

A novel model to accurately predict continental-scale timing of forest green-up



N. Neupane ^{a,*}, M. Peruzzi ^b, A. Arab ^a, S.J. Mayor ^c, J.C. Withey ^d, L. Ries ^{a,1}, A.O. Finley ^{e,1}

^a Georgetown University, Washington DC, USA
^b Duke University, NC, USA
^c Ontario Forest Research Institute, Canada
^d The Evergreen State College, WA, USA
^e Michigan State University, MI, USA

ARTICLE INFO

Keywords:
Spatio-temporal modeling
Bayesian regression
Greenup
Growing-degree-day
Gaussian processes

ABSTRACT

The yearly cycles in vegetation greenness are among the most important drivers of ecosystem processes. Predictive models for the timing of vegetation greenup and senescence are crucial for understanding how biological communities respond to global change. Greenup timing is closely tied to climate and also tracks yearly variability in temperature, and the strength of this relationship varies spatio-temporally. Local studies have been useful in understanding underlying mechanisms but they are insufficient in explaining larger scale variabilities. Large-scale studies using remotely-sensed data have the potential to harness regional dynamics, even if underlying mechanisms remain unknown. Yet predictive power using these approaches is low. Here, we predict vegetation phenology across Eastern North America via a novel class of Bayesian regression model. Our modeling framework provides continental-level peak greenup time predictions with high accuracy using satellite observations from the MODerate resolution Imaging Spectroradiometer (MODIS). In addition to taking into account temporal structure at individual sites, our models make use of information from the entire study extent regardless of their spatial proximity.

Models were built from 2000 to 2016 and showed high prediction accuracy ($R^2 > 95\%$). Out-of-sample predictions for the years 2017 and 2018 showed accuracy within days of the predicted peaks, even though yearly greenup timing can vary by up to 30 days across the study region. Performance was remarkably high across deciduous and mixed forest types. Our method is generalizable to temperate forests across the globe and provides a basis for backcasting and forecasting forest greenup for any time periods where daily temperatures, whether directly measured or modeled, are available.

1. Introduction

The sensitivity of vegetation phenology to climate change has broad implications for many aspects of ecological functions and ecosystem services (Schröter et al., 2005; Lafferty, 2009). Green-up timing is a key determinant of growing season length, so it is vital to productivity in both natural and agricultural systems. Warming temperatures and longer growing seasons may impact insect pest phenologies, sometimes by adding additional generations, which can lead to outbreaks (IPCC, 2007; Altermatt, 2010; Musolin, 2007; Bale et al., 2002). Local studies have shown that when plant phenologies change, it can impact trophic relationships and contribute to phenological mismatches between

plants, herbivores, and predators or between mutualistic species such as plants and pollinators (Kudo and Ida, 2013). Despite its importance, vegetation phenology is rarely well represented by terrestrial biosphere models (Richardson et al., 2012) and reliable forecasting of phenology with climate change has remained elusive. This has hindered research into large-scale processes that rely on phenological information. Here, we develop a model for forest green-up at continental scales that is implemented and parameterized for a particular region. The specific model parameterization can then be used to either forecast phenology dynamics into the future based on global climate models or, using historical climate data, hindcast phenology estimates into past time periods before satellite imagery is available. In either case, this modeling

* Corresponding author.

E-mail address: nn343@georgetown.edu (N. Neupane).

¹ Equal contribution as senior authors.

framework vastly increases the temporal scope at which we can carry out continental-scale studies that include phenological dynamics based on vegetation greenup timing.

Altered vegetation phenologies have both direct and indirect ecological effects: directly, the vegetation surfaces provide food and shelter for herbivores, whereas indirectly it affects predators such as leaf-gleaning birds or mutualistic partners such as ants (Fleming et al., 2021). Timing can also result in phenological mismatches that can ripple through trophic relationships and are thus hypothesized to impact the survival and breeding of many species. This concern has been especially high for migratory species (Both et al., 2006; Hostetler et al., 2015). For instance, many migratory birds (Faaborg et al., 2010) arrive at their breeding grounds in North America after overwintering further south. Such species appear to follow peak vegetation greenness along their migratory pathways to fulfill their food and shelter requirements (La Sorte et al., 2014a,b). Recent reviews also show that the warmer temperatures lead to earlier onset of phenological events for many species (see Gallinat et al., 2021). Such mismatches may be one factor that underlies biodiversity loss in the present day (Both et al., 2006; Johnson et al., 2011) and may continue in the future under various global warming conditions and associated uncertainties (Cleland et al., 2007; Sekercioglu et al., 2008; Wiens, 2016; Neupane et al. 2022).

As disturbances become increasingly global in nature, it is critical to explore the consequences of these shifting dynamics at multiple scales: local, regional, and continental (Heffernan et al., 2014; Youngflesh et al., 2021). For large-scale studies, satellite-based peak greenness timing products are the most tractable data source for studying phenological mismatch; these provide an observable proxy for resource availability across trophic systems. This timing correlates with the maximum abundance of insects and caterpillars that provide food for many secondary consumers in forest systems (Valtonen et al., 2013; Richards and Windsor, 2007). Thus, pinpointing the factors that most accurately predict peak greenness timing at macro-level scales is important to explore trophic dynamics that are linked to phenology (Fu et al., 2014a; Mayor et al., 2017). In Eastern North America, forest greening begins in April in the south and progresses northward into Canada over the late spring and summer (Fig. 1b), where migrant birds, bats, and also insects (e.g., the iconic monarch butterfly, *Danaus plexippus* and important pests such as the armyworm, *Spodoptera frugiperda*) arrive after wintering in the south. Further, there is substantial variation in onset timing (Fig. 2), which differs across space and time. This presents a more challenging landscape when species must adapt their own developmental or migratory timing to match.

In Eastern North American forests, variability in the peak onset of greenness can be as much as 30 days in either direction, although more typically on the order of one to two weeks (Fig. 3). This magnitude of annual variability in observed onset dates has been reported in past studies, and is a function of both true environmental variability and

methodological noise in the data products (Jenerette et al., 2010; Soudani et al., 2008; Tan et al., 2010; Peng et al., 2017; Stanimirova et al., 2019; Xie et al., 2015; Dunn and de Beurs, 2011).

While all the mechanisms that control variability in peak greenness have not yet been fully resolved, it is well known that annual variability in seasonal temperatures is a primary driver (Richardson et al., 2013; Wang et al., 2015; Zhang et al., 2007). However, the evidence on what component of temperature (e.g., minima, means or maxima) is the most critical is mixed. Some studies have shown that vegetation greening is largely dictated by daily maximum, rather than minimum or mean temperatures, especially in the northern hemisphere (Fu et al., 2016; Peng et al., 2013). Yet, other studies have shown that elevated minimum temperatures are associated with increased plant productivity and earlier onset of greening (Myneni et al., 1997; Shen et al., 2016). Rather than focus on average temperature metrics, we use a growing degree day (GDD) model where the accumulation of measured degrees of temperature serves as a proxy of energy available for plant development. GDD models consist of fixing a minimum threshold temperature (the “base” temperature, T_0) and accumulating daily temperature values exceeding that threshold (Morrison et al., 1989; Man and Lu, 2010). For instance, if the base temperature (T_0) is 0°C and the day's average temperature is 20°C, that day 20 GDD would be accumulated. Focal species or systems are thus modeled to accumulate heat (measured in GDD), with specific milestones (in the case of plants, budburst, timing of leaf out and flowering) predicted to occur when GDD accumulation reaches a critical value (k) which varies depending on the species or forest type (Leinonen and Kramer, 2002; Kramer et al., 2000; Chuine, 2000). This approach to measuring temperature was developed in agricultural sciences to allow for appropriate crop choices and timing of planting, harvest and pest control (Diffenbaugh et al., 2008; McMaster and Wilhelm, 1997). More recently, it has been applied to natural systems to understand a wider range of phenological events, such as timing of the budburst, green-up, and flowering season for plants and patterns of abundance for insects and other ectotherms (Aurambout et al., 2009; Cayton et al., 2015; Abarca et al., 2018).

The GDD model is widely recognized for its ability of predicting green-up with models performing well at local scales. GDD models have also been deployed at larger spatial scales in temperate forests (Botta et al., 2000; Yang et al., 2012; Jeong et al., 2012; Cong et al., 2013; Fu et al., 2014b; Piao et al., 2015; Li et al., 2021), including the eastern North America (Klosterman et al., 2018), however, model accuracy has generally been low.

One reason for the reduced performance at large scales, is that researchers are not making predictions for individual species (for which these models are customized) but tracking the greening of entire ecosystems that vary in dominant species across space and time. GDD models vary in critical values such as T_0 and k based on species and even across varietals (Chung et al., 2011). These development models may

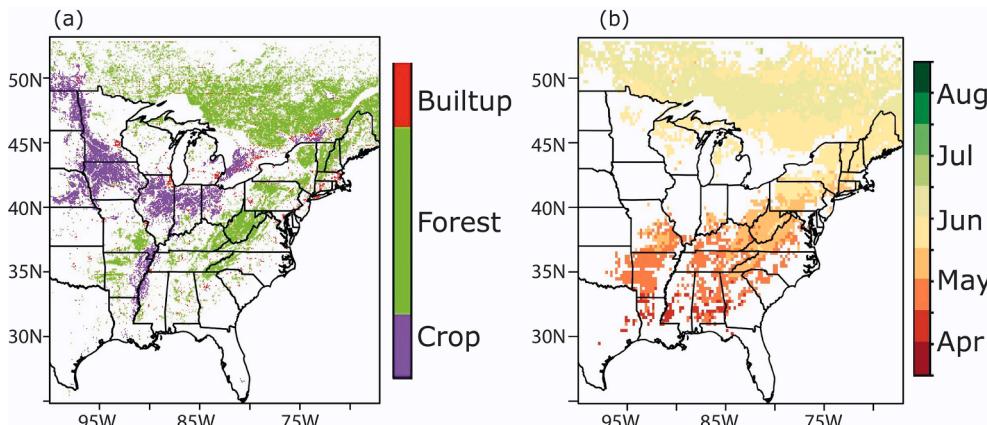


Fig. 1. Land use/land cover classification from the National Land Use Land Cover dataset for 2011, coarsened to align with the Livneh climate product (6-km) (a). Pixels that are dominated by a single cover type (>75%) are shaded: forest (evergreen, deciduous or any mix of the two types), crop, and builtup (cities and suburbs). Unclassified pixels, water bodies, grasslands, and bare grounds are not shown. Peak onset of greenness (POG) timing is shown for forest pixels only (b) and is based on mean values from 2000 to 2016 MODIS satellite observations.

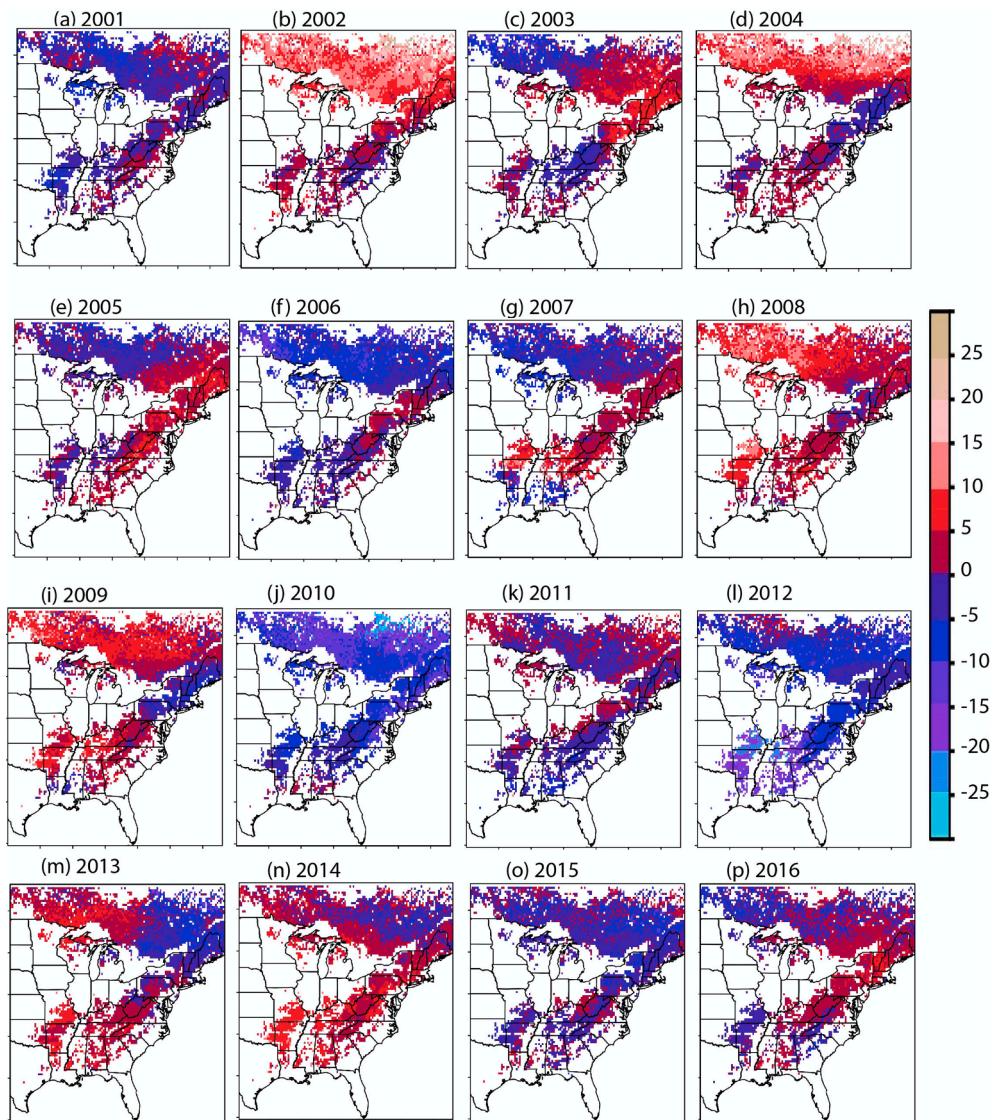


Fig. 2. Differences in the peak onset of greenness (POG) timing from the 2001 to 2016 mean MODIS satellite observations.

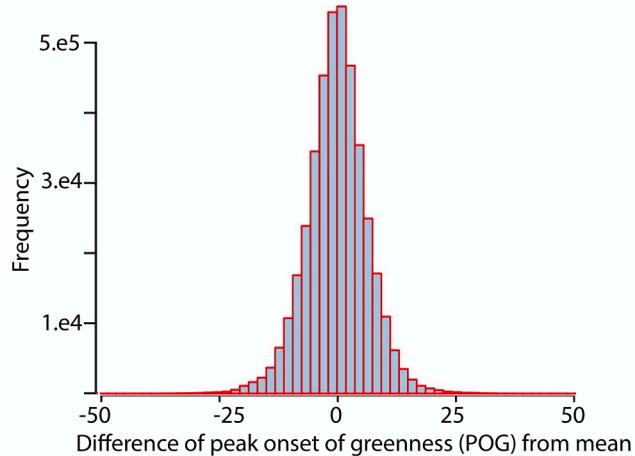


Fig. 3. Histogram of MODIS satellite observed peak onset of greenness (POG) timing differences in annual from the 2001 to 2016 mean (annual MODIS observations - climatological mean). Values outside of ± 50 are truncated.

include additional predictor variables, such as chilling degrees, where low temperatures are accumulated (Harrington et al., 2010; Xie et al., 2015), photoperiod requirements (Way and Montgomery, 2015), or other climate covariates such as precipitation (Laube et al., 2014; Fu et al., 2014b). When making predictions across plant communities rather than for individual species, there is not a single mechanistic model that will work because different species are responding to different environmental factors. However, because the accumulation of heat is a critical component for all species across deciduous, temperate forests (e.g., Wang et al., 2020), including those in eastern North America, model simplicity has led many researchers often use GDD as the sole model predictor (Kim and Wang, 2005; Yang et al., 2012). For instance, several studies have examined the accuracy of GDD model predictions of greenness onset. Xie et al. (2015) tested several GDD models with T_0 values between 4 and 6 °C to predict the Eastern North America greenness onset from satellite imagery and found that a minimum threshold of 6 °C led to the best predictions of onset dates. Fu et al. (2014a) compared various phenological models, the GDD model predictions explained most (67–79%) variability in Eastern North America, but this is still relatively low and limits the ability to make useful projections into time periods with no satellite imagery. Low accuracy is due to substantial spatial structure in performance (Fig. 2) that are likely

keyed to local plant community structure and physical geography.

Our goal is to increase prediction accuracy for phenological greening, using temperate, deciduous forests in Eastern North America (100°W - 67°W ; 25°N - 52°N) as a case study, so that phenological dynamics in those systems can be considered at continental scales and over longer time periods. In addition to potential improvements in the biosphere models that feed into global climate models (Richardson et al., 2012), this will allow the incorporation of phenological timing into broad-scale models of trophic interactions. Trophic dynamics, especially the potential for trophic mismatches under altered climate has received increasing attention in the ecological literature (Visser and Gienapp, 2019). The main tool we introduce to better characterize dynamics of green-up onset is a spatially explicit regression model with latent random effects modeled as gaussian processes (GPs), rather than a more typical implementation where random effects are modeled individually at each location. Indeed, incorporating random effects allows us to vary the model based on local dynamics without understanding the underlying biological mechanism, yet still achieve higher predictability. Modeling random effects using GPs allows us to vastly increase the spatial scope of our studies. This flexible methodology has been developed in the field of geostatistics and its use is widespread today in diverse areas such as biostatistics and machine learning (Gelfand et al., 2010; Banerjee et al., 2014; Rasmussen and Williams, 2006). Due to poor scalability of the GPs, the computational demand is substantial for large scale data and here we overcome that challenge by implementing a recently introduced methodology, “meshed” GPs (MGP, Peruzzi et al., 2020). This allows for scalable inference of random effects at continental scales, which are associated with superior predictive performance for massive multi-dimensional spatial and temporal data arising from satellite imaging. This advancement will allow macro-scale ecologists to project green-up timing into the future using climate models or into the past wherever daily temperature data are available. For instance, daily temperature products currently exist back to 1915 for the USA (Livneh et al., 2013) extending the ability to examine the potential of past impacts of phenology shift on ecological communities. Furthermore, global climate models also project daily temperatures into the future under multiple emissions scenarios (Stocker, 2014). This model thus greatly expands the temporal scope for which we can explore the impacts of phenology and phenological mismatch across multiple trophic levels both forward and backward.

2. Data sources

Our basic model predicts the day of the year when maximum greening is occurring (we call this “peak onset of greenness”, or POG) from satellite observations based on GDD accumulation calculated from daily minimum and maximum temperature values. Environmental datasets are available at various grains and extents. For consistency, we interpolated all the data onto a base Livneh grid (6-km resolution), expanding its extent into Canada up to 52°N . We selected the Livneh grid because this was the coarsest gridded data among all the datasets that we used here and also because it represents the longest-duration daily temperature data across the entire USA and will allow this model to be hindcasted as far back as 1915 (Livneh et al., 2013). Both greenness (see Section 2.1) and temperature (see Section 2.2) were linearly interpolated and harmonized to match the Livneh data set. Our specific study extent is in eastern North America, 100°W - 67°W ; 25°N - 52°N , including regions in both the USA and Canada (Fig. 1a, b). We generated our base grid (from the Livneh grid) which resulted in 528 divisions along longitude and 444 divisions along latitude resulting in a coverage including 234,432 pixels. This region is largely dominated by the temperate forest biomes (Fig. 1a), with generally warm and humid summers and mild to cold winters. The more northern regions (up to 52°N) have the most pronounced greening seasonal cycle because they have the largest seasonal changes in temperature. The southern forest, south of 33°N , towards the Gulf of Mexico is characterized by evergreen

forest and this biome has only a marginal leaf seasonality. We include them in our analysis to determine if our methods are able to capture these more subtle seasonal changes.

2.1. MODIS greenup data

Green-up data are obtained from the MODIS (MODerate Resolution Imaging Spectroradiometer, combined Aqua/Terra) satellite. This data has been developed into a product that tracks phenology. The MODIS Land Cover Dynamics (MCD12Q2, v. 006) product is available since 2001 and it is used to estimate the day of the year when 50% and 100% greening is achieved in each year at 500-m spatial resolution (Friedl et al., 2019). In this product, peak greenness (100%) indicates when greenness reaches the highest level of saturation and signifies a time of year when foliage reaches both its maximum leaf area but at a time when the leaves are still relatively young. We use the day-of-year value when 50% of the peak greenness is reached and define this as our peak onset of greenness (POG), the independent variable in our model. Thus, our predictions are being trained to determine the single day of each year in each pixel when this phenological milestone, POG, is reached. We were able to obtain data through 2018 for this project since data are made available with a 1–2 year lag. We used 2001–2016 data for training and tested our predictions on 2017 and 2018 data. This means, these two years are not included in the model building.

2.2. Land cover

Because our study region includes regions of both the USA and Canada, we used the data with the finest spatial resolution for each country and then harmonized them both to our full study extent. For the USA, we used land cover data from the National Land Cover Database (NLCD, Yang et al., 2018). This is available for the conterminous United States at 30-m spatial resolution for a subset of years from 2001 to 2016. There are altogether 19 land cover types, including the three forest types, deciduous, evergreen, and mixed (Fig. 1a). For Canada, we used the North American Land Change Monitoring System (NALCMS). This product is also available at 30-m spatial resolution for 2005, 2010, and 2015 (Latifovic et al., 2017). To match the period of land cover data available for both USA and Canada, we selected 2011 NLCD data. 2011 is selected because it is around the middle of our study period (2001–2018) and also because the changes in the land use/ forest, especially in the region east of 100°W was relatively low during our study period (Homer et al., 2015; Homer et al., 2016). For this reason, we built our model only with 2011 data, meaning that we do not account for changing land cover type as a variable in our model.

We extracted the land cover types for each pixel by creating a circle with a 2-km radius centered at each pixel centroid. This also allowed us to ensure that distortion over latitudes would not exceed pixel size and thus avoid the problems of overlapping buffers at more northerly latitudes. In our target study regions, we summarized the type of each of the 30-m land cover pixels within the 2-km radius from centroid of each pixel of our base (Livneh) grid. In order to avoid issues of heterogeneous habitat, we only use pixels where $>75\%$ of land cover is forested (Fig. 1 shows selected pixels). Similar threshold approach has also been used in other studies (Xie and Wilson, 2020) and preliminary analyses (not shown) suggested that grids with $<75\%$ forest showed reduced predictive performance. Then, we classified each pixel as the dominant forest type. To calculate our forest pixel type in each 2-km radius buffered circle, we summarized the number of each forest pixel type from the land classification data sets and used the following rules to classify them as deciduous, evergreen or mixed at our coarser grain: (1) calculate the percentage of land cover for all cover types that fall within the enclosed circle (2) if the cover type is $>75\%$ “Deciduous” or “Evergreen” forest, then, we classify those as the dominant forest type, but if the cover type is $< 75\%$ forest but not dominated by either Deciduous or Evergreen, then, we classify those pixels as “Mixed” forest. We also excluded North

America west of 100°W because in that region, precipitation becomes far more important for predicting green-up and thus models based on simple GDD are no longer appropriate (Wang et al., 2003).

Throughout our study region, every year, there are a total of 27,178 pixels with >75% forest cover (Fig. 1a), with 17% predominantly Deciduous, 6% predominantly Evergreen, and 77% a mix of the two. These forest classification types were also included in the model as categorical variables.

2.3. GDD accumulation

GDD values must be calculated from daily temperature values. We used the Daymet daily minimum and maximum temperatures, with a base temperature (T_0) of -5°C . To accumulate GDD, starting on January 1st of each year (as in Fu et al., 2014a), we accumulated all temperature values over -5°C . As a result, each pixel in each year is associated with a sequence of 365 GDD values (366 in a leap year). Rather than choosing a fixed estimate of the critical accumulation threshold ‘ k ’ for green-up, we selected 19 GDD threshold values, from 800 to 2600 at intervals of 100 GDD, where k is 800 GDD, 900 GDD...2600 GDD. Then, we extract the day of the year when each of these threshold values is reached. For example, in the year 2001 for $k = 800$, we accumulated the daily GDD value in every pixel from January 1st through the day of the year when it became equal to 800 GDD. We extract this day of the year as a predictor variable for our model. Then we do the same for the remaining 18 values of k (900, 1000, ..., 2600). Therefore, each pixel has 19 predictors’ values corresponding to the day of year when each of 19 critical thresholds (k) is reached. This range (800 to 2600 GDD) of k values was selected because preliminary analyses showed that GDD accumulation was generally within these limits when peak onset occurred in all pixels throughout our study region. Further, no study has specifically estimated our target milestone (the Fu et al., 2014a estimate of $k = 591$ was for the beginning of onset). In addition, there are different forest types in our study region and using one value of k for the entire domain could result in low predictive value. By allowing all 19 values of k , we can build more accurate models and also leverage the values of ‘ k ’ that are most relevant to the structure of the tree community by region. We are also allowing the model to vary regionally via random processes that are modeled using a Gaussian distribution. Note that this means that each individual pixel (each unique location in our study region), is attributed with a vector of values of k (GDD threshold) of length 19 for each year and with a random effect as well as a factor indicating forest type. Each of these threshold values of k becomes a predictor for our regression model. These 19 variables are certain to be highly correlated but we account for that by our modeling approach (see below). We are also careful not to ascribe any biological meaning to the parameter estimates, our only goal is to build a highly predictive model.

2.4. Model setup

2.4.1. Modeling spatiotemporal dynamics

We consider a regression model to link our outcome variable MODIS peak onset of greenness (POG) to forest types and the GDD predictors listed previously (19 critical GDD thresholds). The simplest and most common regression model to consider first is ordinary least squares or linear regression (LR). If we label the outcome variable, peak onset, for each pixel i as y_i , and the corresponding predictor vector as $x_i = (x_{i,1}, \dots, x_{i,p})$ the conditional distribution of y_i given the predictors is denoted by $p(y_i|x_i)$. The linear regression model assumes that the conditional expectation of the outcome given the predictors is $E[y_i|\beta, x_i] = x_i^\top \beta = \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}$, and the conditional variance $\text{var}[y_i|x_i] = \tau^2$ for all i ; a Gaussian distributional assumption leads to $p(y_i|x_{i,1}, \dots, x_{i,p}) = N(y_i; x_i^\top \beta, \tau^2)$, where $N(y; m, v^2)$ denotes the Gaussian distribution with mean m and variance v^2 evaluated at y .

When a sample of n units (pixels) is collected, the joint likelihood is

the product of n densities $L(\{y_i\}_{i=1}^n) = \prod_{i=1}^n N(y_i; x_i^\top \beta, \tau^2)$. Using vector notation, we denote $y = (y_1, \dots, y_n)^\top$, $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\top$, and $X = (x_1^\top, \dots, x_n^\top)^\top$ as the $n \times p$ matrix of predictors; then, the LR model can be equivalently written as $y = X\beta + \epsilon$ where $\epsilon \sim MVN_n(\mathbf{0}, \tau^2 I_n)$ where I_n is the identity matrix and $MVN_j(\mathbf{m}, V)$ is the j -dimensional multivariate Gaussian distribution with mean \mathbf{m} and covariance matrix V . We take a Bayesian perspective to estimate the unknown parameters β and τ^2 ; this entails representing our uncertainty about β and τ^2 by introducing *prior* distributions $\pi(\beta)$, and $\pi(\tau^2)$. We then obtain the *posterior* distribution of the unknowns, given the observed data, by applying Bayes’ rule to update our prior uncertainty. This LR model, however, is unable to account for spatial or temporal dynamics. The error term $[\epsilon_i = y_i - x_i^\top \beta]$ is assumed to be independent and identically distributed as $N(0, \tau^2)$ random variables. Therefore, a linear regression model corresponds to assuming a linear, static relationship between all predictors and the outcome. In particular, the assumption that random variables are independent and identically distributed (*i.i.d.* assumption) on the measurement error implies that all spatial variability in the outcome must be explained by the regressors. However, such model will fail to capture spatial and temporal variability not measured by the regressors. Understanding this, we develop a spatially varying random intercept model that takes care of such spatio-temporal dynamics.

Our first step in developing a spatio-temporal regression model is the introduction of a spatial and time domain, labeled as $\mathcal{D} \subset \mathbb{R}^2 \times \mathbb{N}$. The spatial coordinates are denoted as $s \in \mathbb{R}^2$ (i.e. longitude and latitude, or easting and northing) whereas time is $t \in \mathbb{N}$. A generic space-time location is thus $\ell = (s, t) \in \mathcal{D}$. Each observational unit $i = 1, \dots, n$ will then be indexed by specific location, i.e. we write $y_i = y(\ell_i)$ as the spatiotemporally-referenced outcome at location ℓ_i . Similarly, $x(\ell_i)$ denotes the row-vector of predictors collected at location ℓ_i .

We model space-time dynamics by introducing a random effect $w(\ell_i)$, stacked into the w vector, which refers to the portion of $y(\ell_i)$ not explained by the predictors, but also not due to measurement error. We may interpret it as the *unobserved* purely spatial and/or temporal effect. We then assume that w arises from a Gaussian Process (GP) with covariance function $C(\cdot, \cdot, \theta) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$. A GP is a prior distribution over functions which leads to tractable inference within a Bayesian paradigm. In spatial settings (with dimension 2), a GP models spatial surfaces and in spatiotemporal settings, it models the continuous temporal dynamics of a spatial surface. The covariance function used to build the GP is also commonly referred to as *kernel*, and controls the smoothness of the surfaces. $C(\cdot, \cdot)$ is a positive-definite function indexed by unknown parameters θ . When the domain includes time, a *separable* covariance function would allow one to write C as the product of spatial-only and time-only covariances, i.e. $C(\ell, \ell') = C_{\text{space}}(s, s')C_{\text{time}}(t, t')$, but this assumption may be restrictive in not allowing for modeling possible interactions. For this reason, we will consider *nonseparable* space-time covariance functions (Gneiting and Guttorp, 2010). Therefore, writing $w(\cdot) \sim GP(\mathbf{0}, C(\cdot, \cdot))$ implies that for any set of locations $\mathcal{T} = \{\ell_1, \dots, \ell_n\}$ it holds that $w = (w(\ell_1), \dots, w(\ell_n))^\top \sim MVN_n(\mathbf{0}, C_{\mathcal{T}})$, where $C_{\mathcal{T}}$ is the $n \times n$ matrix whose (i, j) entry is $C(\ell_i, \ell_j; \theta)$. Then, we write our more general regression model at each location as

$$y(\ell) = \sum_{j=1}^p \beta_j x_{i,j}(\ell) + w(\ell) + \epsilon(\ell), \quad (1)$$

where the symbols in this case represent the variables as indicated within the brackets: y (peak greenness), β (GDD thresholds, forest types), w (varying intercept that also includes the random effect), and ϵ (measurement error), respectively. We can write Eq. 1 in vector form as $y = X\beta + w + \epsilon$. In this regression model, the unobserved space-time random effect w explains y alongside the observed predictors; the measurement errors $\epsilon(\ell_i)$ are independent and identically distributed Gaussian random variables with variance τ^2 . This model is a

spatiotemporally-varying intercept (SVI) model, since the random effects w act as the dynamic intercept term in the model. The random effects can also be interpreted as a way to relax the *i.i.d.* assumption the LR model makes on the measurement errors. In fact, we note that the SVI model can be written as $y = X\beta + \tilde{\epsilon}$ where $\tilde{\epsilon} \sim MVN_n(\mathbf{0}, C_{\mathcal{T}} + \tau^2 I_n)$ are *spatially and temporally correlated* measurement errors. The Bayesian posterior distribution for this model now includes w and can be written as

$$\pi(w, \beta, \tau^2 | y, X) \propto p(y | w, X, \beta, \tau^2) \pi(w | \theta) \pi(\beta, \tau^2, \theta). \quad (2)$$

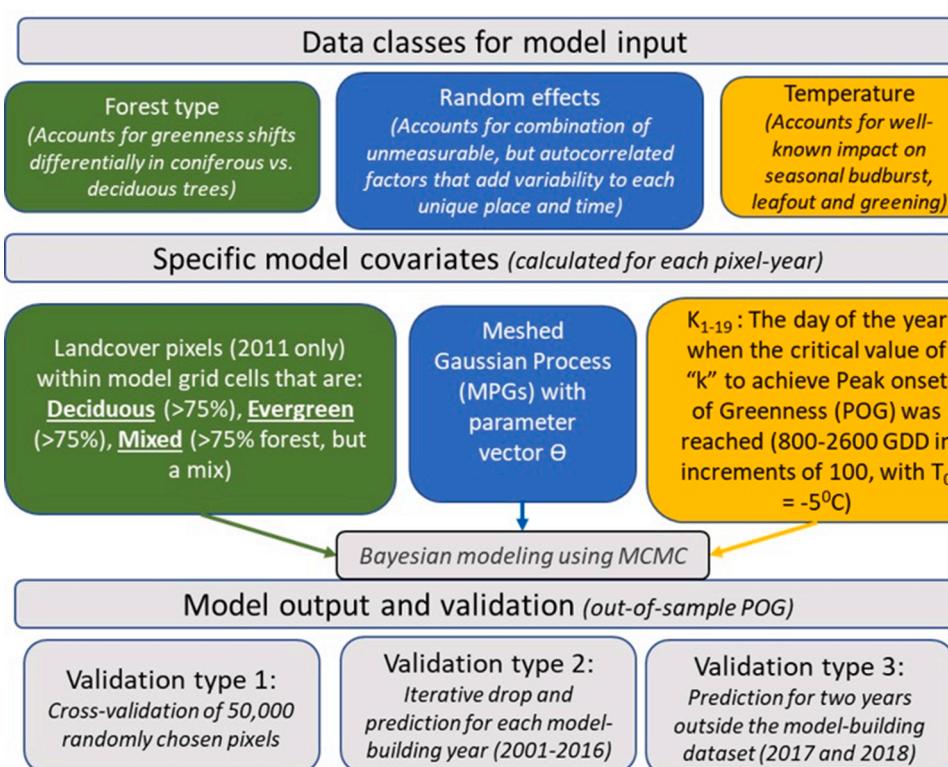
This general regression model is computationally tractable but becomes computationally cumbersome when the number of data locations n is large. When working with extensive satellite generated layers, the massive size of the covariance matrices demand massive computing power to model spatial dependence. In fact, in a standard GP, one needs to compute the inverse and the determinant of $C_{\mathcal{T}}$ which is of size $n \times n$; the cost of these operations grows as $O(n^3)$, leading to prohibitive computational demands. The growing literature on scalable GPs for Bayesian hierarchical modeling addresses this issue by replacing the original GP with a computationally tractable alternative. Here, we implement the meshed Gaussian process (MGP) models introduced by Peruzzi et al. (2020). This particular class of GPs is suitable for Bayesian regression models for space-time data arising from satellite imaging. GPs are convenient tools to model spatial and/or temporal dynamics via parametric covariance functions or kernels, and lead to tractable inference, allow more accurate interpolation of predictions, projections into different time periods, along with a more robust quantification of uncertainty (Banerjee et al., 2008). In addition to the covariance function C , a MGP is based on partitioning the spatiotemporal domain into blocks. Then, each block is mapped in a 1:1 relationship with nodes of a directed acyclic graph (DAG, also called a *Bayesian network*). The DAG is a conceptual model of conditional independence across nodes; when the nodes refer to domain partitions as in a MGP, the DAG is the ruleset that determines how dependence is allowed to flow in space and time across different partitions. The DAG structure for spatio-temporal data,

in this case, is a 3-d mesh (for details see Peruzzi et al., 2020). Using a DAG to limit spatial dependence guarantees that the posterior distribution of all unknowns can be computed even for a much large number of observational locations and times. In this way, we developed our model. The model input and output, including the components of the model are summarized in the flowchart (Fig. 4).

2.5. Estimation and prediction

The posterior distribution in Eq. (2) is not available in closed form, but can be sampled using Markov Chain Monte Carlo (MCMC) methods. In our applications here, we take $\pi(\beta) = N(\mathbf{0}, V_{\beta})$ where $V_{\beta} = 1000I_p$ and $\pi(\tau^2) = IG(a, b)$ – an inverse Gamma distribution – where $a = 2, b = 1$; these are customary choices that correspond to vague priors for these parameters and thus encode high uncertainty *a priori*. MCMC methods construct a Markov chain whose stationary distribution is $\pi(w, \beta, \tau^2 | y, X)$; their output is a correlated sample of size S , denoted as $\{w^{(j)}, \beta^{(j)}, \tau^{2(j)}\}_{j=1}^S$, which can be used to compute approximate summaries of quantities of interest, such as the posterior mean and variance, and make predictions at unobserved or left-out spatial locations. Uncertainty about estimates and predictions can be assessed by constructing credible intervals; these, too, are obtained from $\{w^{(j)}, \beta^{(j)}, \tau^{2(j)}\}_{j=1}^S$. We partition the 2,134,308 space-time locations into 278,338 blocks via axis-parallel partitioning of the three coordinate axes (longitude, latitude, time) into 132, 111, and 19 GDD values, respectively. We choose a “cubic” DAG to link the partitions – each rectangular region in the spatial domain is then modeled to be conditionally independent of all others, once values at its spatial neighbors (including itself in time) are given, as described in Peruzzi et al. (2020). This means that each region is “explained” solely by its spatial and temporal immediate neighbors. We build the MGP model on a nonseparable spacetime correlation defined in Gneiting (2002):

Fig. 4. Data inputs for our model parameterization that predicts the day of the year when greenness is at peak onset (POG) for North America east of the 100th meridian and up to 52°N latitude. Parameterization must be done separately for any focal study region because accounting for autocorrelated random effects is computationally intensive but greatly increases predictive power. Models were validated with out-of-sample predictions in three ways. Projections can then be made into years with no satellite data, but always adding new years as they become available will increase predictive power and decrease the problems of nonstationarity.



$$C((s, t), (s', t'); \theta) = \frac{1}{(\alpha|t - t'| + 1)} \exp \left\{ -\frac{\phi \|s - s'\|}{(\alpha|t - t'| + 1)^{\gamma/2}} \right\}, \quad (3)$$

where $\theta = \{\sigma^2, \alpha, \gamma, \phi\}$ are the unknown covariance parameters. Then, $w(\cdot) \sim MGP(\mathbf{0}, \sigma^2 C(\cdot, \cdot))$, where $\sigma^2 > 0$ is the spatial variance, $\alpha > 0$ and $\phi > 0$ are the time and space decay parameters (their inverses are the temporal and spatial ranges, respectively), and $\gamma \in [0, 1]$ is a parameter controlling the degree of nonseparability between space and time. Our MCMC algorithm estimates σ^2 via parameter expansion (Peruzzi et al., 2021); rather than sampling σ^2 directly, we write the model as $y(\ell) = \mathbf{x}^\top(\ell)\beta + \lambda w(\ell) + \epsilon(\ell)$, where $w(\cdot) \sim MGP(\mathbf{0}, \tilde{\sigma}^2 C(\cdot, \cdot))$. Here, λ and $\tilde{\sigma}^2$ are not separately identifiable as they are introduced with the sole purpose of improving MCMC performance. Ultimately, we find $\sigma^2 = \lambda^2 \tilde{\sigma}^2$. We note that our 19 GDD predictors are highly correlated; correlation across predictors in linear models is concerning when not enough data are collected to reliably estimate their effects on the output. In our case, no such problem arises given the huge dimension of our dataset: in Fig. 5(b), all uncertainty bands for GDD predictors are narrow, indicating that the sample size is large enough (about 1.7 million) to estimate the individual GDD effects with relatively high precision.

2.6. Performance criteria

We evaluate and compare our model on a test set of 50,000 locations spanning all years, in addition to the full year of 2017 and 2018. Predictive performance for model m is measured via the root mean square error (RMSE), defined as $RMSE(m) = \sqrt{\sum_{j=1}^{n_{out}} (y(\ell_j) - \hat{y}_m(\ell_j))^2 / n_{out}}$ where $y(\ell_j)$ is the observed peak onset and $\hat{y}_m(\ell_j)$ is the prediction at location ℓ_j from model m ; the mean absolute percentage error (MAPE), defined as, $MAPE(m) = 100 \sum_{j=1}^{n_{out}} \left| \frac{y(\ell_j) - \hat{y}_m(\ell_j)}{y(\ell_j)} \right| / n_{out}$, and the frequentist coverage of 95% credible intervals of predictions, which is defined as the percentage of data falling into 95% uncertainty intervals identified by the model. Accurate coverage (i.e. near the 95% nominal target) is important as it relates to the accuracy in quantifying posterior uncertainty of predictions. RMSE and MAPE are widely used criteria of model selection in remote sensing (Richter et al., 2012; Liao et al., 2016; Hamm et al., 2015; Kouadio et al., 2014; Wu et al., 2019).

In this way, we develop a spatio-temporal model [LR (fixed only) and SVI (with random effect), as described in Section 2.3 to predict the greenness peak onset times in the Eastern USA. Fixed predictor variables (19 "k" thresholds, as described in Section 2.3 + forest types) are fed into the LR and SVI models. LR is included for comparison with the SVI model. In the first step, we develop models based on all years' POG from MODIS except the year being predicted. For example, for 2010, we developed a model based on MODIS POG from all years from 2001 to 2016 except 2010. Then, using the 2010 GDD, we predict the peak onset for 2010, then we compare this with the MODIS POG for 2010. In this way, every year, we compare the model predicted POG with the MODIS observed POG values. We repeat this for all years. In this way we validated our model using 16 years from our model building step. In the second step, we use the model to predict the near-term future POG, specifically, 2017 and 2018 in this study. These two years are left as test sets. As earlier, we develop models based on all years from 2001 to 2016

MODIS POG, then we use this model to predict POG for 2017 and 2018 and finally compare those with the MODIS POG.

3. Results

Not surprisingly, the random effects model with spatially varying intercept (SVI) models have substantially lower root mean square (RMSE) and mean absolute percentage (MAPE) error values compared with the linear regression (LR) models (Table 1). RMSE measures the absolute deviation of predictions from the observations, whereas MAPE expresses the percentage of deviation of the predictions from the observations; in both cases, smaller values mean a better model fit. All subsequent results are thus based on SVI models only. SVI models that include the effect of dominant forest cover types (Mixed, Deciduous, and Evergreen) outperformed models based only on the 19 k values for GDD, although by a much smaller margin (Table 1). Overall, the best models included the Gaussian process-based (spatiotemporal) random effects, GDD predictors, and forest type and all subsequent results are generated from this model construction.

Spatial distance lags were much more important in the model compared to time lags (Fig. 5a). The spatial effect reduces exponentially as the distance increases, meaning that the farther locations (greater than 1 decimal degree) have a very small effect upon the model prediction at any given location. The effect of time lags is conversely very small. This means all previous years have comparable effect upon prediction in any given year. A combination of several k values had the largest impact on model prediction. In order to incorporate k in our model, we used as predictors the 19 constant GDD thresholds (see Section 2.3). Our results suggest that different thresholds have different effect sizes on model fit, depending on k (Fig. 5b) but that these are additive. For instance, let's focus on a k -value of 1300. The beta coefficient for this accumulation is 0.15. This implies that there is a 0.15-day delay in peak onset for every 1 additional day that is required to reach a threshold of 1300 GDDs. That impact is also influenced by other k thresholds with the magnitude depending on the value of the beta coefficient and whether the credible interval crosses zero. Although our predictors were highly correlated, we did not detect any performance issues due to multicollinearity. In particular, the standard errors of regression coefficients were relatively small (e.g., see the spread in each box plot in Fig. 5b) and we did not detect any switching of signs of coefficient estimates. We believe, in our case, multicollinearity was not an issue in model performance due to the massive sample size ($N = 1747312$).

Our model predicts the MODIS observed POG with very high accuracy within the model-building dataset (2001–2016). Predicted peak onset is significantly correlated with the observed peak onset for all years, with in-sample correlations above 0.93 for all years (Fig. 6). Each data point in this figure represents a single location and there are altogether 27,178 forest locations in our region of interest. There is little spatial structure to the distribution of the errors (Fig. 7) suggesting that our GP random model appropriately accounted for the substantial spatial structure in errors evident in the SLR model (results not shown). Our model was able to project peak onset of greenness well into two subsequent years outside our model-building data set. Relatively, the model performed better for 2017 than 2018 (Fig. 8). Differences between the model predictions and the satellite observations were mostly

Table 1
Model performance summary.

Model	Forest	Validation set				2017 prediction				2018 prediction			
		MAPE	RMSE	95% Covg	Width (days)	MAPE	RMSE	95% Covg	Width (days)	MAPE	RMSE	95% Covg	Width (days)
SVI	no	4.841	8.2384	95.70%	33.8	6.4687	10.1459	95.20%	40.89	5.9737	9.156	97.98%	46.92
SVI	yes	4.8174	8.2227	95.32%	32.7	6.5488	10.1969	95.71%	43.43	6.1102	9.2742	98.42%	51.68
SLR	yes	10.5255	15.4441	93.35%	60.7	10.7402	17.0653	91.10%	60.7	8.1921	12.9737	95.78%	60.68

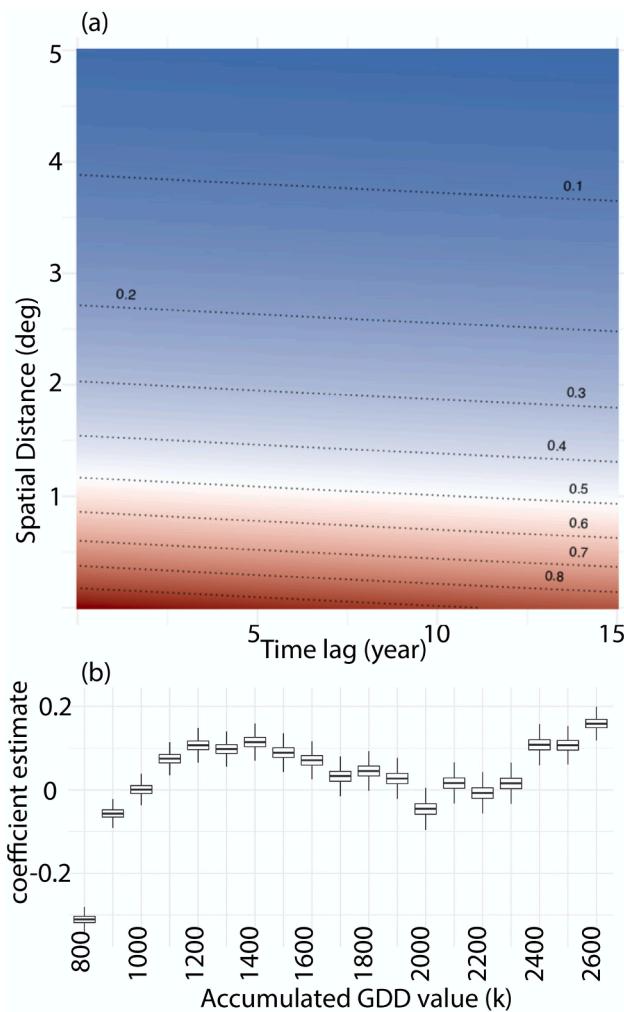


Fig. 5. (a) Spatio-Temporal effect on model prediction (estimated covariance at spacetime lag). (b) Effect of GDD accumulation threshold on the peak onset of greenness (POG). Line in each box indicates the credible interval.

within 3 days in 2017 and 7 days in 2018. In both cases, projections were much more successful for forest stands dominated by deciduous trees (correlations 0.92 and 0.81 for 2017 and 2018, respectively). R^2 value remained high for mixed stands, but fell significantly for stands dominated by evergreen trees (Fig. 9). Most pixels in the study region contain mixed forest, with pure evergreen stands the least common (Fig. 10). Note that prediction error skewed early in both years and for all forest types, with a bigger shift early in 2018.

4. Discussion

We have presented a model that produces highly accurate predictions for green-up timing within mixed and deciduous-dominant temperate forests in Eastern North America (Figs. 5, 6; Table 1). We leveraged an advancement in a computation process, MGPs, (Peruzzi et al., 2020) that allowed us to estimate posterior distributions for both fixed effects (GDD accumulation, k , and forest type) and Gaussian processes (GP) that capture both spatially and temporally autocorrelated random effects over massive spatial and temporal extents. Previously, computational limitations would have made those GPs difficult to estimate; but the use of MGPs (Peruzzi et al., 2020) allowed us to overcome that obstacle. In our model, spatial autocorrelation was much more important than temporal autocorrelation (Fig. 5); this means we can project our model into the past (when daily climate data are available). For example, Livneh that provides data since the early 20th century on

daily time scale), we can also project forward into the future when global climate models provide daily forecasts at an appropriate spatial scale. At some point, issues of nonstationarity may erode predictive power the further we move away from the model-building time-frame (Rollinson et al., 2021); however, continually updating the model as new satellite data become available will greatly reduce this problem (Dietze et al., 2018).

Observations of spring phenology have been made for centuries (Chen, 2003). Long-term changes (e.g., decadal) in green-up timing have been investigated since the 1950s (Schwartz and Reiter, 2000) using the field observations and since 1980s, with satellite data (Jeong et al., 2011). In terms of predicting green-up timing, generally, there has been good correspondence between the field observations, on a local scale, and remotely sensed data (Kross et al., 2011). However, predicting on regional and continental scales has been challenging. Furthermore, spring phenology appears to be more sensitive to region (Reed, 2006; Jeong et al., 2011) and together, these offer more difficulties in building accurate predictive model for larger scales.

Green-up during our study period has generally advanced, for example, it began much earlier in some of the most recent years (as in 2012, 2015, Fig. 2) than in the earlier years (as in 2002–2004). This is consistent with previous study, which shows that timing has advanced by roughly 0.5 days per year over two decades (Keenan et al., 2014) and with reviews of earlier spring phenology associated with warming in temperate forests (Linderholm, 2006; Richardson et al., 2013). For studies at the broadest spatial scales in North America, climate, land cover, and some biodiversity data (especially for birds, but also for high-profile insects such as butterflies) are available since the early 1900s (Wells and Tonkyn, 2014; Dietze et al., 2018). The missing piece for any study of direct and indirect impacts of phenological shifts is tracking vegetation green-up in these natural systems. While Schwartz et al. (2013) provide spring onset indices across North America, these are calibrated to lilac and honeysuckle, herbaceous plants grown as ornamentals that include many species classified as invasive in North America (Beans et al., 2012; Swearingen and Barger, 2016). Lilac and honeysuckle observations for these onset products originate from a large network of on-the-ground volunteer observers (through the USA-National Phenology Network, www.usanpn.org) since they are very easy to observe. However, the relevance for general ecosystem function is currently unknown.

We included the spatiotemporal component in a simple linear regression model (SLR) and also developed a spatially varying intercept (SVI) model. This, implemented using a meshed Gaussian process under a Bayesian framework, helped increase the model's performance substantially and therefore supported this modeling approach despite its higher level of complexity and computational demands. Specifically, goodness of fit tests showed that deviations from expectations were more than halved using the SVI vs. the simpler SLR models. In contrast, including forest types had only marginal performance improvement (Table 1), but that also may be due to the fact that performance was generally high for the two most common forest types, Mixed and Deciduous (Fig. 9). In agreement with this, past studies also showed that the model predictions correlated most with the satellite observations in forest covers (Deciduous and Mixed) of the Eastern North America (Schwartz and Reed, 1999). Overall, within-sample models showed high fits for our study region ($R^2 > 95\%$, Fig. 6). Finally, we used this SVI model to predict POG for two out-of-sample years (2017 and 2018) and the predicted values were within 3 days for most pixels in 2017 and within 7 days for most pixels in 2018 (Fig. 8). Differences were relatively larger (up to 10 days) in the boreal forest of the Northeast and dominant Oak forests of Missouri and Arkansas in 2018. Further research is necessary to improve performance in these region.

For predictors, we used a series of 19 GDD thresholds, k (described in Section 2). k is a parameter that measures the number of GDD that must be accumulated for developmental milestones like budburst, maximum greening, flowering (for plants) or molting and eclosion (for example, in

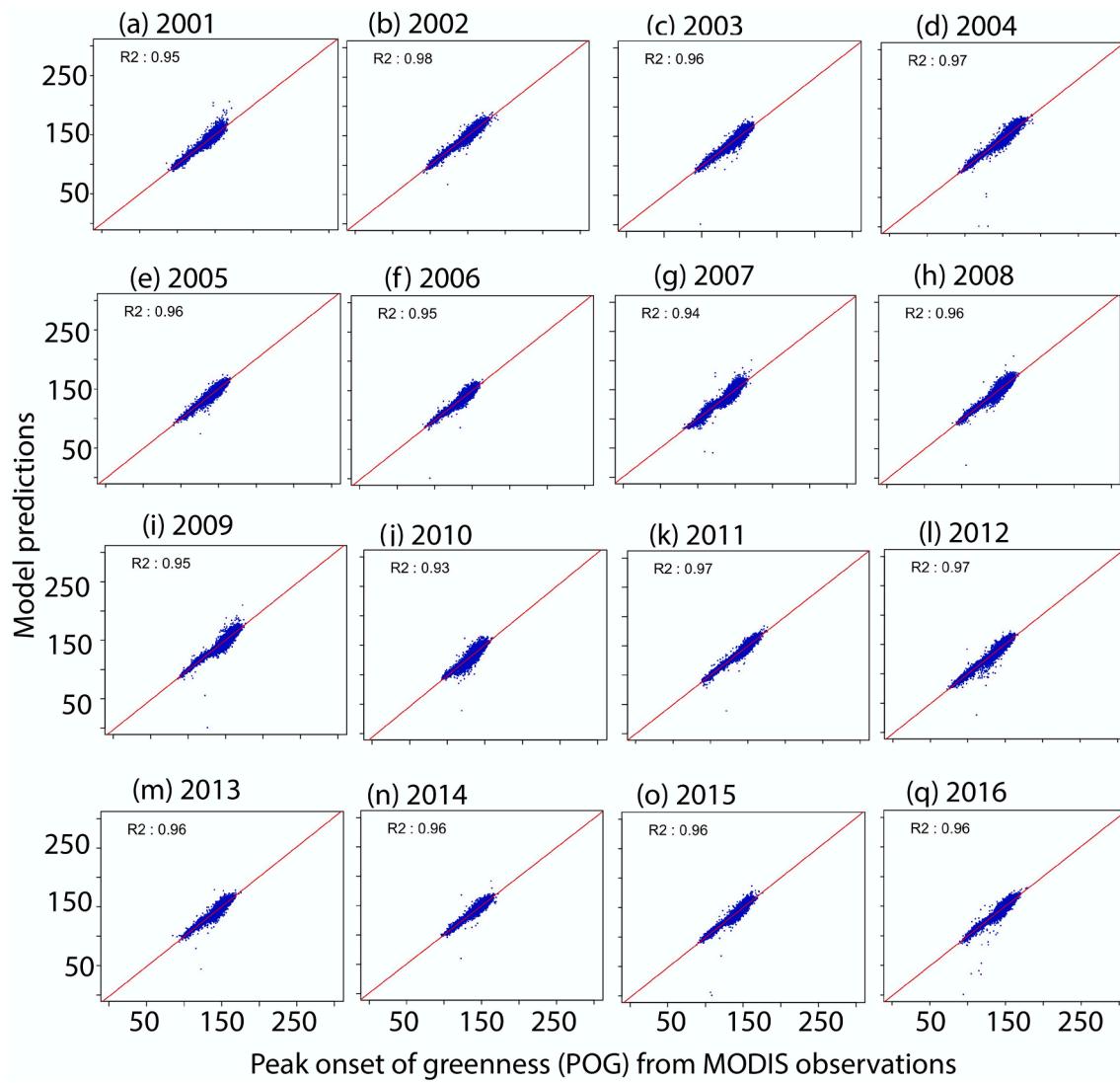


Fig. 6. Scatterplot of the peak onset of greenness (POG) predicted from the model and observations from the MODIS satellite for 2001 through 2016.

holometabolous insects). GDD models have been successful because for ectothermic organisms, development, once triggered, depends most strongly upon ambient air temperatures (Van der Have and De Jong, 1996). In well-studied species (mostly crops and pests), detailed physiological laboratory experiments are conducted to determine the base temperature and k for any individual species (Abarca et al., 2018). In the case of this study, we were tracking development of an ecotype (forest) rather than an individual species, so there is less guidance available on the best critical values to set. So we settled on -5°C after initial exploratory analyses using a range of values typical for these types of systems (see method section). There is much less guidance on appropriate values for k . Our solution was to provide 19 GDD thresholds (parameterized as the POG date when each of those thresholds was reached) and our results suggest that this combination of parameters provided a highly accurate model, although variable coefficients for each of the 19 variables means that we can make little inference about what is driving those values of k . However, previous studies of pests and crops have suggested that " k " may be more locally determined than T_0 , which tends to be highly conserved in most species (Abarca et al., 2018). In our case, we avoided that with the use of multiple thresholds and a spatial model of autoregressive correlations where a spatially varying random intercept allows the model to be tailored to fixed locations throughout Eastern North America. This means, the spatial component vary in a way such that the model provides the best predictions for that

location. Because of this, understanding the physical mechanism becomes difficult in our case. Although the specific parameterizations of this model could not be transferred to any other region, which is fundamentally true for any model with a strong random-effect component, the same approach could be implemented for forest green-up in any forest in other temperate regions of the world.

Our model is solely based on the GDD accumulation and cover types. Xie and Wilson (2020) estimated the green-up timing of the deciduous Eastern USA forest using the MODIS satellite observations from 2001 to 2015. They used a piecewise linear combination of line segments joining four major points that mark the change in the value of greenness index, also called the change point estimation method, to model the green-up timing. They reported a much larger variation in the onset times, especially in the eastern USA (e.g., POG ranged from 51 to 154, which is 20 February–3 June). Also, comparable to our range of predictive accuracy, Jolly et al. (2005) estimated the green-up timing of the Harvard Forest from 1990 to 1997. Their model estimations deviated from the observations by up to 8 days, however, this was only for one location and there are many differences in our method from their method. Their model was based not only on temperatures but also on other variables such as photoperiod and vapor pressure. Likewise, Zhang et al. (2006) developed a piece-wise logistic model based on the MODIS observations and used that to calculate the green-up timing. The calculated times differed by less than 10 days from the in-situ measured greenness index

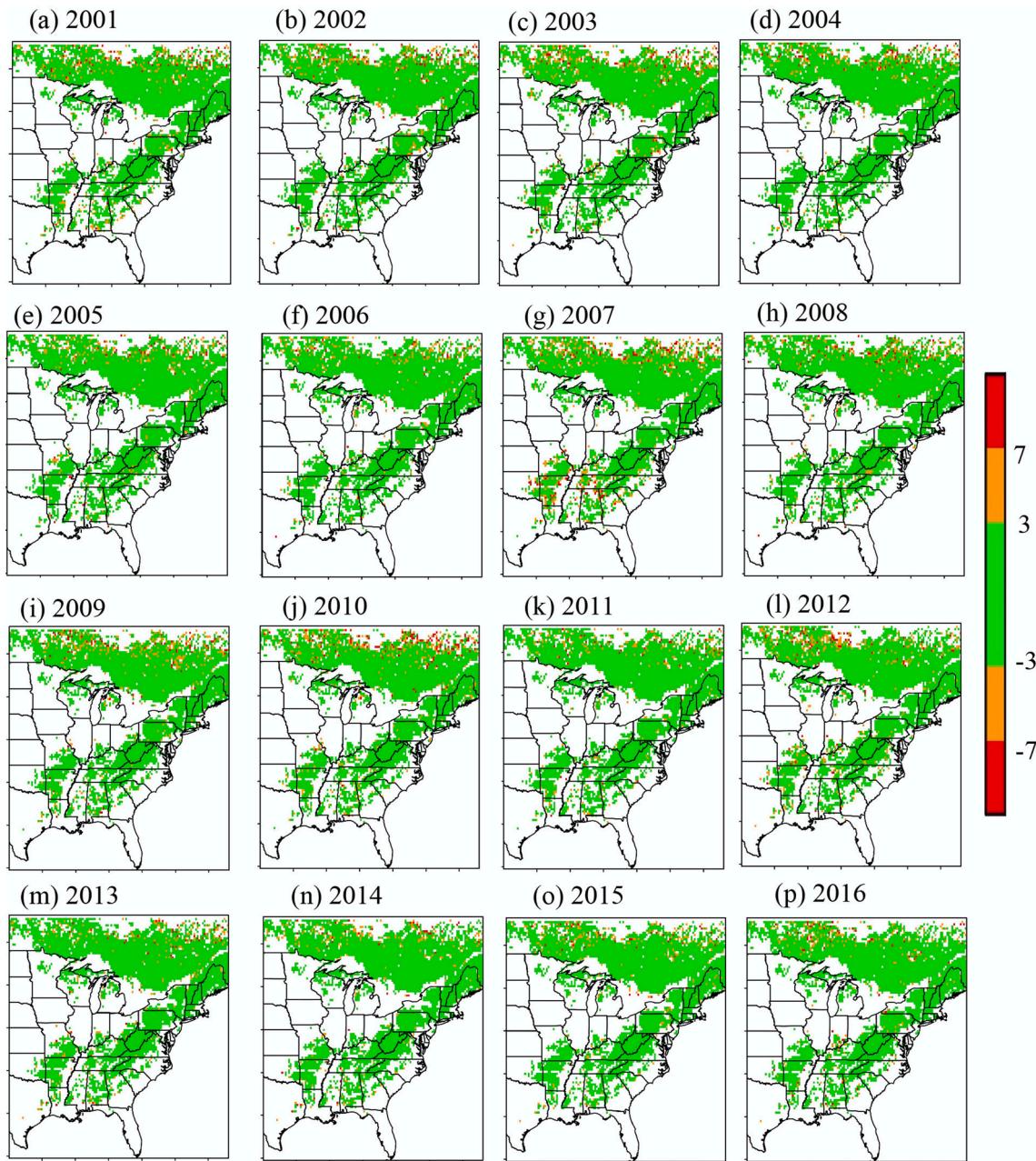


Fig. 7. Peak onset of greenness (POG) timing from SVI Model prediction minus MODIS observation.

in the Harvard Forest.

Compared with the deciduous and mixed forests, SVI model was relatively less efficient in predicting onset peak in the evergreen forest (Fig. 9). This may likely be due to the difficulty in detecting seasonal variation in foliage pigmentation for plants that stay green throughout the year (Filippa et al., 2018; Bowling et al., 2018; Johnson et al., 2011; Kobayashi et al., 2018). For instance Richardson et al. (2018) compared the POG observed from the MODIS satellite to the ground based Phe- noCam data. In their study, the two data sets were highly correlated in deciduous (Pearson's $r = 0.83$) but the correlations were very small in the evergreen ($r = 0.37$) forest. Model performance was generally high for deciduous forests, but for 2018, performance dropped from most estimates being within ± 3 days to ± 7 days or more (Fig. 8b). We did not attempt to determine what factors may have contributed to this drop in performance, but clearly there was a shift in the relationship between temperature and greenup. While future mechanistic studies may be able

to pinpoint the physiological basis for this variability, it is clearly important to include as many years as possible when parameterizing the model to project into years where satellite data are not available. Thus, there would be high value for always updating the model parameters whenever more years become available.

Modeling the year-to-year peak onset variability and its accurate prediction is useful for understanding the phenological relationship across multiple trophic levels (Meineke and Davies, 2019). Trophic mismatches may cascade; for instance, shifts in phenological timing of plant development and associated insects may cause insectivorous birds to either be more or less synchronous with their food source (Brooks et al., 2017). Either situation could cascade throughout the larger ecosystem in ways that are very difficult to measure directly. Although studies of phenological mismatch have become more common, most of those studies are restricted to one or a small set of study sites and are often in simplified ecosystems, such as arctic systems (Visser and Both,

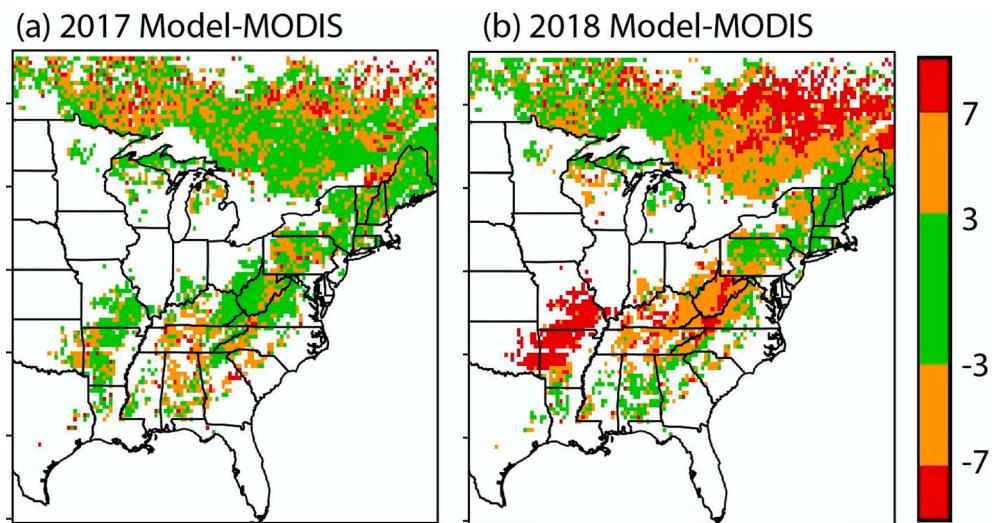


Fig. 8. (a) 2017 peak onset of greenness (POG) timing from SVI Model prediction minus MODIS observation. (b) Is same as (a) but for 2018.

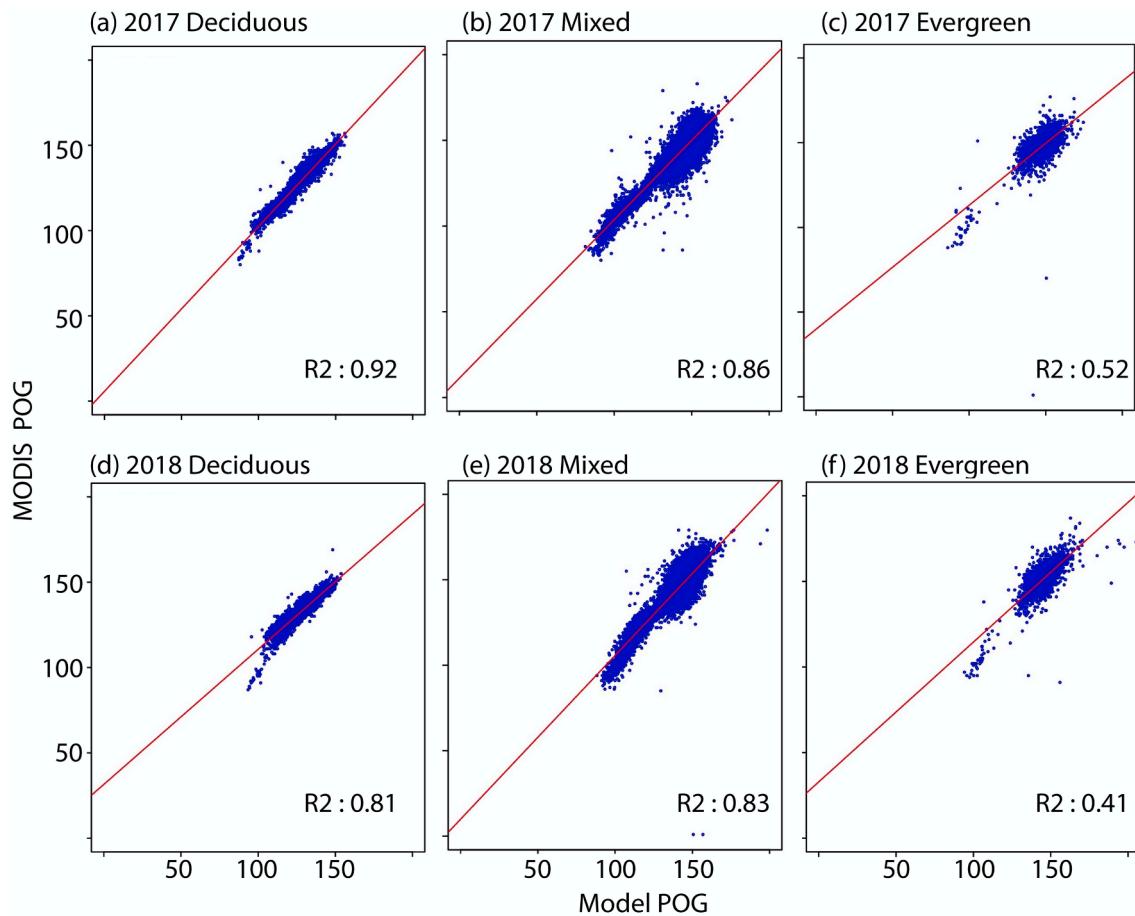


Fig. 9. Peak onset of greenness (POG) timing scatter plots (MODIS, Model) for (a) Deciduous, (b) Mixed, and (c) Evergreen forest locations in 2017. (d, e, f) are same as in (a, b, c) but for 2018.

2005). Large-scale studies of phenological mismatch are rare (but see Mayor et al., 2017; Franks et al., 2018; Youngflesh et al., 2021) because harmonizing data on multiple trophic levels is difficult at the largest spatial scales (Heffernan et al., 2014). The model presented here overcomes that barrier for remotely-sensed plant dynamics at continental scales.

The temperate forest ecoregion in eastern North America is the

breeding ground of many resident and migratory birds, bats, and insects. Every year, migrants arrive there after overwintering in the south, while residents initiate their breeding. The peak onset times of forest greenness correlates with the arrival and breeding times of migrants (La Sorte et al., 2014a,b) and egg-laying times of resident bird species (Phillimore et al., 2016; Franks et al., 2018). However, the synchrony of those shifts differs among species and regions; green-up timing is shifting faster at

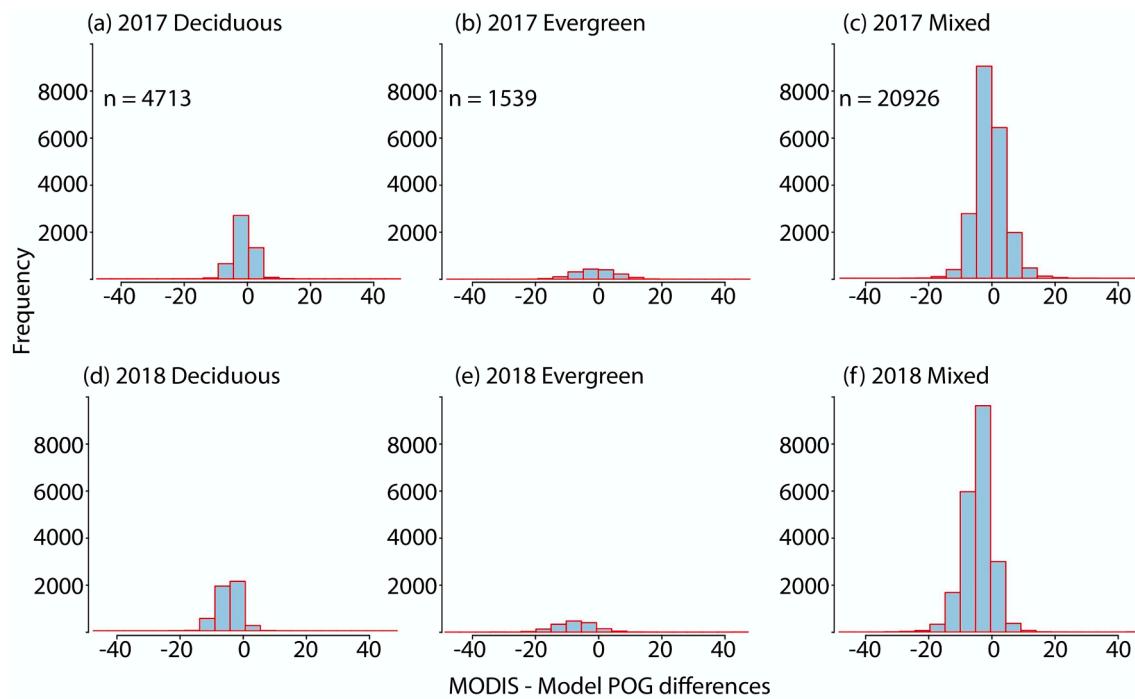


Fig. 10. Peak onset of greenness (POG) timing differences between MODIS and SVI model for Deciduous (a), Evergreen (b), and Mixed (c) forest locations in 2017. (d, e, f) are same as in (a, b, c) but for 2018.

more northern latitudes where synchronization with arriving migratory birds is historically at its highest and mismatches there may have the biggest cross-trophic impacts (Youngflesh et al., 2021). To more rigorously examine these and similar dynamics, expanding the time frame both in the past (hindcasting) and future (forecasting) of these studies is critical. Since daily surface minimum and maximum temperature data is available since the early 20th century (e.g., Livneh), our model will be able to hindcast the green-up times in the past. This will enable us to explore the relationship between insects, birds and vegetation phenologies in the past. Similarly, using the climate models, we can forecast green-up times into the future. Understanding the patterns and mechanisms of the variability and predicting its onset times is important for biodiversity management in the present and preparing the conservation plans under various emerging threats of global warming.

For most years, the frequency of differences between the MODIS observations and the model are centered around zero, and the distribution tapers off towards both sides (Fig. 10). In a few locations, the differences were beyond ± 25 days. It is important to note here that the variability in the MODIS observations were very high (as displayed in Fig. 3). This could be associated with the algorithm used in calculating the POG from the MODIS satellite, and here we do not investigate this further (but see Soudani et al., 2008; Turner et al., 2006). Regardless, our model predicted onset peaks that occurred largely within ± 3 days from the MODIS observed peak. Such high prediction accuracy of green-up timing peaks helps prepare management plans not only for birds and insects but also controlling emergence of pests (Steinbauer, 2011; Deveson, 2013).

5. Conclusions

Vegetation peak onset of greenness (POG) varies in Eastern North America from year to year, as depicted in the MODIS satellite observations. In general, greening begins much earlier in the forests to the south (by early spring) than in the northeast (by late summer). Also, the onset peak varies interannually by up to ± 50 days. Using the daily surface maximum and minimum temperatures from the Daymet observations, we developed a meshed Gaussian process, spatially varying intercept

(SVI) model under the Bayesian framework. Our model was based on the MODIS and Daymet observations from 2001 to 2016. We validated the model performance in this way: first we developed our model based on all years except a year, and in the second step, we predicted the year that was excluded in the model building year. Our model predictions correlated significantly with the observations ($R^2 = 0.97$). The model was also able to capture the spatial and temporal variability in the onset timings and the differences between the model predictions and the observations were mostly ± 3 days. Then, we used this model to predict two years (2017 and 2018) that were not included in our model building. Differences between the SVI model predictions and the observations were still low but with greater error (± 7 days). There were some variations in the model prediction accuracies by forest types. In the deciduous and mixed forest types, the correlations between the model and observations were much larger ($R^2 = 0.92$), while correlations were smaller in the evergreen forest of the south ($R^2 = 0.41$, as in 2018). These results indicate that the greenness timing, in the forest with deciduous trees, can be predicted with high degree of accuracy using the surface temperature observations in our model. This will be useful for projecting greenness timing backward into the past when daily records of surface temperature observations are available (for example, Livneh, since early 20th century) as well as forecasting into the future generally to 2100 using climate models.

The power of this model comes from leveraging the strong influence of local dynamics via a random model based on Gaussian processes. This framework could be applied to any region that is temperate with a strong component of deciduous trees - but each region must be parameterized separately, and updating models each year will result in improved accuracy for forecasts and backcasts.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by Georgetown University under National Science Foundation Award, NSF #1738243 and NSF#2017791. M. P. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 856506), and grant R01ES028804 of the United States National Institutes of Health (NIH).

References

Abarca, M., Larsen, E., Lill, J., Weiss, M., Lind, E., Ries, L., 2018. Inclusion of host quality data improves predictions of herbivore phenology. *Entomol. Exp. Appl.* 166, 648–660.

Altermatt, F., 2010. Climatic warming increases voltnism in European butterflies and moths. *Proc. Biol. Sci.* 277, 1281–1287.

Aurambout, J., Finlay, K., Luck, J., Beattie, G., 2009. A concept model to estimate the potential distribution of the Asiatic citrus psyllid (*Diaphorina citri* Kuwayama) in Australia under climate change—a means for assessing biosecurity risk. *Ecol. Model.* 220, 2512–2524.

Bale, J.S., Masters, G., Hodgkinson, I., Awmack, C., Bezemer, T., Brown, V., Butterfield, J., Buse, A., Coulson, J., Farrar, J., Good, J., 2002. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Glob. Chang. Biol.* 8, 1–16.

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2014. Hierarchical Modeling and Analysis for Spatial Data, second ed. Chapman and Hall/CRC. <https://doi.org/10.1201/b17115>.

Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. *J. Roy. Stat. Soc. Ser. B* 70, 825–848. <https://doi.org/10.1111/j.1467-9868.2008.00663.x>.

Beans, C., Kilkenny, F., Galloway, L., 2012. Climate suitability and human influences combined explain the range expansion of an invasive horticultural plant. *Biol. Invasions* 14, 2067–2078.

Both, C., Bouwhuis, S., Lessells, C.M., Visser, M.E., 2006. Climate change and population declines in a long-distance migratory bird. *Nature* 441 (7089), 81–83.

Botta, A., Viovy, N., Ciais, P., Friedlingstein, P., Monfray, P., 2000. A global prognostic scheme of leaf onset using satellite data. *Glob. Chang. Biol.* 709–725.

Bowling, D., Logan, B., Hufkens, K., Aubrecht, D., Richardson, A., Burns, S., Anderegg, W., Blanken, P., Eiriksson, D., 2018. Limitations to winter and spring photosynthesis of a Rocky Mountain subalpine forest. *Agric. For. Meteorol.* 252, 241–255.

Brooks, S.J., Self, A., Powney, G.D., Pearse, W.D., Penn, M., Paterson, G.L., 2017. The influence of life history traits on the phenological response of British butterflies to climate variability since the late-19th century. *Ecography* 40, 1152–1165.

Cayton, H., Haddad, N., Gross, K., Diamond, S., Ries, L., 2015. Do growing degree days predict phenology across butterfly species? *Ecology* 96, 1473–1479.

Chen, X., 2003. Assessing phenology at the biome level. In: *Phenology: an integrative environmental science*. Springer, Dordrecht.

Chuine, I., 2000. A unified model for budburst of trees. *J. Theor. Biol.* 2007, 337–347.

Chung, U., Mack, L., Yun, J.I., Kim, S.H., 2011. Predicting the timing of cherry blossoms in washington, dc and mid-atlantic states in response to climate change. *PLoS One* 6, e27439.

Cleland, E.E., Chuine, I., Menzel, A., Mooney, H.A., Schwartz, M.D., 2007. Shifting plant phenology in response to global change. *Trends Ecol. Evol.* 22 (7), 357–365.

Cong, N., Wang, T., Nan, H., Ma, Y., Wang, X., Myneni, R., Piao, S., 2013. Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010. *Glob. Chang. Biol.* 19, 881–891.

Deveson, E.D., 2013. Satellite normalized difference vegetation index data used in managing Australian plague locusts. *J. Appl. Remote Sens.* 7, 1–075096.

Dietze, M.C., Fox, A., Beck-Johnson, L., Betancourt, J., Hooten, M., Jarnevich, C., Keitt, T., Kenney, M., Laney, C., Larsen, L., Loeschke, H., 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Nat. Acad. Sci.* 115, 1424–1432.

Diffenbaugh, N., Krupke, C., White, M., Alexander, C., 2008. Global warming presents new challenges for maize pest management. *Environ. Res. Lett.* 3, 044007.

Dunn, A.H., de Beurs, K., 2011. Land surface phenology of North American mountain environments using moderate resolution imaging spectroradiometer data. *Remote. Sens. Environ.* 115 <https://doi.org/10.1016/j.rse.2011.01.005>, 1220–123.

Faaborg, J., Holmes, R., Anders, A., Bildstein, K., Dugger, K., Gauthreaux Jr., S., Heglund, P., Hobson, K., Jahn, A., Johnson, D., Latta, S., 2010. Recent advances in understanding migration systems of New World land birds. *Ecol. Monogr.* 80, 3–48.

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Sonnentag, O., Humphreys, E., Hufkens, K., Ryu, Y., Verfaillie, J., di Cellia, U., Richardson, A., 2018. NDVI derived from near-infrared-enabled digital cameras: Applicability across different plant functional types. *Agric. For. Meteorol.* 249, 275–285.

Fleming, P.A., Wentzel, J., Dundas, S., Kreplins, T., Craig, M., Hardy, G., 2021. Global meta-analysis of tree decline impacts on fauna. *Biol. Rev.* 96, 1744–1768.

Franks, S.E., et al., 2018. The sensitivity of breeding songbirds to changes in seasonal timing is linked to population change but cannot be directly attributed to the effects of trophic asynchrony on productivity. *Glob. Chang. Biol.* 24, 957–971.

Friedl, M., Gray, J., Sulla-Menashe, D., 2019. MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid v006. 2019. Distributed by NASA EOSDIS Land Processes DAAC.

Fu, G., Li, S.W., Sun, W., Shen, Z.X., 2016. Relationships between vegetation carbon use efficiency and climatic factors on the Tibetan Plateau. *Can. J. Remote Sens.* 42 (1), 16–26.

Fu, Y., Zhang, H., Dong, W., Yuan, W., 2014. Comparison of phenology models for predicting the onset of growing season over the Northern Hemisphere. *PLoS One* 9, 10.

Fu, Y.H., Piao, S., Zhao, H., Jeong, S.J., Wang, X., Vitasse, Y., Ciais, P., Janssens, I.A., 2014. Unexpected role of winter precipitation in determining heat requirement for spring vegetation green-up at Northern middle and high latitudes. *Glob. Chang. Biol.* 20 (12), 3743–3755.

Gallinat, A.S., Ellwood, E.R., Heberling, J.M., Miller-Rushing, A.J., Pearse, W.D., Primack, R.B., 2021. Macrophenology: insights into the broad-scale patterns, drivers, and consequences of phenology. *Am. J. Bot.* 108, 2112–2126.

Gelfand, A., Diggle, P., Fuentes, M., Guttorm, P., 2010. *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL. <https://doi.org/10.1201/9781420072884>.

Gneiting, T., 2002. Nonseparable, Stationary Covariance Functions for Space-Time Data. *J. Am. Stat. Assoc.* 97, 590–600. <https://doi.org/10.1198/016214502760047113>.

Gneiting, T., Guttorm, P., 2010. Continuous-parameter Spatio-temporal Processes. In: Gelfand, A.E., Diggle, P., Fuentes, M., Guttorm, P. (Eds.), *Handbook of Spatial Statistics*. Chapman and Hall/CRC, pp. 427–436. <https://doi.org/10.1201/9781420072884>.

Hamm, N., Finley, A., Schaap, M., Stein, A., 2015. A spatially varying coefficient model for mapping PM10 air quality at the European scale. *Atmos. Environ.* 102, 393–405.

Harrington, C.A., Gould, P.J., Clair, J.B.S., 2010. Modeling the effects of winter environment on dormancy release of douglas-fir. *For. Ecol. Manage.* 259 (4), 798–808.

Heffernan, J.B., Soranno, P., Angilletta Jr., M., Buckley, L., Gruner, D., Keitt, T., Kellner, J., Kominoski, J., Rocha, A., Xiao, J., Harms, T., 2014. Macrosystems ecology: understanding ecological patterns and processes at continental scales. *frontiers in ecology and the environment*. *Front. Ecol. Environ.* 12, 5–14.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K., 2015. Completion of the 2011 National Land Cover Database for the conterminous United States representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* 81 (5), 345–354.

Homer, C.G., Dewitz, J.A., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., Auch, R.F., Riitters, K.H., 2016. Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database: ISPRS. *J. Photogramm. Remote Sens.* 162, 184–189.

Hostetler, J.A., Sillett, T.S., Marra, P.P., 2015. Full-annual-cycle population models for migratory birds. the Auk: Ornithological Advances. *Auk Ornithol. Adv.* 132, 433–449.

IPCC, 2007. *Intergovernmental Panel on Climate Change. Climate change 2007: The Physical Science Basis. Contribution of Working Group I to the fourth assessment report of the intergovernmental panel on climate change*. In: M.L. Parry, O.F. Canziani, J.P. Palutikoff, P.J. van der linden, C.E. Hanson (Eds.), Cambridge University Press, pp. 996.

Jenerette, G.D., Scott, R.L., Huete, A.R., 2010. Functional differences between summer and winter season rain assessed with MODIS-derived phenology in a semi-arid region. *J. Veg. Sci.* 21, 16–30.

Jeong, S.-J., Ho, C.-H., Gim, H.-J., Brown, M.E., 2011. Phenology shifts at start vs. end of growing season in temperate vegetation over the northern hemisphere for the period 1982–2008. *Glob. Chang. Biol.* 17, 2385–2399.

Jeong, S.J., Medvige, D., Sheviakova, E., Malyshev, S., 2012. Uncertainties in terrestrial carbon budgets related to spring phenology. *J. Geophys. Res. Biogeosci.* 117, G01030. <https://doi.org/10.1029/2011JG001868>.

Johnson, O.W., Fielding, L., Fox, J.W., Gold, R.S., Goodwill, R.H., Johnson, P.M., 2011. Tracking the migrations of pacific golden-plovers (*Pluvialis fulva*) between hawaii and alaska: New insight on flight performance, breeding ground destinations, and nesting from birds carrying light level geolocators. *Wader Study Group Bull.* 118 (1), 26–31.

Jolly, W.M., Nemani, R., Running, S., 2005. A generalized, bioclimatic index to predict foliar phenology in response to climate. *Glob. Chang. Biol.* 11, 619–632.

Keenan, T., Gray, J., Friedl, M., Toomey, M., Bohrer, G., Hollinger, D., Munger, J., O'Keefe, J., Schmid, H., Wing, I., Yang, B., 2014. Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nat. Clim. Chang.* 4, 598–604.

Kim, Y., Wang, G., 2005. Modeling seasonal vegetation variation and its validation against moderate resolution imaging spectroradiometer (MODIS) observations over North America. *Can. J. Plant Sci.* 110, D04106. <https://doi.org/10.1029/2004JD005436>.

Klosterman, S., Hufkens, K., Richardson, A., 2018. Later springs green-up faster: the relation between onset and completion of green-up in deciduous forests of North America. *Int. J. Biometeorol.* 62, 1645–1655.

Kobayashi, H., Nagai, S., Kim, Y., Yang, W., Ikeda, K., Ikawa, H., Nagano, H., Suzuki, R., 2018. In situ observations reveal how spectral reflectance responds to growing season phenology of an open evergreen forest in Alaska. *Remote Sens.* 10, 1071.

Kouadio, L., Newlands, N.K., Davidson, A., Zhang, Y., Chipanshi, A., 2014. Assessing the performance of MODIS NDVI and EVI for seasonal crop yield forecasting at the ecodistrict scale. *Remote Sens.* 6 (10), 10193–10214.

Kramer, K., Leinonen, I., Loustau, D., 2000. The importance of phenology for the evaluation of impact of climate change on growth of boreal, temperate and Mediterranean forests ecosystems: an overview. *Int. J. Biometeorol.* 44 (2), 67–75.

Kross, A., Fernandes, R., Seaquist, J., Beaubien, E., 2011. The effect of the temporal resolution of ndvi data on season onset dates and trends across canadian broadleaf forests. *Remote Sens. Environ.* 115, 1564–1575.

Kudo, G., Ida, T.Y., 2013. Early onset of spring increases the phenological mismatch between plants and pollinators. *Ecology* 94, 2311–2320.

La Sorte, F., Butchart, S.H., Jetz, W., Böhning-Gaese, K., 2014. Range-wide latitudinal and elevational temperature gradients for the world's terrestrial birds: implications under global climate change. *PLoS One* 9 (5).

La Sorte, F.A., Fink, D., Hochachka, W.M., DeLong, J.P., Kelling, S., 2014. Spring phenology of ecological productivity contributes to the use of looped migration strategies by birds. *Proc. Roy. Soc. B: Biol. Sci.* 281 (1793).

Lafferty, K., 2009. The ecology of climate change and infectious diseases. *Ecology* 7, 888–900.

Latifovic, R., Pouliot, D., Olthof, I., 2017. Circa 2010 land cover of Canada: Local optimization methodology and product development. *Remote Sens.* 9(11), 1098.

Laube, J., Sparks, T.H., Estrella, N., Höfler, J., Ankerst, D.P., Menzel, A., 2014. Chilling outweighs photoperiod in preventing precocious spring development. *Glob. Change Biol.* 20 (1), 170–182.

Leinonen, I., Kramer, K., 2002. Applications of phenological models to predict the future carbon sequestration potential of boreal forests. *Climatic Change* 55 (1–2), 99–113.

Li, W., Xin, Q., Zhou, X., Zhang, Z., Ruan, Y., 2021. Comparisons of numerical phenology models and machine learning methods on predicting the spring onset of natural vegetation across the Northern hemisphere. *Ecol. Ind.* 131, 108126.

Liao, L., Song, J., Wang, J., Xiao, Z., Wang, J., 2016. Bayesian method for building frequent landsat-like NDVI datasets by integrating MODIS and landsat NDVI. *Remote Sens.* 8, 6.

Linderholm, H., 2006. Growing season changes in the last century. *Agric. For. Meteorol.* 137, 1–14.

Livneh, B., Rosenberg, E., Lin, C., Nijssen, B., Mishra, V., Andreadis, K., Maurer, E., Lettenmaier, D., 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions. *J. Clim.* 26, 9384–9392.

Man, R., Lu, P., 2010. Effects of thermal model and base temperature on estimates of thermal time to bud break in white spruce seedlings. *Can. J. Plant Sci.* 40, 1815–1820.

Mayor, S.J., Guralnick, R.P., Tingley, M.W., Otegui, J., Withey, J.C., Elmendorf, S.C., Andrew, M.E., Leyk, S., Pearse, I.S., Schneider, D.C., 2017. Increasing phenological asynchrony between spring green-up and arrival of migratory birds. *Sci. Rep.* 7 (1), 1–10.

McMaster, G., Wilhelm, W., 1997. Growing degree-days: one equation, two interpretations. *Agric. For. Meteorol.* 87, 291–300.

Meineke, E.K., Davies, T.J., 2019. Museum specimens provide novel insights into changing plant–herbivore interactions. *Roy. Soc. B* 374, 20170393.

Morrison, M., McVetty, P., Shaykewich, C., 1989. The determination and verification of a baseline temperature for the growth of Westar summer rape. *Can. J. Plant Sci.* 69, 455–464.

Musolin, D.L., 2007. Insects in a warmer world: ecological, physiological and life-history responses of true bugs (Heteroptera) to climate change. *Glob. Chang. Biol.* 13, 1565–1585.

Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrar, G., Nemani, R.R., 1997. Increased plant growth in the Northern high latitudes from 1981 to 1991. *Nature* 386 (6626), 698–702.

Neupane, N., Zipkin, E.F., Saunders, S.P., Ries, L., 2022. Grappling with uncertainty in ecological projections: a case study using the migratory monarch butterfly. *Ecosphere* 13 (1), 1–18. <https://doi.org/10.1002/ecs2.3874>.

Peng, D., Wu, C., Li, C., Zhang, X., Liu, Z., Ye, H., Luo, S., Liu, X., Hu, Y., Fang, B., 2017. Spring green-up phenology products derived from MODIS NDVI and EVI: Intercomparison, interpretation and validation using National Phenology Network and Ameriflux observations. *Ecol. Ind.* 77, 323–336. <https://doi.org/10.1016/j.ecolind.2017.02.024>.

Peng, S., Piao, S., Ciais, P., Myneni, R.B., Chen, A., Chevallier, F., Dolman, A.J., Janssens, I.A., Penuelas, J., Zhang, G., Vicca, S., 2013. Asymmetric effects of daytime and night-time warming on Northern Hemisphere vegetation. *Nature* 501 (7465), 88–92.

Peruzzi, M., Banerjee, S., Dunson, D.B., Finley, A.O., 2021. Grid-Parametrize-Split (GrIPs) for improved scalable inference in spatial big data analysis. *arXiv: 2101.03579*.

Peruzzi, M., Banerjee, S., Finley, A.O., 2020. Highly scalable Bayesian geostatistical modeling via meshed Gaussian processes on partitioned domains. *J. Am. Stat. Assoc.* <https://doi.org/10.1080/01621459.2020.1833889> in press.

Phillimore, A.B., Leech, D.I., Pearce-Higgins, J.W., Hadfield, J.D., 2016. Passerines may be sufficiently plastic to track temperature-mediated shifts in optimum lay date. *Glob. Chang. Biol.* 24, 3259–3272.

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R.B., Peng, S., Poultier, B., 2015. Detection and attribution of vegetation greening trend in China over the last 30 years. *Glob. Chang. Biol.* 21 (4), 1601–1609.

Rasmussen, C., Williams, C., 2006. Gaussian processes for machine learning. The MIT Press.

Reed, B.C., 2006. Trend analysis of time-series phenology of north america derived from satellite data. *GIScience Remote Sens.* 43, 24–38.

Richards, L.A., Windsor, D.M., 2007. Seasonal variation of arthropod abundance in gaps and the understorey of a lowland moist forest in panama. *J. Trop. Ecol.* 169–176.

Richardson, A., Huffkens, K., Milliman, T., Frolking, S., 2018. Intercomparison of phenological transition dates derived from the phenocam dataset v1. 0 and MODIS satellite remote sensing. *Sci. Rep.* 8, 1–12.

Richardson, A.D., Anderson, R.S., Arain, M.A., Barr, A.G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K.J., Desai, A.R., Dietze, M.C., 2012. Terrestrial biosphere models need better representation of vegetation phenology: results from the North American carbon program site synthesis. *Glob. Chang. Biol.* 18 (2), 566–584.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agric. For. Meteorol.* 169, 156–173.

Richter, K., Hank, T.B., Mauser, W., Arzberger, C., 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. *J. Appl. Remote Sens.* 6, 1.

Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee, S., Dixon Hamil, K.A., Koenig, L. E., Zipkin, E.F., 2021. Working across space and time: nonstationarity in ecological research and application. *Front. Ecol. Environ.* 19, 66–72.

Schröter, D., et al., 2005. Ecosystem service supply and vulnerability to global change in Europe. *Science* 310, 1333–1337.

Schwartz, M., Reed, B., 1999. Surface phenology and satellite sensor-derived onset of greenness: an initial comparison. *Int. J. Remote Sens.* 20, 3451–3457.

Schwartz, M.D., Ault, T.R., Betancourt, J.L., 2013. Spring onset variations and trends in the continental united states: past and regional assessment using temperature-based indices. *Int. J. Climatol.* 33, 2917–2922.

Schwartz, M.D., Reiter, B.E., 2000. Changes in North American spring. *Int. J. Climatol. J. Roy. Meteorol. Soc.* 20 (8), 929–932.

Sekercioglu, C.H., Schneider, S.H., Fay, J.P., Loarie, S.R., 2008. Climate change, elevational range shifts, and bird extinctions. *Conserv. Biol.* 22 (1), 140–150.

Shen, M., Piao, S., Chen, X., An, S., Fu, Y.H., Wang, S., Cong, N., Janssens, I.A., 2016. Strong impacts of daily minimum temperature on the green-up date and summer greenness of the Tibetan Plateau. *Glob. Chang. Biol.* 22 (9), 3057–3066.

Soudani, K., Le Maire, G., Dufrêne, E., François, C., Delpierre, N., Ulrich, E., Cecchini, S., 2008. Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sens. Environ.* 112, 2643–2655. <https://doi.org/10.1016/j.rse.2007.12.004>.

Stanimirova, R., Cai, Z., Melaas, E.K., Gray, J.M., Eklundh, L., Jönsson, P., Friedl, M.A., 2019. An empirical assessment of the MODIS land cover dynamics and TIMESAT land surface phenology algorithms. *Remote Sens.* 11, 2201. <https://doi.org/10.3390/rs11192201>.

Steinbauer, M.J., 2011. Relating rainfall and vegetation greenness to the biology of spur-throated and Australian plague locusts. *Agric. For. Entomol.* 13, 205–218.

Stocker, T. and Eds., 2014. Climate change 2013: the physical science basis: Working group i contribution to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, V.

Swearingen, J., Bargeron, C., 2016. *Invasive plant atlas of the united states*. Cambridge University Press.

Tan, B., Morisette, J.T., Wolfe, R.E., Gao, F., Ederer, G.A., Nightingale, J., Pedelty, J.A., 2010. An enhanced TIMESAT algorithm for estimating vegetation phenology metrics from MODIS data. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 4, 361–371. <https://doi.org/10.1109/JSTARS.2010.2075916>.

Turner, D.P., et al., 2006. Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sens. Environ.* 102, 282–292.

Valtonen, A., Molleman, F., Chapman, C.A., Carey, J.R., Ayres, M.P., Roininen, H., 2013. Tropical phenology: Bi-annual rhythms and interannual variation in an Afrotropical butterfly assemblage. *Ecosphere* 4 (3), 1–28.

Van der Have, T., De Jong, G., 1996. Adult size in ectotherms: temperature effects on growth and differentiation. *J. Theoret. Biol.* 183, 329–340.

Visser, M., Both, C., 2005. Shifts in phenology due to global climate change: the need for a yardstick. *Proc. Roy. Soc. B: Biol. Sci.* 272, 2561–2569.

Visser, M., Gienapp, P., 2019. Evolutionary and demographic consequences of phenological mismatches. *Nat. Ecol. Evol.* 3, 879–885.

Wang, J., Rich, P.M., Price, K.P., 2003. Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *Int. J. Remote Sens.* 24, 2345–2364.

Wang, X., Piao, S., Xu, X., Ciais, P., MacBean, N., Myneni, R.B., Li, L., 2015. Has the advancing onset of spring vegetation green-up slowed down or changed abruptly over the last three decades? *Glob. Ecol. Biogeogr.* 24 (6), 621–631.

Wang, X., Zhou, Y., Wen, R., Zhou, C., Xu, L., Xi, X., 2020. Mapping spatiotemporal changes in vegetation growth peak and the response to climate and spring phenology over Northeast China. *Proc. Natl. Acad. Sci.* 12, 3977.

Way, D.A., Montgomery, R.A., 2015. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant Cell Environ.* 38 (9), 1725–1736.

Wells, C., Tonkyn, D., 2014. Range collapse in the Diana fritillary, *Speyeria diana* (Nymphalidae). *Insect Conserv. Diversity* 7, 365–380.

Wiens, J.J., 2016. Climate-related local extinctions are already widespread among plant and animal species. *PLoS Biol.* 14, e2001104.

Wu, T., Feng, F., Lin, Q., Bai, H., 2019. Advanced method to capture the time-lag effects between annual NDVI and precipitation variation using RNN in the arid and semi-arid grasslands. *Water* 11, 9.

Xie, Y., Ahmed, K.F., Allen, J.M., Wilson, A.M., Silander, J.A., 2015. Green-up of deciduous forest communities of Northeastern North America in response to climate variation and climate change. *Landscape Ecol.* 30, 109–123. <https://doi.org/10.1007/s10980-014-0099-7>.

Xie, Y., Wilson, A.M., 2020. Change point estimation of deciduous forest land surface phenology. *Remote Sens. Environ.* 240, 111698. <https://doi.org/10.1016/j.rse.2020.111698>.

Yang, L., Jin, S., Danielson, P., Homer, C., Gass, L., Case, A., Costello, C., Dewitz, J., Fry, J., Funk, M., Grannemann, B., Riggle, M., Xian, G., 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *J. Photogramm. Remote Sens.* 146, 108–123.

Yang, X., Mustard, J., Tang, J., Xu, H., 2012. Regional-scale phenology modeling based on meteorological records and remote sensing observations. *J. Geophys. Res. Biogeosci.* 117, G3.

Youngflesh, C., Socolar, J., Amaral, B., Arab, A., Guralnick, R., Hurlbert, A., LaFrance, R., Mayor, S., Miller, D., Tingley, M., 2021. Migratory strategy drives species-level variation in bird sensitivity to vegetation green-up. *Nat. Ecol. Evol.* 111, 1–8.

Zhang, X., Friedl, M., Schaaf, C., 2006. Global vegetation phenology from moderate resolution imaging spectroradiometer (MODIS): evaluation of global patterns and

comparison with in situ measurements. *J. Geophys. Res. Biogeosci.* 111 <https://doi.org/10.1029/2006JG000217>.

Zhang, X., Tarpley, D., Sullivan, J.T., 2007. Diverse responses of vegetation phenology to a warming climate. *Geophys. Res. Lett.* 34, 19.