Global Change Biology (2009), doi: 10.1111/j.1365-2486.2009.01910.x

Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006

MICHAEL A. WHITE*, KIRSTEN M. DE BEURS†, KAMEL DIDAN‡, DAVID W. INOUYE§, ANDREW D. RICHARDSON¶, OLAF P. JENSEN‖, JOHN O'KEEFE**, GONG ZHANG*, RAMAKRISHNA R. NEMANI††, WILLEM J. D. VAN LEEUWEN‡‡, JESSLYN F. BROWN§§, ALLARD DE WIT¶¶, MICHAEL SCHAEPMAN¶¶, XIOAMAO LIN‖‖, MICHAEL DETTINGER***, AMEY S. BAILEY†††, JOHN KIMBALL‡‡‡, MARK D. SCHWARTZ§§§, DENNIS D. BALDOCCHI¶¶¶, JOHN T. LEE‖‖‖‖ and WILLIAM K. LAUENROTH****

*Department of Watershed Sciences, Utah State University, Logan, UT, USA, †Department of Geography, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, ‡Institute for the Study of Planet Earth, the University of Arizona, Tucson, AZ, USA, §Department of Biology, University of Maryland, College Park, MD, USA, ¶Complex Systems Research Center, University of New Hampshire, Durham, NH, USA, ‖Center for Limnology, University of Wisconsin-Madison, Madison, WI, USA, **Harvard Forest, Petersham, MA, USA, ††NASA Ames Research Center, Moffett Field, CA, USA, ‡‡Office of Arid Lands Studies & Department of Geography and Regional Development, University of Arizona, Tucson, AZ, USA, §§Land Sciences Division, U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD, USA, ¶¶Centre for Geo-information, Wageningen-UR, Wageningen, The Netherlands, ‖‖Campbell Scientific Inc., Logan, UT, USA, ***US Geological Survey, Scripps Institute Oceanography, La Jolla, CA, USA, †††Hubbard Brook Experimental Forest, Campton, NH, USA, ‡‡‡Flathead Lake Biological Station, Division of Biological Sciences, University of Montana, Polson, MT, USA, §§§Department of Geography, University of Wisconsin-Milwaukee, Milwaukee, WI, USA, ¶¶¶Department of Environmental Science, Policy & Management, University of California, Berkeley, Berkeley, CA, USA, ‖‖‖‖Environmental Physics Group, University of Maine, Dept. PSE, Orono, ME, USA, ****Graduate Degree Program in Ecology & Warner College of Natural Resources, Colorado State University, Fort Collins, CO, USA

Abstract

Shifts in the timing of spring phenology are a central feature of global change research. Long-term observations of plant phenology have been used to track vegetation responses to climate variability but are often limited to particular species and locations and may not represent synoptic patterns. Satellite remote sensing is instead used for continental to global monitoring. Although numerous methods exist to extract phenological timing, in particular start-of-spring (SOS), from time series of reflectance data, a comprehensive intercomparison and interpretation of SOS methods has not been conducted. Here, we assess 10 SOS methods for North America between 1982 and 2006. The techniques include consistent inputs from the 8 km Global Inventory Modeling and Mapping Studies Advanced Very High Resolution Radiometer NDVIg dataset, independent data for snow cover, soil thaw, lake ice dynamics, spring streamflow timing, over 16 000 individual measurements of ground-based phenology, and two temperature-driven models of spring phenology. Compared with an ensemble of the 10 SOS methods, we found that individual methods differed in average day-of-year estimates by ± 60 days and in standard deviation by ± 20 days. The ability of the satellite methods to retrieve SOS estimates was highest in northern latitudes and lowest in arid, tropical, and Mediterranean ecoregions. The ordinal rank of SOS methods varied geographically, as did the relationships between SOS estimates and the cryospheric/hydrologic metrics. Compared with ground observations, SOS estimates were more related to the first leaf and first flowers expanding phenological stages. We found no evidence for time trends in spring arrival from ground- or model-based data; using an ensemble estimate from two methods that were more closely related to ground observations than other methods, SOS

Correspondence: Michael A. White, tel. + 1 435 797 3794, fax + 1 435 797 187, e-mail: mikew.usu@gmail.com

© 2009 Blackwell Publishing Ltd

trends could be detected for only 12% of North America and were divided between trends towards both earlier and later spring.

Keywords: bloom, budburst, climate change, flower, growing season, land surface phenology, seasonality

Received 25 September 2008; revised version received 26 January 2009 and accepted 10 February 2009

Introduction

Phenology, the study of the timing of recurring biological cycles and their connection to climate, is experiencing a renaissance in global change science as it is providing society with an independent measure on how ecosystems are responding to climate change (Badeck *et al*., 2004; Linderholm, 2006; Parmesan, 2006). For example, recent synthesis studies on vegetation phenology are showing a widespread trend towards earlier arrival of spring across the northern hemisphere, as measured by the dates of budbreak, flowering, or partial or full leaf expansion (Chmielewski & Rotzer, 2001; Menzel *et al*., 2006; Schwartz *et al*., 2006). Specifically, long-term phenological data indicate that the date of spring is advancing by, on average, 1–2 days per decade but varies depending on study period (Scheifinger *et al*., 2002; Schaber & Badeck, 2005).

Decadal trends and interannual variability in vegetation phenology are important because they affect carbon, water and energy exchange between the vegetation and the atmosphere. Annually integrated net ecosystem CO_2 exchange of deciduous vegetation is strongly related to length of the carbon uptake period (Baldocchi *et al*., 2001): springtime warming advances phenology, accelerates carbon uptake and reduces the amount of CO_2 in the atmosphere. Springtime advances in phenology also alter the surface energy balance and accelerate transpiration (Wilson & Baldocchi, 2001), which humidifies the atmosphere, alters the rate of growth of the planetary boundary layer and affects clouds and precipitation (Fitzjarrald *et al*., 2001; Schwartz & Crawford, 2001). In some instances, however, earlier spring growth advances soil water depletion, enhancing mid-summer drought, thus counteracting higher early spring carbon assimilation (White & Nemani, 2003; Angert *et al*., 2005). Climate warming and earlier spring growth in combination with forest fuel buildup in the south-western united states also has resulted in an increasing number of large and severe wildfires (Westerling *et al*., 2006).

Methodological advances have accompanied the re-emergence of vegetation phenology research (Cleland *et al*., 2007), with at least five methods now available: (1) networks observing species-specific plants and plant communities, often with concurrent climate observation; (2) phenology modeling; (3) eddy covariance flux towers; (4) global change experiments; and (5) digital cameras and remote sensing – the main topic of this research. Methodologies to estimate phenology from satellite remote sensing are proliferating yet a universally accepted definition of spring arrival does not exist. Within the rubric of the term land surface phenology (LSP) – here defined as the study of the spatio-temporal development of the vegetated land surface as revealed by synoptic spaceborn sensors – terms such as greenup, leaf-out, green wave, and start-of-spring (SOS) appear to be interchangeable but in reality may represent different process or events. Here, we adopt the term SOS and define it conceptually as a rapid sustained increase in remotely sensed greenness after the longest annual period of photosynthetic senescence (SOS has also been defined as start-of-season; we use 'spring' to distinguish our analysis from fall phenology). Critically, in addition to radiation absorption by vegetation canopies, LSP includes the aggregate, confounding influence of atmospheric contamination, cloud cover, snow cover, soil wetness, and bidirectional viewing effects, Therefore, LSP must be considered to be related, but not identical to, plant phenology.

Although efforts have been made to explore differences between SOS methods (Schwartz *et al*., 2002), researchers do not comprehensively understand how the myriad definitions and methods are related to ground-based phenology and related processes such as changes in snow cover, soil thaw, ice, and hydrology. A survey of late 20th and early 21st century phenology literature for North America highlights the conflicting results obtained from satellite-based methods. Researchers have found scattered trends towards earlier and later SOS (Reed, 2006), earlier SOS everywhere except the southeast (Zhang *et al*., 2007), spring greening only in the southeast (Xiao & Moody, 2005), no overall continental trend in spring (Piao *et al*., 2007), a strong continental trend toward earlier spring (Zhou *et al*., 2001), or rare positive trends in June–August greenness in the high arctic (Goetz *et al*., 2005) but almost no trends in spring (Bunn & Goetz, 2006). While we recognize that differing results are likely influenced by variable study periods, satellite platform and atmospheric corrections, study area, compositing schemes, and spatial resolution, we submit that a central and

largely unaddressed uncertainty is the algorithm selected to extract SOS estimates.

Here, we use 10 methods ranging from simple empirical approaches to more complex mathematical models to estimate SOS for North America using a consistently processed remote sensing dataset. We then conducted a four-part analysis: (1) intercomparison of the SOS estimates among methods; (2) interpretation of SOS estimates with cryospheric and hydrologic indicators of interannual variability; (3) assessment of SOS estimates using ground measured phenology and temperature driven phenology models; and (4) an ensemble estimate of 1982–2006 trends in SOS for North American ecoregions based on a selection of SOS methods most consistent with ground data.

Materials and methods

Our overall approach was to compare SOS, cryospheric/hydrologic metrics, and measured and modeled plant phenology, principally at the level of whole ecoregions but sometimes for specific pixels or for all of North America. Here, we present an encapsulated methodology section; full details are available in Appendix A. We used the ecoregion concept to reduce the level of complexity and to simplify the presentation of results (this approach will obscure within-ecoregion differences related to latitude, elevation, or C_3 vs. C_4 species – all of which would be suitable for subsequent intercomparison analyses). For our ecoregion map, we used the level 1 (for most analyses) and level 3 (for detailed trend analysis) US Environmental Protection Agency Ecoregions of North America (http://www.epa.gov/wed/pages/ecoregions/na_eco.htm; there are 15 level 1 and 182 level 3 ecoregions). All analyses were native to or reprojected and resampled to the 8 km Albers Conic Equal Area projection of the remotely sensed data.

When conducting our interpretation and assessment, we used Spearman's correlations or reduced major axis correlation. With numerous statistical tests, the probability of a Type I error is high, but standard Bonferroni's corrections often result in unacceptably high Type II error rates. Other methods for multiple comparison error correction exist but no method is universally accepted and each will promote either Type I or Type II errors. As our central intent is to establish the relative ordinal relationship among SOS estimates, cryospheric/hydrologic metrics, and measured and modeled phenology, formal rejection or acceptance of the null hypothesis is not critical and we here report correlations as being greater than or less than the standard 0.05 cutoff without specific claims of statistical significance.

Satellite SOS estimates

We estimated 1982–2006 SOS with the 8 km (64 km^2 pixels) 15-day composited 1982–2006 normalized difference vegetation index (NDVI) records from the Global Inventory Modeling and Mapping Studies Advanced Very High Resolution Radiometer NDVIg dataset (Pinzon *et al*., 2005). NDVI is a widely used proxy indicator of vegetation canopy function and is related to the absorption of photosynthetically active radiation by plant canopies (Asrar *et al*., 1984). Other processing schemes exist (Smith *et al*., 1997; Kogan & Zhu, 2001; Pedelty *et al*., 2007) but none is free of criticism and NDVIg is widely used for trend analysis. Finer resolution datasets are also available but either lack within- and among-sensor corrections or cover shorter durations (e.g. the Moderate Resolution Imaging Spectroradiometer, Raytheon Space and Airborne Systems, El Segundo, CA, USA). Using NDVIg data, we estimated SOS for North America using 10 methods from four categories (Table 1, Appendix A). All methods also predict the end-of-season day-of-year (DOY) but here we consider only SOS. For SOS methods and related analyses, we refer to the combination of a single pixel and a single year as a pixel-year.

Cryospheric/hydrologic comparisons

We assembled data for snow cover, soil thaw, lake ice dynamics, and spring hydrology (see Appendix A for full details). Each dataset had variable temporal coverage, thus limiting the range of SOS analyses. Briefly, we used: the date of initial snow melt determined from multiple sensors; microwave-detected dates of soil thaw; lake ice breakup from gridded visual observations; and spring snowmelt onset date and the center of flow timing from streamflow records (these hydrologic metrics are designed to track the effects of temperature and should not be taken to solely represent moisture availability). For comparisons between these ancillary datasets and SOS, we calculated ecoregion annual averages and used Spearman's rank correlation coefficient, which relates the ordinal rankings of vectors without distributional assumptions, i.e. we seek to compare early (late) SOS vs., for example, early (late) snowmelt.

Measured and modeled plant phenology comparisons

We used 16 401 records of ground-measured plant phenology from 19 networks or sites covering most of North America (Table 2, Fig. 1). For the purposes of this research, the plant phenology data represent both a tremendously rich data source but also extensive problems including at least five central challenges: (1)

© 2009 Blackwell Publishing Ltd,

Table 1 Start-of-spring methods: acronym, full name, and estimation category and brief algorithm description (see Appendix A for full details)

Method	Full name	Algorithm: SOS estimate
Quadratic	Quadratic	Conceptual-mathematical: first composite period of growing degree accumulation best fitting the observed NDVI time series
NDVI 0.2	NDVI 0.2	Global threshold: NDVI exceeds 0.2
NDVI 0.3	NDVI 0.3	Global threshold: NDVI exceeds 0.3
DMA	Delayed Moving Average	Conceptual-mathematical: smoothed NDVI exceeds expected value of near-term historical NDVI
HANTS-FFT	Harmonic Analyses of NDVI Time-Series – Fast Fourier Transform	Conceptual-mathematical: maximum increase on Fourier approximation of NDVI
Timesat	Timesat	Conceptual-mathematical: high amplitude divergence from a multiple-model NDVI fit
$\text{Midpoint}_{\text{pixel}}$	$\text{Midpoint}_{\text{pixel}}$	Local threshold: NDVI exceeds locally tuned threshold; run for every pixel
PAT	Percent-Above-Threshold	Local threshold: NDVI exceeds locally tuned threshold; run for the group behavior of all pixels within an ecoregion
Gaussian	Gaussian	Hybrid: average date when Gaussian fit of NDVI exceeds three global thresholds
$\text{Midpoint}_{\text{cluster}}$	$\text{Midpoint}_{\text{cluster}}$	Local threshold: NDVI exceeds locally tuned threshold; run for time series aggregated to a cluster level

Table 2 List of phenological observation sites and/or networks

Site/network name	Duration	Sites	Sp	Stages	*n*
Arctic LTER*	1996	1	8	AMB FF FL	19
Chequemegon Ecosystem Atmosphere Study	1999–2006	2	2	FL FE	32
GLOBE	1998–2006	155	71	LB	386
Harvard Forest LTER	1990–2006	1	17	LB 75	530
Howland Research Forest	1990–2006	1	2	LB	28
Hubbard Brook Experimental Forest*†	1989–2005	5	3	FL LB	510
Konza Prairie LTER*†	1982–1987	1	31	FF LB	332
Life Cycles	2001–2004	11	44	AMB FF	47
Long Lake Conservation Center	1998–2006	1	139	AMB FB FF	167
Niwot Ridge LTER*†	1984–1992	1	3	FB FF FI LB	74
North American Lilac Network	1982–2003	392	2	FF FL	5072
Oak Ridge National Laboratory DAAC	1984	2	3	FL	6
OSU Phenology Gardens	2005–2006	28	46	AMB FF	1255
Plantwatch	1991	401	33	FB FF FL	5041
Prairie Wetlands Learning Center	1999–2006	1	80	FF	112
Rocky Mountain Biological Laboratory	1982–2006	1	2	FF	46
Sevilleta LTER Site*†	2000–2006	1	217	FF FL	1795
Shortgrass Steppe LTER*†	1995–2004	1	22	FB FF FL	429
Toolik Snowfence Experiment	1995–2002	2	20	FB FF FL	520

*Where 'first' events such as first leaf or first bloom were measured for multiple plants of the same species at the same location and where observations were taken every few days, we used the earliest observation.
†For sites in which exact geographic coordinates were not present, we averaged observations of the same species.
LTER, long term ecological research; GLOBE, global observations to benefit the environment; DAAC, distributed active archive center; OSU, Oklahoma State University; Sp, number of species; n, number of observations. Records and metadata are available from the corresponding author. Phenological stages are: AMB, anthesis/middle bloom; FB, first flower bud; FF, first flowers expanding; FI, first inflorescence; FL, first leaf; FE, full leaf expansion; LB, leaf budburst or budbreak; 75, 75% of full leaf expansion (used only at Harvard Forest and based on the expert opinion of the observer). LB is the first appearance of leaves from burst buds and is followed by FL, the initial expansion of leaves from buds. Metadata for most sites/networks is available at http://www.uwm.edu/Dept/Geography/npn/.

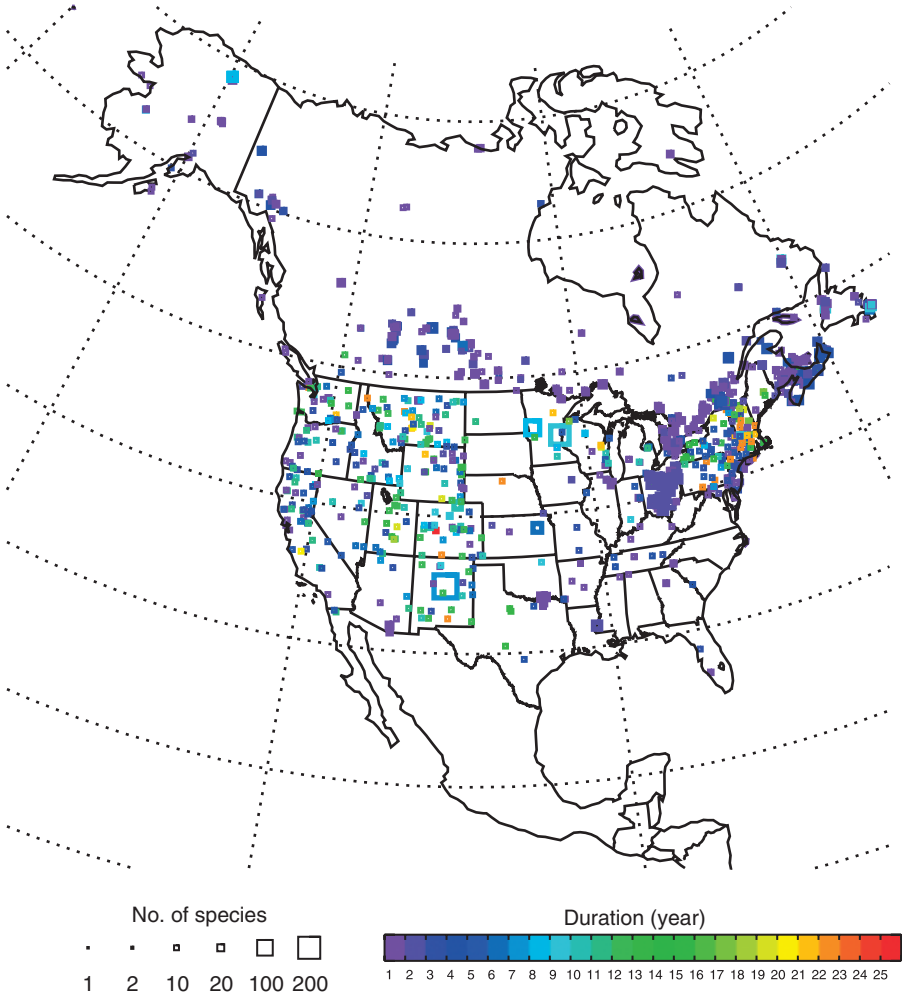

Fig. 1 Ground phenology records. Boxes show: location, center of box; number of species, box size; number of years observed at each site over the 1982–2006 study period, box color.

variable temporal and spatial coverage within and among networks; (2) species monitored may or may not represent general landscape phenology – the classical point vs. pixel problem in remote sensing assessments in which a single point observation may or may not represent the overall pixel characteristics; (3) different measurement protocols among networks; (4) unknown measurement accuracy and errors in data entry; (5) different phenological stages measured (e.g. leaf vs. bloom phenology and differences in how each stage is defined). In order to maximize data usage while minimizing problems, we conducted four separate analyses as follows.

First, we established the overall relationship between plant phenological events and SOS by separating the plant phenology data into the six phenological events (leaf budburst, first leaf, etc. Table 2) and extracting all the satellite SOS estimates for the corresponding pixel-year. We then constructed boxplots for each phenological event and the temporally and spatially co-located SOS estimates.

Second, we assessed the ability of SOS methods to represent individual, detailed plant phenology observations at specific locations. We used daily records of the fraction of photosynthetically active radiation absorbed by plant canopies (FPAR, scaled from 0 at the seasonal minimum to 1 at the seasonal maximum) recorded at the Bartlett Experimental Forest (44°3′52.7″N 71°17′17″W) using above and below-canopy quantum sensors covering $\sim 2000\,m^2$ (Jenkins *et al.*, 2007). Here, FPAR represents a continuous metric of canopy phenological development. We extracted corresponding pixel-year SOS values and compared the satellite DOY estimates against the measured FPAR curves.

Also within the category of specific location comparisons, we extracted all unique individual species

© 2009 Blackwell Publishing Ltd,

records of at least 10 years in duration for a particular phenological event (Table 2). For extractions with more than one site present in a given year, we calculated within-pixel averages and then a time series of annual averages. Next, we calculated correlations between the multi-year plant phenology records and the corresponding SOS pixel-years.

Third, we identified all 8 km pixel-years containing at least two observations of any species and phenological event. For all such pixel-years, we calculated the range and the 90% confidence interval (CI) as:

$$CI = \overline{X} \pm t\left(\frac{s}{\sqrt{n}}\right), \quad (3)$$

where for the plant phenology observations, $\overline{X}$ is the mean, t is the critical value from Student's t distribution for $n-1$ degrees of freedom, s is the sample standard deviation, and n is the number of observations. We categorized the corresponding SOS estimates as before the minimum or CI, within the range or CI, or after the maximum or CI.

Fourth, we initially attempted to construct 1982–2006 ecoregion-specific plant phenology averages but realized that sparse coverage in many ecoregions and variable species demographics made such an approach impracticable. Instead, we calculated annual average plant phenology for the entire study area. While we used all phenological events and species, we attempted to minimize observational or data entry errors by restricting the analysis, as above, to include only pixel-years with at least two observations and averaging multiple observations within a pixel-year. Owing to limited data availability, we restricted this analysis to the 1982–1999 record. For phenological time series with overlapping records for the same species, techniques exist to reduce measurement uncertainty (Linkosalo *et al.*, 1996); most of our data did not meet these criteria and we consequently did not smooth or filter the ground data. Future studies, though, should consider these techniques when appropriate. We then compared the plant phenology time series against SOS time series constructed from the corresponding pixel-years (using reduced major axis Type II regression, as there are uncertainties in both the x and y variables).

As a spatially and temporally continuous adjunct to the sparse plant-based measurements, we simulated the arrival of spring with the Spring Indices (SI) (Schwartz, 1997, 2003) and $D_{\text{leaf-out}}$ (Baldocchi *et al.*, 2005) models (see Appendix A for details). We predicted 1982–2003 first leaf and first bloom with SI and the date of initial overstory leafing with $D_{\text{leaf-out}}$. As both models were developed using data from temperate deciduous species, we restricted the model comparison with the Eastern Temperate Forest ecoregion.

Trend estimates

Based on results of the following interpretation and assessment, we selected the SOS methods most consistent with the interannual variability in cryospheric/hydrologic metrics and ground- and model-based phenology and calculated correlations (year as explanatory variable, SOS as multiple response y-vectors fitted simultaneously and for individual methods).

Results

SOS intercomparison

SOS methods differed in average estimated DOY, standard deviation, and ability to retrieve SOS estimates (Figs 2–4). As a 1982–2006 average, ensemble SOS estimates appeared to match expectations, with an early to late progression from southern to northern North America and latest SOS in the high arctic (Fig. 2, note that while the ecoregion approach has the benefit of simplifying a complex analysis, it tends to minimize the appearance of within-ecoregion variability, that is a strong 'greenwave' is present in the Eastern Temperate Forest but not visually apparent in our maps). Relative to the ensemble, though, individual methods often differed among ecoregions: Quadratic SOS was early in the West and late in the North while DMA exhibited nearly opposite patterns. Some methods were either consistently early (NDVI 0.2 and $\text{Midpoint}_{\text{cluster}}$) or consistently late (HANTS-FFT and PAT) but in most cases, methods exhibited both early and late anomalies, often of up to ± 60 days. Timesat most resembled the ensemble of model SOS estimates. Note that this comparison does not establish the 'correctness' of any one method, only the relative differences in timing.

The variability of SOS estimates, as measured by the ensemble SOS 1982–2006 standard deviation, was low in high latitude snow-dominated systems, high in the North American Desert ecoregion, and highest in Mediterranean California (Fig. 3). For individual methods, standard deviation anomalies calculated across years (Fig. 3) were inconsistent among ecoregions and methods: Quadratic, NDVI 0.2, and NDVI 0.3 were minimally variable in the West and more variable in the North with DMA again exhibiting opposite patterns. HANTS-FFT and $\text{Midpoint}_{\text{pixel}}$ were more consistently highly variable while PAT and Gaussian were more stable. As for average DOY, Timesat had the least SOS DOY variability anomalies compared with the ensemble of model SOS estimates. For most methods, Mediterranean California was usually either maximally or minimally variable.

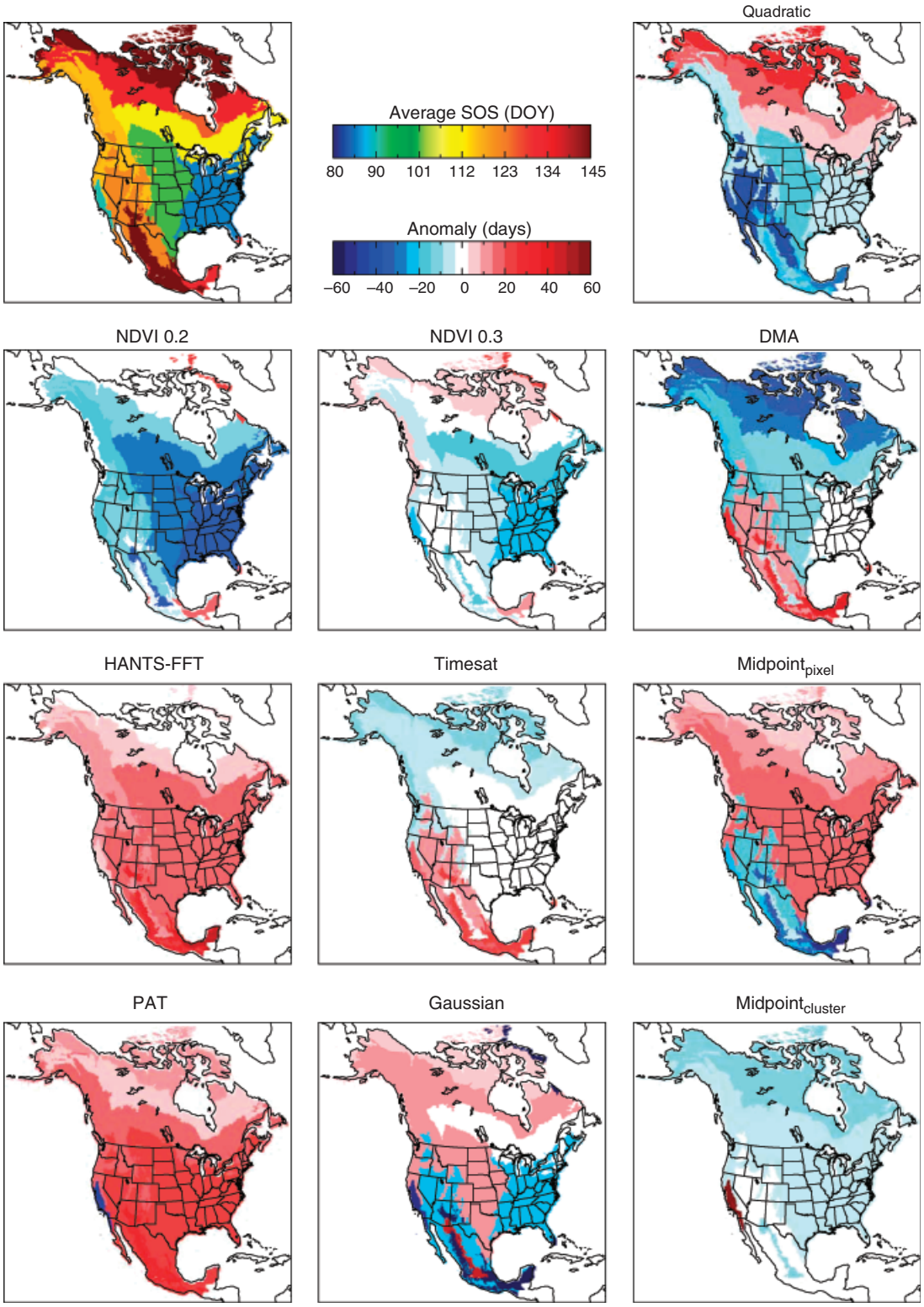

Fig. 2 Ensemble satellite-derived SOS averaged by ecoregion and over the 1982–2006 record (upper left panel, ecoregions visible as color blocks). Remaining panels show the SOS anomaly between individual methods and the ensemble, thus indicating locations in which individual SOS methods are earlier or later than the ensemble.

© 2009 Blackwell Publishing Ltd,

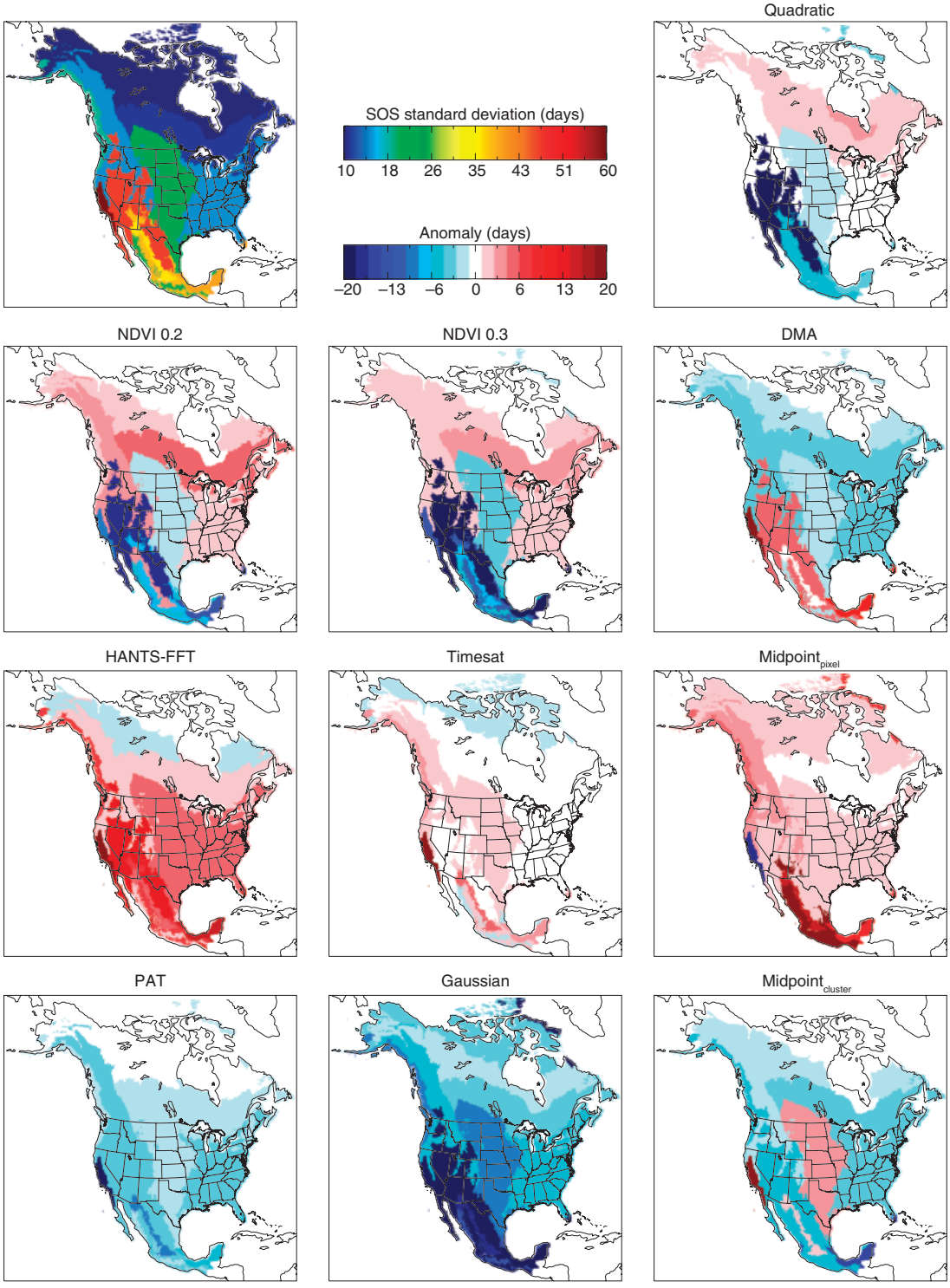

Fig. 3 Ensemble 1982–2006 SOS standard deviation (upper left panel, ecoregions visible as color blocks). Remaining panels show the SOS standard deviation anomaly between individual methods and the ensemble, thus indicating locations in which individual SOS methods were more or less variable than the ensemble.

SOS estimates were retrievable (where a retrieval refers to a successful estimation of SOS – failures occur due to method-specific treatment of missing data, screening, etc.) by all methods in all years only in the Hudson Plain ecoregion (Fig. 4). Retrievals averaged only 7 years in the Arctic Cordillera and were also low

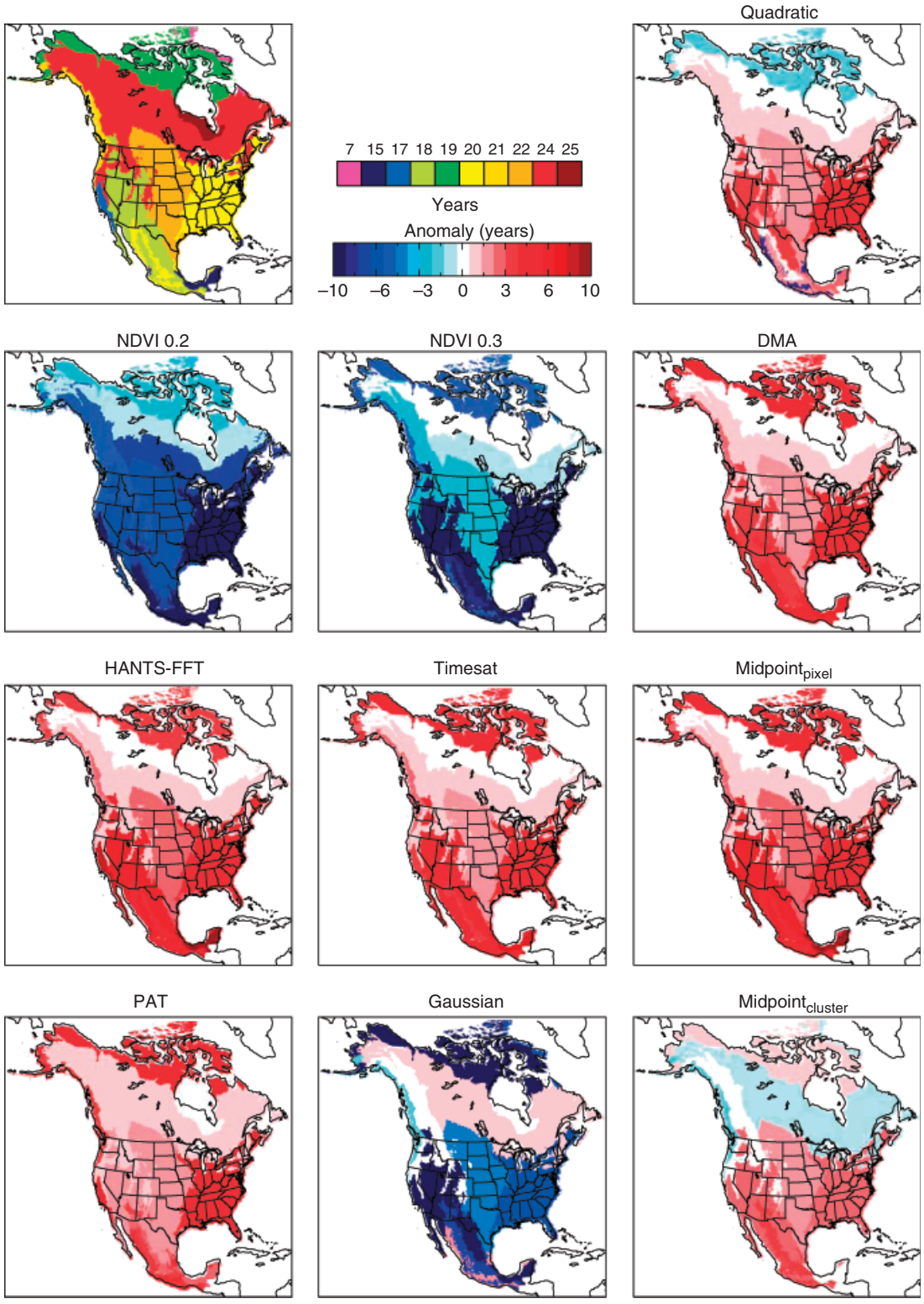

Fig. 4 Ensemble SOS retrieval rate (upper left panel, maximum of 25, ecoregions visible as color blocks). Low values indicate frequent failures to retrieve SOS estimates. Remaining panels show the retrieval rate anomaly between individual methods and the ensemble, thus indicating locations in which individual SOS methods were more or less able to retrieve SOS estimates.

© 2009 Blackwell Publishing Ltd,

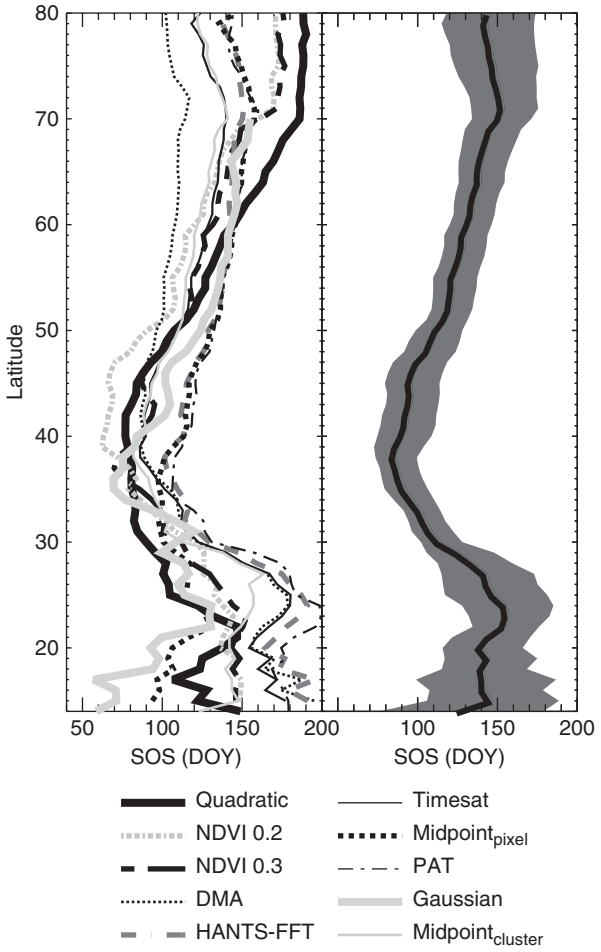

Fig. 5 Average start-of-spring (SOS) calculated by latitude (left panel) and shown as the ensemble mean and standard deviation (right panel).

in Tropical Wet Forests (15 years), Mediterranean California (17 years), and North American Deserts (18 years). Retrieval rates were highest in forested ecoregions with strong annual snow cycles. Among methods, DMA, HANTS-FFT, Timesat, Midpoint$_{pixel}$, PAT and Midpoint$_{cluster}$ were consistently better able to retrieve SOS estimates while all other methods had lower retrieval rates.

SOS methods varied in their ordinal ranking across latitude such that a method consistently early at high latitudes, such as DMA, could become a late method at low latitudes (Fig. 5). Of all the methods, only Timesat and Midpoint$_{cluster}$ tended to maintain approximately the same ordinal ranking. As an ensemble, the latitudinal average showed late SOS at both low and high latitudes and earliest SOS at about 40°. Variability was high above 70° and below 30° and extreme below 20°.

Assessment of annual time series and long-term behavior supports other findings of inconsistent SOS behavior among methods and ecoregion (Figs 6 and 7). The DMA was early and maximally stable in high latitude ecoregions but often late and dynamic in southern ecoregions (compare top and bottom rows of Figs 6 and 7). Among-method variability increased with both aridity (North American Deserts, Southern Semi-arid Highlands) and humidity (Tropical Wet Forests). Consistent with other results, within- and among-method variability was highest for Mediterranean California (Fig. 7).

SOS interpretation with cryospheric/hydrologic metrics

Correlations showed that cryospheric/hydrologic metrics were related to SOS retrievals but that the magnitude and location varied by ecoregion and SOS method (Table 3). Overall, comparisons of cryospheric dynamics were related to SOS while hydrologic dynamics were not. Of the five ecoregions with consistent annual snow cycles, only the Hudson Plain ecoregion had correlations with $P<0.05$ between the date of initial snowmelt and all SOS methods. Correlations were next highest in Northern Forests (mean of 0.49 across SOS methods). Among methods, PAT stood out as being minimally related to initial snow melt. SSM/I soil thaw comparisons were possible in nine ecoregions (Table 3). Here, correlations were high in two northern latitude forested ecoregions (Taiga and Hudson Plain) but also in Tundra. Outside of the high latitude ecoregions, P was >0.05 (two exceptions in Marine West Coast Forest) and were negative in eight out of 10 cases in the Great Plains. As for initial snowmelt and soil thaw, dates of lake ice breakup were related to SOS methods in the colder ecoregion ($P<0.05$ in Northern Forests, except HANTS-FFT) and less so in warmer ecoregions. In contrast to the cryospheric comparisons, the correlations of spring snowmelt onset date and the center of flow timing vs. SOS rarely had $P<0.05$ and were often negative. In the Great Plains, however, center of flow timing was significant and positive for six out of 10 SOS methods.

SOS assessment with plant phenology

We found that while no SOS method exhibited uniformly exceptional performance, the HANTS-FFT and Midpoint$_{pixel}$ methods were consistently more related to measured and modeled plant phenology than were other methods. Boxplots of the six phenological events showed two central patterns (Fig. 8). First, the median HANTS-FFT and Midpoint$_{pixel}$ were usually closest to the median of the spatially and temporally collocated

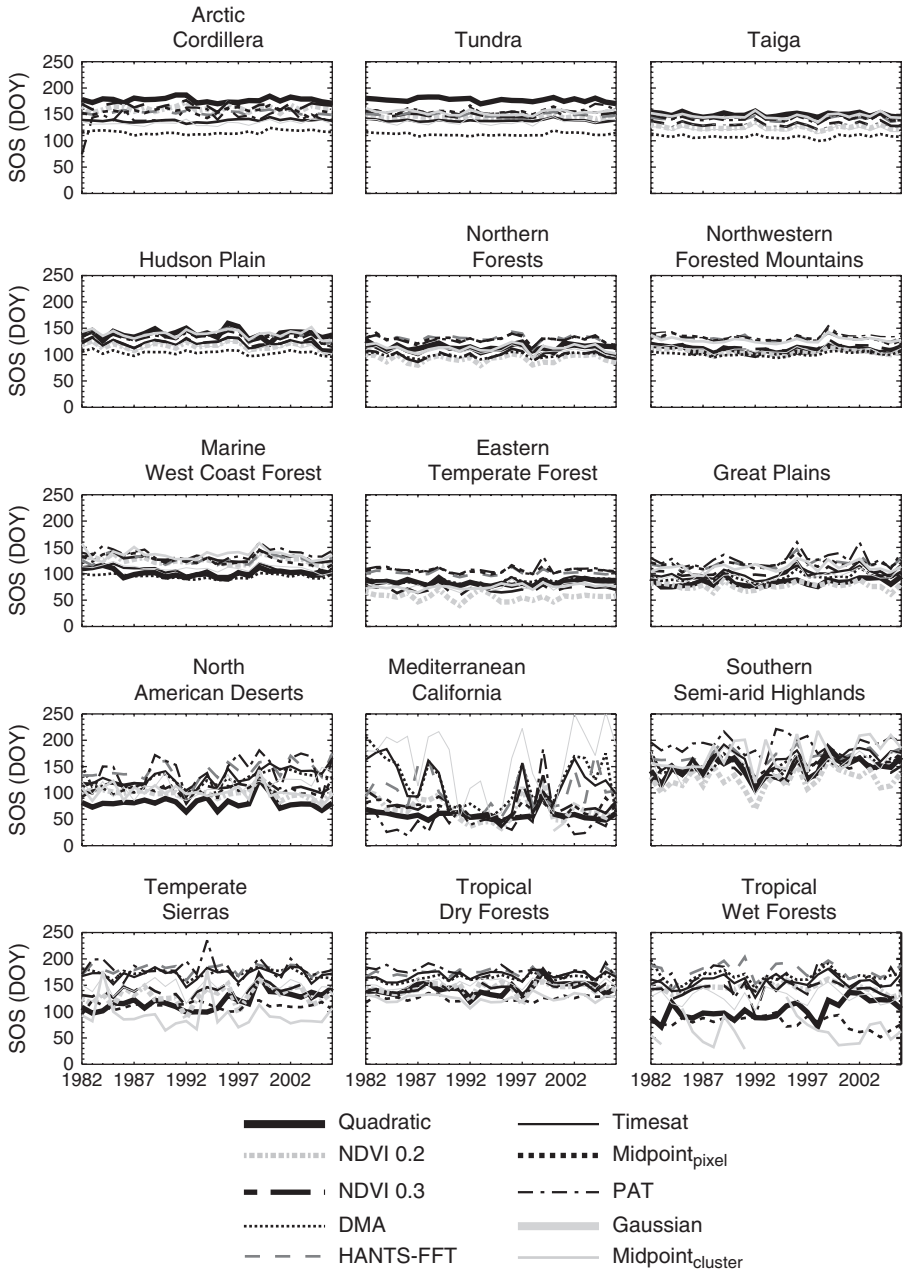

Fig. 6 Annual average start-of-spring (SOS) by ecoregion. Some methods and years are missing, e.g. for the Gaussian technique in Tropical Wet Forests.

measured phenology. Second, when compared across the phenological events, SOS methods were most similar in timing to first leaf such that interquartile ranges overlapped between measured phenology and seven of eight SOS methods (Gaussian and PAT were executed at the ecoregion level and are thus not comparable to ground locations).

When compared against specific measured phenology records, HANTS-FFT and Midpoint$_{pixel}$ again slightly outperformed other methods, but the strength of observed relationships was low (Figs 9 and 10). At the Bartlett Experimental Forest, SOS from the NDVI 0.2 and 0.3 methods was unrelated to the timing of increases in FPAR. In 2004, HANTS-FFT, Timesat, and Midpoint$_{pixel}$ SOS occurred within the measured increases in FPAR, but in 2005 and 2006 only HANTS-FFT and Midpoint$_{pixel}$ SOS were remotely within the FPAR increases. Results between HANTS-FFT and Mid-

© 2009 Blackwell Publishing Ltd,

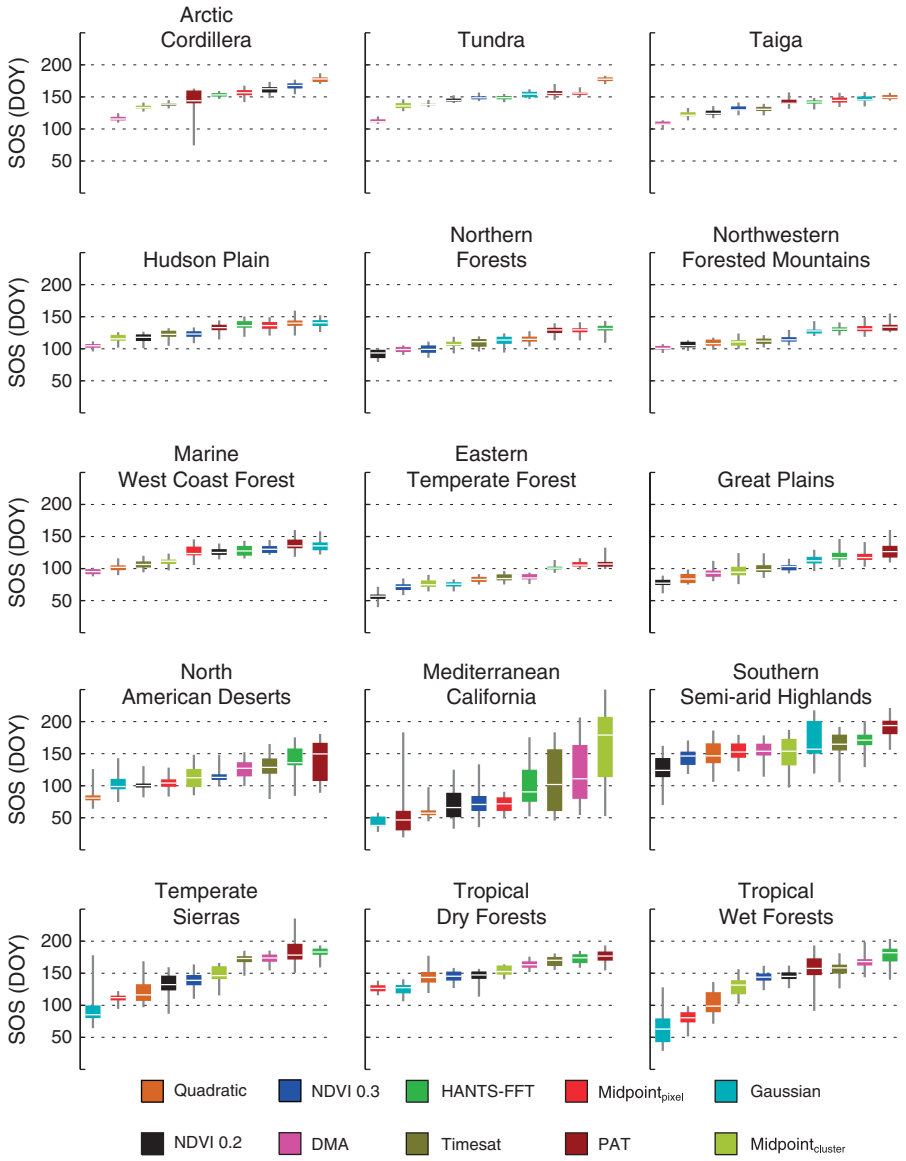

Fig. 7 Boxplots showing the distribution of start-of-spring (SOS) estimates across the 1982–2006 record for each ecoregion (box, interquartile range; white horizontal line, median; thin gray vertical lines, minimum and maximum). In each plots, boxplots are shown ordinally from earliest to latest method.

point$_{pixel}$ were inconsistent, with HANTS-FFT being earlier in 2004 and later in 2005 and 2006, while Midpoint$_{pixel}$ and Timesat had fairly consistent differences.

After screening data to include only time series with at least 10 years of data, we were able to assess 50 correlations between point-based measured phenology and pixel-based SOS (Fig. 10). P was rarely <0.05 (five out of 50 correlations for Quadratic and Midpoint$_{pixel}$, 10 out of 50 for HANTS-FFT, fewer for other methods). Averaged across the 50 time series, the highest mean correlations were for Midpoint$_{pixel}$ (0.35) and HANTS-FFT (0.33). Discounting NDVI 0.2 and NDVI 0.3, which were plagued by missing data, HANTS-FFT and Midpoint$_{pixel}$ also had the fewest number of negative correlations (two) among the SOS methods. The highest correlations for any comparisons were at the Rocky Mountain Biological Laboratory, where results had $P<0.05$ for seven of eight SOS methods.

HANTS-FFT and Midpoint$_{pixel}$ had the highest percentage of SOS estimates within the range and CIs of measured plant phenology for pixel-years with at least two unique plant observations (Fig. 11). For all SOS methods, estimates later than observed phenology CIs were rare (usually $<10\%$) and SOS estimates within the range of ground observed phenology never exceeded 40%. With the less stringent 90% CI comparison, Quad-

© 2009 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2009.01910.x

Table 3 Spearman's rank correlation coefficient for the relationship between the ten SOS methods shown as column headings and five independent cryospheric/hydrologic metrics shown from top to bottom as *shaded* headings and described in Appendix A

	Quadratic	NDVI 0.2	NDVI 0.3	DMA	HANTS-FFT	Timesat	$Midpoint_{pixel}$	PAT	Gaussian	$Midpoint_{cluster}$	Ecoregion average
Initial snowmelt											
Tundra	30	31	28	37	40	38	17	−17	45	12	26
					0.04				0.02		
Taiga	26	36	21	26	30	42	34	13	20	39	29
						0.04					
Hudson Plain	69	89	88	81	84	88	88	89	88	80	84
	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	
Northern Forests	63	64	59	60	28	57	34	39	41	46	49
	<0.01	<0.01	<0.01	<0.01		<0.01			0.04		
Northwestern Forested Mountains	46	44	41	18	12	35	29	11	24	23	28
	0.02	0.03	0.04								
Method average	47	53	47	44	39	52	40	27	44	40	
SSM/I soil thaw											
Arctic Cordillera	6	40	38	−38	47	−35	11	20	NA	13	11
					0.05						
Tundra	49	63	61	−2	54	48	61	65	50	57	51
	0.04	0.01	0.01		0.02	0.04	0.01	<0.01	0.03	0.01	
Taiga	67	66	53	64	57	68	55	58	53	66	61
	<0.01	<0.01	0.03	<0.01	0.01	<0.01	0.02	0.01	0.02	<0.01	
Hudson Plain	83	70	68	65	35	59	54	60	50	48	59
	<0.01	<0.01	<0.01	<0.01		0.01	0.02	0.01	0.04	0.04	
Northern Forests	56	44	48	27	20	48	34	35	30	40	38
	0.02		0.04			0.04					
Northwestern Forested Mountains	−4	32	31	33	30	27	27	27	33	25	26
Marine West Coast Forest	29	35	24	33	41	13	19	48	8	61	31
								0.04		0.01	
Eastern Temperate Forest	40	−3	20	46	14	39	28	34	16	46	28
Great Plains	−2	−30	−6	−6	−17	−12	−4	3	4	−22	−9
Method average	36	35	37	25	31	28	32	39	27	37	
Lake ice breakup											
Northern Forests	62	62	67	57	22	62	44	55	52	54	52
	<0.01	<0.01	<0.01	<0.01		<0.01	0.04	0.01	0.01	0.01	
Eastern Temperate Forest	16	53	40	13	24	24	20	32	0	15	24
		0.01									
Great Plains	5	26	17	11	−2	7	7	0	23	−10	8
Method average	28	47	41	27	15	31	24	29	25	20	
Spring snowmelt onset date											
Northwestern Forested Mountains	−16	−18	−11	23	−27	−14	−40	2	1	−23	−12

Continued

Table 3. (*Contd.*)

	Quadratic	NDVI 0.2	NDVI 0.3	DMA	HANTS-FFT	Timesat	$\text{Midpoint}_{\text{pixel}}$	PAT	Gaussian	$\text{Midpoint}_{\text{cluster}}$	Ecoregion average
							0.05				
Marine West Coast Forest	35	11	13	5	−25	7	−15	17	−21	−12	2
North American Deserts	24	37	36	38	39	34	30	32	29	28	33
Method average	14	10	12	22	−4	9	−8	17	3	−3	
Center of flow timing											
Northern Forests	7	1	−2	−3	−9	−21	−14	−5	−6	−17	−7
Northwestern Forested Mountains	50	39	40	30	11	39	32	41	38	28	35
	0.01		0.05					0.04			
Marine West Coast Forest	7	−4	−3	−10	−52	−13	−34	−18	−33	−49	−21
					0.01					0.01	
Eastern Temperate Forest	−13	25	12	−35	−33	−19	−30	−19	−22	−33	−17
Great Plains	41	35	49	47	42	52	39	49	28	22	40
	0.04		0.01	0.02	0.04	0.01		0.01			
North American Deserts	4	−36	−42	18	21	16	16	25	−10	9	2
			0.04								
Mediterranean California	12	6	11	20	−14	13	21	3	7	4	8
Temperate Sierras	−11	−4	−6	40	24	12	−1	−13	−14	−2	3
				0.05							
Method average	12	8	7	13	−1	10	4	8	−1	−5	

For each method, only those ecoregions containing the relevant information are shown (e.g. no snowmelt data available for most tropics). For visual clarity, coefficients are shown multiplied by 100 and rounded such that a value of 30 represents a value of 0.30. Averages for each ecoregion are shown in the last column and for each SOS method in the last row of each set of comparisons. *P*-values less than the standard cutoff for significance of 0.05 are shown on the second line of each cell (see section 2 for discussion).

© 2009 Blackwell Publishing Ltd, *Global Change Biology*, doi: 10.1111/j.1365-2486.2009.01910.x

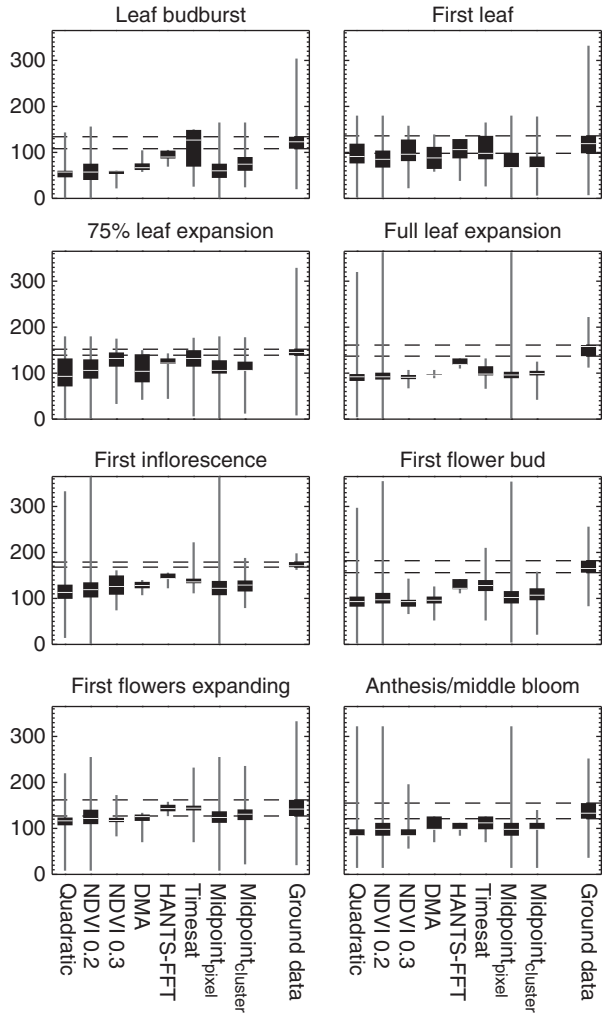

Fig. 8 The relationship between ground-measured phenological stage (panel headings) and SOS estimates. Boxplots: black box, interquartile range; white line, median; thin gray vertical lines, minimum and maximum. Dashed lines show the upper and lower quartile of the ground data. Note that the geographic and temporal coverage of the phenological stages is variable, i.e. the boxplots are not a comparison of SOS vs. ground data at the same location and times and should be taken as an approximate indication of the relative timing of SOS vs. ground phenology: SOS is almost always earlier, often by several weeks.

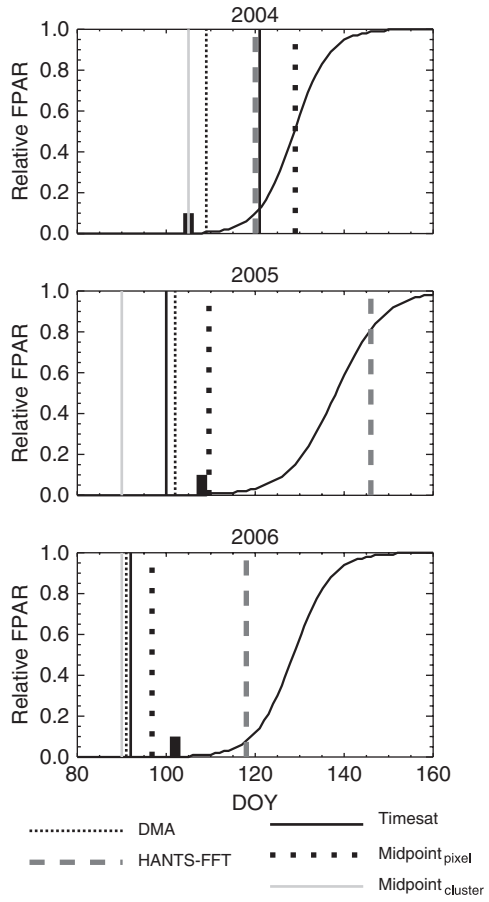

Fig. 9 SOS estimates for the Bartlett Experimental Forest. Black sigmoidal curve shows the site-measured FPAR on a relative scale (0 = annual minimum; 1 = annual maximum). Vertical lines show the SOS estimate from individual methods. SOS estimates for NDVI 0.2, and NDVI 0.3 methods were always earlier than DOY 80 and are not shown; Quadratic was earlier than DOY 80 except in 2006 when it overlapped with DMA and is thus not drawn; PAT and Gaussian methods are not shown as they were implemented at the ecoregion, not pixel, level. Short, thick black line shows date of soil thaw at 5 cm depth.

ratic, NDVI 0.2, NDVI 0.3, DMA, Timesat, and Midpoint$_{cluster}$ had nearly half or more of SOS estimates before the observed CI. Only HANTS-FFT and Midpoint$_{pixel}$ had more than 60% of SOS estimates within the observed CIs (66% and 69%).

In comparisons against the 1982–1999 annual dates of ground-measured spring arrival (all stages, Table 2), only HANTS-FFT and Midpoint$_{pixel}$ had the desired combination of high R^2, a low bias, and a reduced major axis regression slope close to 1 (Fig. 12). Timesat and Midpoint$_{cluster}$ had high R^2 and consistently large biases towards early SOS estimates. All methods besides HANTS-FFT and Midpoint$_{pixel}$ had offsets larger than 3 weeks and only Midpoint$_{pixel}$ had bias <1 week. NDVI 0.2 and NDVI 0.3 had R^2 close to zero. Expressed as time series plots, Midpoint$_{pixel}$ tracked the ground-measured phenology with low bias but some evidence of excessive interannual variability in the mid 1990s (Fig. 12). HANTS-FFT was more biased towards earlier predictions but had interannual variability more consistent with ground-measured phenology.

For the Eastern Temperate Forest comparisons of modeled plant phenology vs. satellite estimates, SOS was related ($P<0.05$) to SI first bloom for six of 10

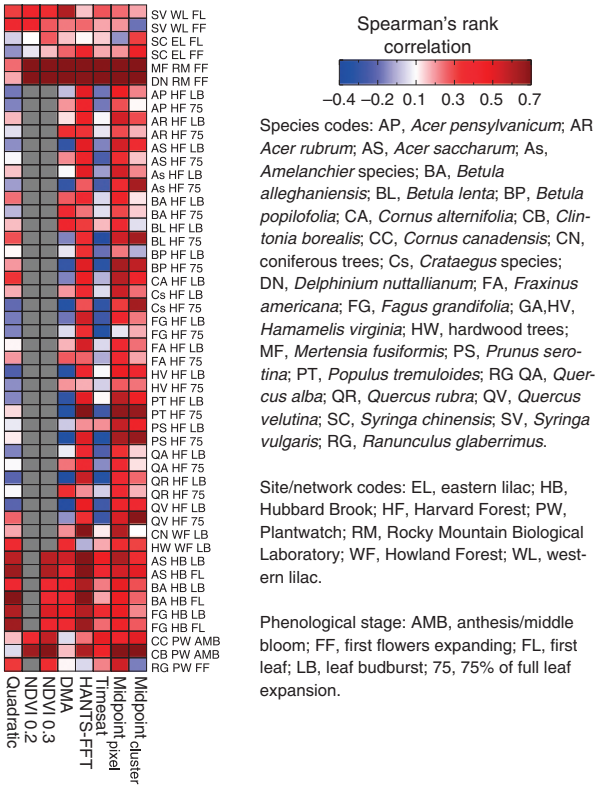

Fig. 10 Spearman's rank correlation coefficients between ground-measured phenology and SOS estimates for the corresponding years and location. Shown are records of at least 10 years in duration. Blue colors indicate negative correlations; red colors positive correlations; gray indicates missing SOS estimates. Each column shows one SOS method; each row shows one ground record. Text to the right of the colored panels shows species, site or network, and phenological stage (see in-figure text for code explanations). For EL, PW, and WL comparisons are based on annual averages at multiple locations; otherwise comparisons are for individual sites. PAT and Gaussian methods not shown as they were implemented at the ecoregion, not pixel, level.

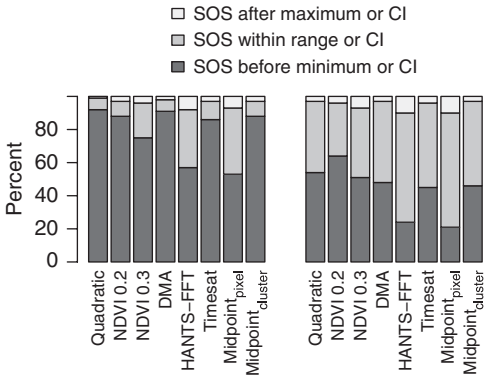

Fig. 11 Comparison of SOS methods against ground data. For each 8 km pixel containing at least two ground observations within a year (any species or phenological event), the range and 90% confidence interval (CI) were constructed for the ground observations; SOS estimates were then categorized as before the minimum or CI, within the range or CI, or after the maximum or CI. Comparisons are range in the left panel and CI in the right panel. All ground data for which valid satellite retrievals existed (ranging from a low of 1848 pixel-year records for NDVI 0.2 to a high of 3046 for HANTS-FFT) were used – a separate assessment was conducted using only pixel-year records found in all eight methods but results were consistent and are not shown. PAT and Gaussian methods not shown as they were implemented at the ecoregion, not pixel, level.

methods with an average value of 0.41 (Table 4). SI first bloom correlations were highest for HANTS-FFT, the three conceptually linked methods ($Midpoint_{pixel}$, $Midpoint_{cluster}$, and PAT), and NDVI 0.3. SI first leaf was related ($P<0.05$) only to NDVI 0.2 while $D_{leaf\text{-}out}$ was related (at $P<0.05$) to NDVI 0.2 and NDVI 0.3. We graphically present results for the HANTS-FFT and $Midpoint_{pixel}$ methods selected as most consistent with these and the preceding ground phenology data (Fig. 13). As for the comparison with measured plant phenology, $Midpoint_{pixel}$ was consistent with the absolute dates of both the SI first bloom and $D_{leaf\text{-}out}$ models while HANTS-FFT was slightly biased towards early estimates. Modeled SI first leaf was approximately 1 month earlier.

Trends

Trends in spring arrival were insignificant for both: (1) measured ground phenology, HANTS-FFT, and $Midpoint_{pixel}$ for North America from 1982 to 1999 (Fig. 12); and (2) SI first leaf and first bloom, $D_{leaf\text{-}out}$, HANTS-FFT, and $Midpoint_{pixel}$ for the US Eastern Temperate Forest from 1982 to 2003 (Fig. 13). Assessed at a more detailed level for the 182 level 3 ecoregions (Supporting Information, Figure S1), trends existed ($P<0.05$) for 30 of 182 ecoregions in HANTS-FFT (20 towards earlier SOS and 10 towards later SOS) and for 24 of 182 ecoregions in $Midpoint_{pixel}$ (15 towards earlier SOS and nine towards later SOS). Only five ecoregions had $P<0.05$ for trends in both methods (two for earlier SOS and three for later SOS). When using HANTS-FFT and $Midpoint_{pixel}$ as simultaneous y-vectors, however, 30 ecoregions had $P<0.05$ for trends (Fig. 14). A total of 12% of land area had a trend with $P<0.05$; earlier trends (7%) slightly exceeding later trends (5%).

Discussion

Our results indicate that given NDVI data with identical duration, satellite correction scheme, geographic region, compositing scheme, and spatial resolution, SOS estimates differed in terms of average DOY by more than 1

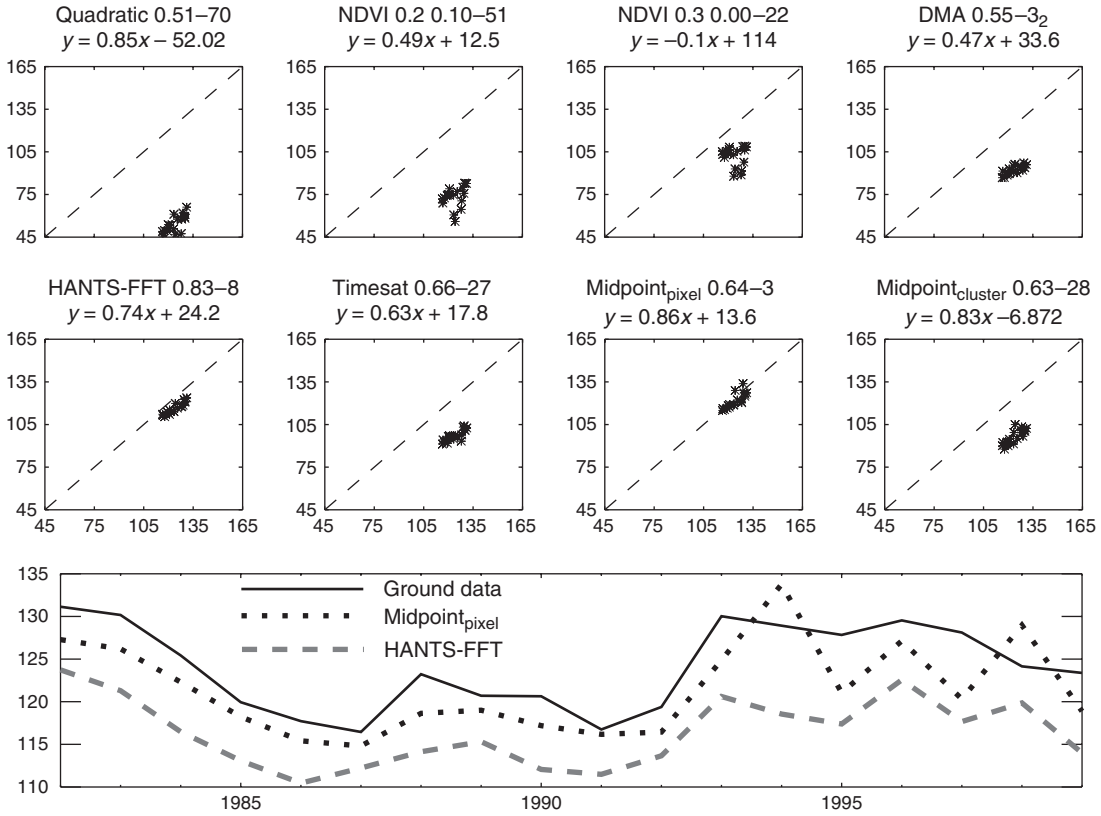

Fig. 12 Comparison of 1982–1999 ground observed phenology and SOS (post-1999 data not used due to very limited ground data availability). Analysis is for all pixel-years containing at least two ground observations of any species and any phenological stage and only for pixel-years present for all SOS methods. Ground data were first averaged to pixel-year and then for all of North America. Small panel titles are: SOS method, R^2, and bias (in days) on the first line and the reduced major axis linear model on the second line. The bottom panel shows the ground data and the two SOS methods with bias close to zero, slope near one, and high R^2. All Spearman's Rank Correlations showed no trends in SOS or ground phenology ($P > 0.05$). Analysis conducted only for those pixels present for all SOS methods.

month (Fig. 2), variability by more than 2 weeks (Fig. 3), retrieval ability by more than one-third (Fig. 4), and ordinal ranking by latitude (Fig. 5) and ecoregion (Figs 6 and 7). Other work has noted that differences exist among SOS methods (Reed *et al*., 2003) and ecoregions (Bradley & Mustard, 2008), but an expectation, or perhaps a hope, has existed in the remote sensing community that SOS methods may have consistent ordinal behavior and may simply be detecting different portions of the annual vegetation phenological developmental cycle. We do not find evidence to support this supposition. Independent of interpretive and assessment data, such an intercomparison of SOS methods would have no rational basis for selecting one method over another method.

When taken in the context of interpretation with cryospheric/hydrologic metrics and assessment with plant phenology observations and models, we believe that our intercomparison may be a useful way of identifying the strengths and weaknesses of particular methods and SOS approaches in general. First, we suggest that methods based on global thresholds (including the hybrid Gaussian method) be abandoned for continental to global applications. The NDVI 0.2 and NDVI 0.3 methods – based on NDVI exceeding a geographically constant threshold – had excessively low retrieval rates (Figs 4 and 10, usually because annual NDVI did not fall below the threshold) and essentially no relationship with measured patterns of interannual phenology variation (Fig. 12). The Gaussian method, which relies on a range of absolute thresholds, had similarly large failure rates for much of North America (Fig. 4). Limited cases of high correlations suggest, i.e. for cryospheric comparisons (Table 3) or $D_{\text{leaf-out}}$ (Table 4), that absolute thresholds may be appropriate for geographically limited application in specific ecosystem; such ability, however, is more than offset by inapplicability over much of North America for the global threshold or hybrid methods.

Second, differences in the implementation of a related method may produce quite different results, as in the case of Midpoint$_{pixel}$ and Midpoint$_{cluster}$. We executed

Table 4 Spearman's rank correlation coefficients for the 1982 through 2003 relationship between the ten SOS methods shown as column headings and modeled phenology

	Quadratic	NDVI 0.2	NDVI 0.3	DMA	HANTS-FFT	Timesat	$Midpoint_{pixel}$	PAT	Gaussian	$Midpoint_{cluster}$	Phenological stage
First leaf	35	46	41	24	24	29	29	38	7	34	21
		0.03									
First bloom	8	37	51	38	55	43	53	56	13	54	41
			0.02		0.01	0.05	0.01	0.01		0.01	
$D_{leaf-out}$	5	44	22	11	44	11	22	22	14	18	31
		0.04			0.04						
Method average	16	42	38	24	41	27	35	39	11	35	

First leaf and first bloom simulations are from the Spring Indices models based on clonal lilac and honeysuckle species; $D_{leaf-out}$ is an estimation of the day at which the ecosystem becomes a net carbon sink as measured by eddy covariance and is related to the timing of leaf-out by overstory trees. Results are shown only for the eastern temperature forest ecoregion in which the models were developed. For visual clarity, coefficients are shown multiplied by 100 and rounded such that a value of 30 represents a value of 0.30. Averages for each phenological stage are shown in the last column and for each SOS method in the last row. *P*-values less than the standard cutoff for significance of 0.05 are shown on the second line of each cell.

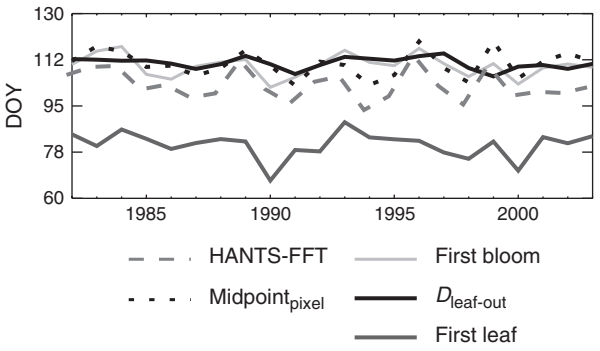

Fig. 13 1982–2003 comparison of SOS estimates and modeled spring phenology. Models (shown by solid lines) are first bloom and first leaf from the Spring Indices models based on clonal lilac and honeysuckly phenology and $D_{leaf-out}$ based on the date at which ecosystems become net carbon sinks, as measured by eddy covariance. Data are shown for the Eastern Temperate Forest ecoregion within the conterminous United States (area covered by meteorological inputs required for models). See Table 3 for statistics for all SOS methods. Note that curves for HANTS-FFT and $Midpoint_{pixel}$ represent a different study area and duration than the curves shown in Fig. 12.

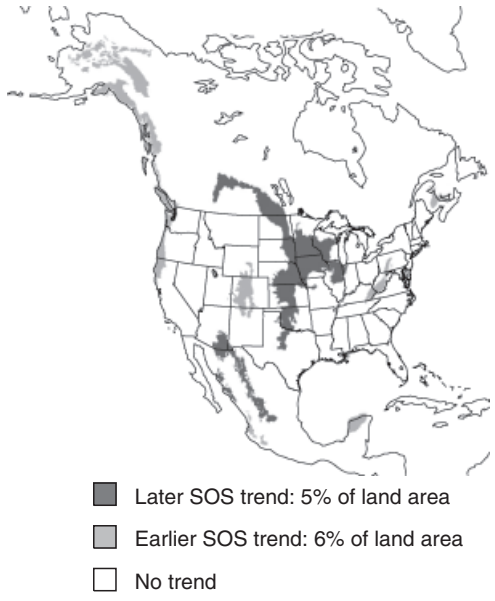

Fig. 14 Location of trends in 1982 to 2006 SOS with $P<0.05$ calculated using the HANTS-FFT and $Midpoint_{pixel}$ methods.

$Midpoint_{pixel}$ for individual pixels and used a spline to fit sub-daily time steps and a detailed removal of undesirable time series; for $Midpoint_{cluster}$, we used a regionalization concept, measures of uncertainty around the threshold, and a 15-day time step (see Appendix A). When retrieved SOS is regressed on ground observations, both methods had similar R^2 and slope but $Midpoint_{cluster}$ had a bias of about 1

month, thus highlighting the importance of implementation details on SOS results (Fig. 12).

Third, SOS methods were frequently incapable of retrieving estimates for desert and tropical ecoregions or ecoregions in which the initiation of growth spans the start of the calendar year (Fig. 4). In these areas, when retrievals were obtained, the variance among methods was high (Figs 6 and 7). In addition, few independent data on cryospheric/hydrologic metrics were available for desert or tropical system. SOS for Mediterranean California, an ecoregion with a pronounced and regular wet and dry season, was inconsistently estimated (Figs 2, 3, 6 and 7). The switch from dry to wet occurs around the end of December to early January, and the SOS methods' variable treatment of calendar years vs. a continual time series likely influenced these results. Whenever possible, we recommend the extraction of SOS estimates from continual time series.

Fourth, we have established that for level 1 ecoregions, SOS estimates are related to cryospheric dynamics, especially in the colder and snowier ecoregions, but less so to hydrologic dynamics (Table 3). Our results support the contention that for evergreen forested ecoregions, the annual cycle from near-total snow cover to a mature canopy provides a distinct and detectable NDVI cycle, arguing for further development of techniques designed to extract a pure vegetation phenology cycle (Delbart *et al.*, 2005).

The spring snowmelt onset date is designed to be a proxy indicator of when temperatures rise above freezing and stay there. It is likely that for Northwestern Forested Mountains and Marine West Coast Forests, the snowmelt metric is too early to track spring phenological development and thus high correlation would be unexpected. In ecoregions with spatially variable snowmelt inputs and/or where a snowmelt pulse does not persistently dominate streamflow, short-term precipitation variability and timing becomes more influential. Thus, in the North American Deserts, where all correlations were positive but had $P > 0.05$, we speculate that phenology is likely to be related to snowmelt timing, unimodal and bimodal precipitation distribution, and moisture availability, and that the low correlations may be related to persistently low SOS retrieval rates (Fig. 4).

For the center of flow timing, the low correlations in forested systems may again reflect the wrong event for comparisons to spring phenology or a strong signal from watersheds with the most snow (usually highest elevation) rather than the most area. In the southern Great Plains, where water limits can be important and many streamflow records are not dominated by a unimodal snowmelt pulse, the timing of water delivery may influence interannual phenological timing, leading to correlations with $P < 0.05$ for some SOS methods (Table 3). It is possible that different hydrologic measures tuned towards different stages of the hydrograph could be more related to SOS; we recommend that further research explore this possibility.

Fifth, based on a suite of information (Figs 8, 9, 11 and 12), we have established that in most cases, SOS estimates occur before measured phenological events. Even in the case where SOS estimates most overlapped with ground observations (first leaf, Fig. 8), satellite dates were usually earlier than ground dates. While the consistent SOS vs. ground measurement bias may be caused by SOS detections being more related to snow dynamics (Table 3; see Fig. 10 for high correlations at Rocky Mountain Biological Laboratory, a site with large NDVI amplitude from snow cover to mature conifer forest) and/or ground networks being biased towards species with relatively late phenologies, our results suggest that observations of first leaf may be most useful for future assessment of SOS methods.

Sixth, although we have attempted to minimize the difficulties inherent in an extremely diverse ground phenology dataset (Fig. 1), we acknowledge that the ground data were not collected with an explicit purpose of satellite assessment; our analysis is thus subject to classical point-vs.-pixel comparison errors. In essence, without remote sensing capable of resolving individual crowns or more detailed sampling schemes, it is not known whether or not the recorded species reflect the overall phenological development of an entire 8 km pixel. While these caveats represent a potentially important source of variation generating unknown uncertainty, our overall results suggest that, in comparison with ground data, the HANTS-FFT and $\text{Midpoint}_{\text{pixel}}$ methods have: about 65% acceptable SOS retrievals (Fig. 11), correlations that are > 0.6, low offsets or bias, and regression slope near 1. For implementation purposes, we note that some methods require complete time series and are best suited for research purposes (i.e. HANTS-FFT which requires data extending well past extracted SOS dates) while others, such as PAT (which is strongly related to $\text{Midpoint}_{\text{pixel}}$ above about 35° and is simple to implement, Fig. 5), are optimized for real-time implementation

Finally, evidence from measured (Fig. 12) and modeled (Fig. 13) phenology supports our findings of very limited SOS trends towards earlier spring arrival (Fig. 14), which are broadly consistent with some satellite results (Reed *et al.*, 2003) and opposite others (Zhang *et al.*, 2007). In our two-way comparison of independent trends estimated between the HANTS-FFT and $\text{Midpoint}_{\text{pixel}}$ methods, we found numerous ecoregions with trends towards both earlier and later SOS ($P < 0.05$), but the locations differed and only five of

182 level 3 ecoregions had $P<0.05$ in both methods. However, for all cases except one, when one method had a $P<0.05$, the sign from the other method was the same, suggesting a measure of consistency that is supported by the 30 ecoregions with $P<0.05$ when using the ensemble approach (Fig. 14). We therefore submit that an ensemble approach of multiple SOS methods may be more powerful for trend estimation than use of single methods alone.

Ground-based observations of cryospheric/hydrologic metrics and plant phenology over longer time periods have tended to show trends consistent with climate warming. For example, trends toward earlier peak snowmelt runoff have been found in the western US during 1948–2002 (Stewart *et al*., 2005) and earlier ice breakup on lakes and rivers across the northern hemisphere has been observed during 1846–1995 (Magnuson *et al*., 2000). Trends towards earlier spring have been found during 1951–2000 for agrometeorological indices in the western US (Feng & Hu, 2004); 1954–1994 for lilac and honeysuckle phenology in the western US (Cayan *et al*., 2001); and 1959–1993 for last −2.2 °C frost date and for SI-modeled first leaf and first bloom (Schwartz & Reiter, 2000). However, results from experimental warming suggest that plants which develop later in the summer may be less likely to respond to climate change by advancing their phenology or may even show trends toward later phenology (Sherry *et al*., 2007).

Satellite SOS trend estimates are limited by a short record (Myneni *et al*., 1997) and are thus often incomparable with longer, climatically driven analyses. We note, however, that measured and remotely sensed estimates for North America both suggest a trend towards earlier spring until the early 1990s followed by a step change to later spring around 1993 – a change that is largely consistent with approximately 0.5 °C decreases in post-1993 December to May temperatures for most of North America except the desert southwest (Figure S2). Other studies have shown trend reversals in measured phenology consistent with seasonal temperature variations or changes in synoptic pressure systems (Scheifinger *et al*., 2002; Schaber & Badeck, 2005).

Conclusions

We conducted an intercomparison, interpretation, and assessment of 10 SOS methods for North America from 1982 to 2006. We demonstrate that SOS estimates vary extensively within and among methods and that independent of other ecosystem information, selecting the strongest method or calculating ensemble methods would be difficult. Based on relationships with independent measures of cryospheric interannual variability and measured and modeled plant phenology, we identify two SOS methods most consistent with currently available corroborating data.

Trend estimates from the SOS methods as well as measured and modeled plant phenology strongly suggest either no or very geographically limited trends towards earlier spring arrival, although we caution that, for an event such as SOS with high interannual variability, a 25-year SOS record is short for detecting robust trends. Increased greenhouse warming since the late 20th century would seem to argue for increased, not decreased, shifts in spring during our study period, indicating that processes such as succession, changes in community structure, land management, or disturbance may be more important than previously recognized. Seasonal temperature changes may also be linked to a trend reversal in SOS in the early 1990s.

Our results highlight both the challenge and potential for integrating remote sensing and ground observations. No other technology besides remote sensing offers wall-to-wall coverage and consistent long-term monitoring, yet few metrics of biospheric response are as unconstrained by appropriate ground data – our study clearly outlines the limitations in using existing historical datasets. Establishing consistent plant phenology monitoring networks (e.g. the USA National Phenology Network, http://www.usanpn.org, (Betancourt *et al*., 2007), or the European Phenology Network) as well as incorporating a broader consideration of non-climatic factors influencing SOS estimates is therefore critical. A specific suggestion is to integrate SOS estimates with ground measurements of first leaf (to which SOS estimates from the two selected methods are most related) in a geographically focused area with broad correspondence among cryospheric/hydrologic metrics and phenology, such as the Hudson Plain ecoregion. A focused approach would also permit assessment of within-ecoregion variability, which was beyond the scope of the current analysis. Similar analyses and study selections could be replicated on other continents to produce a network of phenological monitoring ecoregions.

Acknowledgements

We gratefully acknowledge agency support: MAW, NASA grants NNG04G043G and NNA05CS25A and NSF grant 02-4277; DDB, the Office of Science (BER), US DOE grant DE-FG02-06ER64308 and NSF grant DEB 0639235; MDS, NSF grants ATM-9510342, 9809460, and 0085224; ADR, the Northeastern States Research Cooperative and the U.S. Department of Energy's Office of Science (BER) through the Northeastern Regional Center of the National Institute for Climatic Change Research. We thank Samuel Hiatt for technical assistance. Logistical support and/or data were provided by the Niwot Ridge Long-Term Ecological

Research (LTER) project and the Mountain Research Station (BIR 9115097). Data was supported by the NSF LTER Program at Konza Prairie Biological Station. Data sets were provided by the Arctic LTER. This material is based upon work supported by the National Science Foundation under Grants DEB 981022, 9211775, 8702328; OPP-9911278, 9911681, 9732281, 9615411, 9615563, 9615942, 9615949, 9400722, 9415411, 9318529; BSR 9019055, 8806635, 8507493. Data sets were provided by the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the U.S. Forest Service Pacific Northwest Research Station, Corvallis, Oregon. Significant funding for collection of these data was provided by the National Science Foundation LTER program (NSF Grant numbers BSR 9011663 8811906 0423662; DEB 9632921, 0217631, 9411976, 0080529 and 0217774). Data sets were provided by the Shortgrass Steppe LTER group, a partnership between Colorado State University, United States Department of Agriculture, Agricultural Research Service, and the U.S. Forest Service Pawnee National Grassland. Data sets were provided by the Shortgrass Steppe LTER group, a partnership between Colorado State University, United States Department of Agriculture, Agricultural Research Service, and the U.S. Forest Service Pawnee National Grassland. Data sets were provided by the Sevilleta LTER program.

References

Angert A, Biraud S, Bonfils C *et al*. (2005) Drier summers cancel out the CO_2 uptake enhancement induced by warmer springs. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10823–10827.

Armstrong RL, Brodzik MJ (2005) *Northern Hemisphere EASE-Grid weekly snow cover and sea ice extent version 3. (Digital media)*. National Snow and Ice Data Center, Boulder, Co, USA.

Asrar G, Fuchs M, Kanemasu ET, Hatfield JL (1984) Estimating absorbed photosynthetically active radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal*, **76**, 300–306.

Assel RA, Robertson DM (1995) Changes in winter air temperatures near Lake-Michigan, 1851–1993, as determined from regional Lake-Ice records. *Limnology and Oceanography*, **40**, 165–176.

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

Baldocchi D, Falge E, Gu LH *et al*. (2001) FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, **82**, 2415–2434.

Baldocchi DD, Black TA, Curtis PS *et al*. (2005) Predicting the onset of net carbon uptake by deciduous forests with soil temperature and climate data: a synthesis of FLUXNET data. *International Journal of Biometeorology*, **49**, 377–387.

Betancourt J, Schwartz M, Breshears D *et al*. (2007) Evolving plans for the USA National Phenology Network. *EOS*, **88**, 211.

Bradley BA, Mustard JF (2008) Comparison of phenology trends by land cover class: a case study in the Great Basin, USA. *Global Change Biology*, **14**, 334–346.

Bunn AG, Goetz SJ (2006) Trends in satellite-observed circumpolar photosynthetic activity from 1982 to 2003: the influence of seasonality, cover type, and vegetation density. *Earth Interactions*, **10**, doi: 10.1175/EI190.1.

Cayan D, Kammerdiener S, Dettinger M, Caprio J, Peterson D (2001) Changes in the onset of spring in the western United States. *Bulletin of the American Meteorological Society*, **82**, 399–415.

Chmielewski FM, Rotzer T (2001) Response of tree phenology to climate change across Europe. *Agricultural and Forest Meteorology*, **108**, 101–112.

Cleland EE, Chuine I, Menzel A, Mooney HA, Schwartz MD (2007) Shifting plant phenology in response to global change. *Trends in Ecology & Evolution*, **22**, 357–365.

de Beurs KM, Henebry GM (2008) Northern annular mode effects on the land surface phenologies of Northern Eurasia. *Journal of Climate*, **21**, 4257–4279.

Delbart N, Kergoat L, Le Toan T, Lhermitte J, Picard G (2005) Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sensing of Environment*, **97**, 26–38.

Feng S, Hu Q (2004) Changes in agro-meteorological indicators in the contiguous United States: 1951–2000. *Theoretical and Applied Climatology*, **78**, 247–264.

Fitzjarrald DR, Acevedo OC, Moore KE (2001) Climatic consequences of leaf presence in the eastern United States. *Journal of Climate*, **14**, 598–614.

Goetz SJ, Bunn AG, Fiske GJ, Houghton RA (2005) Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 13521–13525.

Jakubauskas ME, Legates DR, Kastens JH (2001) Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, **67**, 461–470.

Jenkins JP, Richardson AD, Braswell BH, Ollinger SV, Hollinger DY, Smith ML (2007) Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements. *Agricultural and Forest Meteorology*, **143**, 64–79.

Jensen OP, Benson BJ, Magnuson JJ, Card VM, Futter MN, Soranno PA, Stewart KM (2007) Spatial analysis of ice phenology trends across the Laurentian Great Lakes region during a recent warming period. *Limnology and Oceanography*, **52**, 2013–2026.

Jönsson P, Eklundh L (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 1824–1832.

Jönsson P, Eklundh L (2004) TIMESAT – a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, **30**, 833–845.

Kimball JS, McDonald KC, Zhao M (2006) Terrestrial vegetation productivity in the western arctic observed from satellite microwave and optical remote sensing. *Earth Interactions*, **10**, doi: 10.1175/EI187.1.

Kogan FN, Zhu X (2001) Evolution of long-term errors in NDVI time series: 1985–1999. In: *Calibration and Characterization of Satellite Sensors and Accuracy of Derived Physical Parameters*, Vol. 28, pp. 149–153.

© 2009 Blackwell Publishing Ltd,

Landwehr JM, Slack JR (1992) *Hydro-Climatic Data Network (HCDN): Streamflow Data Set, 1874–1988*. Report 93-4076, US Geological Survey, Reston, VA, USA, Digital media.

Leemans R, Cramer W (1991) *The IIASA database for mean monthly values of temperature, precipitation and cloudiness on a global terrestrial grid*. RR-91-18, International Institute for Applied Systems Analysis, Laxenburg, Austria, 61 pp.

Linderholm HW (2006) Growing season changes in the last century. *Agricultural and Forest Meteorology*, **137**, 1–14.

Linkosalo T, Häkkinen R, Hari P (1996) Improving the reliability of a combined phenological time series by analyzing observation quality. *Tree Physiology*, **16**, 661–664.

Loveland TR, Reed BC, Brown JF, Ohlen DO, Zhu J, Yang L, Merchant JW (2000) Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. *International Journal of Remote Sensing*, **21**, 1303–1330.

Magnuson JJ, Robertson DM, Benson BJ *et al*. (2000) Historical trends in lake and river ice cover in the Northern Hemisphere. *Science*, **289**, 1743–1746.

Menzel A, Sparks TH, Estrella N *et al*. (2006) European phenological response to climate change matches the warming pattern. *Global Change Biology*, **12**, 1969–1976.

Mesinger F, DiMego G, Kalnay E *et al*. (2006) North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360.

Myneni RB, Keeling CD, Tucker CJ, Asrar G, Nemani RR (1997) Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, **386**, 698–702.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, **37**, 637–669.

Pedelty JA, Devadiga S, Masuoka E *et al*. (2007) *Generating a Long-term Land Data Record from the AVHRR and MODIS Instruments*. Proceedings of IGARRS 2007, pp. 1021–1025. Institute of Electrical and Electronics Engineers, NY, USA.

Piao SL, Friedlingstein P, Ciais P, Viovy N, Demarty J (2007) Growing season extension and its impact on terrestrial carbon cycle in the Northern Hemisphere over the past 2 decades. *Global Biogeochemical Cycles*, **21**, doi: 10.1029/2006GB002888.

Pinzon J, Brown ME, Tucker CJ (2005) EMD correction of orbital drift artifacts in satellite data stream. In: *The Hilbert-Huang Transform and its Applications* (eds Huang N, Shen S), pp. 167–183. World Scientific Publishing Co., Hackensack, NJ.

Reed BC (2006) Trend analysis of time-series phenology of North America derived from satellite data. *GIScience & Remote Sensing*, **43**, 1–15.

Reed BC, Brown JF, VanderZee D, Loveland TR, Merchant JW, Ohlen DO (1994) Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, **5**, 703–714.

Reed BR, White MA, Brown JF (2003) Remote sensing phenology. In *Phenology: An Integrative Environmental Science* (ed. Schwartz MD), pp. 365–381. Kluwer Academic Publishers, New York, NY.

Roerink GJ, Menenti M, Verhoef W (2000) Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, **21**, 1911–1917.

Schaber J, Badeck F (2005) Plant phenology in Germany over the 20th century. *Regional Environmental Change*, **5**, 37–46.

Scheifinger H, Menzel A, Koch E, Peter C, Ahas R (2002) Atmospheric mechanisms governing the spatial and temporal variability of phenological phases in central Europe. *International Journal of Climatology*, **22**, 1739–1755.

Schwartz MD (1997) Spring Index models: an approach to connecting satellite and surface phenology. In: *Phenology of Seasonal Climates* (eds Leith H, Schwartz MD), Backhuys, the Netherlands.

Schwartz MD (2003) Phenoclimatic Measures. In: *Phenology: An Integrative Environmental Science* (ed. Schwartz M), pp. 331–343. Kluwer Academic Publishers, New York, NY.

Schwartz MD, Ahas R, Aasa A (2006) Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biology*, **12**, 343–351.

Schwartz MD, Crawford TM (2001) Detecting energy-balance modifications at the onset of spring. *Physical Geography*, **21**, 394–409.

Schwartz MD, Reed B, White MA (2002) Assessing satellite-derived start-of-season (SOS) measures in the conterminous USA. *International Journal of Climatology*, **22**, 1793–1805.

Schwartz MD, Reiter BE (2000) Changes in North American spring. *International Journal of Climatology*, **20**, 929–932.

Sherry RA, Zhou XH, Gu SL *et al*. (2007) Divergence of reproductive phenology under climate warming. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 198–202.

Smith PM, Kalluri SNV, Prince SD, DeFries R (1997) The NOAA/NASA Pathfinder AVHRR 8-km land data set. *Photogrammetric Engineering and Remote Sensing*, **63**, 12–31.

Stewart IT, Cayan DR, Dettinger MD (2004) Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. *Climatic Change*, **62**, 217–232.

Stewart IT, Cayan DR, Dettinger MD (2005) Changes toward earlier streamflow timing across western North America. *Journal of Climate*, **18**, 1136–1155.

Thornton PE, Running SW, White MA (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214–251.

van Leeuwen WJD (2008) Monitoring the effects of forest restoration treatments on post-fire vegetation recovery with MODIS multitemporal data. *Sensors*, **8**, 2017–2042.

Wagenseil H, Samimi C (2006) Assessing spatio-temporal variations in plant phenology using Fourier analysis on NDVI time series: results from a dry savannah environment in Namibia. *International Journal of Remote Sensing*, **27**, 3455–3471.

Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW (2006) Warming and earlier spring increase western US forest wildfire activity. *Science*, **313**, 940–943.

White MA, Hoffman F, Hargrove WW, Nemani RR (2005) A global framework for monitoring phenological responses to climate change. *Geophysical Research Letters*, **32**, doi: 10.1029/102004GL021961.

White MA, Nemani RR (2003) Canopy duration has little influence on annual carbon storage in the deciduous broad leaf forest. *Global Change Biology*, **9**, 967–972.

White MA, Nemani RR (2006) Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment*, **104**, 43–49.

White MA, Schwartz MD, Running SW (2000) Young students and satellites team up to study climate-biosphere link. *Earth in Space*, **12**, 4–7.

White MA, Thornton PE, Running SW (1997) A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, **11**, 217–234.

Wilson KB, Baldocchi DD (2007) Comparing independent estimates of carbon dioxide exchange over 5 years at a deciduous forest in the southeastern United States. *Journal of Geophysical Research*, **106**, 34167–34178.

Xiao J, Moody A (2005) Geographical distribution of global greening trends and their climatic correlates: 1982–1998. *International Journal of Remote Sensing*, **26**, 2371–2390.

Zhang XY, Tarpley D, Sullivan JT (2007) Diverse responses of vegetation phenology to a warming climate. *Geophysical Research Letters*, **34**, doi: 10.1029/2007GL031447.

Zhou L, Tucker C, Kaufmann R, Slayback D, Shabanov N, Myneni R (2001) Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *Journal of Geophysical Research-Atmospheres*, **106**, 20069–20083.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Map of the level 3 ecoregions.

Fig. S2. December to May average air temperatures for the 1994 to 2006 minus the 1988 to 1992 period. Purple colors show cooling in the later period. Data and figure obtained through the NCEP/NCAR Reanalysis project (http://www.cdc.noaa.gov/ncep_reanalysis/).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Appendix A

Satellite SOS estimates

SOS category 1: global thresholds

In this simplest of methods, SOS is determined as the DOY that NDVI crosses a threshold in the upward direction where the same threshold is used globally, i.e. for every pixel. To determine at which DOY the threshold is reached, the time series is interpolated to a daily dataset. In this study we have used threshold levels of 0.2 and 0.3 with no filtering or smoothing of input NDVI time series and term the methods NDVI 0.2 and NDVI 0.3.

SOS category 2: local thresholds

Instead of a global threshold, a locally tuned NDVI threshold is used (White *et al*., 1997) wherein the state of the ecosystem is indexed by transforming the NDVI to a 0 to 1 $NDVI_{ratio}$ as

$$NDVI_{ratio} = \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}}, \quad (1)$$

where NDVI is the daily NDVI and $NDVI_{max}$ and $NDVI_{min}$ are the annual maximum and minimum of the NDVI curve. SOS is defined as the DOY when 0.5 $NDVI_{ratio}$ is exceeded (note that an absolute rather than relative threshold may be used as simply the midpoint between the minimum and maximum NDVI). The 0.5 is designed to correspond to the timing of maximum NDVI increase; some evidence suggests that this corresponds to the initial leafing of the overstory canopy (White *et al*., 2000). Here three variations on this method have been applied: $Midpoint_{pixel}$, $Midpoint_{cluster}$, and PAT. For $Midpoint_{pixel}$, we set SOS to missing if any of the following occurred: more than 10% of observations were missing from the total 25 year time series; for any year, at least one observation was missing from composite periods 1, 2, 11, 12, 13, 14, 23, or 24 (determined to be critical for the detection of $NDVI_{max}$ and $NDVI_{min}$); more than three observations were missing during a year. For all the other pixels, we used a cubic smoothing spline to interpolate the composited data to a 0.5-day resolution and calculated $NDVI_{max}$ and $NDVI_{min}$ by a 7-day moving average. In the case of multiple solutions for SOS (e.g. a nonbell-shaped curve), we determined SOS as the earliest day with the increasing rate.

For $Midpoint_{cluster}$, we initially clustered pixels into homogenous clusters (White *et al*., 2005) with similar biological and physical characteristics, as defined by land cover (Loveland *et al*., 2000), monthly temperature and precipitation (Leemans & Cramer, 1991), and GTOPO30 elevation binned to 500 m increments. Within each cluster we averaged NDVI for each composite period and used the midpoint approach where the SOS threshold was defined as the half-maximum NDVI $\pm$ an error threshold (NDVI 0.025 for vegetation with maximum NDVI <0.5, otherwise 0.05). For $Midpoint_{cluster}$ we did not estimate SOS if maximum NDVI was <0.1 and determined SOS DOY as the SOS composite period multiplied by 15 (average composite period length).

We also used a variant of the Midpoint technique called percent-above-threshold [PAT, (White & Nemani, 2006)] in which the behavior of a group of pixels within

a level 3 ecoregion (Fig. S1) is tracked. In PAT, SOS is defined as the date at which 50% of pixels within an ecoregion have exceeded the median ecoregion midpoint NDVI (similar to 0.5 $NDVI_{ratio}$ but defined as an absolute NDVI). We defined PAT SOS only for those ecoregions in which only a single SOS was defined in each of the 25 study years.

SOS category 3: conceptual-mathematical

Here, an assumption is made that a particular mathematical function or suite of functions may be used to represent phenological development. We used two groups of conceptual mathematical models: smoothing methods and model fit methods.

For the smoothing methods, we first determined SOS with the delayed moving average method (DMA, (Reed *et al*., 1994)), in which SOS is the DOY at which a smoothed NDVI time series crosses a curve established from moving average models with an introduced time lag of fifteen composites, i.e. SOS occurs when the true NDVI exceeds the predicted NDVI of the prior composite periods.

Second, in the HANTS-FFT method, we used the HANTS-FFT algorithm (Roerink *et al*., 2000) to iteratively fit a series of frequencies to the NDVI profile (mean, yearly and half-yearly cycle) with the returned fast Fourier transform (FFT) coefficients then used to reconstruct the NDVI profile on a daily basis [reconstruction quality usually increasing with the number of component sinusoidal waves (Jakubauskas *et al*., 2001; Wagenseil & Samimi, 2006)]. SOS is derived as the point of maximum increase on the NDVI profile. Although the HANTS algorithm is robust, the estimation of SOS indicators may become unstable when there is no distinct phenological cycle and dual growing seasons are not detectable in the version of the algorithm used here.

For the model fit methods, we first used the Quadratic method (de Beurs & Henebry, 2008) and a model of the form

$$\mathrm{NDVI} = \alpha + \beta \mathrm{AGDD} + \gamma \mathrm{AGDD}^2, \qquad (2)$$

where AGDD are the accumulated growing degree-days in °C calculated from the North American Regional Reanalysis (Mesinger *et al*., 2006). We applied an exhaustive search algorithm that fits every pixel time series with multiple seasonal windows of differing length and starting period (best fits usually obtained if only the warm season data – usually April to October but variable by pixel – were used and preceding composites with low and static NDVI were excluded). The Quadratic method estimates SOS for each pixel-year as the first composite period of the best fitting model.

Second, in the Timesat method (Jönsson & Eklundh, 2004), we used a model fit consisting of a number of local model functions merged into a global function, thus allowing the fitted function to follow the behavior of the time series (not possible with a simple Gaussian model or lower order Fourier transform (Jönsson & Eklundh, 2002)). In this Timesat implementation, we used a local quadratic polynomial fit and the adaptive Savitzky-Golay filter applied to a moving window size of seven composites. We eliminated NDVI spikes larger than two times the standard deviation of the median values of the closest neighbors in the time series and fitted the remaining upper envelope. SOS is defined from the global model as the interpolated composite period when the NDVI has increased 20% of the seasonal amplitude from the growing season minimum level. Although the threshold level can be adjusted, the 20% threshold has been used effectively (Jönsson & Eklundh, 2002; van Leeuwen, 2008). We estimated SOS DOY values by the interpolated composite period multiplied by 15 days.

SOS category 4: hybrid

In the Gaussian method (closely related to an earlier Weibull curve approach, (Myneni *et al*., 1997)), which we applied at the level 3 ecoregion level (Fig. S1) as opposed to pixel by pixel (as for PAT), we first calculated the mean NDVI for each ecoregion for each composite period. In the next step, we fitted a Gaussian curve to the composites from April 1 until October 31 with SOS determined as the average DOY when the fitted NDVI curve reached 0.30, 0.35 and 0.40 (SOS not determined if the samples contain missing data or if the maximum value of the composites is <0.4). The Gaussian method is thus a hybrid of a conceptual mathematical model and a global threshold model.

Cryospheric/hydrologic comparisons

Snow

We used the 1982–2006 Northern Hemisphere weekly snow cover version 3 product from the National Snow and Ice Data Center [NSIDC, based on visual interpretations of multiple satellite inputs (Armstrong & Brodzik, 2005)]. For each week and ecoregion, we calculated the percent snow free (including NSIDC classes: snow, quality control snow, ice, quality control ice but dominated by variability in snow) and then selected only those ecoregions in which the percent snow free fell below 20% and rose above 80% in all years. For each year, we subsetted a vector from January 1 to the DOY at

which percent snow free exceeded 95% of the annual maximum (95% used because of frequent long plateaus slightly $<100\%$) and then calculated a normalized cumulative distribution function (CDF) such that January 1 was 0 and the date of 95% snow free was 1. We extracted the dates of initial, midway, and complete snowmelt (0.05, 0.5, and 0.95 on the normalized CDF).

Soil thaw

We used 1988–2005 estimates of the date of spring thaw from 19-GHz brightness temperatures recorded by the Special Sensor Microwave/Imager (SSM/I) on both the 06:00 and 18:00 hours equatorial crossing satellites. We compared both the am and pm products; we present results from the am estimates only, for which we found correlations to be consistently higher. Full details are available (Kimball *et al*., 2006) but the method relies on detecting a step change in the landscape dielectric constant as water changes from a frozen to liquid state, with concomitant increases in brightness temperature. The method is functional only in high latitude areas undergoing hard winter freezes.

Lake ice dynamics

We used 1982–2004 maps of ice breakup date created from ground-based observations on 65 water bodies in Minnesota, Wisconsin, Michigan, Ontario, and New York (Jensen *et al*., 2007). We projected water body locations and removed first order spatial (x and y) trends before variogram fitting (spherical model) and kriging (ArcView v8.3, Geostatistical Analyst Extension). The spatial extent of predictions was limited to the x and y extent of the observations and we made no predictions greater than 200 km from the nearest breakup date observation. Five lakes did not freeze in 2002, and one did not freeze in 1998; for these lakes and years, we used an existing method and inferred a breakup date by taking the average midpoint between the freeze and breakup dates of the five winters with the shortest ice durations (Assel & Robertson, 1995).

Spring hydrology

We used 1982 to 2006 indicators of spring hydrology calculated for 1149 stream gages in the US Geological Survey Hydroclimatic Data Network (locations believed to measure flows that are largely devoid of upstream diversions, reservoirs, and land use changes (Landwehr & Slack, 1992). We calculated the spring snowmelt onset date as the DOY when a snow-fed stream begins its rapid rise – defined as the day when the cumulative departure from annual mean flow is minimum – associated with the onset of major snowmelt (Cayan *et al*., 2001) (calculated only for stream gages that are reliably snow-fed, as defined by expert judgment). We also calculated the center of flow timing as the 'center of mass' of the hydrograph for each gage each water year. The center of flow is approximately, but not exactly, the date by which half the annual flow has passed and is described in more detail elsewhere (Stewart *et al*., 2004, 2005). Although both indices are designed to isolate temperature influences, precipitation timing may influence some records, especially in non-mountainous regions. We restricted our analyses to those ecoregions with at least 10 stream gages.

Modeled plant phenology

First, we used the Spring Indices (SI) model (Schwartz, 1997, 2003), which incorporates data from about 190 sites recording lilac (*Syringa chinensis*) and honeysuckle (*Lonicera tatarica*, *L. korolkowii*) phenology in the northeastern US. A step-wise multiple regression model combines the phenology observation with climatic indices (such as accumulation of winter chill and heat accumulation) to predict, among other events, first leaf and first bloom. Second, we used a model based on eddy covariance measurements of CO_2 exchange recorded at 12 deciduous forest sites from 36°N to 53°N. The model assumes that the start of spring ($D_{\text{leaf-out}}$) occurs at the onset of canopy photosynthesis when daily net CO_2 exchange transcends from the winter respiration phase to the spring/summer assimilation phase (Baldocchi *et al*., 2005). Conceptually, $D_{\text{leaf-out}}$ occurs when mean daily soil temperature equals and then surpasses the mean annual air temperature and may be calculated (Baldocchi *et al*., 2005) using air temperature alone:

$$D_{\text{leaf-out}} = 169.3 - 4.84 \times \text{mean annual air temperature}. \quad (4)$$

Since trees are unable to sense the mean annual air temperature *a priori*, we estimated mean annual temperature with a 2-year running mean. For both models, we used meteorological inputs from 1982 to 2003 1 km conterminous US Daymet records of gridded daily maximum, minimum, and average temperatures (Thornton *et al*., 1997). As both models were developed using data from temperate deciduous species, we restricted the model comparison with the Eastern Temperate Forest ecoregion.