Multiscale modeling of spring phenology across Deciduous Forests in the Eastern United States

ELI K. MELAAS[1], MARK A. FRIEDL[1] and ANDREW D. RICHARDSON[2]

[1]*Department of Earth and Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA,* [2]*Department of Organismic and Evolutionary Biology, Harvard University, HUH, 22 Divinity Avenue, Cambridge, MA 02138, USA*

Abstract

Phenological events, such as bud burst, are strongly linked to ecosystem processes in temperate deciduous forests. However, the exact nature and magnitude of how seasonal and interannual variation in air temperatures influence phenology is poorly understood, and model-based phenology representations fail to capture local- to regional-scale variability arising from differences in species composition. In this paper, we use a combination of surface meteorological data, species composition maps, remote sensing, and ground-based observations to estimate models that better represent how community-level species composition affects the phenological response of deciduous broadleaf forests to climate forcing at spatial scales that are typically used in ecosystem models. Using time series of canopy greenness from repeat digital photography, citizen science data from the USA National Phenology Network, and satellite remote sensing-based observations of phenology, we estimated and tested models that predict the timing of spring leaf emergence across five different deciduous broadleaf forest types in the eastern United States. Specifically, we evaluated two different approaches: (i) using species-specific models in combination with species composition information to 'upscale' model predictions and (ii) using repeat digital photography of forest canopies that observe and integrate the phenological behavior of multiple representative species at each camera site to calibrate a single model for all deciduous broadleaf forests. Our results demonstrate variability in cumulative forcing requirements and photoperiod cues across species and forest types, and show how community composition influences phenological dynamics over large areas. At the same time, the response of different species to spatial and interannual variation in weather is, under the current climate regime, sufficiently similar that the generic deciduous forest model based on repeat digital photography performed comparably to the upscaled species-specific models. More generally, results from this analysis demonstrate how *in situ* observation networks and remote sensing data can be used to synergistically calibrate and assess regional parameterizations of phenology in models.

Introduction

The springtime phenology of leaf emergence and development in temperate deciduous broadleaf forests is highly correlated with air temperature (Lechowicz, 1984). Seasonality in photoperiod also influences the timing of leaf development, especially in late successional temperate tree species, although the exact role of photoperiod is not well understood (Chuine *et al.*, 2010; Körner & Basler, 2010). In addition to providing an important diagnostic of climate variation and change, phenology is also a key regulator of seasonal carbon, water, and energy exchanges in many ecosystems (Fitz-jarrald *et al.*, 2001; Richardson *et al.*, 2010). Hence, understanding of mechanistic controls and accurate representation of vegetation phenology in ecosystem models that are used to simulate biosphere–atmosphere interactions is critical (Richardson *et al.*, 2012).

Most widely used land surface models (LSMs) simulate the onset of leaf development using prescribed dates estimated from satellite remote sensing or empirically derived functions based on cumulative chilling and forcing units (Yang *et al.*, 2012). Recent studies, however, have demonstrated that current representations of phenology in LSMs are not realistic. Using *in situ* data from eddy covariance sites located in North American deciduous forests, Richardson *et al.* (2012) showed that LSMs consistently predict the onset of spring phenology early by more than 2 weeks and, in some cases, by as much as 10 weeks. While the source of this bias was different for each model, overly simplified model representations and overfitting of model coefficients (or both) are widely known sources of error in phenological models (Linkosalo *et al.*, 2008; Melaas *et al.*, 2013a).

An additional source of error, which we examine here, is that despite previous research showing substantial variability in the response of different

Correspondence: Eli Melaas, tel. +1 248 343 9278, fax +1 617 353 8399, e-mail: emelaas@bu.edu

temperate forest species to warming, chilling, and photoperiod controls (e.g., Korner & Basler, 2010; Jeong *et al.*, 2012; Migliavacca *et al.*, 2012), most models do not account for species-specific or community-level responses to climate forcing. An obvious solution to this problem is to include species-specific representations of phenology. However, doing so at continental or even regional scales in LSMs poses significant challenges. In particular, current LSMs generally represent vegetation at coarse spatial resolutions (~0.05–1°) using mosaics of highly generalized plant functional types to capture first-order heterogeneity in the ecophysiology of terrestrial vegetation. Refinement of such models to include representations for species-specific responses requires geographically explicit information related to species composition that does not exist for most locations.

In this paper, we propose a solution to this limitation that stratifies plant functional types according to community composition, thereby capturing local controls and geographic variation in springtime phenology to climate forcing. To develop and test our approach, we focus on temperate deciduous forests in the eastern United States and disaggregate the region encompassing these forests into five forest community types. Our justification for this strategy is that geographic variation in forest community composition reflects variation in climate, elevation, soil conditions, and land use history, and by extension, also reflects differential adaptation to bioclimatic conditions including distinct ecophysiological traits that influence spring leaf emergence (Lechowicz, 1984). For example, southern species representative of oak-gum and oak-hickory forests tend to be ring-porous with large diameter vessels prone to embolism from freezing. Northern forest species representative of aspen–birch and maple–beech–birch forests, on the other hand, are overwhelmingly diffuse-porous with narrow-diameter vessels and lower conducting capacity. As a result, ring-porous species tend to leaf out later than diffuse-porous species under the same forcing conditions because ring-porous species require new wood growth to supply water to their newly expanding leaves (Wang *et al.*, 1992).

To test this approach, our analysis makes use of data acquired at multiple scales. Specifically, ground observations, near-surface remote sensing, and satellite remote sensing offer valuable information regarding spatial variation in the timing of spring leaf development at the scale of individual trees [e.g., the USA National Phenology Network (USA-NPN); Betancourt *et al.*, 2007], forest stands (e.g., the PhenoCam network; Richardson *et al.*, 2007), and regions (e.g., the Moderate Resolution Imaging Spectroradiometer (MODIS); Ganguly *et al.*, 2010). Phenology data from individual trees

provide detailed information regarding the nature and magnitude of within-species variability in phenology, but these patterns are hard to generalize at regional and larger scales. Satellite data, on the other hand, provide spatially integrated measurements for the timing of leaf onset over landscapes that include mixtures of species, plant functional types, and land cover types (e.g., forests vs. agriculture) (Badeck *et al.*, 2004; Klosterman *et al.*, 2014). Hence, phenological models that are calibrated to remote sensing data may include biases that depend on the composition of land cover or plant functional types in model grid cells (e.g., Jenkins *et al.*, 2002). At intermediate spatial scales, imagery from repeat digital photography provides data that capture canopy dynamics at scales that range from individual trees to forest stands, which therefore complement observations from both ground-based networks and remote sensing (Hufkens *et al.*, 2012; Klosterman *et al.*, 2014).

Together, these three sources of data provide a rich basis for calibrating species-specific and more generalized chilling and forcing sum models of phenology (e.g., Raulier & Bernier, 2000). To date, however, most modeling efforts focused on deciduous forests in the eastern United States have utilized data from individual sites or observation networks that span relatively small geographic ranges (e.g., Jeong *et al.*, 2012; Migliavacca *et al.*, 2012). As a result, the nature and magnitude of geographic variation in photoperiod and temperature controls on spring phenology in temperate trees remains poorly understood, and models that predict leaf onset and development do not generalize well across sites (e.g., Chuine *et al.*, 1998; Richardson *et al.*, 2006). This limits the utility of models for understanding and predicting how the phenology of temperate forests will be affected by climate change, and by extension, how changes in phenology are likely to affect biosphere–atmosphere interactions in the future (Richardson *et al.*, 2013).

With these issues in mind, the goal of this paper was to develop and test models of springtime leaf onset in temperate deciduous forests that account for community-level differences in phenological behavior. The resulting models offer new insight regarding how ecological controls vary as a function of both climate forcing and community species composition and, by extension, provide a basis for geographically explicit model parameterization of phenological dynamics in LSMs. To accomplish this goal, our analysis includes three main elements. First, we use networks of ground observations for springtime phenology, including data from USA-NPN and PhenoCam, to develop and calibrate process-based spring phenology models for each forest community type (see Table 1 for a complete

Table 1 Summary of observation datasets used to calibrate and test phenology models across the study region

Name	Type	Years	Spatial resolution	Application
USA National Phenology Network	Ground phenology	2009–2013	Individual Trees	Calibrate species-specific GDD model parameters
PhenoCam network	Near-surface digital photography	2000–2012	Forest Canopy	Calibrate PhenoCam GDD model parameters
MODIS	Satellite remote sensing	2001–2012 (excluding 2007)	500 m	Regional testing of GDD models
USHCN	Ground meteorology	2008–2013	Individual Stations	Train species-specific and PhenoCam GDD models
DAYMET	Interpolated surface meteorology	2000–2012	1 km	Regional testing of GDD models
FIA	Species composition and distribution	2012	25 km hexagons	Spatial weighting of species-specific GDD models
NLCD	Tree canopy and land cover	2001 and 2006	30 m	Deciduous forest stratification

summary of observation datasets used in the analysis). Second, we use species composition data to identify community-level clusters of tree species at the regional scale, which are subsequently used to define regions with common species composition. Third, we estimate and apply our spring phenology models over the entire eastern temperate deciduous forest biome using gridded climate data and compare our model results with observations of spring phenology derived from moderate resolution satellite remote sensing data.

Data and methods

Study region

The study region encompassed all temperate forests in the eastern United States located between 68°W–95°W and 30°N–50°N, including most of the eastern temperate forest ecoregion and southern portions of the northern forest ecoregion defined by the Level 2 United States Environmental Protection Agency (http://www.epa.gov/wed/pages/ecoregions.htm) (Fig. 1). Forests in the study region include five main community types: (i) maple–beech–birch forests in the northeast, (ii) aspen–birch forests in the northern Great Lakes region, (iii) oak-hickory forests extending across the Ozark and Ouachita-Appalachian mountain ranges, (iv) oak-gum mixed hardwoods in the south, and (v) elm–ash–cottonwood across the Interior Lowlands. Climate in this region is classified as either humid continental or humid subtropical, with mean annual temperatures and precipitation ranging from −2 to 21 °C and from 500 to 2000 mm, respectively. The region is heavily forested, but also includes substantial urban and agricultural areas, which were excluded from this analysis.

Ground observations

National phenology network. The USA-NPN compiles and distributes observations of spring and autumn phenology collected by citizen scientists across the United States. In spring (nominally March, April, and May), trained observers record dates for key phenophase events (including flowering, leaf emergence, and leaf maturation) using consistent and well-defined protocols (see www.usanpn.org; Denny *et al.*, 2014). As part of their mission, USA-NPN also performs quality control and screening of observations provided by citizen scientists. For this study, we identified all available USA-NPN observations that were collected within our study domain and extracted annual spring onset dates for individual trees based on the first observation in each year when the phenophase attribute was recorded as 'Leaves', indicating leaf emergence or unfolding. This dataset encompassed all available observations for deciduous tree species collected between 2009 and 2013, excluding species with fewer than 30 site-years. The final dataset included 980 site-years for 10 tree species collected at 692 sites in 28 states (Table 2; Fig. 1a). While this dataset does include finite levels of uncertainty, Gerst *et al.* (2015) specifically explored how errors in USA-NPN data might affect results from models. Results from their analysis, which spanned a similar geographic region to this study, showed that at regional scales where sample sizes are large, the quality of USA-NPN data is more than sufficient to support statistical modeling.

To calibrate heat- and chill-sum models that we used to predict the timing of leaf emergence, we downloaded daily minimum and maximum temperatures from 2008 to 2013 for United States Historical Climatology Network (USHCN) stations located in the study region where at least 75% of observations were available (http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html). Using these data, we estimated daily mean temperatures at each USA-NPN site using inverse distance weighting of observations from the 15 closest USHCN stations, corrected for elevation differences between USHCN stations and USA-NPN observation sites using an environmental lapse rate of 6.5 °C km^{-1}.

PhenoCam data. The PhenoCam network uses repeat digital photography to monitor vegetation phenology (http://
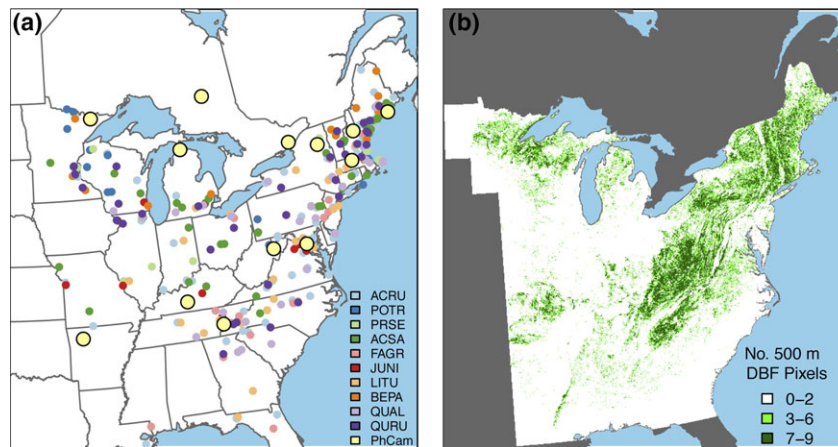
Fig. 1 (a) Map of NPN and PhenoCam observation sites used to parameterize growing degree-day models; (b) map of deciduous forest coverage across the study region. Each 1 km grid cell is labeled according to the number of 500 m pixels classified as deciduous forest in a 3 × 3 window centered on the grid cell. Grid cells with fewer than 3 deciduous pixels were excluded from the study. We used the National Land Cover Database to classify deciduous forest and, therefore, did not provide coverage for Canada.

Table 2 Local test statistics for species-specific growing degree-day models across NPN, PhenoCam, Harvard Forest, and Hubbard Brook observations. Statistics are based on performance across all site-years for each dataset

SPP	SPP Code	No. of obs.	Model	NPN/PhenoCam			Harvard Forest SPP, 1990–2011			Hubbard Brook SPP, 1989–2012		
				RMSE	MBE	COR	RMSE	MBE	COR	RMSE	MBE	COR
Red Maple	ACRU	330	PAR1	9.9	1.2	0.78	5.2	2.2	0.68			
Sugar Maple	ACSA	133	SW4	8.6	0.2	0.74	4.6	−3.2	0.83	4.8	−0.4	0.83
Paper Birch	BEPA	62	SW2	6.4	−1.5	0.79	4.9	−1.6	0.61			
American Beech	FAGR	72	SW2	10.9	−3.8	0.79	8.7	−6.5	0.39	6.8	−3.5	0.69
Quaking Aspen	POTR	63	SW4	8.6	0.7	0.79	5.2	−3.4	0.85			
Black Cherry	PRSE	75	ALT1	10.6	−1.0	0.74	8.1	6.0	0.46			
Black Walnut	JUNI	37	SW2	10.3	−0.1	0.71						
Yellow Poplar	LITU	56	SW2	9.7	1.5	0.74						
White Oak	QUAL	67	ALT1	9.7	0.5	0.81	8.2	6.6	0.57			
N. Red Oak	QURU	75	SW1	9.7	−0.3	0.81	5.1	3.4	0.80			
PhenoCam	PhCam	80	ALT1	8.5	−1.2	0.82						

phenocam.sr.unh.edu). In addition to providing access to the raw imagery, the PhenoCam project has developed methods for screening and preprocessing image data, and for extracting quantitative metrics that are useful for characterizing phenological dynamics in vegetation (Richardson *et al.*, 2007). For this work, we used time series of the green chromatic coordinate (GCC) index to estimate the timing of spring green-up from PhenoCam image time series (Sonnentag *et al.*, 2012). We used all eastern deciduous forest sites in the PhenoCam network with at least 3 years of data and preprocessed GCC time series at each site using a 90th percentile filter applied to 3-day moving windows for manually defined regions of interest in webcam imagery that are selected to provide representative measurements of tree canopy phenology at each site. Application of this filter significantly reduces noise in GCC time series introduced by changes in scene illumination related to

variability in cloud cover and other atmospheric conditions (Sonnentag *et al.*, 2012). Three-day GCC data were interpolated to daily values using cubic smoothing splines, and the timing of leaf emergence was identified as the date when the GCC reached 50% of its springtime amplitude at each site. The resulting dataset included 80 site-years from 13 camera locations (Klosterman *et al.*, 2014).

MODIS observations

The MODIS Nadir Bidirectional Reflectance Distribution Function-Adjusted Reflectance (NBAR) product (MCD43A4) provides surface reflectance measurements at 500 m spatial resolution that are normalized to a consistent nadir view geometry at eight daytime steps using MODIS observations acquired during overlapping 16-day periods (Schaaf *et al.*,

2002). For this work, we computed the enhanced vegetation index (EVI; Huete *et al.*, 2002) from NBAR data collected between 2001 and 2012 and used the resulting NBAR-EVI time series to estimate the timing of spring onset for all deciduous forest pixels located in 6 MODIS tiles spanning the eastern United States. Using a procedure similar to the one we used to estimate the timing of leaf emergence from GCC time series, we generated daily time series of NBAR-EVI using cubic spline interpolation applied to the 8-day data and identified the timing of spring onset as the day of year (DOY) when NBAR-EVI values exceeded 20% of the springtime amplitude in EVI at each pixel. We chose 20% instead of 50% based on results shown by Klosterman *et al.* (2014), who performed an extensive comparison of phenology metrics derived from satellite remote sensing and PhenoCams. In 2007, a widespread frost event occurred, which significantly damaged leaves across the southern half of the study region and subsequently delayed MODIS-detected spring onset by 3–4 weeks (Gu *et al.*, 2008). Because the method to estimate spring onset from MODIS that we describe above does not allow for the possibility of multiple 'green-up' phases in the same spring, retrievals of spring onset in 2007 from MODIS were problematic throughout much of the southern portion of our study region. We therefore excluded 2007 from our analysis.

Daymet meteorological data

The Daymet dataset provides gridded surface meteorological data for the contiguous United States at high spatial and temporal resolution (Thornton *et al.*, 2014; http://daymet.ornl.gov/). The dataset uses 35 years (1980–2014) of minimum and maximum temperature and precipitation measurements interpolated from weather stations in the Global Historical Climatology Network (Williams *et al.*, 2006) to provide daily gridded data at 1 km spatial resolution. For this analysis, we used daily temperature data from 2000 to 2012 for 101 $2° \times 2°$ Daymet tiles located within the study region. To merge Daymet data with MODIS observations, we calculated the median spring onset date for 3×3 MODIS pixel windows centered on each 1-km Daymet grid cell. Only grid cells with at least 3 deciduous forest pixels (based on land cover information, described below) in the 3×3 window were included in the analysis (Fig. 1b).

Species composition and distribution

The Forest Inventory and Analysis (FIA) program of the U.S. Forest Service routinely collects forest stand measurements at nearly 200 000 plots nationwide. In the analysis presented below, we used spatially averaged basal areas for dominant tree species in ~54 000 ha hexagons estimated from plot-level FIA data (~22 plots per hexagon) to characterize geographic variation in forest composition across our study region (Wilson *et al.*, 2013). To do this, we first assigned each deciduous tree species to one of five main forest types defined in Appendix D of the FIA Database User's Manual: (i) oak-hickory, (ii) oak-gum, (iii) elm–ash–cottonwood, (iv) maple–beech–birch, and (v) aspen–birch. We then used the average species-level basal area data to map the proportion of each forest type in each hexagon in our study region (Fig. 2).

Deciduous forest stratification

MODIS NBAR-EVI data have a nominal ground resolution of 500 m. Hence, individual pixels generally contain mixtures of land cover and plant functional types that have different phenological responses to weather and climate forcing. To minimize variability in MODIS time series associated with subpixel heterogeneity in land cover and plant functional types, we used the 2006 National Land Cover Database (NLCD) in association with tree canopy cover information from the 2001 NLCD (Homer *et al.*, 2007; Fry *et al.*, 2011), both at 30 m spatial resolution, to identify 500 m pixels where deciduous forest cover and tree canopy density were both greater than or equal to 60 percent. Pixels that did not meet these criteria were excluded from the analysis.

Spring phenology models

We used USA-NPN, PhenoCam, and MODIS data to calibrate and test 13 models that use thermal forcing or combined thermal forcing and chilling to predict the timing of spring onset. Each model was based on one of four functional forms in which spring onset is predicted to occur when the state of forcing ($S_f(t)$) reaches a critical sum of forcing units (F^*). Below, we describe each of these functional forms. Complete details regarding each of the 13 models that we tested are provided in the Supporting information 1.

Spring warming model. In the spring warming model (SW), accumulated forcing is computed as a function of air temperature using a logistic function (Sarvas, 1974)

$$S_f(t) = \sum_{t_0}^{t_{\text{pheno}}} \max\left[\frac{28.4}{1 + \exp(3.4 - 0.185 * T_{\text{air}})}, 0 \right] \quad (1)$$

where T_{air} is daily mean air temperature, t_0 is the start date when forcing begins to accumulate, and t_{pheno} is the date when $S_f(t)$ exceeds a prescribed threshold (F^*).

Sequential model. The sequential model (SEQ) assumes that forcing accumulation does not occur until a critical threshold of accumulated chilling units (C^*) is reached, where the state of chilling ($S_c(t)$) increases when T_{air} falls below a prescribed temperature threshold, T_c:

$$S_c(t) = \sum_{t_0}^{t_1} \begin{cases} 1 & T_{\text{air}} < T_c \\ 0 & T_{\text{air}} \geq T_c \end{cases} \quad (2)$$

In this model, t_1 is the date when chilling requirements are met and forcing accumulation begins. Unlike the SW model described above, the rate of accumulated forcing in the sequential model is linearly related to air temperature:

$$S_f(t) = \sum_{t_1}^{t_{\text{pheno}}} \max(T_{\text{air}} - T_f, 0) \quad (3)$$

where T_f is a base temperature below which no leaf development is assumed to occur.
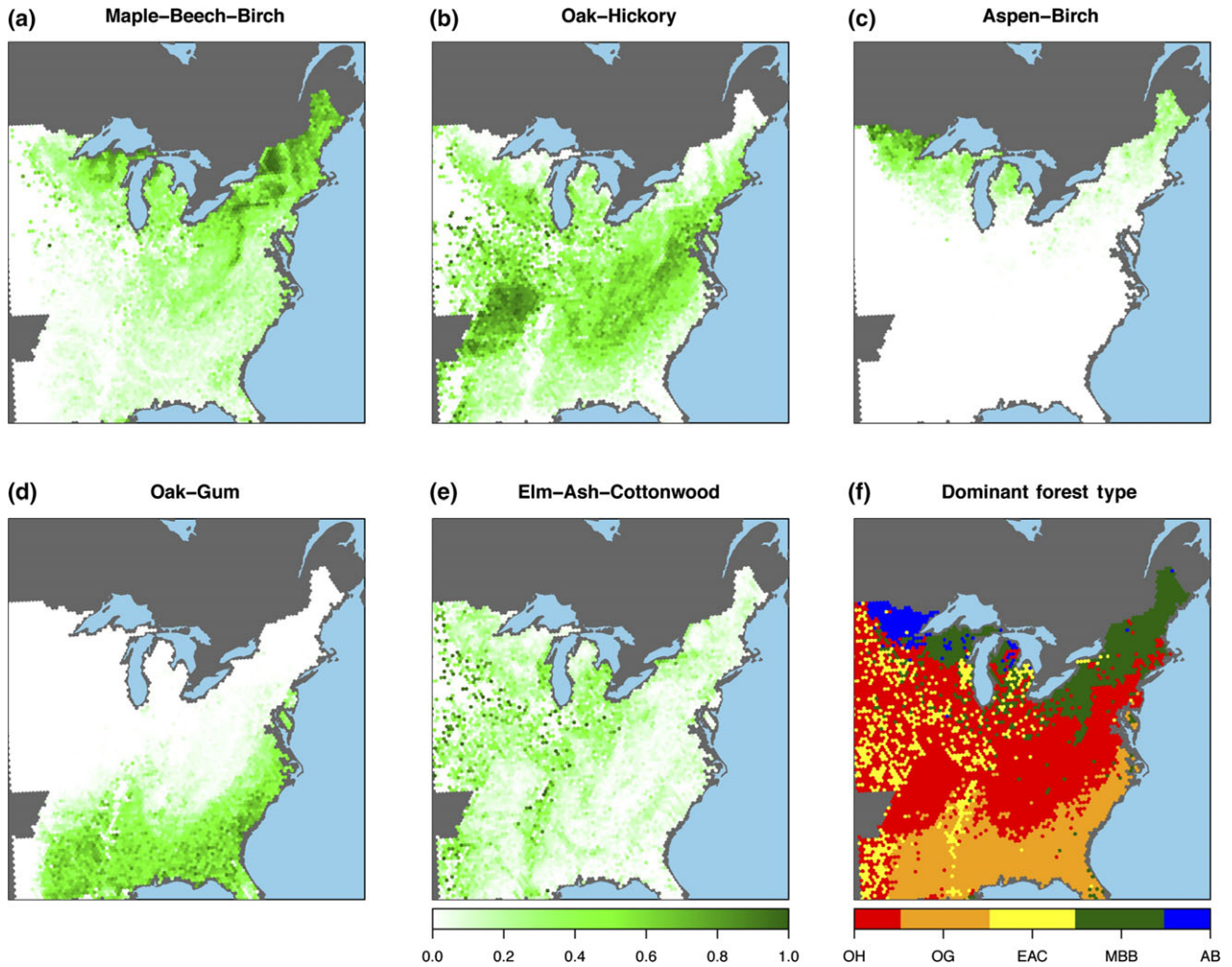
Fig. 2 (a–e) Forest type composition across the study region according to FIA Phase 2 plot-level basal area measurements. Plot-level measurements were spatially averaged within each 25 km hexagon and (f) dominant forest type within each 25 km hexagon.

Alternating and parallel models. The alternating (ALT) and parallel (PAR) models both assume that the forcing required for spring onset decreases exponentially as accumulated chilling increases. In the ALT model, chilling and forcing take turns accumulating from an initial start date relative to a single base temperature. Specifically, chilling and forcing accumulate according to Eqns (2) and (3), respectively, until

$$S_f(t) \geq a + b * \exp(c * S_c(t)) \qquad (4)$$

where *a*, *b*, and *c* are optimized to maximize the predictive accuracy of the model and where $c < 0$. In the PAR model, chilling and forcing occur simultaneously, where forcing accumulation follows Eqn (1) and chilling accumulates according to a triangular function:

$$S_c(t) = \sum_{t_0}^{t_{pheno}} \begin{cases} 0 & T_{air} \geq 10.4 \text{ or } T_{air} \leq -3.4 \\ \frac{x(t)+3.4}{T_{opt}+3.4} & -3.4 < T_{air} \leq T_{opt} \\ \frac{x(t)-10.4}{T_{opt}-10.4} & T_{opt} < T_{air} < 10.4 \end{cases} \qquad (5)$$

where T_{opt} is the optimum chilling temperature. Leaf emergence is predicted to occur when

$$S_f(t) \geq a * \exp(b * S_c(t)) \qquad (6)$$

where *a* and *b* are optimized to minimize model errors, and $b < 0$.

Each of these functional forms has been widely used with site-specific *in situ* phenology observations (Jeong *et al.*, 2012; Migliavacca *et al.*, 2012) and satellite remote sensing observations (Fisher *et al.*, 2007; Yang *et al.*, 2012). Here, we estimate spring onset models using the basic formulations described above and two simple variants: (i) where accumulation of chilling or forcing is based on photoperiod rather than a fixed date (i.e., substitute p_0 for t_0) and (ii) where the base temperatures and chilling and forcing requirements are estimated as functions of mean annual temperature.

Model calibration and validation

To estimate coefficients for each of the 13 models evaluated in the analysis, we used species-specific observations for the timing of leaf emergence derived from USA-NPN observations of

temperate deciduous forest species located within our study region. USA-NPN observations were collected between 2009 and 2013 and included 3 years with relatively normal spring (defined here as March–May) temperatures relative to 20th century climatological means (2009, 2011, and 2013), and 2 years with anomalously warm spring temperatures (2010 and 2012; see Friedl *et al.*, 2014). To evaluate the robustness of estimated models for predicting phenology under projected future climate conditions, we calibrated each model for each species using observations from 2009, 2011, and 2013 (i.e., relatively average years) and then evaluated their performance using observations from 2010 and 2012 (i.e., years with climatologically warm springs).

Following Chuine *et al.* (1998), we optimized coefficient values for each model using Monte Carlo techniques (Metropolis *et al.*, 1953). Once the best parameter set for each model species pair was identified, we also estimated 95% confidence intervals for each parameter using a chi-square test (Franks *et al.*, 1999; Migliavacca *et al.*, 2012). To select the best model for each species, we used the small-sample corrected Akaike information criterion (AIC$_c$) (Burnham & Anderson, 2002) based on the residual sum of squared errors for all observations. The same procedure was applied to parameterize and select an optimal 'species-ignorant' model using PhenoCam observations.

In addition to model calibration and testing using USA-NPN data, we performed two complementary sets of model evaluations. First, we tested the USA-NPN species-specific models using long-term *in situ* observations of phenology collected at the Hubbard Brook Experimental Forest in New Hampshire and the Harvard Forest Long Term Ecological Research site in Massachusetts, focusing on the ability of the models to capture interannual variability in the timing of spring leaf out at each site (see Richardson *et al.*, 2006 for more details on each dataset). Second, we assessed the ability of the species-specific and species-ignorant models estimated from USA-NPN and PhenoCam data, respectively, using daily mean temperature data from Daymet to predict annual spring phenology across the study region between 2000 and 2012. We then calculated the root-mean-square error (RMSE), mean bias error (MBE; here, bias is measured as observed minus predicted), and Pearson's correlation coefficient between the predicted spring onset dates from each of the models and observations of spring onset dates from MODIS.

We characterized regional model performance in two ways. First, we evaluated each species-specific model within each forest type and across the entire study region. Second, we generated predictions at each Daymet grid cell using a weighted average of species-specific model predictions based on FIA forest type composition (Fig. 2). Specifically, in each forest type where sufficient USA-NPN data were available, we used predictions for the most common species (maple–beech–birch – red maple; aspen–birch – quaking aspen; oak-hickory – white oak), while in the other forest types, we used predictions for species with a comparable native range (i.e., for oak-gum, we used a model tuned to yellow poplar; for elm–ash–cottonwood, we used a model tuned to red maple).

Results

Model calibration and validation

Table 2 provides a summary of the best-fit USA-NPN species-specific and PhenoCam models and their relative performance. Across species, AIC$_c$ values suggest that SW models tend to be better supported by the data than more complex chilling models. In particular, SW models were selected for seven of ten species, two of which allowed the critical forcing requirement (F*) to vary as a function of mean annual temperature. Among chilling models, ALT models were selected three times, while the PAR model was selected once. Relative to USA-NPN or PhenoCam observations, models generally predicted spring onset within 8–10 days with relatively low bias and Pearson's correlation coefficients between 0.7 and 0.8.

When the models were assessed using ground observations from Harvard Forest and Hubbard Brook, model performance generally performed best for species that were dominant at each site. At Harvard Forest, the red maple and northern red oak (the two most common species at Harvard Forest) models predicted spring onset within approximately 5 days of observations (on average), with model errors for less common species such as black cherry and American beech that were much larger. At Hubbard Brook, where American beech is more common, the model tuned to NPN data for this species showed better agreement with observations than the same model at Harvard Forest. The sugar maple model, which includes a mean annual temperature control on F*, showed high predictive accuracy at both sites.

Model parameters varied across species in terms of F* and the optimal start date for photoperiod and chilling/forcing accumulation (t_0 or p_0; see Table S2). For example, models for species that tend to be more common in southern parts of the eastern deciduous forest (e.g., yellow poplar and white oak) tended to have earlier start dates or lower photoperiod thresholds and higher F*, while models for more northern species (e.g., American beech, paper birch, and quaking aspen) tended to have later start dates or photoperiod thresholds and lower F* requirements. Photoperiod control was also important in six of ten species models, and most of the species-specific models and the PhenoCam model do not accumulate chilling/forcing until mid-March (~DOY 75).

Regional model evaluation using daymet and MODIS

Following model calibration and local validation against *in situ* measurements, we used Daymet data to

predict spring onset dates for each of the USA-NPN models and the PhenoCam model at 1 km spatial resolution over the study region and then compared these predictions against spring onset dates from MODIS for the period 2001–2012. In addition to evaluating model performance across all years and all Daymet grid cells (i.e., 'Total' in Figs 3 and 4), we assessed how model performance was affected by forest composition, as reflected by the five dominant forest types in the study region (Fig. 2). The resulting models RMSEs and MBEs are shown in Figs 3 and 4, respectively, and boxplots of Pearson's correlation coefficients and model slopes are provided in Figs S1 and S2.

For the entire study region, models predicted spring onset with a median RMSE of 5–10 days (Fig. 3f). The northern red oak (QURU) and sugar maple (ACSA) models provided the highest overall accuracy, while black walnut (JUNI) and yellow poplar (LITU) were least accurate on average. The range in mean bias among all models was ±10 days, while Pearson's correlation coefficients averaged 0.7–0.8.

Closer inspection of model results reveals substantial variation in model performance that depends on forest type (Figs 3a–e and 4a–e). For example, the QURU model consistently predicts spring onset within about 5 days across all forest types. In contrast, other native oak-hickory species models show variable accuracy in predicting oak-hickory forest phenology relative to other forest types. For example, the LITU model predicts spring onset roughly 10 days too early in oak-hickory forests, while the JUNI model shows relatively low RMSE and MBE in oak-hickory forests, but much higher errors in other northern forest types such as aspen–birch and maple–beech–birch. Finally, while the white oak (QUAL) model has a median RMSE less than 7 days across all forest types, its predictions are negatively correlated with observed spring onset in more than half of the oak-gum forest region.

Among aspen–birch species models (red-colored boxplots), the quaking aspen (POTR) model performed consistently well across all forest types, with median RMSEs of less than 7 days, generally low bias, and Pearson's correlation coefficients greater than 0.6. The paper birch (BEPA) model, on the other hand, is more accurate for aspen–birch and maple–beech–birch forests than in more southern elm–ash–cottonwood, oak-hickory, and oak-gum forests. Compared with other species models, both the BEPA and POTR models have relatively low F*. The BEPA model also has a late starting photoperiod (i.e., accumulation of thermal forcing starts later than for other models) and, as a result, significantly underestimates interannual variation in spring onset across southern oak-hickory and oak-gum forests.

The maple–beech–birch species models are also quite variable in terms of their model structure and parameterization across species. The ACSA model, a SW model for which the period over which F* accumulates is a function of mean annual temperature, performs well across all forest types except oak-gum. Similar to BEPA, the American beech (FAGR) model has a relatively late photoperiod trigger, but also requires a relatively large F* to trigger leaf emergence. As a result, this model's performance is relatively weak in the southern portions of the study region. The red maple (ACRU) and black cherry (PRSE) models, which depend on both chilling and forcing temperatures, were relatively robust across all forest types. However, both models were consistently biased early relative to MODIS spring onset dates.

Evaluation of upscaled models

In the final part of our analysis, we evaluated two approaches to parameterizing regional-scale phenological models. First, we computed predictions for the timing of leaf emergence at each Daymet grid cell based on an area-weighted estimate from each of the species-specific models, where the area weights were generated based on FIA basal area data (Fig. 2). Second, we used a single model based on PhenoCam data. Results from these analyses showed that both the area-weighted and PhenoCam models performed well across the study region (Fig. 5k, l). Both models predicted spring onset with a median RMSE of approximately 5 days and with relatively low bias (Figs 4f and 5f). The PhenoCam model was particularly accurate across oak-hickory and maple–beech–birch dominated forests, but was less accurate and significantly biased (late) in oak-gum forests. Figs 5 and 6 show maps illustrating spatial patterns in models RMSE and MBE for the species-specific, PhenoCam, and forest type-weighted models. As these figures show, the PhenoCam and forest type-weighted models accurately predict spring onset in the New England and the upper Great Lakes regions, but are less accurate and biased (early) across northern Michigan and the Allegheny Mountains (eastern-central portion of the study region).

Discussion

There is broad consensus that future warming is likely to influence the growing season of temperate deciduous forests and that these changes have substantial potential to impact ecosystem function and carbon budgets at local- to regional scales (e.g., Keenan et al., 2014. However, recent literature has demonstrated that
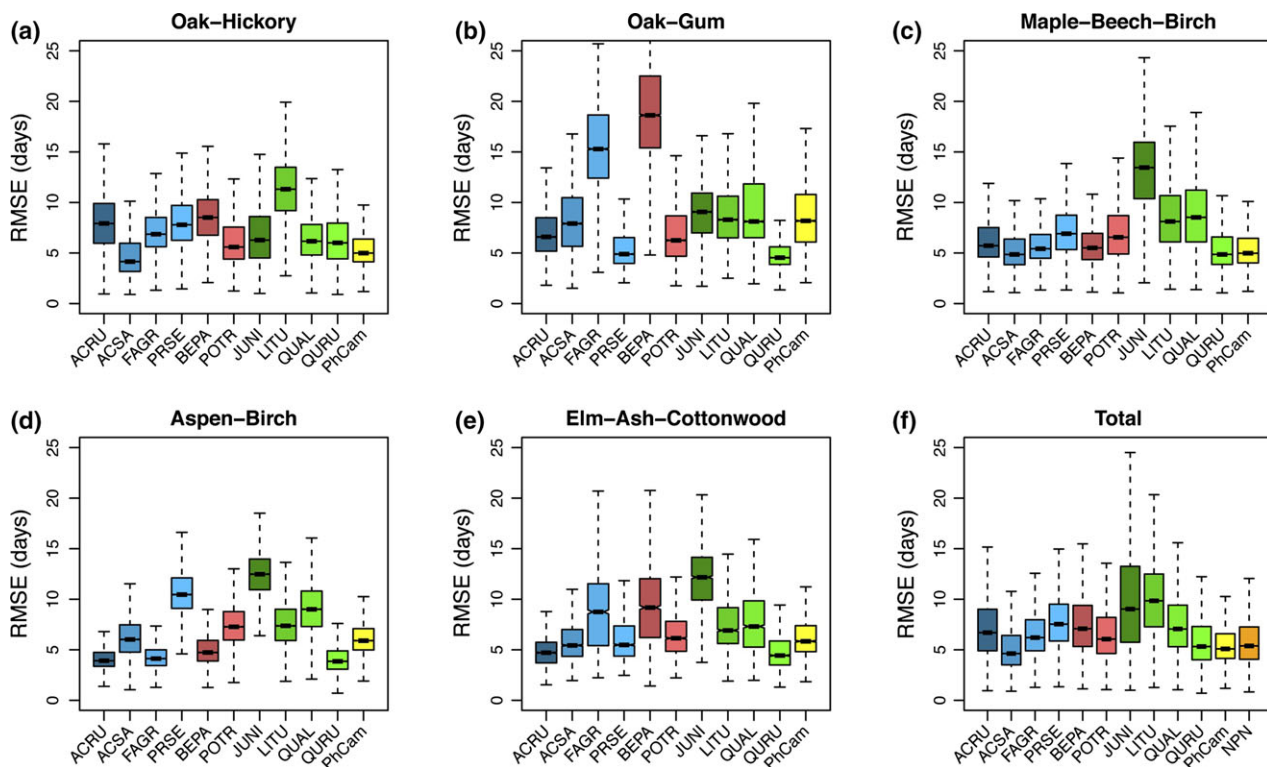
Fig. 3 Boxplots of root-mean-square error (days) between MODIS estimated spring onset and model predicted spring onset across (a–e) pixels dominated by each forest type, according to 25 km FIA hexagon maps (see Fig. 2) or (f) the entire study region.

temperate deciduous forest species have shown a wide range of phenological responses to climate changes during the last half-century (Badeck *et al.*, 2004; Polgar & Primack, 2011; Richardson *et al.*, 2013). Hence, there is considerable uncertainty regarding how, specifically, future changes in temperature will affect key phenophase events such as the timing of leaf emergence and senescence in this biome. Despite evidence showing that species-specific differences in phenological sensitivity to environmental factors can be substantial (e.g., Vitasse *et al.*, 2009; Jeong *et al.*, 2012), current LSMs, which are designed to simulate biological processes that affect long-term energy, water, and carbon budgets, do not account for differences in phenological behavior among species. Indeed, Richardson *et al.* (2012) have shown that bias in the representation of phenology in these models leads to substantial errors in modeled carbon fluxes.

To address this problem, we explored how geographic patterns in species composition influence the timing of spring onset in deciduous forests and assessed methods for parameterizing geographic variation in models of springtime phenology. To do this, we used a combination of *in situ* observations, species composition maps, gridded climate data, and moderate spatial resolution remote sensing data. Our results

reveal differences in the thermal forcing required to initiate leaf emergence among species and, consequently, variable accuracy in species-specific models across forest types. At the same time, some species-specific models – notably the ACSA and QURU models – and the species-independent PhenoCam model showed relatively good accuracy predicting the timing of leaf emergence in deciduous forests across our study region. These latter results suggest that the phenology of both sugar maple and red oak is relatively robust and representative of broader forest community phenology over our study region and that PhenoCam data successfully integrate the phenological behavior of multiple species. More generally, regional weaknesses in species-specific models in combination with the success of both 'up-scaled' modeling approaches (i.e., based on PhenoCam data and weighted species-specific models) demonstrate that realistic representation of forest community composition is important in models that simulate phenological events over large scales.

Our results also illuminate several important functional relationships between air temperatures and the timing of spring onset in eastern deciduous forests of the United States. First, we find that air temperatures in early to late spring (March through May) are sufficient to explain the majority of interannual variability in the
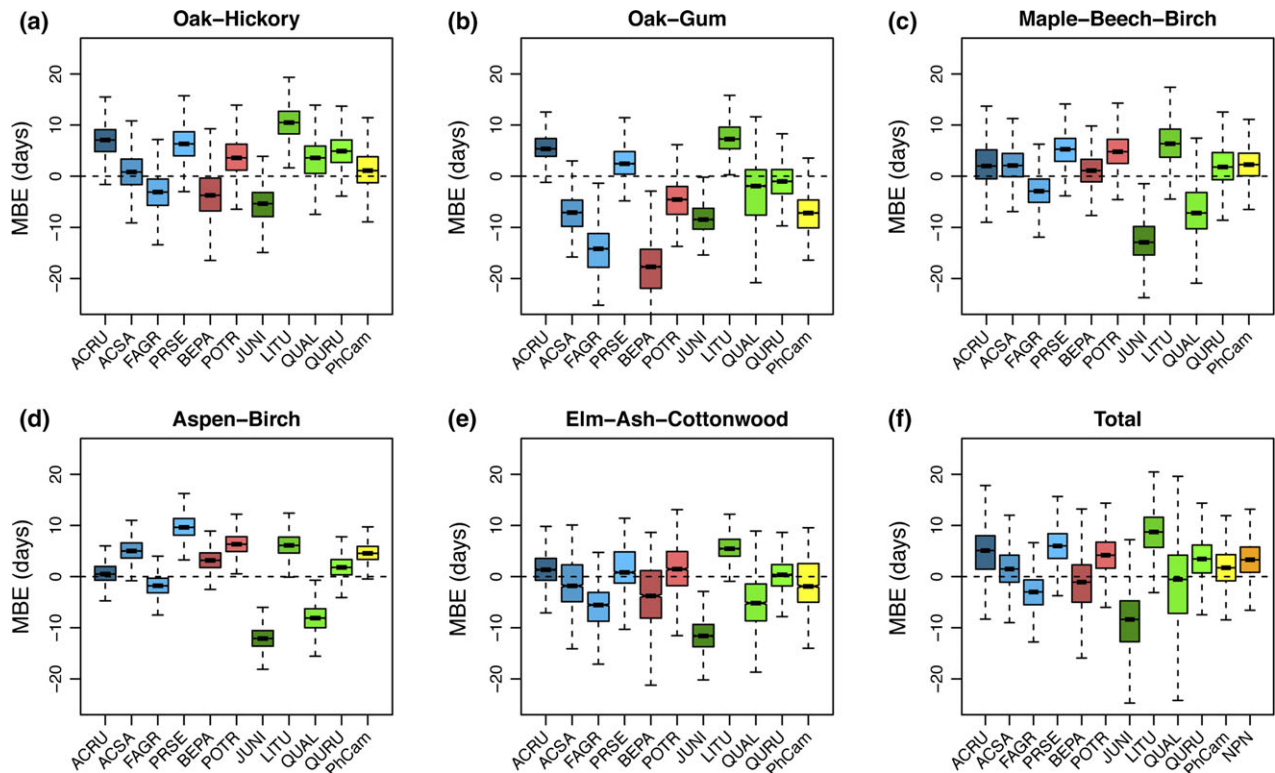
Fig. 4 Boxplots of mean bias error (days) between MODIS estimated spring onset and model predicted spring onset across (a–e) pixels dominated by each forest type, according to 25 km FIA hexagon maps (see Fig. 2) or (f) the entire study region.

timing of spring onset. Previous research using both ground and satellite remotely sensed data has suggested that models incorporating chilling tend to outperform models based solely on forcing requirements (Chiang & Brown, 2007; Kaduk & Los, 2011; Jeong *et al.,* 2012). However, these studies have consistently used arbitrary start dates with no specific biological basis (e.g., January 1) to accumulate forcing. Using a model that allows the start date to vary as a function of photoperiod, Migliavacca *et al.* (2012) found that a relatively simple two-parameter spring warming model outperformed more complex chilling models at Harvard Forest. Our results are consistent with those of Migliavacca *et al.*, as well as other previous experimental and modeling studies (e.g., Linkosalo *et al.,* 2000; Basler & Körner, 2012), and support the conclusion that photoperiod is an important factor that influences when and how thermal forcing triggers leaf emergence. More specifically, results from species-specific models tuned to USA-NPN data largely support the spring warming approach, with start dates for thermal accumulation centered around the spring equinox (~DOY 80, which corresponds to a 12-hour day length everywhere in the study region), but significant differences in forcing requirements among species and forest types. This result implies that winter chilling accumulation is

of secondary importance for most tree species in our study area. Indeed, our model results for 2010 and 2012 suggest that chilling requirements over the study region are being easily met, even in years when winter and early spring temperatures are exceptionally warm. However, it is important to note that as the magnitude of chilling requirements is currently unknown, it is unclear how much longer they will continue to be satisfied under future warming.

A second key result from this work is that relatively coarse spatial resolution information related to forest community composition (in this case, 25 km FIA data) explains meaningful geographic variation in the timing and ecological controls on springtime phenology. At the same time, it is important to note that our analysis includes several import assumptions and phenological datasets include nontrivial levels of uncertainty. For example, we assume that all 500 m MODIS pixels within each 25 km FIA hexagon have roughly uniform species composition. In some regions, this assumption is probably quite robust. In others, especially those with complex topography and recent disturbance, this assumption is weaker (e.g., Hwang *et al.,* 2011). Indeed, this may partly explain why our model predictions do not agree as well with MODIS observations in parts of Appalachia (but see next paragraph). Spatially explicit
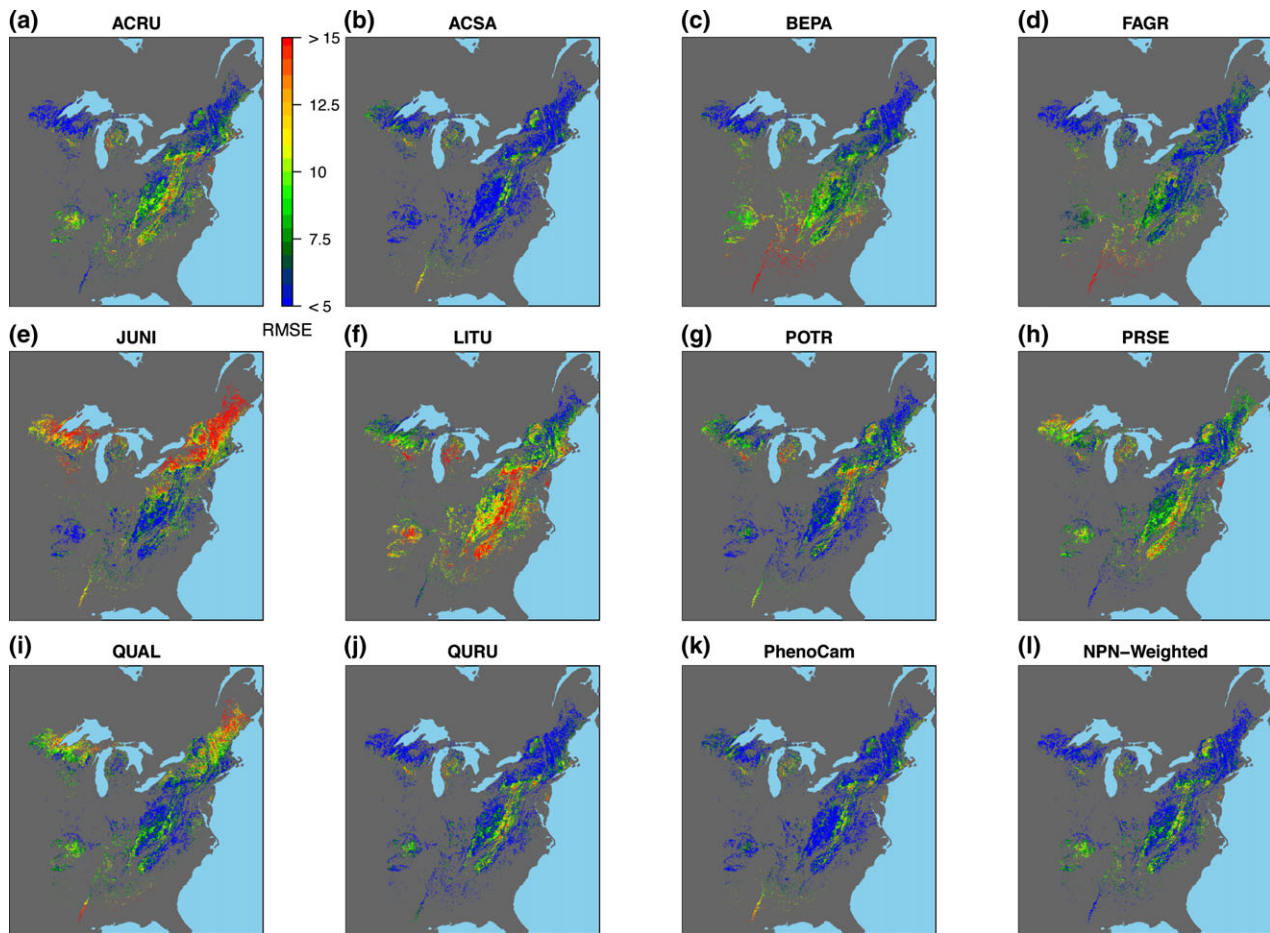
Fig. 5 Maps of root-mean-square error (days) between MODIS estimated spring onset and (a–j) each species-specific model, (k) Pheno-Cam model, and (l) forest type-weighted model predicted spring onset across the study region.

maps of tree species and forest types at higher spatial resolution (i.e., 500 m–1 km) may provide a basis for reducing this source of uncertainty (e.g., Zhu & Evans, 1994; Wilson *et al.*, 2012). Whatever the solution, a key conclusion from the results we report here is that forest composition is important to modeling large-scale phenology (i.e., plant functional types are not sufficient). Hence, new datasets and modeling approaches, be they explicit using species-weighted models or implicit using data from PhenoCams, are required in to realistically model phenology at the spatial resolutions required by large-scale ecosystem and LSMs.

A third key finding from our analysis is that predictions for spring onset from the model calibrated to PhenoCam data had errors that were approximately the same as errors from USA-NPN models tuned to individual species in the northern forest types, even though only 13 sites and 80 site-years were available to calibrate the model. This suggests that PhenoCam data provide a good basis for characterizing the seasonality of species mixtures. In southern areas of the study region dominated by oak-gum, however, PhenoCam model performance was poor and predictions showed significant bias toward later onset dates (Fig. 6k). We believe that this bias is at least partly attributable to the geographic distribution of PhenoCam sites. Specifically, none of the PhenoCam sites were located in southern forests, which probably caused the model to be overfit to northern species and sites. As more sites are added to the PhenoCam network in coming years, it will be important to re-estimate this model using a more geographically and ecological representative sample. Moreover, this result demonstrates that even though previous studies have shown that LSMs poorly represent spring phenology, the performance of these models is not necessarily a by-product of using generalized plant functional types. Rather, the strong performance of the species-neutral PhenoCam model suggests that relatively simple phenological models can be used to predict the timing of leaf emergence over large geographic ranges if they are properly parameterized and that species-level parameterizations
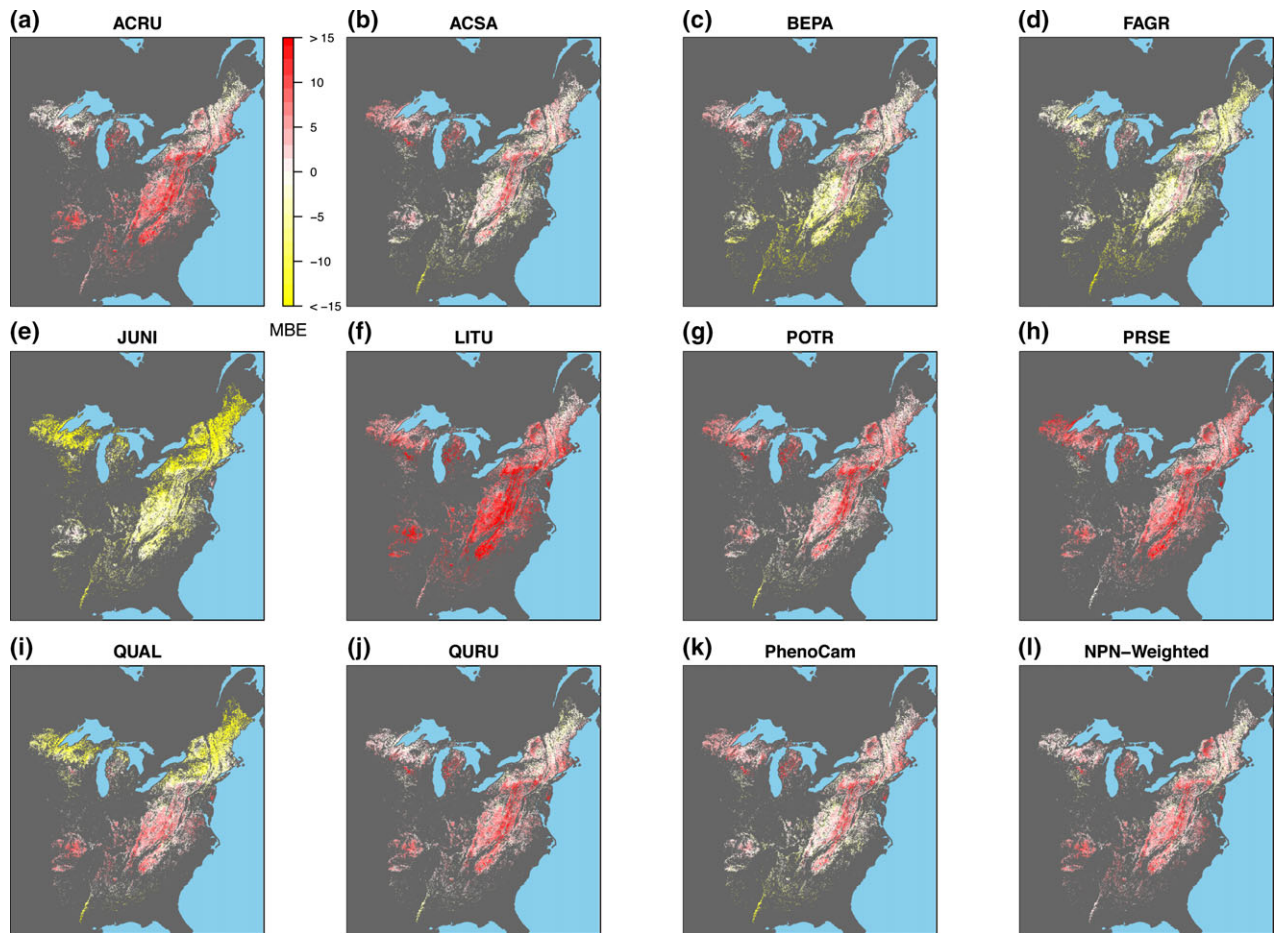
Fig. 6 Maps of mean bias error (days) between MODIS estimated spring onset and (a–j) each species-specific model, (k) PhenoCam model, and (l) forest type-weighted model predicted spring onset across the study region.

are not necessarily needed to improve existing schemes.

Finally, despite their relatively coarse spatial resolution (500 m), results from this analysis demonstrate that MODIS data have substantial utility for estimating and verifying results from upscaled phenology models. Moving forward, there is substantial basis for optimism that new methods and datasets based will soon be available that resolve phenology over large scales at spatial resolutions that are able to capture landscape-scale variation that is not currently resolved from instruments such as MODIS. Specifically, recent work by Elmore *et al.* (2012) and Melaas *et al.* (2013b) has demonstrated that time series of Landsat TM/ETM+ images can successfully estimate spring and autumn phenology at 30 m resolution and therefore have the potential to substantially mitigate this issue.

We used a combination of surface meteorological data, species composition maps, remote sensing, and ground-based observations of phenology to develop and test models that predict the timing of spring leaf emergence for dominant tree species across five different deciduous broadleaf forest types in the eastern United States. Our results identify significant differences in the cumulative forcing requirements and photoperiod cues among individual species, which lead to variation in the accuracy of spring onset model predictions across forest types. Most terrestrial ecosystem models, which rely on accurate predictions of phenophase transitions to simulate season variation in important ecological process, do not account for species-specific phenological dynamics.

To address this shortcoming, we evaluated two strategies for upscaling species-level dynamics to spatial scales commonly used by ecosystem models, which encompass significant variation in species composition and phenological dynamics. The first strategy used information related to the geographic distribution of dominant tree species to provide upscaled predictions for the timing of spring onset based on a linear weighting of species-specific models. The second strategy used time series observations of forest canopies from

repeat digital photography that include mixtures of tree species to calibrate a single model for the entire study region. Both strategies yielded robust predictions of spring onset across a majority of the study region and therefore provide a foundation for improving the representation of spring phenology in ecosystem models. More generally, the issues and strategies we examined in this paper provide the basis for more reliable forecasts of how the phenology of temperate forests will be affected by climate change, and by extension, how changes in climate will impact temperate forests in the coming decades.

Acknowledgements

References

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.

Basler D, Körner C (2012) Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology*, **165**, 73–81.

Betancourt JL, Schwartz MD, Breshears DD *et al.* (2007) Evolving plans for the USA National Phenology Network. *Eos, Transactions American Geophysical Union*, **88**, 211–211.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd edn). Springer, New York.

Chiang JM, Brown KJ (2007) Improving the budburst phenology subroutine in the forest carbon model PnET. *Ecological Modelling*, **205**, 515–526.

Chuine I, Cour P, Rousseau DD (1998) Fitting models predicting dates of flowering of temperate-zone trees using simulated annealing. *Plant Cell and Environment*, **21**, 455–466.

Chuine I, Morin X, Bugmann H (2010) Warming, photoperiods, and tree phenology. *Science*, **329**, 277–278.

Denny EG, Gerst KL, Miller-Rushing AJ *et al.* (2014) Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. *International Journal of Biometeorology*, **58**, 591–601.

Elmore AJ, Guinn SM, Minsley BJ, Richardson AD (2012) Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. *Global Change Biology*, **18**, 656–674.

Fisher JI, Richardson AD, Mustard JF (2007) Phenology model from surface meteorology does not capture satellite-based greenup estimations. *Global Change Biology*, **13**, 707–721.

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

Franks SW, Beven KJ, Gash JHC (1999) Multi-objective conditioning of a simple SVAT model. *Hydrology and Earth System Sciences*, **3**, 477–489.

Friedl MA, Gray JM, Melaas EK *et al.* (2014) A tale of two springs: using recent climate anomalies to characterize the sensitivity of temperate forest phenology to climate change. *Environmental Research Letters*, **9**, 054006.

Fry JA, Xian G, Jin SM *et al.* (2011) National land cover database for the conterminous United Sates. *Photogrammetric Engineering and Remote Sensing*, **77**, 859–864.

Ganguly S, Friedl MA, Tan B, Zhang XY, Verma M (2010) Land surface phenology from MODIS: characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, **114**, 1805–1816.

Gerst KL, Kellermann JL, Enquist CA, Rosemartin AH, Denny EG (2015) Estimating the onset of spring from a complex phenology database: trade-offs across geographic scales. *International Journal of Biometeorology*, 1–10.

Gu L, Hanson PJ, Post WM *et al.* (2008) The 2007 eastern US spring freezes: increased cold damage in a warming world? *BioScience*, **58**, 253–262.

Homer C, Dewitz J, Fry J *et al.* (2007) Completion of the 2001 national land cover database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **73**, 337–341.

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195–213.

Hufkens K, Friedl M, Sonnentag O, Braswell BH, Milliman T, Richardson AD (2012) Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. *Remote Sensing of Environment*, **117**, 307–321.

Hwang T, Song CH, Vose JM, Band LE (2011) Topography-mediated controls on local vegetation phenology estimated from MODIS vegetation index. *Landscape Ecology*, **26**, 541–556.

Jenkins JP, Braswell BH, Frolking SE, Aber JD (2002) Detecting and predicting spatial and interannual patterns of temperate forest springtime phenology in the eastern US. *Geophysical Research Letters*, **29**, 54-1–54-4.

Jeong SJ, Medvigy D, Shevliakova E, Malyshev S (2012) Uncertainties in terrestrial carbon budgets related to spring phenology. *Journal of Geophysical Research-Biogeosciences*, **117.G1**, 1–17.

Kaduk JD, Los SO (2011) Predicting the time of green up in temperate and boreal biomes. *Climatic Change*, **107**, 277–304.

Keenan TF, Gray J, Friedl MA, Toomey M, Bohrer G, Hollinger DY, Richardson AD (2014) Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change*, **4**, 598–604.

Klosterman ST, Hufkens K, Gray JM *et al.* (2014) Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences*, **11**, 1–16.

Korner C, Basler D (2010) Phenology under global warming. *Science*, **327**, 1461–1462.

Lechowicz MJ (1984) Why do temperate deciduous trees leaf out at different times - adaptation and ecology of forest communities. *American Naturalist*, **124**, 821–842.

Linkosalo T, Carter TR, Häkkinen R, Hari P (2000) Predicting spring phenology and frost damage risk of *Betula* spp. under climatic warming: a comparison of two models. *Tree Physiology*, **20**, 1175–1182.

Linkosalo T, Lappalainen HK, Hari P (2008) A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations. *Tree Physiology*, **28**, 1873–1882.

Melaas EK, Friedl MA, Zhu Z (2013a) Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM plus data. *Remote Sensing of Environment*, **132**, 176–185.

Melaas EK, Richardson AD, Friedl MA *et al.* (2013b) Using FLUXNET data to improve models of springtime vegetation activity onset in forest ecosystems. *Agricultural and Forest Meteorology*, **171**, 46–56.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**, 1087–1093.

Migliavacca M, Sonnentag O, Keenan TF, Cescatti A, O'keefe J, Richardson AD (2012) On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. *Biogeosciences*, **9**, 2063–2083.

Polgar CA, Primack RB (2011) Leaf-out phenology of temperate woody plants: from trees to ecosystems. *New Phytologist*, **191**, 926–941.

Raulier F, Bernier PY (2000) Predicting the date of leaf emergence for sugar maple across its native range. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **30**, 1429–1435.

Richardson AD, Bailey AS, Denny EG, Martin CW, O'Keefe J (2006) Phenology of a northern hardwood forest. *Global Change Biology*, **12**, 1174–1188.

Richardson AD, Jenkins JP, Braswell BH, Hollinger DY, Ollinger SY, Smith ML (2007) Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*, **152**, 323–334.

Richardson AD, Black TA, Ciais P *et al.* (2010) Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 3227–3246.

Richardson AD, Anderson RS, Arain MA *et al.* (2012) Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis. *Global Change Biology*, **18**, 566–584.

Richardson AD, Keenan TF, Migliavacca M, Ryu Y, Sonnentag O, Toomey M (2013) Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, **169**, 156–173.

Sarvas R (1974) Investigations on the annual cycle of development of forest trees, Autumn dormancy and winter dormancy. *Communicationes Instituti Forestalis Fenniae*, **84**, 1–101.

Schaaf CB, Gao F, Strahler AH *et al.* (2002) First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, **83**, 135–148.

Sonnentag O, Hufkens K, Teshera-Sterne C *et al.* (2012) Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, **152**, 159–177.

Thornton PE, Thornton MM, Mayer BW *et al.* (2014) *Daymet: Daily Surface Weather Data on a 1-km Grid for North America*, version 2. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN.

Vitasse Y, Delzon S, Dufrêne E, Pontailler JY, Louvet JM, Kremer A, Michalet R (2009) Leaf phenology sensitivity to temperature in European trees: do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology*, **149**, 735–744.

Wang J, Ives NE, Lechowicz MJ (1992) The relation of foliar phenology to xylem embolism in trees. *Functional Ecology*, **6**, 469–475.

Williams CN, Menne MJ, Vose S, Easterling DR (2006) *United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data*. ORNL/CDIAC-118, NDP-070. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN.

Wilson BT, Lister AJ, Riemann RI (2012) A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. *Forest Ecology and Management*, **271**, 182–198.

Wilson BT, Lister AJ, Riemann RI, Griffith DM (2013) *Live Tree Species Basal Area of the Contiguous United States (2000–2009)*. USDA Forest Service, Northern Research Station, Newtown Square, PA.

Yang X, Mustard JF, Tang JW, Xu H (2012) Regional-scale phenology modeling based on meteorological records and remote sensing observations. *Journal of Geophysical Research-Biogeosciences*, **117.G3**, 1–18.

Zhu ZI, Evans DL (1994) United-States forest types and predicted percent forest cover from AVHRR data. *Photogrammetric Engineering and Remote Sensing*, **60**, 525–531.

Supporting Information

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

Figure S1. Boxplots of Pearson correlation coefficients between MODIS estimated spring onset and model predicted spring onset.
Figure S2. Boxplots of model slope.
Table S1. Summary of model parameters.
Table S2. Optimal parameters for each species based on out-of-sample testing results.
Table S3. ΔAIC_c (difference between a particular model's AIC_c and the lowest AIC_c across all candidate models for a species) values for models fit to species-specific NPN or PhenoCam data.
Data S1. Description of phenology models.