NDVI TIME SERIES AND MARKOV CHAINS TO MODEL THE CHANGE OF FUZZY VEGETATIVE DROUGHT CLASSES

S. Ding, C. M. Rulinda, A. Stein, W. Bijker

University of Twente
Department of Earth Observation Science, Faculty of Geo-information Science and Earth Observation
Hengelosestraat 99, 7500 AE, Enschede, The Netherland

ABSTRACT

The objective of this study is to explore the potential of using Markov chains to model the changes of vegetative drought classes. NOAA-AVHRR dekadal NDVI images and fuzzy functions are used to characterize the drought classes while capturing the gradual transition between them. The transition probabilities are estimated using the maximum class membership values at a location. The Markov transition probability matrix is then used to model the changes of vegetative drought classes at selected locations. Future vegetative drought classes are predicted using the estimated transition matrix, then compared with actual data. Twenty pixel locations clustered in four regions of the two main agricultural type in Kenya are selected to implement this approach. Half of the pixels are predicted correctly. 5 of them are predicted either one class higher or lower and 2 of them, two classes higher. We can conclude that Markov chains applied to fuzzy numbers have the potential to model the changes of of vegetative drought classes at a pixel, hence provide a benefit for early warning systems.

***Index Terms*—** Remote Sensing, Drought, East Africa, fuzzy Sets theory.

1. INTRODUCTION

Vegetative drought, i.e. vegetation stress caused by drought, can be detected from anomalies of the Normalized Difference Vegetation Index (NDVI). NOAA-AVHRR derived NDVI data [1], available for almost 30 years cover large geographical areas, can be used to create a time series that can be used to model vegetative drought in space and time as well as to predict the future classes at individual pixel-locations. To do so, we use Markov chains to model the changes of vegetative drought classes. In the past, Markov chains have been used in remote sensing studies for meteorological drought prediction studies [2][3] and for vegetation dynamics studies [4]. Vegetative drought characterisation using NDVI anomaly, however, is uncertain. Therefore we develop a fuzzy modeling approach to define drought classes.

In this study [5] we combine Markov Chains with fuzzy sets theory to predict the future states of vegetative drought. The method is applied on 20 pixel-locations clustered in four different study areas within Kenya [5]: Region 1, centered at **P1** (3.117N, 35.617E); Region 2, centered at **P2** (1.75N, 40.067E); Region 3, centered at **P3** (0.5S, 37.45E), and Region 4, centered at **P1** (3.4S, 38.567E) respectively, as shown in **Fig.1**. The agriculture type of Regions 1 and 2 is pastoral farming, while in Regions 3 and 4 it is cultivation. NDVI data from the years 2004 to 2008, as well as the long term mean calculated from 1982 to 2005 are used.

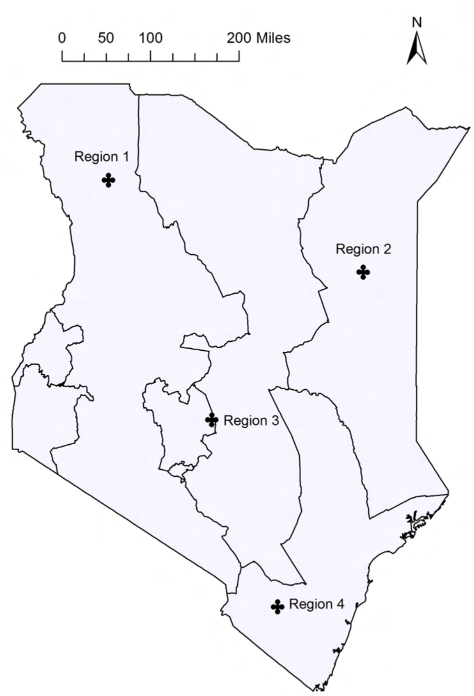

Fig. 1. Map of Kenya with the twenty pixel locations clustered in four regions
(Source from:
http://www.maplibrary.org/stacks/Africa/Kenya/index.php)

 978-1-4577-1203-6/11/$26.00 ©2011 IEEE

2. METHODS

2.1. Correlation between NDVI and RFE

To validate NDVI anomaly as an indicator of vegetative drought, a Pearson (r) correlation test is performed with different time lag between the mean NDVI values and the mean Rainfall Estimates (RFE) calculated between 1995 and 2008. AVHRR derived NDVI images are used to provide data for this indicator, called NDVI$_a$ and calculated by equation (1).

$$NDVI_a = NDVI_i - NDVI_{mean} \quad (1)$$

where $NDVI_i$ is the current $NDVI$ and $NDVI_{mean}$ the long term mean for the same period

2.2. Fuzzy classification

The number of drought classes is decided following the SPI based drought classification method, and the categories of these six classes are: extremely drought, severely drought, moderately drought, dry, wet and moderately wet. In the following context, these six classes are numbered from 1 to 6 sequentially. In **Fig.2**, general membership functions of each drought classes is shown. Range $[a_i, b_i]$, $i \in N$, with N being the total number of classes, is the core zone for each membership function of each drought class, where i represents the index of drought classes. Range $[b_{i-1}, a_i]$ is the transition zone. As there is no standard vegetative drought classification method based on $NDVI_a$ values, in this study, the classification is based on the *same frequency* in the whole dataset. When building the transition matrix for Markov chains, equal frequencies can ensure that the number of observations per class is sufficiently high to obtain reasonable estimates [6]. $[a_i, b_i]$ represents the 95% confidence interval for each original class, which is defined as the *core*, and the rest 5% are assigned to the linear transition zone.

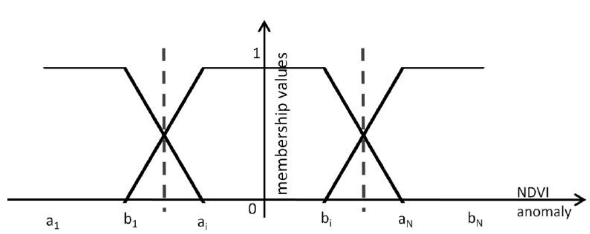

Fig. 2. Membership functions for drought classes, the range $[a_1, b_1]$ is the core zone of the first drought class (*i.e.* the driest), $[a_i, b_i]$, $i \in \{2, 3, \ldots, N-1\}$ is the core zone of middle drought classes, $[a_N, b_N]$ is the core zone of last drought class (*i.e.* the wettest). The dash lines show the classes' boundaries of a crisp classification.

2.3. Markov Chains

A transition matrix is built using transition probabilities between fuzzy classes, calculated as in equation (2). The maximum membership values of all classes for each pixel at each moment of time are used.

$$\mathbf{P} = [p_{ij}] = P(X_t = j | X_{t-1} = i) \quad (2)$$

where $p_{ij} = \frac{n_{ij}}{\sum n_{ij}} \geq 0$, $\sum n_{ij} = 1$ and p_{ij} is the element in the probability transition matrix $\mathbf{P}$, where the pixel is at the current time t in the fuzzy class j, given that it was at the past time $t-1$ in the fuzzy class i. n_{ij} is the number of times a pixel changes from being in class i to class j from time $t-1$ to time t.

2.4. Tests of Markovian properties

The reliability of predicting vegetative drought classes using Markov chains generally depends on two conditions. First, the data-generating process must meet the Markovian property: time homogeneity and time independence. The test statistics can be estimated using the Maximum likelihood ratio (LR) criteria and Pearson χ^2-tests (Q) under specific null and alternative hypotheses. Although the LR and Pearson χ^2 statistics are asymptotically equivalent, in cases of poor asymptotic, they are not equivalent [6]. Therefore, both of the statistics are tested. Second, the estimates have to be based on a number of observations, large enough to be able to rely on the asymptotic properties of the estimators. Otherwise, the accuracy will be rather poor [5].

2.4.1. Test of time homogeneity

The test of time homogeneity, i.e. time stationarity, is conducted to decide whether the transition probabilities are constant over time. The test is done by dividing the entire sample period into M sub-periods and comparing each transition matrix of sub-periods with the one estimated from the entire sample. The LR and Pearson χ^2-tests are shown in equation (3) and (4). The null hypothesis $H_0 : \forall m : p_{ij|m} = p_{ij} (m = 1, 2, \ldots, M)$ and alternative hypothesis $H_a : \exists m : p_{ij|m} \neq p_{ij}$, $\alpha = 0.05$, where $p_{ij} = \frac{n_{ij|m}}{n_{i|m}}$,$n_{ij|m} = \sum_{t \in m} n_{ij|m}(t)$,$n_{i|m} = \sum_{t \in m} n_{i|m}(t-1)$

$$LR^{(M)} = 2 \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j} n_{ij|m}(t) \ln \frac{p_{ij|m}}{p_j} \quad (3)$$

$$Q^{(M)} = \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j} n_{i|m} \frac{(p_{ij|m} - p_{ij})^2}{p_{ij}} \quad (4)$$

2.4.2. Test of time independence

To test the time independent is required to test that the Markov chains are at least of order 1. A test of 0-order Markov chain versus 1-order Markov chain is conducted for this purpose. The LR and Pearson χ^2-tests as shown in equation (5) and (6). The null hypothesis $H_0 : \forall i : p_{ij} = p_j$ and alternative hypothesis $H_a : \exists i : p_{ij} \neq p_j$, $\alpha = 0.05$, where $p_j = n_j/n$

$$LR^{(0)} = 2\sum_{i=1}^{N}\sum_{j} n_{ij}(t)\ln\frac{p_{ij}}{p_j} \tag{5}$$

$$Q^{(0)} = \sum_{i=1}^{N}\sum_{j=1}^{N} n_i(t-1)\frac{(p_{ij}-p_j)^2}{p_j} \tag{6}$$

where $n_i(t-1) = \sum n_{ij}(t)$

2.5. Prediction

The predicted probabilities of vegetative drought classes are estimated as in equation (7). The predicted probabilities μ of the pixel x to the classes at time $t+1$ are obtained as a row vector from equation (2), by multiplying the transition matrix **P** with the membership value at time t.

$$\mu(x)_{t+1} = f(x)_t * \mathbf{P} \tag{7}$$

where $f(x)_t$ is a vector with the membership values of pixel x to the different classes, at time t.

For validation, the predicted values are compared with the actual (reference) data consisting of membership values at time $t+1$.

3. RESULTS AND DISCUSSION

Nineteen out of the twenty pixels show a positive correlation between the long-term mean NDVI and the long-term mean RFE with a minimum lag time of 60 days. Except for the five pixels in Region 1, where the highest $r = 0.4$ (with $\alpha = 0.001$), there is a strong correlation ($r = 0.9$, $\alpha = 0.001$) between the NDVI and RFE data. Region 1 is arid with a lower vegetation density than the other regions. This can explain the lower correlation between the NDVI and RFE in this region, compared to the others. As NDVI is positively correlated to rainfall in most of the regions, NDVI anomaly can indicate a lack of rainfall. Hence, they can be used as an indicator of vegetative drought.

The six classes are produced with an equal frequency classification. This method results in a series of classes from Class 1 with a core zone in range [min, -0.072], Class 2 with a core in range [-0.060, -0.036], Class 3: [-0.024, -0.016], Class 4: [-0.004, 0], Class 5: [0.008, 0.012], Class 6: [0.024, max], where min and max represent the minimum (-0.376) and maximum (0.432) values of the full NDVI anomaly dataset. **Fig.3** and **Fig.4** show respectively the drought classes of pixels **P1** in Region 1 and **P2** in Region 2, **P3** in Region 3 and **P4** in Region 4 from September 2005 until March 2006, based on the maximum class membership values. **P2** remains mostly dry from mid-October 2005 (dekad 5) through the end of March 2006, whereas **P3** and **P4** are dryer from the end of December 2005 (dekad 12 and 13), throughout February 2006 for **P3**, and March for **P4**. **P1** is mostly not dry.

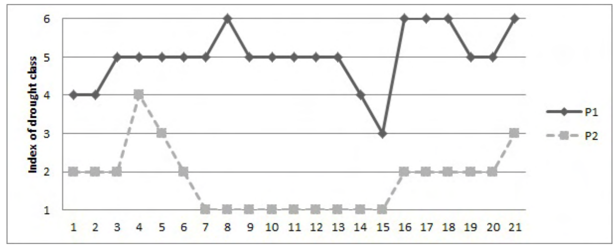

Fig. 3. Extreme drought (1) to wet (6) classes values of **P1** and **P2**, in the pastoral areas. Numbers on the X-axis represent consecutive dekadals from September 2005 through March 2006

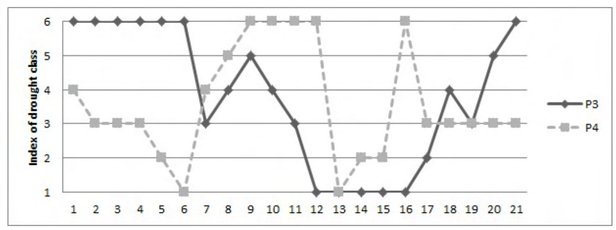

Fig. 4. Extreme drought (1) to wet (6) classes values of **P3** and **P4**, in the cultivated areas. Numbers on the X-axis represent consecutive dekadals from September 2005 through March 2006

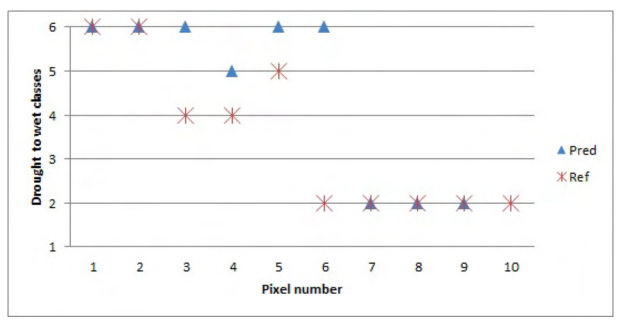

Fig. 5. Comparison for the 10 pixel-locations in Region 1 and Region 2 between the predicted (Pred) and the reference (Ref) classes for the second dekadal of January 2009

To test for time homgeneity, two sub-samples M are used. In the first step, the sample is divided yearly: $M = 5$. Results ($LRprob = 0.157$, and $Qprob = 0.267$) show that the transition probabilities are stationary yearly during the five years. In the second step, the sample is divided monthly: $M = 12$. Results, after excluding the months of February and September from the data sample ($LRprob = 0.517$ and

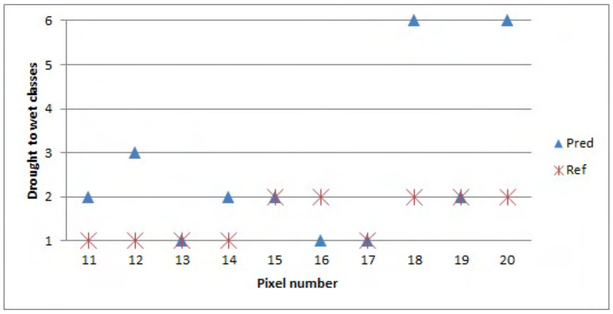

Fig. 6. Comparison for the 10 pixel-locations in Region 3 and Region 4 between the predicted (Pred) and the reference (Ref) classes for the second dekadal of January 2009

$Qprob = 0.978$) show that transition probabilities are stationary during the ten months.

Results of the tests of time independence of order 0 against order 1 (with $LRprob$ and $Qprob$ ¡ 0.05) rejected the order 0. Hence, the dynamics of vegetative drought can be modeled as a first order process, *i.e.* a process depending on the most recent past. The Markov transition probability matrix as in equation (8) model the dynamics of vegetative drought classes, and show that there are higher probabilities to remain to an existing class.

$$\mathbf{P} = \begin{bmatrix} \mathbf{0.596} & 0.211 & 0.059 & 0.027 & 0.033 & 0.074 \\ 0.190 & \mathbf{0.434} & 0.175 & 0.066 & 0.020 & 0.115 \\ 0.108 & 0.194 & \mathbf{0.324} & 0.144 & 0.072 & 0.158 \\ 0.021 & 0.067 & 0.163 & \mathbf{0.351} & 0.247 & 0.151 \\ 0.016 & 0.025 & 0.061 & 0.141 & \mathbf{0.531} & 0.227 \\ 0.018 & 0.036 & 0.032 & 0.021 & 0.099 & \mathbf{0.793} \end{bmatrix} \quad (8)$$

For ten out of twenty pixels, and considering the highest probabilities, the drought classes for the second dekadal of 2009 are predicted correctly (see **Fig.5** and **Fig.6**), using the membership values at the first decadal of January as input data.

A difference between the predicted and actual values of more than one class is observed for five pixels, whereas for the five other classes, the difference is equal one class. Pixel 16 has an actual membership value in class 1 of 0.444 and in class 2 of 0.556, in the first decadal of January 2009; it is predicted to be in class 1 with a probability of 0.38, and class in 2 with a probability of 0.32 at the next time step. Predicted results show that Pixel 16 has a membership value of 0.49 to class 1 and 0.51 to class 2, which means a difference of one class between the predicted and reference data. The positive membership values of Pixel 16 in both classes, and the small difference between the predicted probability values can explain the difference between the predicted and the reference data. A crisp classification is limited in the sense that it cannot capture this nuance.

4. CONCLUSION AND RECOMMENDATIONS

The Markov chains approach applied to membership values shows a potential to model the changes between vegetative drought classes and to the predict vegetative drought classes at time $t + 1$. The use of fuzzy functions to model vegetative drought classes from NDVI anomaly provides the opportunity to account for the gradual changes between these classes. This can also explain the differences between some of the predicted and reference class values, such as for Pixel 16 for instance, where any of the first or second class could be expected.

This study adopts an equal frequency approach to optimize the performance of the transition matrix. The six classes characterise both drought and non-drought vegetation status. This classification can be further optimized to relate only to the status of vegetation stress within a drought context.

Although some pixels classes are correctly predicted, field data or independent satellite images can be used to further validate the results. Further studies incorporating other parameters such as rainfall data and land use, in the prediction model can further explain the differences between the predicted and the actual data.

5. REFERENCES

[1] Compton J. Tucker, Jorge E. Pinzon, Molly E. Brown, Daniel A. Slayback, Edwin W. Pak, Robert Mahoney, Eric F. Vermote, and Nazmi El Saleous, “An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data,” *International Journal of Remote Sensing*, vol. 26, pp. 4485–4498, 2005.

[2] A. Mishra, V. Singh, and V. Desai, “Drought characterization: a probabilistic approach,” *Stochastic Environmental Research and Risk Assessment*, vol. 23, pp. 41–55, 2009.

[3] Pabitra Banik, Abhyudy Mandal, and M. Sayedur Rahman, “Markov chain analysis of weekly rainfall data in determining drought-proneness,” *Discrete Dynamics in Nature and Society*, vol. 7, pp. 231–239, 2002.

[4] Heiko Balzter, “Markov chain models for vegetation dynamics,” *Ecological Modelling*, vol. 126, no. 2-3, pp. 139 – 154, 2000.

[5] S. Ding, “Predicting dynamics of vegetative drought classes using fuzzy markov chains,” M.S. thesis, University of Twente, Faculty of Geo-Information and Earth Observation ITC, The Netherlands, 2011.

[6] F. Bickenbach and E. Bode, “Evaluating the markov property in studies of economic convergence,” *International Regional Science Review*, vol. 26, pp. 363–392, 2003.