## ORIGINAL PAPER

# Comparison of the driving forces of spring phenology among savanna landscapes by including combined spatial and temporal heterogeneity

Likai Zhu · Jane Southworth · Jijun Meng

Received: 13 May 2014 / Revised: 13 December 2014 / Accepted: 14 December 2014 / Published online: 27 December 2014 © ISB 2014

Abstract Understanding spatial and temporal dynamics of land surface phenology (LSP) and its driving forces are critical for providing information relevant to short- and long-term decision making, particularly as it relates to climate response planning. With the third generation Global Inventory Monitoring and Modeling System (GIMMS3g) Normalized Difference Vegetation Index (NDVI) data and environmental data from multiple sources, we investigated the spatio-temporal changes in the start of the growing season (SOS) in southern African savannas from 1982 through 2010 and determined its linkage to environmental factors using spatial panel data models. Overall, the SOS occurs earlier in the north compared to the south. This relates in part to the differences in ecosystems, with northern areas representing high rainfall and dense tree cover (mainly tree savannas), whereas the south has lower rainfall and sparse tree cover (mainly bush and grass savannas). From 1982 to 2010, an advanced trend was observed predominantly in the tree savanna areas of the north, whereas a delayed trend was chiefly found in the floodplain of the north and bush/grass savannas of the south. Different environmental drivers were detected within tree- and grassdominated savannas, with a critical division being represented by the 800 mm isohyet. Our results supported the importance of water as a driver in this water-limited system, specifically preseason soil moisture, in determining the SOS in these

L. Zhu (⊠) · J. Southworth Department of Geography, University of Florida, 3141 Turlington Hall, P.O. Box 117315, Gainesville, FL 32611, USA e-mail: likaizhu@ufl.edu

J. Southworth e-mail: jsouthwo@ufl.edu

#### J. Meng

College of Urban and Environmental Sciences, Peking University, Beijing 100871, China e-mail: jijunm@pku.edu.cn water-limited, grass-dominated savannas. In addition, the research pointed to other, often overlooked, effects of preseason maximum and minimum temperatures on the SOS across the entire region. Higher preseason maximum temperatures led to an advance of the SOS, whereas the opposite effects of preseason minimum temperature were observed. With the rapid increase in global change research, this work will prove helpful for managing savanna landscapes and key to predicting how projected climate changes will affect regional vegetation phenology and productivity.

Keywords Southern Africa  $\cdot$  Savanna  $\cdot$  Land surface phenology  $\cdot$  Spatial panel data models  $\cdot$  The start of the growing season (SOS)

## Introduction

Phenology is the study of the timing of recurring biological cycles and their connection to climate (White and Thomton 1997; White et al. 2009). It can influence the exchange of energy, water vapor, and momentum between the land surface and the atmosphere and, therefore, is critical to understand the global carbon and water cycles (Cong et al. 2013). Changes in phenology also affect the abundance and distribution of species (Enquist et al. 2014). In turn, changes in the timing of vegetation phenology are widely considered to be one of the most sensitive biological responses to climate change and an excellent predictor of changes already underway (IPCC 2013). An earlier occurrence of spring phenology, which closely correlates with rising temperatures, has been observed in northern latitudes from both satellite measurements (Myneni et al. 1997; White et al. 2009; Cong et al. 2013;

Dai et al. 2013) and in situ observations (Menzel and Fabian 1999; Parmesan and Yohe 2003; Menzel et al. 2006). As such, land surface phenology (LSP) research is of great importance and an increase in phenology information is essential as this facilitates the achievement of natural resource management goals and supports informed decision making related to climate change (Enquist et al. 2014).

With the applications of remote sensing in monitoring and characterizing vegetation phenology, the term LSP has been used to refer to the seasonal patterns of variation in vegetated land surfaces, particularly those observed using remote sensing. LSP is distinguished from plant phenology which refers to specific life cycle events such as budburst, flowering, or leaf senescence using in situ observations of individual plants and species (de Beurs and Henebry 2010). LSP is based upon "wall-to-wall" observations of phenology at larger geographic scales instead of plant-specific observations. The relationship between satellite measures of LSP and specific plant phenophases is still ambiguous (White et al. 2009; Beurs and Henebry 2010), but landscape up-scaling approaches are being developed to validate LSP from plot-level observations (Liang et al. 2011). For this reason, the terms and definitions about phenological metrics in LSP, as well as the methods to derive these metrics, are diverse (White et al. 2009; Beurs and Henebry 2010; Atkinson et al. 2012; Cong et al. 2013). For example, the terms such as green-up onset, leaf-unfolding, green wave, and start of the growing season (SOS) appear to be interchangeable in the literature (White et al. 2009). Here, we adopted the term SOS which most authors used to represent the spring phenophase.

Savannas are globally important ecosystems given they support a large proportion of the world's human population and provide food and habitat for both livestock and wildlife. While a characteristic feature of savanna systems is the coexistence of trees and grasses, tree cover is a chief determinant of ecosystem properties. Studies have shown that the factors that determine the structure and function of savanna systems vary across physiographic gradients (Sankaran et al. 2005; Campo-Bescós et al. 2013a, b; Zhu and Southworth 2013). Sankaran et al. (2005) have reported that savannas receiving a mean annual precipitation (MAP) of less than about 650 mm may be considered "stable systems" in which water constrains woody cover and permits grasses to coexist. In contrast, savannas are considered "unstable" above a MAP of about 650 mm and in which disturbances are required for the coexistence of trees and grasses. Campo-Bescós et al. (2013a, b) have found that the environmental drivers of savanna vegetation growth appear to transition from being dominated by precipitation and soil moisture in the grass-dominated regions with MAP <750 mm, to being dominated by fire, potential evapotranspiration (PET), and temperature in tree-dominated regions with MAP >950 mm. Zhu and Southworth (2013) determined that the relationship between mean annual net primary

productivity (NPP) and MAP varies with an increase in MAP, characterized by a linear relationship that changes abruptly when MAP exceeds around 850–900 mm in the same southern Africa region. The potential distinction in environmental factors that determine LSP among different savanna types or across physiographic gradients is worth exploring.

Previous studies have identified the critical role of a single factor (e.g., precipitation, soil moisture, or temperatures) in affecting LSP of savannas. Chidumayo (2001) discovered that the most important determinant of savanna phenology in southern Africa was the interaction between minimum and maximum temperatures. Zhang et al. (2005) found that welldefined thresholds exist in cumulative rainfall that stimulates vegetation green-up in arid and semi-arid African savannas. However, few studies have been conducted to determine how a suite of environmental covariates studied in concert control the changes of phenology in savanna landscapes. In terms of methodology, time-series or cross-sectional data analysis methods are being adopted to explore the intra- and interannual variations in vegetation growth and to examine their driving forces, but no analysis has been made to incorporate time-series and cross-sectional data simultaneously and to consider both spatial and temporal heterogeneity. Our research applied spatial panel data models to explore the spatial and temporal dynamics of the SOS across key physiographic gradients in southern Africa. Spatial panel data offers us extended modeling possibilities as compared to a single time-series or cross-sectional setting (Elhorst 2012).

The purpose of this study is to identify the dominant environmental drivers of SOS in the Okavango–Kwando– Zambezi (OKZ) catchments and how they vary across physiographic gradients. Specifically, we addressed the following research questions: (1) how does SOS change spatially and temporally in the OKZ catchments from 1982 to 2010? and (2) are there any differences in the driving factors of SOS across different savanna landscapes?

## Materials and methods

## Study area

The OKZ catchment covers about 693,000 km<sup>2</sup> in Zambia, Angola, Namibia, and Botswana (Fig. 1a). The elevation ranges from 800 to 1900 m. The combined basins have a MAP range of 400–1400 mm. The high variability of intraand inter-annual rainfall within the OKZ catchment is due to the movements of the Inter-Tropical Convergence Zone (ITCZ), El Niño Southern Oscillation (ENSO) events, and sea surface temperatures in the Indian and Atlantic Oceans (Gaughan and Waylen 2012). The southern portion is characterized by lower MAP and typically unpredictable patterns of precipitation, while the northern portion has higher MAP and Fig. 1 Geographic location of the study area. a The study area is divided into two regions: one with mean annual precipitation (MAP) more than 800 mm and another with MAP less than 800 mm. The spatial units within each region are defined by MAP intervals and boundaries of watersheds. b The spatial pattern of savanna types



lower inter-annual variability (Fig. 1a). The intra-annual distribution of precipitation is uneven although on average the rainy season ranges from October to April and the dry season from May to September. The majority of the Okavango and Kwando catchments and the headwaters of all three basins are located in Angola. Low topography in the south (especially in the Caprivi region of Namibia and in northern Botswana) makes clear hydrologic separation of the catchments difficult. Kalahari sand characterizes a large portion of the region's soil. Human settlements are mainly distributed along the water courses, especially along rivers in the Caprivi region of Namibia, which makes the human–wildlife conflicts acute in the dry season (Chase and Griffin 2009). Varied resource management approaches are applied in response to significant climate variability, for example, the establishment of protected areas.

Savanna is the dominant biome across the study area, in spite of the climatic and edaphic gradients. Figure 1b illustrates the spatial pattern of savanna types as generated by the Joint Research Center, European Commission. The detailed land cover classes were aggregated into three categories: tree savanna, bush savanna, and grass savanna. Tree savanna is defined as a tree-dominated savanna with tree canopy cover more than 15 % and canopy height more than 5 m; bush savanna is defined as a shrub-dominated savanna type with shrub canopy cover greater than 15 % and canopy height less than 5 m with no or a sparse tree layer; and grassland savanna is defined as a grass-dominated savanna type with herbaceous cover greater than 15 % and tree and shrub canopy cover less than 20 %. Tree savannas (tree-dominated) are largely distributed at higher elevations within the upper catchment, while a composition of tree-shrub-grass (grass-dominated savannas) is distributed in the lower catchment or at lower elevations (Fig. 1). The dominant vegetation types are Drachystegia-Julbernardia (miombo) woodland, Colophospermum mopane (mopane) woodland, and Acacia (munga or thorn) woodland. The study area was divided into 56 zones by different precipitation intervals and boundaries of the three watersheds (Fig. 1a). The gradients from north to south can be captured by isohyets and potential variations in SOS and their drivers from east to west (oceanic influence, topography, etc.), may be considered by the three watersheds.

#### Data sets

Our study used the newly available Global Inventory Monitoring and Modeling System (GIMMS3g) Normalized Difference Vegetation Index (NDVI) data which were generated from NOAA's Advanced Very High Resolution Radiometer (AVHRR) data in the framework of the GIMMS project at the NASA Goddard Space Flight Center. The data set was processed in a way, consistent with, and quantitatively comparable to, NDVI generated from improved sensors such as MODIS and SPOT-4 Vegetation and was corrected for dropped scan lines, navigation errors, data dropouts, edge-of-orbit composite discontinuities, and other artifacts (Tucker et al. 2005). The empirical mode decomposition/reconstruction method was applied to minimize the effects of orbital drift by removing common trends between the time series of Solar Zenith Angle and NDVI (Fensholt et al. 2013). Moreover, the data set used the maximum NDVI value over a 15-day period to represent each 15-day interval to minimize corruption of vegetation signals from atmospheric effects, cloud contamination, and scan angle effects at the time of measurement (Bi et al. 2013). The GIMMS3g NDVI data span from July 1981 to December 2011 and have a spatial resolution of around 8 km and a temporal resolution of about 15 days. Despite the corrections and temporal compositing, the GIMMS3g data still contain residual invalid measurements, well indicated by quality flags (de Jong et al. 2013). Any pixel, with a time series having less than 80 % high-quality data points (flag=1 and flag=2) through the whole time span, was excluded from our analysis. For the time points with "poor" quality (flag>2), we used a gap-filling procedure to interpolate the NDVI values of these time points (Jin et al. 2013).

We utilized the Willmott, Matsuura and Collaborators global climatic data of monthly temperature, monthly precipitation, and monthly PET from 1982 through 2010 (Table 1). They have a spatial resolution of  $0.5^{\circ}$  by  $0.5^{\circ}$  with grid nodes centered on 0.25°. These data sets improve on previous ones due to the use of a refined Shepard interpolation algorithm and an increased number of neighboring station points (Legates and Willmott 1990). The grid nodes within and surrounding the study area were interpolated into continuous surfaces at about 8×8 km resolution using the inverse distance weighted interpolation method. We also used the NCEP-DOE Reanalysis II global maximum temperature, minimum temperature, and downward solar radiation flux data with a format of the Global T62 Gaussian grid. Such irregular Gaussian grids were converted into continuous surfaces at 8×8 km resolution also by the inverse distance weighted interpolation method. Downward solar radiation refers to the shortwave radiation flux reaching the Earth's surface. It essentially depends on the solar zenith angle, cloud coverage, and to a lesser extent on atmospheric absorption and surface albedo. The Climate Prediction Center (CPC) global monthly soil moisture data from 1982 to 2010 were used, which are at a spatial resolution of  $0.5^{\circ}$  with grid nodes centered on 0.25°. The data are derived from a one-layer "bucket" water balance model using CPC monthly precipitation and temperatures as inputs (Fan and van den Dool 2004).

## Methods

#### Determining the SOS

Two steps are usually applied to identify the SOS using vegetation indices (VI) time-series (Zhang et al. 2005; White et al. 2009; Cong et al. 2012, 2013). The first step is to depress the noise and fit the shape of the VI curve using methods such

as harmonic analysis and piecewise functions (Jönsson and Eklundh 2002; Jönsson and Eklundh 2004; Beurs and Henebry 2010; Atkinson et al. 2012; Cong et al. 2012). The second step is to calculate the SOS which can be considered as the day of the year (DOY) that the VI crosses a specific threshold of VI or VI ratio in an upward direction (White et al. 2009), the DOY that the positive derivative of the VI curve reaches the highest point (Cong et al. 2012, 2013), or the DOY that the first local maxima of the curvature of the VI curve appears (Zhang et al. 2005). Here, we used piecewise functions to depress noise of the NDVI time series and defined the DOY when the NDVI ratio crossed a threshold of 50 %, as the SOS. The asymmetric Gaussian (AG) function and the double logistic (DL) function within the TIMESAT software was used to smooth the NDVI time series of each pixel within the study area from 1982 to 2010 (Eklundh and Jönsson 2012). A detailed description of the algorithm was given in Eklundh and Jönsson (2012).

The Bayesian information criterion (BIC) was used to measure the model performance by penalizing the number of parameters, whose formula can be written as follows:

$$BIC = T\left(\ln\left[\hat{\sigma}^2\right]\right) + k \cdot \ln(T)$$
(1)

where *T* is the number of input data points,  $\hat{\sigma}^2$  is the error variance, and *k* is the number of parameters. For the AG function, *k* is equal to 7, and for the DL function, *k* is equal to 6 (Atkinson et al. 2012).The lower the BIC value is, the higher preference the model has. For a specific pixel, if the BIC value for the AG function was smaller than that for the DL function, the NDVI time series smoothed by the AG function is selected to determine the SOS, and vice versa. Here, the SOS was defined as the DOY that the NDVI ratio

 Table 1
 Variables used in the spatial panel data models and their respective data sources

Variable	Symbol	Data set	Source
The start of the growing season	SOS	GIMMS3g 15-day NDVI	http://ecocast.arc.nasa.gov/data/pub/gimms/3g
Preseason mean temperature	meanT	Matsuura and Willmott's monthly temperature	http://climate.geog.udel.edu/~climate/html_pages/ download.html#P2011
Preseason mean maximum temperature	maxT	NCEP-DOE Reanalysis II monthly maximum temperature	http://www.esrl.noaa.gov/psd/data/gridded/data.ncep. reanalysis2.gaussian.html
Preseason mean minimum temperature	minT	NCEP-DOE Reanalysis II monthly manimum temperature	http://www.esrl.noaa.gov/psd/data/gridded/data.ncep. reanalysis2.gaussian.html
Preseason total downward solar radiation	Rad	NCEP-DOE Reanalysis II monthly downward solar radiation	http://www.esrl.noaa.gov/psd/data/gridded/data.ncep. reanalysis2.gaussian.html
Preseason total precipitation	Р	Matsuura and Willmott's monthly precipitation	http://climate.geog.udel.edu/~climate/html_pages/ download.html#P2011
Preseason mean potential evapotranspiration	PET	Matsuura and Willmott's monthly PET	http://climate.geog.udel.edu/~climate/html_pages/ download.html#P2011
Preseason mean soil moisture	SoilW	CPC monthly soil moisture	http://www.esrl.noaa.gov/psd/data/gridded/tables/ monthly.html

reaches 50 % of the highest NDVI ratio. The NDVI ratio (NDVI<sub>ratio</sub>(t)), which represents the state of the ecosystem, is transformed from the NDVI by

$$NDVI_{ratio}(t) = \frac{NDVI_t - NDVI_{min}}{NDVI_{max} - NDVI_{min}}$$
(2)

where NDVI<sub>t</sub> is the NDVI value at time t, and NDVI<sub>max</sub> and NDVI<sub>min</sub> are the maximum and minimum values of the annual NDVI curve. A 50 % point suggests that a certain pixel has attained 50 % of its maximum greenness. The justification for the choice of the 50 % threshold is that the increase in greenness is believed to be most rapid at this threshold. Furthermore, the vegetation signals below this level tend to be confounded with soil reflectance (White et al. 1997, 2009; Beurs and Henebry 2010; Cong et al. 2013). Based on the above method, we derived the SOS for each vegetated pixel within our study area from 1982 to 2010.

## Spatial panel data models

Spatial panel data models were used to quantify the empirical relationships between environmental covariates and the SOS, which can allow for both cross-sectional and time-period dependence, and also enables researchers to consider heterogeneity (Lee and Yu 2010). Compared to more traditional methods using cross-sectional or time-series data alone, spatial panel data models can take into consideration interindividual and time-period differences and thus are widely used in agricultural economics, transportation research, economics, and land change science (Lee and Yu 2010; Wang et al. 2013; Li et al. 2013).

A simple linear model between a dependent variable Y and a set of K independent variables X is given as follows:

$$y_{it} = x_{it}\beta + \varepsilon_{it} \tag{3}$$

where *i* (=1, ..., *N*) represents a spatial unit, *t* (=1,...,*T*) denotes a time period, and  $x_{it}$  is an array of observations for m independent variables ([ $N \times T$ ] × *m*).  $\beta$  is a *m*×1 vector which indicates fixed but unknown parameters, and  $\varepsilon_{it}$  is an independently and identically distributed error term for all *i* and *t* values, with zero mean and variance  $\sigma^2$ . The main drawback of this model is the failure to account for spatial and temporal heterogeneity, because spatial units and time periods tend to have spatial or temporal heterogeneity, and space- and time-specific variables do influence the dependent variable.

One remedy is to incorporate a variable intercept  $\mu_i$  and/or  $\gamma_t$  representing the effect of the omitted variables that are peculiar to each spatial unit and/or time period. A simple panel data model with spatial specific effects is

$$y_{it} = \mu_i + x_{it}\beta + \varepsilon_{it} \tag{4}$$

where  $\mu_i$  denotes spatial specific effects. Likewise, a model with time-period specific effects can be expressed as follows:

$$y_{it} = \gamma_t + x_{it}\beta + \varepsilon_{it} \tag{5}$$

where  $\gamma_t$  represents time-period specific effects. The spatial and time-period specific model can be formulated as

$$y_{it} = \mu_i + \gamma_t + x_{it}\beta + \varepsilon_{it} \tag{6}$$

The spatial specific effects might be treated as fixed effects or random effects. In the models with fixed effects,  $\mu_i$  acts as a dummy variable for each spatial unit, while in the models with random effects,  $\mu_i$  is treated as a random variable that is independently and identically distributed with zero mean and variance  $\sigma_{\mu}^2$ . A similar differentiation is applicable to timeperiod-specific effects. A likelihood ratio test is a statistical test used to compare the fit of two models, one of which (the null model) is a special case of the other (the alternative model). Such tests are based on the likelihood ratio, which indicates how many times more likely that the data are under one model than the other (Vuong 1989; Elhorst 2009). A p value or a comparison to a critical value can be computed to decide whether to reject the null model in favor of the alternative model. We ran likelihood ratio (LR) tests to justify the extension of the model with spatial and/or time-period fixed effects.

The specific effects models can be extended to include spatial error autocorrelation and/or spatially lagged dependent variables. For example, the spatial specific model including spatial error autocorrelation can be specified as follows:

$$y_{it} = \mu_i + x_{it}\beta + \phi_{it} \tag{7}$$

$$\phi_{it} = \lambda \sum_{j=1}^{N} w_{ij} \phi_{jt} + \varepsilon_{it}$$
(8)

where  $w_{ij}$  is an element of the spatial weights matrix, indicating the proximity of two observational units.  $\lambda$  is usually called the spatial autocorrelation coefficient. The spatial specific model including a spatially lagged dependent variable can be written as

$$y_{it} = \mu_i + \rho \sum_{j=1}^N w_{ij} y_{jt} + x_{it} \beta + \varepsilon_{it}$$
(9)

where  $\rho$  refers to the spatial autoregressive coefficient. Whether the spatial lag model and/or the spatial error model are more appropriate to describe the data than a model without any spatial interaction effects can be tested using the robust Lagrange multiplier (LM) test. The LM test is a statistical test of a simple null hypothesis that a parameter of interest  $\theta$  is equal to some particular value  $\theta_0$  (Buse 1982). The main advantage of the LM test is that it does not require an estimate of the information under the alternative hypothesis or unconstrained maximum likelihood. The LR and LM test are asymptotically equivalent tests of hypotheses. When testing nested models, the statistics for each test converge to a chi-squared distribution with degrees of freedom equal to the difference in degrees of freedom in the two models. The robust LM test can test for a spatially lagged dependent variable in the presence of spatial error autocorrelation, and for spatial error correlation in the presence of a spatially lagged dependent variable (Elhorst 2012).

The SOS (response variable) and a suite of environmental covariates (explanatory variables) from 1982 through 2010 were incorporated into spatial panel data models (Table 1). The environmental variables include preseason mean temperature (meanT), preseason mean maximum temperature (maxT), preseason mean minimum temperature (minT), preseason total downward solar radiation (Rad), preseason total precipitation (P), preseason mean potential evapotranspiration (PET), and preseason mean soil moisture (soilW). For each spatial unit and each year, we estimated the month in a given

year based on the SOS (the day of the year). If the SOS falls into the first half of the month, the "preseason" is defined as the 2 months before the month which the SOS falls. For example, if the SOS is October 8, the combined August and September are considered as the preseason period. If the SOS falls into the second half of the month, the preseason is defined as the current and last month. For example, if the SOS is October 30, the combined months, September and October, are considered as the preseason period. The 2-month mean or sum values were calculated for each spatial unit and each year. The spatial patterns of the average of these environmental covariates (1982–2010) are illustrated in Fig. 2. The time series of response and explanatory variables were derived from the average values of the 56 zones from 1982 to 2010 (Fig. 1a).

We calculated Pearson's correlation coefficient between the SOS and each environmental driver using pooled spatial panel data (Table 2). The candidate environmental drivers have significant correlation with the SOS in the region with MAP less than 800 mm or in the region with MAP more than 800 mm. As for the region (MAP <800 mm), the variables, meanT, minT, and PET have a positive correlation with the SOS, indicating that their higher values tend to delay the occurrence of the SOS. In contrast, the variables, maxT, P, and SoilW have a negative relationship with the SOS, suggesting that their higher values tend to advance the SOS. In



Fig. 2 Spatial patterns of the mean environmental variables (1982–2010): a preseason temperature, b preseason maximum temperature, c preseason minimum temperature, d preseason downward solar radiation, e preseason precipitation, f preseason potential evapotranspiration, and g preseason soil moisture regard to the region (MAP >800 mm), the driver, PET, has a positive correlation with the SOS, whereas temperature-related variables, Rad and P, have a negative correlation with the SOS.

The spatial panel data of variables should be stationary so that they can be included in the models (Li et al. 2013). Here, we applied the test proposed by Moon and Perron (2004). The null hypothesis is that each individual time series contains a unit root, whereas the alternative hypothesis is that the spatial panel data are stationary. Given space limitations, the detailed description of the test can be found in Moon and Perron (2004). The spatial panel data models, LR, and LM tests used in this study were coded by the authors in MATLAB (Mathworks Inc., Natick, MA, USA), based on the sample MATLAB codes provided in Elhorst (2012).

# Results

#### Spatial and temporal dynamics of the SOS

Figure 3 shows the spatial pattern of the difference between the BIC of the AG function and the BIC of the DL function. The area where the BIC for the AG function is higher than that for the DL function accounts for 69.2 % of the study area, whereas the area where the BIC for the AG function is lower than that for the DL function accounts for 30.8 %. Overall, the DL function is better in reducing noise and fitting the complex shape of the NDVI curve than the AG function in our study area. The selection of the function to smooth the NDVI time series of a specific pixel depends on the performance to remove noise. Figure 4a demonstrates the spatial pattern of the mean SOS for the period 1982-2010. Generally, the SOS in the north occurs earlier than the SOS in the south, suggesting that the SOS is closely linked to the spatial pattern of MAP. The average SOS was calculated for all MAP intervals, and the segmented regression method was applied to fit these data (Fig. 5). A turning point was detected at a MAP of about 805 mm, and the 800-mm isohyet can be considered as an approximate division of grass-dominated and tree-dominated savannas (Campo-Bescós et al. 2013a, b). Below this



**Fig. 3** Spatial pattern of the difference between the Bayesian information criterion (BIC) for the asymmetric Gaussian (AG) and the double logistic (DL) functions

threshold (mainly grass-dominated savannas), there is a significant decreasing trend with a relatively large slope of -0.08 day/mm, while above this threshold (mainly treedominated savannas), there is a significant decreasing trend with a relatively shallower slope of -0.02 day/mm. Figure 4b shows the spatial pattern of changing trends in the SOS from 1982 to 2010. The trends were estimated using nonparametric Mann-Kendall trend tests (Yu et al. 1993; Yue et al. 2002). About 35.5 % of the study area shows a significant change in the SOS. A significantly advanced trend was observed over about 18.6 % of the study area which is chiefly distributed in the tree savanna areas of the north (Figs. 1b and 4b). A significantly delayed trend was observed over about 16.9 % of the study area which is mainly distributed in the floodplain in the north and bush/grass savannas of the south. Figure 6a shows the annual variations in the SOS of the study area as a whole from 1982 to 2010. Trends differ among the three time periods. From 1982 to 1990, there is a significant delay in the SOS with a rate of 2.6 days per year; from 1991 to 1995, there is no significant trend; and from 1996 to 2010, the SOS advances significantly at a rate of -1.2 days per year.

Table 2 Pearson's correlation coefficients between the SOS and environmental drivers using pooled spatial panel data

	meanT	maxT	minT	Rad	Р	PET	SoilW
Region (MAP<800 mm)	0.68*	-0.36*	0.22*	0.02	-0.18*	0.75*	-0.24*
Region (MAP>800 mm)	-0.59*	-0.54*	-0.20*	-0.24*	-0.32*	0.63*	0.04

The positive value suggests that a higher value of the specific driver causes a delay of the SOS, as the SOS is indicated by the day of the year. The negative value implies that a higher value of the specific driver leads to an advance of the SOS p<0.05

Fig. 4 Spatial and temporal patterns the start of the growing season (SOS): **a** the mean SOS and **b** changing trends in the SOS from 1982 to 2010. The positive  $\tau$  indicates a delayed trend of the SOS, whereas the negative  $\tau$  suggests an advanced trend of the SOS



Temporal changes of environmental drivers

Model selection

Figure 6 shows the annual variations in the potential preseason environmental drivers of the SOS from 1982 to 2010. There is a significant increasing trend in meanT at a rate of 0.05 °C per year (Fig. 6b). Figure 6c shows that a turning point in 1997 is identified underlying the annual variations of maxT. The factor, maxT, decreases at a rate of -0.26 °C per year from 1982 to 1997, and an increase at a rate of +0.30 °C per year from 1997 to 2010 is observed. There is a significant increasing trend in minT from 1982 to 2010 at a rate of 0.05 °C per year (Fig. 6d). Figure 6e illustrates the annual variations in radiation from 1982 to 2010. A turning point is detected in 1997, before which a significant decreasing trend is observed and after which there is a significant increasing trend. No significant trend has been witnessed underlying the time series of preseason precipitation (Fig. 6f). Figure 6g indicates that preseason PET increases significantly at a rate 0.45 mm per year from 1982 to 2010. No significant trend is observed for preseason soil moisture (Fig. 6h). Spatial panel data for all the above driving variables were included in the models.



**Fig. 5** The linear relationship between the start of the growing season (SOS) and mean annual precipitation (MAP). Piecewise linear regression was applied to fit the data

Spatial panel data models were run in the region with MAP above 800 mm (defined as tree-dominated savanna region) and the region with MAP below 800 mm (defined as grassdominated savanna region), respectively based on the breakpoint identified in Fig. 5. There were 22 spatial units for the grass-dominated savanna regions and 34 spatial units for the tree-dominated savanna regions. The Global Moran's I for each year of both regions was calculated based on both spatial units and their SOS values. The Global Moran's I can measure spatial autocorrelation to show whether the spatial pattern is clustered, dispersed, or random. Except for the years 1993, 1995, and 2001, the spatial patterns of the SOS for all years show clustered patterns in grass-dominated savannas. Except for the years 2001, 2002, and 2008, the spatial patterns of the SOS for all years also demonstrate clustered patterns. Therefore, it is necessary to include spatial autocorrelation into the regression models, and the spatial panel data models are good options. Table 3 shows the results of the stationarity test for the spatial panel data of response and explanatory variables. The results suggest that all the tests reject the null hypothesis, which implies the spatial panel data for all the variables are stationary.

The variance inflation factor (VIF) quantifies the severity of multicollinearity in a multiple linear regression analysis (Gujarati 2004). It is an index that measures how much the variance of an estimated regression coefficient is increased due to collinearity. A common rule of thumb is that if VIF is more than 10, then multicollinearity is high. To avoid multicollinearity, we ran multiple linear regression models using the pooled spatial panel data of both regions, respectively, and removed the variables with VIF values more than 10. This led to the exclusion of the meanT and PET from models due to their multicollinearity. Table 4 demonstrates the robust LM test results which determine whether the spatial lag model and/or the spatial error model are more appropriate. For Fig. 6 Annual variations in a the start of the growing season and potential environmental driving factors of the SOS: b preseason temperature, c preseason maximum temperature, d preseason minimum temperature, e preseason downward solar radiation, f preseason precipitation, g preseason potential evapotranspiration, and h preseason soil moisture



the grass-dominated savanna region, the hypothesis of no spatially lagged dependent variable or no spatially autocorrelated error term must be rejected on the presence of one of the above variables, irrespective of the inclusion of spatial and/or time-period fixed effects. Similar results were observed for the tree-dominated savanna region. For both regions, the LR tests justify the extension of the model with

both spatial and time-period fixed effects. Overall, the test results imply that the spatial and time-period fixed models, including spatial error autocorrelation and spatially lagged dependent variables, are more appropriate to describe the data.

Stationarity tests for the spatial panel data of response and Table 3 explanatory variables

I able I	The robust Eugrange manipher (Enr) test results which are to
determine	whether the spatially lagged dependent variable (lagged vari-
able) and/c	or spatially autocorrelated error term should be included in the
model	

The robust Lagrange multiplier (LM) test results which are to

	Region (MAP <800 mm)	Region (MAP >800 mm)
SOS	-23.6*	-32.3*
meanT	-25.8*	-42.6*
maxT	-23.4*	-26.0*
minT	-22.5*	-43.2*
Rad	-25.9*	-24.7*
Р	-29.1*	-42.9*
PET	-23.0*	-39.0*
soilW	-28.1*	-35.1*

\*p<0.05

	Region (MA	P <800 mm)	Region (MAP >800 mm)		
	Robust LM (lagged variable)	Robust LM (error term)	Robust LM (lagged variable)	Robust LM (error term)	
Pooled	30.9**	29.9 **	15.8**	33.6**	
Spatial fixed	58.0**	13.6**	17.4**	28.7**	
Time-period fixed	27.0**	6.0*	12.9**	4.4*	
Spatial and time-period fixed	15.1**	4.0*	12.3**	5.4*	

\*p<0.05; \*\*p<0.01

Table 4

1381

Relationships between environmental factors and the SOS

The environmental factors that significantly drive the SOS are different between the two regions (Table 5). In the grassdominated region (MAP <800 mm), minT has a significant and positive correlation with the DOY of the SOS, suggesting that lower minT results in an advance of the SOS. The factor, maxT, has a significant and negative correlation with the DOY of the SOS, implying that higher maxT tends to advance the SOS. Higher soilW relates to an earlier occurrence of the SOS in such water-limited regions as indicated by the negative correlation between soilW and DOY of the SOS. The spatial autoregressive coefficient indicates that the SOS of a spatial unit correlates positively with the SOS of surrounding spatial units. For the region of tree-dominated savannas, both minT and maxT correlate negatively with the DOY of the SOS, such that a decrease in maxT and minT results in an advance in the SOS. It should be noted that the variable, radiation, is also detected as a significant driving factor of the SOS. The spatial autoregressive coefficient implies that the SOS of a spatial unit correlates positively with that of surrounding spatial units.

## Discussion

The accuracy and reliability of techniques to estimate phenophases from remotely sensed time series, whose applications are usually hindered by noise chiefly due to atmospheric conditions and sun-sensor-surface viewing geometrics, is critically important to the global change research community (Atkinson et al. 2012). We evaluated the performance of two different techniques to remove noise from the NDVI time series and then used the technique that better modeled the complex shape of the NDVI curve for each pixel.

 Table 5
 The results of the spatial panel data models for the two regions

	Region (MAP <800 mm)		Region (MAP >800 mm)		
	Coefficient	<i>t</i> value	Coefficient	<i>t</i> value	
minT	0.14*	5.7	-0.38*	13.8	
maxT	-0.35*	-12.3	-0.30*	-14.2	
Rad	-0.01	-0.4	-0.09*	-0.4	
Р	0.02	0.6	0.01	0.4	
soilW	-0.11*	-3.5	-0.04	-1.2	
ρ	0.35*	3.3	0.17*	3.2	
λ	0.05	0.4	0.23*	3.8	

The positive value indicates an increase in the value of environmental variables leading to a delay of the SOS; the negative value indicates an increase in the value of environmental variables leading to an advance of the SOS

\*p<0.05

Our research used the NDVI ratio, which can better represent the state of an ecosystem, to detect the phenophase of spring phenology. Moreover, the increase in greenness is believed to be most rapid at a 50 % threshold (White et al. 1997, 2009; Beurs and Henebry 2010). While numerous methods exist to extract phenological metrics from the time-series satellite data, it is critical to choose the "right" model for the "right" place, as has been determined from multiple method comparisons or evaluations based on plot-level data (White et al. 2009; Liang et al. 2011; Cong et al. 2012). The findings of this study are useful for understanding the complexity of savanna systems. Savanna ecosystems are characterized by the co-dominance of two contrasting plant life forms-trees and grasses. Vegetation phenology of savannas influences both plant production and has profound impacts on several aspects of ecosystem function. Studies have proved that the factors that determine woody cover or NDVI of savanna systems were distinct with different tree cover, and a breakpoint related to MAP was usually detected (Sankaran et al. 2005; Campo-Bescós et al. 2013a, b; Zhu and Southworth 2013). The most conspicuous characteristic of the results is that the linear relationship between the SOS and MAP changes abruptly when MAP exceeds about 805 mm in our study area, as grass-dominated and tree-dominated savannas are largely distributed below and above such a threshold, respectively. Whether this range is applicable to other savanna ecosystems still needs further research. The spatial pattern of the SOS provides information on key aspects of vegetation seasonality. The SOS of treedominated savannas, found here at higher elevations, occurs earlier than that of grass-dominated savannas, which supports previous studies (De Bie et al. 1998; Chidumayo 2001). Therefore, there is a tremendous potential for characterizing, classifying, and mapping vegetation based on such phenological information (Wessels et al. 2011).

Our research determined the spatial and temporal variations of the SOS for the entire study area using a robust Mann-Kendall trend test. A significant advance of the SOS was mainly found in tree-dominated savannas of the north from 1982 to 2010, and a delay of the SOS was observed in the grass-dominated savannas in the south. The change in the SOS over time may affect vegetation primary production, but this is worth much more research as other phenophases and environmental factors are also subject to change (Richardson et al. 2010). The implications of these results are important not only for the vegetation but the ecosystem as a whole. The new leaves are the major source of food for many browsers at the end of dry season, a time when all herbivores are struggling to fulfill their metabolic requirements. Therefore, a delay or advance of the SOS can influence the spatio-temporal variability of forage availability, and the survival and reproductive success of many herbivores (Archibald and Scholes 2007).

Understanding the driving forces of land surface phenology is critical to assess the impact of future climate change on vegetation growth. Two regions divided by this threshold approximately correspond to tree- and grass-dominated savannas in our study area. Our study highlights the use of the spatial panel data models to figure out the differences in driving factors of the SOS between tree-dominated savannas (MAP >800 mm) and grass-dominated savannas (MAP <800 mm). The variables maxT and minT are identified as significant influencing factors of the SOS. Higher maxT tends to result in an earlier occurrence of the SOS in both regions, suggesting that higher daytime maximum temperatures can prompt more rapid vegetation green-up in spring. The opposite impacts of minT are observed in tree- and grass-dominated savannas. The reason might be that the higher temperatures indicated by minT are more likely to enhance water evaporation, causing a decrease in plant water availability in these already water-limited, grass-dominated savannas. Increased water demands resulting from higher temperatures can be met in tree-dominated savannas where water supplies are relatively more abundant. Our research identified preseason soil moisture as a significant influencing factor in grass-dominated savannas. The importance of water has also been proven by previous studies (Zhang et al. 2005; Jolly et al. 2005), and soil water exerts a more direct influence on plant water availability. In tree-dominated savannas, spring radiation makes a partial contribution to the change in the SOS, as radiation can affect vegetation photosynthetic activity and further vegetation primary production (Nemani et al. 2003). Beyond the dynamic environmental factors included in the models, Brown et al. (2010) found that the SOS correlates significantly with the large-scale oscillations indicated by different measures such as the Multivariate ENSO Index (MEI) and the Indian Ocean Dipole (IOD). However, we did not consider these potential factors, as they are closely correlated to precipitation fluctuations which we have included in our models (Gaughan and Waylen 2012). The model tests justify the inclusion of the spatial and temporal fixed effects, which indicates that spatial and temporal heterogeneity does affect the robustness of statistical models. The models in our study accounted for spatial autocorrelation terms, but did not consider serial autocorrelation of the dependent variables as no significant serial autocorrelation was detected underlying the SOS time series (Lee and Yu 2010).

In conclusion, based on remotely sensed data and environmental data from multiple sources, we derived the spatial and temporal dynamics of the SOS and then disentangled its relationship with environmental factors across the physiographic gradients in southern African savanna systems from 1982 to 2010. The research identified the spatial panel data model with spatial and temporal fixed effects and including spatial error autocorrelation and spatially lagged dependent variables to be more appropriate to describe the data. The SOS occurs earlier in the north in areas of higher MAP and dense tree cover (mainly tree savannas), and later in the south, in areas of lower MAP and sparse tree cover (mainly grass or bush savannas). From 1982 to 2010, an advancing trend was observed in the tree savanna areas of the north, whereas a delayed trend was found in the floodplain of the north and bush/grass savannas of the south. The study permitted analysis of shared effects of potentially important environmental variables on the SOS within tree-dominated and grass-dominated savannas, which were divided by the isohyet of 800 mm. In contrast, most previous studies of phenology in southern Africa have focused on individual relationships between phenology and one or two variables, highlighting some of the trends relating to individual factors. Different environmental drivers were detected within two different savanna types. Our results supported the importance of the water-related factor, preseason soil moisture, in affecting the SOS significantly only in water-limited and grass-dominated savannas. Also, our research pointed to more often overlooked environmental drivers, specifically preseason maximum temperature and minimum temperature, which were identified as significant, influencing factors in both regions. Further research is needed to understand underlying mechanisms and to validate the results in savanna landscapes. In terms of climate variability and future climate change across this region, both vegetation type and environmental factors are essential to understand different savanna landscapes and the impacts of likely future climate change. This study can better evaluate and understand landscape level changes in land surface phenology. Furthermore, the examinations of changing mechanisms of spring phenology are useful to managers and researchers and key to predicting how projected climate changes will affect regional vegetation phenology and productivity.

Acknowledgments The authors wish to thank Range Myneni, Jorge Pinzón, and Zaicun Zhu for the provision of the GIMMS3g NDVI data. This work was supported by the NASA Land Cover/Land Use Change Program under Grant No. NNX09AI25G, titled "Understanding and predicting the impact of climate variability and climate change on land use/land cover change via socioeconomic institutions in Southern Africa" (PI: Jane Southworth, University of Florida).

#### References

- Archibald S, Scholes RJ (2007) Leaf green-up in a semi-arid African savanna: separating tree and grass responses to environmental cues. 18:583–594
- Atkinson PM, Jeganathan C, Dash J, Atzberger C (2012) Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote Sens Environ 123:400–417
- Bi J, Xu L, Samanta A et al (2013) Divergent Arctic-boreal vegetation changes between North America and Eurasia over the Past 30 Years. Remote Sens 5:2093–2112
- Brown ME, de Beurs K, Vrieling A (2010) The response of African land surface phenology to large scale climate oscillations. Remote Sens Environ 114:2286–2296

- Buse A (1982) The likelihood ratio, Wald, and Lagrange multiplier tests: an expository note. Am Stat 36:153–157
- Campo-Bescós MA, Muñoz-Carpena R, Kaplan DA et al (2013a) Beyond precipitation: physiographic gradients dictate the relative importance of environmental drivers on Savanna vegetation. PLoS One 8:e72348
- Campo-Bescós MA, Muñoz-Carpena R, Southworth J et al (2013b) Combined spatial and temporal effects of environmental Controls on long-term monthly NDVI in the Southern Africa Savanna. Remote Sens 5:6513–6538
- Chase MJ, Griffin CR (2009) Elephants caught in the middle: impacts of war, fences and people on elephant distribution and abundance in the Caprivi Strip, Namibia. Afr J Ecol 47:223–233
- Chidumayo EN (2001) Climate and phenology of savanna vegetation in southern Africa. J Veg Sci 12:347–354
- Cong N, Piao S, Chen A et al (2012) Spring vegetation green-up date in China inferred from SPOT NDVI data: a multiple model analysis. Agric For Meteorol 165:104–113
- Cong N, Wang T, Nan H et al (2013) Changes in satellite-derived spring vegetation green-up date and its linkage to climate in China from 1982 to 2010: a multimethod analysis. Glob Chang Biol 19:881–891
- Dai J, Wang H, Ge Q (2013) The spatial pattern of leaf phenology and its response to climate change in China. Int J Biometeorol. doi:10.1007/s00484-013-0679-2
- De Beurs KM, Henebry GM (2010) Spatio-temporal statistical methods for modelling land surface phenology. In: Hudson IL, Keatley MR (eds) Phenol. Res. Springer, Netherlands, pp 177–208
- De Bie S, Ketner P, Paasse M, Geerling C (1998) Woody plant phenology in the West Africa savanna. 25:883–900
- De Jong R, Verbesselt J, Zeileis A, Schaepman M (2013) Shifts in global vegetation activity trends. Remote Sens 5:1117–1133
- Eklundh L, Jönsson P (2012) TIMESAT 3.1 Software Manual 1-82
- Elhorst JP (2009) Spatial panel data models. In Fischer MM, Getis A (Eds.) Handbook of applied spatial analysis, 377–407
- Elhorst JP (2012) Matlab Software for spatial panels. Int Reg Sci Rev 1-22
- Enquist CAF, Kellermann JL, Gerst KL, Miller-Rushing AJ (2014) Phenology research for natural resource management in the United States. Int J Biometeorol. doi:10.1007/s00484-013-0772-6
- Fan Y, van den Dool H (2004) Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. J Geophys Res 109:D10102
- Fensholt R, Rasmussen K, Kaspersen P et al (2013) Assessing land degradation/recovery in the African Sahel from long-term earth observation based primary productivity and precipitation relationships. Remote Sens 5:664–686
- Gaughan AE, Waylen PR (2012) Spatial and temporal precipitation variability in the Okavango–Kwando–Zambezi catchment, southern Africa. J Arid Environ 82:19–30
- Gujarati DN (2004) Basic econometrics (4th eds). McGraw-Hill Education, New York
- IPCC (2013) In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauel A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- Jin C, Xiao X, Merbold L et al (2013) Phenology and gross primary production of two dominant savanna woodland ecosystems in Southern Africa. Remote Sens Environ 135:189–201
- Jolly WM, Nemani R, Running SW (2005) A generalized, bioclimatic index to predict foliar phenology in response to climate. Glob Chang Biol 11:619–632
- Jönsson P, Eklundh L (2002) Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans Geosci Remote Sens 40:1824–1832

- Jönsson P, Eklundh L (2004) TIMESAT—a program for analyzing timeseries of satellite sensor data. Comput Geosci 30:833–845
- Lee L, Yu J (2010) Some recent developments in spatial panel data models. Reg Sci Urban Econ 40:255–271
- Legates DR, Willmott CJ (1990) Mean seasonal and spatial variability in gauge-corrected, global precipitation. Int J Climatol 10:111–127
- Li S, Xie Y, Brown DG et al (2013) Spatial variability of the adaptation of grassland vegetation to climatic change in Inner Mongolia of China. Appl Geogr 43:1–12
- Liang L, Schwartz MD, Fei S (2011) Validating satellite phenology through intensive ground observation and landscape scaling in a mixed seasonal forest. Remote Sens Environ 115:143–157
- Menzel A, Fabian P (1999) Growing season extended in Europe. Nature 397:659
- Menzel A, Sparks TH, Estrella N et al (2006) European phenological response to climate change matches the warming pattern. Glob Chang Biol 12:1969–1976
- Moon HR, Perron B (2004) Testing for a unit root in panels with dynamic factors. J Econom 122:81–126
- Myneni RB, Keeling CD, Tucker CJ et al (1997) Increasing plant growth in the northern high latitude from 1981 to 1991. Nature 386:698–702
- Nemani RR, Keeling CD, Hashimoto H et al (2003) Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science 300:1560–1563
- Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. Nature 421:37–42
- Richardson AD, Black TA, Ciais P et al (2010) Influence of spring and autumn phenological transitions on forest ecosystem productivity. Philos Trans R Soc Lond B Biol Sci 365:3227–3246
- Sankaran M, Hanan NP, Scholes RJ et al (2005) Determinants of woody cover in African savannas. Nature 438:846–849
- Tucker C, Pinzon J, Brown M et al (2005) An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int J Remote Sens 26:4485–4498
- Vuong QH (1989) Likelihood ratio tests for model selection and nonnested hypothesis. Econometrica 57:307–333
- Wang J, Brown DG, Chen J (2013) Drivers of the dynamics in net primary productivity across ecological zones on the Mongolian Plateau. Landsc Ecol 28:725–739
- Wessels K, Steenkamp K, von Maltitz G, Archibald S (2011) Remotely sensed vegetation phenology for describing and predicting the biomes of South Africa. Appl Veg Sci 14:49–66
- White A, Thomton PE (1997) A continental responses phenology model climatic for monitoring variability vegetation to interannual. 11: 217–234
- White M, Thomton P, Running S (1997) A continental phenology model for monitoring vegetation responses to interannual climatic variability. Glob Bogeochem Cycle 11:217–234
- White MA, de BEURS KM, Didan K et al (2009) Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. Glob Chang Biol 15:2335–2359
- Yu YS, Zou SM, Whittemore D (1993) Nonparametric trend analysis of water-quality data of rivers in Kansas. J Hydrol 150:61–80
- Yue S, Pilon P, Cavadias G (2002) Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. J Hydrol 259:254–271
- Zhang X, Friedl MA, Schaaf CB, Strahler AH (2005) Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments. J Geophys Res 110: D12103. doi:10.1029/2004JD005263
- Zhu L, Southworth J (2013) Disentangling the relationships between net primary production and precipitation in Southern Africa savannas using satellite observations from 1982 to 2010. Remote Sens 5: 3803–3825