Remote Sensing of Environment 148 (2014) 97–107

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Remote sensing of spring phenology in northeastern forests: A comparison of methods, field metrics and sources of uncertainty

Katharine White [a], Jennifer Pontius [a,b,*], Paul Schaberg [b]

[a] University of Vermont, Rubenstein School of Environment and Natural Resources, 81 Carrigan Drive, Burlington, VT 05405, United States
[b] USDA Forest Service, Northern Research Station, 81 Carrigan Drive, Burlington, VT 05405, United States

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form 3 March 2014
Accepted 14 March 2014
Available online xxxx

Keywords:
Landsat
Green leaf phenology
Vegetation indices
Field to sensor scaling
EVI
NDVI

ABSTRACT

Current remote sensing studies of phenology have been limited to coarse spatial or temporal resolution and often lack a direct link to field measurements. To address this gap, we compared remote sensing methodologies using Landsat Thematic Mapper (TM) imagery to extensive field measurements in a mixed northern hardwood forest. Five vegetation indices, five mathematical fits to model a continuous temporal response, and a suite of threshold estimates for "start of spring/season" (SOS) assessments were compared to field measurements of bud burst stage and hemispherical photo derived canopy structural metrics (transparency, leaf area index, greenness). Results indicated that a four-parameter logistic model based on at least five spring coverages of the Enhanced Vegetation Index (EVI) and a SOS threshold of 0.3 was most closely related to field metrics and most accurate in predicting the date of full leaf out. Plot level SOS was predicted with a mean absolute error of 11 days for all species and elevation combinations, but improved to 9 days for hardwood dominated plots and 7 days for sugar maple dominated plots. Mean absolute error was improved to 8 days when forest type (mixed, conifer hardwood) was used to refine predictions. The consistency of prediction errors across forest types indicates that while overall accuracy across pixels may be low, inter-annual comparisons of changes in phenology on a pixel basis may provide accurate assessments of changes in phenology over time. This was confirmed by application to seven years of independent phenology data predicted with 12 days of mean absolute error. However, image availability will be a limiting factor in areas of frequent cloud cover.

1. Introduction

With an increasing interest in how climate impacts forest ecosystems, studies of plant phenology have become more common, spanning from single site field studies (Badeck et al., 2004) to continental remote sensing assessments (White, Thornton, & Running, 1997). In the northeastern United States and northern latitudes these studies indicate that warming trends have extended the growing season, with both an earlier spring and later autumn (Badeck et al., 2004; McNeil, Denny, & Richardson, 2008; Menzel, 2002; Tucker et al., 2001; White, Running, & Thornton, 1999; Zhang, Friedl, Schaaf, & Strahler, 2004), with implications for productivity (Badeck et al., 2004; White et al., 1999), carbon cycling (Cleland, Chuine, Menzel, Mooney, & Schwartz, 2007), nutrient and water cycling (McNeil et al., 2008), species interactions (Badeck et al., 2004), and disturbance regimes (Dukes et al., 2009; Kramer & Hänninen, 2009).

The majority of remote sensing efforts across the region have fit vegetation growth curves to temporal composites using coarse spatial resolution satellites, such as AVHRR and MODIS (deBeurs & Henebry, 2010; Fisher, Mustard, & Vadeboncoeur, 2006; Liang, Schwartz, & Fei, 2011). White et al. (1997) compared AVHRR derived Normalized Difference Vegetation Index (NDVI) to lilac leaf out with a mean absolute error of 26 days. A MODIS derived Enhanced Vegetation Index (EVI) was used by Zhang, Friedl, and Schaaf (2006) to predict full canopy cover at Hubbard Brook with a mean absolute error of 10 days. Liang et al. (2011) improved accuracy to a mean absolute error of 2 days using MODIS derived EVI and weighting field observations by dominant forest community type. While these studies offer a regional perspective, there are several scale-based limitations. Because of the large degree of heterogeneity in species composition, soil type, forest fragmentation, land use (mixed pixels) and elevation found in northeastern forests, the use of such coarse resolution assessments limits the evaluation of spatial variability in phenology (Fisher et al., 2006; Ibanez et al., 2010; Kramer & Hänninen, 2009). The success and interpretation of such efforts are also limited by the difficulty in linking plot level field measurements to satellite derived pixel values; as well as uncertainty in which phenological field metrics match sensor reflectance metrics (Fisher et al., 2006).

* Corresponding author at: USDA Forest Service, Northern Research Station, 81 Carrigan Drive, Burlington, VT 05405, United States. Tel.: +1 802 656 3091.
E-mail addresses: kmwhite@uvm.edu (K. White), Jennifer.pontius@uvm.edu (J. Pontius), pschaberg@fs.fed.us (P. Schaberg).

http://dx.doi.org/10.1016/j.rse.2014.03.017
0034-4257/

The use of higher resolution imagery, such as from the Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM+), offers both the spatial resolution necessary to address these limitations, with the added advantage of a two decade archive that can provide a context for changes in phenology (Reed, Schwartz, & Xiao, 2009). While Landsat has been utilized for many vegetation applications, assessments of phenology are limited. In southern New England, Fisher et al. (2006) used an 18-year sequence of Landsat imagery to quantify spatial patterns in phenology. While they found a strong relationship between field measured phenological stage and satellite derived vegetation greenness, analyses were limited to hardwood dominated plots only. Most recently, Melaas, Friedl, and Zhu (2013) used a historical archive of Landsat TM imagery to examine patterns and variability in phenology at the Harvard Forest. While their results were promising ($r^2 = 0.80$), their assessment was limited to highly variable inter-annual assessments on a subset of trees contained within one Landsat pixel.

The goal of this study is to test previously published approaches to quantify phenology in northeastern hardwood and mixed forests using Landsat TM/ETM+ imagery, including the ability to differentiate subtle differences in phenology driven by species demographics and topographical variability. Because there is no standard method for phenology assessment using satellite imagery, and even less exploration of methodological options using Landsat, this paper focuses on the comparison of many reflectance products, curve fits, threshold values and field metrics in order to identify the most accurate approach to quantify the start of spring across a heterogeneous landscape. Our specific objectives include to:

- Identify the most accurate vegetation index, logistic model, and threshold for detecting the start of spring (SOS).
- Assess the impact of temporal frequency of imagery on SOS predictions.
- Examine the relationships between vegetation indices and various phenological field metrics to better understand what the sensor is characterizing and which phenological stages are most accurately predicted.
- Characterize potential sources of error in remote sensing SOS estimates.

2. Methods

2.1. Study area

Sampling a variety of sites is important in the northeast because of the heterogeneity of species composition and site characteristics across the landscape. To this end, we established 32 0.1 ha plots that were assessed every 2–4 days from March to July of 2011 across north/central Vermont (Fig. 1). This region was selected based on the overlap of Landsat paths 13 and 14 (row 29), effectively doubling the frequency of image dates available for analysis. Plot selection was designed to capture a range of elevations (60 to 780 m), forest cover types, and landscape positions while still providing repeated sampling of dominant canopy compositions. This included 19 northern hardwood stands (dominated by *Acer saccharum* Marsh., *Acer rubrum* L., *Betula alleghaniensis* Britton and *Fagus grandifolia* Ehrh.) and 12 mixed stands (including the species listed above and *Quercus rubra* L., *Pinus strobus* L., *Tsuga canadensis* L., and *Abies balsamea* L.). One pure conifer plot (*Pinus strobus*) was included as a reference, to see how vegetation index curves appear for species that retain foliage year-round. At each site, stand composition and topography were consistent across a larger 100 m area to minimize noise from misregistration of plot locations to image pixels.

2.2. Field phenology assessments

Field phenology methods were adapted to match as closely as possible guidelines from several active phenology monitoring programs conducted by the Vermont Department of Forest Parks and Recreation, Vermont Monitoring Cooperative, Proctor Maple Research Center, and Hubbard Brook Experimental Forest. Our modified approach categorized a tree into one of five phenology field rank (PFR) categories, from dormancy to full leaf out (Table 1) based on the dominant phenological stage assessed across the entire canopy. We recorded observations for all dominant, codominant and understory trees within 20 m of plot center. To scale PFR from individual trees to the plot level we calculated a plot average based on a species average weighted by its plot percent basal area. This ensured that variation within trees of the same species and variation among species on a plot were captured. While the initial tree level phenology rankings are ordinal in scale, both plot average PFR and vegetation indices are continuous in scale and therefore allow a potentially more precise comparison between field and satellite metrics (Reed et al., 2009).

To include canopy structural metrics, we collected digital images at each plot using a Nikon D90 digital camera with a Sigma 8 mm 180 degree circular fisheye lens. In order to minimize common errors reported in canopy closure estimates due to contrast confusion, slope or time of day (Breda, 2003; Frazer, Fournier, Trofymow, & Hall, 2001; Jonckheere et al., 2004; Sonnentag et al., 2012), each photo was taken at a fixed height and location over plot center, under cloudy conditions or early morning/late afternoon, with consistent auto focus, aperture, and ISO settings.

We used Gap Light Analyzer (GLA) (v2.0) software (Simon Fraser University (SFU), Burnaby, British Columbia, Canada and Institute of Ecosystem Studies (IES), Millbrook, New York, USA) to calculate percent canopy openness and Leaf Area Index (LAI 4 ring – effective leaf area index integrated over zenith angles 0–60°, and LAI 5 ring – effective leaf area integrated over the zenith angles 0–75°) (Frazer, Canham, & Lertzman, 1999). Images were isolated to the blue color plane in order to improve the contrast between foliage and sky (Frazer et al., 1999; Leblanc, Chen, Fernandes, Deering, & Conlye, 2005), with canopy/sky thresholds optimized using SideLook 1.1 (Jarčuška, Kucbel, & Jaloviar, 2010; Minkova & Logan, 2007; Nobis & Hunziker, 2005). Canopy "greenness" was approximated using software available from Digital Earth Watch (http://www.globalsystemsscience.org/software/download). This approach utilizes a normalized color channel brightness (percent red, green and blue) to quantify "green-up" associated with spring phenology (Richardson, Braswell, Hollinger, Jenkins, & Ollinger, 2009; Richardson et al., 2007).

2.3. Image processing

We downloaded Level 1T Landsat TM and ETM+ imagery from the USGS Global Visualization Viewer (U.S. Department of the Interior, 2012) for all image dates between March 10 and July 25, 2011 with less than 60% cloud cover. In order to maximize image availability, all scenes from the overlap of paths 13 and 14 were reprojected to UTM zone 18N for use in this study. This resulted in 15 images (out of 37 possible dates) for analyses. However, because of partial clouds across the study area, the highest number of image dates at any of our established field plots was 12. Level 1T processing includes an orthorectification, but to ensure consistency across image dates in our region of high terrain variability, each image was georegistered to a cloud-free, mid-summer 2002 control image using a 3rd order polynomial warping transform and nearest neighbor resampling. Reported fit of registration was below 0.2 pixels (RMSE 6m), based on a minimum of 50 ground control points located across the full extent of each image.

Raw DN values for bands 1–5 and 7 were converted to reflectance using ENVI's Landsat TM calibration tool (ITT Visual Information Solutions, 2009). This transformation accounts for solar illumination angle, minimum/maximum radiance and gain for each band, and scene specific atmospheric path length (Chander, Markham, & Helder, 2009). In order to minimize differences in atmospheric conditions between image acquisition dates we applied a dark object subtraction in ENVI that used the lowest DN that first represents a sharp increase in the number of pixels (Chavez, 1988; Soudani, Francois, Le Marie, Le

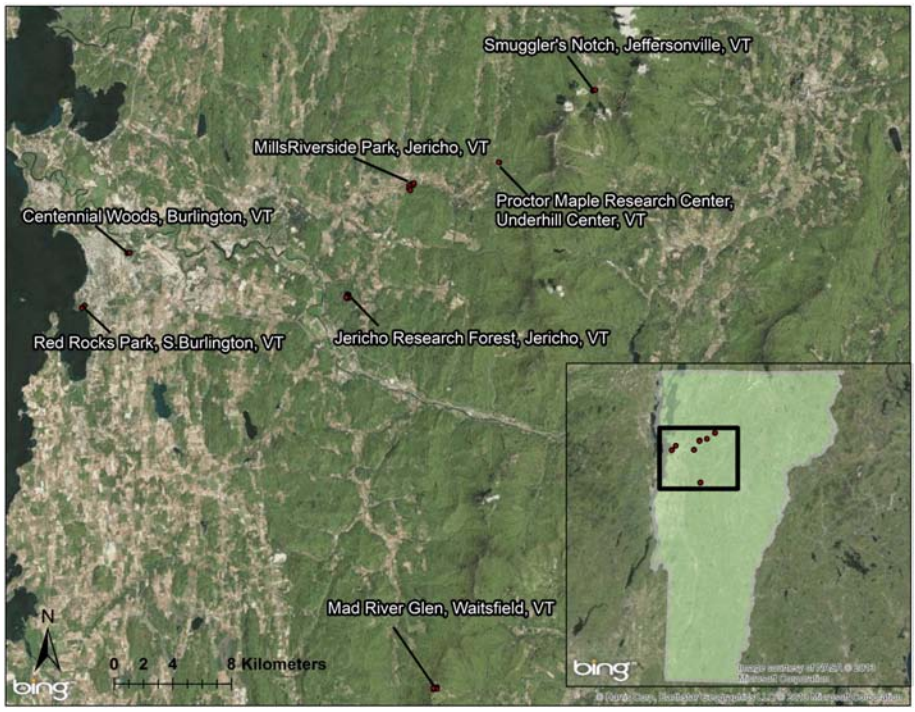

Fig. 1. The network of 32 plots assessed in this study. Plots were concentrated in the overlapping area of Landsat Path 13 and 14 (row 29) and selected to cover a range of elevations and forest types common to the northern hardwood mixed forests.

Gantec, & Dufrene, 2006). This was determined from image histogram statistics as the lowest data value that represents 0.1% of all pixels.

Clouds and haze were masked manually using thresholds determined individually for each image using the thermal band (TM band 6) (Fisher et al., 2006). The thermal band was also used to mask pixels dominated by snow, which can significantly affect vegetation index calculations (Huete et al., 2002; Zhang et al., 2003). Cloud and topographic shadow was removed using the Enhanced Vegetation Index (EVI), which clearly distinguished shadow from coniferous and other relatively dark vegetation.

2.4. Phenology modeling

2.4.1. Vegetation indices

The effectiveness of vegetation indices varies with local climate, soil, vegetation conditions, and purpose of the study (Broge & Leblanc, 2000; Hufkens et al., 2012; Rondeaux, Steven, & Baret, 1996). The Normalized Difference Vegetation Index (NDVI) and EVI are most commonly used to monitor phenology (Fisher et al., 2006; Hmimina et al., 2013; Melaas et al., 2013; Zhang et al., 2003). The normalized difference water index (NDWI) and other mid-infrared based indices are less commonly used, but have shown potential in monitoring phenology in regions of high snowfall and across varying regeneration stages (Boyd, Foody, Curran, Lucas, & Honzak, 1996; Broge & Leblanc, 2000). In this study we compared five different multi-spectral vegetation indices (Table 2).

Table 1
Phenology field rank (PFR).
Adapted from the Vermont Monitoring Corporative and Hubbard Brook Experimental Forest phenology methods.

Phenology field ranks (PFR)	Bud/leaf characteristics
0	Dormancy, no change from winter condition
1	Bud is swollen/larger than winter condition. No visible green leaf emerging from scales
2	Budbreak, green tip of leaf showing
3	Leaf emergence, recognizable form but often wrinkled and not fully developed
4	Leaves no longer wrinkled, fully expanded, may still be developing in size but large enough to be in final orientation position

2.4.2. Logistic model fit

Because reflectance values are not available every day, curves are fit to the available data in order to approximate daily values. Sigmoidal logistic models are commonly used to fit a time series for daily phenology measurements (Elmore, Guinn, Minsley, & Richardson, 2012; Melaas et al., 2013; Zhang et al., 2003). While logistic models begin with user-specified starting parameters that are then adjusted iteratively until the curve fit converges to best match the data, there are many ways to parameterize and fit them. We tested five common sigmoidal models (Table 3) on each of the field and satellite metrics at our 32 field plots. The quality of model fit for each plot was assessed by identifying the model with the lowest mean RMSE across all 45 plots. As a secondary comparison that included all plots and all image dates in one fit metric, we further compared the Spearman's rho correlation between actual vegetation index values and the corresponding fitted vegetation index values. This allowed us to identify the most robust model across forest types, pixel illumination and geographic locations.

2.4.3. Start of spring (SOS) thresholds

Although phenology gradually changes over several weeks, most phenological metrics are reported as a single date, or day of year (DOY) for events such as “bud burst” or “leaf out” (Fisher et al., 2006).

Table 2
The five vegetation indices used in previous phenology studies that were compared to assess which is most accurate for northern forest assessments.

Equation	Reference
$EVI = 2.5 \times \frac{(b4-b3)}{(b4+6\times b3-7.5\times b1+1)}$	Fisher et al. (2006), Zhang et al. (2003)
$NDVI = \frac{(b4-b3)}{(b4+b3)}$	Bradley, Jacob, Hermance and Mustard (2007), Jönsson, Eklundh, Hellström, Bärring, and Jönsson (2010), Van Leeuwen, Davison, Casady, and Marsh (2010), Nagai et al. (2010)
$NDWI = \frac{(b3-b5)}{(b3+b5)}$	Delbart et al. (2006)
$Red-MIR\ ratio = \frac{b3}{b7}$	Boyd et al. (1996), Broge and Leblanc (2000)
$Thermal-Red-MIR\ ratio = \frac{(b2\times b6)}{b7}$	Boyd et al. (1996), Broge and Leblanc (2000)

Table 3
The five sigmoid fits to model daily index values that were compared. The strongest and most consistent fit across forest types was achieved with the Zhang four-parameter logistic fit.

	Model	#Parameters	Citation	EVI ρ	NDVI ρ
Deciduous	Fisher	2	Fisher et al. (2006)	0.689	0.743
	Logistic	3	JMP SAS Institute	0.692	0.743
	Weibull	4	JMP SAS Institute	0.764	0.824
	Zhang	4	Zhang et al. (2003)	0.745	0.829
	Richard	5	JMP SAS Institute	0.723	0.761
Mixed	Fisher	2	Fisher et al. (2006)	0.639	0.796
	Logistic	3	JMP SAS Institute	0.639	0.796
	Weibull	4	JMP SAS Institute	0.7	0.796
	Zhang	4	Zhang et al. (2003)	0.716	0.801
	Richard	5	JMP SAS Institute	0.717	0.818

Reporting a single DOY provides the temporal resolution necessary to assess small changes in the timing of phenology from place to place or year to year. However, the obvious disconnect between the actual rate and variability in physiological changes associated with phenology and the desire to quantify a single date limits accuracy and interpretation of results, and has resulted in a myriad of ways to pinpoint DOY values. One common approach is to identify the day that the interpolated vegetation index curve first passes some threshold value (deBeurs & Henebry, 2010; Nagai, Nasahara, Muraoka, Akiyama, & Tsuchida, 2010; Zhang et al., 2006). Other approaches utilize the two points of maximum rate of curvature change in the interpolated logistic curve (Soudani et al., 2008; Zhang et al., 2003). In this way, each pixel's changing phenology is compared to itself instead of an arbitrary threshold that may not be applicable to all forest types (Zhang et al., 2006).

For each of our vegetation indices and sigmoidal model fits, we compared 3 different index thresholds, 3 different proportions of maximum index value, and 3 different slope based thresholds (Table 4, Fig. 2) in predicting SOS. Accuracy is quantified here as the mean absolute difference between the field measured and threshold predicted day a plot first reaches a given phenology field rank, averaged across all plots. This was repeated for phenology field ranks 2 (bud burst), 3 (leaf expansion) and 4 (full leaf out) in order to determine which phenology field rank was most accurately detected by each of the threshold options. Because some index/threshold combinations estimate SOS better for specific plot types and worse for others, we also included the standard deviation of the absolute difference between predicted and field measured SOS to assess the consistency of prediction across a variety of stand characteristics.

2.4.4. Accuracy and errors in the final phenology model

Once the most accurate combinations of vegetation index, sigmoidal fit and SOS threshold had been identified we examined how accuracy differed among plots of mixed hardwood/conifer, mixed hardwood, and sugar maple dominated composition to assess the impact of mixed species on prediction accuracy. To further explore sources of variability and error in the model, we compared site characteristics (elevation, slope and aspect) and various metrics of species composition (partitioning variation based on the percent hardwood, percent conifer, percent diffuse and ring porous species, forest type (hardwood, conifer, mixed), number of tree species, and Shannon Weiner diversity) to the accuracy of prediction across all field sites. Spearman's Rho correlations between each static variable and accuracy were used to identify potential sources of error. To test this method over many years, we compared sugar maple phenology data collected at the Proctor Maple Research Center in Underhill, VT (Wilmot, 2012) to SOS predictions following the remote sensing method described here. The Proctor field assessments used an 8-class scale, with daily assessments of sugar maple only at two locations dominated by sugar maple, but with other hardwood species occurring in small proportions. A spearman's Rho based on Proctor field measurements, and Landsat predicted phenology is used to assess our ability to apply this method in other years.

Table 4
Previously published start of spring (SOS) vegetation index thresholds, with expected phenology characteristics.

SOS metric	Spring ecological interpretation	Citation
0.3, 30%	Leaf expansion	Nagai et al. (2010)
0.5, 50%	Most leaves likely emerged	Bradley et al. (2007), Fisher et al. (2006), Fisher and Mustard (2007), White et al. (1997)
0.7, 70%	Initial leaf expansion	Nagai et al. (2010)
First max change in curvature	Bud burst or "greenup onset"	Zhang et al. (2003)
Second max change in curvature	Full leaf appearance, "maturity onset"	Zhang et al. (2003)
Second plateau curvature	Canopy maturity	Zhang et al. (2003)

2.4.5. Impact of image availability

To accurately observe rapid changes in phenology, sensors such as Landsat (16 day revisit) may not provide the temporal resolution needed (Busetto, Meroni, & Colombo, 2008). This is further complicated in areas such as the northeast, where cloud cover in the spring is common. To investigate the influence of image availability on SOS we compared the accuracy of SOS predictions in two ways. First we simulated various data availability scenarios using modeled daily phenology metrics based on our best phenology model (EVI, Zhang four-parameter fit, 0.3 threshold). This model was rerun on simulated return intervals of 1, 2, 4, 8, 16, and 32 days. The accuracy of each temporal resolution simulation was compared to field measurements using analysis of variance in JMP (v. 9.0.0 SAS Institute).

As a more direct examination of image availability on SOS predictions, phenology was modeled with various configurations of image availability (Table 5) at all field plots with the full complement of available imagery (from the full set of 12 images, to a minimum of 5 input image dates). In addition to testing the impact of temporal resolution, this allowed us to confirm that using a combination of Landsat TM and ETM+ sensors, from 2 different paths could be combined where available.

3. Results and discussion

3.1. Field phenology assessments

Mean bud burst across the 32 field sites was DOY = 127 (May 7), ranging from DOY = 117 (April 27) to DOY = 144 (May 24). Mean date of full leaf out was DOY = 158 (June 7), with a range between DOY 146 (May 26) and 193 (July 12). While exact timing varied between sites, spring greenup, from bud burst to full leaf out, typically occurred over two to three weeks. Rates of change between phenological field rankings were faster from bud burst (PFR2) to initial leaf emergence (FPR3) (mean 7.9 days), than from leaf emergence (PFR3) to full leaf development (PFR4) (mean 15.6 days). Phenological stages were observed first in lower elevation stands, with one day delay per 29 m gain in elevation to reach PFR 4, one day per 30 m gain to reach PFR3, and one day per 32 m gain to reach PFR2 (r^2 = 0.45, 0.79, and 0.73 respectively). This is consistent with Hopkins' Law of a one day delay in the onset of spring for every 30 m elevation increase (Fitzjarrald, Acevedo, & Moore, 2001; Richardson, Bailey, Denny, Martin, & O'Keefe, 2006) for the eastern United States.

Canopy photo metrics (LAI, canopy closure, and greenness) were not sensitive to early physiological changes such as bud swelling or bud burst. These digital field metrics peaked near initial leaf emergence, saturating before leaves were fully developed (Fig. 3). The resulting logistic fit to photo metrics is a much steeper and narrower region of rapid change centered near PFR 3 (Fig. 3). This indicates that hemispherical

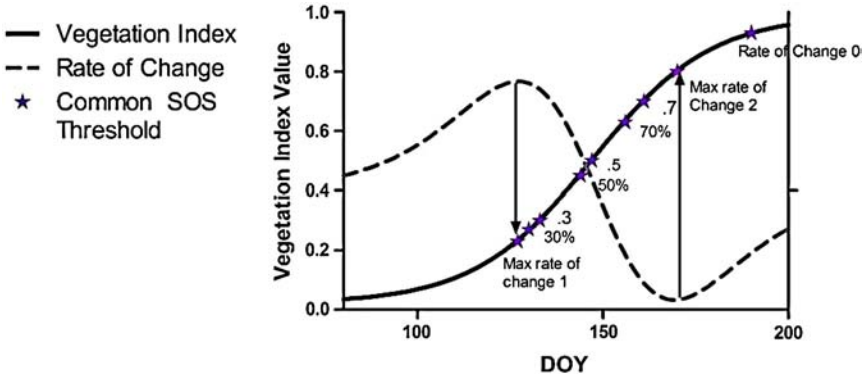

Fig. 2. The nine common start of spring (SOS) thresholds derived from interpolated vegetation index curves that were compared to field phenology rankings for accuracy assessment.

photo derived field metrics of canopy structure may not be representative of the changes witnessed in remote sensing vegetation indices.

In order to understand how much variability exists within plots we examined field rankings for canopy dominant hardwoods at each of the field sites. We found that the difference between the date that the first and last trees reached full leaf out spanned 19 days on average. For some plots, this was upwards of 30 days. Considering that so much variation on the ground exists, it is understandable that remote sensing prediction accuracies can border 10–15 days. Further, on the date when most trees on a plot had reached full leaf out (average PFR rounds to 4), the average standard deviation of field ranks was 0.54 for the 0–4 class scale. This indicates that even when a plot can be considered as having reached a PFR4 defined SOS, up to a third of individuals on a plot still lagged considerably. This variability was even higher earlier in the spring, with an average standard deviation of 0.81 when the first trees reached PFR4.

3.2. Phenology modeling

3.2.1. Comparison of indices and field metrics

Of the five indices tested, only the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) consistently matched field phenology metrics over the spring season (Spearman's Rho correlation of NDVI and EVI to PFR was ρ 0.95 and 0.96 respectively). The Normalized Difference Water Index (NDWI) decreased in early spring, likely as a result of snowmelt, before increasing in response to vegetation emergence. Coincident timing of bud burst and snow melt, particularly at higher elevations, may limit the use of NDWI in this region (Delbart, Toan, Kergoat, & Fedotova, 2006). Similarly, the red-middle infrared ratio (RW) and thermal-red-middle infrared ratio (TMIR) (Broge & Leblanc, 2000), decreased in early spring and spiked mid-spring, with no clear response coinciding with greenup across our plot network. Because of these inconsistencies with field metrics, NDWI, RW and TMIR were not considered in later steps of method development.

Table 5
Average fit (Spearman's ρ) and SOS prediction accuracy based on total number and timing of imagery dates. A minimum of 5 images was required for logistic fit.

# image dates	Images removed	EVI fit ρ	SOS error (days)
12	None	0.713	12
11	Post-SOS image	0.625	15.25
11	Pre-SOS image	0.634	15.25
10	Pre- and post-SOS images	0.634	12.5
8	All Landsat 7	0.681	19.5
6	Every other date	0.684	18.25
6	Path 14 images	0.628	14
6	Path 13 images	0.76	21.25
5	Every other date and pre-SOS	0.625	14.5

Both NDVI and EVI were significantly correlated with all of the field phenology metrics, but changes in NDVI over the spring deviated from field metrics in several ways. At our plots, NDVI typically increased from its winter baseline prior to plot-level leaf development (Fig. 3). This could be beneficial by providing the ability to detect the earliest onset of bud burst, but more likely results from sensitivity to the greening up of understory vegetation, which typically precedes the forest canopy (Richardson & O'Keefe, 2009). NDVI is also known to saturate at higher biomass (Huete et al., 2002; Reed et al., 2009), decreasing its ability to characterize the latter stages of canopy development and leaf maturity.

The closest relationship between field and satellite metrics was between EVI and optical phenology field ranks ($r^2 = 0.913$). The strength of this relationship is similar to results reported by Fisher et al. (2006) relating optical phenology estimates to greenness modeled from Landsat imagery in pure hardwood stands ($r^2 = 0.91$) and consistent with the approach used by Melaas et al. (2013). The ability of EVI to maintain the strength of this relationship across our mixed plots indicates that this approach may be robust for application across larger, more complex forested landscapes.

While it is somewhat surprising that the visual phenology metrics more closely matched remote sensing indices than digital photo metrics, it is likely that the latter may be strongly influenced by vegetation density and less sensitive to subtle spectral changes that accompany leaf development and changing canopy chemistry. Digital metrics saturated before full leaf expansion, limiting their sensitivity to the final stages of canopy maturation. This indicates that while it may be desirable to switch from subjective ocular estimates to automated methods, visual assessments of phenological stage are the closest match to remote sensing indices and most sensitive to subtle changes in canopy physiology.

Aside from comparisons of vegetation indices to field measurements, there were several interesting patterns observed in vegetation index curves over the spring. EVI values were lower than expected (zero for hardwood plots (Fig. 4A), and close to 0.1 for most mixed forest plots (Fig. 4B)) compared to other studies (Zhang et al., 2003 report EVI ~ 0.2 for deciduous broadleaf forests, and Liang et al., 2011 report EVI ~ 0.2 in a mixed temperate forests). This likely results from the relatively low evergreen (canopy and understory) composition for our plots compared to other studies, and a late winter snowfall across the region in 2011. This significant late season storm delayed snowmelt and understory vegetation growth. The result was a relatively low winter EVI and NDVI baseline, followed by a short snow-free period before trees broke bud.

Also interesting was the continued increase in EVI after field phenology metrics reached peak levels (Fig. 3). This could be due to the lag between full anatomical development and functional maturity (e.g., stable foliar chemistry and peaks in photosynthetic rates (Bassow & Bazzaz, 1998)). Anatomical leaf development precedes maximum photosynthetic capacity because the latter is dependent on the accumulation of chlorophyll, activation of Rubisco enzymes, and the completion of

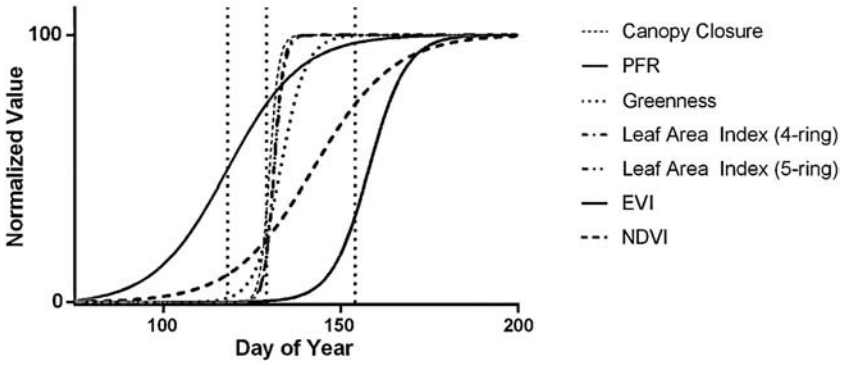

Fig. 3. Example plot field metrics and vegetation indices interpolated over the spring season using the Zhang logistic fit. Curves were normalized to compare the timing of metrics, where 0 was lowest data value and 100 was the highest. The vertical dotted lines represent the date that plot average phenology reached PFR 2, 3 and 4 chronologically.

electron transport systems of photosystem I (Eichelmann et al., 2008). In addition, cell wall and leaf cuticular thickening can continue beyond cellular expansion (Dengler, Mackay, & Gregory, 1975; Riederer & Schönherr, 1988). Vegetation indices have been related to physiological canopy changes after leaves have appeared fully developed (White et al., 1997). Unlike NDVI, this indicates that the EVI is not susceptible to saturation at high canopy densities and should be able to characterize changes in physiology through full leaf maturity (Huete et al., 2002; Liang et al., 2011).

3.2.2. Logistic fit models

While none of the 5 sigmoid logistic models tested was able to fit all field and index phenology metrics across all plots, the four-parameter logistic fit described in Zhang et al. (2003) had the most consistent (significant at 25 of the 31 mixed and deciduous broadleaf stands) and the most accurate overall fit (average $\rho = 0.85$, Table 3). While other four- and five-parameter models were able to fit EVI on hardwood dominated plots (Fig. 4A), the Zhang model was better able to fit mixed plots and plots with limited image dates (Fig. 4B). This validates the assumption that this four-parameter model is flexible enough to work across complex spatial variations in forest composition (Liang et al., 2011) without the loss of statistical power associated with higher order models fit to sparse data (deBeurs & Henebry, 2010).

3.2.3. Start of spring (SOS) thresholds

Because both NDVI and EVI consistently approximated the changes in field phenology metrics using the Zhang sigmoidal fit, we tested the accuracy of both indices at each phenological field rank (PFR2 through PFR4) using nine established thresholds (Table 4). Phenology field rank 2, bud burst, could not be accurately predicted using either index at any of the common thresholds tested. An examination of field rank to index curves (Fig. 3) showed that none of the indices tested increased

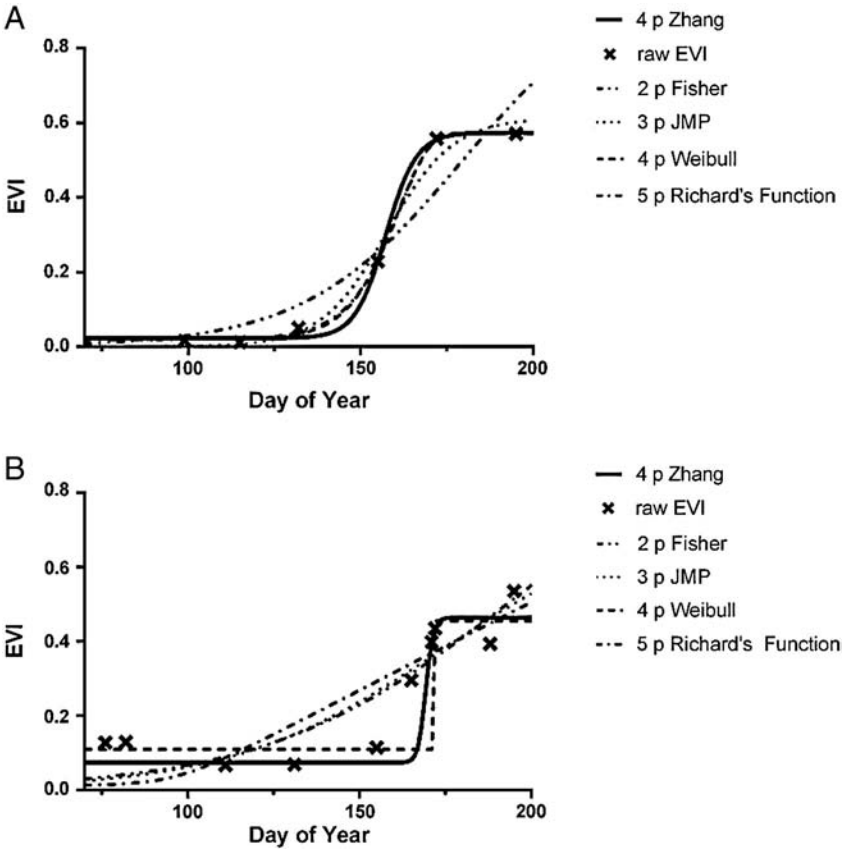

Fig. 4. Modeled logistic fits of EVI on sample hardwood (A) and mixed (B) plots. The four-parameter logistic model most consistently fit EVI.

during early phenology (PFR 1). NDVI typically responds first (PRF 2) with EVI initiating a sharp increase after leaves begin to emerge (PFR 3). This index insensitivity to early phenological changes is likely due to several factors. Reflectance at every pixel is the conglomeration of a variety of vegetation and background features, which may obscure observations of initial bud burst (Reed et al., 1994). While field observers may be able to note bud swelling and emergence, our data and others (e.g., Liang et al., 2011) suggest that the reflectance signal of bud burst is not sufficient to alter the reflectance signal from the larger mixed pixel because of the radiometric sensitivity of Landsat.

Accuracy for PFR 3 (leaf expansion) improved across all thresholds (Table 4), but was highly variable across plots. NDVI was most sensitive to canopy changes associated with PFR3, but still reported a mean absolute error of 18.5 days, and was highly variable across plots (stdev = 11.3 using a threshold 0.5) (Table 4). At PFR3, EVI was typically just beginning to increase, explaining its reduced accuracy compared to NDVI (mean absolute error = 20 days, stdev = 11.3 days using 30% max threshold) (Table 4).

PFR 4, the first day that the plot average phenology field rank reached full leaf out, was most consistently and accurately predicted by both vegetation indices. An EVI threshold of 0.3 was able to predict PFR 4 with a mean absolute error of 11.2 days (stdev = 8.4). Because some plots were overpredicted, and others underpredicted, the average error across the entire scene for the 2011 season was only one day later than field observations.

To determine if there might be a better threshold for determining the SOS, we identified the index value for EVI at the actual date of the field-measured PFR 4, which was very close to the standard 0.3 threshold (0.295 for deciduous broadleaf plots and 0.278 for mixed plots). Using this custom threshold instead of the common 0.3 EVI value improved mean absolute error from 11.2 to 9.74 days for deciduous broadleaf stands. Accuracy at mixed stands did not change using the custom threshold.

3.3. Accuracy and errors in the final phenology model

Across a variable landscape, our "best" method to predict spring phenology was based on the EVI index, fit with the sigmoid model proposed by Zhang et al. (2003), to predict PFR4 (full leaf out) using a 0.3 EVI threshold. In order to apply this on a pixel basis, across the landscape we developed a custom IDL script (Dick Jackson Software Consulting; available upon request). The resulting image for 2011 SOS across our study region is shown in Fig. 7.

The accuracy of this approach (field to index metric $r^2 = 0.95$, mean absolute error for all plot types = 11.2 days) was similar to, or an improvement over other studies (26 days with AVHRR, (White et al., 1997); 10 days with MODIS, (Zhang et al., 2006)) (Table 6). Accuracy was improved when considering only homogeneous stands. For example, in pure deciduous stands it was possible to make predictions of full leaf out with a mean absolute error of 9 days. The accuracy of SOS prediction at our four "pure" sugar maple stands (over 60% of the basal area) improved to 7 days mean absolute error.

In order to identify sources of variability and error in the model, we compared site characteristics and species composition to the accuracy of prediction across field sites. Spearman's Rho correlations indicated that SOS accuracy was higher at higher elevations. This could result from a more rapid greenup at upper elevations. We also found that accuracy increased with increasing percentages of hardwoods in general, and diffuse porous hardwoods specifically. This indicates that uniformity within hardwood species also reduces error. Similarly, accuracy was higher on plots with more balanced species composition (quantified as the standard deviation of species' percent basal area; $p = 0.04$ and $p = 0.03$ respectively). This indicates that when mature trees at low species density are present on a plot there may be a mismatch between basal area weighted field metrics and plot level spectral signatures. It is possible that a different weighting method based on canopy size or canopy fraction may better capture the contribution of anomalous species better than basal area based weightings.

Fisher et al. (2006) also found that in coarse-resolution satellite phenology studies, coniferous and mixed forests are difficult to segregate, resulting in compositional uncertainty. Our results indicate that mixed composition is still an issue at finer resolutions. Lower accuracy at mixed stands is expected because of inherent differences in coniferous and deciduous broadleaf phenology. Conifers maintain greenness throughout the winter, dampening changes in reflectance in the spring. The result is that the more subtle phenological changes are masked by the larger reflectance signal of existing vegetation. Instead of foliage appearance and expansion, biochemical signals are stronger drivers of the seasonal pattern in vegetation indices of coniferous forests (Richardson et al., 2009). Reflectance patterns may therefore differ between deciduous and conifer forest types and one global method of quantifying phenological patterns may not be appropriate.

One possible solution is to include ancillary information of species composition to help refine phenological assessments. While highly accurate species distribution maps are hard to come by at this resolution, there are sources of high quality hardwood, conifer and mixed forest coverages. When forest type (conifer/hardwood/mixed) was included to refine satellite predictions of SOS, mean absolute error decreased from 11 days to 8 days across all plots. This is similar to the improved accuracy found by Liang et al. (2011) when forest community type was included in assessments.

However, our results also indicate that error in SOS prediction arises when there is a mixture of species within hardwood-dominated stands. Different deciduous broadleaf species are adapted to leaf out and break

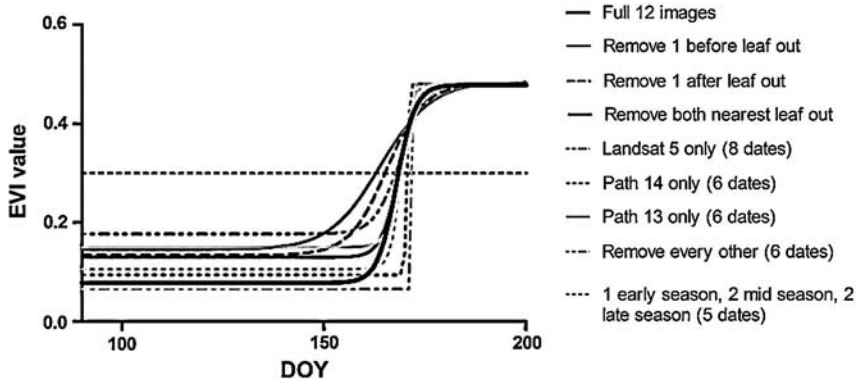

Fig. 5. While there was no significant difference in accuracy among various temporal resolution scenarios, predicted SOS based on the EVI 0.3 threshold (dotted horizontal line) varied up to 9 days depending on image availability.

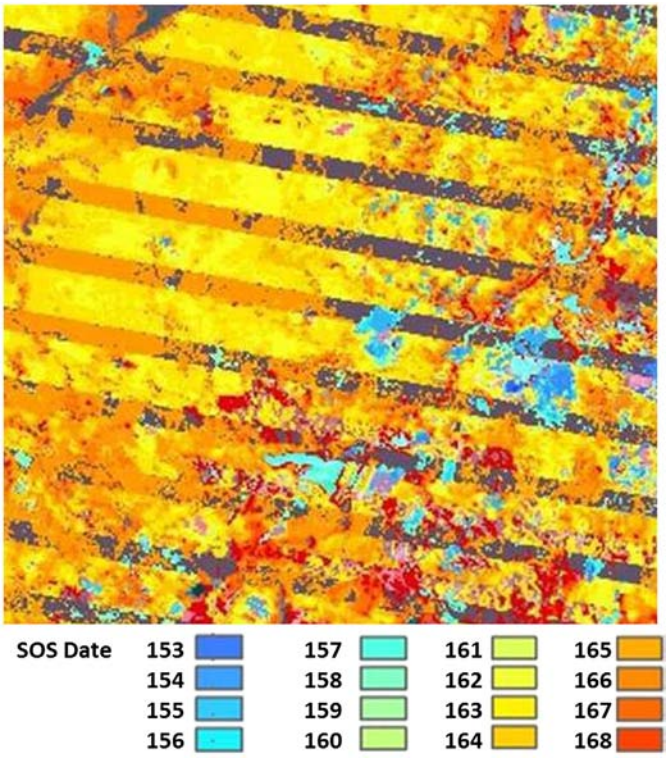

Fig. 6. The striped pattern in this predicted SOS image matches the scan line correction error in the June 4 Landsat 7 image. This highlights that even though resulting predictions are not significantly different, image availability can alter the final SOS prediction, in this case by two days between scan line error locations.

bud at different times (Lechowicz, 1984). If one species leafs out early while another species on the same plot remains dormant, this will be reflected in the mixed spectra observed by the sensor. Richardson et al. (2009) found phenology metrics of sub-regions within a single camera view at the Bartlett Research Forest varied by almost a week. On our hardwood dominated plots, PFR dates for individual trees varied by 19 days on average, but up to 33 days on mixed hardwood plots (see Section 3.1 field phenology assessments).

Predicting an event that does not occur uniformly even at a single location is further complicated by variability resulting from sensor, atmosphere, misregistration, calibration or curve fits. Considering this, it seems improbable to achieve accuracies better than the 10–30 days typically reported in the literature. While it may be unrealistic to predict SOS with higher accuracy, precision should remain high when using a consistent approach. This indicates that while independent assessments of SOS should be interpreted conservatively, it is more likely that comparisons of SOS changes across landscapes or over time following a uniform methodology could still provide useful information about interannual trends and spatial patterns in phenology in spite of poor accuracy. More specifically, if interested in analyzing possible changes in phenology at any one location over time, many sources of error in SOS predictions (e.g., landscape position and subsequent sensor view angle, and likely even species composition barring a catastrophic event) would be expected to be controlled — partially accounted for because many of these factors do not change substantially over the period of analysis. While the SOS in a mixed conifer/hardwood stand may be underestimated using remote sensing approaches, it is at least consistently underestimated. Therefore, comparisons of the same pixel over several years may provide a precise assessment of changing phenology, even if the accuracy of SOS predictions for any given year is poor.

To test the precision of SOS assessments at a given location over time, we compared sugar maple phenology data collected at the Proctor Maple Research Center in Underhill, VT (Wilmot, 2012) to Landsat predictions following the methods proposed here. The Proctor field assessments used an 8-class scale, with daily assessments of sugar maple trees at two locations where sugar maple dominated, but with other hardwood species occurring in small proportions. Because of frequent cloud cover at this location, only 7 of the 20 years of field data had corresponding SOS predictions from Landsat imagery. Predicted mean absolute error was 12.6 days for Wilmot's field measured phenology class 7 (initial leaf expansion) and 13.9 days of field measured phenology class 8 (full leaf expansion), consistent with accuracies reported for our field plots in 2011. Spearman's correlations showed a moderate correlation between the field measured and predicted phenology ($\rho = 0.54$) for the seven years where data overlaps. However, due to low sample size, this correlation was not significant ($p = 0.2$). A power analysis ($\alpha = 0.05$, $\beta = 0.8, \rho = 0.5$) indicated that 23 years of data would be required to establish a significant relationship. Considering that this independent field phenology data was collected on plots where not all trees were assessed and species composition was not uniform, the strength of this correlation, in combination with the 12.6 day mean absolute error indicates that interannual assessments of SOS can be successfully estimated using this approach.

3.4. Impact of timing and availability of imagery

It has been suggested that sensors such as Landsat may lack the temporal resolution required to accurately model vegetation phenology (Busetto et al., 2008). This is compounded by the common presence of cloud cover during the spring season. In order to test the impact of image availability and timing on the fit and accuracy of SOS predictions, we compared SOS accuracy across a variety of modeled and collected temporal resolutions. A comparison of field plots with different numbers of available calibration images (due to cloud cover) showed no significant difference in accuracy using our final phenology model ($F_{(6, 23)} = 0.976$, $p = 0.463$). The number of images within 16 days of the field measured SOS date (0, 1, 2 or 3 images) also did not significantly alter the SOS prediction accuracy ($F_{(3, 28)} = 0.371$, $p = 0.775$).

To control for inherent differences in SOS accuracy between plots, we also simulated the impact of temporal resolution at the four plots with the full 12 image dates by sequentially removing dates and re-predicting SOS. Again, there was no significant difference in the accuracy of 9 different temporal resolutions for predicted SOS ($F_{(8,27)} = 0.539$, $p = 0.82$) (Table 5).

However, even though there was no significant statistical difference between temporal resolution scenarios overall, there was variability in predicted SOS resulting from different configurations of image availability. For example, predicted SOS could vary as much as 9 days at a single plot (Fig. 5) based on different simulated temporal resolutions. While this variability is not significant, it can result in patterns across the landscape that are due to image availability and not topographic, climatic or species related differences in actual phenology. In our study area, this included differences in SOS predictions by up to 13 days between neighboring pixels (Fig. 6). This is especially evident when using Landsat ETM+ imagery with the scan line correction error. The striped pattern in predicted SOS for 2011 in Fig. 6 results from missing June 4 data in the ETM+ calibration image. This is similar to results presented by Melaas et al. (2013), indicating that cloud, cloud shadow and other sources of noise could bias estimated spring onset by 7 days. To minimize the potential impact of missing image dates in regions of frequent cloud cover, it is essential to conduct a high quality atmospheric correction and cloud screening. These results indicate that additional masking may be required to exclude any pixels biased by image availability.

4. Conclusions

The ability to monitor phenology at a landscape scale is an important tool to better understand long-term trends and spatial patterns in forest

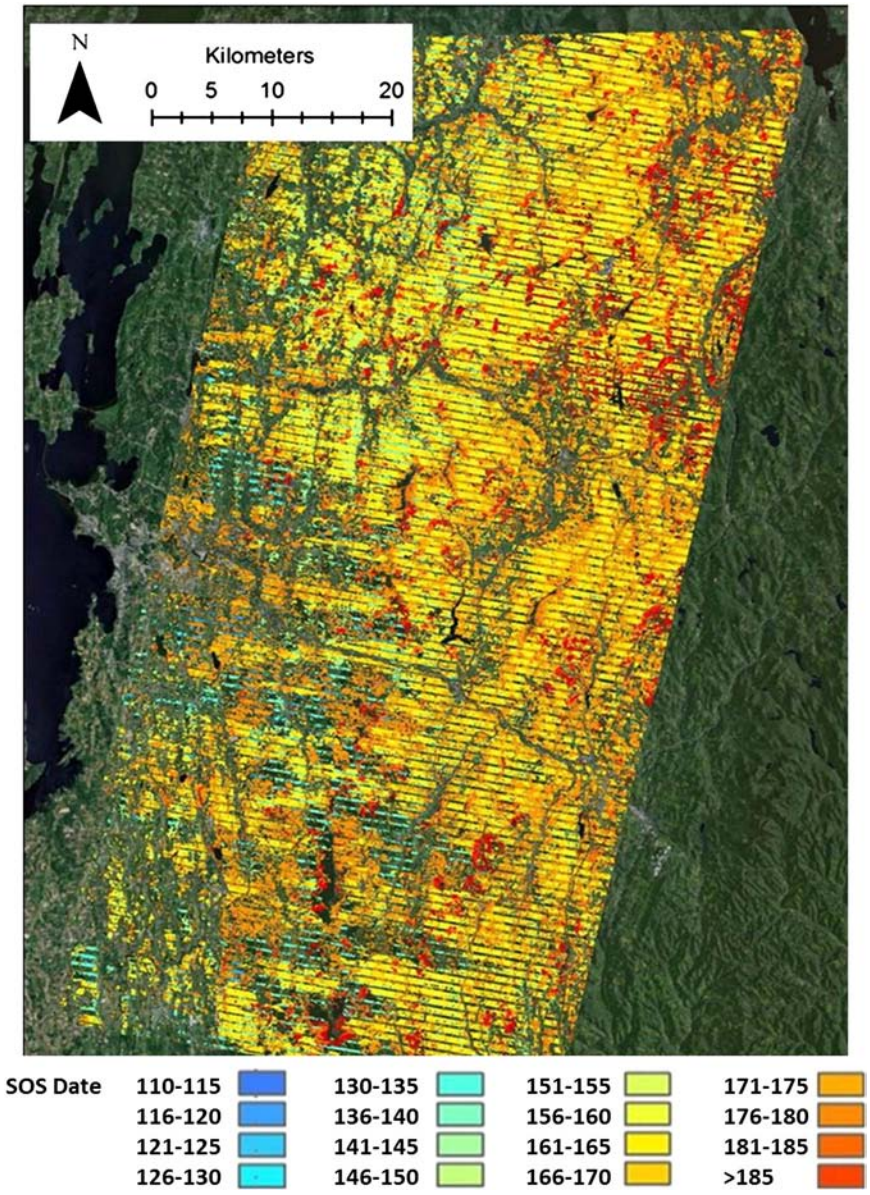

Fig. 7. Predicted SOS in the path 13/14 (row 29) study area using the final best method (EVI, fit by Zhang to predict PFR4 with a 0.3 threshold) and masking images based on QAQC guidelines covered in Section 3.4.

response to changing climate. Remote sensing methods provide this scale of analysis but need to represent meaningful phenology events as seen in the field, with appropriate guidance for interpretation of the accuracy of results. By comparing a suite of remote sensing-derived indices and mathematical fits to extensive field measurements we propose a standardized model for Landsat-based assessments of phenology. While other methods may prove more accurate in other regions, the method summarized here represents a challenging test in a region of persistent snow, high cloud frequency, broad range of elevations and mixed forest type. While this method was stable across forest types, mean absolute error decreased from 11 to 7 days when differences in species composition and elevation were controlled.

This represents improved (White et al., 1997) or comparable (Zhang et al., 2006) accuracy over other studies matching in situ phenology records to satellite metrics, but our study also provided a link to specific phenology characteristics measured on the ground across a range of forest composition. We found that vegetation indices are more closely related to visual assessments of phenology than to hemispherical photo derived canopy metrics such as canopy density or LAI. Indices were most accurate in their predictions of full leaf out, as opposed to bud burst or leaf emergence phenology stages.

Because plots of similar species composition and landscape position were consistently predicted, the accuracy of *changes* in SOS timing are likely higher than the accuracy within a given year across a heterogeneous landscape. The combined effect of standardizing multiple factors at once (i.e., species composition and landscape position) as would occur when assessing relative changes at the same pixel over time, suggests that temporal assessments of changes in spring phenology are valid in spite of limited accuracy across a heterogeneous landscape.

Although pixel-based measurements result in a loss of the individual tree or species-specific phenology that is so aptly captured in field studies, it does allow for an analysis of spatial and temporal patterns across the landscape (Reed et al., 2009). Our ongoing efforts involve applying the methodology outlined here to a 25-year archive of Landsat imagery in order to investigate long-term changes, inter-annual variability, and spatial patterns in phenology. This information will help quantify the impact of changing climate on the forested landscape.

Table 6
Prediction accuracy of various thresholds using the Zhang et al. (2003) sigmoid model. Threshold performance is reported as the average days' difference between actual and estimated PFR stages across all 32 plots. The most accurate assessment was obtained using EVI at a threshold value of 0.3 as an approximation of the first day of full leaf out (PFR 4).

Index	Threshold	Average error (days)			Stdev of error (days)		
		PRF 2	PRF 3	PRF 4	PRF 2	PRF 3	PRF 4
EVI	0.3	32.3	24.1	11.2	10.7	10.0	8.4
	0.5		47.4	26.5		18.2	17.5
	0.7						
	30% max	58.2	20.2	14.6	12.9	11.3	11.7
	50% max		23.6	12.8			8.7
	70% max		29.9	13.7		13.9	12.1
	Max slope	28.2	23	20.2	15.8	15.4	20.8
	Min slope		33.1	14.8		12.8	11.1
	Slope to 0			17.9			11.5
NDVI	0.3	19.2			14.1		
	0.5	22	18.5	19.5	14.6	11.3	14.0
	0.7			16.1			10.8
	30% max	20			15.5		
	50% max	22.3	19.1	19.8	15.0	11.2	14.1
	70% max			18.2			14.1
	Max slope	29.8	27.9		20.2	22.0	
	Min slope	30.6	24.3	18.2	17.9	15.2	9.3
	Slope to 0			18.3			8.3

References

Badeck, F. -W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J., et al. (2004). Responses of spring phenology to climate change. *New Phytologist*, *162*, 295–309.

Bassow, S. L., & Bazzaz, F. A. (1998). How environmental conditions affect canopy leaf-level photosynthesis in four deciduous tree species. *Ecology*, *79*, 2660–2675.

Boyd, D. S., Foody, G. M., Curran, P. J., Lucas, R. M., & Honzak, M. (1996). An assessment of radiance in Landsat TM middle and thermal infrared wavelengths for the detection of tropical forest regeneration. *International Journal of Remote Sensing*, *17*, 249–261.

Bradley, B. A., Jacob, R. W., Hermance, J. F., & Mustard, J. F. (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment*, *106*, 137–145.

Breda, N. J. J. (2003). Ground-based measurements of leaf area index: A review of methods, instruments, and current controversies. *Journal of Experimental Botany*, *54*, 2403–2417.

Broge, N. H., & Leblanc, E. (2000). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, *76*, 156–172.

Busetto, L., Meroni, M., & Colombo, R. (2008). Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. *Remote Sensing of Environment*, *112*, 118–131.

Chander, G., Markham, B., & Helder, D. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, *113*, 893–903.

Chavez, P.S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, *24*, 459–479.

Cleland, E. E., Chuine, I., Menzel, A., Mooney, H., & Schwartz, M.D. (2007). Shifting plant phenology in response to global change. *Trends in Ecology and Evolution*, *22*.

deBeurs, K. M., & Henebry, G. M. (2010). Spatio-temperoral statistical methods for modeling land surface phenology. In I. L. Hudson, & M. R. Keatley (Eds.), *Phenological research* (pp. 177–208). Springer Science & Business Media.

Delbart, N., Toan, T. L., Kergoat, L., & Fedotova, V. (2006). Remote sensing of spring phenology in boreal regions: A free of snow-effect method using NOAA-AVHRR and SPOT-VGT data (1982–2004). *Remote Sensing of Environment*, *101*, 52–62.

Dengler, N. G., Mackay, L. B., & Gregory, L. M. (1975). Cell enlargement and tissue differentiation during leaf expansion in beech, *Fagus grandifolia*. *Canadian Journal of Botany*, *53*, 2846–2865.

Dukes, J. S., Pontius, Jennifer, Orwig, David, Garnas, Jeffrey R., Rodgers, Vikki L., Brazee, Nicholas, et al. (2009). Responses of insect pests, pathogens, and invasive plant species to climate change in the forests of northeastern North American: What can we predict? *Canadian Journal of Forest Research*, *39*, 231–248.

Eichelmann, H., Oja, V., Rasulov, B., Padu, E., Bichele, I., Pettai, H., et al. (2008). Development of leaf photosynthetic parameters in *Betula pendula* Roth leaves: Correlations with photosystem I density. *Plant Biology*, *6*, 307–318.

Elmore, A. J., Guinn, S. M., Minsley, B. J., & Richardson, A.D. (2012). Landscape controls on the timing of spring, autumn and growing season length in mid-Atlantic forests. *Global Change Biology*, *18*, 656–674.

Fisher, J. I., & Mustard, J. F. (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment*, *109*, 261–273.

Fisher, J. I., Mustard, J. F., & Vadeboncoeur, M.A. (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment*, *100*, 265–279.

Fitzjarrald, D. R., Acevedo, Otavio C., & Moore, Kathleen E. (2001). Climatic consequences of leaf presence in the Eastern United States. *Journal of Climate*, *14*, 598–614.

Frazer, G. W., Canham, C. D., & Lertzman, K. P. (1999). *Gap Light Analyzer (GLA): Imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, users manual and program documentation.* In: Simon Fraser University, Burnaby, British Columbia, and Institute of Ecosystem Studies, Milbrook, New York.

Frazer, G. W., Fournier, R. A., Trofymow, J. A., & Hall, R. J. (2001). A comparison of digital and film fisheye photography for analysis of forest canopy structure and gap light transmission. *Agricultural and Forest Meteorology*, *109*, 249–263.

Hmimina, G., Dufrene, E., Pontailler, J. -Y., Delpierre, N., Aubinet, M., Caquet, B., et al. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment*, *132*, 145–158.

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 Sensing of Environment*, *83*, 195–213.

Hufkens, K., Friedl, Mark, Sonnentag, Oliver, Braswell, Bobby H., Milliman, Thomas, & Richardson, Andrew D. (2012). Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment*, *117*, 307–321.

Ibanez, I., Richard, B., Primack, Abraham J., Miller-Rushing, Elizabeth Ellwood, Higuchi, Hiroyoshi, Lee, Sang Don, et al. (2010). Forecasting phenology under global warming. *Philosophical Transactions of the Royal Society B*, *365*, 3247–3260.

ITT Visual Information Solutions (2009). *ENVI Versions 4.7.* (In).

Jarčuška, B., Kucbel, S., & Jaloviar, P. (2010). Comparison of output results from two programmes for hemispherical image analysis: Gap Light Analyser and WinScanopy. *Journal of Forest Science*, *56*, 147–153.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., et al. (2004). Review of methods for in situ leaf area index determination. Part I. Theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, *121*, 19–35.

Jönsson, A. M., Eklundh, L., Hellström, M., Bärring, L., & Jönsson, P. (2010). Annual changes in MODIS vegetation indices of Swedish coniferous forests in relation to snow dynamics and tree phenology. *Remote Sensing of Environment*, *114*, 2719–2730.

Kramer, K., & Hänninen, Heikki (2009). The annual cycle of development of trees and process-based modelling of growth to scale up from tree to the stand. In Asko Noormets (Ed.), *Phenology of ecosystem processes* (pp. 201–227). New York: Springer.

Leblanc, S., Chen, J., Fernandes, R., Deering, D., & Conlye, A. (2005). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology*, *129*, 187–207.

Lechowicz, M. J. (1984). Why do temperate deciduous trees leaf out at different times? Adaptation and ecology of forest communities. *The American Naturalist*, *124*, 821–842.

Liang, L., Schwartz, Mark D., & Fei, Songlin (2011). Validating satellite phenology through intensive ground observation and landscape scaling in mixed seasonal forest. *Remote Sensing of Environment*, *115*, 143–157.

McNeil, B. E., Denny, E., & Richardson, A.D. (2008). Coordinating a northeast regional phenology network. *Bulletin of the Ecological Society of America*, *89*, 188–190.

Melaas, E. K., Friedl, M.A., & Zhu, Z. (2013). Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data. *Remote Sensing of Environment*, *132*, 176–185.

Menzel, A. (2002). Phenology: Its importance to the global change community. *Climatic Change*, *54*, 379.

Minkova, T. V., & Logan, C. J. (2007). In W.S.D.o.N. Resources (Ed.), *Comparing spherical densitometry and hemispherical photography for estimating canopy closure* (Olympia, WA).

Nagai, S., Nasahara, K. N., Muraoka, H., Akiyama, T., & Tsuchida, S. (2010). Field experiments to test the use of the normalized-difference vegetation index for phenology detection. *Agricultural and Forest Meteorology*, *150*, 152–160.

Nobis, M., & Hunziker, U. (2005). Automatic thresholding for hemispherical canopy-photographs based on edge detection. *Agricultural and Forest Meteorology*, *128*, 243–250.

Reed, B. C., Brown, Jesslyn F., VanderZee, Darrel, Loveland, Thomas R., Merchant, James W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *International Association of Vegetation Science*, *5*, 703–714.

Reed, B. C., Schwartz, Mark D., & Xiao, Xiangmin (2009). Remote sensing phenology: Status and the way forward. In A. Noormets (Ed.), *Phenology of ecosystem processes: Applications in global change research* (pp. 231–246). New York: Springer.

Richardson, A.D., Bailey, A. S., Denny, E. G., Martin, C. W., & O'Keefe, J. (2006). Phenology of a northern hardwood forest canopy. *Global Change Biology*, *12*, 1174–1188.

Richardson, A.D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P., & Ollinger, S. V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, *19*, 1417–1428.

Richardson, A.D., Jenkins, J. P., Braswell, B. H., Hollinger, D. Y., Ollinger, S. V., & Smith, M. -L. (2007). Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Ecosystem Ecology*, *152*, 323–334.

Richardson, A.D., & O'Keefe, John (2009). Phenological differences between understory and overstory: A case study using the long-term Harvard forest records. In A. Noormets (Ed.), *Phenology of ecosystem processes* (pp. 87–117). New York: Springer.

Riederer, M., & Schönherr, J. (1988). Development of plant cuticles: Fine structure and cutin composition of *Clivia miniata* Reg. leaves. *Planta*, *174*, 127–138.

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*, 95–107.

Sonnentag, O., Hufkens, Koen, Teshera-Sterne, Cory, Young, Adam M., Friedl, Mark A., Braswell, Bobby H., et al. (2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, *152*, 159–177.

Soudani, K., Francois, C., Le Marie, G., Le Gantec, V., & Dufrene, E. (2006). Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. *Remote Sensing of Environment*, *102*, 161–175.

Soudani, K., le Maire, Guerric, Dufrene, Eric, Francois, Christophe, Delpierre, Nicolas, Ulrich, Erwin, et al. (2008). Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sensing of Environment*, *112*, 2643–2655.

Tucker, C. J., Slayback, D. A., Pinzon, J. E., Los, S. O., Myneni, R. B., & Taylor, M. G. (2001). Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *International Journal of Biometeorology*, *45*, 184–190.

U.S. Department of the Interior (2012). In Earth Resources Observation and Science Center (EROS) (Ed.), *USGS Global Visualization Viewer*.

Van Leeuwen, W. J., Davison, J. E., Casady, G. M., & Marsh, S. E. (2010). Phenological characterization of desert sky island vegetation communities with remotely sensed and climate time series data. *Remote Sens*, *2*, 388–415.

White, M.A., Running, S. W., & Thornton, P. E. (1999). The impact of growing-season length variability on carbon assimilation and evapotranpiration over 88 years in the eastern US deciduous forest. *International Journal of Biometeorology*, *42*, 139–145.

White, M.A., Thornton, P. E., & Running, S. W. (1997). A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, *11*, 217–234.

Wilmot, S. (2012). Sugar maple bud development, Proctor Maple Research Center. *The Vermont Monitoring Cooperative Online Database* (Available: http://www.uvm.edu/vmc/research/summary.php?id=13. Accessed February 23, 2014).

Zhang, X., Friedl, M.A., & Schaaf, C. B. (2006). Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *Journal of Geophysical Research*, *111*.

Zhang, X., Friedl, M.A., Schaaf, C. B., & Strahler, A. H. (2004). Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. *Global Change Biology*, *10*, 1133–1145.

Zhang, X., Friedl, Mark A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., et al. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*, 471–475.