Triplet Markov Chain for Modeling a Non-Stationary NDVI Time Series

Ali Ben Abbes 1,2,* , Imed Riadh Farah 1,3 and Vincent Barra 2

1 National School of Computer Science, SIVIT-RIADI Laboratory, University of Manouba, Manouba 2010, Tunisia
2 LIMOS Laboratory, University of Blaise Pascal, 63170 Aubière, France
3 ITI Department, Telecom Bretagne, Technopôle Brest-Iroise - CS 83818, 29238 BREST CEDEX 3, France
* Correspondence: ali.benabbes@yahoo.fr; Tel.: +216-97-385-504

Abstract: In this paper, we propose a Triplet Markov Chain (TMC) based technique to study vegetation monitoring using remotely sensed data. TMCs are a generalization of Hidden Markov Model (HMM). This latter has proved its ability to represent multi-temporal satellite images as well as to analyze vegetation dynamics on large scales. The main idea of using HMM is to relate the varying spectral response along the crop cycle with plant life. However, there has been considerable dissatisfaction, because it has been found unable to study a non-stationary data. An interesting feature for the application of TMC is to use auxiliary processes which model the non-stationarity. The primary purpose of this paper is to present a novel methodology based on TMC for modeling a non-stationary Normalized Difference Vegetation Index (NDVI) time series. In order to assess the performance of the proposed model experiments are carried out using Moderate Resolution Imaging Spectroradiometer (MODIS) NDVI time series of the northwestern region of Tunisia. Moreover, the developed model is compared to two models HMM and Seasonal Auto Regressive Integrated Moving Average (SARIMA). Our results show the efficacy of our model with a precision of 82.36%.

Keywords: NDVI time series; vegetation; HMM; non-stationary; TMC

1. Introduction

Recently, due to the considerable advances in remotely sensed data in their temporal, spatial, and spectral resolutions, it has become possible to make use of tools for a wide variety of environmental applications e.g. global vegetation studies, crop management, land cover studies and climate studies [1–4]. In order to take the best decisions, natural resource managers, policy makers and researchers avail of remote sensing data in various fields [1,5,6]. Earth components have different physical properties which generate different spectral responses. Remote sensing community exploits this important physical particularity in order to provide several products. Particularly, the Normalized Difference Vegetation Index (NDVI) proposed by [7] is considered as the most popular and the most widely used index in remote sensing applications. NDVI serves to quantify the vegetation condition, being directly related to the photosynthetic capacity and energy absorption of plant canopies [7–9]. It is defined as the difference between near-infrared and red reflectance divided by their sum. Generally, the dynamic analysis of vegetation through remotely sensed images involves various considerations, processes and techniques. The proposal of powerful methodologies for the NDVI time series is one of the most challenging issues that the remote sensing communities will face. In the last decade, a number of different methods have been developed to adopt an advent model, among them we can find: Principal Component Analysis [10], Curve Fitting [11], Fourier Analysis [12], Harmonic Analysis [13], Seasonal Autoregressive Integrated Moving Average (SARIMA) [14] and Wavelet Decomposition [15]. Most of these methods cannot get details “from, to” information, which is provided by stochastic models (e.g. Hidden Markov Model).

In this work, we are particularly interested on analyzing the vegetation monitoring using Triplet Markov Model (TMC). In fact, TMCs are a generalization of HMM [16,17]. This latter is a powerful statistical model for temporal information modeling and forecasting, which has been successfully
applied to various applications [18,19]. Several works presented the advantages of using an HMM [20,21], which has strong statistical background, and has a power to handle new data robustly and forecast similar patterns expeditiously. Due to the availability of established training algorithms, we can develop and evaluate the estimation parameters. In the literature, many previous works have used HMMs in remotely sensed images to study vegetation dynamics of different covers type with several points of view. Viovy and Saint [22] illustrated one of those points. The authors proposed a methodology to classify and to extract various dynamics parameters as well as to detect phenological anomalies using NDVI time series to study West African Savanna area. In [23], a novel approach has been developed to classify agricultural crops. The method uses HMM to relate the varying spectral responses along the crop cycle with plant phenology, for different crop classes. In [24,25], the authors have developed a methodology for vegetation ground cover classification in Norway, using phenology knowledge into the classification process. Shen et al. [26] proved that HMM is a suitable tool for estimating the corn progress percentages in the real-time combining multisource features. Essid et al. [27] suggested an approach based on HMM in order to predict spatio-temporal phenomena. Miguel et al. [28] used HMM to define phenological parameters from MODIS derived NDVI time series data in order to study a semi-arid Mediterranean region. In the last example in [29], authors have modeled the vegetation dynamic over an agricultural area in Greece.

According to the works listed above [22–29], we observe a vegetation evolution as states. The HMM includes two processes: the external process which represents the radiometric observations i.e. NDVI values, and the internal process of states that represents the phenology stage of plant species (e.g. dormancy, greenup, maturity, senescence).

An HMM contains a: $N$ the number of states in the model, stages of plant life are denoted as $O=\{O_1, O_2, ..., O_N\}$ and the stage at time $t$ as $q_t$, $M$ the number of distinct values that an observable variable may have per state and emission values are denoted as $V = \{v_1, v_2, ..., v_M\}$. In our problem, the basic HMM consists of three sets of parameters:

- The vector emission probabilities: $b_{jk}$ stands for the probability that the value $v_k$ is observed in state $O_j$, represent spectral values.

$$b_{jk} = P[v_k \text{ at } t | q_t = O_j] \text{ where } 1 \leq j \leq N \text{ and } 1 \leq k \leq M$$  

(1)

- The state transition probabilities: $a_{ij}$ is the probability of a system to be in state $O_j$ in the subsequent time instant, given that its current state is $O_i$:

$$a_{ij} = P[q_{t+1} = O_j | q_t = O_i] \text{ where } 1 \leq i, j \leq N$$  

(2)

- The initial probability distribution: $\pi_i$ stands for the probability of the system being in a given state $O_i$ at the initial time instant.

$$\pi_i = P[q_1 = O_i] \text{ where } 1 \leq i \leq N$$  

(3)

Generally, all these methods have been applied on short time series without noticing the crucial non-stationary behavior of the NDVI time series. As a matter of fact, one of the most important HMM backgrounds is its inability of modeling time variant data i.e. the state transitions are stationary over time [16,17,30–33]. To overcome the limitation of HMM, many scientists and researchers concentrate on this aspect in various fields. Some tried to enhance HMM itself which is known as Non Stationary HMM (NSHMM) in case where researchers proposed a time variant transition probabilities (as functions of time or multiple transition matrix) [34] or Variable Duration HMM (DHMM) when they included a state duration variable [35]. Recently, HMM have been generalized to Triplet Markov Chain [16,17]. Basically, a TMC is defined as triplet $T=(R,S,Y)$, and by assuming that the process is Markovian. In which a third auxiliary process $R$ is introduced. In fact, the third process allows the modeling of non-stationary data behavior. The joint distribution of TMC $t$ may be expressed by:
The major interest of TMC lies in the utility of auxiliary process $R$ which can be used to model different structure. It models different stationarity of distribution of $(S, Y)$ and thus $R$ models switches parameters defining the distribution of $(S, Y)$. TMCs have been used in various problems that occurred in different areas such as: signal processing [36], 3D MRI brain segmentation [37], data classification [38], Filtering [39], biometry [40], image segmentation [41,42] and change detection [43,44].

More precisely, the purpose of this paper is to establish a new approach of Triplet Markov Chain which attempts to overcome some limitations of using HMM. The approach adopted in this paper is based on TMC. Since their introduction in the literature in 2003, the TMC has proved its usefulness in studying of non-stationary time series. This stochastic method is beneficial when compared to the HMM because the latter requires stationarity.

The proposed approach has been applied on real datasets of northwest of Tunisia. More specifically, the outline of the paper is as follows: Section 1 presents the study area and date sets. Section 2 introduces TMC for monitoring vegetation using non-stationary data. In section 3, we present experimental results of our model with a comparative study with HMM and SARIMA model. To this end, we draw some future works.

2. Study Area and Data

2.1. Study area

The study area situated in the north-western region of Tunisia as shown in Fig. 1, inside a geographical region with the following coordinates (Latitude: 36° 45’ 00” N, Longitude: 8° 45’ 00” E). Agriculture is the main activity in this area. Furthermore, the region has a plain to slightly undulated relief, a tropical climate with rainy winter, annual average temperature of 24.9 ° C and annual precipitations of 1389 mm.

2.2. MODIS Time Series Data

The NDVI time series of the covered study area is extracted from 16-day MODIS NDVI composites with a 250 m spatial resolution (MOD13Q1 collection 5). The MOD13Q1 images were acquired from February 2000 to the end of December 2012. The frequency is 23 images per year. These data describe the red and the near infrared spectral wavelengths where the NDVI index is already calculated. It is worthy to mention that the generation of MODIS images is based on a constrained view angle maximum NDVI value compositing technique [45]. The advantages of using MODIS data include their large-scale coverage, high temporal resolution, and free open data. Fig. 2 shows the NDVI image acquired in April, 2005.

3. Methodology

In this section we propose a new methodology to model a vegetation dynamic using TMC model. The first step is to specify a TMC model, then to estimate of TMC parameters from the past dataset using an extension of Baum-Welch algorithm.

3.1. General model description

Before model training, the structure of the TMC should be fitted. The main idea is to relate the varying spectral response along the crop cycle, plant life with time. A TMC is thus defined as $T=(R,S,Y)$ where $S=\{S_1, S_2, \ldots, S_n\}$ is a hidden process where each $S_n$ takes its values in the state set $\Omega =\{\omega_1, \omega_2, \ldots, \omega_n\}$, $Y = (Y_n)_{N-1}^{n}$ is the observed one where each $Y_n \in \mathbb{R}$ and $R=\{R_1, R_2, \ldots, R_n\}$ is
an auxiliary process where each $R_n$ takes its values in a finite set $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_M\}$. Fig. 3 shows the TMC model fitted in our case study.

In order to define an TMC completely, the following elements are needed:

- The Hidden states: Let us consider $S$ the hidden states denoted by $S = \{S_1, S_2, ..., S_n\}$ where each $S_n$ takes its values in the state set $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$. During the study of the evolution of changes in vegetation photosynthesis, we could distinguish different variations. Changes usually differ for each plant species. For identification we calculate the NDVI values at $t-1$ and $t$. Consequently, the ups, downs and stable variations could represent the NDVI dynamics at the same time. As mentioned in Fig. 3, we have used three states in our case i.e $S_1$: Rising, $S_2$: Decreasing and $S_3$: Stable.

- The observed states: $Y = (Y_t)_{t=1}^N$, in our case, represent the probability of observing specific values of NDVI that occurred at a time $t$. Our NDVI times series describe seasonal variations with a minimum and maximum value. In fact, if the inputs are discrete. The observed states are symbolized by the number of possible cases. However, in our case, the input is a set of continuous values where the observation probability could be written as follows:

$$b_i(o_t) = f(o_t; k_i)$$  \hspace{1cm} (5)

Depending on the observation at time $t$, $f(o_t)$, which depends on a gaussian model whose parameters are $K_i$. The Gaussian distribution is defined by:

$$b_i(o_t) = \frac{1}{\sqrt{2\pi} \sum_i} e^{-\frac{(o_t - \mu_i)^2}{2 \sum_i}}$$  \hspace{1cm} (6)
We assume that \( p(y_n, u_n|x_n) \) are Gaussian. For \( M \) classes \( \Omega = \{ \omega_1, \ldots, \omega_M \} \), we have estimated \( K \) means \( \mu_1, \ldots, \mu_M \) and \( M \) variances \( \sigma^2_1, \ldots, \sigma^2_M \) of \( M \) Gaussian densities \( p(y_n, u_n|x_n = 1), \ldots, p(y_n, u_n|x_n = m) \).

Fig. 4 shows the division of intervals. This curve describes a stable NDVI variation. Surely, downs or aberrant peaks could be manifest. However, we recall that this division depends only on the value of the observation independently of production time.
Figure 4. Phenological development of a vegetation ground cover class when observed in the form of NDVI

- The auxiliary process: Let us consider $R$ the auxiliar states denoted by $R = \{R_1, R_2, ..., R_n\}$ where each $R_n$ takes its values in a finite set $\Lambda = \{\lambda_1, \lambda_2, ..., \lambda_M\}$. NDVI time series is usually non-stationary. This is due to the dependence of the plant to climate scenarios. To model this important statistical property, we have used a seasonal division $R = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Each plant is distinguished by a unique and very special behavior from one season to another.

This model is fully defined through the parameters set $\Theta = (\pi, A, B, C, \mu, \rho)$ where

- $\pi$ is a distribution on $\Omega$ such as $\pi_i = p(s_1 = \omega_i)$
- $A$ is a transition matrix defined on $\Omega^2$ as $a_{ij} = p(s_n = \omega_j | s_{n-1} = \omega_i)$
- $B$ is an emission probability defined on $\Omega \times \lambda$ as $b_{ij} = p(y_n | s_n = \omega_i)$
- $C$ which is an auxiliary probability defined on $\lambda$ as $c_{ij} = p(y_{n+k}^n | s_n^{n+k} = (\omega_2, ..., \omega_2, p_n = 1)$

4. Validation

This section involves the performance assessment of the proposed model. As a case study, the proposed model is applied to predict change of vegetation species. We present experimental results and finally we compare the proposed model to HMM and SARIMA model in order to evaluate its performance.

Our methodology used in this experimentation consisted of three steps: (1) classification of images and filtering, (2) specification of TMC model and (3) parameters estimation. In this study, for the classification phase we have used ISODATA (Iterative SelfOrganizing Data Analysis) as unsupervised classification, which calculates class means evenly distributed in the data space then iteratively clusters the remaining pixels using minimum distance techniques [46]. Each iteration recalculates means and reclassifies pixels with respect to the new means. This process continues until the number of pixels in each class changes by less than the selected pixel change threshold or the maximum number of iterations is reached. The vegetation species uses in this experimentation are: Agriculture land 1, dense forest, herbaceous, sparse and agriculture land 2. The process classification of NDVI time series of our study requires the extensive skills of an expertise. An expert of the ministry of agriculture of Tunisia has selected a several reference locations in the study area at different dates which, allows for determining the crop class and corresponding phenological stage. Once results are provided and compared by the expert investigated again the NDVI time series and decided on a final classification. Fig. 5 presents a segmentation of NDVI image acquired in April 2005 and table 1 illustrates the input of the five vegetation species in the studied area.
Figure 5. Segmentation of NDVI image acquired in April 2005

Table 1. Input of the five vegetation species in the studied area in April 2005

<table>
<thead>
<tr>
<th>Input</th>
<th>Agriculture Land 1</th>
<th>Dense Forest</th>
<th>Herbaceous</th>
<th>Sparse</th>
<th>Agriculture Land 2</th>
<th>Other Cover type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8%</td>
<td>24%</td>
<td>16%</td>
<td>23%</td>
<td>21%</td>
<td>8%</td>
</tr>
</tbody>
</table>

The extracted NDVI time series shows a high variability caused by different noise sources (see Fig. 6). In order to smooth our time series, we have chosen Savitsky Golay filter. In fact, this filter is proved efficient [47]. In addition, it is provided by Timesat program [2]. A time series of a selected pixel of forest is illustrated in Fig. 6 (the red curve). The black curve represents the smoothing one. In order to model the vegetation dynamics of a different vegetation species, TMC is learned for each one. As for forest, the observable values are divided into three intervals based on an expert knowledge related to the vegetation cover type: 1 = [0.47, 0.55]; 2 = [0.55, 0.7]; 3 = [0.7, 0.84]. These intervals represent the emission states of our fitted TMC. As we have already mentioned, three variations which are considered as the hidden states: \( \omega_1 = \text{"Rising (R)"} \), \( \omega_2 = \text{"Decreasing (D)"} \) or \( \omega_3 = \text{"Stable (S)"} \) of NDVI values. Typically, these three variations depend on time. Therefore, our TMC contains four stationary matrices: \( \lambda_1 = \text{"Winter (S1)"} \), \( \lambda_2 = \text{"Spring (S2)"} \), \( \lambda_3 = \text{"Summer (S3)"} \) and \( \lambda_4 = \text{"Autumn (S4)"} \) where each auxiliary state represents a season. Firstly, the auxiliary process \( R \) is generated; it is given by a distribution of four Markov states which are characterized by an initial distribution \( \pi = (1/4, 1/4, 1/4, 1/4) \) and the following transitions generated by Baum-Welch algorithm presented in subsection ??:

\[
M_R = \begin{pmatrix}
0.77 & 0.23 & 0 & 0 \\
0 & 0.77 & 0.23 & 0 \\
0 & 0 & 0.77 & 0.23 \\
0.47 & 0 & 0 & 0.63 \\
\end{pmatrix}
\]

The hidden state \( S \) is conditionally simulated to the auxiliary process \( R \) from:

\[
p(s_1 | r_1) = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}
\]
To answer the prediction problem, the TMCs were used to find an optimal hidden state sequence for an observed sequence of NDVI values for each pixel of vegetation species. The Viterbi algorithm was applied to forecast the optimal state sequence [48,49].

In order to evaluate prediction errors between proposed and real changes 7 experiments were carried out for the region at 7 different periods. The same previously mentioned methodology has been repeated. Predicted NDVI changes for these 7 periods were estimated through the proposed

\[ p(s_{t+1}|s_t, r_{t+1} = \lambda_i) = M_{\lambda_i} \]  

\[
M_{S1} = \begin{pmatrix}
0.42 & 0.21 & 0.37 \\
0.51 & 0.36 & 0.13 \\
0.09 & 0.17 & 0.74 \\
\end{pmatrix}
\]

\[
M_{S2} = \begin{pmatrix}
0.04 & 0.04 & 0.92 \\
0.33 & 0.33 & 0.34 \\
0.37 & 0.02 & 0.61 \\
\end{pmatrix}
\]

\[
M_{S3} = \begin{pmatrix}
0.71 & 0.21 & 0.08 \\
0.04 & 0.93 & 0.03 \\
0.45 & 0.09 & 0.46 \\
\end{pmatrix}
\]

\[
M_{S4} = \begin{pmatrix}
0.08 & 0.45 & 0.47 \\
0.17 & 0.73 & 0.1 \\
0.33 & 0.34 & 0.33 \\
\end{pmatrix}
\]

\[
A = \begin{pmatrix}
0.47 & 0.32 & 0.31 \\
0.29 & 0.51 & 0.2 \\
0.33 & 0.16 & 0.51 \\
\end{pmatrix}
\]
approach. Then, real NDVI values were evaluated being based on NDVI images representing the same dates in each period. Table 2 depicts the error calculated between proposed and real changes for each period.

**Table 2.** Average error for the prediction of land covers for 7 period tests.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average error (%)</td>
<td>1.25</td>
<td>1.34</td>
<td>1.36</td>
<td>1.45</td>
<td>1.89</td>
<td>3.56</td>
<td>2.01</td>
</tr>
</tbody>
</table>

We compare the proposed prediction changes of this study area between 2005 and 2006 as illustrated in Fig. 5 and Fig. 7 respectively to the prediction changes obtained using SARIMA, HMM and the real changes. This assumes that the image in 2006 wasn’t available. Thus, the significant changes are related to the forest and sparse. The surface of the urban area has grown from around 3% of the entire surface of the study area between 2005 and 2006.

Fig. 7 illustrates the obtained results for each vegetation species. The error related to dense forest prediction is 8%.

In order to assess the performance of the proposed model against the other proposed models, we compare the TMC model to HMM and SARIMA models.

SARIMA model included additional seasonal terms in the ARIMA models. The SARIMA model is a traditional statistical model. It has been highly successful for forecasting seasonal time series in many fields. According to the Box-Jenkins notation system: SARIMA model is defined by \((p,d,q)(P,D,Q)_s\). The first part \((p,d,q)\) represents the parameters of the non-seasonal part of series with \(p\) is a non-seasonal AR order, \(d\) is a non-seasonal differencing, \(q\) is a non-seasonal MA order. The second one, \((P,D,Q)_s\) describes the seasonal part with \(P\) is a seasonal AR order, \(D\) is a seasonal differencing, \(Q\) is a seasonal MA order and \(s\) is the length of the seasonal period (=4 in our case). More details are given in [14,50–52].

According to these results, we can conclude that TMC model has a precision of 82.36% for the five classes, whereas the SARIMA and HMM have a precision of 76.42% and 72.28% respectively. In fact, we can affirm that the TCM model is more accurate than SARIMA and HMM with difference by 5.94% and 10.08% respectively. Then, the table 3 bellow shows the Accuracy Rate (AR) using these models.
5. Conclusion and future works

In this paper we have presented and evaluated the use of Triplet Markov Chain in a new way to monitor vegetation. Precisely, we obtained a system that describes the vegetation evolving through a series of different states. In this system, the auxiliary matrix allowed the modelization of seasonal specific time dynamics. In fact, the obtained results proves the ability of TMC to model non-stationary data. The multiple auxiliary matrices generated different probabilities through time. On one hand, this overcomes the famous HMM drawback which is the fixed transition probability. On the other hand, it depicts an exceptional way of describing the vegetation dynamics according to the seasonal variation. The experimental are carried out in case of prediction based on a sequence of 276 MODIS images of northwestern region of Tunisia. The TMC provided better predictive ability compared to HMM and SARIMA. A first example of using the first application of our model is to estimate trends of a vegetation site and to follow if there are possible disasters (e.g. erosion, deforestation). The proposed approach can be applied to a variety of fields, and is not restricted to vegetation. This
modeling can be used in several applications such as classification of vegetation canopies, prediction. This research could be extended to include three main considerations:

- Taking advantage of decomposition methods (e.g. Breaks For Additive Seasonal and Trend analysis [47], Detecting Breakpoints and Estimating Segments in Trend analysis [53]) in order to: interpret and identify the different variations (seasonal and progressive) [4], and to take into account other types of non-stationary (i.e random variations)
- Recent trends aim to overcome the drawbacks of using one sensor’s data poor temporal resolution (e.g. MODIS data) by using the Multi sensor data fusion techniques.
- Applying the proposed model in other land cover changes (e.g. desertification, and urban growth) and in other applications such as classification.

Acknowledgments: The authors would like to thank the Ministry of agriculture of Tunisia for providing us with useful information.

References


46. Tou, J.T.; Gonzalez, R.C. Pattern recognition principles **1974**.


© 2016 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).