International Journal of Applied Earth Observations and Geoinformation 116 (2023) 103148

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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

Validating remotely sensed land surface phenology with leaf out records from a citizen science network

Logan M. Purdy [*], Zihaohan Sang, Elisabeth Beaubien, Andreas Hamann

Department of Renewable Resources, University of Alberta, 751, General Services Building, Edmonton, AB T6G 2H1, Canada

ARTICLE INFO

Keywords:
Remote sensing
Ground truthing
Plant phenology
Citizen science
NDVI
EVI2
MODIS

ABSTRACT

Vegetation phenology indices derived from multispectral remote sensing data are used to estimate primary productivity, track impacts of climate change and predict fire seasons. Such indices may, however, lack accuracy due to effects of snow and water, different vegetation types, and parameter choices for determining green-up and green-down. Here, we compare remotely sensed green-up dates with an extensive database of 57,000 leaf out and flowering observations from the Alberta PlantWatch citizen science network. We evaluate older global 5 km resolution VIP-NDVI and VIP-EVI2 v4 and v5 products, a regional 250 m resolution MOD09Q1-NDVI v6 product specifically designed for Alberta, and a recent 500 m resolution MCD12Q2-EVI2 v6 product. Overall, we find that MCD12Q2-EVI2 had the highest precision and least bias relative to ground observations, representing a significant advance over earlier phenology products. Different vegetation types showed a staged remotely sensed phenology in Alberta, with deciduous forest green-up first, followed by grasslands about 5 days later, and conifer forests green-up with a 10-day delay, allowing for corrections for different vegetation types. All products showed reduced interannual variability compared to ground observations, which may also lead to underestimating impacts of directional climate change. However, also in this respect MCD12Q2-EVI2 was superior, maintaining approximately 60% of the interannual variability. Nevertheless, the analysis shows that remotely sensed time series estimations of advances in leaf out may benefit from bias correction.

1. Introduction

Remote sensing is a valuable approach to monitor plant phenology, namely the timing of green-up of vegetation in spring, and senescence in fall from local to global scales. Remotely sensed phenology estimates have important and diverse applications, such as monitoring climate change effects on ecosystems (e.g., Garonna et al., 2015; Richardson et al., 2013), improving estimates of carbon sequestration (e.g., Leinonen and Kramer, 2002; White et al., 1999), determining optimal timing of pest management in agriculture (e.g., Adan et al., 2021), predicting start or end of the wildfire season (e.g., Bajocco et al., 2015; De Angelis et al., 2012), or pollen forecasting for allergy risk assessments (e.g., Li et al., 2022; Scheifinger et al., 2013).

Land surface phenology products are derived from time series of vegetation indices such as the Normalized Difference Vegetation Index (NDVI) (Rouse Jr et al., 1973) or the Enhanced Vegetation Index (EVI) (Liu and Huete, 1995). NDVI is a unitless index between −1 and 1 that represents the difference of Near Infrared (NIR) – Red bands divided (normalized) by the sum of NIR + Red bands. The 3-band EVI (Liu and Huete, 1995) and the 2-band EVI2 (Jiang et al., 2008) are an enhanced variant of this metric that minimises the influence of aerosol variations and bare soil, and avoids saturation of index values over dense vegetation (Huete et al., 2002). To derive land surface phenology products, which estimate the seasonal timing of vegetation changes, time series data with repeat periods of up-to 3 days in the mid-latitudes, can be summarized to single layers describing various phenophase transitions, such as green-up, mid-greenup, maturity, senescence, etc. (Bolton et al., 2020; Gao et al., 2021). Such phenology products may be developed at different spatial resolution (typically ranging from 30 m to 5 km in grid size), using different parameter (e.g. 15%, 35% and 50% ratios of vegetation index amplitudes) and different interpolation methods or smoothing functions (e.g. cubic spline, double-logistic functions) to arrive at day-of-year estimates for remotely sensed land surface phenology events (Bolton et al., 2020; Gao et al., 2021; Moon et al., 2022; Reed et al., 2009).

Remotely sensed land surface phenology records have several advantages over ground-based monitoring. Historical remote sensing data now span decades and have global coverage. Overlapping sensor sets

* Corresponding author.
E-mail address: lmpurdy@ualberta.ca (L.M. Purdy).

https://doi.org/10.1016/j.jag.2022.103148
Received 21 June 2022; Received in revised form 6 December 2022; Accepted 7 December 2022
Available online 20 December 2022

ensure consistent and comparable data coverage over space and time (Reed et al., 2009). In contrast, ground observations of phenology vary widely in methodological approaches, observation protocols, duration, and spatial coverage. Nevertheless, observational records are subjectively preferred, sometimes referred to as “true phenology” while remote sensing records are sometimes described as “apparent phenology”, (e.g., Younes et al., 2021). Yet, both metrics are equally “real”: ground observations are point records, whereas remote sensing grid cells integrate phenology over sizable landscape units. Notably, point observations on the ground are not expected to be representative of the average phenology of a landscape unit corresponding to 250 m to 5 km grid cell (Wu et al., 2019). Discrepancies between a limited number of point-based ground observations of plant phenology are therefore common and expected (Bornez et al., 2019; Delbart et al., 2015; Pouliot et al., 2011; Schwartz et al., 2002; White et al., 2014).

Besides this conceptual limitation with regards to matching point-based phenology observations to remote sensing data, there are also other sources of error that affect the precision and accuracy of land surface phenology estimates. The phenology of anthropogenic vegetation types can be difficult to model (Delbart et al., 2015). Cloud cover creates temporal and spatial gaps in data that can be particularly problematic for some ecosystems and geographic regions (White et al., 2014; Zhang et al., 2006). Spring snowmelt can bias the vegetation index upward, creating a false early greening-up signal (White et al., 2009). Lastly, varying curve smoothing approaches and parameter choices have been shown to produce green-up dates that vary by up to two months for the same grid cells (Helman, 2018). Therefore, establishing relationships between remotely sensed phenology estimates and ground-based phenological observations remains an important task of land surface phenology research.

The importance of “true phenology” records, despite their limitations, is also highlighted by different inferences on climate change impacts, when comparing ground phenology data and land surface phenology (Badeck et al., 2004). Ground phenology studies have documented relatively strong trends in temperate zones across the planet, with spring advancing at a typical rate of 2–7 days per decade, depending on the species (see histograms in Fu et al., 2015; Menzel et al., 2006; Piao et al., 2019). In contrast, remotely sensed trend estimates are usually much smaller in magnitude, typically ranging from 0 to 3 days per decade (Garonna et al., 2015; Jeong et al., 2011; Park et al., 2016), but assessments using more recent phenology products such as MODIS collection 5 show stronger trends (e.g., Karkauskaite et al., 2017). The slower rate of spring advance detected from land surface phenology may be due to reduced interannual variability of remotely sensed land surface phenology compared to ground phenology (Fisher and Mustard, 2007), which is particularly prevalent in lower resolution products (Peng et al., 2018; White et al., 2009). However, a systematic amplitude discrepancy can be corrected and some land surface phenology products may be largely free of such issues.

Here, we contribute a comparison of remotely sensed green-up dates with an extensive database of 57,000 leaf out and flowering observations from the Alberta PlantWatch citizen science network. We evaluate older global AVHRR/MODIS-based 5 km resolution VIP-NDVI and VIP-EVI2 Version 4 and 5 product, a regional 250 m resolution MOD09Q1-NDVI Version 6 product specifically designed for Alberta, and the recent 500 m resolution MCD12Q2-EVI2 Version 6 product. To address the issue of representation of point observations, we stratify the study area into ecological zones and vegetation types, and evaluate mean estimates for ecoregion × vegetation combinations (each represented by hundreds or thousands of phenology observations). Our objectives are to (1) compare several remote sensing products for their precision and bias of phenology estimates, (2) investigate if statistics vary with vegetation types and region, and (3) to test the hypothesis that older, lower resolution and multi-day composite phenology products may underestimate the magnitude of climate change effects on phenology.

2. Materials and methods

2.1. Alberta PlantWatch phenology data

The ground phenology data used in this study were collected by the Alberta PlantWatch citizen science network, which was initiated in 1987 to track spring plant phenology in the province of Alberta, Canada (Beaubien and Hamann, 2011a). Volunteers observe and report the calendar date of the following phenophases for common and easily identifiable plant species across Alberta: first bloom (first flowers open in three different places of a woody shrub/tree, or first flowers open in a patch of herbaceous plants), mid bloom (50% of flower buds open), full bloom (90% of flower buds open), and leaf out (first leaves unfurled in three places on the tree/shrub).

As of 2016, the Alberta PlantWatch database includes over 57,000 records for 30 species taken by roughly 700 observers (Fig. 1a). We evaluated a subset of six climatically homogenous ecoregions (Natural Regions Committee, 2006) with sufficient sample locations: Central Parkland (726 locations), Dry Mixedwood (478), Foothills Parkland (144), Central Mixedwood (165), Montane (258), and Grasslands (313). The Grasslands region is a combination of the Foothills Fescue, Dry Mixedgrass, and Mixedgrass natural subregions. We evaluated observations for three bloom phases for the nine most reported species, and aspen leaf out (Table 1), for a total of approximately 35,000 phenology observations, which represents 60% of the PlantWatch database. This selection was necessary to arrive at complete time series for ecosystem-species combinations for the 1987–2016 study period.

Rather than focusing on individual locations, we evaluate the correlation between the mean ground observations and its corresponding remote sensing phenology estimations per ecoregion. While correlations of individual point observations are sensitive to random noise, hundreds or thousands of phenology observations aggregated can more reliably represent vegetation types within the landscape unit. Mean phenology observations by species, year, and ecoregions were determined through best linear unbiased estimates (BLUEs) from a mixed effects model, taking advantage of collinearity among phenological phases (first, mid, and full bloom, and leaf out), collinearity among species, and collinearity among adjacent ecoregions to improve the accuracy of the estimated mean ground phenology dates. The predictor variables year, ecoregion, and species were specified as fixed effects (i.e. we estimate BLUEs for their combinations), and bloom phase was specified as a random effect in the mixed model, implemented with the ASReml package (Butler et al., 2009) for the R programming environment (R Core Team, 2013).

Although we evaluate remotely sensed phenology against all available phenology time series from different species (Appendix Table S1), we focus on aspen leaf out as primary benchmark because it is the most abundant tree species in Alberta and can also serve as a general phenology proxy with an intermediate phenology timing.

2.2. Remotely sensed land surface phenology

We compared six remote sensing products (Table 2) including three using the normalized difference vegetation index (NDVI) and three products based on a 2-band modified enhanced vegetation index (EVI2). This includes two older global 0.05° resolution Vegetation Index and Phenology (VIP) global datasets, released by the Vegetation Index and Phenology Laboratory at the University of Arizona (Didan and Barreto, 2015; Didan and Barreto, 2016) for the period 1981–2014. We refer to this dataset as VIP-NDVI and VIP-EVI2 in this study. The dataset was created using daily surface reflectance from the Advanced Very High Resolution Radiometer (AVHRR) LTDR v4 (1981–1999) and the Moderate Resolution Imaging Spectroradiometer (MODIS) Collection 5 (2000–2014). Time series data were subjected to a temporal smoothing algorithm to remove noise. To better represent green-up regions with a slowly emerging growing season, a 0.35 ratio was used in VIP to identify

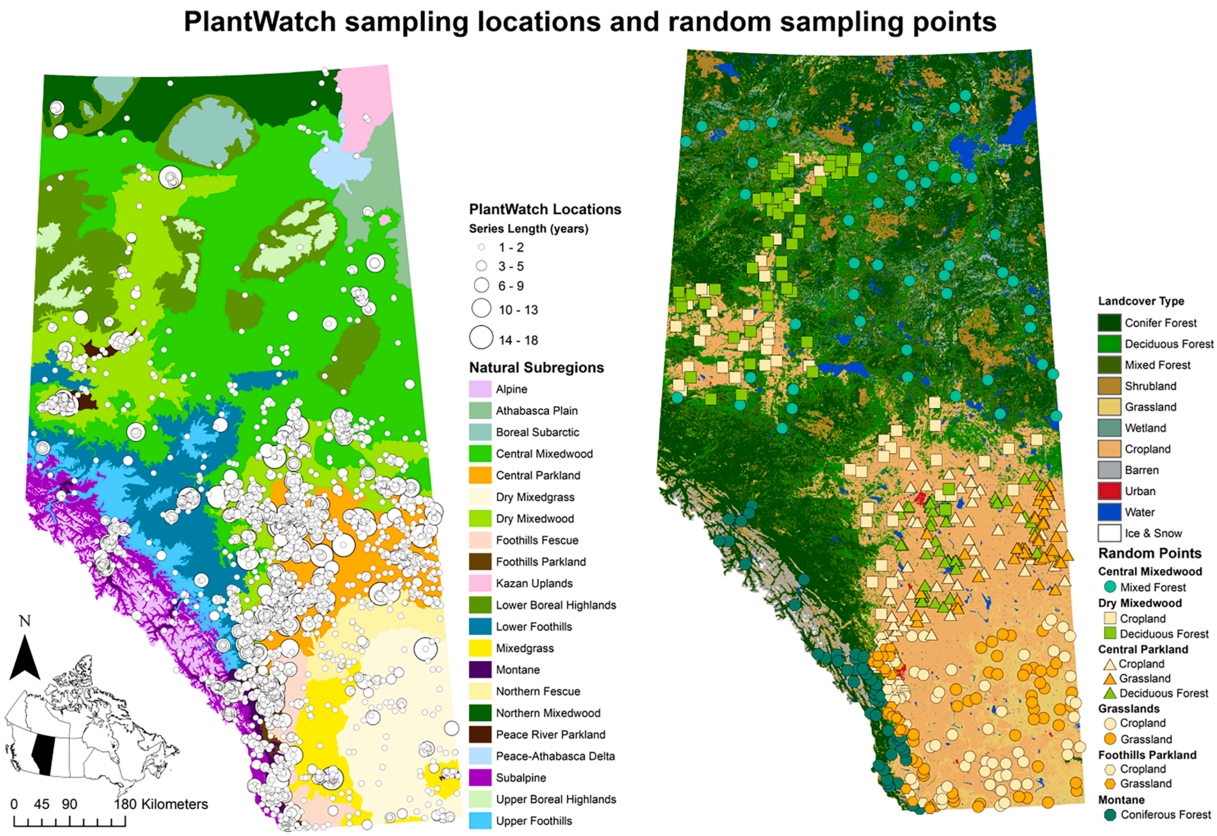

Fig. 1. Ecoregions of Alberta and locations of PlantWatch observer locations, with observer series length is indicated by the size of circles (right panel), and position of sampling locations for remote sensing data by ecoregions and vegetation types (left panel). Fifty sampling locations were random placed within each ecoregion and vegetation type combination, but then moved slightly to areas of homogenous vegetation types.

Table 1
Number of Alberta PlantWatch records from six different ecoregions by phenophase for the nine species used in this analysis.

Species		Bloom-phase				
Common name	Scientific name	First	Mid	Full	Leaf out	Total
Saskatoon	*Amelanchier alnifolia*	2044	1818	1577		5439
Early blue violet	*Viola adunca*	1608	1420	1291		4319
Prairie crocus	*Anemone patens*	1550	1285	1112		3947
Aspen poplar	*Populus tremuloides*	1320	1126	819	601	3866
Chokecherry	*Prunus virginiana*	1419	1226	1009		3654
Northern bedstraw	*Galium boreale*	1344	1219	1025		3588
Golden bean	*Thermopsis rhombifolia*	1357	1203	1004		3564
Yarrow	*Achillea millefolium*	1370	1171	948		3489
Solomon's seal	*Maianthemum stellatum*	1310	1089	852		3251

the start of growing season instead the standard half-max (0.5 ratio) method (Didan et al., 2018).

The recent 500 m resolution MCD12Q2-EVI2 Collection 6 product (Friedl et al., 2019), was developed from the combined Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, providing global land surface phenology metrics from 2001 to 2019. Vegetation phenology metrics identify up to two growing cycles per year to capture uncommon phenological regimes (e.g., wet-seasons or phenology of agricultural systems). The MCD12Q2-EVI2 product reports green-up and mid-green-up using 0.15 and 0.5 of the time series maximum minus the minimum annual EVI2 deviations. Here, we evaluate the mid-greenup layers.

We also evaluated a higher resolution (250 m pixel size) NDVI-based land surface phenology product developed by Pickell et al. (2017) based on MOD09Q1 Version 6 (Vermote, 2015). The MOD09Q1 product is based on the MODIS/Terra sensor, consisting of 8-day composites at 250 m resolution. Pickell et al. (2017) implemented a noise removal procedure, converting NDVI observations to a z-score and removing outliers based on a one-sided critical α level of 0.1. Subsequently, a one-dimensional cubic spline was fitted to interpolate 8-day composites, and green-up was determined with the half-maximum or 0.5 ratio method. This phenology product was designed for Alberta forest lands, with the purpose of determining the onset and end of wildfire seasons.

Similar to the stratification of ground phenology data, all sample points representing the remotely sensed green-up date were averaged by ecoregion × vegetation combinations (Fig. 1b). Vegetation classes were determined with MODIS vegetation data (Commission for Environmental Cooperation, 2013). Each ecoregion × vegetation combination was represented by 50 sample points that were first randomly placed, but then manually moved away from boundaries into areas of homogenous vegetation classes in order to minimize the probability of a sample pixel being composed of multiple vegetation types. We further replaced a small number of sample points that showed remotely sensed green-up dates earlier than the 60th day of year in any of the tested products.

Table 2
Remote sensing products evaluated in this study. They differed in vegetation index used, including the normalized difference vegetation index (NDVI) and the 2-band modified enhanced vegetation index (EVI2), spatial and temporal resolution of the gridded product, data coverage, and method to determine mid-greenup: either a 0.35 ratio or the standard half-max (0.5 ratio) method.

Product	Index	Grid	Data coverage	Mid-greenup
VIP AVHRR v4	NDVI	0.05°	1981–1999	0.35 ratio
VIP AVHRR v4	EVI2	0.05°	1981–1999	0.35 ratio
VIP MODIS v5	NDVI	0.05°	2000–2014	0.35 ratio
VIP MODIS v5	EVI2	0.05°	2000–2014	0.35 ratio
MOD09Q1 v6	NDVI	250 m	2000–2016	0.5 ratio
MCD12Q2 v6	EVI2	500 m	2001–2019	0.5 ratio

Earlier dates are considered erroneous for Alberta climate (Cui et al., 2017; Pickell et al., 2017). Annual green-up dates were then aggregated across the 50 sample points representing an ecoregion × vegetation combination.

2.3. Statistical analysis

Pearson correlation coefficients (r), and root mean square error (RMSE) were used to evaluate the strength of the relationship between remotely sensed green-up dates and phenology observations, following Willmott (1982). We also calculate mean bias error (MBE), which represents the average difference of remote sensing – ground observations (positive MBE values indicate a lag of the remote sensing estimate and negative MBE values indicate that remote sensing estimates precede the ground observations). The r, RSME and MBE statistics for the VIP-NDVI and VIP-EVI2 datasets were evaluated separately from 1987 to 1999 (AVHRR data) and 2000–2014 (MODIS data). Since all phenology data contribute to each BLUE estimate of each species based on collinearity, we use aspen leaf out as a ground observation which would be expected to be unbiased for the deciduous forest vegetation type. Aspen is one of the most common and widespread trees throughout Alberta and is expected to have a considerable direct influence on remotely sensed phenology. Correlations of remote sensing products with phenology time series of other plant species, spanning from approximately day-of-year 100 to 170, are reported in the Electronic Supplement Table S1.

3. Results

3.1. Evaluation of ground observations versus remote sensing

Bloom dates for all the nine species across Alberta range from early April to the end of June. Species blooming at similar times show close correlations in interannual phenology (Fig. 2). The sequence of spring phenology for the chosen species begins with the synchronized bloom of aspen and prairie crocus, followed by aspen leaf out approximately two to three weeks later, and the bulk of most phenology observations occurring within an additional two to three weeks thereafter. In the following, we focus on aspen leaf out as primary benchmark for evaluation as it is the most common tree species in Alberta, serving as a relevant direct observation for deciduous and mixed forests. Because of its intermediate timing and occurrence throughout all ecosystem types, it is also a useful general proxy for ground phenology of other land cover types, and it should be noted that all other autocorrelated phenology time series contribute to the precision of aspen leaf out estimates, as they would to an average phenology proxy.

Remote sensing products have different accuracy and precision of the aspen green-up estimations. Correlations between ground-observed aspen leaf out and remotely sensed phenology are the highest for the MCD12Q2-EVI2 product across all combinations of ecoregions and vegetation types, except for deciduous forest in the Dry Mixedwood (which is slightly weaker than MOD09Q1-NDVI model, Table 3). This also holds true for bloom dates of most other species evaluated in this study (Table S1). Root mean square errors (RMSEs), which measure both precision and bias are generally one of the lowest for MCD12Q2. RMSEs of this model are over 50% lower than others in Montane and Foothills Parkland ecoregions (Table 3). The association between aspen leaf out and MCD12Q2-EVI2 estimations varies only slightly among vegetation types. Generally, the strongest associations were found for deciduous, mixed forests and grasslands (~0.83), with slightly lower values for coniferous forests and croplands (~0.77).

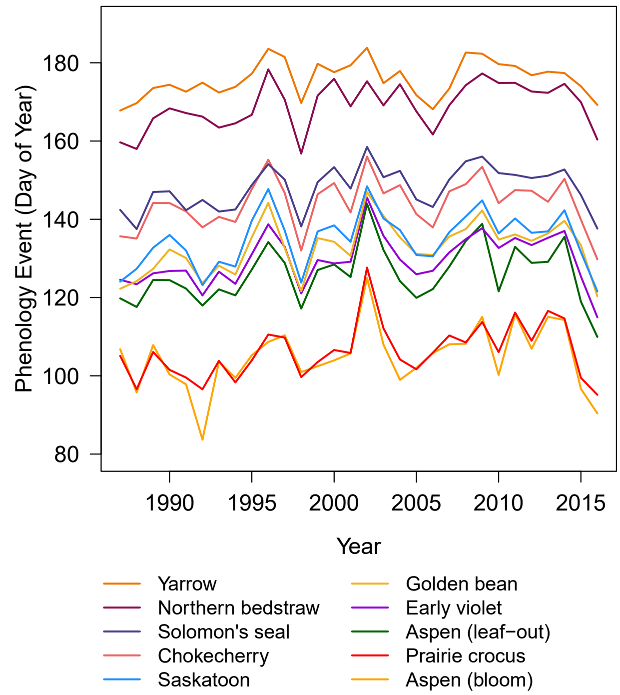

Fig. 2. Phenology sequence of the mean date estimate for first bloom of 9 plant species, plus aspen leaf out, in the Central Parkland region of Alberta. The legend is ordered from the latest (Yarrow) to earliest (Aspen bloom) phenological event.

Compared to MCD12Q2-EVI2, the other phenology products showed weaker correlations with ground observations. The MOD09Q1-NDVI only showed strong associations for deciduous forests (~0.80), with lower values for coniferous and mixed forests (0.5 to 0.6). The phenology of croplands and grasslands are not well captured by this dataset, except for the Dry Mixedwood ecoregion (0.63, Table 3). Green-up date estimates from the VIP-series products (VIP-NDVI based on AVHRR and MODIS) have even weaker correlations with ground phenology. Correlations are low across the majority of ecoregion and vegetation combinations, except Foothills parkland. These two products also have higher RMSE values, indicating strong bias of estimations. However, the older 0.05 degree NDVI-subset from the AVHRR sensor with estimates prior the year 2000 has moderate to good correlations for the cropland and grassland vegetation types compared to other products. Correlations with blooming observations across all 9 species revealed similar patterns across remote sensing products (Table S1).

3.2. Lags in the timing of green-up estimates

Mean Bias Error (MBE) represents the lag time between remotely sensed green-up and ground observations of phenology (Table 3 and visualized in Fig. 3). Although the correlation of ground observations and remote sensing may be strong, lags or systematic bias need to be quantified for accurate estimates of vegetation green-up. We find that green-up dates estimated by VIP-NDVI MODIS are much earlier than the corresponding aspen leaf out (Fig. 3, dotted lines vs black line after 2000), with MBE values ranging from 16.3 to 37.9 days earlier (Table 3).

Table 3
Statistics for the accuracy and precision of the green-up estimate relative to the mean date of aspen leaf out in each ecoregion-vegetation combination. The root mean squared error (RMSE) is used as a measure of bias, and the Pearson correlation coefficient (r) is reported as a measure of precision for the relationship between the remotely sensed sensing green-up and observed phenology (* denotes statistical significance of correlations for $p < 0.05$, and ** for $p < 0.01$). Mean bias error (MBE) are in units of days between the remotely sensed green-up date minus the observed date for aspen leaf out. A negative MBE value indicates that the green-up estimate precedes the ground observed date for that phenophase.

		VIP-NDVI AVHRR			VIP-NDVI MODIS			MOD09Q1-NDVI			MCD12Q2-EVI2		
Ecoregion	Vegetation	RMSE	MBE	r	RMSE	MBE	r	RMSE	MBE	r	RMSE	MBE	r
Central Parkland													
	Deciduous	20.6	−20.1	0.37	30.9	−30.2	0.3	5.8	−2.4	0.75**	6.6	−4.1	0.83**
	Cropland	12.2	−11.2	0.16	21.1	−19.9	0.32	16.9	15.4	0.46*	11.3	9	0.58*
	Grassland	23	−22.6	0.4	33.2	−32.3	0.18	13.7	−10.4	0.37	9.9	−8.6	0.80**
Dry Mixedwood													
	Deciduous	15.2	−14.5	0.29	28.9	−28.3	0.52*	4	0.3	0.86**	4.5	−0.8	0.83**
	Cropland	11	−10.2	0.56*	23.1	−22.3	0.50*	13.7	12.3	0.63**	11.2	9.6	0.75**
Foothills Parkland													
	Cropland	21.2	−20.6	0.75**	26.8	−26.2	0.66**	12.1	8.7	0.47*	5.3	−2.4	0.89**
	Grassland	19.1	−18.5	0.89**	22.3	−21.6	0.64*	12.5	8.6	0.33	5.4	−1.7	0.86**
Grasslands													
	Cropland	13.1	−12.3	0.55*	17.9	−16.3	0.03	24.2	22.3	0.06	17.1	16.1	0.86**
	Grassland	32.2	−31.9	0.56*	38.8	−37.9	−0.06	16	−9	0.17	8.3	−6.2	0.82**
Central Mixedwood													
	Mixed forest	25.8	−25.4	0.29	38.2	−37.7	0.48*	16.1	−14.5	0.63**	7.5	−5.5	0.83**
Montane													
	Conifer forest	17.5	−16.5	0.08	27.6	−26.9	0.66**	26	−24.3	0.52*	11.5	−9.6	0.77**

For natural vegetation types, the remotely sensed green-up date is approximately one month earlier, whereas remotely sensed cropland green-up is two to three weeks earlier than ground observations. Early green-up estimates can also be found in VIP-NDVI AVHRR, but the average difference between ground-observed aspen leaf out and remotely sensed green-up is 40% less before the year 2000 (Fig. 3., vertical dashed line). Further, we note that VIP-EVI2 showed very similar correlations and lags as VIP-NDVI. Statistics for VIP-EVI2 are not reported in Table 3 and Fig. 3 for conciseness, but correlation statistics for VIP-EVI2 can be found in Appendix Table S1.

MCD12Q2-EVI2 and MOD09Q1-NDVI products are closer to the ground observations of aspen leaf out dates (Fig. 3; solid lines and dash-dot lines). These products use a 0.5 ratio to estimate the date of green-up (Table 1), but are still somewhat early. The most relevant statistic for a general adjustment of phenology would be observed aspen leaf out against "Deciduous" landcover in the "Central Parkland" and "Dry Mixedwood" ecoregions where aspen is dominant (Table 3, rows 1 and 4). Other vegetation types showed consistent lags across different ecoregions, with grassland green-up first, followed by aspen deciduous forest green-up, and later, cropland green-up (Fig. 3). BMEs for MOD09Q1-NDVI based green-up showed similar but stronger bias based on aspen leaf out dates. Remote sensing estimates were delayed for croplands and also lagged for natural forested areas (especially for Montane conifer forest). MBEs of MOD09Q1-NDVI have the same direction (sign), but with large absolute values than the corresponding values of MCD12Q2-EVI2.

3.3. Interannual variability of green-up estimates

Interannual variability is lower for all remotely sensed phenology estimates than for ground-observed phenology, indicated by slopes less than 1 (Fig. 4). This figure only shows slopes for the remote sensing product with the strongest correlations with ground observations. The range of slope values for other products was 0.02 to 0.47 for the VIP-NDVI AVHRR, −0.03 to 0.52 for the VIP-NDVI MODIS, 0.02 to 0.65 for the MOD09Q1-NDVI, and 0.25 to 0.82 for the MCD12Q2-EVI2 dataset. This can also be seen in Fig. 3, where VIP-NDVI AVHRR data from1980 to 1999 has nearly flat-lines for interannual variation despite some strong correlations for grasslands and cropland vegetation types (Table 3). The VIP-NDVI MODIS product shows more pronounced interannual variation than the older AVHRR data, but the newer products, MOD09Q1-NDVI and MCD12Q2-EVI2, preserve interannual variation best. Vegetation types also vary in their degree of interannual variation, with croplands showing the least variance and slope parameters as low as 0.25 (Figs. 3 and 4, brown lines).

4. Discussion

4.1. Strong associations when using aggregate statistics

We generally find high correlations when aggregating both phenology ground observations by ecoregions, and remote sensing data by ecoregion × vegetation type. This type of aggregation reduces noise in any statistical setting, but here it also serves the useful purpose to make the datasets more representative of a predictor or response variable. We can investigate statistics for specific vegetation types within specific ecoregions, which show different behaviours with respect to both bias and precision. This method of aggregation yields stronger correlations and lower RMSE values than other studies have observed (Delbart et al., 2015; Delbart et al., 2005; White et al., 2009).

Of the phenophases included in this study, leaf out is likely to have the greatest direct influence on satellite observed green-up. In contrast, the flowering phenophases observed through the PlantWatch program are not expected to strongly influence landscape-level reflectance directly. Rather, they can and have been used here as indicators and proxy for the timing of leaf out of other species and green-up in general. It has previously been shown that phenophases of different species that occur approximately at the same day of year are highly correlated and can therefore serve as proxies for green-up of vegetation throughout the spring (Beaubien and Hamann, 2011b; Menzel et al., 2006). In contrast,

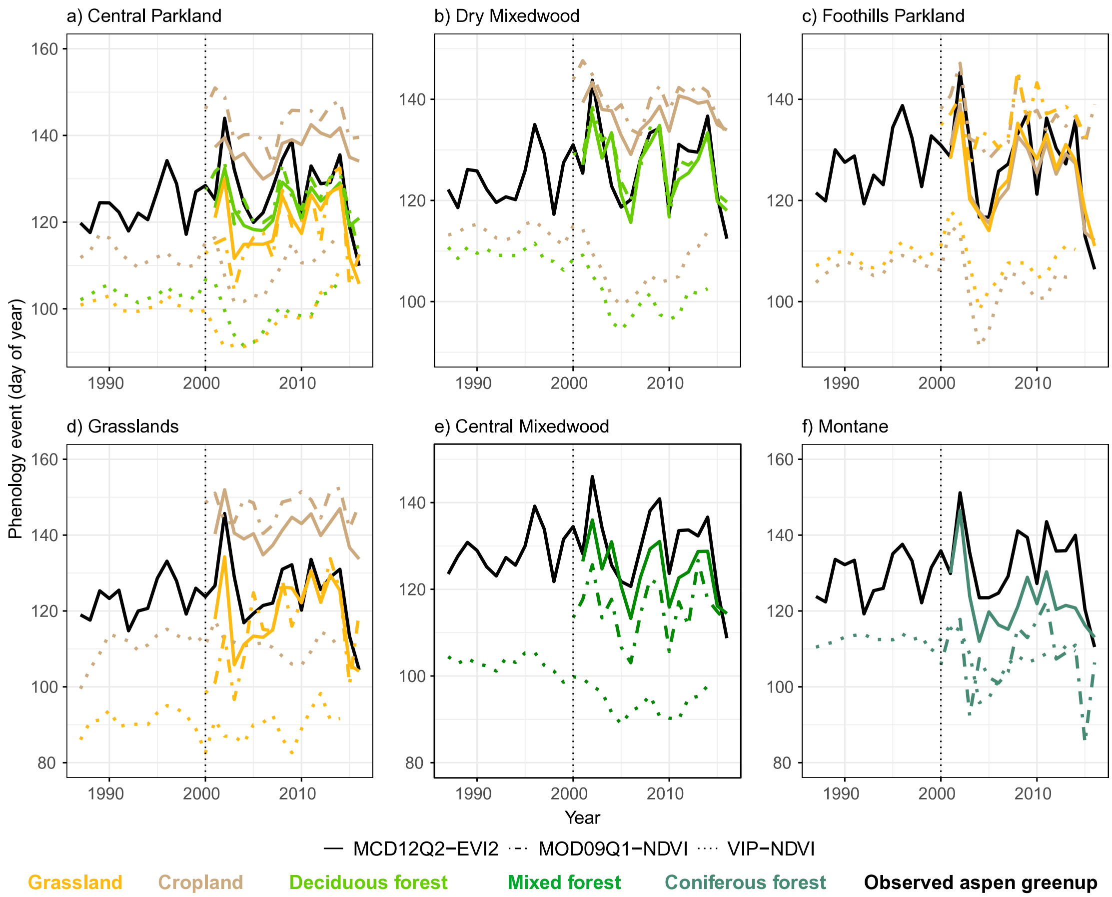

Fig. 3. Time series showing the remote sensing green-up dates for the three different remote sensing products by ecoregion (panels) and vegetation type (colour coded). The black line represents the ground-observed date for aspen leaf out in each ecoregion. The vertical dashed line is the year when the VIP-NDVI product switched from relying on AVHRR to MODIS sensors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the earliest and latest ground observations may not be well correlated and may therefore yield different correlation statistics with for different phenology products. The highest correlations would be expected for the best temporal matches between ground and remotely-sensed green-up estimates of different products. Appendix Table S1 is therefore useful in evaluating the overall quality of remote sensing products: those that never correlate high with any of the phenology time series (listed from early to late) should be considered the weakest.

4.2. Vegetation specific lag correction

Bias-statistics (RSME, MBE) are strongly influenced by analytical choices, such as the type of ground observation used (different species and phenology pahses), and ratios used to represent remotely sensed green-up (0.35 and 0.5). As such, bias statistics are not an indicator of the “quality” of the product but can be used for corrections. A relevant metric for general adjustment of phenology products from this study, would be observed aspen leaf out against “Deciduous” landcover in the “Central Parkland” and “Dry Mixedwood” ecoregions where aspen is dominant (rows 1 and 4 in Table 3). VIP-MODIS estimates appear very early, even considering the 0.35 ratio used. Mid-greenup estimates from MOD09Q1 and MCD12Q2 are quite close to aspen leaf out ground observations, perhaps still one week early, given that they are based on a 0.5 ratio method and given first leaf unfolding observations. Our results do conform to other research that concluded that the half-max (0.5 ratio) method corresponds well to the initial leafing of the canopy (Misra et al., 2016, White et al., 1997). Because the 0.5 ratio method is already somewhat early with respect to observed aspen leaf out in this study area, we would not recommend phase transitions based on the even earlier 0.15 ratio “greenup” product, which is also available for some phenology products. When making bias corrections against ground observations, the temporally closest phase transition should be chosen to maximise linearity in the bias correction.

Green-up estimates for cropland land are challenging (Delbart et al., 2015), with changes in reflectance of croplands being significantly influenced by crop type and management, as changes in NDVI due to tree foliation are weaker (Zhang et al., 2006). We observe green-up for agricultural areas to be generally later than forests and grasslands in the same region, which other North American studies also found (Zhang et al., 2006; Zhang et al., 2017). A possible cause of excessive negative bias can be false early green-up due to snow-melt (White et al., 2009). However, this does not seem to be the case for the MCD12Q2 product. Grasslands show a slight negative bias and deciduous and mixed forests are largely unbiased relative to ground observed phenology. The late phenology in croplands may be driven by local agricultural management, where farmers seed late, or it may have ecological causes, where grasslands have high heatsum requirements to avoid drought periods in spring. In Alberta, sufficient moisture for agricultural activities and green-up of grasslands come with early summer rains, relatively late in May.

Nevertheless, bias relative to any specific phenology event of interest can be corrected, if ground observations are available. In this case study for Alberta with aspen leaf out observations as a reference value, deciduous forest green-up estimated by the MCD12Q2-EVI2 product

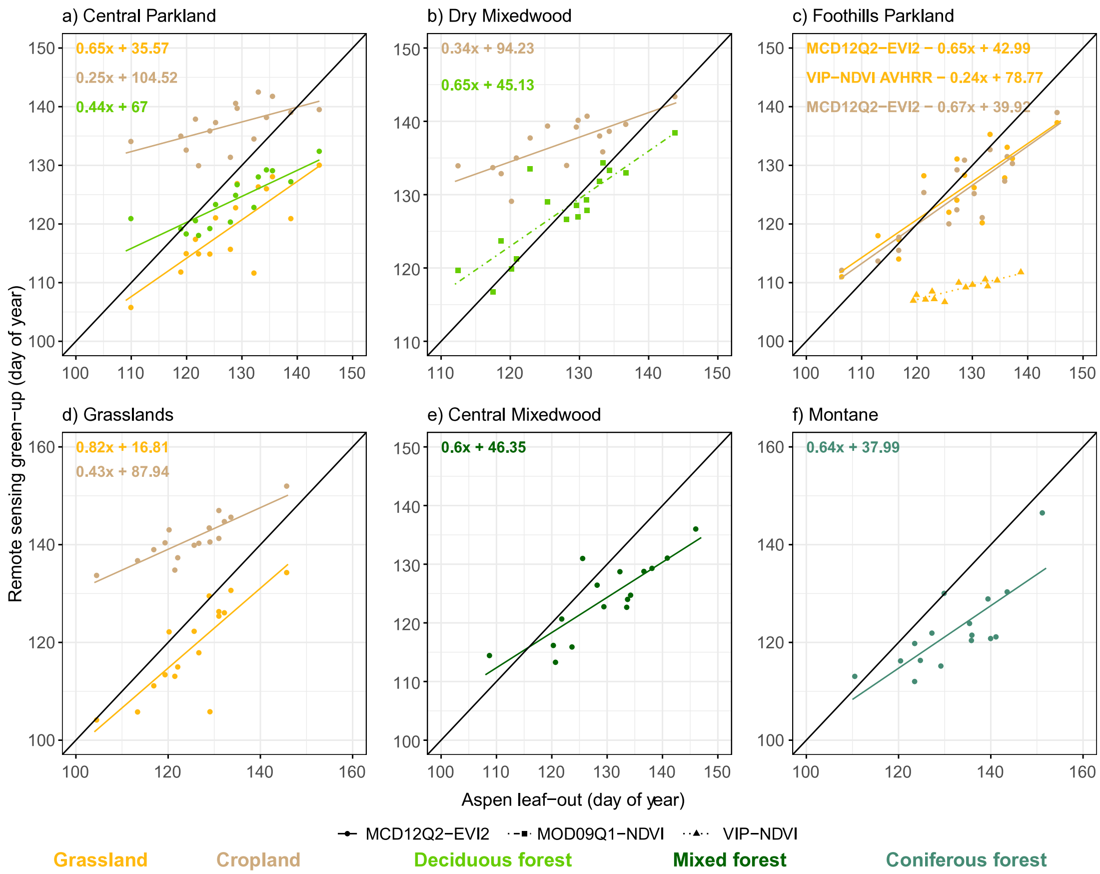

Fig. 4. Regressions of aspen leaf out from ground observations versus remotely sensed green-up date. The plots show the best correlation for each vegetation type (colors) by ecoregion (panels). The black diagonal is the 1:1 line of day of year. All other correlations are shown in Appendix 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

appears largely unbiased, grassland phenology can be linked to aspen leaf out by adding 5 days, and conifer forests green-up by adding 10 days.

4.3. Climate change response monitoring

All land surface phenology products evaluated in this study significantly underestimated the interannual variability of phenology, which has also been noted in other studies (Fisher and Mustard, 2007; Peng et al., 2018; White et al., 2009). Heterogeneous pixels generally show reduced interannual variability (Doktor et al., 2009), which can be caused by vegetation heterogeneity, cloud contamination, atmospheric variability, and sensor view angle (Zhang et al., 2003; Zhang et al., 2001). Without corrections from ground data, remotely sensed assessments of climate change impacts may therefore yield underestimates of the true change to land surface phenology. This may explain the difference of reported advances in spring phenology by 2–7 days per decade based on ground-observed plant phenology (Ahas et al., 2002; Badeck et al., 2004; Beaubien and Hamann, 2011b; Menzel et al., 2006; Root et al., 2003; Schwartz et al., 2006; Schwartz and Reiter, 2000), compared to remotely sensed estimates, typically ranging from 0 to 3 days per decade (Garonna et al., 2015; Jeong et al., 2011; Karkauskaite et al., 2017; Park et al., 2016).

Our interpretation is that the discrepancy is caused by spatial averaging and temporal compositing or smoothing algorithms. The least interannual variability was observed in the lowest resolution (approximately 5 km) VIP-EVI2 and VIP-NDVI products, which were further subjected to temporal smoothing algorithms. The AVHRR sensor data from these products from 1981 to 1999 was further limited by fewer spectral bands than the MODIS sensor, with the NIR channel sensitive to noise (Rao and Chen, 1996) and calibration difficulties (Van Leeuwen et al., 2006). The next best product with regards to maintaining interannual variability was the highest resolution (250 m) MOD09Q1-NDVI product that was based on an 8-day temporal composite time series. The 500-m resolution MCD12Q2-EVI2 used missing value imputation rather than temporal compositing, and maintained the highest interannual variability (usually with a slope of 0.5 to 0.8 for various ecoregion $\times$ vegetation type combinations. In summary, avoidance of temporal compositing and high resolution datasets appear to maintain interannual variability best. Nevertheless, to track trends of phenology in response to climate warming, remote sensing estimates of advances in green-up dates need to be adjusted with the appropriate regression equation to make these estimates comparable to ground phenology.

5. Conclusions

Overall, we find that the newer land surface phenology products, notably MCD12Q2-EVI2, exhibit significantly improved precision and less bias relative to ground observations, representing a significant advance over earlier phenology products. Different vegetation types showed a staged remotely sensed phenology in Alberta, with deciduous forest green-up first, followed by grasslands about 5 days later, and conifer forests green-up with a 10-day delay, allowing for corrections for different vegetation types. Our results do conform to other research that concluded that the half-max (0.5 ratio) method corresponds well to the initial leafing of the canopy and is therefore recommended for deciduous and mixed forests. All products showed reduced interannual variability compared to ground observations, which may lead to underestimating impacts of directional climate change. However, also in this respect MCD12Q2-EVI2 was superior, maintaining approximately 60% of the interannual variability in ground observations. Nevertheless, the analysis shows that remotely sensed time series of advances in leaf out may benefit from bias correction if ground observations are available.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgements

Funding to carry out the analysis presented in this paper was provided by the NSERC Discovery Grant RGPIN-330527 to AH. We thank all citizen scientists that contributed to the Alberta PlantWatch program led by EB, and we appreciate their enthusiasm and continued support of this program. We also thank the scientists and custodians of the NASA EOSDIS Land Processes Distributed Active Archive Center, and Paul Pickell for providing the land surface phenology data sets evaluated in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2022.103148.

References

Adan, M., et al., 2021. Use of earth observation satellite data to guide the implementation of integrated pest and pollinator management (IPPM) technologies in an avocado production system. Remote Sens Appl 23, 100566.
Ahas, R., Aasa, A., Menzel, A., Fedotova, V.G., Scheifinger, H., 2002. Changes in European spring phenology. Int. J. Climatol. 22 (14), 1727–1738.
Badeck, F.-W., et al., 2004. Responses of Spring Phenology to Climate Change. New Phytol. 162 (2), 295–309.
Bajocco, S., Guglietta, D., Ricotta, C., 2015. Modelling fire occurrence at regional scale: does vegetation phenology matter? European Journal of Remote Sensing 48 (1), 763–775.
Beaubien, E., Hamann, A., 2011a. Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada. Int. J. Biometeorol. 55 (6), 833–841.
Beaubien, E., Hamann, A., 2011b. Spring flowering response to climate change between 1936 and 2006 in Alberta. Canada. BioScience 61 (7), 514–524.
Bolton, D.K., et al., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sens. Environ. 240.
Bornez, K., Descals, A., Verger, A., Penuelas, J., 2019. Land surface phenology from VEGETATION and PROBA-V data. Assessment over deciduous forests. Int. J. Appl. Earth Obs. Geoinf. 84, 101974.
Butler, D., Cullis, B.R., Gilmour, A., Gogel, B., 2009. ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries, Brisbane.
Commission for Environmental Cooperation, 2013. 2010 Land Cover of North America at 250 meters - National Geospatial Data Asset (NGDA) Land Use Land Cover. Produced by Natural Resources Canada/ The Canada Centre for Mapping and Earth Observation (NRCan/CCMEO), United States Geological Survey (USGS); Insituto Nacional de Estadística y Geografía (INEGI), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) and Comisión Nacional Forestal (CONAFOR). Access 2019-03-05 from https://www.sciencebase.gov/arcgis/rest/services/Catalog/57d9d4b1e4b090824ffb1026/MapServer/.
Cui, T., Martz, L., Guo, X., 2017. Grassland Phenology Response to Drought in the Canadian Prairies. Remote Sens. (Basel) 9 (12), 1258.
De Angelis, A., Bajocco, S., Ricotta, C., 2012. Modelling the phenological niche of large fires with remotely sensed NDVI profiles. Ecol. Model. 228, 106–111.
Delbart, N., Beaubien, E., Kergoat, L., Le Toan, T., 2015. Comparing land surface phenology with leafing and flowering observations from the PlantWatch citizen network. Remote Sens. Environ. 160, 273–280.
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 (1), 26–38.
Didan, K. and Barreto, A., 2015. NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology NDVI Yearly Global 0.05Deg CMG . NASA EOSDIS Land Processes DAAC. Accessed 2019-10-02 from 10.5067/MEaSUREs/VIP/VIPPHEN_NDVI.004. (Oct 2, 2019).
Didan, K. and Barreto, A., 2016. NASA MEaSUREs Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG . NASA EOSDIS Land Processes DAAC. Accessed 2019-10-02 from 10.5067/MEaSUREs/VIP/VIPPHEN_EVI2.004. (Oct 2, 2019).
Didan, K. et al., 2018. Algorithm Theoretical Basis Document and User Guide. version 4.1. Vegetation Index & Phenology Lab, the University of Arizona. Accessed from https://vip.arizona.edu/VIP_ATBD_UsersGuide.php.
Doktor, D., Bondeau, A., Koslowski, D., Badeck, F.-W., 2009. Influence of heterogeneous landscapes on computed green-up dates based on daily AVHRR NDVI observations. Remote Sens. Environ. 113 (12), 2618–2632.
Fisher, J.I., Mustard, J.F., 2007. Cross-scalar satellite phenology from ground, Landsat, and MODIS data. Remote Sens. Environ. 109 (3), 261–273.
Friedl, M., Gray, J. and Sulla-Menashe, D., 2019. MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006 . NASA EOSDIS Land Processes DAAC. Accessed 2021-01-02 from 10.5067/MODIS/MCD12Q2.006.
Fu, Y.S.H., et al., 2015. Declining global warming effects on the phenology of spring leaf unfolding. Nature 526 (7571), 104–107.
Gao, X.J., Gray, J.M., Reich, B.J., 2021. Long-term, medium spatial resolution annual land surface phenology with a Bayesian hierarchical model. Remote Sens. Environ. 261.
Garonna, I., Jong, R., Schaepman Michael, E., 2015. Variability and evolution of global land surface phenology over the past three decades (1982–2012). Glob. Chang. Biol. 22 (4), 1456–1468.
Helman, D., 2018. Land surface phenology: What do we really 'see' from space? Sci. Total Environ. 618, 665–673.
Huete, A., et al., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1), 195–213.
Jeong, S.-J., Ho, C.-H., Gim, H.-J., Brown Molly, E., 2011. Phenology shifts at start vs. end of growing season in temperate vegetation over the Northern Hemisphere for the period 1982–2008. Glob. Chang. Biol. 17 (7), 2385–2399.
Jiang, Z.Y., Huete, A.R., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sens. Environ. 112 (10), 3833–3845.
Karkauskaite, P., Tagesson, T., Fensholt, R., 2017. Evaluation of the Plant Phenology Index (PPI), NDVI and EVI for Start-of-Season Trend Analysis of the Northern Hemisphere Boreal Zone. Remote Sens. (Basel) 9 (5), 485.
Leinonen, I., Kramer, K., 2002. Applications of Phenological Models to Predict the Future Carbon Sequestration Potential of Boreal Forests. Clim. Change 55 (1), 99–113.
Li, L., et al., 2022. Satellite-based phenology products and in-situ pollen dynamics: A comparative assessment. Environ. Res. 204, 111937.
Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens. 33 (2), 457–465.
Menzel, A., et al., 2006. European phenological response to climate change matches the warming pattern. Glob. Chang. Biol. 12 (10), 1969–1976.
Misra, G., Buras, A., Menzel, A., 2016. Effects of different methods on the comparison between land surface and ground phenology—A methodological case study from south-western Germany. Remote Sensing 8 (9), 753.
Moon, M., Richardson, A.D., Milliman, T., Friedl, M.A., 2022. A high spatial resolution land surface phenology dataset for AmeriFlux and NEON sites. Sci. Data 9 (1).
Natural Regions Committee, 2006. Natural Regions and Subregions of Alberta. Compiled by D.J. Downing and W.W. Pettapiece. Government of Alberta. Pub. No. T/852. Accessed from https://www.albertaparks.ca/media/2942026/nrsrcomplete_may_06.pdf. (T/852).
Park, T., et al., 2016. Changes in growing season duration and productivity of northern vegetation inferred from long-term remote sensing data. Environ. Res. Lett. 11 (8), 084001.
Peng, D., et al., 2018. Scaling up spring phenology derived from remote sensing images. Agric. For. Meteorol. 256–257, 207–219.
Piao, S.L., et al., 2019. Plant phenology and global climate change: Current progresses and challenges. Glob. Chang. Biol. 25 (6), 1922–1940.
Pickell, P.D., et al., 2017. An early warning system to forecast the close of the spring burning window from satellite-observed greenness. Sci. Rep. 7, 14190.
Pouliot, D., Latifovic, R., Fernandes, R., Olthof, I., 2011. Evaluation of compositing period and AVHRR and MERIS combination for improvement of spring phenology detection in deciduous forests. Remote Sens. Environ. 115 (1), 158–166.
R Core Team, 2013. R: A language and environment for statistical computing. Accessed from. https://cran.r-project.org/.
Rao, C., Chen, J., 1996. Post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft. Int. J. Remote Sens. 17 (14), 2743–2747.
Reed, B.C., Schwartz, M.D., Xiao, X., 2009. Remote sensing phenology. Phenology of ecosystem processes. Springer 231–246.
Richardson, A.D., et al., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156–173.
Root, T.L., et al., 2003. Fingerprints of global warming on wild animals and plants. Nature 421, 57–60.
Rouse Jr, J., Haas, R.H., Deering, D., Schell, J. and Harlan, J.C., 1974. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Type II Report for the Period April 1973 - September 1973. Goddard Space Flight Center, Greenbelt, Maryland.
Scheifinger, H., et al., 2013. Monitoring, modelling and forecasting of the pollen season, Allergenic pollen. In: Sofiev, M., Bergmann, K.-.-C. (Eds.), Allergenic pollen. Springer, pp. 71–126.

Schwartz, M.D., Ahas, R., Anto, A., 2006. Onset of spring starting earlier across the Northern Hemisphere. Glob. Chang. Biol. 12 (2), 343–351.

Schwartz, M.D., Reed, B.C., White, M.A., 2002. Assessing satellite-derived start-of-season measures in the conterminous USA. Int. J. Climatol. 22 (14), 1793–1805.

Schwartz, M.D., Reiter, B.E., 2000. Changes in North American Spring. Int. J. Climatol. 20, 929–932.

Van Leeuwen, W.J., Orr, B.J., Marsh, S.E., Herrmann, S.M., 2006. Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. Remote Sens. Environ. 100 (1), 67–81.

Vermote, E., 2015. MOD09Q1 MODIS/Terra Surface Reflectance 8-Day L3 Global 250m SIN Grid V006 . NASA EOSDIS Land Processes DAAC. Accessed 2022-03-02 from 10.5067/MODIS/MOD09Q1.006.

White, K., Pontius, J., Schaberg, P., 2014. Remote sensing of spring phenology in northeastern forests: A comparison of methods, field metrics and sources of uncertainty. Remote Sens. Environ. 148, 97–107.

White, M., A., et al., 2009. Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. Glob. Chang. Biol. 15 (10), 2335–2359.

White, M.A., Running, S.W., Thornton, P.E., 1999. The impact of growing-season length variability on carbon assimilation and evapotranspiration over 88 years in the eastern US deciduous forest. Int. J. Biometeorol. 42 (3), 139–145.

Willmott, C.J., 1982. Some Comments on the Evaluation of Model Performance Bulletin American Meteorological Society 63(22): 1309-1313.

White, M.A., Thornton, P.E., Running, S.W., 1997. A continental phenology model for monitoring vegetation responses to interannual climatic variability. Global Biogeochemical Cycles 11 (2), 217–234.

Wu, X.D., Xiao, Q., Wen, J.G., You, D.Q., Hueni, A., 2019. Advances in quantitative remote sensing product validation: Overview and current status. Earth Sci. Rev. 196, 102875.

Younes, N., Joyce, K.E., Maier, S.W., 2021. All models of satellite-derived phenology are wrong, but some are useful: A case study from northern Australia. Int. J. Appl. Earth Obs. Geoinf. 97, 102285.

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. Biogeo. 111 (G4), G04017.

Zhang, X., et al., 2003. Monitoring vegetation phenology using MODIS. Remote Sens. Environ. 84 (3), 471–475.

Zhang, X. et al., 2001. Global vegetation phenology from AVHRR and MODIS data, IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217). IEEE, pp. 2262-2264.

Zhang, X., et al., 2017. Exploration of scaling effects on coarse resolution land surface phenology. Remote Sens. Environ. 190, 318–330.