed and other ancillary input data for mapping wetlands and classifying wetland type in Northern Minnesota.

High Gini scores correspond to features that are consistently found more often and higher up in the splits of individual decision trees. Given that the importance of features may vary from tree to tree, we computed the overall feature importances of the forest, as well as inter-tree variability. Because strong correlations among individual Landsat bands would be expected to lower Gini scores for features with similar information content, the full spectral-temporal feature set was reduced to include only TCB, TCG, TCW, and thermal harmonic features and phenology features.

A series of additional RF classifications using this reduced feature set were performed to assess feature importances for the following combinations of forest classes: (a) all eight classes (for reference), (b) Central Hardwoods paired with White pine (CH vs. Wp), which frequently co-occur in the eastern portion of the study area, (c) Northern Hardwoods paired with Spruce-Fir (NH vs. SF), which frequently co-occur at the higher elevations in the western portion of the study area, (d) Northern Hardwoods and Central Hardwoods (hardwoods only), (e) White pine, Hemlock-White pine, Spruce-Fir, and Pitch pine (conifers only), and (f) Hardwood swamp and Softwood swamp (wetlands only). For each combination of classes, we constructed a single RF with 1000 trees using only data from the specified classes for training. For example, the CH vs. Wp feature importance classification was performed using an RF trained only on pixels from Central Hardwoods and White pine polygons. This approach allowed us to consider how feature importances differed for discriminating among specific forest types.

### 2.7. Mapped results

In order to explore spatial patterns in forest type predictions, the trained RF classifiers used in this study were used to generate wall-to-wall maps of categorical forest type and class probabilities. The time and expertise required to produce a validation dataset was beyond the scope of this study, therefore we were unable to validate mapped results following good practices for accuracy assessment and area-based estimation (Olofsson et al., 2014). However, we include mapped results for the best-performing feature set to show the predicted distributions of the eight forest types considered in this study.



**Fig. 4.** Distribution of three pixel-scale agreement scores (Out-of-Bag [OOB] from Random Forest and independently assessed overall and area-weighted agreements), measured as percentage correct relative to reference labels. Results are displayed as notched box plots, with the median shown as a red line, the 95% confidence intervals around the median indicated by the notches, and maximum and minimum agreement values shown by the ticks. Results are grouped by classification type. Note: ST = spectral-temporal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3. Results

#### 3.1. Pixel-scale agreement

Pixel-scale agreement assessments exhibited a large range of variability across classifications using single-date, multi-date, and spectral-temporal feature sets (Fig. 4). The OOB agreement score for each classification is consistently higher than the overall agreement score and exhibits a much smaller range of variability than overall and area-weighted measures of agreement. Of the two independently assessed measures of agreement, overall agreement scores were consistently higher than area-weighted agreement.

Agreement scores for single-date feature sets exhibited a large range of variability, and we found that images from the transition seasons (i.e. spring onset and autumn offset) significantly outperformed all leaf-on imagery. While there was no significant difference between the late-fall (November 8, 2001) leaf-off image, the early spring image (May 8, 2001), and the early autumn image (October 12, 2000), the late autumn image (October 20, 2000) significantly outperformed all other single-date classifications with an overall agreement  $74.4 \pm 1.3\%$  and an area-weighted agreement of  $64.4 \pm 2.0\%$ . This image, which was acquired during peak fall foliage and has cloud cover of < 2%, represents the ideal scenario for single-date classification, as phenological differences among both deciduous and coniferous forest types and the two hardwood classes (*Northern Hardwoods* and *Central Hardwoods*) are maximized.

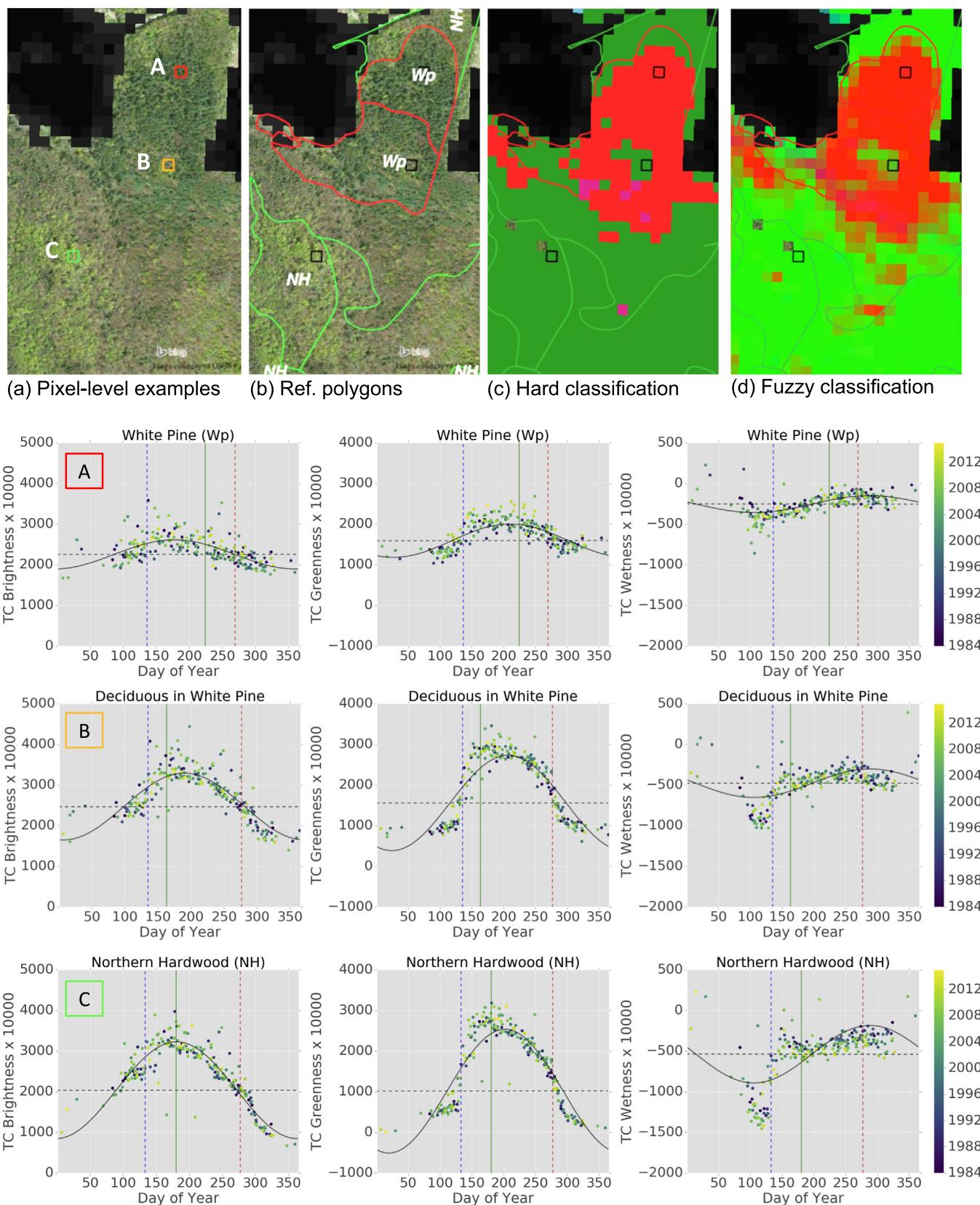
Multi-date feature sets consistently and significantly (paired *t*-test,  $p \ll 0.01$ ) outperformed all single-date feature sets. While many other image pairings would be possible, we chose to compare the best performing leaf-on image (July 27, 2000) with each of the autumn transition images to assess potential added information from the growing season. We also paired the May 8, 2001 image with both October images to assess the utility of information from both transitional periods (autumn and spring). As would be expected based on the single-date results, pairings with the ideal late-autumn image (October 20, 2000) produced higher levels of agreement compared with pairings with a slightly earlier autumn image (October 12, 2000). Furthermore, we observed that pairings of autumn imagery with leaf-on imagery

produced significantly higher results than pairings with early spring imagery, suggesting there may be redundancy in the information content of features from different transition periods. Overall, the highest multi-date agreement scores were for the July 27 and October 20, 2000 pairing, with an overall agreement of  $78.5 \pm 1.2\%$  and an area-weighted agreement of  $71.6 \pm 2.1\%$ .

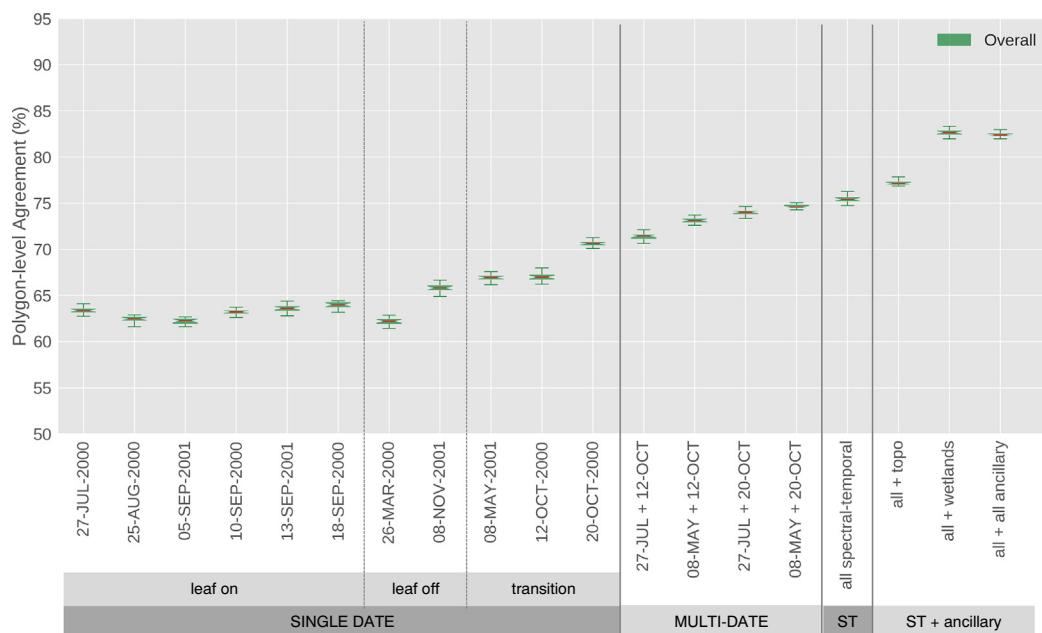
The best results were achieved for spectral-temporal features, which consistently and significantly (paired *t*-test,  $p \ll 0.01$ ) outperformed all features sets derived from individual images and multi-date combinations in all measures of agreement. Using spectral-temporal features alone, we achieved a mean pixel-scale overall agreement of  $80.50 \pm 2.23\%$  and an area-weighted agreement of  $75.17 \pm 3.58\%$ . When ancillary data are added, pixel-scale agreements significantly increased (paired *t*-test,  $p \ll 0.01$ ) to  $83.39 \pm 2.31\%$  overall agreement and  $80.41 \pm 3.44\%$  area-weighted agreement. The addition of the wetlands probability layer resulted in a notable decrease in the difference between overall and area-weighted agreements that can be attributed to improved discrimination of forested wetlands and thus improved class-level agreements for wetland types.

#### 3.2. Polygon-scale agreement

Because the decision rules used to create the reference dataset (Table 1) were defined at the polygon scale and typically represent species or assemblages that are > 75% dominant in the canopy, there is the possibility that forest types within the reference polygons are not homogenous. Inspection of mapped predictions revealed that classification results may be more accurate at the pixel scale than the corresponding reference polygons, as illustrated in Fig. 5. In Fig. 5(b) we show a set of example polygons from the reference dataset. Though the hand-digitized polygon boundaries describe the general pattern of forest patches and adhere to the class decision rules outlined in Table 1 (i.e. *White pine*  $\geq 75\%$  of polygon), the pixel-based results shown in Fig. 5(c) and (d) appear to show sub-polygon detail in dominant forest type. Results of the eight-class hard classification reveal areas of *Northern Hardwoods* within the reference polygons (Fig. 5(c)). Examination of fuzzy results representing class probabilities of deciduous, coniferous, and wetland types reveals even more complex patterns,



**Fig. 5.** Pixel versus polygon labeling and spectral-temporal signatures. Map (a) shows the location of three reference pixels. Map (b) shows the boundaries and labels of polygons used for training and testing. Map (c) shows the output of the eight-class hard classification, while (d) is a composite of fuzzy class probabilities for hardwood (green), conifer (red), and forested wetland classes (blue). Classifications were produced using all spectral-temporal features, topography and wetlands ancillary datasets. Spectral-temporal signatures for three example pixels illustrate differences among spectral-temporal features across hardwood and conifer forest types. The first column shows observed data and model fits in Tasseled Cap Brightness, the second in Tasseled Cap Greenness and the third in Tasseled Cap Wetness. Deciduous-dominated pixels (*Northern Hardwood* and *Northern Hardwood* within a *White pine* stand) exhibit greater intra-annual variability compared to the conifer-dominated (*White pine*) example. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Distribution of polygon-scale overall agreement scores, measured as percentage correct relative to reference labels. Results are displayed as notched box plots, with the median shown as a red line, the 95% confidence intervals around the median indicated by the notches, and maximum and minimum agreement values shown by the ticks. Results are grouped by classification type. Note: ST = spectral-temporal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Mean of polygon-scale confusion matrices. Values correspond to number of polygons correctly or incorrectly labeled based on agreement between reference data and the plurality class as determined by pixel-scale results.

	Hsw	Ssw	NH	CH	Wp	HeWp	SF	Pp	Map total
Hsw	386.87	10.8	16.23	4.57	1.4	0.17	0	0	420.0
Ssw	15.37	260.2	0.67	0	2.17	2.83	2.9	0	284.1
NH	55.03	15.57	778.83	46.6	20.23	20.53	23.57	1.87	962.2
CH	26.3	0.1	84.4	492.37	15.17	5.03	0.93	1.1	625.4
Wp	11.93	6.23	5.77	2.57	267.07	37.77	2.17	4.5	338.0
HeWp	0.5	10.13	3.07	1.3	17.5	71.37	0.73	2.1	106.7
SF	0	7.97	0.03	0	1.23	1.7	33.7	0	44.6
Pp	0	0	0	0.6	0.23	0.6	0	19.43	20.9
Ref. total	496.0	311.0	889.0	548.0	325.0	140.0	64.0	29.0	2802.0

capturing gradients rather than hard boundaries between these high-level thematic classes. Seasonal profiles of TCB, TCG, and TCW and corresponding spectral-temporal features visualized for three example pixels support the classification results, with pixels predicted as *Northern Hardwoods* within the *White pine* polygons (e.g. Example B) exhibiting features that are more similar to nearby pixels within a *Northern Hardwood* polygon (e.g. Example C) than to other pixels within the *White pine* polygon (Example A).

The polygon-based agreement assessment, which is based on a plurality rule, showed much less variability than assessments at the pixel scale (Fig. 6). More similar scores across leaf-on feature sets suggest that polygon-scale aggregation of pixel-based results has an expected smoothing effect. Leaf-off and transition season imagery maintained similar patterns, though differences at the polygon scale are perhaps even more apparent than at the pixel scale. The late-fall (October 20, 2001) image continued to significantly outperform all other single-date feature sets at the polygon-scale, and resulted in higher levels of agreement when combined with imagery from other phenological periods.

Overall agreement using spectral-temporal features at the polygon scale was  $75.48 \pm 0.70\%$  (95% CI), which is surprisingly lower than overall agreement at the pixel scale. However, the polygon-scale overall agreement was found to be greater than area-weighted agreements at the pixel scale. The impact of ancillary data appears to be greater at the

polygon scale, with the addition of two ancillary datasets resulting in a 7% improvement in agreement, to  $82.44 \pm 0.64\%$ . It should be noted, however, that agreement achieved using all ancillary data is not significantly different (paired *t*-test,  $p = 0.08$ ) than using only the wetlands probability dataset. In general, we found that the polygon-scale results were consistent with the pixel-scale results in that classifications using spectral-temporal features produced significantly greater (paired *t*-test,  $p \ll 0.01$ ) agreement scores than all other single-date and multi-date feature sets.

### 3.3. Agreement by forest type

Confusion matrices provide insight into both the nature and magnitude of disagreement between the reference and predicted labels for individual classes. Though the mean and standard deviation confusion matrices were generated for all feature set classifications, we present only the results for the best-performing feature set at the polygon scale (Tables 2 and 3). We also visualize this data as producer's and user's agreement (Fig. 7(a) and (b)).

In general, we observed relatively high agreement for hardwood and forested wetland classes, but lower agreement for conifer classes. There was substantial confusion between the *Hemlock-White pine* and *White pine* classes across both measures of agreement, which is to be expected given overlapping class definitions. In terms of Producer's

**Table 3**

Standard deviation of polygon-scale confusion matrices. Values correspond to number of polygons correctly or incorrectly labeled based on agreement between reference data and the plurality class as determined by pixel-scale results.

	Hsw	Ssw	NH	CH	Wp	HeWp	SF	Pp
Hsw	4.01	1.14	1.56	1.58	0.8	0.37	0	0
Ssw	1.91	1.72	0.7	0	0.73	0.78	0.91	0
NH	4.09	1.33	3.92	3.43	1.48	2.17	1.71	0.56
CH	2.38	0.3	3.05	2.85	1.57	1.11	0.25	0.54
Wp	1.21	1.36	1.05	0.76	3.39	2.56	0.64	1.15
HeWp	0.62	1.56	0.96	0.59	3.22	3.55	0.57	0.6
SF	0	1.05	0.18	0	0.62	0.64	1.83	0
Pp	0	0	0	0.49	0.42	0.49	0	1.23

agreement, the greatest confusion was between *Spruce-Fir* and *Northern Hardwoods*, which are known to co-occur in the western portion of the study area. There was also notable confusion between the two pine-dominated classes, *Pitch pine* and *White pine*. In terms of User's agreement, the greatest confusion was between *Spruce-Fir* and *Softwood swamp*, classes that likely have relatively similar species composition in some cases. Thus, despite confusion between taxonomically related and/or co-occurring classes, the results of class-level agreement assessment are promising, and suggest that spectral-temporal features provide stable feature sets capable of distinguishing among forest types more accurately and at a greater level of thematic detail than currently possible using multi-date imagery.

#### 3.4. Feature importances

Feature importances varied considerably depending on the classes used in training the RF classifier (Fig. 8). When the classifier was trained on the full set of eight forest types, the annual amplitude of TCG had the highest overall importance (0.10), followed by the annual amplitude of TCW (0.09) and the intercept of TCW (0.08). However, the overall range of importances remained relatively low, with all features exhibiting an importance of at least 0.03.

Differentiation among feature importances was more apparent when fewer classes were used in training. When considering only co-occurring hardwood and conifer forest type (e.g. *Northern Hardwoods* paired with *Spruce-Fir* and *Central Hardwoods* paired with *White Pine*),

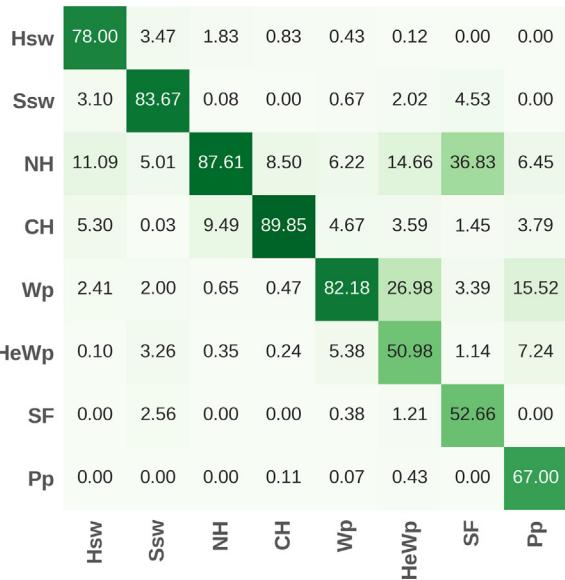
the RMSE of the TCG harmonic fit and the annual amplitude of TCG are the most important features. Additionally, the length of the growing season appears to be more important for separating *Spruce-Fir* and *Northern Hardwoods* examples, while the annual amplitude and RMSE of TCW are more important for separating *White pine* and *Central Hardwoods* examples. When considering only the two deciduous forest types (*Northern Hardwoods* and *Central Hardwoods*), the timing of autumn onset was the most important feature, followed by the annual amplitude of TCG. When considering the four conifer forest types, there was less differentiation among features, though the annual amplitude of TCB appears to be most important, followed by the annual amplitude of TCW. Perhaps not surprisingly, TCW features were most important when considering the separability of the wetlands classes (*Hardwood swamp* and *Softwood swamp*), with the RMSE of the TCW harmonic model ranked as most important, followed by TCW annual amplitude and TCW intercept.

Comparing feature importances across all combinations of forest types, no one feature appears to be consistently ranked among the most important. However, our results do suggest that the intercept and RMSE of the TCB harmonic model, thermal features (intercept, amplitude and RMSE), timing of spring onset, and the timing and value of peak EVI appear to be relatively unimportant for any of the forest type discriminations considered. In general, our analysis of feature importance indicates that different sets of spectral-temporal features are useful for distinguishing among different forest types.

#### 3.5. Mapped results

Maps of class probability estimates for the eight relatively pure forest classes characterized in this study are shown in Fig. 9. At broad scales, the forest probability maps show relatively distinct spatial patterns. For example, both types of forested wetlands appear to form small patches, and have a low-level, but persistent presence across the landscape (Fig. 9, maps (A) and (C)). In contrast, *Northern Hardwoods* and *Central Hardwoods*, which form a more significant portion of the forest matrix in this region, generally show higher probabilities and more widespread distribution than wetland and conifer classes (Fig. 9, maps (B) and (D)). *Northern Hardwoods* (e.g. beech, birch, maple, ash and aspen species) exhibit a clear east-west gradient, as would be expected given the underlying gradient in elevation, with higher

#### a. Mean Producer's Accuracy



#### b. Mean User's Accuracy

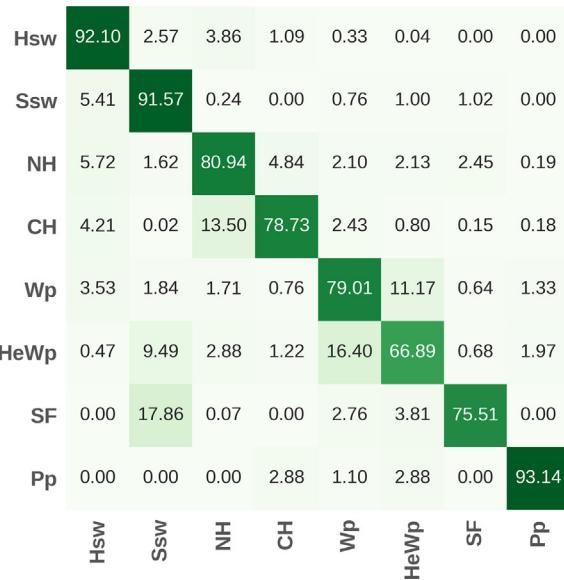
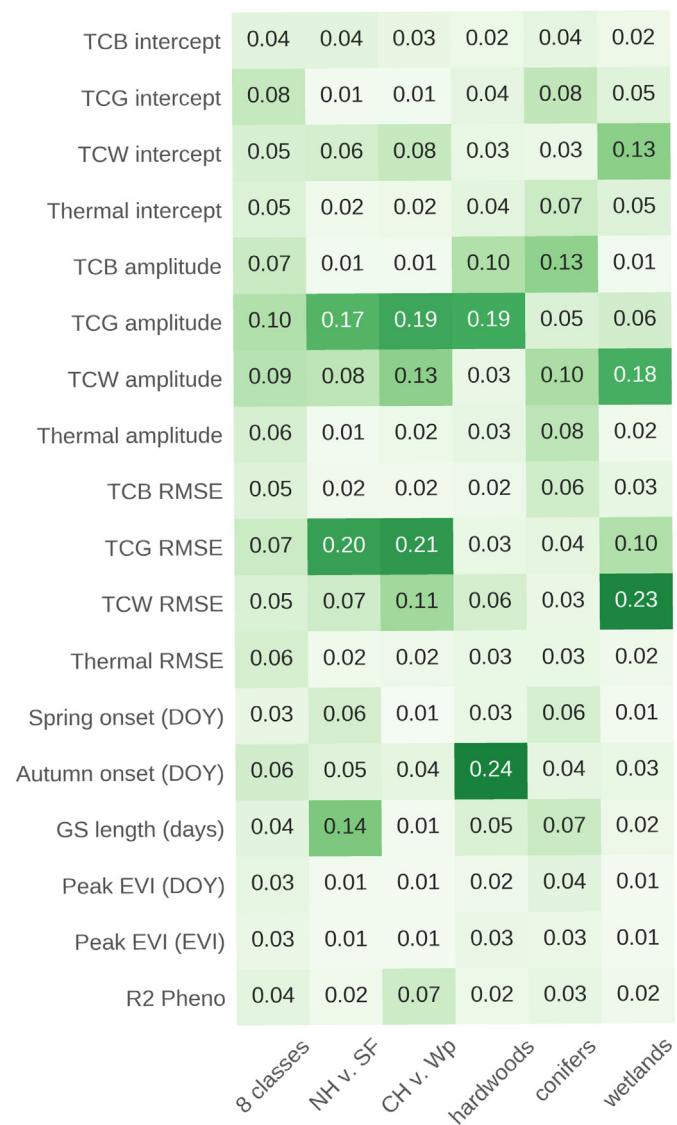


Fig. 7. Class-level agreement for feature set that includes all spectral-temporal features, as well as features from two ancillary datasets (topography and wetlands probability). Confusion matrices are shown as heat maps, with darker colors corresponding to higher scores.



**Fig. 8.** Feature importances for discriminating among various combinations of forest types. Higher feature importances suggest higher placement across trees and thus greater discriminatory power.

elevations found toward the western boundary with New York (as shown in Fig. 3). Central Hardwoods (e.g. oak-hickory communities) tend to show higher probability estimates in the southeastern portion of the study area, capturing both elevational and to some degree latitudinal patterns. White pine and Hemlock-White pine classes show similar distributions (Fig. 9 maps (E) and (F)), as would be expected given their overlapping class definitions. Spruce-Fir and Pitch pine classes show a more limited distribution, with Spruce-Fir present in the higher elevations of the western portion of the study area, and Pitch pine showing high probabilities only in a highly localized area known to support an exemplary dry Pitch pine outwash community. These results suggest that despite limitations and potential errors in the reference dataset, spectral-temporal features derived from Landsat time series, coupled with ancillary datasets, can be used to produce reasonable predictions of both common and rare forest types. While forest type maps such as those shown in Fig. 9 must be fully validated before being used for statistical inference or incorporated into modeling efforts, these preliminary results justify further investigation into using spectral-temporal features for species distribution mapping.

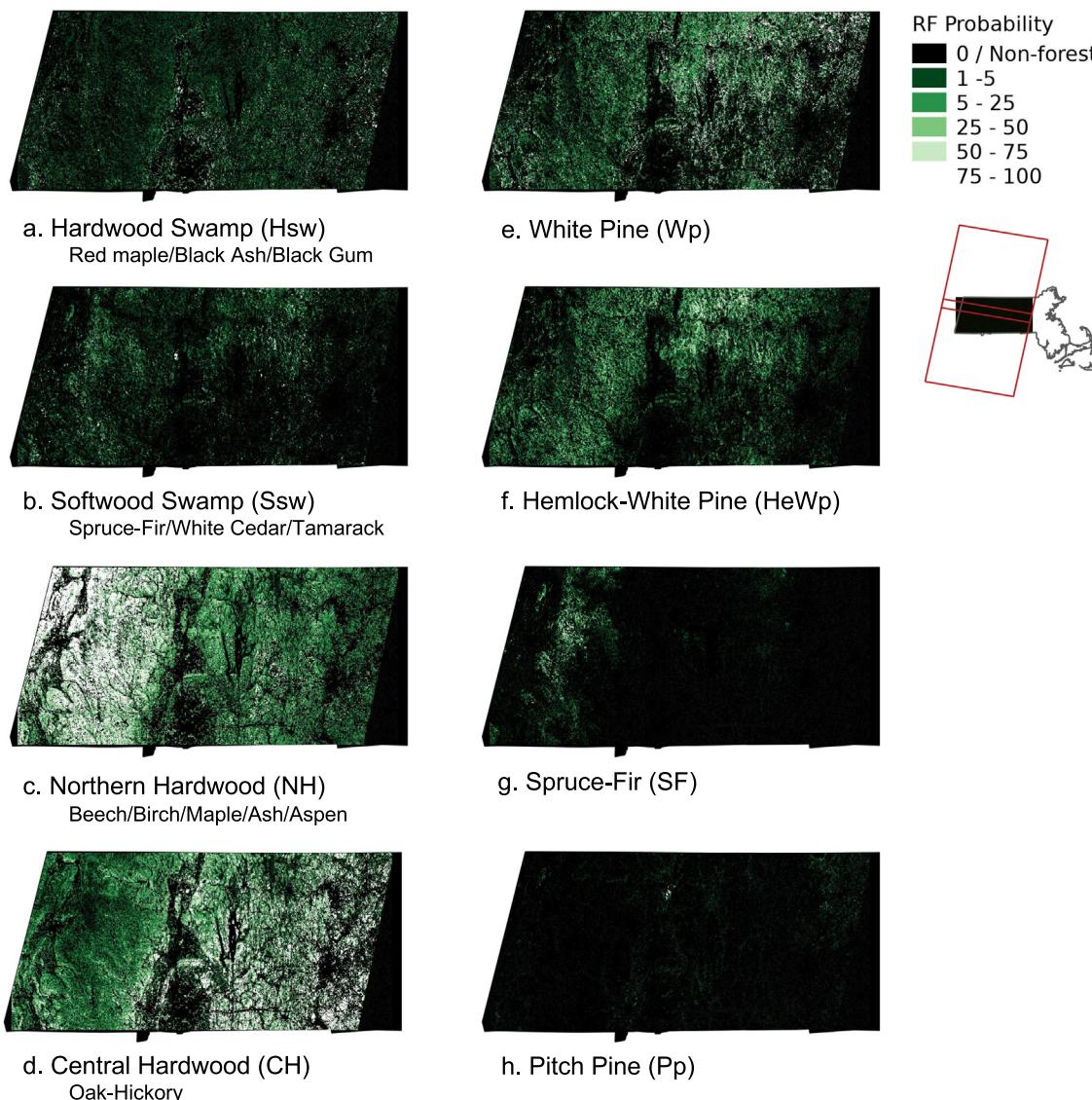
#### 4. Discussion

We evaluated the utility of single-date, multi-date, and spectral-temporal features for classifying eight forest types in Western Massachusetts, USA. Spectral-temporal features derived from Landsat time series produced superior discrimination of relatively pure forest communities relative to both single-date imagery and two-date combinations of the best available imagery across multiple metrics of agreement. The approach presented should be generalizable to a broader range of land cover classes and, because spectral-temporal features rely on all high-quality observations and not on single acquisitions, the approach can also be extended over larger areas. Given recent interest in time series-based metrics for classification (Gómez et al., 2016), this study provides empirical evidence for the utility of such metrics for improved discrimination of forest types.

The observed variability in single-date results highlights the uncertainty of using any individual image for discriminating among forest types, and may have important implications for using best-available-pixel composites for classification. Despite the visual consistency of leaf-on imagery (Fig. 3), classifications using only single-date growing season reflectance varied widely and often significantly in their agreement with reference data at both the pixel and polygon scale. Leaf-off imagery, which would be expected to aid in discriminating among deciduous and conifer classes, also showed varied results, with the November acquisition significantly outperforming the March acquisition, especially at the polygon scale. As anticipated, imagery from the spring and fall transitions seasons produced the highest levels of agreement of all single-date feature sets. These results clearly demonstrate the importance of image timing for classification efforts. The October 20, 2000 image, which was acquired during peak fall foliage that year, produces significantly higher levels of agreement than imagery acquired just eight days prior on October 12, 2000. This suggests that a well-timed image can be very valuable for classification over smaller areas. However, while mosaicking or compositing images from different dates may be suitable for change detection approaches where the objective is to determine whether or not there has been a dramatic, long-term shift in forest structure (e.g., Hermosilla et al., 2015), our classification results suggest that differences in image acquisition dates, even with an optimal 8-day repeat time, could introduce notable differences in pixel-scale predicted results.

Previous work has clearly demonstrated the utility of using multi-seasonal imagery to distinguish among forest types (see Wolter et al., 1995; Cohen and Goward, 2004 for reviews). We found that that multi-date feature sets consistently and significantly outperformed all single-date feature sets across all metrics of agreement, re-affirming the importance of the temporal dimension for forest type classification (Iverson et al., 1989; Cohen and Goward, 2004). Yet, as in single-date results, there was significant variability across different multi-date pairings. This result again suggests that when using inputs from individual dates of imagery, the timing of image acquisition has a significant effect on its utility for predicting forest composition. While other transformations in the spectral domain, such as multi-date differencing or orthogonalization approaches may improve differentiation among forest types (e.g. Wolter et al., 1995; Maiersperger et al., 2001; Dymond et al., 2002) and potentially achieve greater levels of agreement than those presented here, the limitations of working with individual images remain.

Spectral-temporal features derived from time series of observations offer a number of advantages over reflectance values from individual images. Though overall gains in agreement using spectral-temporal feature sets compared to multi-date feature sets were modest in magnitude (1–5%, depending on measure of agreement and scale of assessment), these gains were statistically significant across all metrics used in this study, suggesting that spectral-temporal inputs improve classification results relative to more traditional image-based inputs. Furthermore, unlike observations from a single image (single point in



**Fig. 9.** Unvalidated maps of forest type probabilities in the Western Massachusetts study area. Maps were produced using a feature set consisting of all spectral-temporal features and both ancillary datasets. Continuous probability estimates have been binned into aggregate ordinal ranges for visualization purposes. Darker green represents low probability while lighter tones represent higher probabilities of a given class, given the voting behavior of the RF classifier. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

time), spectral-temporal features are image-independent, characterizing patterns in the seasonal reflectance and timing of phenological events for different forest communities based on all clear observations for an individual pixel. This pixel-based perspective (Kennedy et al., 2014) allows for the derivation of equivalent features over large areas, and our results indicate that these features outperform more conventional Landsat inputs for distinguishing among forest types.

Of the spectral-temporal features considered in this study, annual amplitude features were often ranked highly in the feature importance analysis, indicating that this time series metric is generally useful for discriminating among forest types. Feature importance analysis suggests autumn offset is particularly useful for distinguishing among different hardwood communities. Previous multi-date efforts to classify thirty forest classes in Connecticut with a focus on the wide variety of deciduous species produced a genus-level fuzzy accuracy of 73%, but an absolute match accuracy of only 13% (Mickelson et al., 1998). The results of the eight-class hard classification presented here suggests that discrimination at the deciduous assemblage level seems realistic, with the highest Producer's agreement for deciduous forest classes. We find

that phenology values, specifically growing season length, are important for differentiating among deciduous and conifer forest types, demonstrating the applicability of phenology analysis for characterizing both highly seasonal and relatively aseasonal forest assemblages. In general, we do find lower class-level accuracies for conifer forest types. Previous work has suggested that species-level conifer characterization is challenging (Walsh, 1980), and the presence of mixed pixels may complicate assessment of individual conifer trees within predominantly hardwood stands (Fig. 5).

It is worth noting that in testing the influence of both the type and number of spectral-temporal features used for classification, we found that individual features sets (e.g. only intercepts, only annual amplitudes, only phenology features) produced significantly lower levels of agreement than feature sets that combined multiple types of features (results not shown). Thus, the power of spectral-temporal features lies in the combination of features that characterize both spectral variability and temporal variability in reflectance properties. This finding is further supported by the results of the feature importance analysis, which revealed that both harmonic and phenology features played an

important role in distinguishing among different forest types.

Efforts to map land cover and land cover change for the United States using all available Landsat observations are already underway (Zhu et al., 2016), and combining the approach presented in this study with Landsat Analysis Ready Data (ARD) and a large-scale reference dataset like the US Forest Service (USFS) Forest Inventory Analysis (FIA) data would make it possible to extend Landsat-resolution forest composition mapping to regional and/or national scales. Previous studies have used FIA data to generate imputed maps of forest type at both regional and continental extents using coarser resolution imagery (e.g. Zhu and Evans, 1994; Ruefenacht et al., 2008; Wilson et al., 2012; Duvaneck et al., 2015), and three-season Landsat 7 imagery has also been used to produce maps of forest parameters, including basal area and trees per acre, for 175 tree species based on FIA plot data (Ellenwood et al., 2015). However, to our knowledge FIA data have not yet been combined with Landsat imagery for forest type classification at national scales. Though restrictions on the use of true FIA plot locations represent a significant challenge (Ruefenacht et al., 2008), a plot-based dataset like the FIA could be used to extend the spectral-temporal classification approach presented here, providing a sample of forest types to be used as ground reference data. Given the wide variety of variables measured on FIA plots, further development of classification methods using FIA information would help constrain what level of thematic detail may be achieved using spectral-temporal feature sets. Examining the relationship between spectral-temporal properties and age, structure and disturbance in addition to forest composition would aid in building more meaningful forest classification models. With a clearly defined sampling design, the FIA data could also be used to validate forest type maps. Thus, incorporating plot-based forest inventory data like that collected by the FIA is a natural next step toward producing a validated map of forest types across regional and eventually national extents.

There are also opportunities to increase the temporal frequency of observations through harmonization of Landsat and Sentinel-2 time series (Flood, 2017). The combined use of Landsat and Sentinel-2 imagery enables acquisitions every 3–5 days globally (Skakun et al., 2017), and increasing the number of clear observations used in time series analysis will help better resolve spectral-temporal patterns, particularly seasonal variability. More research into feature variability across time, space and forest community communities will be essential for operational large area forest type mapping using moderate-resolution time series data, but the work presented here suggests the utility of spectral-temporal features for forest type classification.

Ultimately, the development of improved maps of forest type and other forest parameters from Landsat spectral-temporal features will have important implications for a variety of forest monitoring efforts. Many forest research applications rely on locally accurate information on the distribution and general characteristics of forests, including habitat mapping and species-habitat modeling (e.g. Simons-Legarda et al., 2016), carbon and biomass estimation (e.g. Goulden et al., 1996; Hadley and Schedlbauer, 2002), land change simulation modeling (e.g. Schneider and Pontius, 2001; Thompson et al., 2011), and assessment of climate impacts on forest distribution and composition (e.g. Iverson and Prasad, 2001; Archetti et al., 2013). The spectral-temporal approach presented in this study represents a new means for generating maps of forest properties needed to parameterize and model this diverse array of ecosystem processes, and Landsat-resolution forest type maps are expected to serve many different user groups from both the research and land management communities.

## 5. Conclusions

The results of this study indicate that features derived from time series of all available Landsat observations quantify meaningful spectral and temporal variability in relatively homogenous forest communities in the complex forested landscape of Western Massachusetts. Though a

well-timed image may be very valuable for distinguishing among forest types, spectral-temporal features consistently produce the highest levels of agreement with the forest community reference dataset across all metrics used, while also overcoming key limitations of single-date imagery or combinations of single date images. Spectral-temporal features derived from time series of surface reflectance data provide stable, wall-to-wall features; have greater dimensionality; and can (and should) be combined with change detection approaches so that features can be extracted for relatively stable time segments. Overall agreement scores of > 80% were achieved at both the pixel and polygon scale for eight-class hard classifications of forest types that included spectral-temporal features and ancillary datasets, and improved levels of agreement and thematic detail may be obtained with plot-scale reference data. After decades of reliance on only a very limited subset of imagery to distinguish among forest types, use of the full Landsat spectral-temporal domain represents an exciting new direction for large-scale forest type mapping.

## Acknowledgements

This work was supported by the US Forest Service (grant number 13-CR-11221638-146), US Geological Survey (grant number G12PC00070), NASA (grant number NNX13AP42), and the Boston University Department of Earth & Environment. The content of this document does not necessarily represent the views or policies of the Department of the Interior, nor does mention of trade names, commercial products or organizations imply endorsement by the U.S. Government. Landsat Surface Reflectance products courtesy of the U.S. Geological Survey Earth Resources Observation and Science Center. We would like to thank Brian Hawthorne of the Massachusetts Division of Fisheries and Wildlife for providing the forest type data used in this study. We would also like to thank Robert Kennedy, Pontus Olofsson, and Mark Friedl for their insightful comments on an earlier version this manuscript.

## References

- Archetti, M., Richardson, A.D., O'Keefe, J., Delpierre, N., 2013. Predicting climate change impacts on the amount and duration of autumn colors in a New England forest. *PLoS One* 8 (3), e57373. <http://dx.doi.org/10.1371/journal.pone.0057373>.
- Belgiu, M., Drägut, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <http://dx.doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Breiman, L., 1996. Out-of-bag estimation. In: *Technical Report*. Department of Statistics: University of California, Berkeley.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Bromley, S.W., 1935. The original forest types of southern New England. *Ecol. Monogr.* 5 (1), 61–89.
- Brooks, E.B., Wynne, R.H., Thomas, V.A., Blinn, C.E., Coulston, J.W., 2014. On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *IEEE Trans. Geosci. Remote Sens.* 52 (6), 3316–3332. <http://dx.doi.org/10.1109/TGRS.2013.2272545>.
- Brown de Colstoun, E., 2003. National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sens. Environ.* 85 (3), 316–327. [http://dx.doi.org/10.1016/S0034-4257\(03\)00010-5](http://dx.doi.org/10.1016/S0034-4257(03)00010-5).
- Chen, C., Liaw, A., Breiman, L., 2004. Technical Report 666. Department of Statistics, University of California, Berkeley, CA, USA (Using random forest to learn imbalanced data).
- Cohen, W.B., Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. *Bioscience* 54 (6), 535. [http://dx.doi.org/10.1641/0006-3568\(2004\)054\[0535:LRIEAO\]2.0.CO;2](http://dx.doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2).
- Corcoran, J., Knight, J., Gallant, A., 2013. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota. *Remote Sens.* 5 (7), 3212–3238 (doi:3390/rs5073212).
- Crist, E.P., 1985. A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* 17 (3), 301–306. [http://dx.doi.org/10.1016/0034-4257\(85\)90102-6](http://dx.doi.org/10.1016/0034-4257(85)90102-6).
- DeVries, B., Verbesselt, J., Kooistra, L., Herold, M., 2015. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sens. Environ.* 161, 107–121. <http://dx.doi.org/10.1016/j.rse.2015.02.012>.
- DeVries, B., Prathist, A.K., Verbesselt, J., Kooistra, L., Herold, M., 2016. Characterizing forest change using community-based monitoring data and Landsat time series. *PLoS One* 11 (3), e0147121. <http://dx.doi.org/10.1371/journal.pone.0147121>.

- Dodge, A.G., Bryant, E.S., 1976. Forest type mapping with satellite data. *J. For.* 74 (8), 526–531.
- Duveneck, M.J., Thompson, J.R., Wilson, B.T., 2015. An imputed forest composition map for New England screened by species range boundaries. *For. Ecol. Manag.* 347 (C), 107–115. <http://dx.doi.org/10.1016/j.foreco.2015.03.016>.
- Dymond, C.C., Mladenoff, D.J., Radloff, V.C., 2002. Phenological differences in tasseled cap indices improve deciduous forest classification. *Remote Sens. Environ.* 80 (3), 460–472. [http://dx.doi.org/10.1016/S0034-4257\(01\)00324-8](http://dx.doi.org/10.1016/S0034-4257(01)00324-8).
- Ellenwood, J.R., Krist, F.J., Romero, S.A., 2015. National Individual Tree Species Atlas. United States Department of Agriculture, Forest Service, Fort Collins, CO Available at: [http://www.fs.fed.us/foresthealth/technology/pdfs/FHTET\\_15\\_01\\_National\\_Individual\\_Tree\\_Species\\_Atlas.pdf](http://www.fs.fed.us/foresthealth/technology/pdfs/FHTET_15_01_National_Individual_Tree_Species_Atlas.pdf), Accessed date: 17 December 2016.
- Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* 15 (1), 3133–3181.
- Fisher, J., Mustard, J., Vadéboncoeur, M., 2006. Green leaf phenology at Landsat resolution: scaling from the field to the satellite. *Remote Sens. Environ.* 100 (2), 265–279. <http://dx.doi.org/10.1016/j.rse.2005.10.022>.
- Flood, N., 2017. Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sens.* 9 (7), 659. <http://dx.doi.org/10.1016/j.rse.2015.12.024>.
- Foster, D.R., 1992. Land-use history (1730–1990) and vegetation dynamics in central New England, USA. *J. Ecol.* 753–771. <http://dx.doi.org/10.2307/2260864>.
- Foster, D.R., Motzkin, G., 1998. Ecology and conservation in the cultural landscape of New England: lessons from nature's history. *Northeast. Nat.* 111–126.
- Fralish, J.S., Ponder Jr., F., 2003. In: Van Sambeek, J.W., Dawson, J.O., Loewenstein, E.F., Fralish, J.S. (Eds.), The Central Hardwood Forest: Its Boundaries and Physiographic Provinces. Proceedings, 13th Central Hardwood Forest Conference; 2002 April 1–3; Urbana, IL. Gen. Tech. Rep. NC-234 U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN (565 p.).
- Franklin, J., Logan, T.L., Woodcock, C.E., Strahler, A.H., 1986. Coniferous Forest classification and inventory using Landsat and digital terrain data. *Geosci. Remote Sens. IEEE Trans.* 24 (1), 139–149.
- Friedl, M.A., Woodcock, C., Gopal, S., Muchoney, D., Strahler, A.H., Barker-Schaaf, C., 2000. A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data. *Int. J. Remote Sens.* 21, 1073–1077.
- Gerhardt, F., Foster, D., 2002. Physiographical and historical effects on forest vegetation in central New England, USA. *J. Biogeogr.* 29, 1421–1437.
- Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. *ISPRS J. Photogramm. Remote Sens.* 116 (C), 55–72. <http://dx.doi.org/10.1016/j.isprsjprs.2016.03.008>.
- Gonçalves, A.C., 2017. Multi-species stand classification: definition and perspectives. In: *Forest Ecology and Conservation*. InTech.
- Goulden, M.L., Munger, J.W., Fan, S.M., Daube, B.C., Wofsy, S.C., 1996. Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. *Science* 271 (5255).
- Hadley, J.L., Schedlbauer, J.L., 2002. Carbon exchange of an old-growth eastern hemlock (*Tsuga canadensis*) forest in central New England. *Tree Physiol.* 22 (15–16), 1079–1092.
- Hall, B., Motzkin, G., Foster, D.R., Syfert, M., Burk, J., 2002. Three hundred years of forest and land-use change in Massachusetts, USA. *J. Biogeogr.* 29 (10–11), 1319–1335.
- He, H.S., Mladenoff, D.J., Radloff, V.C., Crow, T.R., 1998. Integration of GIS data and classified satellite imagery for regional forest assessment. *Ecol. Appl.* 8 (4), 1072–1083. [http://dx.doi.org/10.1890/1051-0761\(1998\)008\[1072:IOGDAC\]2.0.CO;2](http://dx.doi.org/10.1890/1051-0761(1998)008[1072:IOGDAC]2.0.CO;2).
- Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sens. Environ.* 158, 220–234. <http://dx.doi.org/10.1016/j.rse.2014.11.005>.
- Holden, C.E., Woodcock, C.E., 2016. An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. *Remote Sens. Environ.* 1–21. <http://dx.doi.org/10.1016/j.rse.2016.02.052>.
- Holden, C.E., Arevelo, P., Pasquarella, V., 2016. Yet Another Time Series Model (YATSM): v0.6.1. Zenodo. <http://dx.doi.org/10.5281/zenodo.51336>.
- Iverson, L.R., Prasad, A.M., 2001. Potential changes in tree species richness and forest community types following climate change. *Ecosystems* 4 (3), 186–199. <http://dx.doi.org/10.1007/s10021-001-0003-6>.
- Iverson, L.R., Graham, R.L., Cook, E.A., 1989. Applications of satellite remote sensing to forested ecosystems. *Landscape Ecol.* 3 (2), 131–143.
- Kennedy, R.E., Andréfouët, S., Gómez, C., Griffiths, P., Hais, M., Healey, S., Helmer, E.H., Hostert, P., Lyons, M., Meigs, G.W., Pflugmacher, D., Phinn, S., Powell, S., Scarth, P.F., Sen, S., Schroeder, T.A., Schneider, A.-M., Sonnenchein, R., Vogelmann, J.E., Wulder, M.A., Zhu, Z., 2014. Bringing an ecological view of change to Landsat-based remote sensing. *Front. Ecol. Environ.* 12 (6), 339–346. <http://dx.doi.org/10.1890/130066>.
- Kerr, J.T., Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* 18 (6), 299–305. [http://dx.doi.org/10.1016/S0169-5347\(03\)00071-5](http://dx.doi.org/10.1016/S0169-5347(03)00071-5).
- Khatami, R., Mountrakis, G., Stehman, S.V., 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. *Remote Sens. Environ.* 177, 89–100.
- Loveland, T.R., Dwyer, J.L., 2012. Landsat: building a strong future. *Remote Sens. Environ.* 122, 22–29. <http://dx.doi.org/10.1016/j.rse.2011.09.022>.
- Maiersperger, T.K., Cohen, W.B., Ganio, L.M., 2001. A TM-based hardwood-conifer mixture index for closed canopy forests in the Oregon Coast Range. *Int. J. Remote Sens.* 22 (6), 1053–1066. <http://dx.doi.org/10.1080/01431160117436>.
- Markham, B.L., Helder, D.L., 2012. Forty-year calibrated record of earth-reflected radiance from Landsat: a review. *Remote Sens. Environ.* 122 (C), 30–40. <http://dx.doi.org/10.1016/j.rse.2011.06.026>.
- Martin, M.E., Newman, S.D., Aber, J.D., Congalton, R.G., 1998. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* 65 (3), 249–254.
- Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* 3 (1), 68–72. <http://dx.doi.org/10.1109/LGRS.2005.857030>.
- MassWildlife Staff, 2013. DFW Biomonitoring Data [ArcGIS geodatabase]. (Unpublished raw data).
- Melaas, E.K., Friedl, M.A., Zhu, Z., 2013. Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sens. Environ.* 132, 176–185. <http://dx.doi.org/10.1016/j.rse.2013.01.011>.
- Melaas, E.K., Sulla-Menashe, D., Gray, J.M., Black, T.A., Morin, T.H., Richardson, A.D., Friedl, M.A., 2016. Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. *Remote Sens. Environ.* 186, 452–464. <http://dx.doi.org/10.1016/j.rse.2016.09.014>.
- Mickelson, J.G., Civco, D.L., Silander, J.A., 1998. Delineating forest canopy species in the northeastern United States using multi-temporal TM imagery. *Photogramm. Eng. Remote. Sens.* 64 (9), 891–904.
- Mora, B., Wulder, M.A., White, J.C., 2010. Identifying leading species using tree crown metrics derived from very high spatial resolution imagery in a boreal forest environment. *Can. J. Remote. Sens.* 36 (4), 332–344.
- National Oceanic and Atmospheric Administration (NOAA), Office for Coastal Management, 2014. “National Wetland Potential”. Coastal Change Analysis Program (C-CAP) Regional Land Cover. NOAA Office for Coastal Management, Charleston, SC. [www.coast.noaa.gov/ccapft](http://www.coast.noaa.gov/ccapft), Accessed date: May 2017.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* 148, 42–57. <http://dx.doi.org/10.1016/j.rse.2014.02.015>.
- Pasquarella, V.J., Holden, C.E., Kaufman, L., Woodcock, C.E., 2016. From imagery to ecology: leveraging time series of all available Landsat observations to map and monitor ecosystem state and dynamics. *Remote Sens. Ecol. Conserv.* 2 (3), 152–170. <http://dx.doi.org/10.1002/rse.24>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Plourde, L.C., Ollinger, S.V., Smith, M.L., 2007. Estimating species abundance in a northern temperate forest using spectral mixture analysis. *Photogramm. Eng. Remote. Sens.* 73 (7), 829–840.
- Reese, H.M., Lillesand, T.M., Nagel, D.E., Stewart, J.S., Goldmann, R.A., Simmons, T.E., et al., 2002. Statewide land cover derived from multiseasonal Landsat TM data - a retrospective of the WISCLAND project. *Remote Sens. Environ.* 82 (2–3), 224–237. [http://dx.doi.org/10.1016/S0034-4257\(02\)00039-1](http://dx.doi.org/10.1016/S0034-4257(02)00039-1).
- Rodriguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012. Random forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* 121 (C), 93–107. <http://dx.doi.org/10.1016/j.rse.2011.12.003>.
- Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalsky, V., Hansen, M., et al., 2010. Web-enabled Landsat data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sens. Environ.* 114 (1), 35–49. <http://dx.doi.org/10.1016/j.rse.2009.08.011>.
- Ruefenacht, B., Finco, M.V., Nelson, M.D., Czaplewski, R., Helmer, E.H., Blackard, J.A., et al., 2008. Conterminous US and Alaska forest type mapping using forest inventory and analysis data. *Photogramm. Eng. Remote. Sens.* 74 (11), 1379–1388. <http://dx.doi.org/10.14358/PERS.74.11.1379>.
- Sader, S.A., Ahl, D., Liou, W.-S., 1995. Accuracy of Landsat-TM and GIS rule-based methods for forest wetland classification in Maine. *Remote Sens. Environ.* 53 (3), 133–144.
- Schneider, L.C., Pontius Jr., R.G., 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agric. Ecosyst. Environ.* 85 (1–3), 83–94. [http://dx.doi.org/10.1016/S0167-8809\(01\)00189-X](http://dx.doi.org/10.1016/S0167-8809(01)00189-X).
- Simons-Legard, E.M., Harrison, D.J., Legard, K.R., 2016. Habitat monitoring and projections for Canada lynx: linking the Landsat archive with carnivore occurrence and prey density. *J. Appl. Ecol.* 53, 1260–1269. <http://dx.doi.org/10.1111/1365-2664.12611>.
- Skakun, S., Vermote, E., Roger, J.-C., Franch, B., 2017. Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. *AIMS Geosci.* 3 (2), 163–186. <http://dx.doi.org/10.3934/geosci.2017.2.163>.
- Thompson, J.R., Foster, D.R., Scheller, R., Kittredge, D., 2011. The influence of land use and climate change on forest biomass and composition in Massachusetts, USA. *Ecol. Appl.* 21 (7), 2425–2444. <http://dx.doi.org/10.1890/10-2383.1>.
- Thompson, S.D., Nelson, T.A., White, J.C., Wulder, M.A., 2015. Mapping dominant tree species over large forested areas using Landsat best-available-pixel image composites. *Can. J. Remote. Sens.* 41 (3), 203–218. <http://dx.doi.org/10.1080/07038992.2015.1065708>.
- U.S. Geological Survey, 2015. Product Guide: Landsat Surface Reflectance Products. Courtesy of the U.S. Geological Survey. pp. 1–27. [http://landsat.usgs.gov/documents/cdr\\_sr\\_product\\_guide.pdf](http://landsat.usgs.gov/documents/cdr_sr_product_guide.pdf).
- Walsh, S.J., 1980. Coniferous tree species mapping using Landsat data. *Remote Sens. Environ.* 9 (1), 11–26.
- Westveld, M., 1956. Natural forest vegetation zones of New England. *J. For.* 54 (5), 332–338.
- White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., et al.,

2014. Pixel-based image compositing for large-area dense time series applications and science. *Can. J. Remote. Sens.* 40 (3), 192–212. <http://dx.doi.org/10.1080/07038992.2014.945827>.
- Wickham, J., Homer, C., Vogelmann, J., McKerrow, A., Mueller, R., Herold, N., Coulston, J., 2014. The multi-resolution land characteristics (MRLC) consortium — 20 years of development and integration of USA national land cover data. *Remote Sens.* 6 (8), 7424–7441. <http://dx.doi.org/10.3390/rs6087424>.
- Williams, D.L., Nelson, R.F., 1986. Use of remotely sensed data for assessing forest stand conditions in the Eastern-United-States. *IEEE Trans. Geosci. Remote Sens.* 24 (1), 130–138.
- Wilson, B.T., Lister, A.J., Riemann, R.I., 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. *For. Ecol. Manag.* 271 (C), 182–198. <http://dx.doi.org/10.1016/j.foreco.2012.02.002>.
- Wolter, P.T., Mladenoff, D.J., Host, G.E., Crow, T.R., 1995. Improved forest classification in the northern lake-states using multitemporal Landsat imagery. *Photogramm. Eng. Remote. Sens.* 61 (9), 1129–1143.
- Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al., 2008. Free access to Landsat imagery. *Science* 320 (5879). [\(1011a–1011a\).](http://dx.doi.org/10.1126/science.320.5879.1011a)
- Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* 122, 2–10. <http://dx.doi.org/10.1016/j.rse.2012.01.010>.
- Zhu, Z., Evans, D.L., 1994. US forest types and predicted percent forest cover from AVHRR data. *PE & RS- Photogramm. Eng. Remote. Sens.* 60 (5), 525–531.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <http://dx.doi.org/10.1016/j.rse.2011.10.028>.
- Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* 144 (C), 152–171. <http://dx.doi.org/10.1016/j.rse.2014.01.011>.
- Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sens. Environ.* 122, 75–91. <http://dx.doi.org/10.1016/j.rse.2011.10.030>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015a. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <http://dx.doi.org/10.1016/j.rse.2014.12.014>.
- Zhu, Z., Woodcock, C.E., Holden, C., Yang, Z., 2015b. Generating synthetic Landsat images based on all available Landsat data: predicting Landsat surface reflectance at any given time. *Remote Sens. Environ.* 162, 67–83.
- Zhu, Z., Gallant, A.L., Woodcock, C.E., Pengra, B., Olofsson, P., Loveland, T.R., ... Auch, R.F., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS J. Photogramm. Remote Sens.* 122, 206–221.