Remote Sensing of Environment 229 (2019) 179–192

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Complex network-based time series remote sensing model in monitoring the fall foliage transition date for peak coloration

Chunyuan Diao

Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

ARTICLE INFO

Keywords:
Complex network
MODIS time series
Autumn phenology
Peak coloration
Deciduous forest

ABSTRACT

Vegetation phenological events, especially peak foliage coloration, are among the ecological phenomena that are most sensitive to climate change. Compared to spring seasonally recurring events, fall phenology remains much less understood. Remotely sensed monitoring of fall phenology provides a wealth of opportunities to understand the underlying processes and mechanisms. However, the gradual change of foliage color in the fall season makes it challenging to remotely estimate critical phenological transition dates. Particularly, the transition date for foliage peak coloration cannot be adequately captured via conventional curve fitting-based phenological models. Also the lack of consensus among the conventional models makes it desirable to explore new remotely sensed representations of the fall phenological process. In this study, we developed an innovative complex network-based phenological model, namely "pheno-network", to estimate the fall foliage transition date for peak coloration. The pheno-network model characterizes the phenological process through analyzing the collective changes of spectral signatures along the temporal trajectory. A network measure, moving average bridging coefficient, is newly designed to estimate the phenological transition date. With Harvard Forest and Hubbard Brook Forest as reference sites, the results demonstrated that the transition date estimated through the devised pheno-network model corresponds well with the peak coloration period of the reference sites. The unique structure of the pheno-network formulated via spectral similarities differentiates the various roles of vegetation spectral signatures at different phenological stages. This study is the first attempt at introducing network science to time series remote sensing in modeling the complex phenological processes of vegetation. The innovative network-based phenological representation shows great potential in improving remotely sensed phenological monitoring and shedding light on the subsequent modeling of vegetation phenological responses to climate change.

1. Introduction

Vegetation phenology plays a crucial role in shaping carbon, water, and energy cycles, and regulating ecosystem functions and biotic interactions (Morisette et al., 2009; Peñuelas and Filella, 2001; Richardson et al., 2013). This seasonal variation in vegetation foliage parameterizes land surface process models, numerical weather prediction models, and ecohydrological models for understanding land biophysicis, biogeochemistry, and ecosystem dynamic processes (Chen et al., 2016; Gutman and Ignatov, 1998). In temperate deciduous forests, vegetation phenology has been viewed as a first-order control on biosphere-atmosphere interactions, and is among the ecological phenomena that are most sensitive to climate and environmental changes (Keenan et al., 2014; Kramer et al., 2000; Xie et al., 2015). Over the past several decades, most attention has been devoted to exploring phenological events at the start of the growing season (e.g., budburst, leaf-out, and flowering) (Badeck et al., 2004; Clark et al., 2014; Polgar and Primack, 2011). In contrast, fall phenological responses to environmental changes remain poorly understood (Gallinat et al., 2015; Richardson et al., 2013; Xie et al., 2015). However, fall phenology regulates the length of photosynthetically active periods, mediates vegetation feedbacks to the climate system, and has significant implications for ecosystem management activities (Estiarte and Peñuelas, 2015; Garonna et al., 2014). There is increasing recognition that additional work is required for this area (Dragoni and Rahman, 2012; Richardson et al., 2013). As there are generally large intraspecies variabilities in leaf coloration at early senescence stages, the timing of peak foliage coloration has been found to be more responsive to environmental conditions (Delpierre et al., 2009; Jeong and Medvigy, 2014). According to United States Weather Channel (https://weather.com/) and field foliage network (https://www.foliagenetwork.com/), the peak foliage coloration denotes the stage when most/all leaves

E-mail address: chunyuan@illinois.edu.

https://doi.org/10.1016/j.rse.2019.05.003
Received 26 October 2018; Received in revised form 23 April 2019; Accepted 4 May 2019

change to yellow or red color, and is defined when the percentage of colored leaves reaches 60%–95% in this study. The peak foliage coloration stage is also of enormous economic importance for the multi-billion dollar fall foliage tourism industry in northeastern United States (Rustad et al., 2012). Remotely sensed monitoring of fall phenology, particularly peak foliage coloration, provides unprecedented opportunities to understand the underlying processes and mechanisms.

During the past decade, a variety of methods have been developed to quantify seasonal patterns of vegetation growth and senescence at regional or global scales using Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High Resolution Radiometer (AVHRR), and SPOT Vegetation (VGT). Typically, time series of satellite-derived vegetation index (e.g., normalized difference vegetation index [NDVI]) throughout the year is constructed to characterize phenological stages of vegetation. Satellite time series are mostly preprocessed using curve-fitting based phenological approaches to minimize the effects of cloud contamination and noise. Representative curve-fitting based approaches include double logistic functions (Beck et al., 2006; Zhang et al., 2003), asymmetric Gaussian functions (Jonsson and Eklundh, 2002), adaptive Savitzky-Golay functions (Jönsson and Eklundh, 2004), high order annual spline (Hermance et al., 2007), Harmonic analysis (Diao and Wang, 2014; Roerink et al., 2000), and wavelet analysis (Sakamoto et al., 2005). The smoothed time series can then be employed to estimate phenological transition dates (e.g., the start of the growing season and the end of the growing season), using threshold-based methods (Lloyd, 1990; White et al., 1997), inflection point methods (Moulin et al., 1997), change rate of curvature (Zhang et al., 2003), autoregressive moving average methods (Reed et al., 1994), etc. However, the phenological transition dates identified from different approaches can vary dramatically, which affects the subsequent modeling efforts and may yield conflicting results in tracking vegetation responses to climate change (White et al., 2009). Compared to spring phenology, the gradual change of foliage color in autumn makes it more challenging to capture critical phenological transition dates. Further, previous investigations of fall phenology have mainly focused on the onset of leaf senescence or dormancy (Xie et al., 2015; Zhang et al., 2003), few efforts have been dedicated to the peak coloration stage of foliage. Zhang and Goldberg (2011) developed a temporally-normalized brownness index to measure the status of leaf senescence by combining a linear mixture model with satellite NDVI time series. It assumes that NDVI in a pixel can be linearly mixed by that of land surface components (i.e., green and brown materials). Yet the linear mixture model assumption of NDVI may be violated, especially in large-scale mapping.

Remotely sensed phenological monitoring has mainly been conducted by tracking the temporal trajectory of vegetation index (e.g., NDVI and enhanced vegetation index [EVI]). To date, a wide variety of vegetation indices have been developed, with each emphasizing a particular vegetation property. It is often challenging to select an appropriate one, as the use of vegetation index may be limited by its sensitivity to soil and atmospheric conditions, and saturation in dense vegetation areas. Delbart et al. (2005) used NDVI and normalized difference water index (NDWI) to capture the onset of leaf senescence in boreal regions, and found that these indices were both in poor agreement with in situ measurements. Hufkens et al. (2012) found that estimation of phenological transition dates was largely affected by the selection of vegetation indices (i.e., NDVI, EVI, and excess green index) in deciduous forest areas. The leaf phenological process is complex with the concurrent changes of a multitude of vegetation properties (e.g., canopy pigment content, cell structure, and water content). Those vegetation properties can be characterized by a wide range of spectral wavelengths (e.g., canopy pigment content from visible bands, cell structure from near-infrared bands, and water content from shortwave infrared bands). Thus tracking the collective change of spectral signatures, with consideration of combining dynamics of those vegetation properties, may provide an alternative solution to model the complex leaf phenological processes. Diao and Wang (2018) developed a Multiyear Spectral Angle Clustering (MSAC) model to analyze the temporal patterns of spectral signatures for constructing a composite image, where each pixel was acquired from the leaf senescence stage to facilitate the detection of invasive species. Yet time series of this expanded range of spectral information needs to be further explored in characterizing vegetation phenological stages.

The recent development of network science has revolutionized our understanding of complex systems (Barabási, 2009; Borgatti et al., 2009; Newman, 2003; Strogatz, 2001). Network science focuses on the behaviors and dynamics of complex networks, and has gained a significant use in a large number of fields, including physical, biological, and social sciences. A complex network is often represented as a graph with a set of nodes interconnected together in pairs by edges, such as social networks, information networks, and biological networks (Eagle et al., 2009; Isalan et al., 2008; Onnela et al., 2011; Pastor-Satorras et al., 2014; Rubinov and Sporns, 2010). It provides an abstract network representation capturing the most fundamental patterns and structures of a complex system (Newman, 2010). Through examining the structure of relationships between their constituents, complex networks present a new viewpoint of predicting the behaviors of complex systems (Albert and Barabási, 2002). The flexibility of constructing the networks shows great potential to model the complex phenological process and to estimate the critical phenological transition dates. The structure of collections of spectral signatures, characterized by complex networks, may further our understanding of plant phenological development along the temporal trajectory. However, the network representation has seldom been studied in time series remote sensing for phenological estimations.

The objective of this research is to develop an innovative complex network-based phenological model, namely “pheno-network”, to estimate the fall foliage transition date for peak coloration. Network science, for the first time, is introduced into time series remote sensing to construct the pheno-network model. Through uncovering the temporal relationships between spectral signatures, the pheno-network model provides a new prospective in analyzing the temporal trajectory of plant growth and development. It offers an innovative network-based representation to estimate critical phenological transition dates, which can further shed light on how vegetation phenology responds to climate change and environmental stress conditions. This study uses MODIS time series data spanning from 2002 to 2013 to demonstrate this new concept in estimating autumn peak coloration of foliage, with field reference data from Harvard Forest and Hubbard Brook Forest.

2. Data and methods

2.1. Field measurements of fall foliage peak coloration

To demonstrate the concept of complex networks in estimating fall foliage peak coloration, this study focuses on two temperate deciduous forest regions (i.e., Harvard Forest and Hubbard Brook Forest), of which abundant field measurements of vegetation phenophases have been collected in a systematic manner. Harvard Forest is located in central Massachusetts, the northeastern United States (42.54°N, 72.18°W, and 335 to 365 m elevation). The climate is moist temperate, with annual mean precipitation of 110 cm. Mean July temperature of the site is 20 °C, and mean January temperature is −7 °C. Harvard Forest is a long term ecological research site, in which spring and autumn phenology of a suite of woody plant species has been recorded since 1991. For our studying periods (2002 − 2013), the percentage of colored leaves and the percentage of fallen leaves for 14 dominant species in the fall phenology were observed at 3–7 day intervals (http://harvardforest.fas.harvard.edu/data-archives). The 14 species were sugar maple (*A. saccharum*), red maple (*Acer rubrum*), striped maple (*A. pensylvanicum*), red oak (*Q. rubra*), black oak (*Q. velutina*), white oak (*Quercus alba*), black birch (*B. lenta*), paper birch (*B. papyrifera*), yellow birch (*Betula alleghaniensis*), white ash (*Fraxinus americana*), black cherry (*Prunus*

serotina), beech (*Fagus grandifolia*), black gum (*Nyssa sylvatica*), and shadbush (*Amelanchier laevis*). Those phenological observations were collected for two to five individuals of each dominant species located within 1.5 km of the Harvard Forest headquarters, with a total of 57 individuals. For most of dominant species, both overstory and understory individuals were represented in the collected data. To match with satellite-derived phenological measures, the species-based field observations were aggregated to the community level. The average percentages of leaf coloration and leaf fall were calculated at each observation date. Further, the temporal percentage of colored leaves between observation dates was linearly interpolated to quantify the extent of foliage coloration and to represent the continuous leaf senescence process.

Hubbard Brook Forest is located within the White Mountain National Forest in central New Hampshire, the northeastern United States (43.95°N, 71.70°W, and 222 to 1015 m elevation). The climate is moist temperate, with annual mean precipitation of 140 cm. Mean July temperature of the site is 18 °C, and mean January temperature is −9 °C. Hubbard Brook Forest is within the long term ecological research network. Spring and fall phenological observations in the site have been collected since 1989 for three dominant species, namely sugar maple (*Acer saccharum*), yellow birch (*Betula alleghaniensis*), and American beech (*Fagus grandifolia*). The phenological status of those species was recorded at 7 day intervals, at nine locations with varying topographical and watershed characteristics (http://hubbardbrook.org/data/dataset.php?id=51). Three representative trees of each species were surveyed at each location. Unlike the continuous percentage of foliage coloration recorded at Harvard Forest, the foliage coloration at Hubbard Brook Forest was measured according to the stage of senescence with an index value (Table 1). At each location, the index was averaged for the nine trees (i.e., three trees for each dominant species) for each observation date. The average phenological index was used to represent the community-level leaf coloration extent. Besides, the index of colored leaves between observation dates was linearly interpolated to represent the continuous leaf senescence process.

To date, most fall phenological studies have focused on documenting intra- or inter-annual variations in the onset of leaf senescence and dormancy. These two phenophases estimated using time series of EVI were also provided in the MODIS Land Cover Dynamic product (MCD12Q2). However, fall phenology, ranging from scattered leaves changing color to all leaves falling down, is a gradual senescent movement process. Multiple foliage coloration stages, including little/no change, low coloration, moderate coloration, peak coloration, and post-peak coloration, have been defined by United States Weather Channel and field foliage network. The peak coloration stage, focused in this study, is defined by synthesizing those two sources. At Harvard Forest, the foliage peak coloration stage is defined when the percentage of colored leaves reaches 60%–95%. It usually lasts about two weeks. At Hubbard Brook Forest, the peak coloration stage is defined when the phenological index value achieves 1–2 (i.e., most/all leaves becoming yellow or red, with a few to half of fallen leaves). To assess the capability of the pheno-network model in predicting the satellite-derived peak foliage coloration dates, the surveyed field peak coloration dates at Harvard Forest and Hubbard Brook Forest will be compared to those estimated from the MODIS time series from 2002 to 2013.

Table 1
Definition of phenological index of foliage coloration at Hubbard Brook Forest[a].

Phenological index	Description
4	Only scattered leaves or branches have changed color
3	Many leaves have noticeable reddening or yellowing color
2	Most leaves have become yellow or red, with a few fallen leaves
1	No more green leaves on the canopy. Half of the leaves have fallen
0.5	Most leaves have fallen
0	All leaves have fallen, except remnants on beech

[a] Adapted from routine seasonal phenology measurements at Hubbard Brook Forest (http://hubbardbrook.org/data/dataset.php?id=51).

2.2. MODIS data and pre-processing

Time series of MODIS Nadir Bidirectional reflectance distribution function Adjusted Reflectance (NBAR) data (MCD43A4 H12V04, version 6, with a spatial resolution of 500 m) from 2002 to 2013 were acquired from Land Processes Distributed Active Archive Center (LP DAAC). MCD43A4 utilizes a bidirectional reflectance distribution function (BRDF) to model the reflectance under a nadir view, accounting for view and solar geometries. It is a daily 16-day product, of which surface reflectance at each date is determined based on a 16-day retrieval period (including its eight preceding days and seven succeeding days). For each date, the reflectance value is estimated by weighting all acquisitions from Terra and Aqua sensors during the retrieval period. The estimated reflectance value has the advantage of reducing the influence of atmospheric interference and noise, and is taken as the best representative value for each observation date. With its daily optimized observation, MCD43A4 time series facilitates the remotely sensed estimation of key phenological transition dates of vegetation. Each year, MCD43A4 reflectance data covering fall phenophases (starting from the date when NDVI begins to decrease) were used to estimate the phenological transition dates for fall foliage peak coloration. Seven spectral bands (bands 1–7), ranging from visible to shortwave infrared regions, were obtained. These spectral bands, characteristic of leaf color, leaf cell structure, and leaf water content of vegetation, were used to construct the pheno-network model.

Two quality assurance layers were employed to pre-process the time series of MCD43A4 reflectance data. First, the MODIS BRDF/Albedo Quality data (MCD43A2, version 6) provide a snow and ice quality layer of the MCD43A4 reflectance data. It denotes if the MCD43A4 reflectance is retrieved on a snow-free or snow-covered surface. MCD43A2 was used to flag the pixels that were contaminated by snow for each observation date. Second, the MODIS land surface temperature (LST) data (MOD11A1, version 6) provide a daytime surface skin temperature layer to locate winter periods when snow cover could appear. MOD11A1 records daytime LST on a daily basis at a spatial resolution of 1000 m. It was spatially resampled to 500 m using a nearest neighbor algorithm to match with that of the MCD43A4 reflectance data. The missing values of LST were temporally replaced by the moving average of the nearest preceding and subsequent values. These two quality control layers were considered together to distinguish the irregular outlying MCD43A4 reflectance in the time series. Specifically, a pixel would be flagged as invalid at each observation date if it was marked as a snow-covered one by MCD43A2, or if its daytime LST was less than 5 °C (Zhang et al., 2006; Zhang and Goldberg, 2011). Invalid reflectance observations were then replaced by the moving average of the nearest preceding and succeeding good quality neighbors in the MCD43A4 time series. Consequently, the pre-processed MCD43A4 reflectance data smoothed out abrupt irregular changes caused by snow and cloud contamination, and were more robust to outlying observations in the time series.

The MODIS Land Cover Type data (MCD12Q1, version 5.1) provide annual global maps of land cover at 500 m spatial resolution in a variety of classification schemes (e.g., International Geosphere Biosphere Programme [IGBP] global vegetation classification scheme, University of Maryland scheme, and MODIS-derived Net Primary Production (NPP) scheme). IGBP global vegetation classification scheme is a primary land cover scheme that includes 17 land cover classes (i.e., 11 natural vegetation classes, 3 developed land classes, and 3 non-vegetation classes). It was adopted in this study to locate areas covered by deciduous broadleaf forest (IGBP class label 4) and mixed forest (IGBP class label 5). Fall phenology and foliage peak coloration are of

particular interest and importance for these two land cover classes. The quality control layer of the MCD12Q1 assesses the confidence level of land cover classification, and only the pixels with the classification confidence greater than 50% for all the years were kept in this study. For those deciduous broadleaf and mixed forest areas within the extent of the MODIS H12V04 tile, the spatial distributions of mean and standard deviation of fall foliage peak coloration dates throughout 2002–2013 would be mapped using the proposed model.

The MODIS Land Cover Dynamic data (MCD12Q2, version 5) provide annual global estimation maps of the timing of vegetation phenology (e.g., vegetation growth, maturity, and senescence) at 500 m spatial resolution. The seasonal phenological dynamics of vegetation were estimated through a logistic function fitted to time series of smoothed EVI (Ganguly et al., 2010; Zhang et al., 2003). The phenological measures of MCD12Q2 relevant to fall phenology include the onset date of EVI decrease and the onset date of EVI minimum. These phenological transition dates correspond to the timing of vegetation senescence and dormancy, respectively. As a global vegetation phenological product, MCD12Q2 may be a good source for defining a temporal constraint range for estimating the peak coloration dates, which may further reduce the influence of outlying observations in the time series. In this study, the role of this phenological constraint would be evaluated on the estimation results of the pheno-network model.

2.3. Complex network-based time series remote sensing model

In this study, we developed an innovative complex network representation of satellite time series to estimate the fall foliage transition date for peak coloration. Specifically, the complex network is constructed on a per-pixel basis with the spectral signatures acquired during the fall phenology. The spectral signatures of a pixel taken on each day are represented as nodes (i.e., spectral nodes) and the similarities between the nodes are represented as edges (Section 2.3.1). Based on this representation, we model the collective behavior of the fall phenological progress through network measures (Section 2.3.2).

2.3.1. Innovative complex network representation of fall phenological progress

The proposed model constructs an undirected network for each pixel based on its spectral signatures acquired during the fall phenology of a year (referred to as a pheno-network). Specifically, the spectral signature of a pixel obtained on each day of the fall phenology is represented as a spectral node. Those spectral nodes that share similar spectral signatures are connected using edges. For a pair of nodes, the similarity between their spectral signatures is measured using a cosine distance function (Salton and McGill, 1986):

$$\mathrm{D}(A,B) = cos^{-1}\left(\frac{A \bullet B}{\|A\|\,\|B\|}\right) \tag{1}$$

where A and B are vectors representing spectral signatures of the pixel obtained on two different days, and $D(A,B)$ is the cosine distance between A and B. $\|A\|$ and $\|B\|$ are the norms of the vectors A and B, respectively. The cosine distance is shorter between nodes with more similar spectral signatures. The cosine similarity is selected in this study, because it measures the similarity of orientation between the vectors representing spectral signatures, and is invariant to the magnitude difference. Therefore, the cosine similarity can help reduce the influence of albedo and illumination effects.

The way to formulate the relationships between pairs of nodes according to their spectral similarities determines the structure of a pheno-network. In the pheno-network representation, edges between spectral nodes with shorter cosine distance are more likely to be edges between nearby dates, and thus provide more information about the temporal dynamics of phenological changes. Based on this assumption, we define a critical control parameter, called "tolerance value" (d_0), to retain edges with short cosine distance and filter out edges with long cosine distance. When constructing the pheno-network representation of the fall phenological process, only those pairs of nodes with a cosine distance shorter than d_0 are connected by edges. The structure of the network differs with different tolerance values. A proper d_0 value is essential in constructing a pheno-network that could capture the relationships between the nodes. On one hand, if d_0 is too large, all nodes in the network are connected, resulting in a fully connected network. Such a network cannot differentiate the role of different nodes. On the other hand, if d_0 is too small, the resultant network may be too sparse to be representative of the relationships between the nodes. To handle this issue, a tolerance selection algorithm is devised and will be discussed in Section 2.3.3 to help select proper d_0 values. For the following discussion, we assume an arbitrary d_0 is used to construct the pheno-network.

2.3.2. Network measures of fall foliage peak coloration

Characterized by canopy pigment content, cell structure, and water content, the spectral signature of forest canopy usually changes gradually before and after the phenological transition period, but more dramatically during the transition period in the fall. From a network perspective, those nodes representing days before the transition period share similar spectral signatures and tend to form a densely connected group among themselves, the "pre-transition group". Those nodes representing days after the transition period would show similar characteristics and tend to form the "post-transition group", but these two groups would have different spectral signatures and appear to be apart in the network. As the change of spectral signature during the fall phenology is a continuous process, those nodes representing days during the transition period would form the "transition group", which stands between the pre- and post-transition groups.

In the pheno-network, those nodes representing the late days in the pre-transition group would share similar spectral signatures with those nodes representing the early days in the transition group. Similarly, those nodes representing the late days in the transition group would share similar spectral signatures with those nodes representing the early days in the post-transition group. Both the pre- and post-transition groups are connected to the transition group, but the edges between different groups are far fewer than the edges within each group. There would be few direct edges between the pre- and post-transition groups, but only the transition group serves as a bridge connecting both the pre- and post-transition groups. The spectral signature of forest canopy changes more dramatically during the transition period, compared to that during the pre- and post-transition periods. Hence, the edges would be most dense within the pre- and post-transition groups, and most sparse within the transition group. Fig. 1 shows an example of a pheno-network with three groups. Fig. 1 provides a visualization of a pheno-network from a network view, using the force-directed graph-drawing algorithm (Fruchterman and Reingold, 1991). The algorithm visualizes a network in two-dimensional space by assigning spring forces among its nodes based on their relative positions. The forces include attractive forces between directly connected nodes and repulsive forces between all nodes. The network is projected to the two-dimensional space by simulating the motion of its nodes and edges using these forces. While the axes in Fig. 1 bare no ecological meaning, the visualization of the pheno-network illustrates the relationships between spectral nodes and the changes of spectral signatures over time.

Based on these characteristics of the pheno-network, this study devises an innovative bridging coefficient to identify nodes within the transition group for estimating the critical phenological transition period in the fall season. For a node i, its bridging coefficient, noted as bc_i, is defined as the ratio of its betweenness centrality to clustering coefficient:

$$bc_i = \frac{b_i}{c_i} \tag{2}$$

where b_i and c_i are the betweenness centrality and clustering coefficient of node i, respectively. A pair of nodes in a network could be connected

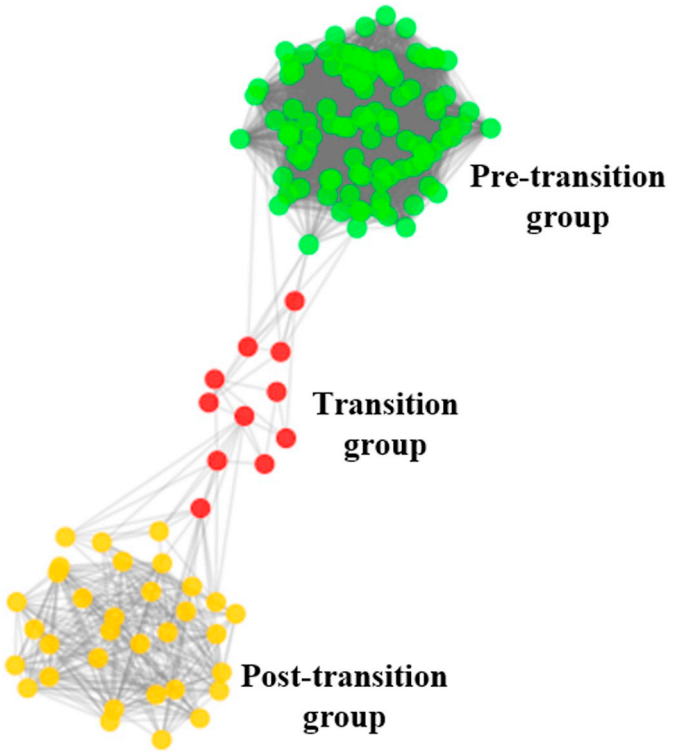

Fig. 1. Pheno-network of fall phenological progress with three node groups.

by either a direct edge or a chain of edges (referred to as "paths"). There may be multiple paths between a pair of nodes, with each path consisting of different number of edges. The path with the lowest number of edges is called the shortest path between the node pair. The betweenness centrality of node i is the fraction of all shortest paths between all pairs of nodes that pass through node i (Freeman, 1978):

$$b_i = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{3}$$

where σ_{st} is the total number of shortest paths between a pair of nodes s and t, and $\sigma_{st}(i)$ is the number of these paths that pass through node i. The betweenness centrality represents the degree to which a node stands between all other nodes. A node with a high betweenness centrality value serves as a hub to connect other nodes together in the network. A number of studies have used betweenness centrality to identify critical nodes in various networks (Newman, 2010). For example, nodes with the highest betweenness centrality value in a social network are considered to be individuals who connect different social groups. These individuals always receive the most attention in network-based intervention strategies against infectious diseases (Gao et al., 2016; Luo et al., 2018; Valente, 2012). As discussed above, all paths connecting the pre- and post-transition groups have to go through the transition group. Nodes in the transition group have the highest betweenness centrality in the pheno-network.

Besides being the hub of the pheno-network, the transition group has another important characteristic, i.e. edges are sparse within the group. Therefore, we combine the betweenness centrality with another network measure, clustering coefficient, to help identify the transition nodes. The clustering coefficient of a node is given by the total number of edges between its neighbors divided by the maximum number of edges that could possibly exist between the neighbors (Watts and Strogatz, 1998):

$$c_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)} \tag{4}$$

where nodes j and k are any two neighbors of node i and e_{jk} is the edge between these two nodes, and k_i is the total number of neighbors of node i. The clustering coefficient measures the degree to which neighbors of a node cluster together. For a node with a low clustering coefficient value, its neighbors are sparsely connected between themselves. Given the structure of the pheno-network, nodes in the transition group tend to have the highest betweenness centrality and relatively low clustering coefficient, which results in the highest bridging coefficient.

The spectral signature of a pixel on each observation day is acquired from the daily 16-day MODIS MCD43A4 product in this study. Despite the spectral signature on each date being smoothed and optimized over its 16-day retrieval period, the node-based representation of spectral signatures via the pheno-network may be subject to the influence of irregular (or outlying) spectral signatures. A node with irregular (or outlying) spectral signature may yield a high bridging coefficient if its spectral signature is similar to that of phenological transition dates. To reduce the influence of irregular and outlying spectral signatures, a seven-day moving average window is applied to this network representation to calculate the moving average of bridging coefficients for each observation day. Specifically, for a given day, the moving average is calculated as the average of the bridging coefficients for a seven day moving retrieval period (including three preceding days and three succeeding days). As there is no bona fide boundary of the transition period, we focus on the identification of the most critical phenological transition date, the date with the highest moving average bridging coefficient. By adopting the moving average, the network-based phenological model is expected to provide a more robust representation of the fall phenological process.

2.3.3. Tolerance selection algorithm

As discussed in Section 2.3.1, different tolerance values d_0 may yield different pheno-networks. We devised a tolerance selection algorithm to help select proper tolerance values. The algorithm starts by creating candidate edges between all possible pairs of nodes. For each pair, the candidate edge between the nodes is labelled by the cosine distance between them. All candidate edges are sorted based on their associated cosine distance. The tolerate selection algorithm will test a range of d_0 values for constructing the pheno-networks. In this study, we chose the 5th, 10th, 15th, …, 95th percentiles of the cosine distance values associated with all candidate edges. For example, when d_0 is set to the 5th percentile, the 5% of candidate edges with the shortest cosine distance are selected and added to the pheno-network. In the case that a small value is chosen as the tolerance value, the resultant network may be highly fragmented. To avoid this situation, an additional constraint on the size of the "giant component" is designed.

Networks typically contain a number of connected components, each consisting of a subset of nodes. Any two nodes within a connected component can be connected through either a direct edge or a path between them. In contrast, all connected components are disconnected from each other, meaning that there is no edge or path connecting nodes in different components. Among the connected components a network may have, the one with the largest number of nodes is defined as the giant component. The number of nodes in the giant component, as a fraction of the total number of nodes in the network (S), is an important measure for the cohesiveness of the network (Newman, 2010). In real world, functional components of a complex system are mostly interrelated, and thus networks representing these systems usually have high S values, mostly over 50% and sometimes over 90% (Gao and Bian, 2016; Newman et al., 2011; Newman, 2010). For pheno-networks, the change of the spectral signatures of a pixel throughout the year is a continuous process. It is reasonable to assume that the majority of the nodes in the pheno-network can be connected through edges or paths, and the S value would be relatively high. Following this principle, the tolerance selection algorithm will discard any d_0 value that will yield a network with an S value lower than 85%. This constraint ensures that the resulting pheno-network is cohesive with a large

giant component. The required S value is not too high to allow the algorithm certain tolerance to missing and outlying data.

For a tolerance value that satisfies the above constraint, the moving average bridging coefficient will be calculated for all nodes, and the one with the highest value will be identified. The algorithm then compares all identified nodes among all pheno-networks and selects the node with the highest moving average bridging coefficient as the critical transition date. With adequate computing resources, this algorithm could potentially be configured into an exhaustive search on all possible d_0 values and produces best predictions of the transition dates.

2.4. Accuracy assessment

The network-retrieved phenological transition dates would be compared to field-based phenological measurements at Harvard Forest and Hubbard Brook Forest from 2002 to 2013. To match the field measurements with satellite pixels, the foliage coloration extents of surveyed trees species at each forest were averaged to represent the foliage status of the vegetation community. It is noted that the foliage coloration at the community level may be more appropriately represented using the areal average if the species abundance and coverage are known (Liang et al., 2011). The field dataset at Harvard Forest was mainly used for the accuracy assessment purpose, as it recorded the detailed leaf development status of representative species in the vegetation community during the fall phenology. The dataset at Hubbard Brook Forest covered a variety of watersheds and topography, and was less accurately measured in the field. Thus it was only used to roughly assess the phenological transition dates retrieved from the pheno-network model.

Several network-based measures were assessed using the field observations. First, the phenological transition date detected through the devised bridging coefficient was compared to that from the betweenness centrality, in which the node with the maximum moving average betweenness centrality was utilized to estimate the transition date. As the bridging coefficient is built upon the betweenness centrality, and the betweenness centrality evaluates the role of a node standing between all other nodes in a network, the comparison between these two network measures would further our understanding of pheno-network models in retrieving critical phenological transition dates. Second, the network measures were assessed according to the size of moving window. The structure of the devised pheno-network may be affected by nodes of irregular spectral signatures. To investigate the influence of those observations, a series of moving windows with sizes ranging from one to seven days were designed to calculate the moving average bridging coefficient and to estimate the corresponding phenological transition dates. Third, the role of the phenological constraint, defined through the MCD12Q2 product, on the network estimation results was evaluated. MCD12Q2 records the onset dates of EVI decrease and EVI minimum, indicative of the range of fall phenological process. This phenological constraint assumes that the peak coloration transition date lies between the onset of the leaf senescence stage and the onset of the dormancy stage. With the phenological constraint, the node with the highest moving average bridging coefficient within this temporal range was captured to estimate the transition date. Fourth, the network-based phenological estimation results for both the pixel of Harvard Forest headquarters and the pixels within the $6 * 6$ window (about 1.5 km) of the headquarters were assessed. Due to the geo-location uncertainty, the pixel of Harvard Forest headquarters at 500 m may not match well with field survey geographical coverage area. The mean and standard deviation of the estimated transition dates of its $6 * 6$ pixel window were calculated.

Finally, the network-based measures were compared with the conventional MODIS NDVI time series-based phenological measures. The NDVI time series was constructed using the MCD43A4 data and smoothed through following the MCD12Q2 protocol (Ganguly et al., 2010). Double logistic function, as one of the most widely used phenological curve fitting methods, was employed to further smooth the NDVI time series in tracking the seasonal phenological patterns of vegetation. The phenological transition dates were estimated using a variety of conventional phenophase detection methods, including the threshold-based method, the curve derivative method, and the curvature change rate method based on the fitted curve (White et al., 2009; Zhang et al., 2003). The threshold-based method estimates the phenological transition dates through a user-defined threshold of seasonal development of NDVI. A multitude of threshold values were experimented in the study and the 50% of the amplitude of NDVI was selected as the threshold to retrieve the peak coloration dates. The curve derivative method calculates the local extremes in the first derivative of the fitted curve for phenological estimations. The curvature change rate method captures the transition dates by calculating the local extremes in the rate of curvature change of the fitted curve. The correlation coefficient (R) and mean absolute difference (MAD) were utilized to evaluate the performance of these phenological methods. The comparison between the pheno-network model and conventional phenological measures would provide insight into the role of network science in modeling the remotely sensed phenological process.

3. Results

3.1. Pheno-network of fall phenological process

The rationale for modeling the fall phenological process through pheno-networks is that the structure of relationships between spectral nodes can infer the phenological transition dates. As the tolerance value determines how the spectral nodes are connected to each other and the subsequent network structure, selecting an appropriate tolerance value is critical. In this study, the role of the tolerance value was demonstrated using the MODIS pixel that covers the Harvard Forest field site (Fig. 2). For this MODIS pixel, its spectral signatures acquired from day of year (DOY) 180 to 365 of year 2010 were represented as nodes in the network. Three tolerance values were considered, including the 10th, 40th, and 70th percentiles of the cosine distance values among all

(a) 10th percentile of the cosine distance values (b) 40th percentile of the cosine distance values (c) 70th percentile of the cosine distance values

Fig. 2. Pheno-networks constructed with different tolerance values.

candidate edges (as discussed in Section 2.3.3). For instance, the tolerance value of the 10th percentile of the cosine distance values denotes that the 10% of candidate edges with the shortest cosine distance are utilized to construct the pheno-network. For the low tolerance value, nodes formed a number of small connected components, including isolated singletons (components with one node). The resultant network structure was too fragmented to infer the phenological process (Fig. 2a). As more edges were added to the network with the increasing tolerance value, the connected components (including singletons) began to merge with each other. The giant component, whose size became significantly larger than that of the others, began to appear. Yet when the tolerance value was too high, almost all the nodes in the network were connected to each other. This dense network structure made it difficult to differentiate the roles of spectral nodes in the network (Fig. 2c). With the tolerance selection algorithm, the formulated pheno-network that could best represent the fall phenology was shown in Fig. 2b. Three node groups could be roughly identified, with the transition group serving as the hub linking other two groups together.

Several network measures were calculated based upon the formulated network to uncover the underlying phenological process, including betweenness centrality, clustering coefficient, and bridging coefficient. As an example, those network measures of the pixel of Harvard Forest headquarters for the year of 2010 were shown in Fig. 3. Ranging from DOY 180 to 365, the nodes with the high betweenness centrality values were located within the temporal period of DOY 290 to 300. It indicated that the spectral signatures within this range played a central role in connecting all the spectral signatures in the fall season, and had significant controls on the spectral dynamic changes transitioning from one period to another. The clustering coefficient, as a local measure, quantified the degree to which neighbors of a node clustered together. Due to the drastic change of spectral signatures during the transition period, only transition nodes in nearby dates shared spectral similarity. This was represented as the sparse edges between the transition nodes and their neighbors.

The identified nodes with high bridging coefficient values were similar to those with high betweenness centrality values. Yet relative rankings of these nodes via those two measures slightly differed, as the bridging coefficient also accounted for the extent of neighborhood connections of a node. The grey-shaded region (DOY 282 to 298) denoted the field-observed peak coloration period at Harvard Forest when the mean percent of colored leaves achieved 60% to 95% in 2010. All the high betweenness centrality and bridging coefficient values lay within this temporal period, indicating that the spectral signatures during the peak coloration period served as the transition nodes in the constructed pheno-network.

3.2. Field evaluation of fall foliage coloration

Field observations of fall phenology at Harvard Forest were typically conducted from DOY 240 to 330. The leaf foliage of tree species started to change color in early September, achieved the half of maximum coloration around early October, and continued to full coloration around mid-November. The mean and standard deviation of the percentage of colored leaves of 57 sampling trees along the fall phenological trajectory were shown in Fig. 4. The mean percentage of colored leaves generally followed a logistic curve for the mapping years, with a gradual change at the beginning and end of the coloration process and a more drastic change in the middle. During the fall phenological process, phenological variation in foliage coloration existed both within and between tree species in the community. The standard deviation of the colored leaf percentage was relatively low (below 10%) in early September and mid-November, and then increased to about 30% around mid-October due to a wider leaf coloration extent of tree species. This large variation indicated that some tree species changed color considerably while other species might still maintain green color. The timing to achieve a certain percentage of colored leaves also varied across individual trees. For instance, the timing that an individual tree attained 70% of leaf coloration spanned from DOY 268 to 301 in 2004. For some species, this might differ slightly from DOY 283 to 289 (e.g., *A. pensylvanicum*), while other species might have a larger variation ranging from DOY 268 to 288 (e.g., *Fraxinus americana*). The average timing of 70% of leaf coloration differed inter-annually at Harvard Forest, with the standard deviation of the timing across years being 3.4 days. Similarly, the timing of leaf coloration extent varied across trees at Hubbard Brook Forest. The standard deviation of the timing corresponding to the phenological index 2 across years was 3.13 days.

3.3. Network measures of fall phenological transition dates

For the constructed pheno-networks of the pixel of Harvard Forest headquarters, the bridging coefficient and betweenness centrality values of each observation date from 2002 to 2013 were shown in Fig. 5. The grey-shaded regions in the figure denoted the field-observed peak coloration period (about two weeks). Almost all the maximum bridging

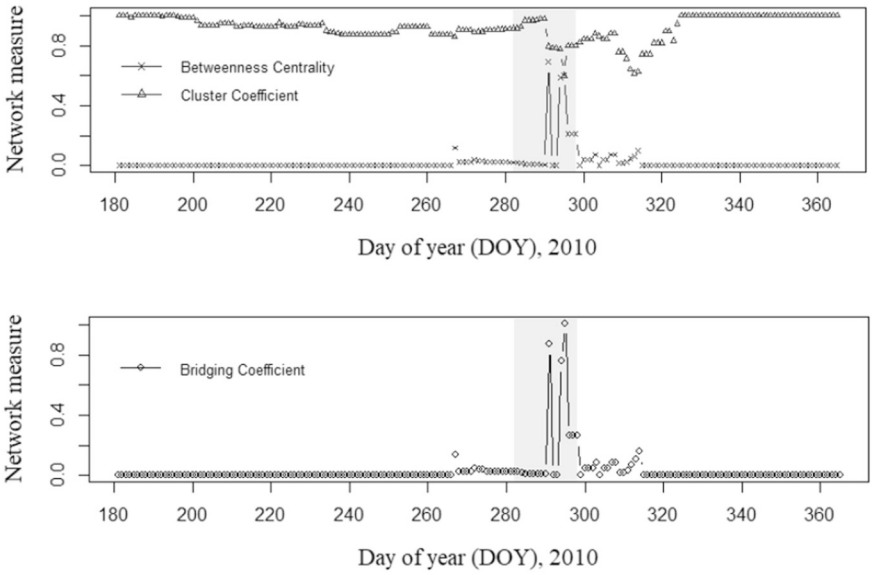

Fig. 3. Network measures of the fall phenology of the pixel of Harvard Forest headquarters in 2010.

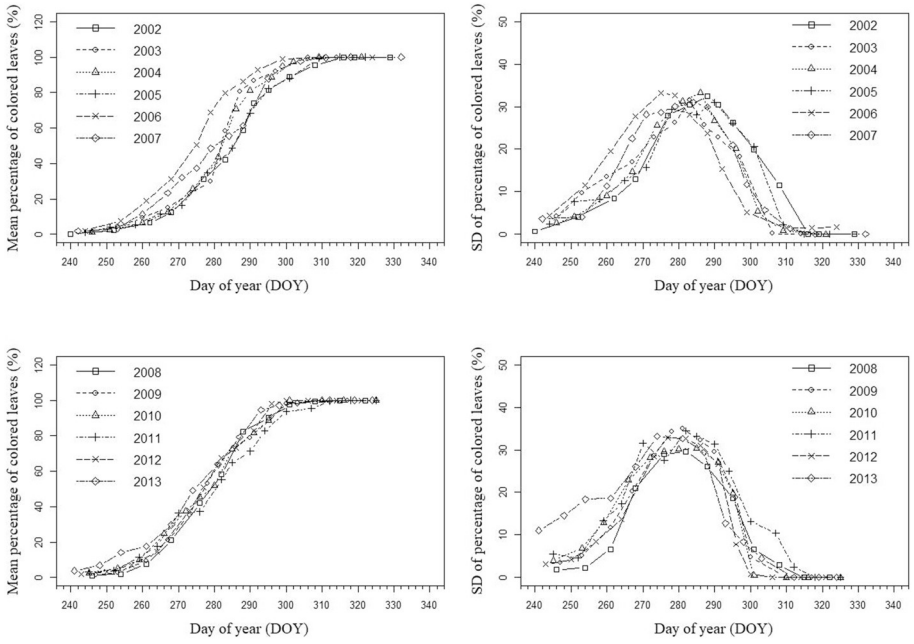

Fig. 4. The mean and standard deviation of the percentage of colored leaves at Harvard Forest.

coefficient and betweenness centrality values were located within the grey-shaded peak coloration regions, which indicated the central role of spectral signatures in this region in connecting all the spectral nodes within the network.

The measures of bridging coefficient and betweenness centrality were compared in Fig. 5. Comparable betweenness centrality values might exist on multiple observation dates along the fall phenological trajectory. For example, the betweenness centrality values on DOY 276 and 292 were close in 2008, and so were the betweenness centrality values on DOY 292 and 315 in 2011. The betweenness centrality measure was able to identify the spectral signatures that had more controls in linking the spectral information within the network, but might not be sufficiently appropriate to capture the transition nodes connecting one phenological stage with another. Of those dates with comparable betweenness centrality values, the ones with higher bridging coefficient values were more inclined to lie within the grey regions. As an integrated measure of betweenness centrality and clustering coefficient, the bridging coefficient weighted the connection effects of neighboring nodes to facilitate the identification of nodes serving as the transition hub between different phenological stages. The patterns of spectral nodes during the fall phenological process and the structure of the corresponding network could be captured. Almost all

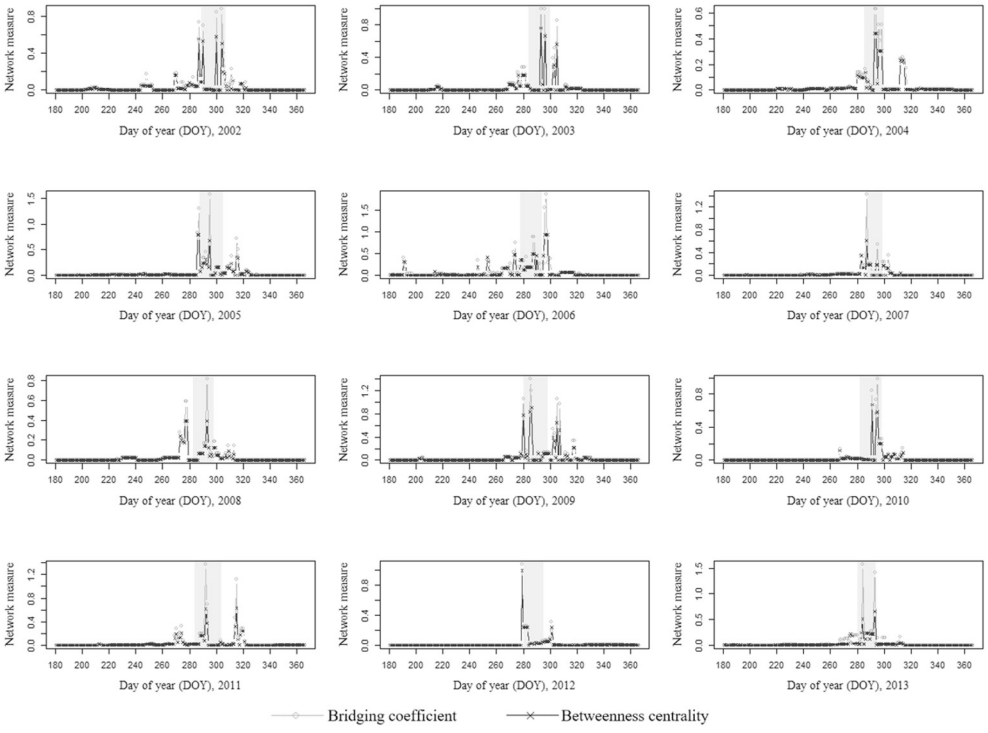

Fig. 5. The bridging coefficient and betweenness centrality values of each observation date from 2002 to 2013 for the pixel of Harvard Forest headquarters.

the phenological transition dates, estimated through the maximum bridging coefficient, lay within the field-observed peak coloration periods at Harvard Forest from 2002 to 2013. Despite the importance of betweenness centrality in evaluating the role of a node standing between all other nodes, the comparison between these two network measures indicated that bridging coefficient may be more appropriate in retrieving critical phenological transition dates and modeling the fall phenological process.

Despite the good performance of bridging coefficient in estimating the phenological transition dates, this node-based measure might be subject to the influence of irregular spectral signatures. As the phenonetwork was constructed according to the spectral similarities between the nodes, an irregular signature similar to those during the transition dates might attain an abrupt high bridging coefficient value in comparison to that of its neighboring observation dates. To reduce the influence of those irregular spectral signatures, both bridging coefficient and betweenness centrality were averaged within a temporal moving window. A series of moving window sizes, ranging from one to seven, were experimented in the study to calculate the moving average bridging coefficient and moving average betweenness centrality.

Of a variety of moving window sizes, the seven-day moving average was found to be the best tradeoff among the spectral representation of nodes, spectral smoothness, and duration of in-situ foliage peak-coloration period. Compared to single day network measures, the resulting seven-day moving average bridging coefficient and betweenness centrality had much smoother and clearer temporal patterns during the fall phenology (Fig. 6). The phenological transition dates were much easier to be estimated via maximum moving average network measures. Between these two network measures, the moving average bridging coefficient had more distinct values than the moving average betweenness centrality values among the observation dates. It could more accurately identify the transition nodes within the network, similar to the single date network measure comparisons. Almost all the phenological transition dates, estimated through moving average bridging coefficients, lay within the grey-shaded field peak coloration regions.

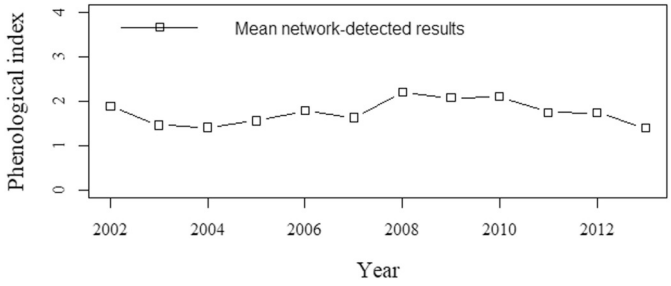

Fig. 7. The mean network-detected phenological index results from 2002 to 2013 at Hubbard Brook Forest.

The field observations at Hubbard Brook Forest were also utilized to evaluate the network-based phenological estimations. As the phenological surveys were conducted for three dominant species and field sites were located in the areas of varying topography and watersheds, the average phenological index of corresponding MODIS pixels was roughly employed to assess the community-level leaf coloration extent. Specifically, the phenological transition dates of the MODIS pixels covering the sites were estimated through the network model, and the ground-based phenological measures corresponding to estimated transition dates were averaged as the mean network-detected phenological index results (Fig. 7). The majority of average phenological index results were close to or below level 2 during 2002–2013, indicating that most leaves changed to intense reddish or yellow color at the estimated phenological transition dates. This field-based evaluation, along with the Harvard Forest assessment results, implied that the phenological transition dates estimated through the network model aligned well with the foliage peak coloration period, a critical phenophase to understand vegetation phenological responses to climate and environmental changes. As the field measurements at Hubbard Brook Forest were relatively rough (e.g., three species in ordinal scales), the further assessment of network-based measures was mainly conducted using the field data at Harvard Forest.

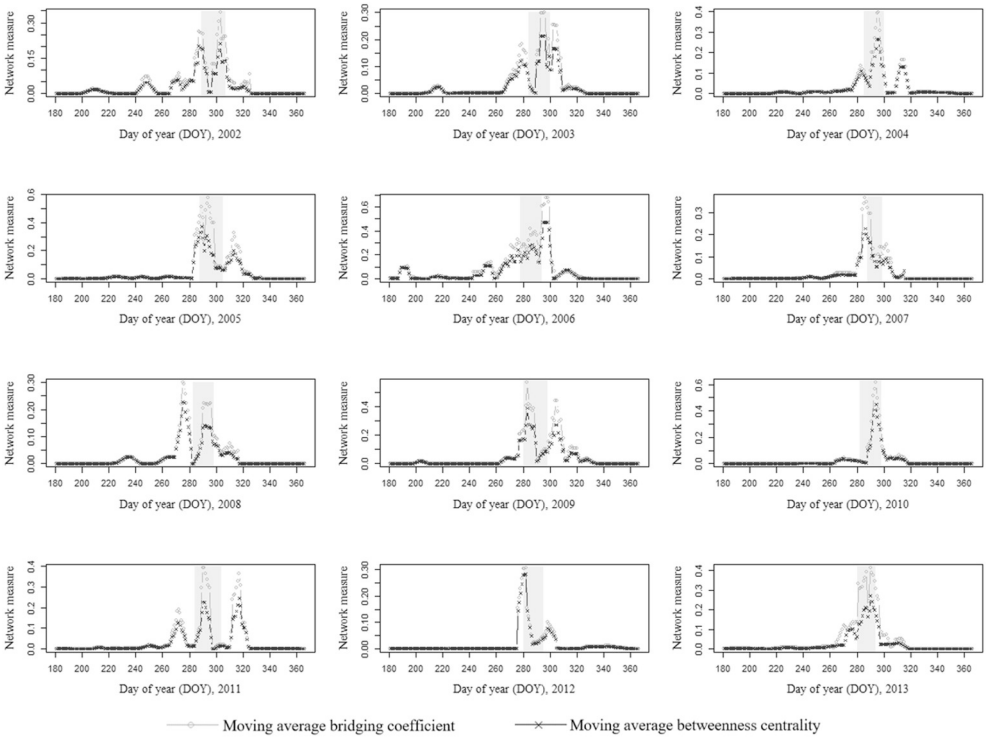

Fig. 6. The moving average bridging coefficient and betweenness centrality values of each observation date from 2002 to 2013 for the pixel of Harvard Forest headquarters.

Table 2
The phenological transition dates estimated for the pixel of Harvard Forest headquarters and its 6 * 6 pixel window, with and without phenological constraints via moving average bridging coefficients.

Year	Field-based mean peak coloration period	Phenological transition dates estimated via pheno-network (moving average bridging coefficient)			
		Phenological constraint		No phenological constraint	
		The pixel of Harvard Forest headquarters	6 * 6 pixel window (mean ± standard deviation)	The pixel of Harvard Forest headquarters	6 * 6 pixel window (mean ± standard deviation)
2002	[289, 307]	303	292 ± 6	303	292 ± 8
2003	[284, 300]	296	290 ± 7	296	290 ± 8
2004	[285, 300]	296	290 ± 4	296	294 ± 6
2005	[288, 305]	294	292 ± 7	294	296 ± 16
2006	[278, 294]	297	284 ± 6	298	286 ± 8
2007	[288, 299]	286	288 ± 3	286	287 ± 5
2008	[283, 298]	275	283 ± 2	275	281 ± 4
2009	[280, 298]	283	285 ± 3	283	285 ± 4
2010	[282, 298]	294	292 ± 5	294	292 ± 7
2011	[284, 304]	290	294 ± 6	290	296 ± 8
2012	[280, 295]	282	285 ± 6	282	282 ± 7
2013	[280, 294]	290	289 ± 5	290	289 ± 8

The impact of the phenological constraint, defined through the MCD12Q2 product, on the network estimation results was evaluated at Harvard Forest. By leveraging the devised network model, the moving average bridging coefficient values were calculated both with and without phenological constraints (Table 2). The phenological transition dates estimated under these two scenarios were almost identical, lying within or close to in-situ two week peak coloration periods. It suggested that the network model devised in this study was not affected much by the varying number of nodes, caused by the change of the temporal phenological range.

Due to the geolocation uncertainties of the MODIS pixel and surveyed tree species (within 1.5 km of the Harvard Forest headquarters), the phenological transition dates were also estimated for all the pixels in the 6*6 window of the Harvard Forest headquarters. The mean and standard deviation of the estimated transition dates were in Table 2. The mean phenological transition dates for this 6*6 pixel window aligned better with field mean peak coloration periods than that estimated from the pixel of Harvard Forest headquarters, particularly for years 2006 and 2008. The mean phenological estimation results of the 6*6 pixel window were also comparable with and without phenological constraints, though the network with phenological constraints might help reduce the standard deviation caused by irregular spectral observations. The results indicated that the devised network structure is characteristic of the complex fall phenological process, particularly to predict the transition hub in connecting all nodes of different stages together.

Conventionally, the remotely sensed phenological transition dates were estimated through time series of vegetation index over the course of a year. Specifically, double logistic function was employed in this study to fit the time series of NDVI data at Harvard Forest and Hubbard Brook Forest. The phenological transition dates were extracted using three phenophase extraction methods, namely the threshold-based method, the curve derivative method, and the curvature change rate method. At Harvard Forest, those satellite-derived mean phenological estimates (6 * 6 pixel window with phenological constraints) were compared to the mean dates of 75% ground leaf coloration using correlation coefficient and MAD (Fig. 8). Among the proposed network model and conventional phenophase extraction methods, the network model achieved the highest accuracy in estimating foliage peak coloration dates, with the highest correlation coefficient (0.68) and the lowest MAD (3.68 days) values. The correlation coefficient and MAD values of the threshold-based method were 0.6 and 6.33 days, respectively. For the curve derivative method, the correlation coefficient was 0.52 and the MAD was 8.5 days. As regards the curvature change rate method, the correlation coefficient and MAD values were 0.42 and

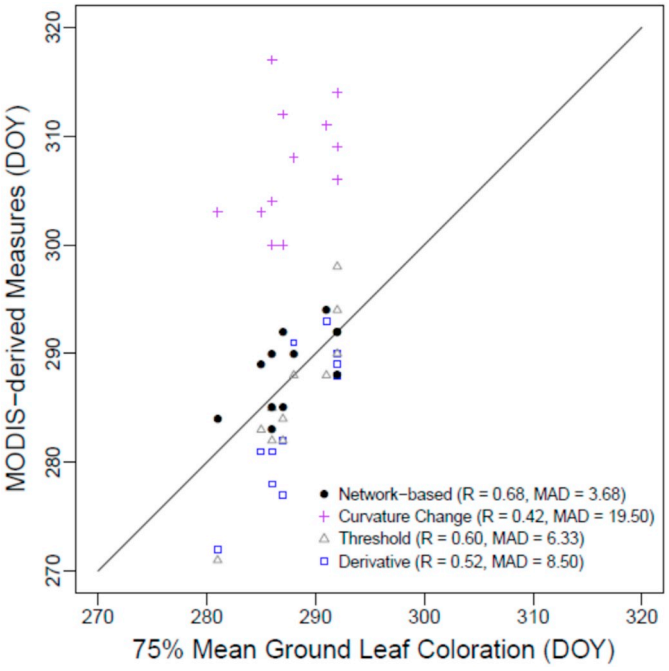

Fig. 8. The comparisons of the network model and three conventional NDVI curve-fitting based phenophase extraction methods (i.e., threshold-based, curve derivative, and curvature change rate) in estimating mean phenological transition dates, using mean dates of 75% ground leaf coloration during 2002 to 2013 at Harvard Forest.

19.5 days, respectively. By comparison, the network measures aligned better with field observed peak coloration periods. The inter-annual variation in foliage peak coloration was more appropriately explained by the network model with a correlation coefficient of 0.68. The mean difference between network-retrieved and field-observed phenological measures was smaller than the time interval of field observations (7 days). The transition dates estimated by threshold-based and curve derivative methods tended to be earlier than the field peak coloration dates, with most satellite-field pairs distributed below the 1:1 line in Fig. 8. On the contrary, the transition dates extracted by the curvature change rate method were much later than the field peak coloration dates, with a mean difference of about 20 days. The comparison between those phenological measures indicated that this network representation could uncover fall foliage peak coloration characteristics, complementary to the conventional NDVI measures.

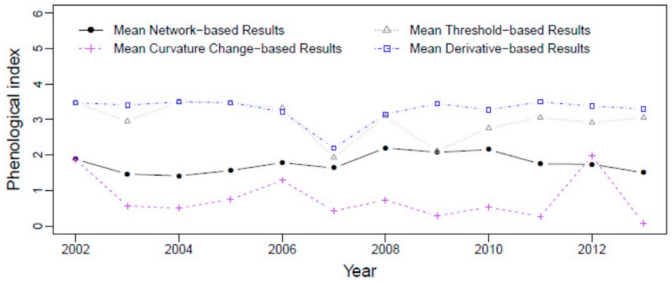

Fig. 9. The mean field-level phenological index retrieved from the network model and three conventional NDVI curve-fitting based phenophase extraction methods (i.e., threshold-based, curve derivative, and curvature change rate) during 2002 to 2013 at Hubbard Brook Forest.

At Hubbard Brook Forest, those satellite-derived phenological estimates were compared and evaluated according to their corresponding mean field-level phenological indices (Fig. 9). During 2002–2013, the mean field-level phenological indices retrieved by the network model were mostly close to index 2, indicating that the mean network-based estimates lay within the peak coloration period in which most leaves changed to intense reddish or yellow color. The mean phenological indices estimated by the curve derivative and threshold-based methods were mostly between index 3 and 4, implying that the retrieved phenological dates tended to be within low- to moderate- foliage coloration periods. By contrast, the mean phenological indices derived by the curvature change rate method were mostly below index 1, suggesting that the estimated dates were during the post-peak coloration period with most leaves fallen. The retrieved phenological periods from those four methods at Hubbard Brook Forest were comparable to those estimated at Harvard Forest. The comparisons among the methods in Fig. 8 and Fig. 9 indicated that the spatial and inter-annual variations in foliage peak coloration could be more appropriately explained by the network model.

With the devised network model, the phenological transition dates for peak coloration were estimated for the deciduous broadleaf and mixed forest areas within the extent of MODIS H12V04 tile. The spatial distributions of mean and standard deviation of network-retrieved peak coloration dates from 2002 to 2013 were shown in Fig. 10. The estimated mean peak coloration dates were around mid-October (DOY 290) in the southern area, and shifted to earlier dates around the end of September (DOY 270) in the north. They exhibited a mean change of 2.3 days per latitude degree. The network detected results revealed a gradual phenological shift from north to south along the latitudinal gradient, following the Hopkins bioclimatic law (Hopkins, 1938). It suggested that temperature decreased with increasing latitude and would likely be an important force driving the foliage phenological variations over wide geographic regions, though a variety of abiotic and biotic factors (e.g., elevation, local climate, and species interactions) might contribute to local variations in foliage phenological status (Schwartz and Reiter, 2000; Zhang and Goldberg, 2011). Along the latitudinal gradient, the inter-annual variation in the estimated average timing of peak coloration (the standard deviation map in Fig. 10) was generally less than 10 days across 2002–2013, which was comparable to the inter-annual variations in field observations of validation sites. A more comprehensive phenological spatial pattern analysis in response to climatic and environmental factors was out of scope of this study, due to the uncertainties in MODIS IGBP land cover classification results, snow and temperature quality flags, cloud and atmospheric interferences, and NBAR spectral data. The understanding of complex phenological patterns would benefit from future investigations of the proposed network model across a wide variety of ecosystems and regions, including low illumination conditions at high latitude regions.

4. Discussion

This study is the first attempt at introducing network science to time series remote sensing in modeling the complex phenological process of vegetation. The structures of pheno-networks characterize the collective changes of vegetation spectral signatures along the phenological trajectory. The overall performance of the pheno-network model, along with its opportunities and limitations, is discussed below.

Over the last decade, complex networks have been increasingly utilized in a wide variety of disciplinary fields, ranging from social to physical sciences, to capture the fundamental patterns and structures of complex systems. In this study, the formulation of network structures through spectral similarity measures presents a new representation of the complex phenological process of vegetation. This representation characterizes the vegetation phenological status through tracking the continuing changes of vegetation spectral signatures along the temporal trajectory. It models the fall phenology of deciduous forests to estimate the critical phenological transition dates using betweenness centrality, clustering coefficient, and bridging coefficient measures. With Harvard Forest and Hubbard Brook Forest as the reference sites, the network-retrieved phenological transition dates were found to align well with the peak coloration periods of the field observations. This indicates that the spectral signatures at the peak coloration period serve as a connecting hub and play a significant role in linking all the spectral signatures during the fall phenological process. The network measures can differentiate the various roles of vegetation spectral signatures on observation dates. Compared to betweenness centrality, the bridging coefficient network measure can better identify the transition nodes connecting different phenological stages through weighting the clustering coefficient. The peak coloration period of deciduous forests usually lasts about two weeks, and cannot be adequately captured by conventional curve fitting-based phenological approaches. The devised model extracts this critical phenological characteristic by modeling the collective change of spectral signatures with the network representation, complementary to current phenological models.

The conventional phenological models are typically based on curve fitting. The basic assumption is that the phenological trajectory of

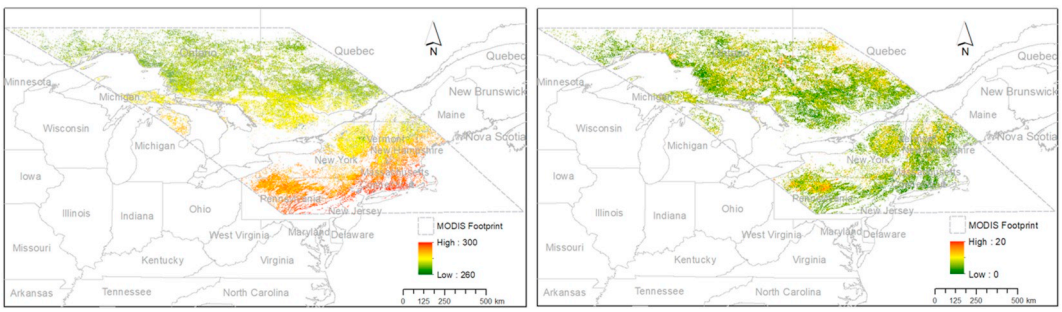

Fig. 10. The spatial distributions of mean (left) and standard deviation (right) of estimated peak coloration dates from 2002 to 2013 for the deciduous broadleaf and mixed forest areas in the MODIS H12V04 tile using the pheno-network model.

vegetation can be modeled via specific forms of mathematical equations (e.g., double logistic functions, asymmetric Gaussian functions, and Fourier transform). They fit a smoothing curve to raw observations of vegetation index throughout the year, and extracts the critical phenological stages of vegetation through characterizing the change of the curve. Yet the curve fitting-based approaches do not exploit all available spectral information and assume key phenological information can be captured through the mathematical equations for the curve. The phenological stages characterized by different equations may vary considerably, yielding conflicting results in uncovering vegetation responses to climate and environmental changes. Further, phenological observations over the course of a year are typically required in the curve fitting-based approaches. In this study, the proposed network model does not limit the complex phenological process to specific mathematical equations. This non-parametric model represents the complex system through defining phenology-relevant spectral nodes and edges, and analyzes the structure of collections of spectral nodes along the phenological trajectory with network measures. It extends one-dimensional vegetation index-based analysis to multiple-dimensional spectral analysis (Diao and Wang, 2016a, 2016b). The flexibility in constructing network models enables the characterization of critical phenological transition dates without requiring the year-long information. This network-based phenological model opens up unique opportunities to understand phenological dynamics of complex terrestrial ecosystems within certain temporal growth windows, or the ecosystems in which phenological processes are challenging to be defined via specific mathematical forms.

Remotely sensed phenological monitoring has mainly been conducted by tracking the temporal dynamics of vegetation index (e.g., NDVI, EVI, NDWI, and excess green index). Despite the promising results of those studies, the phenological transition dates extracted via different vegetation indices may vary considerably. Vegetation phenology is characterized by a collective combination of vegetation properties. As different vegetation indices have emphases on different vegetation properties, selecting an appropriate vegetation index to characterize the phenological process is difficult. The pheno-network model devised in this study offers a new angle for phenological characterization. It utilizes cosine distance to measure the similarities between spectral signatures along the vegetation development trajectory. Spectral signatures within a phenological stage are inclined to be more similar compared to those from different stages. A number of networks can be constructed with varying levels of spectral similarities. The underlying structure of the network depends on the way in which the spectral signatures are interconnected. As a sparsely or densely connected network may not adequately capture gradual changes of vegetation spectral signatures over time, the tolerance value is employed to locate a cohesive yet representative network that portrays phenological transition processes. The tolerance value controls the level of connectivity between spectral nodes in the network, and hence identifying an appropriate tolerance value is crucial in representing the phenological process. As the first attempt in bridging network science with time series remote sensing, the pheno-network model shows great potential in estimating the phenological transition dates for peak coloration. Yet the formulation of edges connecting the spectral nodes via the tolerance value is not as straightforward as that of complex networks in other disciplines (e.g., social networks). The capability of complex networks in tracking the collective dynamics of vegetation properties along the phenological trajectory, particularly the ways to define spectral nodes and principles for connecting those nodes, need to be further explored in future studies.

Due to the influence of cloud, snow, and atmospheric interferences, outlying observations in the time series may gravely affect the phenological retrieving results from satellites. The conventional curve fitting-based phenological approaches fit smoothing curves to raw observations of vegetation index to reduce the outlying effects. Similarly, several efforts have been made in the pheno-network to diminish the influence of outliers. First, the MCD43A4 product used in the study is a NBAR daily 16-day product, which pinpoints the best representative reflectance value for each observation date on a 16-day retrieval period. Second, the ancillary quality assurance layers, including snow and land surface temperature layers, are utilized to filter out the spectral observations that are contaminated by snow or low daytime temperature. Third, the formulation of the pheno-network in connecting the spectral nodes is built on spectral similarities. The outlying observations, which typically maintain distinct spectral values from those of other dates, will not be connected to other nodes in the network. Those separate nodes will not be considered in calculating the network measures (e.g., bridging coefficient). Fourth, the moving average of network measures is calculated for each observation date with a seven day temporal window. This moving average measure accommodates the temporal sequence of spectral nodes. It reduces the influence of irregular and outlying spectral signatures and locates the transition period that has consistently high bridging coefficient values. Fifth, the phenological constraint, defined by onset dates of EVI decrease and EVI minimum, is employed to constrain the analysis to the desired phenological periods to mask out irrelevant temporal observation periods. The pheno-network constructed in this study is unweighted and undirected, with the temporal sequence of spectral observations accommodated via the moving average measures. Other types of networks (e.g., directed or weighted networks), which are out of scope of this study, may also be good candidates for representing the complex phenological processes.

The accuracy of phenological transition dates retrieved by the network model was affected by several factors. The network model was constructed using the daily MODIS NBAR data generated from 16-day observations. The temporal compositing process, imperfect atmospheric corrections, and presence of snow and cloud cover may bring some uncertainties to the phenological detection results. Besides, the missing data during the fall phenological process may also affect the modeling results. For example, about 27% of the land surface was contaminated by clouds in 2001 to prevent acquisition of NBAR data for at least one 16-day period (Zhang et al., 2006). The extent of missing data varies geographically and depends on the cloud cover distribution. Though those missing data were replaced by the moving average of the nearest preceding and succeeding good quality neighbors in the time series, the temporal interpolation may cause some uncertainties and the detection confidence for each pixel should be evaluated in future studies. Complementary to MODIS, the recently launched Visible Infrared Imager Radiometer Suite (VIIRS) and Geostationary Operational Environmental Satellites (GOES) also observe the earth surface with high temporal resolutions. The synthesis of those data sources can provide daily spectral observations that are less affected by cloud contaminations to further improve the phenological detection results.

Validating the remotely sensed phenology through ground observations is one of the most significant challenges in phenological studies. Evaluation of estimated phenological transition dates using field observations at Harvard Forest and Hubbard Brook Forest indicates that the pheno-network model can capture well the foliage peak coloration phase. To sufficiently validate the performance of the network model on retrieving the vegetation phenology, more ground phenological observations across ecosystems and species are needed in future studies. Long term ground-based phenological observations are still rare in the US. The recent initiatives in obtaining ground or near surface remote sensing based phenological data, such as PhenoCam Network (https://phenocam.sr.unh.edu/webcam/) and National Phenological Network (https://www.usanpn.org/), will increase the data repository for conducting the validation. However, reconciling remotely sensed derived phenology with in-situ measurements is challenging, due to the lack of field data accounting for species abundance or spatial heterogeneity within the vegetation community.

The field phenological observations are collected for representative species in both validation sites (e.g., three tree species in Hubbard Brook Forest). Yet the MODIS-retrieved phenology represents the

overall phenological development of the vegetation community within a pixel, which may not correspond to that of individual species within the pixel. To be compatible with remotely sensed phenological monitoring, more accurate community-level in-situ phenological measures could be achieved through weighting the species-specific phenological status by species abundance (or cover). Due to the limited species distribution data and geo-location uncertainties in the current sites, the average percentages of leaf coloration of dominant tree species were calculated for both sites to assess the average phenological estimates of corresponding MODIS pixels. Further examination of the promising results achieved by the network model would benefit from more in-situ measurements collected at spatial scales commensurate with MODIS-retrieved community-level phenology. More comprehensive validation will be possible in future through upscaling the intensive field species-specific phenological observations to landscape or vegetation community scales, and through tracking the continuous vegetation growth with near surface remote sensing techniques (e.g., webcams and unmanned aerial vehicle).

5. Conclusions

This study aims to develop a pheno-network model to retrieve the phenological transition dates of vegetation during the fall season. The pheno-network model characterizes the phenological process through tracking the spectral signature changes over time. It identifies the spectral signatures during the transition period as the connecting hub in linking all other spectral nodes in the network. This unique network structure formulated via spectral similarities differentiates the various roles of spectral nodes at different phenological stages. It is one of the few studies to model the remotely sensed phenological process in a non-parametric fashion, complementary to conventional curve fitting-based phenological approaches. Our findings demonstrate that the phenological transition dates estimated through the pheno-network model correspond well with the peak coloration period of deciduous forests, with Harvard Forest and Hubbard Brook Forest as the reference sites. More efforts are needed in the future to evaluate this model over a broader range of ecosystems and conditions. Vegetation phenology has been documented to affect a wide range of ecosystem properties, including plant species competition, surface albedo, gross primary productivity, and ecosystem carbon uptake. To date, there is still a lack of consensus on the relationships between environmental factors (e.g., temperature) and the timing of leaf senescence, which is an indicator of the sophisticated mechanisms driving the fall phenological process. The devised pheno-network presents a unique representation to estimate critical phenological transition dates. This can subsequently improve the understanding of underlying phenological theories and mechanisms, and facilitate more accurate phenological representations in vegetation-climate interaction models to infer vegetation responses to climate change.

Acknowledgements

This research is partially supported by NSF Office of Advanced Cyberinfrastructure award (1849821) and Campus Research Board (RB18146) at the University of Illinois at Urbana-Champaign. This research is also part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. We thank J. O'Keefe for making the Harvard Forest phenology data available (http://harvardforest.fas.harvard.edu/), and Amey Bailey for providing the phenological data of Hubbard Brook Experimental Forest (https://hubbardbrook.org/).

References

Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47.
Badeck, F.W., Badeck, F.-W., Bondeau, A., Bottcher, K., Doktor, D., Lucht, W., Schaber, J., Sitch, S., 2004. Responses of spring phenology to climate change. New Phytol. 162, 295–309.
Barabási, A.-L., 2009. Scale-free networks: a decade and beyond. Science 325, 412–413.
Beck, P.S.A., Atzberger, C., Høgda, K.A., Johansen, B., Skidmore, A.K., 2006. Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. Remote Sens. Environ. 100, 321–334.
Borgatti, S.P., Mehra, A., Brass, D.J., Labianca, G., 2009. Network analysis in the social sciences. Science 323, 892–895.
Chen, M., Melaas, E., Gray, J., Friedl, M., Richardson, A., 2016. A new seasonal-deciduous spring phenology submodel in the community land model 4.5: impacts on carbon and water cycling under future climate scenarios. Glob. Chang. Biol. 22, 3675–3688.
Clark, J., Melillo, J., Mohan, J., Salk, C., 2014. The seasonal timing of warming that controls onset of the growing season. Glob. Chang. Biol. 20, 1136–1145.
Delbart, N., Kergoat, L., Le Toan, T., Lhermitte, J., Picard, G., 2005. Determination of phenological dates in boreal regions using normalized difference water index. Remote Sens. Environ. 97, 26–38.
Delpierre, N., Dufrêne, E., Soudani, K., Ulrich, E., Cecchini, S., Boé, J., François, C., 2009. Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. Agric. For. Meteorol. 149, 938–948.
Diao, C., Wang, L., 2014. Development of an invasive species distribution model with fine-resolution remote sensing. Int. J. Appl. Earth Obs. Geoinf. 30, 65–75.
Diao, C., Wang, L., 2016a. Incorporating plant phenological trajectory in exotic saltcedar detection with monthly time series of Landsat imagery. Remote Sens. Environ. 182, 60–71.
Diao, C., Wang, L., 2016b. Temporal partial unmixing of exotic salt cedar using Landsat time series. Rem. Sens. Lett. 7, 466–475.
Diao, C., Wang, L., 2018. Landsat time series-based multiyear spectral angle clustering (MSAC) model to monitor the inter-annual leaf senescence of exotic saltcedar. Remote Sens. Environ. 209, 581–593.
Dragoni, D., Rahman, A.F., 2012. Trends in fall phenology across the deciduous forests of the Eastern USA. Agric. For. Meteorol. 157, 96–105.
Eagle, N., Pentland, A.S., Lazer, D., 2009. Inferring friendship network structure by using mobile phone data. Proc. Natl. Acad. Sci. 106, 15274–15278.
Estiarte, M., Peñuelas, J., 2015. Alteration of the phenology of leaf senescence and fall in winter deciduous species by climate change: effects on nutrient proficiency. Glob. Chang. Biol. 21, 1005–1017.
Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Soc. Networks 1, 215–239.
Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. Softw. Pract. Exp. 21, 1129–1164.
Gallinat, A., Primack, R., Wagner, D., 2015. Autumn, the neglected season in climate change research. Trends Ecol. Evol. 30, 169–176.
Ganguly, S., Friedl, M., Tan, B., Zhang, X., Verma, M., 2010. Land surface phenology from MODIS: characterization of the collection 5 global land cover dynamics product. Remote Sens. Environ. 114, 1805–1816.
Gao, J., Barzel, B., Barabási, A.-L., 2016. Universal resilience patterns in complex networks. Nature 530, 307.
Gao, P., Bian, L., 2016. Scale effects on spatially embedded contact networks. Comput. Environ. Urban. Syst. 59, 142–151.
Garonna, I., de Jong, R., de Wit, A.J.W., Mücher, C., Schmid, B., Schaepman, M., 2014. Strong contribution of autumn phenology to changes in satellite-derived growing season length estimates across Europe (1982–2011). Glob. Chang. Biol. 20, 3457–3470.
Gutman, G., Ignatov, A., 1998. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. Int. J. Remote Sens. 19, 1533–1543.
Hermance, J.F., Jacob, R.W., Bradley, B.A., Mustard, J.F., 2007. Extracting phenological signals from multiyear AVHRR NDVI time series: framework for applying high-order annual splines with roughness damping. IEEE Trans. Geosci. Remote Sens. 45, 3264–3276.
Hopkins, A.D., 1938. Bioclimatics: A Science of Life and Climate Relations. (US Department of Agriculture).
Hufkens, K., Friedl, M., Sonnentag, O., Braswell, B.H., Milliman, T., Richardson, A.D., 2012. Linking near-surface and satellite remote sensing measurements of deciduous broadleaf forest phenology. Remote Sens. Environ. 117, 307–321.
Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., Serrano, L., 2008. Evolvability and hierarchy in rewired bacterial gene networks. Nature 452, 840–845.
Jeong, S.-J., Medvigy, D., 2014. Macroscale prediction of autumn leaf coloration throughout the continental United States. Glob. Ecol. Biogeogr. 23, 1245–1254.
Jonsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40, 1824–1832.
Jönsson, P., Eklundh, L., 2004. TIMESAT—a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.
Keenan, T., Gray, J., Friedl, M., Toomey, M., Bohrer, G., Hollinger, D., Munger, J.W., Okeefe, J., Schmid, H., Wing, I., Yang, B., Richardson, A., 2014. Net carbon uptake has increased through warming-induced changes in temperate forest phenology. Nat. Clim. Chang. 4, 598–604.
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, 67–75.
Liang, L., Schwartz, M.D., Fei, S., 2011. Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. Remote Sens. Environ. 115, 143–157.
Lloyd, D., 1990. A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. Int. J. Remote Sens. 11, 2269–2279.
Luo, W., Gao, P., Cassels, S., 2018. A large-scale location-based social network to understanding the impact of human geo-social interaction patterns on vaccination strategies in an urbanized area. Comput. Environ. Urban. Syst. 72, 78–87.
Morisette, J., Richardson, A., Knapp, A., Fisher, J., Graham, E., Abatzoglou, J., Wilson, B., Breshears, D., Henebry, G., Hanes, J., Liang, L., 2009. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Front. Ecol. Environ. 7, 253–260.
Moulin, S., Kergoat, L., Viovy, N., Dedieu, G., 1997. Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements. J. Clim. 10, 1154–1170.
Newman, M., Barabasi, A.-L., Watts, D.J., 2011. The Structure and Dynamics of Networks. Princeton University Press.
Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45, 167–256.
Newman, M.E.J., 2010. Networks: An Introduction. Oxford University Press, Oxford.
Onnela, J.-P., Arbesman, S., González, M.C., Barabási, A.-L., Christakis, N.A., 2011. Geographic constraints on social network groups. PLoS One 6, e16939.
Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2014). Epidemic Processes in Complex Networks. arXiv preprint arXiv:1408.2701.
Peñuelas, J., Filella, I., 2001. Phenology. Responses to a warming world. Science 294, 793–795.
Polgar, C., Primack, R., 2011. Leaf-out phenology of temperate woody plants: from trees to ecosystems. New Phytol. 191, 926–941.
Reed, B.C., Brown, J.F., VanderZee, D., Loveland, T.R., Merchant, J.W., Ohlen, D.O., 1994. Measuring phenological variability from satellite imagery. J. Veg. Sci. 5, 703–714.
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.
Roerink, G.J., Menenti, M., Verhoef, W., 2000. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. Int. J. Remote Sens. 21, 1911–1917.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–1069.
Rustad, L., Campbell, J., Dukes, J.S., Huntington, T., Lambert, K.F., Mohan, J., Rodenhouse, N., 2012. Changing Climate, Changing Forests: The Impacts of Climate Change on Forests of the Northeastern United States and Eastern Canada.
Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., Ohno, H., 2005. A crop phenology detection method using time-series MODIS data. Remote Sens. Environ. 96, 366–374.
Salton, G., McGill, M.J., 1986. Introduction to Modern Information Retrieval. McGraw-Hill, Inc.
Schwartz, M.D., Reiter, B.E., 2000. Changes in North American spring. Int. J. Climatol. 20, 929–932.
Strogatz, S.H., 2001. Exploring complex networks. Nature 410, 268–276.
Valente, T.W., 2012. Network interventions. Science 337, 49–53.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. Nature 393, 440–442.
White, M., de Beurs, K., Didan, K., Inouye, D., Richardson, A., Jensen, O., O'Keefe, J., Zhang, G., Nemani, R., van Leeuwen, W.J.D., Brown, J., de Wit, A., Schaepman, M., Lin, X., Dettinger, M., Bailey, A., Kimball, J., Schwartz, M., Baldocchi, D., Lee, J., Lauenroth, W., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. Glob. Chang. Biol. 15, 2335–2359.
White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. Glob. Biogeochem. Cycles 11, 217–234.
Xie, Y., Wang, X., Silander, J.A., 2015. Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts. Proc. Natl. Acad. Sci. 112, 13585–13590.
Zhang, X., Goldberg, M.D., 2011. Monitoring fall foliage coloration dynamics using time-series satellite data. Remote Sens. Environ. 115, 382–391.
Zhang, X., Friedl, M., Schaaf, C., Strahler, A., Hodges, J.C.F., Gao, F., Reed, B., Huete, A., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84, 471–475.
Zhang, X., Friedl, M.A., Schaaf, C.B., 2006. Global vegetation phenology from moderate resolution imaging Spectroradiometer (MODIS): evaluation of global patterns and comparison with in situ measurements. J. Geophys. Res. Biogeosci. 111.