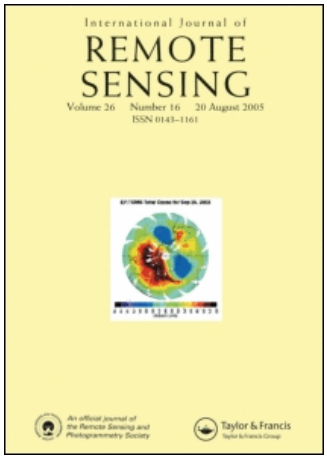

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722504

A statistical framework for the analysis of long image time series

K. M. de Beurs [a]; G. M. Henebry [a]
[a] Center for Advanced Land Management Information Technologies (CALMIT), School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, 68588-0517 USA

Online Publication Date: 01 April 2005
To cite this Article: de Beurs, K. M. and Henebry, G. M. (2005) 'A statistical framework for the analysis of long image time series', International Journal of Remote Sensing, 26:8, 1551 - 1573
To link to this article: DOI: 10.1080/01431160512331326657
URL: http://dx.doi.org/10.1080/01431160512331326657

International Journal of Remote Sensing
Vol. 26, No. 8, 20 April 2005, 1551–1573

A statistical framework for the analysis of long image time series

K. M. DE BEURS and G. M. HENEBRY*
Center for Advanced Land Management Information Technologies (CALMIT), School of Natural Resources, University of Nebraska-Lincoln, 102 East Nebraska Hall, Lincoln, NE, 68588-0517 USA

(*Received 22 January 2004; in final form 4 October 2004*)

Coarse spatial resolution satellites are capable of observing large swaths of the planetary surface in each overpass resulting in image time series with high temporal resolution. Many change-detection strategies commonly used in remote sensing studies were developed in an era of image scarcity and thus focus on comparing just a few scenes. However, change analysis methods applicable to images with sparse temporal sampling are not necessarily efficient and effective when applied to long image time series. We present a statistical framework that gathers together: (1) robust methods for multiple comparisons; (2) seasonally corrected Mann–Kendall trend tests; (3) a testing sequence for quadratic models of land surface phenology. This framework can be applied to long image time series to partition sources of variation and to assess the significance of detected changes. Using a standard image time series, the Pathfinder AVHRR Land (PAL) NDVI data, we apply the framework to address the question of whether the institutional changes accompanying the collapse of the Soviet Union resulted in significant changes in land surface phenologies across the ecoregions of Kazakhstan.

1. Introduction

The response of the Earth system to natural variations and anthropogenic forcings on the land surface has become an increasingly important question in recent years due to progress in the numerical modelling of weather and climate. Several Earth observation programmes have been launched as means to address significant knowledge gaps about the dynamics of the Earth system. Today, a multitude of space-borne sensors generate streams of imagery at various spatial, temporal and spectral scales. The image time series record patterns generated by land surface processes that have been affected by various biogeophysical processes and disturbance events, both anthropogenic and otherwise, in addition to the typical range of variability in meteorological forcings (Potter *et al.* 2003). To distinguish change from variation, it is necessary to partition the variability observed in an image time series into contributions attributable to sensor artifacts, inter-annual climate variation and anthropogenic disturbances such as changes in policy or institutions (de Beurs and Henebry 2004a).

Coarser spatial resolution sensors (e.g. the Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectrometer (MERIS)) are able to observe

*Corresponding author. Email: ghenebry@calmit.unl.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160512331326657

large swaths of the planetary surface in each overpass, resulting in data records with high temporal resolution. Many change-detection strategies commonly used in remote sensing studies were developed in an era of image scarcity with relatively high spatial resolutions and thus focus on comparing just a few scenes (Jensen 1996). Change analysis techniques applicable to shorter image time series are not necessarily efficient or effective when applied to longer image time series. For example, a shorter image time series may consist of a chronosequence selected at comparable phenological periods over the span of a few years. Disturbance events may be poorly localized in time, if the gap between sequential images is too long or the number of scenes is too few. Thus, discontinuities resulting from disturbance events may become indistinguishable from trends, when there is a mismatch between the pace of change and the frequency of image acquisitions. The risk of confounding variability with change is high with shorter image time series.

In contrast, the long image time series can offer sufficient temporal sampling in terms of duration and frequency that allow detection of changes amidst substantial variation. Here, we present a statistical framework within which to discriminate between discontinuities and trends in a long univariate image time series.

Frequently, changes in univariate image time series are summarized by regression trend analysis (Fuller 1998, Kogan and Zhu 2001, Lee *et al.* 2002, Slayback *et al.* 2003). Typically, yearly integrated data are regressed against time and trend is expressed as the slope of the regression curve. Other common methods are autoregressive-moving average (ARMA) linear models (Piwowar and LeDrew 2002), Principal Component Analysis (PCA) (Eastman and Fulk 1993, Shabanov *et al.* 2002) and Fourier analyses (Jakubauskas *et al.* 2001, Moody and Johnson 2001). However, stationarity assumptions, data quality, sensor noise and the complicatedness of the methods can make it a challenge to quantify the separate sources of information that influence the signal and to determine what constitutes a significant change (Nightingale and Phinn 2003). Kaufmann *et al.* (2000) analysed from theoretical and empirical perspective the Pathfinder AVHRR Land (PAL) Normalized Difference Vegetation Index (NDVI) dataset to determine whether the effects of orbital drift and sensor changes had sufficiently contaminated the PAL NDVI data to render it useless for change analysis. While these analyses are powerful, they are not readily adaptable to explore other sources of change and variation, even within the same dataset. In summary, although a broad range of methods have been developed to detect and describe changes in long image time series, there is a dearth of general techniques to assess the statistical significance of these identified changes.

We compiled a framework consisting of a suite of relatively straightforward standard statistical analyses that can be applied to partition the sources of variation in a long image time series and to assess the significances of detected changes. Two questions motivate the framework:

1. Are there discontinuities in the image time series?
2. Are there trends in the image time series?

The second question distinguishes between (a) increasing and/or decreasing trends that are independent of observed seasonality within the image time series and (b) trends in phenology. These questions are addressed in § 5 and § 6 of this paper, respectively. Phenological trends are here defined as shifts in phenological patterns, such as changes in the onset and timing of events, as summarized by

changes to model parameter coefficients rather than to the model structure. To determine these changes, phenological models are fit for two observational periods and their parameter estimates are compared to assess significant differences.

The statistical framework we present here can be applied to a variety of long image time series – it is not restricted to NDVI data. However, to demonstrate its power, we apply the framework to the PAL NDVI data recorded for Kazakhstan. We show that it is possible to distinguish between the several sources of variation embedded in the PAL dataset and to attribute some of the variation to the institutional changes following the independence of Kazakhstan in 1991.

2. Image time series

Myriad space-borne sensors with high temporal resolutions (1–3 days) and relatively coarse spatial resolutions (1–8 km), such as AVHRR, Along Track Scanning Radiometer (ATSR) on ERS-1&2, Systeme Pour l'Observation de la Terre (SPOT) Vegetation, MERIS on Envisat, and MODIS on Terra and Aqua, have been recording comparable data over the planetary surface for several years. However, most of these sensors have not been in operation long enough to record a sufficient number of annual image times series to make long-term statistical change analysis feasible. The AVHRR PAL image dataset provides one of the longer image time series collected over the past two decades. The PAL NDVI dataset consists of 8×8 km global 10-day composite images for the period of July 1981 to October 2001. Thus, the PAL dataset provides a model image time series with which to demonstrate a method of change analysis. Vegetation indices such as the NDVI have been widely used in studies of land surface condition, including phenological studies (Myneni *et al.* 1997, Chen *et al.* 2001, Tucker *et al.* 2001, Shabanov *et al.* 2002), crop development identification (Jakubauskas *et al.* 2001), weather impacts on crop growth (Chen *et al.* 2001, Wang *et al.* 2001, Labus *et al.* 2002, Peters *et al.* 2002) and land cover fraction identification based on land surface phenology (Moody and Johnson 2001, Lu *et al.* 2003).

Sensor systems are subject to wear and thus are only expected to be active for a limited period. Aging satellites are generally replaced after a few years by newer sensor generations and data producers try to maintain the accumulating dataset as consistently as possible through calibration techniques. For example, the PAL NDVI data from 1982 to 1999 were recorded by four different AVHRR sensors on a sequence of National Oceanic and Atmospheric Administration (NOAA) satellites (NOAA-7, NOAA-9, NOAA-11 and NOAA-14). When analysing satellite time series collected by more than one sensor, it is important to ensure that detected differences and inferred trends reflect land or ocean surface trends, not simply discontinuities among sensors.

3. Methods

We propose a sequence of three distinct methods to partition the sources of variation that are captured by the PAL NDVI image time series. The first phase tests for discontinuities, which we call step changes, in the time series. Specifically, the suggested tests can evaluate whether average values from multiple periods are significantly different (figure 1, A). The second phase tests for trends within periods (figure 1, B). We suggest a powerful and robust alternative to the common practice

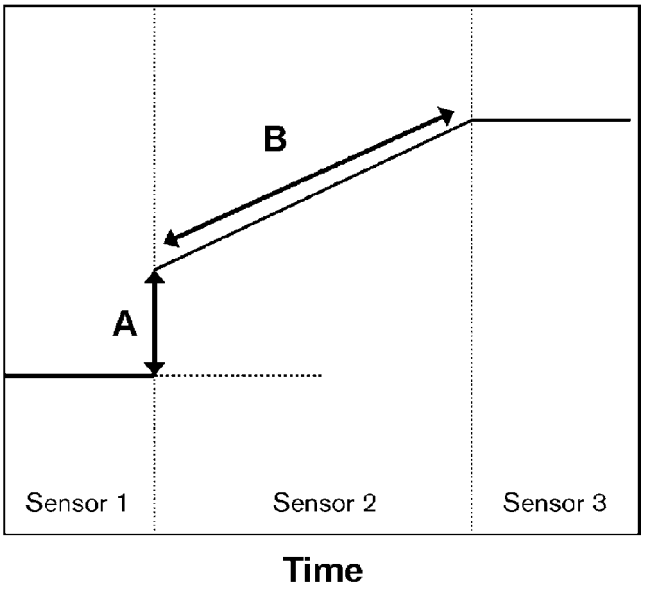

Figure 1. Discontinuities and trends in responses among generic sensors. There is a step change (A) between observing periods of sensors 1 and 2. Sensor 2 exhibits a linear trend during the observation period (B).

of trend identification via simple linear regression. The third phase tests for changes in land surface phenology by modelling NDVI as a quadratic function of accumulated growing degree–days (AGDDs). By accounting for seasonal (intra-annual) variation, it is possible to test for significant differences in the onset and tempo of phenological development.

There are two types of errors that can occur when testing hypotheses. Rejecting the null hypothesis when in fact it is true is called a type 1 error (α). The error made when the test does not reject the null hypothesis when in fact it is false is called a type 2 error (β). The power of a test, $1-\beta$, signifies the probability of correctly rejecting the null hypothesis in favour of the alternative hypothesis (Zar 1984, Conover 1999).

4. Are there discontinuities (step changes) in the image time series?

Step changes in image time series may be visualized as different averages of the time series for two or more periods. Discontinuities can occur due to differences in sensors and/or can be caused by sudden events in the land surface environment. Independent of the cause for the step changes, it is to be expected that the variances change with the average values. Furthermore, the data in an image time series often do not follow a normal distribution and they are not always divisible into periods of equal duration; thus, statistical methods employed must be capable of handling non-normal data and unequal numbers of observations.

The following three general assumptions influence both the type 1 and type 2 error rates of standard parametric tests (Dunnet 1980, Day and Quinn 1989, Keselman *et al.* 1998).

1. *Equal population variances*. The variance of each period should be more or less equal. The greater the variance differences between periods, the greater the type 1 error.
2. *Normality*. The data should follow a normal distribution. If the data do not follow a normal distribution an appropriate data transformation might be applied. An alternative solution is to apply a non-parametric test. However, since parametric tests generally are more powerful than non-parametric tests, data transformation is preferable to using a non-parametric test simply to avoid testing for a normal distribution.
3. *Equal sample size*. Periods should have an equal number of observations, so the periods should be of equal duration. Violation of this assumption is only damaging in the case of unequal population variances and/or non-normality of the data.

To avoid distorted results when testing for significant differences between multiple periods of an image time series, it is necessary to select methods that are robust against all three violated assumptions. If the data follow a normal distribution, we propose use of the C-method, a parametric method that is robust for unequal variances and unequal sample sizes. When the data are not normally distributed and a data transformation fails to yield the desired results, we suggest the Fligner–Policello test.

4.1 *The C-method*

The C-method is an extension of Cochran's test under unequal variances and is considered the best way to describe comparison problems with large sample sizes (Dunnet 1980, Stoline 1981, Day and Quinn 1989). The critical value for the test should be calculated for each pair of periods separately. If the difference between two means is larger than the critical value, the null hypothesis is rejected. The critical value of the C-method is given by:

$$\text{Critical value} = \left[Q_{a(m,df)} s_i^2 / n_i + Q_{a(m,df)} s_j^2 / n_j \right] \Big/ \sqrt{2\left(s_i^2 / n_i + s_j^2 / n_j \right)} \qquad (1)$$

The test incorporates the variances (s_i^2 and s_j^2) and the sample sizes (n_i and n_j) of both periods (i, j). The critical value is based on the 'Studentized range', $Q_{\alpha(m,df)}$, and is dependent upon the α-level, the total number of means that are being tested (m) and the error degrees of freedom (df) of the analysis. $Q_{\alpha(m,df)}$ follows the q-distribution and tables can be found in most standard statistical software packages and texts (e.g. Zar 1984).

4.2 *The Fligner–Policello test*

Non-normal data should be submitted to the Fligner–Policello (FP) test. The FP test is a non-parametric test for unplanned comparisons and unequal sample sizes. It is one of the few tests that permits unequal variances. The test statistic of the FP test is given by:

$$\text{Test statistic} = \left(\sum P_i - \sum P_j \right) \Big/ \left[2\sqrt{\sum \left(P_i - \overline{P}_i \right)^2 + \sum \left(P_j - \overline{P}_j \right)^2 + \overline{P}_i \overline{P}_j} \right] \qquad (2)$$

The test statistic should be calculated with period i longer than j. This test statistic is based completely on the 'placement' (P) of observations within the two periods, i.e. the rank of that observation for the combined sample of both periods. With large sample sizes, as is usually the case when we test differences of long image time series, the critical values are based on the standard normal distribution. To maintain an exact α-level, the test should be corrected for the total number of comparisons performed with the Dunn–Sidák method (Fligner and Policello 1981, Day and Quinn 1989), where we set the significance levels as:

$$b = 1 - (1 - a)^{1/m} \tag{3}$$

where m is the number of comparisons. With $\alpha = 0.05$ and six comparisons, $b = 0.00851$, so that $(1–0.00851)^6 = 0.95$. With an adjusted significance level of $b = 0.00851$ we manage to keep the α-level of all six tests simultaneously at 0.05.

5. Are there linear trends in the image time series?

Many authors report trends in image time series as the slope parameter resulting from regression analysis (Fuller 1998, Chen *et al*. 2001, Kogan and Zhu 2001, Tucker *et al*. 2001). Linear regression always results in parameter estimates but these parameters are not always significantly different from zero. Many studies that report the slope parameter fail to report the associated error and the overall error of the linear regression model. Furthermore, there are four assumptions that are generally violated when trend lines are estimated from long image time series using regression:

1. All Y values should be independent of each other.
2. Residuals should be random, follow a normal distribution, and be independent of the explanatory (independent) variable.
3. The mean of the residual distribution should be zero.
4. The variance of the residuals should be an equal constant for all values of X.

Parameter estimates generated by linear regression generally remain unbiased and linear even if these assumptions are violated; however, the calculated significance of the estimated parameters becomes unreliable. For long-term image time series this translates into an increased probability of falsely rejecting the null hypothesis of no trend.

Statistical trend tests provide a more appropriate way to describe trends in long time series (Dietz and Killeen 1981, Hirsch and Slack 1984). However, most image time series are dependent on processes that are temporarily correlated, causing the failure of standard trend tests (von Storch and Navarra 1999). Serial dependence is often referred to as temporal autocorrelation (ρ_k), which is the correlation between residuals of observations of the same variable at different points in time. To test for significant trends in satellite time series, it is crucial to use a test that is corrected for serial dependency. The seasonal Mann–Kendall trend test is completely rank-based and therefore robust against non-normality, missing values, seasonality and, if corrected, serial dependence as well.

Hirsch and Slack (1984) provide a test with correction for serial dependence in seasonal data. The time series is first decomposed into a series of non-overlapping

subsets of equal length (here, 10-day periods) and an X-matrix is generated in the following form:

$$X = \begin{bmatrix} X_{11} & X_{12} & \dots & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & \dots & X_{2p} \\ \dots & \dots & & & \dots \\ \dots & \dots & & & \dots \\ X_{n1} & X_{n2} & & & X_{np} \end{bmatrix} \tag{4}$$

where p is the number of observations within one year, or the number of subsets, and n is the number of years.

The Mann–Kendall test statistic for a particular subset g is based on the observation ranks of that same subset across all years in the time series. The test statistic is calculated by summing the number of times a particular year has a higher value than any previous year. With

$$\begin{aligned} \operatorname{sgn}(x) &= +1 \quad x>0 \\ \operatorname{sgn}(x) &= 0 \quad\;\; x=0 \\ \operatorname{sgn}(x) &= -1 \quad x<0 \end{aligned} \tag{5}$$

the Mann–Kendall test statistic for each subset g has the form

$$S_g = \sum_{i<j} \operatorname{sgn}\left(X_{jg} - X_{ig}\right) \quad g = 1,2,\dots,p \tag{6}$$

where i and j are respectively the ith and jth year in the time series ($i=1 \dots n-1$; $j=2 \dots n$). For example, to calculate the Mann–Kendall test statistic for a particular 10-day period based on five years of observations ($n=5$), we determine the number of times that the observations of all the years are larger than any previous year. A positive trend with all observations increasing would yield a maximal value for the Mann–Kendall test statistic ($S_g=10$ for $n=5$).

The seasonal Kendall test statistic for the complete time series is defined as the sum of all Mann–Kendall statistics from each subset:

$$S' = \sum_{g=1}^{p} S_g \tag{7}$$

where S' is asymptotically normal with a mean of 0 and the variance is defined as the sum of the variances of every subset plus the sum of the covariances of every combination of subsets:

$$\operatorname{var}[S'] = \sum_{g} \sigma_g^2 + \sum_{\substack{g,h \\ g \neq h}} \sigma_{gh} \tag{8}$$

where g and h are defined as subsets in the time series. If the data from all subsets are mutually independent, the covariance is zero. However, in the case of dependency between the subsets, the covariance between the subsets is defined (Dietz and Killeen 1981) as

$$\hat{\sigma}_{gh} = \frac{K_{gh}}{3} + \frac{\left(n^3 - n\right) r_{gh}}{9} \tag{9}$$

where

$$K_{gh} = \sum_{i<j} \operatorname{sgn}\left[\left(X_{jg} - X_{ig}\right)\left(X_{jh} - X_{ih}\right)\right] \tag{10}$$

n is the number of years in the time series, r_{gh} is Spearman's correlation coefficient for subsets g and h, and i and j represent the years. Larger sample sizes generate more conservative tests but only five years' data with 12 observation per year (60 observations, $n=5$, $p=12$) can generate reliable results (Hirsch and Slack 1984).

6. How does the phenology of a landscape change? Are there changes in seasonality?

Many authors have found relationships between NDVI, temperature and precipitation (Di *et al.* 1994, Richard and Poccard 1998, Potter and Brooks 2001, Wang *et al.* 2001, Foody 2003, Ji and Peters 2003). Several studies have found significant relations between climate variables and NDVI globally (Myneni *et al.* 1997, Kawabata *et al.* 2001, Potter and Brooks 2001, Potter *et al.* 2003), while others have monitored the relationship between climate variables and NDVI in specific regions, such as Australia (Hill and Donald 2003, Nightingale and Phinn 2003), China (Chen and Pan 2002, Li *et al.* 2002), and Central Asia (Lee *et al.* 2002, Suzuki *et al.* 2003). In every study the authors found significant relationships of both temperature and precipitation with NDVI.

We accept the relation between NDVI and climate as amply demonstrated; what we are interested in here are land surface phenology changes through time. We make the distinction here between a change in phenology and a phenological change. The former relates to change in the kind of model used to describe the phenological pattern, such as a change from woodlands to croplands. The latter refers to shifts in the parameter coefficients of the model used to describe the phenological pattern. These shifts can result in significant differences in the onset and timing of phenological events, but the general shape of the phenological pattern remains the same.

Plant phenology models relate thermal regimes of the growing season with events in plant development (Schwartz 2003). The thermal regime of the growing season can be measured as accumulated growing degree-days (AGDDs) by summing growing degree–days from some consistent start date until a specific subsequent date. For most wheat varieties a base temperature of 0°C is chosen (Rickman and Klepper 1991). Assuming a standard Northern Hemisphere start date of 1 January and an effective temperature threshold of 0°C, AGDDs are calculated as follows:

$$\begin{aligned} \mathrm{GDD} &= \frac{(T_{\max} + T_{\min})}{2} \\ \mathrm{AGDD} &= \sum_{01\,\mathrm{Jan}}^{31\,\mathrm{Dec}} \mathrm{GDD}, \quad \text{if } \mathrm{GDD} > 0 \end{aligned} \tag{11}$$

Thermal-based regression models using AGDDs as the explanatory variable have been regularly used in crop phenology studies to describe and predict the green-up, flowering, fruiting and senescence stages of crops and grasslands and to compare multiple crop varieties (Goodin and Henebry 1997, Mitchell *et al.* 2001, Smart *et al.* 2001, Davidson and Csillag 2003). Here, we propose to use this method to analyse changes in long image time series using AGDDs instead of calendar dates. This allows us to align the imagery using a temporal metric that is relevant to land

surface phenology instead of an arbitrary anthropocentric calendar. Although we demonstrate how to compare quadratic models, the method of comparison is generic and can be easily applied to higher order or more complicated models and it is not restricted to the regression of NDVI with AGDDs.

We summarize the development of NDVI as a function of AGDD using a quadratic model:

$$\mathrm{NDVI} = \alpha + \beta \mathrm{AGDD} + \gamma \mathrm{AGDD}^2 \tag{12}$$

where NDVI is a vector with all NDVI values for a certain period and AGDD is a vector with all AGDD values for the same period. The intercept parameter α measures the greenness expressed as NDVI at low AGDD. An increased intercept points to an increase in greenness at the beginning of the observational season. (It is important to distinguish between the beginning of the growing season and the beginning of the observational season as they may not coincide, as is the situation in the case study presented below.) The slope parameter β measures the sum of growing degree–days necessary to reach the seasonal peak of NDVI. The quadratic parameter γ determines the shape of the model with smaller γ values producing broader curves and a longer season.

Quadratic phenology models (equation (12)) regressing NDVI with AGDD have been demonstrated to provide a parsimonious model to describe seasonal NDVI variability (Goodin and Henebry 1997, de Beurs and Henebry 2004a). Furthermore, each model parameter has a ready ecological interpretation. Phenology regression models can only reveal changes in phenological pattern between periods if the model explains a significant proportion of NDVI variation. The fraction of all variance in NDVI that is explained by multiple regression models can be expressed by the coefficient of determination adjusted for model complexity (R^2_{adj}). Once a model with a good statistical fit has been identified, values of the parameter coefficients can give insight into the processes that drive land surface phenology.

We propose to fit the given model (equation (12)) for each period of interest so that the parameter estimates from the models enable detection of change in the pattern of land surface phenology. The testing sequence to compare parameter coefficients begins with a standard F-test for equality of the highest order parameters of two periods. The test procedure ends when a significant difference between the parameter coefficient estimates for two periods is found. If a pair of parameter coefficient estimates is found to be not significantly different, the two parameter coefficients are weighted by the sum of squares of the observations from both periods and this new single coefficient is used to re-estimate the lower order parameter coefficients in each period. The test procedure is followed until the lowest order parameter coefficient, typically the intercept, is tested. If no significant change is found for any parameter coefficient, we can conclude that the estimated phenology models for both periods are statistically equivalent. This method can be applied to higher order polynomials; however, we found that second-order polynomials explain the variation sufficiently.

The coefficients of two quadratic models can differ in 15 distinct ways from one period to the next. We divide these 15 possibilities into four overall change behaviours (Type I, II, III and IV in figure 2). First, we have the possibility that there is no phenological change from one period to the next (Type I). In the second case (Type II), we only find a change in the intercept coefficient between the two

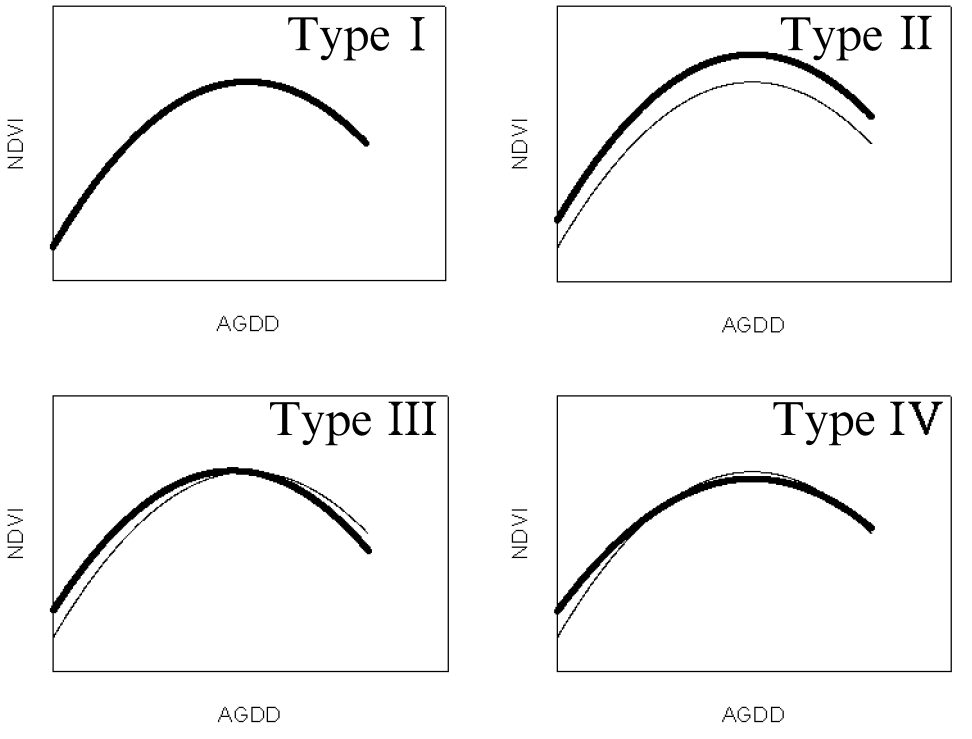

Figure 2. Four possible phenological change behaviours: Type I, no change; Type II, overall increase in NDVI; Type III, increased NDVI at lower AGDD and earlier peak; Type IV, increased NDVI at lower AGDD, earlier peak and longer period of elevated NDVI.

periods. The intercept can increase or decrease between periods, resulting in two change options within Type II. Type III supports differing intercepts and linear coefficients with a constant quadratic coefficient. Since two coefficients can increase or decrease, there are four options within Type III. The last change behaviour (Type IV) presents the possibility that all coefficients change from one model to the next. Three variable coefficients yield eight change options within Type IV.

7. Residual analyses

Residuals of regression models are good indicators of model fit. Well-fit models produce normally distributed random errors with zero mean that exhibit no patterns of relationship with the independent variables. Well-fit models produce no significant differences in average residuals and no trend in any period.

Figure 3 gives the complete scheme of our analysis framework, which we divided into a sequence of three sections. The first section (*a*) demonstrates the statistical method used to analyse significant differences in averages from multiple time series. There are n groups of input data for which normality is first evaluated. If all input data are normally distributed, the C-method is applied. If the data are not normally distributed, the FP test is applied. If there is a significant difference between two groups, we call this a step change. The second section (*b*) demonstrates the application of the trend test. The p-values give an indication of the significance of detected trends. The last section (*c*) demonstrates the testing sequence to determine phenological differences between two time series. First, a phenological model is fit

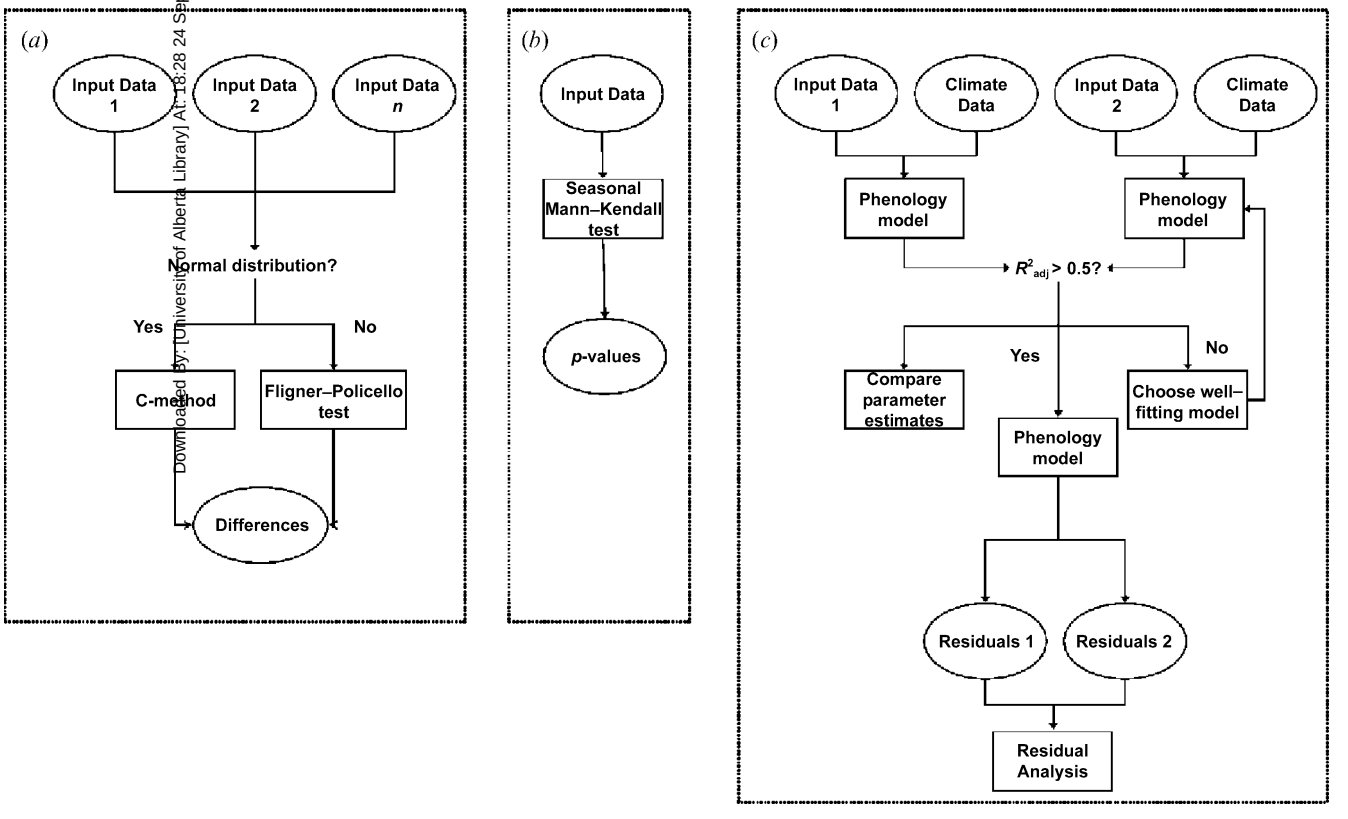

Figure 3. Schematic diagram of the complete analysis to partition sensor-driven variation, weather-driven variation and institutional change. (*a*) The statistical method used to analyse significant differences in averages from multiple time series. (*b*) Application of the trend test. (*c*) The testing sequence used to determine phenological differences between two time series.

for each time series separately. If the phenological model fits well, the parameter coefficients of the models are compared. The residuals of each phenological model are then submitted to a residual analysis to confirm correct model fits.

8. Case study

8.1 *Background*

Kazakhstan is the ninth largest country in the world covering the area from the Caspian Sea in the west to the Chinese border in the east, Uzbekistan in the south and Russian Siberian in the north. The total surface area is 2.72 million km^2, divided into nineteen terrestrial ecoregions according to the World Wildlife Fund (Olson *et al.* 2001).

During the Virgin Lands Programme, between 1953 and 1956, Kazakhstan's principal export products became cereal grains (mostly wheat), wool and meat. This continued to independence in 1991, at which time Kazakhstan supplied 27% of the Russian demand for wheat (Kaser 1997). Most of the wheat supply was spring wheat. Until the 1980s, most of the cultivated area belonged either to collectives or state farms, but the management was always highly centralized (Johnson and McConnel Brooks 1983). Most decisions were developed from Moscow, based on annual and five-year plans. This centralized planning led to large inter-annual variation in agricultural output (Johnson and McConnel Brooks 1983, see also Brada 1986).

As a result of the collapse of the Soviet Union in 1991, centralized planning ceased, inter-governmental trade agreements were postponed, and there was a decline in Russian demand of goods from Kazakhstan. The agricultural sector responded to these shocks with a decrease in wheat area cultivated, sharp declines in cattle and sheep numbers, large decreases in pesticide and fertilizer use, and an eventual decline in agricultural machinery (Baydildina *et al.* 2000).

8.2 *Data preparation*

PAL NDVI data have been corrected for changes in sensor calibration, ozone absorption, Raleigh scattering and sensor degradation after pre-launch calibration, and have been normalized for changes in solar zenith angle (Kaufmann *et al.* 2000). Additionally, cloud contamination was minimized using the maximum value compositing technique by generating 10-day composites. A version of the modified best index slope extraction algorithm (Lovell and Graetz 2001) was applied to the data to remove remaining cloud contamination and artifacts left by the maximum value compositing technique. We selected the NDVI composites for Kazakhstan during the growing season from the last 10-day period of April through September from two sensors: NOAA-9 (1985–1988, $n=64$ images) before institutional change and NOAA-14 (1995–1999, $n=80$ images) after institutional change. Composites from NOAA-7 and NOAA-11 were excluded because of documented sensor artifacts that could result in the detection of spurious trends (de Beurs and Henebry 2004b). From these data we arbitrarily selected 19 continuous representative subsets of 1600 km^2, one for each ecoregion in Kazakhstan (figure 4). We averaged the pixels within each subset to create 19 separate time series each spanning 144 image composites. Ecoregions as defined by the WWF reflect the *potential* vegetation (Olson *et al.* 2001); however, vegetation can change within a region under

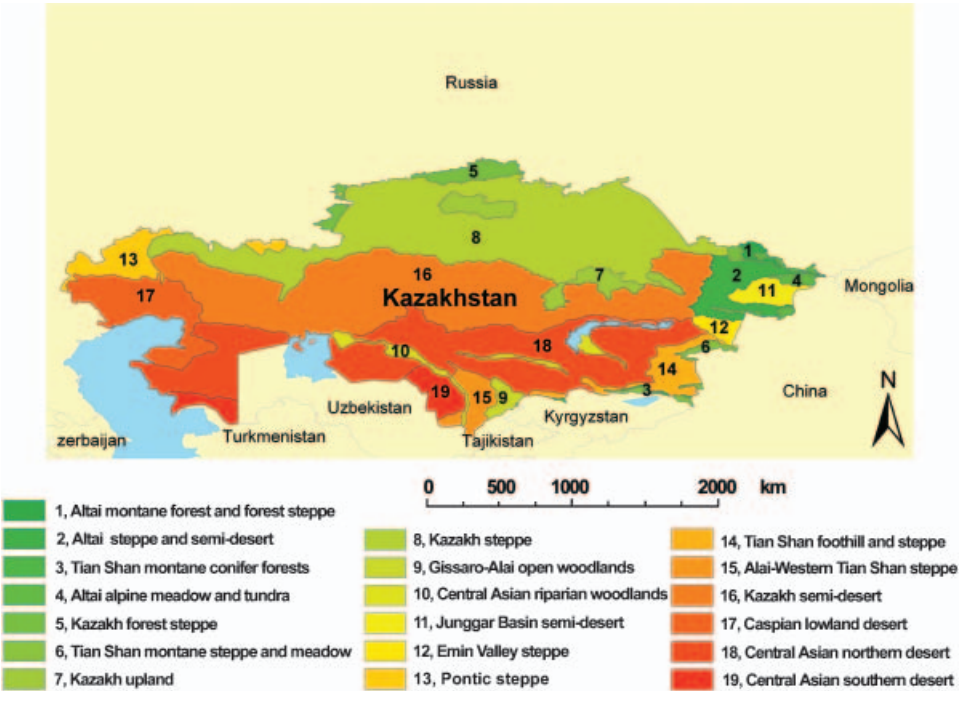

Figure 4. Kazakhstan as partitioned into 19 WWF ecoregions. The larger ecoregions are desert and steppe regions. The ecoregions are colour coded according to their NDVI values with greener shades indicating higher average NDVI.

anthropogenic influence and, indeed, croplands are the dominant land cover in certain ecoregions.

In figure 4 the ecoregions are colour coded: dark green represents the ecoregion with the highest average NDVI and yellow the ecoregion with the lowest NDVI. The highest average NDVI occurs in the Altai Mountains of eastern Kazakhstan. The largest ecoregions are desert regions in central and south Kazakhstan. While both the 'Pontic steppe' and the 'Kazakh steppe' are steppe areas in Kazakhstan, much of these grasslands have been converted to dryland agriculture.

8.3 *Auxiliary climate data*

Climate in several parts in the world is observable only through sparse networks of weather stations. Kazakhstan is one such area. Many of the observational records at available stations are short or patchy and fail to cover the entire observational period from 1985 to 1999. Furthermore, most of the available data have been summarized into monthly averages. There are not enough weather station data available to fit phenological models reliably at finer (sub-monthly) resolutions. To address this data gap, we used air temperature data from the National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP–NCAR) Reanalysis Project (Kalnay *et al.* 1996, Kistler *et al.* 2001) as a surrogate dataset. The NCEP–NCAR Reanalysis dataset provides daily maximum and minimum temperature data (in K) at 2 m with global coverage, albeit at coarse spatial resolution. The Gaussian grid for a global coverage consists of 192×94 pixels, corresponding to a resolution of $1.875^\circ \times 1.91^\circ$, or roughly 2° by 2° lat/long.

We summarized the daily data into 10-day composites that correspond to the 10-day NDVI composites. We tested for significant differences using the FP test as well as for trends.

9. Results

9.1 *Is the average NDVI before and after institutional change significantly different?*

We applied the C-method to the seven ecoregions that were log-normally distributed; data from the remaining ecoregions were submitted to the non-parametric FP test. Table 1 shows significances resulting from the comparisons between data before and after institutional change. The ecoregions are rank ordered from highest average NDVI to lowest average NDVI.

We found four ecoregions with significant differences ($\alpha=0.10$). There were no ecoregions with significantly different growing degree–days between the two periods. We concluded that temperature increases were not responsible for the observed changes.

9.2 *Is there a trend in the image time series?*

Table 2 presents the *p*-values resulting from the corrected seasonal Mann–Kendall test for NDVI. The ecoregions with significant trends in growing degree–days are shown in bold font.

Table 1. Significant differences from the comparisons of NOAA-9 and NOAA-14 extracted from the multiple comparisons of all sensors. In the case of a normal distribution of the data in both periods, the C-method was applied. In the case of non-normality, the non-parametric Fligner–Policello test (FP) was applied. (**=significant at 0.10 level)

	Ecoregion	Test	Significance
1	Altai montane forest and forest steppe	FP	**
2	Altai steppe and semi-desert	FP	
3	Tian Shan montane conifer forest	FP	**
4	Altai alpine meadow and tundra	FP	
5	Kazakh forest steppe	FP	
6	Tian Shan montane steppe and meadow	FP	**
7	Kazakh upland	FP	
8	Kazakh steppe	C	
9	Gissaro-Alai open woodlands	C	**
10	Central Asian riparian woodlands	FP	**
11	Junggar Basin semi-desert	FP	
12	Emin Valley steppe	C	
13	Pontic steppe	FP	
14	Tian Shan foothill arid steppe	FP	
15	Alai-Western Tian Shan steppe	C	
16	Kazakh desert	C	
17	Caspian lowland desert	C	**
18	Central Asian northern desert	C	**
19	Central Asian southern desert	C	

Note: The Fligner–Policello test has a significance level with Dunn–Sidak correction, $b=1-(1-a)^{1/m}$, with m the number of comparisons made. The α-level was chosen as 0.10, and there were six comparisons, resulting in a significance level of $b=0.017$. All *p*-values smaller than the significance level were considered significant.

Table 2. The *p*-values of the corrected seasonal Mann–Kendall test from the periods of NOAA-9 and NOAA-14 for all ecoregions. Values in bold font represent a significant trend ($p<0.05$) in the growing degree–days from the same periods. There are only two ecoregions with a significant trend in both NDVI and growing degree–days simultaneously.

	Ecoregion	NOAA-9 (1985–1988)	NOAA-14 (1995–1999)
1	Altai montane forest and forest steppe	0.06	0.41
2	Altai steppe and semi-desert	0.10	0.38
3	Tian Shan montane conifer forest	**0.12**	<0.01
4	Altai alpine meadow and tundra	**0.09**	0.30
5	Kazakh forest steppe	0.04	0.24
6	Tian Shan montane steppe and meadow	0.11	0.18
7	Kazakh upland	0.19	0.46
8	Kazakh steppe	0.22	0.19
9	Gissaro-Alai open woodlands	0.04	<**0.01**
10	Central Asian riparian woodlands	0.22	0.13
11	Junggar Basin semi-desert	0.50	0.23
12	Emin Valley steppe	0.04	0.05
13	Pontic steppe	0.07	0.14
14	Tian Shan foothill arid steppe	**0.02**	0.02
15	Alai-Western Tian Shan steppe	0.25	0.11
16	Kazakh desert	0.06	0.11
17	Caspian lowland desert	0.19	0.01
18	Central Asian northern desert	0.06	0.16
19	Central Asian southern desert	0.15	0.43

We found four ecoregions with a significant trend either before or after institutional change. There were two ecoregions with significant trends both before and after: the Gissaro-Alai open woodlands and the Tian Shan foothill arid steppe. These ecoregions also displayed trends in growing degree–days in one of the time periods.

9.3 *Did the land surface phenologies in Kazakhstan change after independence?*

In this section we restrict our attention to the ecoregions for which the regression models explained a significant proportion of NDVI variation ($R^2_{adj}>0.5$) for at least one of the periods (NOAA-9 or NOAA-14). Table 3 presents the final models with R^2_{adj} and the model type found for each ecoregion.

Not surprisingly, we found that the variability in the NDVI from ecoregions with higher values is better explained by AGDDs than the variability of the NDVI from desert and arid steppe regions. In contrast to the well-vegetated regions in northern Kazakhstan, the arid regions in the south are usually moisture limited.

We were able to identify twelve ecoregions with well fitting models. Four ecoregions followed Type IV behaviour with all parameter coefficients different between the first and the second period (figure 5(*a*)–(*d*)). These ecoregions are located in the foothills and higher elevations of the Altai Mountains in eastern Kazakhstan and are the regions with highest average NDVI. Altai montane forest and forest steppe was the only ecoregion that peaked at a later bioclimatological time (+ 300°C AGDD) after institutional change. The other three ecoregions with Type IV behaviour all peaked earlier (−16°C, −63°C and −71°C). We found that all four regions displayed an increased intercept (36%, 48%, 158% and 122%),

Table 3. Phenological models for all ecoregions with well fitting quadratic models ($R^2_{adj}>0.5$ in at least one period). The first model is based on NOAA-9 data and the second model is based on NOAA-14 data. The models are divided into four change types from the period of NOAA-9 to NOAA-14 (I=no change, II=intercept change, III=change in intercept and slope, IV= all parameters change).

	Ecoregion	Model (NOAA-9 and NOAA-14)	R^2_{adj}	Type
1	Altai montane forest and forest steppe	$0.285+1.089\times10^{-3}$ AGDD -5.958×10^{-7} AGDD^2 $0.388+6.767\times10^{-4}$ AGDD -2.791×10^{-7} AGDD^2	0.86 0.74	IV
2	Altai steppe and semi-desert	$0.264+7.297\times10^{-4}$ AGDD -3.288×10^{-7} AGDD^2 $0.391+4.908\times10^{-4}$ AGDD -2.244×10^{-7} AGDD^2	0.78 0.69	IV
4	Altai alpine meadow and tundra	$0.111+9.709\times10^{-4}$ AGDD -4.064×10^{-7} AGDD^2 $0.287+6.841\times10^{-4}$ AGDD -3.024×10^{-7} AGDD^2	0.77 0.74	IV
5	Kazakh forest steppe	$0.184+6.284\times10^{-4}$ AGDD -2.412×10^{-7} AGDD^2 $0.263+5.846\times10^{-4}$ AGDD -2.412×10^{-7} AGDD^2	0.75 0.74	III
6	Tian Shan montane steppe and meadow	$0.134+8.647\times10^{-4}$ AGDD -3.855×10^{-7} AGDD^2 $0.298+6.167\times10^{-4}$ AGDD -2.935×10^{-7} AGDD^2	0.83 0.64	IV
7	Kazakh upland	$0.150+5.799\times10^{-4}$ AGDD -2.002×10^{-7} AGDD^2 $0.243+5.298\times10^{-4}$ AGDD -2.002×10^{-7} AGDD^2	0.72 0.84	III
8	Kazakh steppe	$0.120+5.579\times10^{-4}$ AGDD -2.002×10^{-7} AGDD^2 $0.212+4.894\times10^{-4}$ AGDD -2.002×10^{-7} AGDD^2	0.76 0.68	III
9	Gissaro-Alai open woodlands	$0.446-2.830\times10^{-8}$ AGDD^2 $0.491-2.830\times10^{-8}$ AGDD^2	0.73 0.59	II
10	Central Asian riparian woodlands	$0.062+2.981\times10^{-4}$ AGDD -6.480×10^{-8} AGDD^2 $0.119+2.981\times10^{-4}$ AGDD -6.480×10^{-8} AGDD^2	0.74 0.86	II
11	Junggar Basin semi-desert	$0.192+3.207\times10^{-4}$ AGDD -1.540×10^{-7} AGDD^2 same model	0.35 0.76	I
14	Tian Shan foothill arid steppe	$0.153+2.212\times10^{-4}$ AGDD -8.440×10^{-8} AGDD^2 same model	0.20 0.60	I
15	Alai-Western Tian Shan steppe	$0.304-7.418\times10^{-5}$ AGDD $+8.300\times10^{-9}$ AGDD^2 same model	0.67 0.46	I

indicating higher NDVI values at the end of April when our seasonal observational period commenced.

We found Type III behaviour for the three most northern ecoregions, Kazakh forest steppe, Kazakh upland and Kazakh steppe (figure 5(*e*)–(*g*)). Estimates for the quadratic coefficients were equal in both periods, while the linear coefficients decreased and the intercepts increased. The lower linear coefficients resulted in a peak NDVI at fewer AGDD (-91°C, -171°C and -125°C, respectively) and the intercepts increased (43%, 77% and 62%, respectively). These three ecoregions are located in Kazakhstan's spring wheat belt. Prior to institutional change there were vast areas of dryland spring wheat cultivation throughout the region. Following institutional change spring wheat cultivation was reduced and, in marginal lands, abandoned. These areas were not directly added to the rangelands, since there was a concomitant decline in livestock. We concluded that the seasonal advancement and accompanying increase in NDVI at the beginning of the observed growing season is

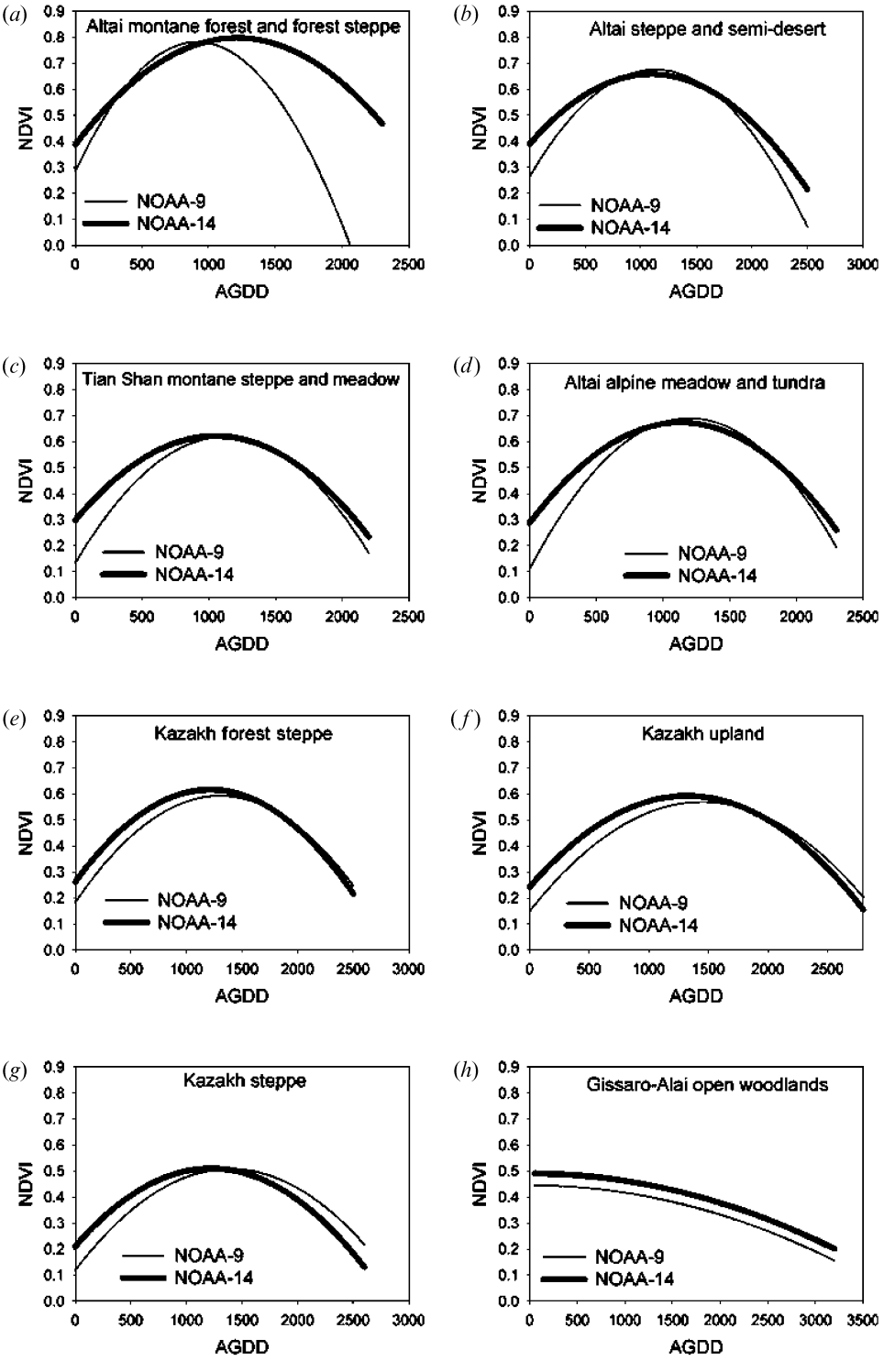

Figure 5. Final models of ecoregions with well fitted models in at least one period ($R^2_{adj}>0.5$). The ecoregions are ordered by average NDVI with the high NDVI ecoregion first.

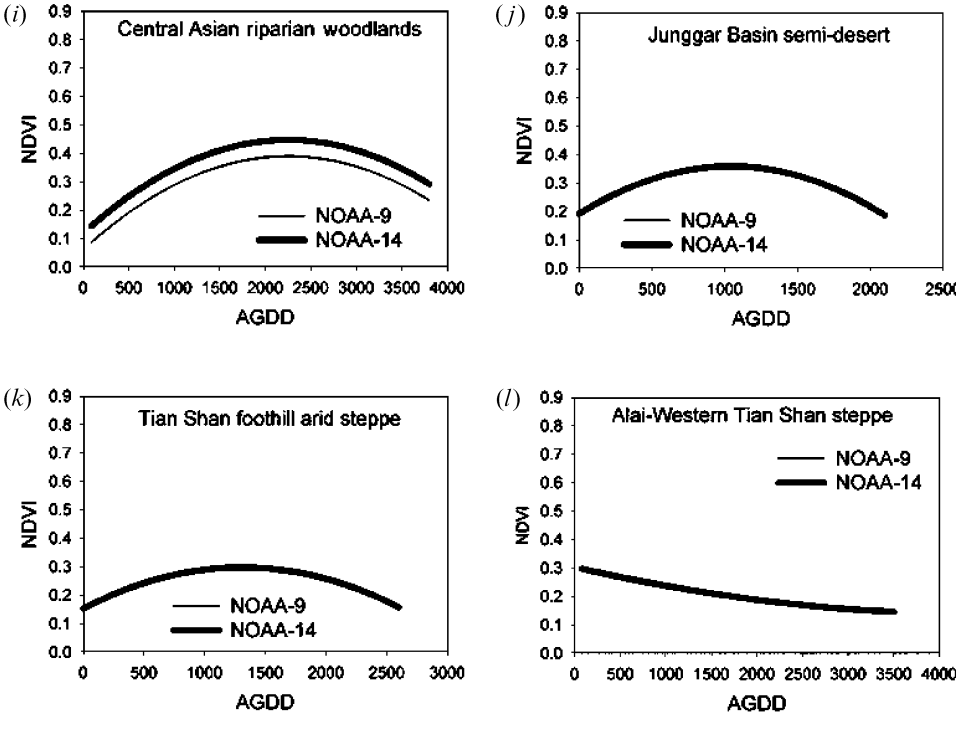

Figure 5. (*Continued.*)

a result of the decline in wheat area in the selected regions. In a related study we focused on the land cover changes in agricultural areas of Kazakhstan (de Beurs and Henebry 2004a) and, using a different spatial partitioning that divided Kazakhstan into functional response regions, we demonstrated that there have been significant changes in land surface phenology in northern Kazakhstan.

Both of the woodland ecoregions, Gissaro-Alai open woodlands and Central Asian riparian woodlands, exhibit Type II behaviour (figure 5(*h*), (*i*)). We found an increase in intercept after institutional change, while the other two parameter coefficients remained constant. The intercept increase was only 10% in the Gissaro-Alai open woodlands in contrast to 92% in the other region. The Central Asian riparian woodlands are located around the Syr-Darya River in southern Kazakhstan and are mostly used for irrigated rice and cotton cultivation.

We found Type I behaviour in Junggar Basin semi-desert, Tian Shan foothill arid steppe and Alai-Western Tian Shan steppe, indicating no change in phenology between the two periods (figure 5(*j*)–(*l*)). There were no significant differences in average NDVI and no significant trends in these regions, except in the Tian Shan foothill arid steppe. Although we interpret these models with caution because the R^2_{adj} is lower than in the agricultural areas, we suggest that the stability of these phenological models supports the assumption of stable desert land cover in these regions.

Well fitting regression models should result in no trends in the residuals and an expected average residual value of zero. The residuals of the models from all ecoregions and each period are slightly smaller in the middle of the growing season around the peak NDVI. However, this decrease is very small and, since the residuals

are random for the rest of the growing season and have an expected mean of zero, we concluded that the models fit the data well. Model fits of the two time periods combined result in residuals that reveal patterns similar to what we have shown with the phenological models. Further residual analyses did not reveal additional information and therefore the results are not reported.

10. Conclusions

In this paper we have presented a statistical framework for the analysis of long image time series. First, we have discussed the importance of the proper use of parametric and non-parametric statistical methods for long image time series analyses. We indicated that the probability of type 1 and type 2 inferential errors can increase dramatically if the assumptions of commonly used statistical methods are violated. We suggested both a parametric and a non-parametric method for the analyses of discontinuities in the image time series, robust against unequal sample sizes and unequal variances. Furthermore, we discussed the risks of trend analysis by simple linear regression and offered an alternative procedure for trend testing that is robust against serial dependence, seasonality and non-normality. We also discussed a method to test for significant changes in phenological pattern between time periods.

To demonstrate these methods, we applied the statistical framework to PAL NDVI data over the ecoregions of Kazakhstan. The primary question in this application was whether the institutional changes in Kazakhstan were of significant magnitude to alter the land surface phenologies of the various ecoregions. The ecoregions with significant phenological models for at least one time period revealed only three general change types. Two ecoregions had a higher overall greenness, three ecoregions displayed earlier peaks and higher greenness at low growing degree–days, and three ecoregions displayed earlier peaks and higher greenness for low growing degree-days with extended green periods. We argued that the changes as revealed by the statistical framework do not result from temperature variability or from sensor artifacts but are understandable from the available socio-economic literature, as we have discussed for agricultural areas elsewhere (de Beurs and Henebry 2004a).

Agricultural areas were expected to show the effects of institutional change most clearly. As institutional changes affected agricultural policy and planning, the resulting changes in the extent and intensity of cultivation would affect land surface phenology. While a previous study used a functional partitioning of selected agricultural regions (de Beurs and Henebry 2004a), here the ecoregional partitioning of the entire country provided an equitable analysis of the differential effects on land surface phenologies. Both studies have produced comparable findings, although in this study we found that the changes were more dramatic in agriculturally dominated ecoregions than in woodlands or deserts – a finding that is in accordance with our expectations.

Yet, another piece of the puzzle remains. Precipitation is a key factor in most ecoregions in Kazakhstan. Droughts have a major impact in land surface phenology, especially in the spring wheat belt of northern Kazakhstan leading to high inter-annual variation in grain production (Spivak *et al.* 1997, Morgounov *et al.* 2000). Here we did not treat variability resulting from variation in precipitation because available datasets generally did not have a fine enough spatial or temporal resolution or did not extend over the entire period of study. As a result there is a

portion of the variation in the data that we cannot adequately explain, but which could likely be explained in large part by seasonal and inter-annual variation in growing season precipitation, were these data available.

Data limitations aside, the statistical framework we have presented in this paper is capable of partitioning the variation observed in image time series into variations due to sensor artifacts, inter-annual temperature variation and human-induced variation such as policy change. Although we have restricted our demonstration of the statistical framework to subsampled averages, it is possible to apply the framework to other aspects of image time series, such as the spatial structure. In the context of social debates on the pace and extent of global change and amidst the confounding effects arising from a lack of observational continuity, it is critical that studies of land cover and land use change follow statistical methodologies that are able to discern with confidence the average from the unusual and expected variation from significant change.

Acknowledgments

We thank Eric Boer for careful readings of an earlier version and for feedback from two anonymous reviewers. This research was sponsored by the NASA Land Cover Land Use Change program. Data used by the authors include data produced through funding from the Earth Observing System Pathfinder Program of NASA's Mission to Planet Earth in cooperation with National Oceanic and Atmospheric Administration. The data were provided by the Earth Observing System Data and Information System (EOSDIS) Distributed Active Archive Center at Goddard Space Flight Center, which archives, manages and distributes this dataset. A contribution of the University of Nebraska Agricultural Research Division, Lincoln, NE, Journal Series No. 14454.

References

ANNE, H., SCHISTAD, S., STORVIK, G., SOLBERG, R. and VOLDEN, E., 1999, Automatic detection of oil spills in ERS SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, pp. 1916–1924.

BAYDILDINA, A., AKSHINBAY, A., BAYETOVA, M., MKRYTICHYAN, L., HALIEPESOVA, A. and ATAEV, D., 2000, Agricultural policy reforms and food security in Kazakhstan and Turkmenistan. *Food Policy*, **25**, pp. 733–747.

BRADA, J.C., 1986, The variability of crop production in private and socialized agriculture: evidence from Eastern Europe. *The Journal of Political Economy*, **94**, pp. 545–563.

CHEN, X. and PAN, W., 2002, Relationships among phenological growing season, time-integrated normalized difference vegetation index and climate forcings in the temperate region of eastern China. *International Journal of Climatology*, **22**, pp. 1781–1792.

CHEN, X., XU, C. and TAN, Z., 2001, An analysis of relationships among plant community phenology and seasonal metrics of Normalized Difference Vegetation Index in the northern part of the monsoon region of China. *International Journal of Biometeorology*, **45**, pp. 170–177.

CONOVER, W.J., 1999, *Practical Nonparametric Statistics*, 3rd edn (New York: Wiley).

DAVIDSON, A. and CSILLAG, F., 2003, A comparison of three approaches for predicting C4 species cover of northern mixed grass prairie. *Remote Sensing of Environment*, **86**, pp. 70–82.

DAY, R.W. and QUINN, G.P., 1989, Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs*, **59**, pp. 433–463.

de Beurs K.M. and Henebry, G.M., 2004a, Land surface phenology, climatic variation, and institutional change: analyzing agricultural land cover change in Kazakhstan. *Remote Sensing of Environment*, **89**, pp. 497–509.

de Beurs K.M. and Henebry, G.M., 2004b, Trend analysis of the Pathfinder AVHRR Land (PAL) NDVI data for the deserts of Central Asia. *Geoscience and Remote Sensing Letters*, **1**, pp. 282–286.

Di, L., Rundquist, D.C. and Han, L., 1994, Modeling relationships between NDVI and precipitation during vegetative growth cycles. *International Journal of Remote Sensing*, **15**, pp. 2121–2136.

Dietz, E.J. and Killeen, T.J., 1981, A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association*, **76**, pp. 169–174.

Dunnet, C.W., 1980, Pairwise multiple comparisons in the unequal variance case. *Journal of the American Statistical Association*, **75**, pp. 796–800.

Dong, Y., Forster, B.C. and Milne, A.K., 2003, Comparison of radar image segmentation by Gaussian- and Gamma-Markov random field models. *International Journal of Remote Sensing*, **24**, pp. 711–722.

Eastman, J.R. and Fulk, M., 1993, Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering and Remote Sensing*, **59**, pp. 1307–1312.

Fligner, M.A. and Policello II G.E., 1981, Robust rank procedures for the Behrens–Fisher problem. *Journal of the American Statistical Association*, **76**, pp. 162–168.

Foody, G.M., 2003, Geographical weighting as a further refinement to regression modeling: An example focused on the NDVI-rainfall relationship. *Remote Sensing of Environment*, **88**, pp. 283–293.

Fuller, D.O., 1998, Trends in NDVI time series and their relation to rangeland and crop production in Senegal, 1987–1993. *International Journal of Remote Sensing*, **19**, pp. 2013–2018.

Goodin, D.G. and Henebry, G.M., 1997, A technique for monitoring ecological disturbance in tallgrass prairie using seasonal NDVI trajectories and a discriminant function mixture model. *Remote Sensing of Environment*, **61**, pp. 270–278.

Hill, M.J. and Donald, G.E., 2003, Estimating spatio-temporal patterns of agricultural productivity in fragmented landscapes using AVHRR NDVI time series. *Remote Sensing of Environment*, **84**, pp. 367–384.

Hirsch, R.M. and Slack, J.R., 1984, A nonparametric trend test for seasonal data with serial dependence. *Water Resources Research*, **20**, pp. 727–732.

Jakubauskas, M.E., Legates, D.R. and Kastens, J.H., 2001, Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, **67**, pp. 461–470.

Jensen, J.R., 1996, *Introductory Digital Image Processing*, 2nd edn (Englewood Cliffs, NY: Prentice-Hall).

Ji, L. and Peters, A.J., 2003, Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. *Remote Sensing of Environment*, **87**, pp. 85–98.

Johnson, D.G. and McConnel Brooks, K., 1983, *Prospects for Soviet Agriculture in the 1980s*, 1st edn (Bloomington, IN: Indiana University Press).

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D., 1996, The NCEP–NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, pp. 437–471.

Kaser, M., 1997, *The Economics of Kazakhstan and Uzbekistan*, 1st edn (London: The Royal Institute of International Affairs, Russia and Eurasia Programme).

KAUFMANN, R.K., ZHOU, L., KNYAZIKHIN, Y., SHABANOV, N.V., MYNENI, R.B. and TUCKER, C.J., 2000, Effect of orbital drift and sensor changes on the time series of AVHRR Vegetation Index data. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, pp. 2584–2597.

KAWABATA, A., ICHII, K. and YAMAGUCHCI, Y., 2001, Global monitoring of interannual changes in vegetation activities using NDVI and its relationships to temperature and precipitation. *International Journal of Remote Sensing*, **22**, pp. 1377–1382.

KESELMAN, H.J., LIX, L.M. and KOWALCHUK, R.K., 1998, Multiple comparison procedures for trimmed means. *Psychological Methods*, **3**, pp. 123–141.

KISTLER, R., KALNAY, E., COLLINS, W., SAHA, S., WHITE, G., WOOLLEN, J., CHELLIAH, M., EBISUZAKI, W., KANAMITSU, M., KOUSKY, V., VAN DEN DOOL, H., JENNE, R. and FIORINO, M., 2001, The NECP–NCAR 50-year reanalysis: Monthly mean CD-ROM and documentation. *Bulletin of the American Meteorological Society*, **82**, pp. 247–267.

KOGAN, F.N. and ZHU, X., 2001, Evolution of long-term errors in NDVI time series: 1985–1999. *Advances in Space Research*, **28**, pp. 149–153.

LABUS, M.P., NIELSEN, G.A., LAWRENCE, R.L. and ENGEL, R., 2002, Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, **23**, pp. 4169–4180.

LEE, R., YU, F., PRICE, K.P., ELLIS, J. and SHI, P., 2002, Evaluating vegetation phenological patterns in Inner Mongolia using NDVI time-series analysis. *International Journal of Remote Sensing*, **23**, pp. 2505–2512.

LI, B., TAO, S. and DAWSON, R.W., 2002, Relations between AVHRR NDVI and ecoclimatic parameters in China. *International Journal of Remote Sensing*, **23**, pp. 989–999.

LOPES, A., TOUZI, R. and NEZRY, E., 1990, Adaptive speckle filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, pp. 992–1000.

LOVELL, J.L. and GRAETZ, R.D., 2001, Filtering Pathfinder AVHRR Land NDVI data for Australia. *International Journal of Remote Sensing*, **22**, pp. 2649–2654.

LU, H., RAUPACH, M.R., MCVICAR, T.R. and BARRETT, D.J., 2003, Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment*, **86**, pp. 1–18.

MITCHELL, R., FRITZ, J., MOORE, K., MOSER, L., VOGEL, K., REDFEARN, D. and WESTER, D., 2001, Predicting forage quality in switchgrass and big bluestem. *Agronomy Journal*, **93**, pp. 118–124.

MOODY, A. and JOHNSON, D.M., 2001, Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment*, **75**, pp. 305–323.

MORGOUNOV, A., KARABAYEV, M., BEDOSHVILI, D. and BRAUN, H.J., 2000, Wheat production in Central Asia and the Caucasus. *Research Highlights of the CIMMYT Wheat Program 1999–2000* (El Batan, Mexico: CIMMYT).

MYNENI, R.B., KEELING, C.D., TUCKER, C.J., ASRAR, G. and NEMANI, R.R., 1997, Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature*, **386**, pp. 698–702.

NIGHTINGALE, J.M. and PHINN, S.R., 2003, Assessment of relationships between precipitation and satellite derived vegetation condition within South Australia. *Australian Geographical Studies*, **41**, pp. 180–195.

OLSON, D.M., DINERSTEIN, E., WIKRAMANAYAKE, E.D., BURGESS, N.D., POWELL, G.V.N., UNDERWOOD, E.C., D'AMICO, J.A., ITOUA, I., STRAND, H.E., MORRISON, J.C., LOUCKS, C.J., ALLNUTT, T.F., RICKETTS, T.H., KURA, Y., LAMOREUX, J.F., WETTENGEL, W.W., HEDAO, P. and KASSEM, K.R., 2001, Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, **51**, pp. 933–938.

PETERS, A.J., WALTER-SHEA, E.A., JI, L., VIŇA, A., HAYES, M. and SVOBODA, M.D., 2002, Drought monitoring with NDVI-based standardized vegetation index. *Photogrammetric Engineering and Remote Sensing*, **68**, pp. 71–75.

PIWOWAR, J.M. and LEDREW, E.F., 2002, ARMA time series modeling of remote sensing imagery: a new approach for climate change studies. *International Journal of Remote Sensing*, **23**, pp. 5225–5248.

POTTER, C., TAN, P., STEINBACH, M., KLOOSTER, S., KUMAR, V., MYNENI, R. and GENOVESE, V., 2003, Major disturbance events in terrestrial ecosystems detected using global satellite data sets. *Global Change Biology*, **9**, pp. 1005–1021.

POTTER, C.S. and BROOKS, V., 2001, Global analysis of empirical relations between annual climate and seasonality of NDVI. *International Journal of Remote Sensing*, **19**, pp. 2921–2948.

RICHARD, Y. and POCCARD, I., 1998, A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. *International Journal of Remote Sensing*, **19**, pp. 2907–2920.

RICKMAN, R.W. and KLEPPER, E.L., 1991, Tillering in wheat. In *Predicting Crop Phenology*, T. Hodges (Ed.), pp. 73–85 (Boca Raton, FL: CRC Press).

SCHWARTZ, M.D., 2003, *Phenology: An Integrative Environmental Science*, 1st edn (Boston: Kluwer).

SHABANOV, N.V., ZHOU, L., KNYAZIKHIN, Y., MYNENI, R.B. and TUCKER, C.J., 2002, Analysis of interannual changes in northern vegetation activity observed in AVHRR data from 1981 to 1994. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, pp. 115–130.

SLAYBACK, D.A., PINZON, J.E., LOS, S.O. and TUCKER, C.J., 2003, Northern hemisphere photosynthetic trends 1982–1999. *Global Change Biology*, **9**, pp. 1–15.

SMART, A.J., SCHACHT, W.H. and MOSER, L.E., 2001, Predicting leaf/stem ratio and nutritive value in grazed and nongrazed big bluestem. *Agronomy Journal*, **93**, pp. 1243–1249.

SPIVAK, L., TEREHOV, A., MURATOVA, N. and ARKHIPKIN, O., 1997, The method of early drought detection with AVHRR/NOAA data. *Digest of IGARSS 1997* (Piscataway NJ: IEEE), pp. 281–283.

STOLINE, M.R., 1981, The status of multiple comparisons: simultaneous estimation of all pairwise comparison in one-way ANOVA designs. *The American Statistician*, **35**, pp. 134–141.

SUZUKI, R., NOMAKI, T. and YASUNARI, T., 2003, West–east contrast of phenology and climate in northern Asia revealed using a remotely sensed vegetation index. *International Journal of Climatology*, **47**, pp. 126–138.

TUCKER, C.J., SLAYBACK, D.A., PINZON, J.E., LOS, S.O., MYNENI, R.B. and TAYLOR, M.G., 2001, Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *International Journal of Biometeorology*, **45**, pp. 184–190.

VON STORCH, H. and NAVARRA, A., 1999, *Analysis of Climate Variability/Applications of Statistical Techniques*, 1st edn (Berlin: Springer-Verlag).

WANG, J., PRICE, K.P. and RICH, P.M., 2001, Spatial patterns of NDVI response to precipitation and temperature in the central Great Plains. *International Journal of Remote Sensing*, **22**, pp. 3827–3844.

ZAR, J.H., 1984, *Biostatistical Analysis*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall).