Remote Sensing of Environment 194 (2017) 155–160

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A Bayesian hierarchical model for estimating spatial and temporal variation in vegetation phenology from Landsat time series

Cornelius Senf [a,*], Dirk Pflugmacher [a], Marco Heurich [b], Tobias Krueger [a,c]

[a] Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[b] Bavarian Forest National Park, Freyungerstr. 2, 94481 Grafenau, Germany
[c] Integrative Research Institute on Transformation of Human-Environment Systems (IRI THESys), Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 4 November 2016
Received in revised form 12 March 2017
Accepted 24 March 2017
Available online 30 March 2017

Keywords:
Vegetation phenology
Hierarchical model
Climate change
Bavarian Forest

ABSTRACT

Phenology is a key indicator of vegetation response to global change. Satellite observations from coarse resolution sensors (i.e., MODIS and AVHRR) have been widely used to study impacts of climate anomalies on vegetation phenology. The advantage of coarse resolution sensors are daily observations across the entire globe, but the coarse spatial resolution, as well as the relatively short time span covered by MODIS, are significant drawbacks in analyzing landscape-scale trends in phenology. Time series data from the medium resolution sensor Landsat may overcome these issues. However, because of Landsat's lower observation frequency, phenological methods developed with coarse resolution data are not directly transferable. Here, we demonstrate a new Bayesian hierarchical modeling framework for estimating inter-annual variation in vegetation phenology from Landsat time series while controlling for spatial variation. The method pools all available observations to estimate the spatial variation in phenological parameters, while specifically modeling inter-annual variation as random effect terms. The advantage of a Bayesian approach is the ability to incorporate prior knowledge from other phenology and climate observations to reduce variability in the estimates, as well as a more robust estimation of uncertainty. We demonstrate and evaluate the modeling framework with a case study of changing spring phenology in broad-leaved trees in the Bavarian Forest National Park in southern Germany. Results show that the model estimated the spatial and temporal variation in phenological parameters precisely. Temporal variation in start of season showed overall strong agreement with ground-based measures of bud-break variability ($r = 0.82$ [0.80–0.84]). Our proposed modeling framework will help to better monitor and understand changes in vegetation phenology at scales yet unexplored by the phenological community.

1. Introduction

Vegetation phenology describes the study of the timing of recurring plant life cycles and their connection to climate (Forrest and Miller-Rushing, 2010). Monitoring vegetation phenology is of great importance for the study, reporting, and management of global change impacts on vegetation (Skidmore et al., 2015). Vegetation phenology can be monitored in-situ (Chmielewski et al., 2013) and by earth observation (Fitchett et al., 2015). While the former allows for detailed measurements of biological events on single plants (Menzel et al., 2006), the latter enables a globally consistent and holistic view on the phenology of entire plant communities (Tang et al., 2016).

Phenological studies that utilize earth observation data commonly make use of coarse spatial resolution sensors such as the MODerate Resolution Imaging Spectroradiometer (MODIS; Keenan et al., 2014) and the Advanced Very-High Resolution Radiometer (AVHRR; White et al., 2009). MODIS and AVHRR provide daily observations of the Earth's surface, which allows for fitting phenological models to each year's time series to estimate annual phenological parameters (Jonsson and Eklundh, 2002). While MODIS and AVHRR have been the sensors most often used in remote sensing based phenological analysis, they have important drawbacks. First, they lack the spatial detail necessary for studying vegetation phenology in heterogeneous landscapes (Fisher et al., 2006). Second, since MODIS data are only available from 2000 on, the time period for studying trends in phenology is rather short to be able to detect patterns related to global climate change (Hamunyela et al., 2013). Third, estimating annual phenological parameters from MODIS or AVHRR assumes that each year's phenological parameters are sampled from an independent model, with different variances between years (i.e., an un-pooled model). By risking over-fitting the data within each year, an un-pooled model might lead to wrong conclusions about trends underlying inter-annual variation in phenological parameters (White et al., 2014; White et al., 2009). To

* Corresponding author.
E-mail address: cornelius.senf@geo.hu-berlin.de (C. Senf).

http://dx.doi.org/10.1016/j.rse.2017.03.020
0034-4257/

improve upon these drawbacks in phenological modeling, Landsat's long-term data record of fine-resolution observations holds a unique potential that is still not fully explored.

With the opening of the United States Geological Survey (USGS) archive in 2008 (Wulder et al., 2012), Landsat time series have become a new potential data source for monitoring vegetation phenology (Fisher et al., 2006). The Landsat sensor family spans >30 years of continuous earth observation data at a spatial resolution of 30 m and a potential 8-day repeat cycle in periods where different Landsat sensors overlap (Roy et al., 2014). However, in practice the data availability is much lower, because most regions outside the United States have not been continuously imaged, due to archive restrictions, cloud cover, and Landsat 7's failed Scan-Line Corrector in 2003 (Ju and Roy, 2008; Wulder et al., 2016). As a consequence, Landsat's cloud-free observation frequency ranges from bi-weekly to bi-monthly or even less observations per year, hampering the analysis of trends in phenological parameters using un-pooled phenological models (Coops et al., 2012). One solution to this problem is to pool all data into one model to estimate average phenological parameters independent of the year of observation (Fisher et al., 2006). Pooled models assume that all observations are sampled from one underlying model with equal variance. To quantify the inter-annual variations in model parameters, studies have used the annual deviations between the observations and the pooled model averaged over all pixels (Melaas et al., 2013). Unfortunately, this approach is strongly dependent on data availability, leading to imprecise estimates of inter-annual variability in phenological parameters that likely obscure the true underlying phenological trend (Nijland et al., 2016).

Building on the research by Melaas et al. (2013), we here demonstrate a Bayesian partially-pooled hierarchical model for estimating spatial and temporal variations in phenological parameters from Landsat time series. A key advantage of a partially-pooled hierarchical model is that model parameters are themselves defined as a model, whose parameters are also estimated from data (Gelman, 2006). Thus, while the full observation density is used for estimating average phenological parameters (as in a pooled model), the inter-annual variation in model parameters (e.g., in the start of season) is also specifically modeled from the data. The partial pooling hereby allows for individual variances to be drawn for each year, though all variances arise from a common underlying process. In this way, the estimation of inter-annual variation in a model parameter is less sensitive to inter-annual variations in data density, since in data scarce years the estimate of variance will approach the overall mean (Gelman, 2006). Finally, the Bayesian approach allows for quantifying uncertainty in parameter estimates that go beyond classical frequentist methods (Ellison, 2004), and it allows the incorporation of prior information such as land cover, climate, or topographic data to further improve the precision of parameter estimates.

The aim of this paper is to demonstrate and evaluate a new Bayesian hierarchical modeling approach for estimating trends in phenological parameters from Landsat time series. We utilize the newly developed modeling approach for estimating the spatial and temporal variation in spring phenology of broad-leaved forests in the Bavarian Forest National Park in southern Germany. Specifically, we estimate inter-annual variation in the start of season (green-up) and compare those estimates to ground based estimates of bud-break. We further discuss remaining challenges, potential enhancements, and practical issues with the proposed model. We conclude by linking this research with potential applications in global change research.

2. Model development

We used a five-parameter logistic function to model spring phenology as in Melaas et al. (2013):

$$g(t;\beta_i) = \beta_{1[i]} + \left(\beta_{2[i]} - \left(\beta_{5[i]} * t\right)\right) * \frac{1}{\left(1 + e^{-\beta_{3[i]} * \left(t - \beta_{4[i]}\right)}\right)} \quad (1)$$

where $\beta_i = \{\beta_{1[i]}, \beta_{2[i]}, \beta_{3[i]}, \beta_{4[i]}, \beta_{5[i]}\}$ is a five-dimensional vector of model parameters for each pixel i with β_1 the minimum (*seasonal minimum*), β_2 the difference between minimum and maximum (*seasonal amplitude*), β_3 the inflection point (*green-up rate*), and β_4 the timing of the inflection point (*start of season*). The seasonal minimum describes the off-season (i.e. winter) spectral background value. The seasonal amplitude describes the spectral difference between off-season and on-season activity of the observed vegetation. The start of season describes the approximate timing of transition from off-season to on-season condition, whereas the green-up rate describes the speed of transition. The parameter β_5 controls for an often observed decline in vegetation greenness in the summer months (*greendown factor*; Elmore et al., 2012). Model residuals are modeled as normally distributed with equal variance σ^2, yielding for each time step t and each pixel i:

$$y_{ti} = N\left(g(t, \beta_i), \sigma^2\right) \quad (2)$$

Since the five phenological parameters are likely correlated with each other, the vector β_i is assumed to follow a multivariate normal distribution $MVN(\mu_\beta, \Sigma_\beta)$ with μ_β being the vector of mean model parameters and Σ_β being the variance-covariance matrix of the five model parameters. Since the five model parameters are measured on very different scales (i.e., day of year versus spectral index values), which can cause instability in sampling the posterior distribution, we re-parameterize the model to a non-centered formulation (Bernardo et al., 2003). Therefore, we re-write the multivariate normal formulation of β_i as:

$$\beta_i = \left(diag\left(\sigma^2_\beta * \tau\right) * L_\beta * z_\beta\right)^T + \mu_\beta \quad (3)$$

where σ^2_β is a vector of variances for each model parameter, which are scaled by a scale vector τ; L_β is the Cholesky decomposition of a 5×5 correlation matrix C_β with $C_\beta = L_\beta{}^* L_\beta{}^T$; and z_β is a vector of five $N(0,1)$ random variables. We assign a weakly-informative $LKJ(2)$ prior on the correlation matrix C_β and weakly-informative $half\text{–}Cauchy(0,1)$ priors on the model parameter variances. Adding the vector μ_β centers the posterior distribution of the five model parameters according to prior guesses on their approximate location. The centering vector μ_β and the scale vector τ must be given by the user and we give recommendations on how it can be determined in the later sections.

To account for variations in the start of season parameter β_4 between years j, we add a hierarchical level to our model:

$$\begin{aligned} y_{ijt} &= N\left(g\left(t, \beta'_{ij}\right), \sigma^2\right) \\ \beta'_{ij} &= \left\{\beta_{1[i]}, \beta_{2[i]}, \beta_{3[i]}, \beta_{4[i]} + \phi_{[j]}, \beta_{5[i]}\right\} \end{aligned} \quad (4)$$

where ϕ_j is a random effect allowing for inter-annual variation in the start of season parameter. The model specified in Formula (4) thus allows for estimating per pixel phenological parameters (β_i) while simultaneously accounting for variation in the start of season among years (ϕ_j). We here like to note that for each of the five phenological parameters a random effect might be added, though for demonstration purposes we here present only one annually varying phenological parameter (see the Discussion for further information). The random effect in start of season is assumed to follow a normal distribution $N(0, \sigma^2_\phi)$ with zero mean and variance σ^2_ϕ, which is assigned a weakly informative $half\text{–}N(0,5)$ prior. The model is made complete by assigning a weakly informative $half\text{–}Cauchy(0,1)$ prior on σ^2.