DE GRUYTER OPEN RTU Information Technology and Management Science doi: 10.1515/itms-2015-0009

2015 / 18

Applying Markov Chains for NDVI Time Series Forecasting of Latvian Regions

Arthur Stepchenko[1], Jurij Chizhov[2]
[1, 2] *Riga Technical University*

Abstract **– Time series of earth observation based estimates of vegetation inform about variations in vegetation at the scale of Latvia. A vegetation index is an indicator that describes the amount of chlorophyll (the green mass) and shows the relative density and health of vegetation. NDVI index is an important variable for vegetation forecasting and management of various problems, such as climate change monitoring, energy usage monitoring, managing the consumption of natural resources, agricultural productivity monitoring, drought monitoring and forest fire detection. In this paper, we make a one-step-ahead prediction of 7-daily time series of NDVI index using Markov chains. The choice of a Markov chain is due to the fact that a Markov chain is a sequence of random variables where each variable is located in some state. And a Markov chain contains probabilities of moving from one state to other.**

Keywords **– Forecasting, Markov chains, NDVI.**

I. Introduction

Human activities affect ecosystems, including the natural vegetation cover. Vegetation cover change is an important factor that affects ecosystem condition and function. A change of vegetation cover may have a long-term impact on sustainable food production, freshwater and forest resources, the climate and human welfare. Documenting changes occurring in vegetation cover at periodic intervals is very important to providing information about the stability of vegetation.

The use of satellite-based remote sensor data has been widely applied to provide a cost-effective means to develop land cover coverages over large geographic regions. Vegetation cover is an evident part of land cover. Change detection has become a widespread application of remotely sensed data because of repetitive wide coverage, short revisit intervals and good image quality. Change detection is the process of identifying differences in the state of an object or phenomenon by observing it at different times. The main prerequisite of using remote sensing data for vegetation change detection is that changes in land cover result in changes in radiance values, and changes in radiance due to land cover change are large with respect to radiance change caused by others factors, such as differences in atmospheric conditions, differences in soil moisture and differences in sun angles [1].

Vegetation indices calculated from satellite images can be used for monitoring temporal changes associated with vegetation. Vegetation indices (VIs) are combinations of digital numbers (DNs) or surface reflectance at two or more wavelengths designed to take out a particular property of vegetation. Each of the VIs is designed to emphasise a particular vegetation property. Analysing vegetation using remotely sensed data requires knowledge of the structure and function of vegetation and its reflectance properties. This knowledge enables the linking of vegetative structures and their condition to their reflectance behaviour in an ecological system of interest [2]. The normalised difference vegetation index (NDVI) is developed for estimating vegetation cover from the reflective bands of satellite data. The NDVI is an indicator that quantifies the amount of green vegetation. Past studies have demonstrated the potential of using NDVI to study vegetation dynamics. The NDVI data layer is defined as:

$$NDVI = (NIR - R)/(NIR + R)\,, \qquad (1)$$

where *NIR* represents the spectral reflectance in near-infrared band and *R* represents red band. Greener and dense vegetation has low red light reflectance and high near-infrared reflectance, and thus high NDVI values. The NDVI real values, by definition, would be between −1 and +1, where increasing positive values indicate increasing green vegetation, but low positive values and negative values indicate non-vegetated surface features such as water, barren land, rock, ice, snow, clouds or artificial materials [3]. The NDVI has also the ability to reduce external noise factors, such as topographical effects and sun-angle variations. When analysed through time, NDVI can reveal where vegetation is thriving and where it is under stress, as well as changes in vegetation due to human activities such as deforestation, natural disturbances such as wild fires, or changes in plant phenological stage. Therefore, the NDVI index is an important variable for vegetation forecasting and management of various problems, such as climate change monitoring, energy usage monitoring, managing the consumption of natural resources, agricultural productivity monitoring, drought monitoring and forest fire detection.

Time series analysis of remotely sensed data, as shown earlier, has gained special attention supported by availability of wide-coverage, high temporal satellite data. NDVI time series data have been employed to predict the NDVI variable beyond the time span. Box-Jenkins methods, including univariate autoregressive integrated moving average (ARIMA) models, are widely used for univariate time series forecasting, also for the NDVI time series [4]. However, these models are parametric. Markov chains are used in this study as an alternative to Box-Jenkins methods.

II. Study Area and Data Acquisition

A. Study Area

Ventspils Municipality is located in the western part of Courland, Latvia, with total area of 2 472 km^2 (Fig. 1).

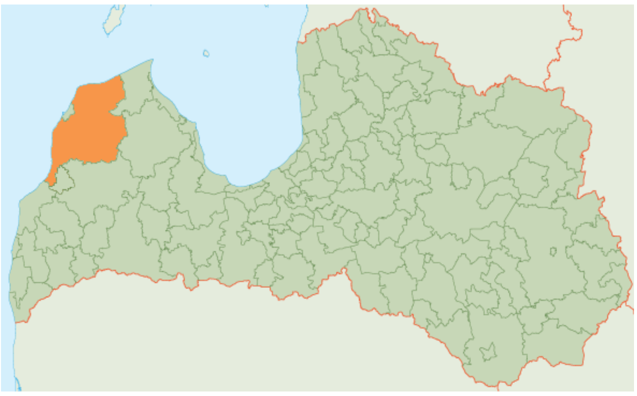

Fig. 1. Location of the Ventspils Municipality.

One pixel (Fig. 2) from Ventspils Municipality images was selected as a test site.

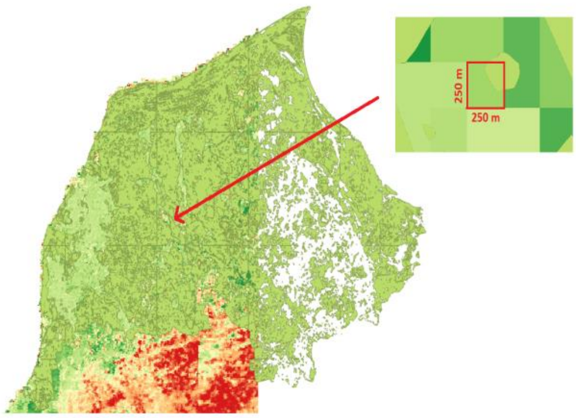

Fig. 2. Test pixel in Ventspils Municipality.

Test pixel size is 250 m x 250 m. The climate at the test site is determined by a temperate climate zone with significant maritime features. Approximately half of the area is covered by forests. Latvia lies on the border between two different forest types: the northern coniferous zone and the broad-leaved trees of the temperate zone, so the tree species, characteristic of the both forest types, can be found in the landscape.

B. NDVI Data Set

This study explored the use of multi-temporal MODIS Terra NDVI composite data with spatial resolution 250 m and produced on 7-day intervals (Fig. 3). Data were obtained from data service platform for MODIS vegetation index time series processing at BOKU, Vienna [5]. The used data were smoothed and gap-filled using the Whittaker smoothing algorithm with smoothing parameter $\lambda = 15$ and two filtering iterations [6]. Iterative filtering was used, because undetected clouds and poor atmospheric conditions decreased the observed NDVI values.

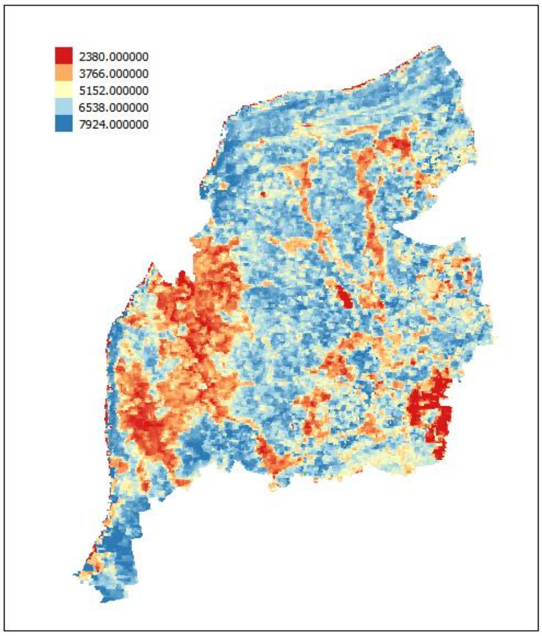

Fig. 3. MODIS Terra NDVI satellite image.

The NDVI data set consists of 814 smoothed NDVI images obtained every 7 days over 14 years. NDVI values of these images were obtained for corresponding pixel and used as NDVI time series observations (Fig. 4).

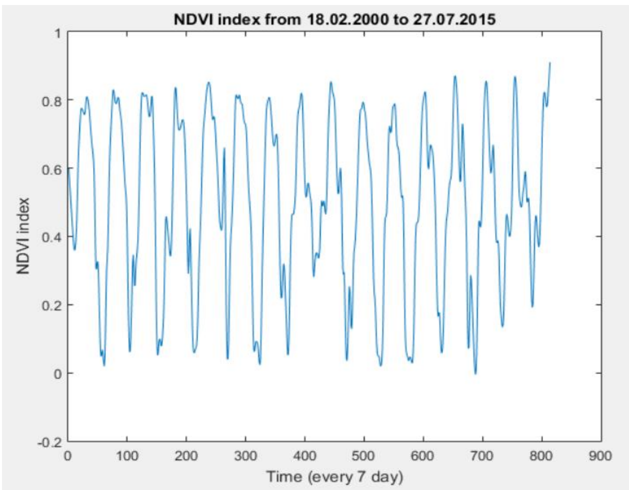

Fig. 4. Smoothed NDVI time series from 18 February 2000 to 27 July 2015.

The NDVI time series data provide a seasonal trajectory – time series show pronounced seasonal oscillations, which correspond to the vegetation phenological cycles where maximum NDVI values are observed between May and August. Variations in the NDVI values are seen to be −0.0050 to 0.9109 units. NDVI trends are not always monotonic but can change. A positive trend can change, for example, into a negative one and reversely.

III. Markov Chains

Many decisions are accepted within the context of randomness. In order to calculate, understand, and predict the effects of randomness, one special type of stochastic process called a Markov chain is introduced in this paper [7]. Markov chains are usually used in modelling many practical problems and are useful in studying the evolution of systems where the state of the system in any particular period cannot be determined with certainty [8]. They are also effective in modelling time series. If a Markov chain can model the time series accurately, then good predictions and optimal planning in a decision process can be made [9].

Assume that there is a sequence of discrete states. From this sequence, we can calculate transition probabilities between the two states [10]. Simple Markov chain is a random process that undergoes transitions from one state to another on a state space. It must possess a property that is usually characterised as "memorylessness": the probability distribution of the next state depends only on the current state rather than the sequence of events in past. The stochastic process $X = \{X_n;\ n = 0, 1, \ldots\}$ with discrete state space **S** is a first-order discrete-time *Markov chain* if the following holds for each $j \in S$ and $n = 0, 1, \ldots, N$:

$$\Pr\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_n = i_n\} = \Pr\{X_{n+1} = j \mid X_n = i_n\} \tag{2}$$

for any set of states $i_0, \ldots, i_n$ in the state space [11]. The possible values of X_i come from a finite or countable set **S** called the state space of the chain. Equation (2) is a mathematical statement of the Markov property. Furthermore, the Markov chain is said to have *stationary* transition probabilities if:

$$\Pr\{X_1 = j \mid X_0 = i\} = \Pr\{X_{n+1} = j \mid X_n = i\}. \tag{3}$$

As the probabilities are stationary, the only information needed to describe the process is the initial conditions: the one-step transition probabilities. A square matrix is used for the transition probabilities and is often denoted by the capital letter **P**:

$$P = \begin{vmatrix} P(1,1) & P(1,2) & \ldots & P(1,r) \\ P(2,1) & P(2,2) & \ldots & P(2,r) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ P(r,1) & P(r,2) & \ldots & P(r,r) \end{vmatrix}, \tag{4}$$

where,

$$P(i, j) = \Pr\{X_1 = j \mid X_0 = i\}, \tag{5}$$

and r is the number of states. Since the matrix **P** contains probabilities, it is always nonnegative and the sum of the elements in each row equals one. Any nonnegative matrix with row sums equal to one is called a *Markov matrix*. Transition probabilities are used to describe the manner in which the system makes transitions from one period to the next. It helps us to determine the probability of the system being in a particular state at a given period of time.

IV. Experimental Procedure

The aim of this experiment is to investigate the capability and accuracy of Markov chains in the NDVI time series forecasting when states are used that include only NDVI values without exogenous variables. This experiment shows how much information about future values of a NDVI time series current values of a NDVI time series contain. At the first stage, values of NDVI index of every NDVI image pixel were calculated and the NDVI time series was created. The second stage employed a Markov chain for time series modelling and forecasting.

The data set was divided into two sets, training and testing data sets by 85/15 principle, namely, 85 % of the NDVI data (a total of 692 observations) were used as a training data set and the remaining NDVI data (a total of 122 observations) were used as a testing data set.

The values of NDVI index can be classified into 10 possible states **S** = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10). NDVI values are expressed as:

- The first state (NDVI value ≤ 0);
- The second state (0 < NDVI value ≤ 0.1);
- The third state (0.1 < NDVI value ≤ 0.2);
- The fourth state (0.2 < NDVI value ≤ 0.3);
- The fifth state (0.3 < NDVI value ≤ 0.4);
- The sixth state (0.4 < NDVI value ≤ 0.5);
- The seventh state (0.5 < NDVI value ≤ 0.6);
- The eight state (0.6 < NDVI value ≤ 0.7);
- The ninth state (0.7 < NDVI value ≤ 0.8);
- The tenth state (0.8 < NDVI value ≤ 1).

We use a confusion matrix (or an error matrix) to measure prediction errors. Each column of the confusion matrix represents the instances in a predicted state while each row represents the instances in an actual state. Overall prediction accuracy can be calculated by [9]:

$$r = \frac{1}{N-n} \sum_{i=n+1}^{N} a_i \times 100\ \%, \tag{6}$$

where

$$a_i = \begin{cases} 1, if\ \hat{y} = y \\ 0, otherwise \end{cases}. \tag{7}$$

In (7) n is a size of train data, $\hat{y}$ is a predicted state, but y is an observed state.

V. Results

Calculated one-step transition probabilities are summarised in Table I and a confusion matrix is given in Table II.

TABLE I
ONE-STEP TRANSITION PROBABILITIES

State	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	1	0	0	0	0	0	0	0	0
2	0.0012	0.8427	0.1461	0	0	0	0	0	0	0
3	0	0.2857	0.3810	0.3333	0	0	0	0	0	0
4	0	0.0217	0.2609	0.3913	0.3261	0	0	0	0	0
5	0	0	0.0143	0.2143	0.5571	0.2143	0	0	0	0
6	0	0	0	0	0.1818	0.6104	0.2078	0	0	0
7	0	0	0	0	0.0244	0.1829	0.5854	0.2073	0	0
8	0	0	0	0	0	0	0.2179	0.5897	0.1923	0
9	0	0	0	0	0	0	0	0.1087	0.7971	0.0942
10	0	0	0	0	0	0	0	0	0.1912	0.8088

TABLE II
A CONFUSION MATRIX

		Predicted									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Actual	**1**	**0**	0	0	0	0	0	0	0	0	0
	2	0	**0**	0	0	0	0	0	0	0	0
	3	0	0	**6**	2	0	0	0	0	0	0
	4	0	0	2	**2**	2	0	0	0	0	0
	5	0	0	0	3	**8**	4	0	0	0	0
	6	0	0	0	0	5	**21**	3	0	0	0
	7	0	0	0	0	0	4	**15**	3	0	0
	8	0	0	0	0	0	0	4	**5**	2	0
	9	0	0	0	0	0	0	0	3	**7**	3
	10	0	0	0	0	0	0	0	0	4	**14**

The overall prediction accuracy of the proposed model is $r = 0.6393$. It means that 63.93 % of the states are predicted correctly. 36.07 % of the states are predicted incorrectly; however, as seen from Table II, these incorrect states are predicted either one state higher or lower in relation to the correct value of the state.

VI. CONCLUSION

In this paper, one-step-ahead prediction of the normalised difference vegetation index (NDVI) data recorded by satellites over Ventspils Municipality in Courland, Latvia, were obtained using a Markov chain. The Markov chain prediction method is purely a probability forecasting method as the predicted results are probability of a certain state of NDVI values in the future. This study showed how the Markov model fitted the data and its ability to predict future values due to its memoryless property and random walk capability. Each state could be achieved directly by every other state in the transition matrix, consequently giving satisfactory results with overall prediction accuracy 0.6393. Overall prediction accuracy can be higher if states that are more complex will be used. These states can include information about time series direction or exogenous variables that well correlated with NDVI index.

REFERENCES

[1] M. M. Badamasi, S. A. Yelwa, M. A. AbdulRahim and S. S. Noma, “NDVI threshold classification and change detection of vegetation cover at the Falgore Game Reserve in Kano State, Nigeria,” *Sokoto Journal of the Social Sciences*, vol. 2, no. 2, pp. 174–194.

[2] N. B. Duy and T. T. H. Giang, “Study on vegetation indices selection and changing detection thresholds selection in Land cover change detection assessment using change vector analysis,” *presented at International Environmental Modelling and Software Society (iEMSs)*, Sixth Biennial Meeting, Leipzig, Germany, 2012.

[3] E. Sahebjalal and K. Dashtekian, “Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods,” *African J. of Agricultural Research*, vol. 8, no. 37, pp. 4614-4622, September 26, 2013.

[4] M. Manobavan, N. S. Lucas, D. S. Boyd and N. Petford, “Forecasting the interannual trends in terrestrial vegetation dynamics using time series modelling techniques,” *presented at the ForestSAT Symposium Heriot Watt University*, Edinburgh, United Kingdom, 5th - 9th August 2002.

[5] University of Natural Resources and Life Sciences, Vienna. Data service platform for MODIS Vegetation Indices time series processing at BOKU, Vienna. Available: http://ivfl-info.boku.ac.at/, [Accessed August 25, 2015].

[6] F. Vuolo, M. Mattiuzzi, A. Klisch and C. Atzberger, “Data service platform for MODIS Vegetation Indices time series processing at BOKU Vienna: current status and future perspectives,” *Proc. SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III*, 85380A (October 25, 2012). http://dx.doi.org/10.1117/12.974857

[7] R. M. Feldman and C. Valdez-Flores, *Applied Probability and Stochastic Processes, 2nd Edition*, Springer-Verlag Berlin Heidelberg 2010, p. 141. http://dx.doi.org/10.1007/978-3-642-05158-6_5

[8] D. R Anderson, D. J. Sweeney, T. A. Williams, J. D. Camm and J. J. Cochran, *Quantitative Methods for Business, 13th Edition*, South-Western College Pub 2015, p. 771.

[9] T. Liu, "Application of Markov Chains to Analyze and Predict the Time Series," *Modern Applied Science*, vol. 4, no. 5, pp. 162-166, May 2010. http://dx.doi.org/10.5539/mas.v4n5p162

[10] V. Soloviev, V. Saptsin and D. Chabenko, "Markov Chains Application To The Financial-Economic Time Series Prediction," *Computer Modelling and New Technologies*, vol. 14, no. 3, pp. 16–20, 2011.

[11] G. Rakocevic, T. Djukic, N. Filipovic and V. Milutinovic, *Computational Medicine in Data Mining and Modeling*, Springer-Verlag New York 2013, p. 205. http://dx.doi.org/10.1007/978-1-4614-8785-2

[12] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2nd Edition*, Springer-Verlag Berlin Heidelberg 2011, p. 280. http://dx.doi.org/10.1007/978-3-642-19460-3

Arthur Stepchenko is a second-year Doctoral Student majoring in Information Technology at Riga Technical University (RTU). He received his *Mg. sc. comp.* degree in Computer Science from Ventspils University College (VUC) in 2013.
He is a Research Assistant at the Department of Space Technology of the Engineering Research Institute "Ventspils International Radio Astronomy Centre", Ventspils University College (VIRAC).
His research interests include machine learning, data mining, digital image processing and programming.
E-mail: Arturs.Stepcenko@edu.rtu.lv

Jurij Chizhov is currently a Lecturer and Leading Researcher at the Department of Modelling and Simulation, Riga Technical University (RTU). He received his *Dr. sc. ing.* from Riga Technical University in 2012.
His major research interests include a number of techniques related to computational intelligence, in particular, cluster analysis, ontology building, evolutionary algorithms, reinforcement learning, artificial neural network.
Now he focuses on the learned artificial neural network simplification and translating to a set of rules.
E-mail: jurijs.cizovs@rtu.lv