8

Relating Time-Series of Meteorological and Remote Sensing Indices to Monitor Vegetation Moisture Dynamics

J. Verbesselt, P. Jönsson, S. Lhermitte, I. Jonckheere, J. van Aardt, and P. Coppin

CONTENTS

8.1 Introduction

The repeated occurrence of severe wildfires, which affect various fire-prone ecosystems of the world, has highlighted the need to develop effective tools for monitoring fire-related parameters. Vegetation water content (VWC), which influences the biomass burning processes, is an example of one such parameter [1–3]. The physical definitions of VWC vary from water volume per leaf or ground area (equivalent water thickness) to water mass per mass of vegetation (dry or fresh matter and fuel moisture content) [4]. Therefore, AQ1
VWC could also be used to infer vegetation water stress and to assess drought conditions that linked with fire risk [5]. Decreases in VWC due to the seasonal decrease in available

soil moisture can induce severe fires in most ecosystems. VWC is particularly important for determining the behavior of fires in savanna ecosystems because the herbaceous layer becomes especially flammable during the dry season when the VWC is low [6–7].

Typically, VWC in savanna ecosystems is measured using labor-intensive vegetation sampling. Several studies, however, indicated that VWC can be characterized temporally and spatially using meteorological or remote sensing data, which could contribute to the monitoring of fire risk [1,4]. The meteorological Keetch–Byram drought index (KBDI) was selected for this study. This index was developed to incorporate soil water content in the root zone of vegetation and is able to assess the seasonal trend of VWC [3,8]. The KBDI is a cumulative algorithm for the estimation of fire potential from meteorological information, including daily maximum temperature, daily total precipitation, and mean annual precipitation [9–10]. The KBDI also has been used for the assessment of VWC for vegetation types with shallow rooting systems, for example, the herbaceous layer of the savanna ecosystem [8,11].

The application of drought indices, however, presents specific operational challenges. These challenges are due to the lack of meteorological data for certain areas, as well as spatial interpolation techniques that are not always suitable for use in areas with complex terrain features. Satellite data provide sound alternatives to meteorological indices in this context. Remotely sensed data have significant potential for monitoring vegetation dynamics at regional to global scale, given the synoptic coverage and repeated temporal sampling of satellite observations (e.g., SPOT VEGETATION or NOAA AVHRR) [12–13]. These data have the advantage of providing information on remote areas where ground measurements are impossible to obtain on a regular basis.

Most research in the scientific community using optical sensors (e.g., SPOT VEGETATION) to study biomass burning has focused on two research areas[4]: (1) the direct estimation of VWC and (2) the estimation of chlorophyll content or degree of drying as an alternative to the estimation of VWC. Chlorophyll-related indices are related to VWC based on the hypothesis that the chlorophyll content of leaves decreases proportionally to the VWC [4]. This assumption has been confirmed for selected species with shallow rooting systems (e.g., grasslands and understory forest vegetation) [14–16], but cannot be generalized to all ecosystems [4]. Therefore, chlorophyll-related indices, such as the normalized difference vegetation index (NDVI), only can be used in regions where the relationship among chlorophyll content, degree of curing, and water content have been established.

Accordingly, a remote sensing index that is directly coupled to the VWC is used to investigate the potential of hyper-temporal satellite imagery to monitor the seasonal vegetation moisture dynamics. Several studies [4,16–17] have demonstrated that VWC can be estimated directly through the normalized difference of the near infrared reflectance (NIR, 0.78–0.89 μm) ρ_{NIR}, influenced by the internal structure and the dry matter, and the shortwave infrared reflectance (SWIR, 1.58–1.75 μm) ρ_{SWIR}, influenced by plant tissue water content:

$$\mathrm{NDWI} = \frac{\rho_{NIR} - \rho_{\mathrm{SWIR}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR}}} \tag{8.1}$$

The NDWI or normalized difference infrared index (NDII) [19] is similar to the global vegetation moisture index (GVMI) [20].

The relationship between NDWI and KBDI time-series, both related to VWC dynamics, is explored. Although the value of time-series data for monitoring vegetation moisture dynamics has been firmly established [21], only a few studies have taken serial correlation into account when correlating time-series [6,22–25]. Serial correlation occurs when data collected through time contains values at time t, which are correlated with observations at

time t–1. This type of correlation in time-series, when related to VWC dynamics, is mainly caused by the seasonal variation (dry–wet cycle) of vegetation [26]. Serial correlation can be used to forecast future values of the time-series by modeling the dependence between observations but affects correlations between variables measured in time and violates the basic regression assumption of independence [22]. Correlation coefficient of serially correlated data cannot be used as indicators of goodness-of-fit of a model as the correlation coefficients are artificially inflated [22,27].

The study of the relationship between NDWI and KBDI is a nontrivial task due to the effect of serial correlation. Remedies for serial correlation include sampling or aggregating the data over longer time intervals, as well as further modeling, which can include techniques such as weighted regression [25,28]. However, it is difficult to account for serial correlation in time-series related to VWC dynamics using extended regression techniques. The time-series related to VWC dynamics often exhibit high non-Gaussian serial correlation and are more significantly affected by outliers and measurement errors [28]. A sampling technique therefore is proposed, which accounts for serial correlation in seasonal time-series, to study the relationship between different time-series. The serial correlation effect in time-series is assumed to be minimal when extracting one metric per season (e.g., start of the dry season). The extracted seasonal metrics are then utilized to study the relationship between time-series at a specific moment in time (e.g., start of the dry season).

The aim of this chapter is to address the effect of serial correlation when studying the relationship between remote sensing and meteorological time-series related to VWC by comparing nonserially correlated seasonal metrics from time-series. This chapter therefore has three defined objectives. Firstly, an overview of time-series analysis techniques and concepts (e.g., stationarity, autocorrelation, ARIMA, etc.) is presented and the relationship between time-series is studied using cross-correlation and ordinary least square (OLS) regression analysis. Secondly, an algorithm for the extraction of seasonal metrics is optimized for satellite and meteorological time-series. Finally, the temporal occurrence and values of the extracted nonserially correlated seasonal metrics are analyzed statistically to define the quantitative relationship between NDWI and KBDI time-series. The influence of serial correlation is illustrated by comparing results from cross-correlation and OLS analysis with the results from the investigation of correlation between extracted metrics.

8.2 Data

8.2.1 Study Area

The Kruger National Park (KNP), located between latitudes 23°S and 26°S and longitudes 30°E and 32°E in the low-lying savanna of the northeastern part of South Africa, was selected for this study (Figure 8.1). Elevations range from 260 to 839 m above sea level, and mean annual rainfall varies between 350 mm in the north and 750 mm in the south. The rainy season within the annual climatic season can be confined to the summer months (i.e., November to April), and over a longer period can be defined by alternating wet and dry seasons [7]. The KNP is characterized by an arid savanna dominated by thorny, fine-leafed trees of the families *Mimosaceae* and *Burseraceae*. An exception is the northern part of the KNP where the Mopane, a broad-leafed tree belonging to the *Ceasalpinaceae*, almost completely dominates the tree layer.

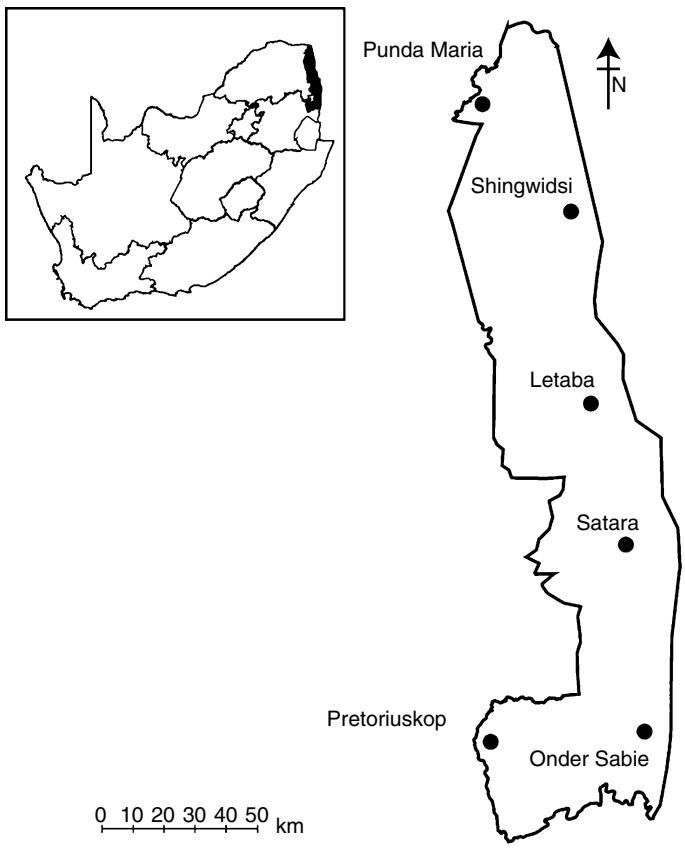

FIGURE 8.1
The Kruger National Park (KNP) study area with the weather stations used in the analysis (right). South Africa is shown with the borders of the provinces and the study area (top left).

8.2.2 Climate Data

Climate data from six weather stations in the KNP with similar vegetation types were used to estimate the daily KBDI (Figure 8.1). KBDI was derived from daily precipitation and maximum temperature data to estimate the net effect on the soil water balance [3]. Assumptions in the derivation of KBDI include a soil water capacity of approximately 20 cm and an exponential moisture loss from the soil reservoir. KBDI was initialized during periods of rainfall events (e.g., rainy season) that result in soils with maximized field capacity and KBDI values of zero [8]. The preprocessing of KBDI was done using the method developed by Janis et al. [10]. Missing daily maximum temperatures were replaced with interpolated values of daily maximum temperature, based on a linear interpolation function [29]. Missing daily precipitation, on the other hand, was assumed to be zero. A series of error logs were automatically generated to indicate missing precipitation values and associated estimated daily KBDI values. This was done because zeroing missing precipitation may lead to an increased fire potential bias in KBDI. The total percentage of missing data gaps in rainfall and temperature series was maximally 5% during the study period for each of the six weather stations. The daily KBDI time-series were transformed into 10-daily KBDI series, similar to the SPOT VEGETATION S10 dekads (i.e., 10-day periods), by taking the maximum of

each dekad. The negative of the KBDI time-series (i.e., –KBDI) was analyzed in this chapter such that the temporal dynamics of –KBDI and NDWI were related (Figure 8.2). The –KBDI and NDWI are used throughout this chapter. The Satara weather station, centrally positioned in the study area, was selected to represent the temporal vegetation dynamics. The other weather stations in the study area demonstrate similar temporal vegetation dynamics.

8.2.3 Remote Sensing Data

The data set used is composed of 10-daily SPOT VEGETATION (SPOT VGT) composites (S10 NDVI maximum value syntheses) acquired over the study area for the period April 1998 to December 2002. SPOT VGT can provide local to global coverage on a regular basis (e.g., daily for SPOT VGT). The syntheses result in surface reflectance in the blue (0.43–0.47 μm), red (0.61–0.68 μm), NIR (0.78–0.89 μm), and SWIR (1.58–1.75 μm) spectral regions. Images were atmospherically corrected using the simplified method for atmospheric correction (SMAC) [29]. The geometrically and radiometrically corrected S10 images have a spatial resolution of 1 km.

The S10 SPOT VGT time-series were preprocessed to detect data that erroneously influence the subsequent fitting of functions to time-series, necessary to define and extract metrics [6]. The image preprocessing procedures performed were:

- Data points with a satellite viewing zenith angle (VZA) above 50° were masked out as pixels located at the very edge of the image (VZA $>$ 50.5°) swath are affected by re-sampling methods that yield erroneous spectral values.
- The aberrant SWIR detectors of the SPOT VGT sensor, flagged by the status mask of the SPOT VGT S10 synthesis, also were masked out.
- A data point was classified as cloud-free if the blue reflectance was less than 0.07 [31]. The developed threshold approach was applied to identify cloud-free pixels for the study area.

NDWI time-series were derived by selecting savanna pixels, based on the land cover map of South Africa [32], for a 3 $\times$ 3 pixel window centered at each of the meteorological

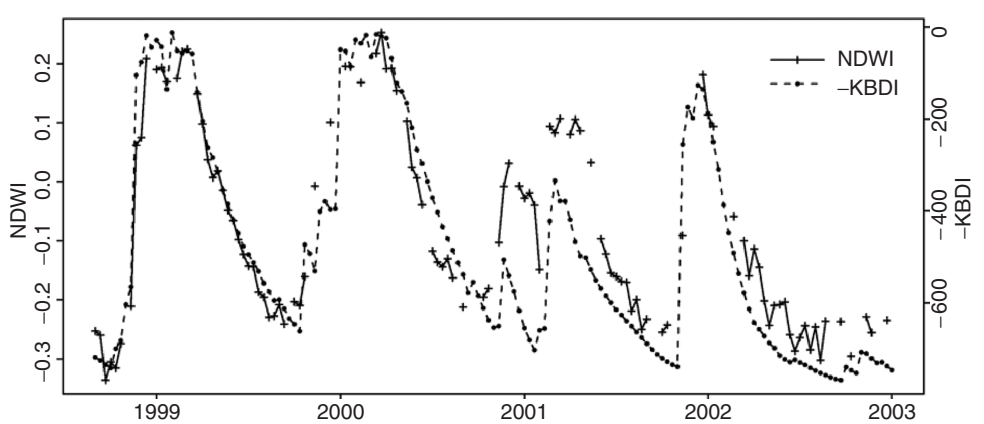

FIGURE 8.2
The temporal relationship between NDWI and –KBDI time-series for the ''Satara'' weather station (Figure 8.1).

stations to reduce the effect of potential spatial misregistration (Figure 8.1). Median values of the 9-pixel windows then were retained instead of single pixel values [33]. The median was preferred to average values as it is less affected by extreme values and therefore is less sensitive to potentially undetected data errors.

8.3 Serial Correlation and Time-Series Analysis

Serial correlation affects correlations between variables measured in time, and violates the basic regression assumption of independence. Techniques that are used to recognize serial correlation therefore are discussed by applying them to the NDWI and –KBDI time-series. Cross-correlation analysis is illustrated and used to study the relationship between time-series of –KBDI and NDWI. Fundamental time-series analysis concepts (e.g., stationarity and seasonality) are introduced and a brief overview is presented of the most frequently used method for time-series analysis to account for serial correlation, namely autoregression.

8.3.1 Recognizing Serial Correlation

This chapter focuses on discrete time-series, which contain observations made at discrete time intervals (e.g., 10 daily time steps of –KBDI and NDWI time-series). A time-series is defined as a set of observations, x_t, recorded at a specific time, t [26]. Time-series of –KBDI and NDWI contain a seasonal variation which is illustrated in Figure 8.2 by a smooth increase or decrease of the series related to vegetation moisture dynamics. The gradual increase or decrease of the graph of a time-series is generally indicative of the existence of a form of dependence or serial correlation among observations.

The presence of serial correlation systematically biases regression analysis when studying the relationship between two or more time-series 25. Consider the OLS regression line with a slope and an intercept:

$$Y(t) = a_0 + a_1 X(t) + e(t) \tag{8.2}$$

where t is time, a_0 and a_1 are the respective OLS regression intercept and slope parameter, $Y(t)$ the dependent variable, $X(t)$ the independent variable, and $e(t)$ the random error term. The standard error of each parameter is required for any regression model to define the confidence interval and to derive the significance of parameters in the regression equation. The parameters a_0, a_1, and the confidence intervals, estimated by minimizing the sum of the squared ''residuals,'' are valid only if certain assumptions related to the regression and $e(t)$ are met [25]. These assumptions are detailed in statistical textbooks [34] but are not always met or explicitly considered in real-world applications. Figure 8.3 illustrates the biased confidence intervals of the OLS regression model at a 95% confidence level. The standard error (SE) term of the regression model is underestimated due to serially correlated residuals and explains the biased confidence interval (CI), where CI $=$ mean $\pm$ 1.96 $\times$ SE.

The Gauss–Markov theorem states that the OLS parameter estimate is the best linear unbiased estimate (BLUE); that is, all other linear unbiased estimates will have a larger variance, if the error term, $e(t)$, is stationary and exhibits no serial correlation. The Gauss–Markov theorem consequently points to the error term and not to the time-series

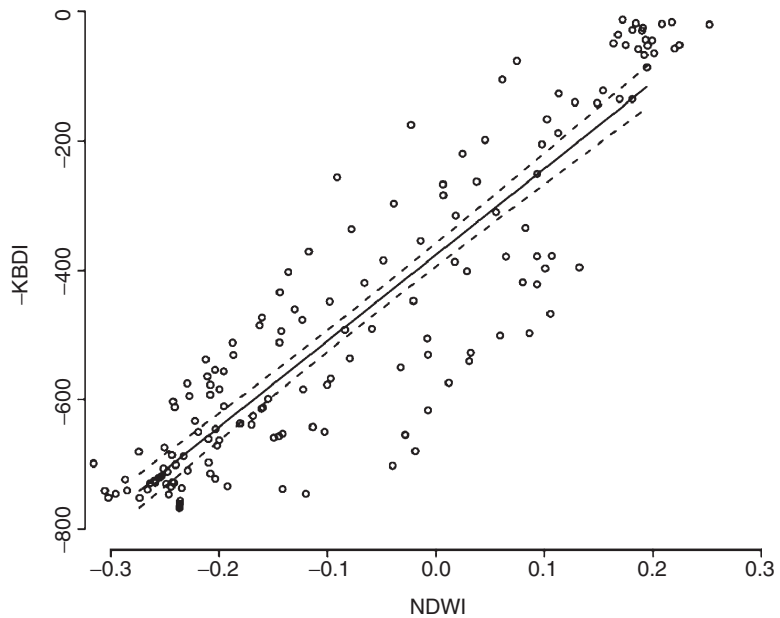

FIGURE 8.3
Result of the OLS regression fit between –KBDI and NDWI as dependent and independent variables, respectively, for the Satara weather station ($n = 157$). Confidence intervals (- - -) at a 95% confidence level are shown, but are ''narrowed'' due to serial correlation in the residuals.

themselves as the critical consideration [35]. The error term is defined as stationary when it does not present a trend and the variance remains constant over time [27]. It is possible that the residuals are serially correlated if one of the dependent or independent variables also is serially correlated, because the residuals constitute a linear combination of both types of variables. Both dependent and independent variables of the regression model are serially correlated (KBDI and NDWI), which explains the serial correlation observed in the residuals.

A sound practice used to verify serial correlation in time-series is to perform multiple checks by both graphical and diagnostic techniques. The autocorrelation function (ACF) can be viewed as a graphical measure of serial correlation between variables or residuals. The sample autocorrelation function is defined when $x_1, \ldots, x_n$ are observations of a time-series. The sample mean of $x_1, \ldots, x_n$ is [26]:

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{8.3}$$

The sample autocovariance function with lag h and time t is:

$$\hat{\gamma}(h) = n^{-1}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x}),\ -n < h < n \tag{8.4}$$

The sample ACF is:

$$\hat{\rho} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},\ -n < h < n \tag{8.5}$$

Figure 8.4 illustrates the ACF for time-series of –KBDI and NDWI presented from the Kruger park data. The ACF clearly indicates a significant autocorrelation in the

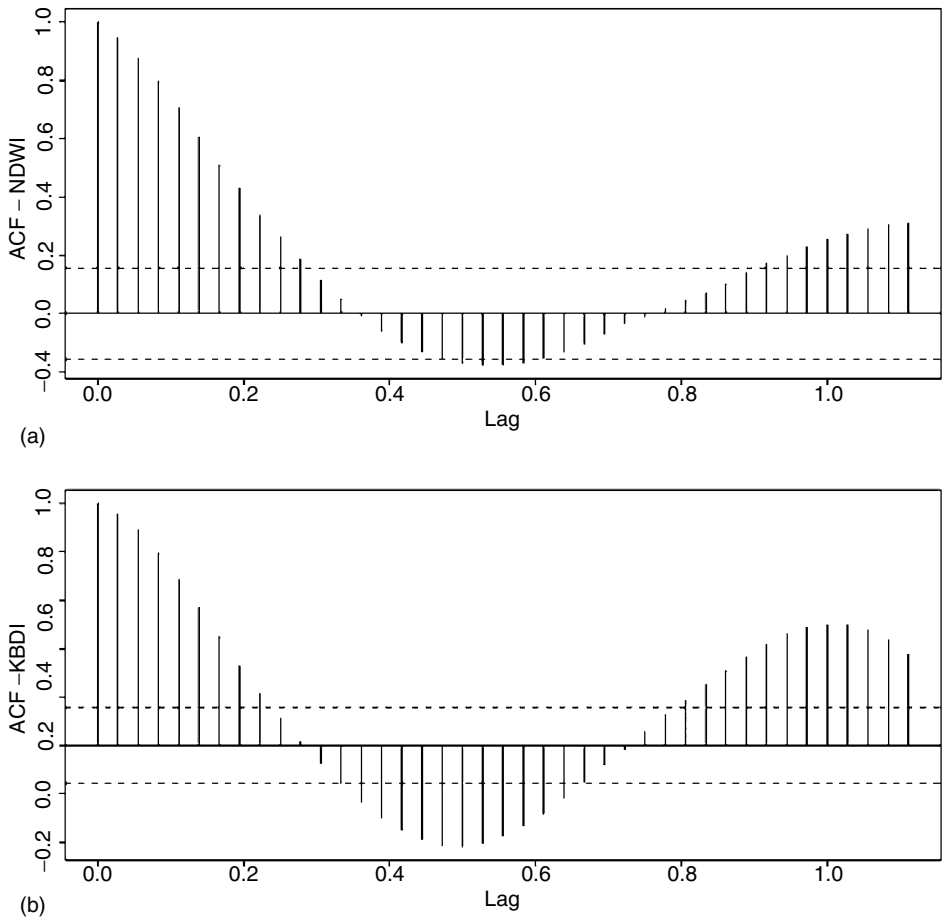

FIGURE 8.4
The autocorrelation function (ACF) for (a) –KBDI and (b) NDWI time-series for the Satara weather station. The horizontal lines on the graph are the bounds $= \pm 1.96/\sqrt{n}$ ($n = 157$).

time-series, as more than 5% of the sample autocorrelations fall outside the significance bounds $= \pm 1.96/\sqrt{n}$ [26]. There are also formal tests available to detect autocorrelation such as the Ljung–Box test statistic and the Durbin–Watson statistic [25,26].

8.3.2 Cross-Correlation Analysis

The cross-correlation function (CCF) can be derived between two time-series utilizing a technique similar to the autocorrelation function applied for one time-series [27]. Cross-correlation is a measure of the degree of linear relationship existing between two datasets and can be used to study the connection between time-series. The CCF, however, can only be used if the time-series is stationary [27]. For example, when all variables are increasing in value over time, cross-correlation results will be spurious and subsequently cannot be used to study the relationship between time-series.

Nonstationary time-series can be transformed to stationary time-series by implementing one of the following techniques:

- Differencing the time-series by a period d can yield a series that satisfies the assumption of stationarity (e.g., $x_t - x_{t-1}$ for $d = 1$). The differenced series will

contain one point less than the original series. Although a time-series can be differenced more than once, one difference is usually sufficient.

- Lower order polynomials can be fitted to the series when the data contains a trend or seasonality that needs to be subtracted from the original series. Seasonal time-series can be represented as the sum of a specified trend, and seasonal and random terms. For example, for statistical interpretation results, it is important to recognize the presence of seasonal components and to remove them to avoid confusion with long-term trends. Figure 8.5 illustrates the seasonal trend decompositioning method using locally weighted regression for the NDWI time-series [36].
- The logarithm or square root of the series may stabilize the variance in the case of a nonconstant variance.

Figure 8.6 illustrates the cross-correlation plot for stationary series of –KBDI and NDWI. –KBDI and NDWI time-series became stationary after differencing with $d = 1$. The stationarity was confirmed using the ''augmented Dickey–Fuller'' test for stationarity [26,29] at a confidence level of 95% ($p < 0.01$; with stationarity as the alternative hypothesis). Note that approximately 95% confidence limits are shown for the autocorrelation plots of an independent series. These limits must be regarded with caution, since there exists an *a priori* expectation of serial correlation for time series [37].

Table 8.1 illustrates the coefficients of determination (i.e., multiple R^2) of the OLS regression analysis with serially correlated residuals –KBDI as dependent and NDWI as

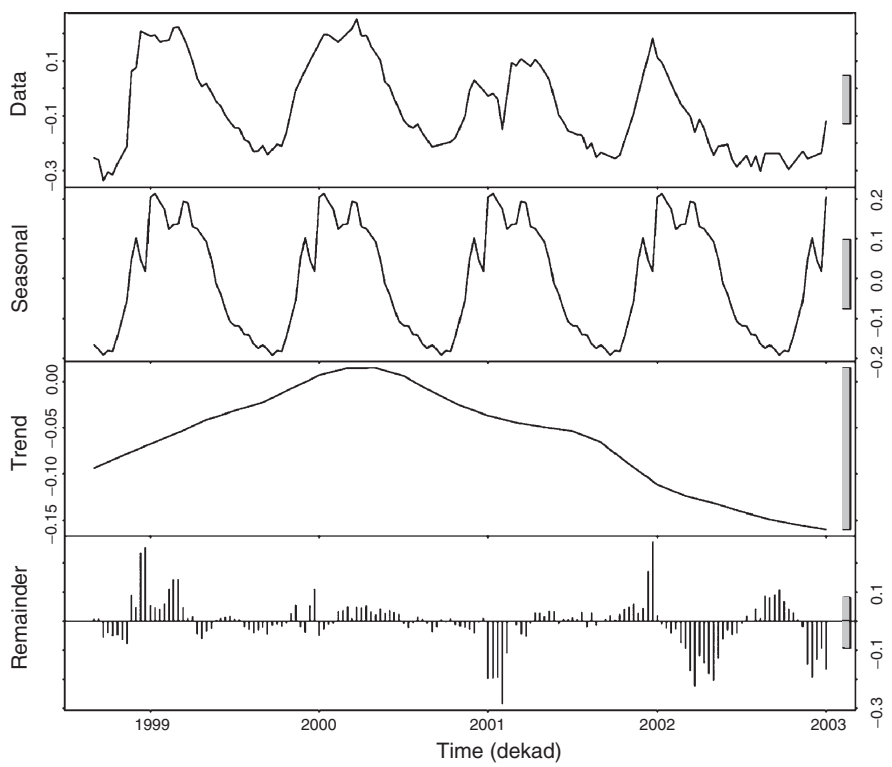

FIGURE 8.5
The results of the seasonal trend decomposition (STL) technique for NDWI time-series of Satara weather station. The original series can be reconstructed by summing the seasonal, trend, and remainder. In the *y*-axes the NDWI values are indicated. The grey bars at the right-hand side of the plots illustrate the relative data range of the time-series.

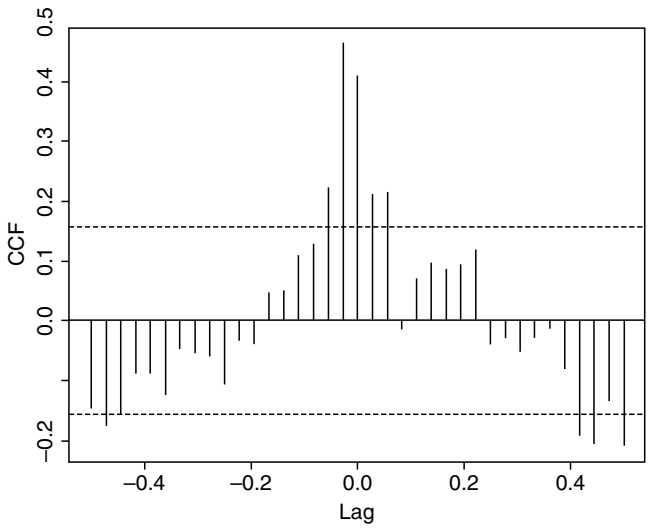

FIGURE 8.6
The cross-correlation plot between stationary –KBDI and NDWI time-series of the Satara weather station, where CCF indicates results of the cross-correlation function. The horizontal lines on the graph are the bounds ($= \pm 1.96/\sqrt{n}$) of the approximate 95% confidence interval.

independent variable for all six weather stations in the study area. The Durbin–Watson statistic indicated that the residuals were serially correlated at a 95% confidence level ($p < 0.01$). These results will be compared with the method presented in the ''results and discussion'' section of this chapter. Table 8.1 also indicates the time lags at which correlation between time-series was maximal, as derived from the cross-correlation plot. A negative lag indicates that –KBDI reacts prior to NDWI, for example, in the cases of Punda Maria and Shingwedzi weather stations, and subsequently can be used to predict NDWI. This is logical since weather conditions, for example, rainfall and temperature, change before vegetation reacts. NDWI, which is related to the amount of water in the vegetation, consequently lags behind the –KBDI. The major vegetation type in savanna vegetation is the herbaceous layer, which has a shallow rooting system. This explains why the vegetation in the study area quickly follows climatic changes and NDWI did not lag behind –KBDI for the other four weather stations.

8.3.3 Time-Series Analysis: Relating Time-Series and Autoregression

A remedy for serial correlation, apart from applying variations in sampling strategy, is modeling of the time dependence in the error structure by autoregression (AR). AR most

TABLE 8.1

Coefficients of Determination of the OLS Regression Model Between –KBDI and NDWI are Shown ($n = 157$). The time expressed in dekads of maximum correlation of the cross-correlation between –KBDI and NDWI is also indicated

Station	R^2	Time Lag
Punda Maria	0.74	−1
Letaba	0.88	0
Onder Sabie	0.72	0
Pretoriuskop	0.31	0
Shingwedzi	0.72	−1
Satara	0.81	0

often is used for purposes of forecasting and modeling of a time-series [25]. The simplest AR model for Equation 8.2, where ρ is the result of the sample autocorrelation function at lag 1, is:

$$e_t = \rho e_{t-1} + \varepsilon_t \tag{8.6}$$

where ε_t is a series of serially independent numbers with mean zero and constant variance. The Gauss–Markov theorem cannot be applied and therefore OLS is not an efficient estimator of the model parameters if ρ is not zero [35].

Many different AR models are available in statistical software systems that incorporate time-series modules. One of the most frequently used models to account for serial correlation is the autoregressive- integrated-moving-average model (ARIMA) [26,37]. Briefly stated, ARIMA models can have an autoregressive term (AR) of order p, a differencing (integrating) term (I) of order d, and a moving average term (MA) of order q. The notation for specific models takes the form of (p,d,q) [27]. The order of each term in the model is determined by examining the raw data and plots of the autocorrelation function of the data. For example, a second-order AR ($p = 2$) term in the model would be appropriate if a series has significant autocorrelation coefficients between x_t, and x_{t-1}, and x_{t-2}. ARIMA models that are fitted to time-series data using AR and MA parameters, p and q, have coefficients Φ and θ to describe the serial correlation. An underlying assumption of ARIMA models is that the series being modeled is stationary [26–27].

8.4 Methodology

The TIMESAT program is used to extract nonserially correlated metrics from remote sensing and meteorological time-series [38,39]. These metrics are utilized to study the relationship between time-series at specific moments in time. The relationship between time-series, in turn, is evaluated using statistical analysis of extracted nonserially correlated seasonal metrics from time-series (–KBDI and NDWI).

8.4.1 Data Smoothing

It often is necessary to generate smooth time-series from noisy satellite sensors or meteorological data to extract information on seasonality. The smoothing can be achieved by applying filters or by function fitting. Methods based on Fourier series [40–42] or least-square fits to sinusoidal functions [43–45] are known to work well in most instances. These methods, however, are not capable of capturing a sudden, steep rise or decrease of remote sensing or meteorological data values that often occur in arid and semiarid environments. Alternative smoothing and fitting methods have been developed to overcome these problems [38]. An adaptive Savitzky–Golay filtering method, implemented in the TIMESAT processing package developed by Jönsson and Eklundh [39], is used in this chapter. The filter is based on local polynomial fits. Suppose we have a time-series (t_i, y_i), $i = 1, 2, \ldots, N$. For each point i, a quadratic polynomial

$$f(t) = c_1 + c_2 t + c_3 t^2 \tag{8.7}$$

is fit to all $2k+1$ points for a window from $n = i-k$ to $m = i+k$ by solving the system of normal equations

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}c = \mathbf{A}^{\mathrm{T}}b \tag{8.8}$$

where

$$\mathbf{A} = \begin{pmatrix} w_n & w_n t_n & w_n t_n^2 \\ w_{n+1} & w_{n+1} t_{n+1} & w_{n+1} t_{n+1}^2 \\ & \vdots & \\ w_m & w_m t_m & w_m t_m^2 \end{pmatrix} \text{ and } b = \begin{pmatrix} w_n y_n \\ w_{n+1} y_{n+1} \\ \vdots \\ w_m y_m \end{pmatrix} \tag{8.9}$$

The filtered value is set to the value of the polynomial at point i. Weights are designated as w in the above expression, with weights assigned to all of the data values in the window. Data values that were flagged in the preprocessing are assigned weight ''zero'' in this application and thus do not influence the result. The clean data values all have weights ''one''. Residual negatively biased noise (e.g., clouds) may occur for the remote sensing data and accordingly the fitting was performed in two steps [6]. The first fit was conducted using weights obtained from the preprocessing. Data points above the resulting smoothed function from the first fit are regarded more important, and in the second step the normal equations are solved using the weight of these data values, but increased by a factor 2. This multi-step procedure leads to a smoothed function that is adapted to the upper envelope of the data (Figure 8.7). Similarly, the ancillary metadata of the meteorological data from the preprocessing also were used in the iterative fitting to the upper envelope of the –KBDI time-series [6].

The width of the fitting window determines the degree of smoothing, but it also affects the ability to follow a rapid change. It is sometimes necessary to locally tighten the window even when the global setting of the window performs well. A typical situation occurs in savanna ecosystems where vegetation, associated remote sensing, and meteorological indices respond rapidly to vegetation moisture dynamics. A small fitting window can be used to capture the corresponding sudden rise in data values.

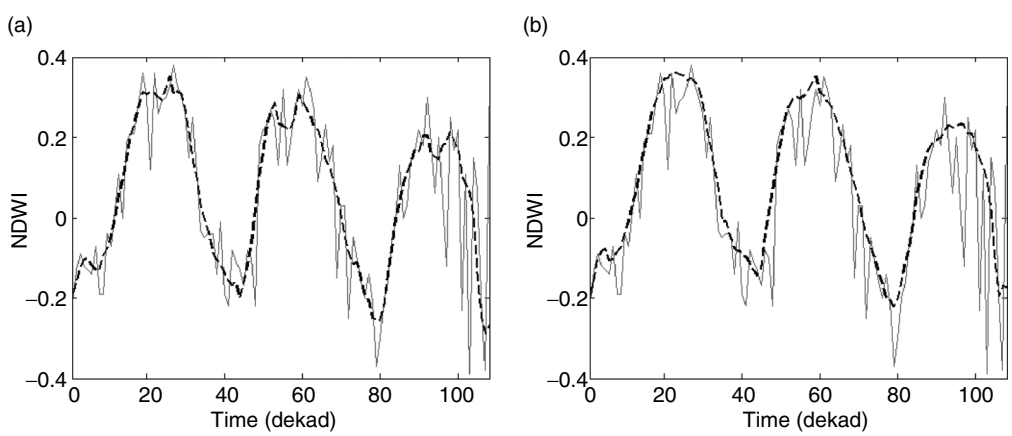

FIGURE 8.7
The Savitzky–Golay filtering of NDWI (——) is performed in two steps. Firstly, the local polynomials are fitted using the weights from the preprocessing (a). Data points above the resulting smoothed function (– – –) from the first fit are attributed a greater importance. Secondly, the normal equations are solved with the weights of these data values increased by a factor 2 (b).

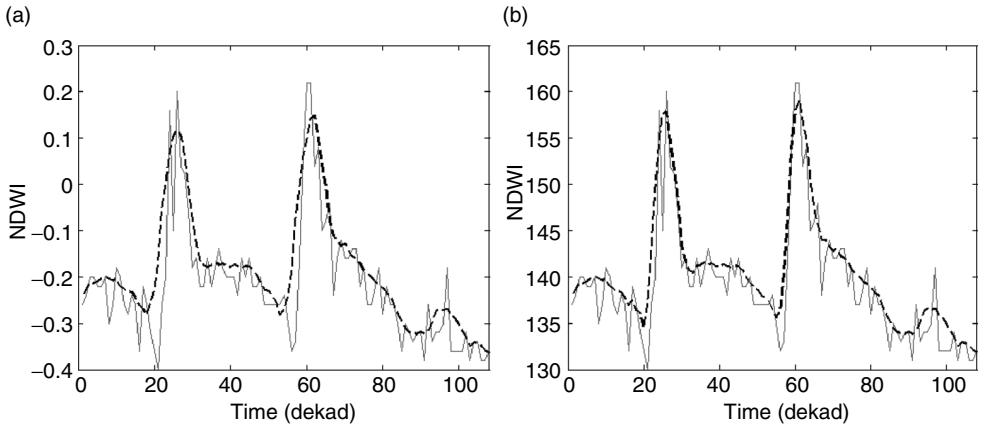

FIGURE 8.8
The filtering of NDWI (——) in (a) is done with a window that is too large to allow the filtered data (– – –) to follow sudden increases and decreases of underlying data values. The data in the window are scanned and if there is a large increase or decrease, an automatic decrease in the window size will result. The filtering is then repeated using the new locally adapted size (b). Note the improved fit at rising edges and at narrow peaks.

The data in the window are scanned and if a large increase or decrease is observed, the adaptive Savitzky–Golay method applied an automatic decrease in the window size. The filtering is then repeated using the new locally adapted size. Savitzky–Golay filtering with and without the adaptive procedure is illustrated in Figure 8.8. In the figure it is shown that the adaptation of the window improves the fit at the rising edges and at narrow seasonal peaks.

8.4.2 Extracting Seasonal Metrics from Time-Series and Statistical Analysis

Four seasonal metrics were extracted for each of the rainy seasons. Figure 8.9 illustrates the different metrics per season for NDWI and KBDI time-series. The beginning of a season, that is, 20% left of the rainy season, is defined from the final function fit as the point in time for which the index value has increased by 20% of the distance between the left minimum level and the maximum. The end of the season is defined in a similar way as the point 20% right of the rain season. The 80% left and right points are defined as the points for which the function fit has increased to 80% of the distance between, respectively, the left and right minimum levels and the maximum. The current technique used to define metrics also is used by Verbesselt et al. [6] to define the beginning of the fire season.

The temporal occurrence and the value of each metric were extracted for further exploratory statistical analysis to study the relationship between time-series. The SPOT VGT S10 time-series consisted of four seasons (1998–2002) from which four occurrences and values per metric type were extracted. Twenty-four occurrence–value combinations per metric type ultimately were available for further analysis since six weather stations were used.

Serial correlation that occurs in remote sensing and climate-based time-series invalidates inferences made by standard parametric tests, such as the Student's *t*-test or the Pearson correlation. All extracted occurrence–value combinations per metric type were tested for autocorrelation using the Ljung–Box autocorrelation test [26]. Robust nonparametric techniques, such as the Wilcoxon's signed rank test were used in case of non-normally distributed data. The normality of the data was verified using the Shapiro–Wilkinson normality test [29].

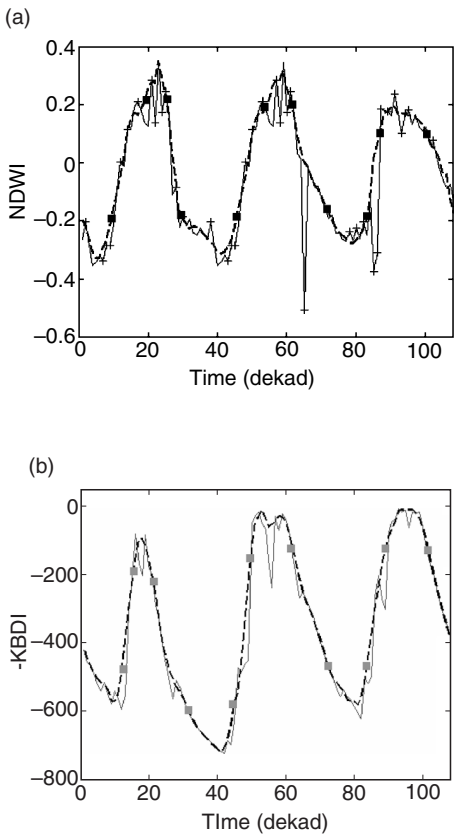

FIGURE 8.9
The final fit of the Savitzky–Golay function (– – –) to the NDWI (a) and –KBDI (b) series (——), with the four defined metrics, that is, 20% left and right, and 80% left and right (■), overlaid on the graph. Points with flagged data errors (+) were assigned weights of zero and did not influence the fit. A dekad is defined as a 10-day period.

Firstly, the distribution of the temporal occurrence of each metric was visualized and evaluated based on whether or not there was a significant difference between the temporal occurrence of the four metric types extracted from –KBDI and NDWI time-series. Next, the strength and significance of the relationship between –KBDI and NDWI values of the four metrics types were assessed with an OLS regression analysis.

8.5 Results and Discussion

Figure 8.9 illustrates the optimized function fit and the defined metrics for the –KBDI and NDWI. Notice that the Savitzky–Golay function could properly define the behavior of the different time-series. The function was fitted to the upper envelope of the data by using the uncertainty information derived during the preprocessing step. The results of the statistical analysis based on the extracted metrics for –KBDI and NDWI are presented. The Ljung–Box statistic indicated that the extracted occurrences and values were not significantly autocorrelated at a 95% confidence level. All p-values were greater than 0.1, failing to reject the null hypothesis of independence.

8.5.1 Temporal Analysis of the Seasonal Metrics

Figure 8.10 illustrates the temporal distribution of temporal occurrence from extracted metrics from time-series of –KBDI and NDWI. The occurrences of extracted metrics

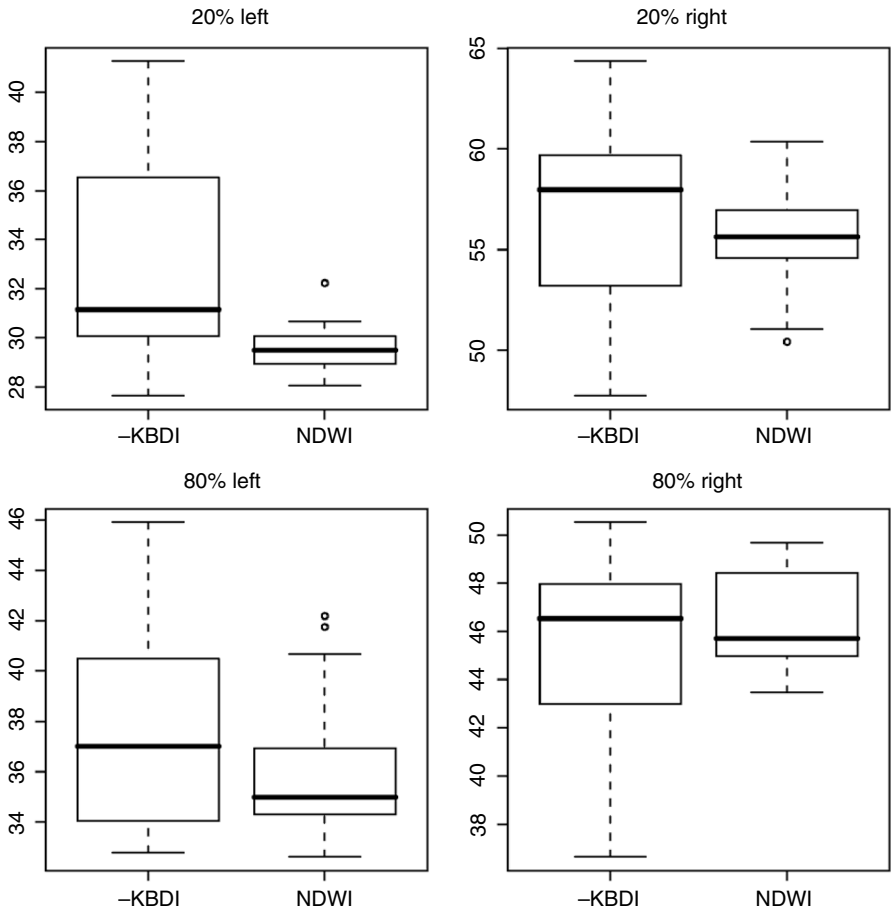

FIGURE 8.10
Box plots of the temporal occurrence of the four defined metrics, that is, 20% left and right and 80% left and right, extracted from time-series of –KBDI and NDWI. The dekads (10-day period) are shown on the *y*-axis and are indicative of the temporal occurrence of the metric. The upper and lower boundaries of the boxes indicate upper and lower quartiles. The median is indicated by the solid line (—) within each box. The whiskers connect the extremes of the data, which were defined as 1.5 times the inter-quartile range. Outliers are represented by (o).

were significantly non-normally distributed at a 95% confidence level ($p > 0.1$), indicating that the Wilcoxon's signed rank can be used. The Wilcoxon's signed rank test showed that –KBDI and NDWI occurrences of the 80% left and right, and 20% right were not significantly different from each other at a 95% confidence level ($p > 0.1$). This confirmed that –KBDI and NDWI were temporally related. It also corroborated the results of Burgan et al. and Ceccato et al. [4,11] who found that –KBDI and NDWI both were related to the seasonal vegetation moisture dynamics, as measured by VWC.

Figure 8.10, however, illustrates that the start of the rainy season (i.e., 20% left occurrence), derived from the –KBDI and NDWI time-series, was different. The Wilcoxon's signed rank test confirmed that the –KBDI and NDWI differed significantly from each other at a 95% confidence level ($p < 0.01$). This phenomenon can be explained by the fact that vegetation in the study area starts growing before the rainy season starts, due to an early change in air temperature (N. Govender, Scientific Service Kruger National Park,

South Africa, Personal Communication). This explained why the NDWI reacted before the change in climatic conditions as measured by the –KBDI, given that the NDWI is directly related to vegetation moisture dynamics [4].

8.5.2 Regression Analysis Based on Values of Extracted Seasonal Metrics

The assumptions of the OLS regression models between values of metrics extracted from –KBDI and NDWI time-series were verified. The Wald test statistic showed nonlinearity to be not significant at a 95% confidence level ($p > 0.15$). The Shapiro–Wilkinson normality test confirmed that the residuals were normally distributed at a 95% confidence level ($p < 0.01$) [6,29]. Table 8.2 illustrates the results of the OLS regression analysis between values of metrics extracted from –KBDI and NDWI time-series. The values extracted at the ''20% right'' position of the –KBDI and NDWI time-series showed a significant relationship at a 95% confidence level (p-values $<$ 0.01). The other extracted metrics did not exhibit significant relationships at a 95% confidence level (p-values > 0.1). A significant relationship between the –KBDI and NDWI time-series was observed only at the moment when savanna vegetation was completely cured (i.e., 20% right-hand side). The savanna vegetation therefore reacted differently to changes in climate parameters such as rainfall and temperature, as measured by KBDI, depending on the phenological growing cycle. This phenomenon could be explained because a living plant uses defense mechanisms to protect itself from drying out, while a cured plant responds to climatic conditions[46]. These results consequently indicated that the relationship between extracted values of –KBDI and NDWI was influenced by seasonality. This is in corroboration with the results of Ji et al. [23], who indicated that seasonality had a significant effect on the relationship between vegetation as measured by a remote sensing index and drought index. These results further illustrated that the seasonal effect needs to be taken into account when regression techniques are used to quantify the relationship between time-series related to vegetation moisture dynamics. The seasonal effect also can be accounted for by utilizing autoregression models with seasonal dummy variables, which take the effect of serial correlation and seasonality into account [23,26]. However, the proposed method to account for serial correlation by sampling at specific moments in time had an additional advantage; the influence of seasonality could be studied by extracting metrics at the specified moments, besides the fact that serial correlation was taken into account.

Furthermore, it was shown that serial correlation caused an overestimation of the correlation coefficients when results from Table 8.1 and Table 8.2 were compared. The

TABLE 8.2

Coefficients of Determination of the OLS Regression Models (NDWI $\sim$ –KBDI) for the Four Extracted Seasonal Metric Values Between –KBDI and NDWI Time-Series of ($n = 24$ per Metric)

NDWI $\sim$ KBDI	R^2	*p-Values*
20% left	0.01	0.66
20% right	0.49	<0.01
80% left	0.00	0.97
80% right	0.01	0.61

coefficient of determination (R^2) of Table 8.1 all were significant with an average value of 0.7, while only the correlation coefficient at the end of the rainy season (20% right-hand side) was significant ($R^2 = 0.49$). This confirmed the importance of accounting for serial correlation and seasonality in the residuals of a regression model, when studying the relationship between two time-series.

8.5.3 Time-Series Analysis Techniques

Time-series analysis models most often are used for purposes of describing current conditions and forecasting [25]. The models use the serial correlation in time-series as a tool to relate temporal observations. Future observations can be predicted by modeling the serial correlation structure of a time-series [26]. Time-series analysis techniques (e.g., ARIMA) can be used to model a time-series by using other, independent time-series [27]. ARIMA subsequently can be used to study the relationship between –KBDI and NDWI. However, there are constraints that have to be considered before ARIMA models can be applied to a time-series (e.g., drought index) by using other predictor time-series (e.g., satellite index). AQ2

Firstly, the goodness-of-fit of an ARIMA model will not be significant when changes in the satellite index precede or coincide with those in the drought index. The CCF can be used in this context to verify how time-series are related to each other. Time lag results in Table 8.1 indicate that the –KBDI (drought index) precedes or coincides with the NDWI time-series. This illustrates that ARIMA models cannot directly be used to predict the KBDI, with NDWI as predictor variable. Consequently, other more advanced time-series analysis techniques are needed to model vegetation dynamics because they will precede or coincide with the dynamics monitored by remote sensing indices in most of the cases. Such more advanced time-series analysis techniques, however, are not discussed since they are outside the scope of this chapter.

Secondly, availability of data is limited for time-series analysis, namely, from 1998 to 2002. This is an important constraint because two separate data sets are needed to parameterize and evaluate an ARIMA model. One set is needed for parameterization, while the other is used to forecast and validate the ARIMA model through comparison of the observed and expected values. Accordingly, it is necessary to interpolate missing satellite data that were masked out during preprocessing to ensure adequate data are available for parameterization.

Thirdly, the proposed sampling strategy made investigation of the time lag and correlation at a defined instant in time possible, as opposed to ARIMA or cross-correlation analysis, through which only the overall relationship between time-series can be studied [25]. The applied sampling strategy is thus ideally suited to study the relationship between time-series of climate and remote sensing data, characterized by seasonality and serial correlation. The sampling of seasonal metrics minimized the influence of serial correlation, thereby making the study of seasonality possible.

8.6 Conclusions

Serial correlation problems are not unknown in the field of statistical or general meteorology. However, the presence of serial correlation, found during analysis of a variable sampled sequentially at regular time intervals, seems to be disregarded by

many agricultural meteorologists and remote sensing scientists. This is true despite abundant documentation available in the traditional meteorological and statistical literature. Therefore, an overview of the most important time-series analysis techniques and concepts was presented, namely, stationarity, autocorrelation, differencing, decomposition, autoregression, and ARIMA.

A method was proposed to study the relationship between a meteorological drought index (KBDI) and remote sensing index (NDWI), both related to vegetation moisture dynamics, by accounting for the serial correlation effect. The relationship between –KBDI and NDWI was studied by extracting nonserially correlated seasonal metrics, for example, 20% and 80% left- and right-hand side metrics of the rainy season, based on a Savitzky–Golay fit to the upper envelope of the time-series. Serial correlation between the extracted metrics was shown to be minimal and seasonality was an important factor influencing the relationship between NDWI and –KBDI time-series. Statistical analysis using the temporal occurrence of the extracted metrics revealed that NDWI and –KBDI time-series are temporally connected, except at the beginning of the rainy season. The fact that the savanna vegetation starts re-greening before the start of the rainy season explains this inability to detect the beginning of the rainy season. The values of the extracted seasonal metrics of NDWI and –KBDI were significantly related only at the end of the rainy season, namely, at the 20% right-hand side value of the fitted curve. The savanna vegetation at the end of the rainy season was cured and responded strongly to changes in climatic conditions monitored by the –KBDI, such as rain and temperature. The relationship between –KBDI and NDWI consequently changes during the season, which indicates that seasonality is an important factor that needs to be taken into account. Moreover, it was shown that correlation coefficients estimated by OLS regression analysis were overestimated due to the influence of serial correlation in the residuals. This confirmed the importance of taking serial correlation of the residuals into account by sampling nonserially correlated seasonal metrics when studying the relationship between time-series.

The serial correlation effect consequently was taken into account by the extraction of seasonal metrics from time-series. The seasonal metrics in turn could be used to study the relationship between remote sensing and ground-based time-series, such as meteorological or field measurements. A better understanding of the relationship between remote sensing and *in situ* observations at regular time intervals will contribute to the use of remotely sensed data for the development of an index that represents seasonal vegetation moisture dynamics.

Acknowledgment

The SPOT VGT S10 data sets were generated by the Flemish Institute for Technological Development. The climate data were provided by the Weather Services of South Africa, whereas the National Land Cover Map (1995) was supplied by the Agricultural Research Centre of South Africa. We acknowledge the support of L. Eklundh, as well as the Crafoord and Bergvall foundations. The authors wish to thank N. Govender from the Scientific Services of Kruger National Park for scientific input.

References

1. Camia, A. et al., Meteorological fire danger indices and remote sensing, in *Remote Sensing of Large Wildfires in the European Mediterranean Basin*, E. Chuvieco, Ed., Springer-Verlag, New York, 1999, p. 39.
2. Ceccato, P. et al., Estimation of live fuel moisture content, in *Wildland Fire Danger Estimation and Mapping: The Role of Remote Sensing Data*, E. Chuvieco, Ed., World Scientific Publishing, New York, 2003, p. 63.
3. Dennison, P.E. et al., Modeling seasonal changes in live fuel moisture and equivalent water thickness using a cumulative water balance index, *Remote Sens. Environ.*, 88, 442, 2003.
4. Ceccato, P. et al., Detecting vegetation leaf water content using reflectance in the optical domain, *Remote Sens. Environ.*, 77, 22, 2001.
5. Jackson, T.J. et al., Vegetation water content mapping using landsat data derived normalized difference water index for corn and soybeans, *Remote Sens. Environ.*, 92, 475, 2004.
6. Verbesselt, J. et al., Evaluating satellite and climate data derived indices as fire risk indicators in savanna ecosystems, *IEEE Trans. Geosci. Remote Sens.*, In Press, 2006. AQ3
7. Van Wilgen, B.W. et al., Response of Savanna fire regimes to changing fire-management policies in a large African national park, *Conserv. Biol.*, 18, 1533, 2004.
8. Dimitrakopoulos, A.P. and Bemmerzouk, A.M., Predicting live herbaceous moisture content from a seasonal drought index, *Int. J. of Biometeorol.*, 47, 73, 2003.
9. Keetch, J.J. and Byram G.M., A drought index for forest fire control, *U.S.D.A. Forest Service, Asheville NC.*, SE-38, 1988.
10. Janis, M.J., Johnson, M.B., and Forthun, G., Near-real time mapping of Keetch–Byram drought index in the south-eastern United States, *Int. J. Wildland Fire*, 11, 281, 2002.
11. Burgan, E.R., Correlation of plant moisture in Hawaii with the Keetch–Byram drought index, *USDA. Forest Service Research Note*, PSW-307, 1976.
12. Chuvieco, E. et al., Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating, *Remote Sens. Environ.*, 92, 322, 2004.
13. Maki, M., Ishiahra, M., and Tamura, M., Estimation of leaf water status to monitor the risk of forest fires by using remotely sensed data, *Remote Sens. Environ.*, 90, 441, 2004.
14. Paltridge, G.W. and Barber, J., Monitoring grassland dryness and fire potential in Australia with NOAA AVHRR data, *Remote Sens. Environ.*, 25, 381, 1988.
15. Chladil, M.A. and Nunez, M., Assessing grassland moisture and biomass in Tasmania—The application of remote-sensing and empirical-models for a cloudy environment, *Int. J. Wildland Fire*, 5, 165, 1995.
16. Hardy, C.C. and Burgan, R.E., Evaluation of NDVI for monitoring live moisture in three vegetation types of the western US, *Photogramm. Eng. Remote Sens.*, 65, 603, 1999.
17. Chuvieco, E. et al., Estimation of fuel moisture content from multitemporal analysis of LANDSAT thematic mapper reflectance data: Applications in fire danger assessment, *Int. J. Remote Sens.*, 23, 2145, 2002.
18. Fensholt, R. and Sandholt, I., Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment, *Remote Sens. Environ.*, 87, 111, 2003. AQ4
19. Hunt, E.R., Rock, B.N., and Nobel, P.S., Measurement of leaf relative water content by infrared reflectance, *Remote Sens. Environ.*, 22, 429, 1987.
20. Ceccato, P., Flasse, S., and Gregoire, J.M., Designing a spectral index to estimate vegetation water content from remote sensing data—Part 2: Validation and applications, *Remote Sens. Environ.*, 82, 198, 2002.
21. Myneni, R.B. et al., Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, 386, 698, 1997.
22. Eklundh, L., Estimating relations between AVHRR NDVI and rainfall in East Africa at 10-day and monthly time scales, *Int. J. Remote Sens.*, 19, 563, 1998.
23. Ji, L. and Peters, A.J., Assessing vegetation response to drought in the northern great plains using vegetation and drought indices, *Remote Sens. Environ.*, 87, 85, 2003.

24. De Beurs, K.M. and Henebry, G.M., Land surface phenology, climatic variation, and institutional change: Analyzing agricultural land cover change in Kazakhstan, *Remote Sens. Environ.*, 89, 497, 2004.
25. Meek, D.W. et al., A note on recognizing autocorrelation and using autoregression, *Agricult. Forest Meterol.*, 96, 9, 1999.
26. Brockwell P.J. and Davis R.A., *Introduction to Time Series and Forecasting*, 2nd ed., Springer-Verlag, New York, 2002, p. 31.
27. Ford, C.R. et al., Modeling canopy transpiration using time series analysis: A case study illustrating the effect of soil moisture deficit on Pinus Taeda, *Agricult. Forest Meterol.*, 130, 163, 2005.
28. Montanari, A., Deseasonalisation of hydrological time series through the normal quantile transform, *J. Hydrol.*, 313, 274, 2005.
29. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org, 2005.
30. Rahman, H. and Dedieu, G., SMAC—A simplified method for the atmospheric correction of satellite measurements in the solar spectrum, *Int. J. Remote Sens.*, 15, 123, 1994. AQ5
31. Stroppiana, D. et al., An algorithm for mapping burnt areas in Australia using SPOT-VEGETATION data, *IEEE Trans. Geosci. Remote Sens.*, 41, 907, 2003.
32. Thompson, M.A., standard land-cover classification scheme for remote-sensing applications in South Africa, *South Afr. J. Sci.*, 92, 34, 1996.
33. Aguado, I. et al., Assessment of forest fire danger conditions in southern Spain from NOAA images and meteorological indices, *Int. J. Remote Sens.*, 24, 1653, 2003.
34. von Storch, H. and Zwiers, F.W., *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 1999, p. 483.
35. Thejll, P. and Schmith, T., Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies, *J. Geophys. Res. Atmos.*, 110, 2005. AQ6
36. Cleveland, R.B. et al., STL: A seasonal-trend decomposition procedure based on Loess., *J. Off. Stat.*, 6, 3, 1990.
37. Venables, W.N. and Ripley, B.D., *Modern Applied Statistics with S*, 4th ed., Springer-Verlag, New York, 2003, p. 493.
38. Jönsson, P. and Eklundh, L., Seasonality extraction by function fitting to time-series of satellite sensor data, *IEEE Trans. Geosci. Remote Sens.*, 40, 1824, 2002.
39. Jönsson, P. and Eklundh, L., TIMESAT—a program for analyzing time-series of satellite sensor data, *Comput. Geosci.*, 30, 833, 2004.
40. Menenti, M. et al., Mapping agroecological zones and time-lag in vegetation growth by means of Fourier-analysis of time-series of NDVI images, *Adv. Space Res.*, 13, 233, 1993.
41. Azzali, S. and Menenti, M., Mapping vegetation–soil–climate complexes in Southern Africa using temporal Fourier analysis of NOAA-AVHRR NDVI data, *Int. J. Remote Sens.*, 21, 973, 2000.
42. Olsson, L. and Eklundh, L., Fourier-series for analysis of temporal sequences of satellite sensor imagery, *Int. J. Remote Sens.*, 15, 3735, 1994.
43. Cihlar, J., Identification of contaminated pixels in AVHRR composite images for studies of land biosphere, *Remote Sens. Environ.*, 56, 149, 1996.
44. Sellers, P.J. et al., A global 1-degrees-by-1-degrees NDVI data set for climate studies. The generation of global fields of terrestrial biophysical parameters from the NDVI, *Int. J. Remote Sens.*, 15, 3519, 1994.
45. Roerink, G.J., Menenti, M., and Verhoef, W., Reconstructing cloudfree NDVI composites using Fourier analysis of time series, *Int. J. Remote Sens.*, 21, 1911, 2000.
46. Pyne S.J., Andrews P.L., and Laven R.D., *Introduction to Wildland Fire*, 2nd ed., John Wiley & Sons, New York, 1996, 117.

Author Queries

[AQ1] Please advise whether this change is okay.
[AQ2] Please approve this edit.
[AQ3] Please update this reference.
[AQ4] Please cite reference number 18 in the text.
[AQ5] Please cite reference number 30 in the text.
[AQ6] Please provide volume number/page numbers.