Trends and natural variability of North American spring onset as evaluated by a new gridded dataset of spring indices

Toby R. Ault*

Mark D. Schwartz

Department of Geography, University of Wisconsin-Milwaukee

Raul Zurita-Milla

Faculty of Geo-Information Science and Earth Observation, University of Twente

Jake F. Weltzin

USA National Phenology Network, U.S. Geological Survey

Julio L. Betancourt

U.S. Geological Survey

*Corresponding author address: Dept. of Earth and Atmospheric Sciences, Cornell University, Bradfield Hall, Ithaca, NY.

E-mail: tobyault@cornell.edu
Climate change is expected to modify the timing of seasonal transitions this century, impacting wildlife migrations, ecosystem function, and agricultural activity. Tracking seasonal transitions in a consistent manner across space and through time requires indices that can be used for monitoring and managing biophysical and ecological systems during the coming decades. Here a new gridded dataset of spring indices is described and used to understand interannual, decadal, and secular trends across the coterminous US. This dataset is derived from daily interpolated meteorological data, and results are compared with historical station data to ensure the trends and variations are robust. Regional trends in the first leaf index range from $-0.6$ to $-1.7$ days per decade, while first bloom index trends are between $-0.2$ and $-1.4$ for most regions. However, interannual variability is substantial and coherent across regional (100s of Km) scales, and shorter-term trends (even as long as 35 years) are dominated by the imprint of decadal variability. These findings emphasize the important role large-scale climate modes of variability play in modulating spring onset on interannual to multi-decadal timescales. Finally, there is some potential successful sub-seasonal forecasts of spring onset, as indices from most regions are significantly correlated with antecedent large-scale modes of variability.
1. Introduction

Understanding the effects of climate change on seasonality requires metrics that track seasonal transitions through time and across large spatial scales. Like drought indices such as the Palmer Drought Severity Index (PDSI), these metrics should be “geographically independent,” meaning that they indicate the same event in the seasonal cycle regardless of climate differences in the regions from where they originate. Phenological data afford particular promise in developing such metrics because a limited suite of “indicator species” can be observed and modeled to represent the phenological response to climate of many temperature-sensitive native and crop species at mid-to-high latitudes (e.g., Schwartz et al. 2012). Phenological events are also thought to integrate weather and climate (Lieth 1974; Schwartz et al. 2012), making both their long term trends and interannual variations powerful indicators of climate change impacts (IPCC 2007; Parmesan 2006). Moreover, humans organize their lives around the seasonal cycle, and as forecasting capabilities improve, they will increasingly adapt by anticipating both variability and change in the timing of seasonal transitions.

In-situ phenological data are collected by professional scientists (Cook et al. 2012), networks of thousands of volunteers coordinated by national phenological networks (Vliet et al. 2003; Schwartz et al. 2012), and by passive and active sensors mounted on Earth-observing and orbiting satellites (e.g., White et al. 2009; de Beurs and Henebry 2010; Zurita-Milla et al. 2013). Although these datasets are invaluable for documenting what is being observed on the ground from year to year and decade to decade, they are not ideal for understanding large-scale patterns of decadal variability or secular changes because they either lack spatial continuity, as is typically the case of the observer-based networks, or long temporal coverage because satellite data is only available since the late 1970s.
As an alternative to strictly observational datasets of phenology, a number of weather-based indices have been developed and employed to understand the patterns and causes of timing variations and changes in seasonal transitions (e.g., Jolly et al. 2005; Schwartz et al. 2006). Such indices have the advantage of being calculable anywhere that meteorological data are available, enabling researchers to use them across much wider geographic domains and longer time periods than those available from observational data alone. One such suite of indices, the “Spring Indices” (Schwartz et al. 2006), has been particularly useful in this regard because the algorithms require only daily temperature minima ($T_{\text{min}}$) and maxima ($T_{\text{max}}$) as inputs, which are widely available from many meteorological stations throughout the world for most of the 20th century (e.g., Klein Tank et al. 2002). Spring indices calculated from station data have recently been employed to understand hemispheric-scale changes in spring onset (Schwartz et al. 2006), the relationship between spring onset and snowmelt in the Western US (McCabe et al. 2013), the interannual-to-decadal predictability of spring onset across North America (McCabe et al. 2011), and the large-scale patterns of trends and interannual variability in spring onset and “false springs” across North America (Ault et al. 2011; Schwartz et al. 2013; Ault et al. 2013; Peterson and Abatzoglou 2014).

Our primary goal in this study is to document regional and continental scale trends and variability in the timing of spring using spring indices calculated from both station data and a new gridded $T_{\text{min}}/T_{\text{max}}$ product—the Berkley Earth dataset (Mueller 2013)—for the coterminous United States. By focusing on gridded data with long temporal coverage (1920-2013), we hope to gain better insight into the regional patterns and trends, as well as establish new benchmarks for climate models, especially those used to understand and prepare for climate change (e.g., Taylor et al. 2012). These findings are especially important to consider in seeking a better understanding of climate
change impacts on seasonality, particularly at regional scales and over time horizons of 30 years or more.

2. Methods & Data

The original Spring Indices (SI-o) were developed as phenological models of lilac and honeysuckle leafing and blooming (Schwartz et al. 2002, 2006). As such, they incorporated “chilling requirements,” or the minimum exposure (measured in chilling hours) that some plants need to initiate their dormant period in winter. Subtropical climates rarely satisfy such chilling requirements, precluding the calculation of SI-o across all of the coterminous US. Recently, however, Schwartz et al. (2013) lifted the chilling requirements and found that these “extended” spring indices (SI-x) were equally valid in regions where SI-o had been used previously, but that SI-x could also be employed in subtropical regions to track phenological variations through time of several local species. Accordingly, Schwartz et al. (2013) suggested that the SI-x provided a metric of spring that was at least consistent through time and cued in to interannual variability in ways that could be readily connected to large-scale and regional fluctuations in the climate system and local phenological responses. Here we extend this approach further by calculating SI-x from gridded observational datasets and a much denser network of station data than has been employed previously (e.g., as in Schwartz et al. 2013; Ault et al. 2013).

A detailed explanation of the underlying assumptions, algorithm, and code for calculating SI-x is provided in Ault et al., (submitted; available online at: http://ecrl.eas.cornell.edu/node/11). Briefly, the indices are based on averages of phenological models for three plants (one species of lilac and two species of honeysuckle), validated with widespread phenophase observational data, and require only daily $T_{\text{min}}$ and $T_{\text{max}}$ as input (as well as latitude to compute day length). As such, they can be calculated across broad geographic domains and provide a consistent metric of
the same event (e.g., the emergence of spring foliage) through time. The two primary variables
comprising SI-x are the “first leaf index” (the average date when the first leaves appear on the three
plants) and “first bloom index” (the average date when blossoms first appear).

Here we calculate SI-x from the following datasets:

1. Daily station data from 2858 sites in the Global Historical Climatology Network (GHCN)
   (Klein Tank et al. 2002). Sites were used if they retained sufficient daily data from each
   year (no more than 30 days total missing during any given spring; no gaps longer than 10
   consecutive days) to calculate SI-x from at least 25 years out of the 30 year reference period
   from 1981-2010.

2. Gridded $T_{min}/T_{max}$ from the “Berkeley Earth” data product at 1 degree spatial resolution
   (Mueller 2013). Although daily temperatures from this product are available going back to
   1880, only the 1920-2013 portion was used because of heterogeneity in the spatial coverage
   of the data before that time. The focus of most of our analysis, however, is from 1955-2013
   for comparison with Schwartz et al. (2006).

Monthly indices of large-scale modes of climate variability were obtained from the National
Oceanic and Atmospheric Administration’s (NOAA) Physical Sciences Division (PSD) of the
Earth System Research Laboratory (ESRL) “climate indices” webpage: http://www.esrl.noaa.gov/psd/data/climateindices/list/. The indices considered here include: (a) the
NINO3.4 time series, as an indication of the state of the El Niño/Southern Oscillation (ENSO),
defined as average sea surface temperature (SST) from the Pacific Ocean between 5°S to 5°N
and 120°W to 170°W; (b) the Pacific Decadal Oscillation (PDO) (Zhang et al. 1997; Mantua
et al. 1997), defined as the first principal component of Pacific SST poleward of 20°N; (c) the
Atlantic Multidecadal Oscillation (AMO), computed as the detrended, area weighted average of
SST in the Atlantic between 0°N and 70°N (Enfield et al. 2001); (d) the Pacific North American (PNA), identified from 500mb height anomalies in NCEP reanalysis http://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/month_pna_index2.shtml; and finally (e) the North Atlantic Oscillation (Hurrell 1996), defined again using the 500mb height anomalies and methods as for the PNA (see http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/history/method.shtml for more information).

We used several simple metrics to compare SI-x station data with gridded data from the same region over the reference period of 1981-2010. First, we estimated the bias of gridded mean values by identifying each station encompassed by each grid cell, then subtracting the grid cell average from each station’s average. Second, we calculated root mean squared errors between the station data time series and the grid, again at each cell. Third, we correlated station data and gridded data at each station within each cell. Finally, linear trends were computed over the 1981-2010 period and removed to compute standard deviations of detrended data.

The gridded first leaf and bloom data for the time period from 1955-2013 were grouped into regional domains using a k-means clustering procedure with nine spatial clusters (e.g., Wilks 2011) applied in two different ways. Under the first application, clusters are identified using the mean index values, which isolates regions by offsets in their climatological spring onset dates. Under the second application, the clustering is applied to first leaf index anomalies by subtracting the long-term (1955-2013) mean from each grid point. The centroids of these clusters are therefore anomaly time series, which show regionally coherent departures from the long-term mean, but they are not directly compared with climate indices or atmospheric variables. Instead, these anomalies were used to identify the clusters, but we then averaged the raw first leaf and first bloom index values across each cluster to explore correlations of spring onset with large-scale climate modes and atmospheric variations. This step is necessary for identifying the seasonal window, which
varies year by year, that is most relevant to variations in index values. For example, correlations
with, say, April geopotential heights could yield misleading results for a region that experiences
leaf out dates that vary between the end of March and the beginning of May (as many locations
do). By taking average index dates from the entire cluster, we circumvent this problem because it
allows us to identify a “seasonally-appropriate” window (explained below) in time for each cluster
during each year.

Seasonally-appropriate 250mb geopotential height and wind fields are derived from the National
Center for Environmental Prediction (NCEP) daily reanalysis product (Kalnay et al. 1996). In to-
tal, nine of these seasonally-adjusted fields were produced (one for each region) by identifying the
temporal window surrounding the first leaf date for each region during each year. We identified
these windows as the days within one positive or negative standard deviation before the first leaf
index event. This produces a field of geopotential height values with dimensions of latitude, longi-
tude, year, for each region, which is then correlated with the corresponding time series of averaged
first leaf dates (not anomalies) for that area. The 250mb level was used to connect regional-scale
variations in spring timing to large-scale patterns and teleconnections of atmospheric variability.

Finally, each regional cluster is correlated with the large-scale climate indices described earlier.
As most of these indices are defined at monthly resolution, the month prior to the mean onset date
for that region is used. While this decision is in part necessary because only climatic information
leading up to an event can be relevant to it, this choice also allows us to explore the potential
predictability of spring onset on sub-seasonal, and possibly longer, lead times.
3. Results

a. Station versus gridded Spring Indices

Average first leaf and first bloom index values from station (GHCN) and gridded (Berkeley Earth) data are shown in Figure 1. First leaf index averages for both datasets follow elevational and latitudinal gradients in the western US, with greater mean values (later dates) occurring at higher latitudes and altitudes. In the eastern US, the gradients are primarily dominated by latitude. The first bloom index dates exhibit a similar geographic pattern, albeit with greater values (later dates) because blooming occurs after leafing in these lilac and honeysuckle species. The underlying gridded data and the station data exhibit similar spatial patterns.

Comparisons between the GHCN and the gridded datasets reveal biases in mean dates in the gridded data that are generally greater in the mountainous west than in the east (Figure 2). Specifically, throughout the Cascade and Rocky Mountain ranges north of about 38°N, many of the stations experience first leaf and first bloom dates that are earlier (warm colors in Figure 2) than the underlying grid. In the mountain ranges south of 38°N, the biases are reversed, with station dates being later than the underlying grid. Biases further east are small and generally less than 5 days, which is within the uncertainty of the original lilac and honeysuckle models (e.g., Schwartz et al. 2006).

Root mean squared error (RMSE) values between the time series of each station and each corresponding grid are illustrated in Figure 3. Again, the greatest discrepancies between the gridded product and the station data are found in the mountainous regions of the western US, with values between 20 and 30 days of error. Despite the biases and larger RMSE values between some stations and the gridded data, correlations between the two products are quite strong throughout the domain (Figure 4, although they are still lowest in mountainous western regions).
Detrended standard deviations (Figure 5) are lower for the gridded data than for the raw station data, especially for the first leaf index between 45°N and 50°N. Nonetheless, both products exhibit broadly similar geographic patterns, with higher standard deviations found in a belt stretching diagonally from the Pacific Northwest to the central Atlantic coast. Standard deviations are lower in California and throughout the southern portion of the domain. First Bloom index standard deviations are similar to those of the first leaf index, but slightly lower (Figure 5).

The patterns of trends in the first leaf and first bloom indices are very similar for the station and gridded data (Figure 5). In the Northwest Cascades, trends are positive (towards later onset dates), whereas they are negative throughout most of the rest of the domain. The area of greatest negative trend is centered on the Great Lakes for the first leaf index, and displaced slightly to the south of the Great Lakes for the first bloom index.

Overall, the geographic patterns of the means, standard deviations and trends are similar between the gridded and station-based SI-x. Likewise, the interannual variations of both products covary quite strongly—so much so that most correlation values are close to one, implying that any biases in mean values (leading also to high RMSE values) do not influence unduly the interannual variations of either product on regional scales. As such, we now move on to our analysis of regional clusters using only the gridded data.

**b. Cluster Analysis**

Gridpoints in each of the nine clusters identified from *raw* first leaf index values are shown in Figure 7 (results from applying the clustering to first bloom indices were similar and are not show). Defined in this way, the clusters loosely follow elevational and latitudinal gradients in the west, and follow more decidedly latitudinal gradients in the east. Given the similarities between these clusters and the contours of mean index dates (Figure 1), we focus instead on clustering first leaf
index anomalies to emphasize regionally synchronous covariations (Figure 8). The western US is nominally subdivided into four regions: (1) The Pacific Northwest; (2) the Southwest; (3) the Southern Rockies; and (4) the Northern Rockies. In the east, clusters tend to be more longitudinally elongated and are comprised of the following nominal regions: (1) East Northern; (2) East Central; (3) Texas & Gulf Coast; (4) Southeast & Atlantic Coast; and (5) the Far Northeast.

Cluster averages of the first leaf and bloom indices are plotted in Figure 9 for each of the nine regions. To generate these time series, the average raw first leaf and first bloom index dates are calculated from each cluster, weighted by their latitude, and averaged together (see methods). Statistics summarizing the means, standard deviations, and trends of each region are provided in Table 1.

Correlations between each cluster’s average leaf index and 250mb geopotential heights of seasonally-appropriate anomalies (see methods) are shown in Figure 10. In this figure, negative correlations (red) associate high geopotential height anomalies (indicative of ridging) with early spring, and positive correlations relate low height anomalies (troughs) with late spring. In the Pacific Northwest, for example, interannual variations in spring onset are negatively correlated with the central subtropical Pacific and Caribbean sectors (red and orange, Figure 10), while positively correlated with heights in the central north Pacific and eastern half of the continent (blue colors, Figure 10). The Southwest, in contrast, is negatively correlated with regional height anomalies (e.g., high heights correspond to early spring and low heights to late spring). Correlation patterns for both the northern Rockies, and southern Rockies are similar, but with the first of these two regions seeing weaker local negative correlations that are also displaced slightly to the north of the Southern Rocky negative correlating center. The East Northern and East Central regions do not appear at all strongly correlated with local variations in geopotential height, although there are significant positive correlations further afield. In the Southeast & Atlantic Coast cluster, negative
correlations with geopotential height occur across a longitudinal band from just west of 100°W to about 30°W. Positive correlations occur north and south of this band, and a similar pattern is present for the Texas & Gulf Coast cluster. Finally, the pattern in the Far Northeast also has a longitudinal band of significant correlations that is similar to the one in the previous two clusters, but of the opposite (positive) sign. These correlation patterns are interpreted further in Section 4 using composite maps of geopotential heights and winds during early and late years.

Other methodological choices to identify regional clusters produced qualitatively comparable results. For example, we calculated regional averages based on the US climate divisions (http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-divisions.php) as well as Principal Component (PC) time series, and identified similar regionally-coherent patterns of variability (not shown). Moreover, although the choice of nine clusters is somewhat arbitrary, we found it to be a good compromise between identifying truly unique regions of covariability while at the same time minimizing the total number of regions to make our correlation analysis meaningful. It also corresponds to the number of climate divisions in the coterminous US; although the regions do not overlap one-to-one with those US climate regions, they do provide insight into the trends, variability, and mechanism relevant to each location.

c. Correlations with climate indices

Table 2 summarizes the results of correlating each regional first leaf index time series with antecedent large-scale modes of climate variability, while Table 3 reports the first bloom index results. Significant negative correlations (positive climate index values with early spring) are found between the Pacific Northwest and NINO3.4, the PDO, and the PNA for both the first leaf index and first bloom index time series. Conversely, first leaf and first bloom indices from SE & Atlantic Coast as well as Texas & Gulf Coast are positively correlated with these same three modes, and
negatively correlated with the NAO. Moreover, the NAO correlates negatively with all but the Pacific NW and Far Northeast regions.

4. Discussion

a. Consistency in gridded and station-based indices

Results from calculating spring indices from GHCN station data as well as a gridded $T_{min}/T_{max}$ dataset yield a qualitatively similar picture of the means, trends, and variability in the timing of spring in the coterminous US. Spring onset (as indicated by the first leaf index) occurs between January and February for the most southern regions, and starts a bit later (March) for the west coast and much of the central part of the continent (Figure 1). Mean onset dates also follow meridional and elevational gradients, with later dates (April, May, and June) occurring for both the first leaf and first bloom index at greater elevations and higher latitudes. Biases in the gridded product, however, do show systematic regional differences, with most grid cells indicating dates that are later than the station data for the mountainous west north of 38°N (Figure 2). South of that latitude, high elevation sites tend to indicate earlier dates than the underlying grid. East of the Rockies biases are small (within 5 days) and are in fact within the uncertainty of the original first leaf and first bloom models (Schwartz et al. 2006). Despite these biases (see also Figure 3), correlations between station data and local grid points are quite strong and almost universally positive and close to one (Figure 4). Hence interannual variations in spring onset are very similar between the station data and the grid even when they are offset in time from one another. The patterns of interannual variability (Figure 5) and trends (Figure 6) between 1981-2010 likewise support this interpretation, hence we focus the remainder of our discussion on results obtained from the gridded SI-x data.
b. Regional patterns of gridded spring indices

Regional clusters and their time series help identify areas that covary on interannual timescales (Figures 8 & 9). We emphasize that these clusters are based on anomaly first leaf indices, not mean values (as in Figure 7), so that the clusters reflect primarily regions of spatial synchrony. Nonetheless, clusters in the southern and western regions correspond to grid cells with some of the earliest first leaf index dates. As a consequence, they are effectively isolated from the more northern and higher elevation regions because they have passed the first leaf index thresholds earlier in the year than those sites with later dates. Put differently, the early spring clusters cannot “feel” the effects of atmospheric variations later in the year. As a result our clusters depict the combined effects of geography, which determines the mean value of each region, and interannual variability, which governs regional synchrony.

In other studies, “phenoregions” have been identified using remote sensing data and k-clustering techniques (White 2005; Zhang et al. 2012), but these classification efforts are confounded by land use, and rely primarily on similarities in seasonal timing and other factors, and not on the spatial coherence of the interannual variability. A potential application for our SI-x anomaly-driven phenoregions include explanation of some biogeographic sutures (e.g., the clustering of species contact zones, hybrid zones, phylogeographic breaks as in Swenson and Howard 2005) through disruption of reproductive synchrony across plant metapopulations. A more obvious use would be to rely on patterns of long-term synchrony to guide plant phenological monitoring efforts in the U.S. (Schwartz et al. 2012).

c. Role of atmospheric circulation anomalies

Correlations (Figure 10) and especially composite maps (Figures 11 & 12) of each regional time series highlight regional and remote influences of the structure of the atmosphere on the timing of
spring onset. Beginning with the Pacific Northwest region, early spring onset dates are associated with an enhanced “Aleutian Low” (a climatological low pressure feature over the central north Pacific ocean), enhanced zonal wind just south of 40°N, and slightly more poleward meridional winds off the West Coast coast. Late spring onset dates correspond to positive anomalies of the opposite sign in geopotential height throughout most of the domain (Figure 12). During these years, zonal winds around 50°N are anomalously strong around the window of time associated with onset, and winds south of 40°N are weaker. Enhanced southward meridional flow is also evident in the eastern North Pacific region.

High geopotential height anomalies throughout the Southwest are present during early spring onset dates in that region, along with weak anticyclonic wind anomalies (Figure 11). Late spring, by contrast, occurs with low heights across much of western north America, cyclonic circulation anomalies, and a weakened Aleutian low in the Pacific (Figure 12). Similar patterns are present for both the southern and northern Rockies, but with the local high or low pressure anomaly displaced somewhat northward.

In both the E. Northern and E. Central regions, northward meridional anomalies are present during early spring years regions (Figure 11). Late onset dates in the East Northern region, however, appear linked to enhanced westerly flow and anomalously low heights over much of central Canada. In the E. Central region, however, the atmospheric anomalies are very weak.

Composites of early spring years for the Southeast & Atlantic Coast and Texas & Gulf Coast clusters are similar. In each region, relative highs emerge over the Pacific Basin and southeastern US, with a low anomaly over northwestern Canada in conjunction with early onset dates (Figure 11). Late springs in both regions likewise correspond to low heights and regional cyclonic circulation patterns. Anomalies in geopotential height are also present over the Atlantic basin, with
relative highs (and corresponding wind anomalies) occurring slightly the west of Greenland, and
lows in the central Atlantic.

Finally, the patterns of 250mb geopotential height and wind associated with early spring in
the Far Northeast show low anomalies off of the eastern US coast and over Alaska, with a high
centered on Hudson Bay. The late anomalies are associated with anomalous highs in a band across
the coterminous US and central Atlantic.

d. Climate indices

The results in Tables 2 and 3 lend themselves to a relatively straightforward interpretation, espe-
cially in light of the composite and correlation results discussed above. The NINO3.4, PDO, and
PNA time series are connected with each other although they also vary independently (Wallace
values in any of these climate indices are indicative of anomalous high pressure over western
Canada and lower than average pressure in the southeast. Such conditions favor more warm air
intrusion earlier into the Pacific Northwest connecting high NINO3.4, PDO, or PNA index values
with early spring. At the same time, the positive phases of these modes also allows a greater num-
ber of cold events to prolong spring onset in the southeast. Anomalies of the opposite sign occur
during the negative phases of these modes, thus favoring late spring in the Pacific Northwest and
early spring in the SE & Atlantic Coast and Texas & Gulf Coast clusters.

The role of the NAO in modulating spring onset is different from that of the PNA in two im-
portant regards. First, it operates primarily over the Atlantic and western Europe, but also exerts
its influence on winter climate in North America (Thompson and Wallace 2001). Second, low in-
dex values are key to understanding its potential influence on spring climate because during these
phases cold air outbreaks are more likely as the jetstream exhibits a more meridional structure
(Thompson and Wallace 2001). Hence, the negative correlations between most regional first leaf index time series, and almost all first bloom index time series, with the NAO suggests that cold outbreaks during its low index periods can delay spring, while the absence of such events during its warm phase may favor earlier spring.

**e. Regional Trends**

While regional and hemispheric variations in the atmosphere help us to interpret interannual variations in spring onset, the length (and quality) of reanalysis data before the satellite era make it difficult to conduct the same kind of analysis over longer time horizons with the same level of detail (e.g., Schneider et al. 2013, https://climatedataguide.ucar.edu/). Importantly, however, the patterns of trends in the gridded SI-x data are quite different for longer time horizons than they are for the relatively short period of the satellite era. Over the period from 1979 to 2013 (i.e., when satellite data could be available), trends in the coterminous US show an apparent dipole pattern with “centers of action” in the Northwest Cascade and Great Lakes Regions (boxes in Figure 13). The maps for the longer 1920 to 2013 period, on the other hand, are very different for both the first bloom and first leaf indices with more uniform warming across the continent. There is also evidence of a “warming hole” (Pan et al. 2004; Kunkel et al. 2006; Meehl et al. 2012; Schwartz et al. 2013), a region where spring onset was either delayed or did not advance as fast as elsewhere in the US, in the Southeast & Atlantic Coast and Texas & Gulf Coast.

To investigate these trends further, we show time series from each of the two centers of action identified above (Figure 14). The Northwest Cascade time series of average first leaf and first bloom index values show pronounced multidecadal variability, with relatively early dates occurring in the 1920s, 1930s as well as the 1980s and 1990s. In contrast, the Great Lakes trends towards earlier values start sometime in the late 1950s and appear to decrease steadily with time,
with the first bloom index dates trending slightly more quickly than the first leaf index ones. These results suggest that the mechanisms for generating the dipole pattern seen in Figure 13 (and also Figure 6) may be different in the Northwest Cascades and Great Lakes regions (e.g., they may not be driven by the same climate process).

The Pacific Decadal Oscillation’s (PDO) influence on spring onset has been documented and discussed elsewhere (McCabe et al. 2013). It is worth mentioning here insomuch as it helps explain the trends in the Northwest Cascades, a region strongly correlated to the PDO, especially during the spring (Figure 15). Correlations with the PDO are negative (see also Table 2) indicating the positive phase of the PDO favors early spring, while the negative phase favors later spring (as documented in McCabe et al. 2013). As the PDO has been in a negative phase recently, the positive (later spring) trends in the Northwest appear to be following suit.

Variability in the Pacific Ocean does not appear as strongly linked to the Great Lakes region (Figure 15), nor do the first leaf and first bloom time series exhibit multidecadal variations like those in the Northwest time series (Figure 14). Therefore the negative (later spring) trends in the Great Lakes region cannot be readily explained by multidecadal variations in the PDO (nor can they be explained by the Atlantic Multidecadal Oscillation; not shown). Instead, given the steady, post-1950s trend evident in Figure 14, we interpret changes in this region to reflect the secular long-term warming trend. Hence we suggest that the apparent dipole in North America is not driven by a single mode of climate variability, but instead the combined effects of two different process. In the west, the PDO appears to suppress, or even counteract, the role of the warming trend; in the Great Lakes and eastern US, the secular trend towards earlier spring dominates multidecadal and recent timescales.
5. Conclusion & future work

We have described continental and regional means, variability, and trends in spring onset for the coterminous United States using extended spring indices (SI-x) calculated from a dense network of historical station data and a spatially-complete (gridded) daily $T_{\text{min}}/T_{\text{max}}$ product (Mueller 2013). We found that both the station data and the gridded product provide a consistent view of the means, variances, and trends in spring onset across the continent. However, there are some key differences between the two that could have important ramifications for certain areas of inquiry. For example, the gridded product is generally biased towards later values in the mountainous western regions. It is also necessarily smoother (in space) than the underlying station data, and consequently the interannual variability (as measured by detrended standard deviations) is lower in the gridded SI-x than the station-based values. Nonetheless, the spatial patterns of standard deviations, as well as trends and means, are all very similar between the two types of underlying data.

Despite the considerable warming experienced globally and across the coterminous US during recent decades (e.g., Melillo et al. 2014), spring onset is not advancing towards uniformly earlier values as might be expected. Instead, interannual to decadal variations dominate regional trends and the overall spatial patterns. Specifically, the coterminous US is characterized by a dipole of spring onset trends with “centers of action” in the Northwest Cascades and Great Lakes regions from 1981-2013, with later dates seen in the former and earlier ones in the latter. Over the longer 1920-2013 period the trends are more uniform, but still exhibit large geographic differences. We interpret these results to be reflective of primarily the PDOs influence on the climate of the Pacific Northwest, which drives trends towards later dates most prominently since the late 1970s. The rest of the domain is seeing earlier dates, but not in connection to any one mode or combination of
modes. Hence the apparent dipole is likely more reflective of the suppression of early spring onset in the Northwest, and the influence of climate change elsewhere. These results apply to both the first leaf and first bloom indices.

Other regional trends calculated from regional clusters are between -0.6 and -1.7 days per decade for the first leaf index and -0.2 and -1.4 days per decade on average from 1955-2013. These estimates are quite consistent with those published earlier for continental and global scale phenometrics, and phenologies, of “shrubs” that respond to early spring (Schwartz et al. 2006; Parmesan 2006; Schwartz et al. 2013). Interestingly, the trends of the first leaf index, which responds to early spring warmth, outpace those of the first bloom index in most areas. This finding could have certain consequences for agriculture and ecosystem dynamics, if the period of spring warming is becoming longer as a consequence of climate change.

Our findings also highlight some of the challenges and opportunities in working with data for which dependent values are days of the year instead of physical quantities. As such, interpreting the interannual variability of SI-x requires standard methods in the atmospheric and climate sciences to be used somewhat differently than is typical. For example, applying a k-means clustering algorithm to the raw first leaf data emphasizes regions that differ from each other primarily in their mean dates, not their interannual variability. By applying the clustering to the first leaf index anomalies instead, we identified regions that tend to co-vary on interannual timescales regardless of their actual date. Nonetheless, the mean date is still taken into account in some capacity because these areas are effectively temporally isolated from the large-scale processes driving their variations: once spring has sprung in March in the Southeast, for example, variability in that region is not longer relevant to the clusters further north.

Finally, there is some potential opportunity for exploring the predictability of spring onset a month or so in advance (Tables 2 & 3). Significant correlations between spring indices and large-
scale modes of variability a month in advance occur in most regions. Hence, as suggested in McCabe et al. (2013) and Ault et al. (2011), early spring in the Northwest should be more likely during warm phases of ENSO, the PDO, or the positive phase of the PNA. Likewise, nearly all regions appear influenced by the NAO, which likely acts to delay spring by increasing the number of cold outbreaks at the end of winter (Thompson and Wallace 2001).

In addition to exploring the predictability of spring onset on sub-seasonal and longer timescales, future work should also investigate the atmospheric dynamics and thermodynamics linked to early and late spring in each of the regions identified here. Likewise, the SI-x could be readily calculated from global climate models to characterize and anticipate future changes in climate. Similar analysis could also be conducted at the hemispheric or global scale. By publishing the first, to our knowledge, *gridded* spring index dataset, this work opens the door to new “apples to apples” comparisons of observable changes (the timing of phenological events) to outputs from climate model simulations.

**Acknowledgments.** This project was facilitated in large part by the USA National Phenology Network (USA-NPN; https://www.usanpn.org/). We are especially grateful to Alyssa Rose-martin and the National Coordinating Office of the USA-NPN for providing logistical support and feedback. The SI-x models employed here were developed using phenological data that now reside at, and are available through, the National Phenology Database at the USA National Phenology Network. The project described in this publication was supported by Grant/Cooperative Agreement Number G13AC00248 from the United States Geological Survey. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the USGS.
References


IPCC, 2007: Chapter 1: Assessment of Observed Changes and Responses in Natural and Manage-
Systems. Working Group II: Impacts, Adaption, and Vulnerability, Cambridge University
Press, United Kingdom and New York, NY, USA.

Jolly, W. M., R. Nemani, and S. W. Running, 2005: A generalized, bioclimatic index to predict

American Meteorological Society, 77 (3), 437–471.

and precipitation series for the European Climate Assessment. International Journal of Climat-

Kunkel, K. E., X.-Z. Liang, J. Zhu, and Y. Lin, 2006: Can CGCMs simulate the twentieth-century


oscillation with impacts on salmon production. Bulletin of the American Meteorological Society,
78, 1069–10079.

of the El Niño Southern Oscillation and the Pacific Decadal Oscillation on the timing of the
//dx.doi.org/10.1002/joc.3400.


Table 1. Summary statistics for each of the nine time series of regional clusters in Figure 9. The first column is the cluster’s nominal regional identification in Figure 8, while the next three columns report the mean (integer value), standard deviation, and trend for the first leaf index, respectively. The next three columns report the mean, standard deviation, and trend for the first bloom index averages for each region. Units of the mean and standard deviation values are in days, while the units of the trends are in days per decade. All values are computed over the 1955-2013 period used for clustering.

Table 2. Correlation coefficients between each regional leaf index time series and several well-known indices of large-scale climate variability for the previous month. The previous month is used here to investigate the potential for predictability on sub-seasonal timescales. The month used from each index is indicated in parenthesis in the second column. Correlations that are significant at the 95% confidence limit are shown in bold.

Table 3. Correlation coefficients between each regional bloom index time series and several well-known indices of large-scale climate variability for the previous month. The previous month is used here to investigate the potential for predictability on sub-seasonal timescales. The month used from each index is indicated in parenthesis in the second column. Correlations that are significant at the 95% confidence limit are shown in bold.
TABLE 1. Summary statistics for each of the nine time series of regional clusters in Figure 9. The first column is the cluster’s nominal regional identification in Figure 8, while the next three columns report the mean (integer value), standard deviation, and trend for the first leaf index, respectively. The next three columns report the mean, standard deviation, and trend for the first bloom index averages for each region. Units of the mean and standard deviation values are in days, while the units of the trends are in days per decade. All values are computed over the 1955-2013 period used for clustering.

<table>
<thead>
<tr>
<th>Region</th>
<th>Leaf Index</th>
<th>Bloom Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Pacific NW</td>
<td>77</td>
<td>9.6</td>
</tr>
<tr>
<td>Southwest</td>
<td>32</td>
<td>2.7</td>
</tr>
<tr>
<td>N. Rockies</td>
<td>114</td>
<td>6.4</td>
</tr>
<tr>
<td>S. Rockies</td>
<td>90</td>
<td>7.5</td>
</tr>
<tr>
<td>E. Northern</td>
<td>110</td>
<td>6.8</td>
</tr>
<tr>
<td>E. Central</td>
<td>79</td>
<td>7.8</td>
</tr>
<tr>
<td>SE &amp; Atlantic Coast</td>
<td>49</td>
<td>9.1</td>
</tr>
<tr>
<td>Texas &amp; Gulf Coast</td>
<td>34</td>
<td>5.6</td>
</tr>
<tr>
<td>Far Northeast</td>
<td>130</td>
<td>5.5</td>
</tr>
</tbody>
</table>
TABLE 2. Correlation coefficients between each regional leaf index time series and several well-known indices of large-scale climate variability for the previous month. The previous month is used here to investigate the potential for predictability on sub-seasonal timescales. The month used from each index is indicated in parenthesis in the second column. Correlations that are significant at the 95% confidence limit are shown in bold.

<table>
<thead>
<tr>
<th>Region</th>
<th>Prev. Month</th>
<th>Climate Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NINO34</td>
</tr>
<tr>
<td>Pacific NW</td>
<td>(Mar)</td>
<td>-0.51</td>
</tr>
<tr>
<td>Southwest</td>
<td>(Feb)</td>
<td>0.1</td>
</tr>
<tr>
<td>N. Rockies</td>
<td>(Apr)</td>
<td>-0.22</td>
</tr>
<tr>
<td>S. Rockies</td>
<td>(Apr)</td>
<td>0.13</td>
</tr>
<tr>
<td>E. Northern</td>
<td>(Apr)</td>
<td>-0.11</td>
</tr>
<tr>
<td>E. Central</td>
<td>(Mar)</td>
<td>0.12</td>
</tr>
<tr>
<td>SE &amp; Atlantic Coast</td>
<td>(Feb)</td>
<td><strong>0.36</strong></td>
</tr>
<tr>
<td>Texas &amp; Gulf Coast</td>
<td>(Feb)</td>
<td><strong>0.36</strong></td>
</tr>
<tr>
<td>Far Northeast</td>
<td>(May)</td>
<td>-0.18</td>
</tr>
</tbody>
</table>
TABLE 3. Correlation coefficients between each regional bloom index time series and several well-known indices of large-scale climate variability for the previous month. The previous month is used here to investigate the potential for predictability on sub-seasonal timescales. The month used from each index is indicated in parenthesis in the second column. Correlations that are significant at the 95% confidence limit are shown in bold.

<table>
<thead>
<tr>
<th>Region</th>
<th>Prev. Month</th>
<th>NINO34</th>
<th>PDO</th>
<th>AMO</th>
<th>NAO</th>
<th>PNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pacific NW</td>
<td>(May)</td>
<td>-0.59</td>
<td>-0.72</td>
<td>-0.23</td>
<td>-0.24</td>
<td>-0.37</td>
</tr>
<tr>
<td>Southwest</td>
<td>(Feb)</td>
<td>0.16</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.48</td>
<td>0.03</td>
</tr>
<tr>
<td>N. Rockies</td>
<td>(May)</td>
<td>-0.31</td>
<td>-0.35</td>
<td>-0.01</td>
<td>-0.53</td>
<td>-0.25</td>
</tr>
<tr>
<td>S. Rockies</td>
<td>(May)</td>
<td>0.03</td>
<td>-0.13</td>
<td>0.18</td>
<td>-0.45</td>
<td>0.19</td>
</tr>
<tr>
<td>E. Northern</td>
<td>(May)</td>
<td>-0.03</td>
<td>-0.13</td>
<td>-0.08</td>
<td>-0.35</td>
<td>-0.23</td>
</tr>
<tr>
<td>E. Central</td>
<td>(Apr)</td>
<td>0.14</td>
<td>0.04</td>
<td>0.1</td>
<td>-0.31</td>
<td>0.18</td>
</tr>
<tr>
<td>SE &amp; Atlantic Coast</td>
<td>(Mar)</td>
<td>0.31</td>
<td>0.35</td>
<td>0.21</td>
<td>-0.57</td>
<td>0.34</td>
</tr>
<tr>
<td>Texas &amp; Gulf Coast</td>
<td>(Mar)</td>
<td>0.42</td>
<td>0.37</td>
<td>0.13</td>
<td>-0.54</td>
<td>0.36</td>
</tr>
<tr>
<td>Far Northeast</td>
<td>(Mar)</td>
<td>-0.05</td>
<td>-0.31</td>
<td>0.01</td>
<td>0.12</td>
<td>-0.16</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1. Mean values of the “first leaf” (top) and “first bloom” (bottom) indices calculated from the GHCN historical station data and statistically interpolated (gridded) data (Mueller 2013). Values of both indices are colorized according to their calendar day of the year (starting from Jan. 1), with the station data values displayed as open circles overlying the gridded index values. Means are calculated over the 1981-2010 base period. 34

Fig. 2. Left: Differences in mean index values between GHCN station data and for the first leaf (top) and first bloom (bottom) indices. In both panels the mean values of the overlying grid are subtracted from each station’s mean value, so that negative numbers (warm colors) indicate locations where the stations experience earlier onset dates than the gridded data. Right: Histogram summarizing the biases. Mean bias values are grouped into nine evenly spaced bins from -20 to 20 days at 5 day intervals; the fraction of all stations in each of those bins is shown as the y-axis. Roughly 3.6% of the stations had biases lower than -20 days and about 1.9% had biases greater than 20 days for the first leaf index, and about 3.71% and 2.31%, respectively, for the first bloom index. 35

Fig. 3. Same as Figure 2, but for root mean squared error (RMSE) scores between station data and gridded data. Approximately 2.6% of the stations had first leaf RMSE scores greater than 30, and 2.9% of the first bloom RMSE scores were greater than 30. 36

Fig. 4. Correlation coefficients between station-based and gridded spring indices over the period 1981-2010. Each open circle is the Pearson linear correlation between a GHCN station and the corresponding gridded index value where that station is located. Only 0.3% and 0.2% of the stations had first leaf index and first bloom index, respectively, correlations below zero. 37

Fig. 5. Standard deviations of linearly-detrended first leaf (top row) and first bloom (bottom row) indices calculated for both the gridded (left) and station (right) data. Standard deviations are estimated over the 1981-2010 reference period. Units are in days (i.e., of interannual variability). 38

Fig. 6. Same as Figure 5, but for the linear trends in the gridded (left) and station (right) data. Units are days per decade (with negative values indicating trends towards earlier onset dates), calculated over 1981-2010. 39

Fig. 7. Regions identified by k-means clustering applied to the first leaf index matrix over the period 1955-2013. Shading indicate the membership of each gridpoint in a given cluster. 40

Fig. 8. Same as figure 7, but for the anomaly first leaf matrix. 41

Fig. 9. Time series of mean annular first leaf (green) and first bloom (red) index dates for each regional cluster shown in Figure 8. Note that the values of the y-axes differ among panels, but they all cover a range of 90 days. 42

Fig. 10. Correlations between the time averaged first leaf index values of each cluster in Figure 8 and seasonally-appropriate geopotential height anomalies from NCEP reanalysis (Kalnay et al. 1996) over the period 1955 to 2013. Seasonally-appropriate geopotential height fields are generated from each year for each cluster by averaging daily values over a short period of time leading up to the mean first leaf index date for a given region and year. Correlations below the 95% (p<0.05) confidence limit are masked out with light pale blue. 43

32
Average geopotential height anomalies during the low first leaf index (e.g., early spring) years of each cluster. Geopotential height anomalies from each year are only averaged over a temporal window that is close to onset date during each year (as in Figure 10). Anomalies were determined by removing the long-term mean from each day using NCEP’s 1981-2010 geopotential height climatology.

Fig. 11.

Same as figure 11, but for the positive first leaf index (late spring) years.

Fig. 12.

Linear trends in the first leaf (left panels) and first bloom (right) indices over two different time periods: 1979-2013 (top) and 1920-2013 (bottom). The boxes show the Northwest Cascades (43°N to 48°N; 122°W to 110°W) and Great Lakes (38°N to 45°N; 90°W to 75°W) “centers of action” (the regions with the strongest trends as discussed in e).

Fig. 13.

Time series of indices from 1920-2013 in each of the two “centers of action” in the coterminous US shown in Figure 13 with the strongest positive (later) and negative (earlier) spring onset dates. The y-axis in both panels span 60 days, but their values differ.

Fig. 14.

Correlation coefficients between the spring (AMJ) PDO time series and the first leaf (top) and first bloom (bottom) indices for all of North America north of 20°.
Fig. 1. Mean values of the “first leaf” (top) and “first bloom” (bottom) indices calculated from the GHCN historical station data and statistically interpolated (gridded) data (Mueller 2013). Values of both indices are colorized according to their calendar day of the year (starting from Jan. 1), with the station data values displayed as open circles overlying the gridded index values. Means are calculated over the 1981-2010 base period.
Fig. 2. **Left:** Differences in mean index values between GHCN station data and for the first leaf (top) and first bloom (bottom) indices. In both panels the mean values of the overlying grid are subtracted from each station’s mean value, so that negative numbers (warm colors) indicate locations where the stations experience earlier onset dates than the gridded data. **Right:** Histogram summarizing the biases. Mean bias values are grouped into nine evenly spaced bins from -20 to 20 days at 5 day intervals; the fraction of all stations in each of those bins is shown as the y-axis. Roughly 3.6% of the stations had biases lower than -20 days and about 1.9% had biases greater than 20 days for the first leaf index, and about 3.71% and 2.31%, respectively, for the first bloom index.
Fig. 3. Same as Figure 2, but for root mean squared error (RMSE) scores between station data and gridded data. Approximately 2.6% of the stations had first leaf RMSE scores greater than 30, and 2.9% of the first bloom RMSE scores were greater than 30.
Fig. 4. Correlation coefficients between station-based and gridded spring indices over the period 1981-2010. Each open circle is the Pearson linear correlation between a GHCN station and the corresponding gridded index value where that station is located. Only 0.3% and 0.2% of the stations had first leaf index and first bloom index, respectively, correlations below zero.
FIG. 5. Standard deviations of linearly-detrended first leaf (top row) and first bloom (bottom row) indices calculated for both the gridded (left) and station (right) data. Standard deviations are estimated over the 1981-2010 reference period. Units are in days (i.e., of interannual variability).
**Fig. 6.** Same as Figure 5, but for the linear trends in the gridded (left) and station (right) data. Units are days per decade (with negative values indicating trends towards earlier onset dates), calculated over 1981-2010.
Fig. 7. Regions identified by k-means clustering applied to the first leaf index matrix over the period 1955-2013. Shading indicate the membership of each gridpoint in a given cluster.
FIG. 8. Same as figure 7, but for the anomaly first leaf matrix.
FIG. 9. Time series of mean annular first leaf (green) and first bloom (red) index dates for each regional cluster shown in Figure 8. Note that the values of the y-axes differ among panels, but they all cover a range of 90 days.
FIG. 10. Correlations between the time averaged first leaf index values of each cluster in Figure 8 and seasonally-appropriate geopotential height anomalies from NCEP reanalysis (Kalnay et al. 1996) over the period 1955 to 2013. Seasonally-appropriate geopotential height fields are generated from each year for each cluster by averaging daily values over a short period of time leading up to the mean first leaf index date for a given region and year. Correlations below the 95% (p<0.05) confidence limit are masked out with light pale blue.
FIG. 11. Average geopotential height anomalies during the low first leaf index (e.g., early spring) years of each cluster. Geopotential height anomalies from each year are only averaged over a temporal window that is close to onset date during each year (as in Figure 10). Anomalies were determined by removing the long-term mean from each day using NCEP’s 1981-2010 geopotential height climatology.
FIG. 12. Same as figure 11, but for the positive first leaf index (late spring) years.
FIG. 13. Linear trends in the first leaf (left panels) and first bloom (right) indices over two different time periods: 1979-2013 (top) and 1920-2013 (bottom). The boxes show the Northwest Cascades (43°N to 48°N; 122°W to 110°W) and Great Lakes (38°N to 45°N; 90°W to 75°W) “centers of action” (the regions with the strongest trends as discussed in e).
FIG. 14. Time series of indices from 1920-2013 in each of the two “centers of action” in the coterminous US shown in Figure 13 with the strongest positive (later) and negative (earlier) spring onset dates. The y-axis in both panels span 60 days, but their values differ.
Fig. 15. Correlation coefficients between the spring (AMJ) PDO time series and the first leaf (top) and first bloom (bottom) indices for all of North America north of 20°.