DE GRUYTER OPEN
RIGA TECHNICAL UNIVERSITY
*Information Technology and Management Science*
ISSN 2255-9094 (online)
ISSN 2255-9086 (print)
December 2016, vol. 19, pp. 39–44
doi: 10.1515/itms-2016-0009
*https://www.degruyter.com/view/j/itms*

# Markov Chain Modelling for Short-Term NDVI Time Series Forecasting

Artūrs Stepčenko[1], Jurijs Čižovs[2]
[1, 2] *Riga Technical University*

***Abstract* – In this paper, the NDVI time series forecasting model has been developed based on the use of discrete time, continuous state Markov chain of suitable order. The normalised difference vegetation index (NDVI) is an indicator that describes the amount of chlorophyll (the green mass) and shows the relative density and health of vegetation; therefore, it is an important variable for vegetation forecasting. A Markov chain is a stochastic process that consists of a state space. This stochastic process undergoes transitions from one state to another in the state space with some probabilities. A Markov chain forecast model is flexible in accommodating various forecast assumptions and structures. The present paper discusses the considerations and techniques in building a Markov chain forecast model at each step. Continuous state Markov chain model is analytically described. Finally, the application of the proposed Markov chain model is illustrated with reference to a set of NDVI time series data.**

***Keywords* – Continuous state space, Markov chains, NDVI, short-term forecasting.**

## I. Introduction

Remote sensing has shown great opportunities in vegetation mapping and monitoring over the past decades. Remote sensing data such as satellite images can be obtained fast, regularly; images cover large areas and remote sensing is cost effective compared to traditional vegetation mapping and monitoring methods. Vegetation indices obtained from satellite images are used to describe vegetation properties [2]. One of the most popular vegetation indices is the normalised difference vegetation index (NDVI). The NDVI is a numerical indicator that uses the ratio between spectral reflectance measurements obtained in red and infrared bands of the electromagnetic spectrum [1]. The NDVI has found a wide application in vegetative studies such as forestry, agriculture, food security and water management. In order to obtain NDVI values, the red band values are subtracted from the near-infrared band values and divided by the sum of near-infrared and red bands:

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

where NIR is the spectral reflectance value in near-infrared band and R is the spectral reflectance value in red band [3]. The NDVI takes values between −1 and 1; however, the typical range is between about 0 to 0.9. Negative values of NDVI (values close to −1) correspond to water. Values close to zero correspond to barren areas of rock, sand, or snow. Low and medium positive values represent shrub and grassland, while high values (close to 1) correspond to dense vegetation, for example, forests [15]. When analysed through time, the NDVI can show where vegetation is healthy and where it is under stress, as well as changes in vegetation due to human activities such as deforestation or natural disasters such as wild fires.

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 processes named Markov chains is examined in this paper. 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 cannot be determined with certainty [4]. Therefore, Markov chains are often used for capturing dynamic behaviour with a large stochastic component [5]. 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 [6].

In their previous research [3], the authors used discrete time, discrete state first order Markov chains in order to obtain short-term forecasts of the NDVI time series. The aim of the experiment described in this paper is to examine accuracy of the discrete time, continuous state *m*-th order Markov chains combined with feature selection method – stepwise regression and feature extraction method – principal component analysis as data pre-processing methods in the NDVI time series forecasting problem.

## II. Study Area and Data Acquisition

### *A. Study Area*

Ventspils Municipality is the municipality in Courland, Latvia. Its area is 2472 km$^2$ (Fig. 1).

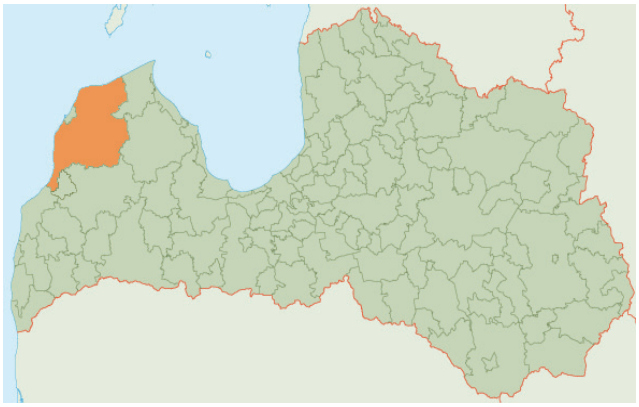

Fig. 1. Ventspils Municipality.

One pixel with spatial resolution of 250 m (Fig. 2) was selected as a test site.

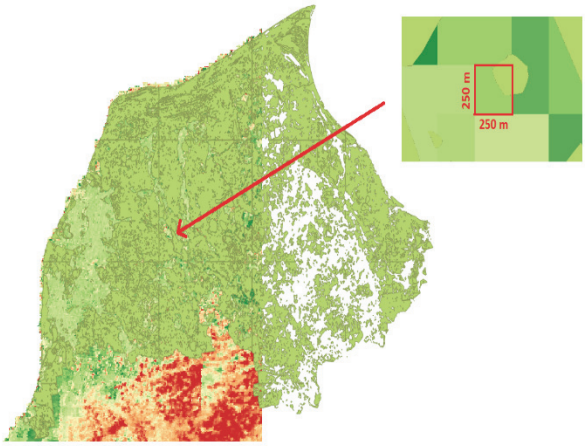


Fig. 2. Test pixel with spatial resolution of 250 m.

Therefore, the size of a test site is 250 m by 250 m. A moderate climate zone with significant maritime elements is dominating in this area.

*B. NDVI Data Set*

Multi-temporal MODIS Terra NDVI images with spatial resolution of 250 m and temporal resolution of 7 days (Fig. 3) were used in this study. Data were obtained from data service platform for MODIS vegetation index time series processing at BOKU, Vienna [7]. The used data were smoothed and gap-filled using the Whittaker smoothing algorithm with smoothing parameter $\lambda = 15$ and two filtering iterations [8]. Iterative filtering was used because undetected clouds and poor atmospheric conditions decreased the observed NDVI values.

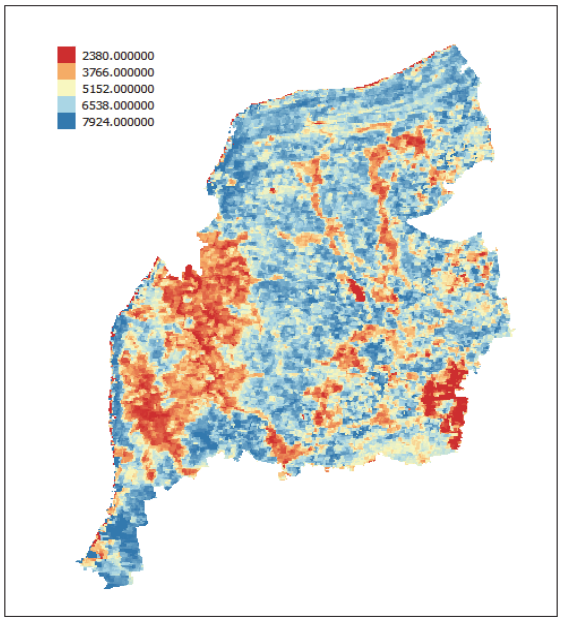


Fig. 3. MODIS Terra NDVI satellite image.

The data set contains 814 smoothed NDVI images obtained every 7 days within 15 years. NDVI values of these images were collected for corresponding test pixel and NDVI time series was obtained (Fig. 4).

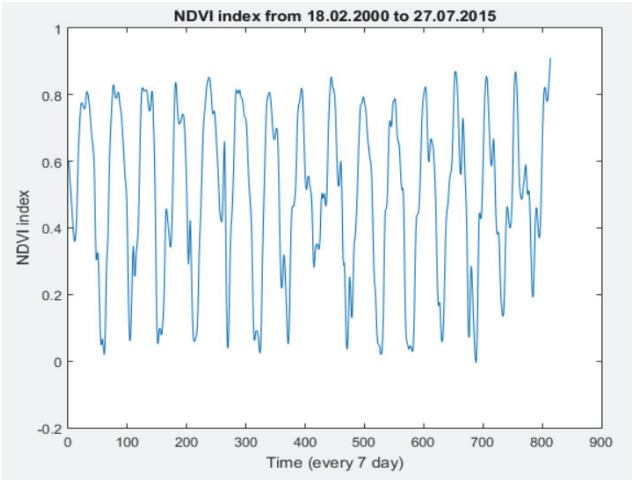


Fig. 4. Smoothed NDVI time series from 18 February 2000 to 27 July 2015.

Table I summarises the descriptive statistics of the NDVI time series.

TABLE I
DESCRIPTIVE STATISTICS

| | **NDVI** |
|---|---|
| **Observations** | 814 |
| **Mean** | 0.4965 |
| **Median** | 0.5171 |
| **Maximum** | 0.9109 |
| **Minimum** | −0.0050 |
| **Standard deviation** | 0.2492 |
| **Skewness** | −0.3591 |
| **Kurtosis** | 1.9533 |

The NDVI time series data show seasonal oscillations, which correspond to the vegetation phenological cycles. Maximum NDVI values conform to the period between May and August, and minimum values conform to the period between November and February.

## III. MARKOV CHAINS

A Markov chain is a stochastic process $X = \{X_n; n = 0, 1, \ldots\}$ that sequentially moves from one state to another in the state space [9]. A Markov chain consists of state space $S$, which is a set of values that the chain can take and a transition operator that determines the probability of moving from one state to another. Transition operator can help to determine the probability that system will go to a certain state in the next period.

If the state space of a Markov chain takes on a finite number of distinct values, then the transition operator can be defined by a matrix that it is always nonnegative and where the sum of the elements in each row equals one. Markov chain is said to have *stationary* transition probabilities if:

$$\Pr\{X_1 = j \mid X_0 = i\} = \Pr\{X_{n+1} = j \mid X_n = i\}. \qquad (2)$$

A Markov chain can also have a continuous or uncountable state space that can take all real numbers. In this case, the transition operator cannot be defined simply as a probability transition matrix. Continuous state space Markov chain also has a stationary distribution. However, the stationary distribution will also be over a continuous set of variables. For a continuous state Markov chain, the true underlying transition probability for any current state is unique, but expected to be smoothly varying with changes in the initial state [10].

Higher-order Markov chain is a stochastic process, in which the probability to get to the next state value depends not only on the current, but also on the sequence of several previous states or history [11]. The number of states in history that is used to calculate the probability is the order of the Markov chain.

## IV. Reconstructed Phase Space

A phase space of a dynamical system is a space where all possible states of this system are represented. Each possible state is one unique point in the multidimensional phase space. The evolution of system in time establishes a phase space trajectory for the system through the high-dimensional space [12]. In order to estimate the internal system information from complex time series data, time-delay phase space can be used. From the original time series $Y$ with length $N$:

$$Y = \{y(1), y(2), ..., y(N)\}, \tag{3}$$

$i$-th state vector or delay vector can be obtained by:

$$S_i = [(y(t_i), y(t_i + \tau), y(t_i + 2\tau), ..., y(t_i + (m-1)\tau)], \tag{4}$$

where $m$ is an embedding dimension and $\tau$ is a time delay. Therefore, one-dimensional time series measurements in time are transformed into a sequence of $m$-dimensional state vectors and reconstructed phase space $PhS$ is given by:

$$PhS = \begin{bmatrix} y(1) & y(1+\tau) & y(1+2\tau) & ... & y(1+(m-1)\tau) \\ y(2) & y(2+\tau) & y(2+2\tau) & ... & y(2+(m-1)\tau \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ y(M) & y(M+\tau) & y(M+2\tau) & ... & y(M+(m-1)\tau \end{bmatrix}, \tag{5}$$

where $M$ is a number of points or states in the reconstructed phase space.

## V. Stepwise Regression

Stepwise regression is a sequential feature selection method that can be used for least-squares regression in which the choice of predictive variables is implemented by an automatic procedure [13].

Initial regression model includes a constant variable. Sequence of F-tests is used in order to compare explanatory power of incrementally larger and smaller models that are obtained adding or removing features to initial model. At each step, the p value of an F-statistic is computed to test models with and without a potential feature. If the F-test values of the new model are better than that of the first model, the new model is saved and the third feature is added. If the new model performs worse compared to the first one, the first feature is removed, the second variable is kept and the next model is created that contains the second and third features. This procedure repeats until all two variable combinations are tested, the best performing model that contains two variables (features) is selected as the final model at this stage. Then the procedure is repeated with adding or removing the third variable. The process ends when all significant features are included in the model.

## VI. Principal Component Analysis

Principal component analysis (PCA) is a statistical feature extraction method that uses an orthogonal transformation to transform a set of possibly correlated features $x_1, x_2, x_3, \ldots, x_p$ into a set of linearly uncorrelated features $y_1, y_2, y_3, \ldots, y_m$ called principal components [14]. Here $p$ is the dimension of the original data set. The principal components are chosen so that the first principal component $y_1$ contains the maximum variance, the second principal component $y_2$ has the second greater variance and it is uncorrelated with the first principal component, and so on. Therefore, the goal of PCA is to find a set of orthogonal features that minimise the error in the transformed data set.

The first step in the PCA algorithm is to normalise the features so that they have zero mean and unity variance. Then the second step is to compute the principal components of the normalised features using orthogonalisation method. The principal components are orthogonal because they are the eigenvectors of the sample covariance matrix, which is symmetric and positive semi-definite. Sample covariance matrix is given by:

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^T (x_i - \mu), \tag{6}$$

where $x_i$ is $i$-th original feature (component), $\mu$ is the sample mean and $N$ is the number of samples. This leads to:

$$Cy_i = l_i y_i, i \in 1, ..., m, \tag{7}$$

where $l_i$ is the $i$-th largest eigenvalue of covariance matrix $C$ and $m$ is the number of principal components [14]. Therefore, the PCA method can also help reduce dimensionality of the original data set.

## VII. Experimental Procedure

The data set was divided into three parts: training, validation and testing data set where 70 % of the NDVI data (or 568 observations) were used as a training data set, 15 % of the NDVI data (or 122 observations) were used as a validation data set and last 15 % of the NDVI data (or 122 observations) were used as a testing data set.

It can be assumed that the measurements of time series $Y$ are obtained by a time-discrete sampling of an instantaneous and

arbitrary projection of the state vectors $S_i$ of some underlying Markov process [10]. We want to construct a stochastic process whose properties are compatible with the observed data $Y$. A continuous state Markov chain, whose order $m$ is a Markov model parameter representing the memory, approximated the stochastic dynamics of the time series $Y$. The present state of the underlying Markov process, $X_N$, can be replaced by the last $m$ measurements of the observed time series. This dependence of the one-step-ahead value of the variable $y(N+1)$ on its current and past values $y(N)$, $y(N-1)$, $y(N-2)\ldots$ can be described by using a conditional probability distribution function (pdf), so that the probability to observe $v$ in the next period is approximated by [10]:

$$\Pr\{y(N+1)=v \mid y(N), y(N-1), \ldots, y(N-m+1)\}. \quad (8)$$

In our constructed Markov chain, each state is equal to real number. Since we used $m$-th order Markov chain, last $m$ states or last $m$ time series observations were used in order to forecast the next state. A combination of these $m$ Markov chain states established a vector that was formally identical to state vector in the reconstructed phase space. Geometrically, the current state vector $S_N$ is a point in an $m$ dimensional phase space. Neighbouring points in this phase space represent similar state vectors. It can be assumed that similar points in phase space have a similar probability distribution. Variable $\Phi_\varepsilon(S_N)$ represents a neighbourhood around the vector $S_N$ bounded by small diameter ε. Variable $|\Phi_\varepsilon(S_N)|$ represents the number of vectors $S_k$, $k < N$, in this neighbourhood where these vectors $S_k$ are taken from the past measurements of the time series. For these vectors the future values $S_{k+1}(y(t_{k+1}+(m-1)\tau))$ were examined and variable $N(v)$ represented their number within the neighbourhood. Then the conditional probability was obtained by [10]:

$$\Pr\{v \mid S_N\} \approx \frac{N(v)}{|\phi_\varepsilon(S_N)|}. \quad (9)$$

In practice, it is enough to predict first or the second moment (a mean and a variation) of conditional probability. In this study we predicted first moment of a conditional probability.
The optimal prediction was given by the first moment of estimated conditional probability obtained by (9):

$$E[v] \approx \frac{1}{|\phi_\varepsilon(S_N)|} \sum_{k \in \phi_\varepsilon(S_N)} S_{k+1}(y(t_{k+1}+(m-1)\tau)). \quad (10)$$

The first moment of conditional probability (10) is an average value of current state $S_N$ possible future values. It minimises the root mean square error by means of maximum likelihood.

Stepwise regression was applied to initial input data set where 100 last values of the NDVI time series were used as features in order to reduce input data dimensionality and improve continuous state Markov chain predictability. In the experiments it was found that the optimal number of input data was two last successive values of the NDVI time series and the phase space was reconstructed with $m = 2$ and $\tau = 1$. Reconstructed phase space is shown in Fig. 5.

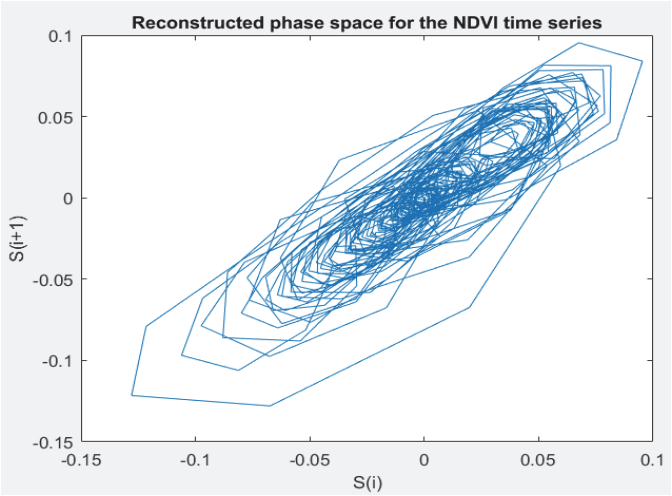


Fig. 5. The reconstructed phase space for the NDVI time series.

Then the PCA method was applied to phase space and linearly uncorrelated data set was obtained. After data pre-processing procedure, a continuous state second order Markov chain was used in order to obtain probabilistic forecasts. As diameter the Euclidean distance was used, and using cross-validation it was found that the optimal diameter ε = 0.06.

Several error measurements were chosen as performance criteria. The root mean square error (RMSE) is the standard deviation of the residuals (prediction error between observed and predicted time series) and is given by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}}, \quad (11)$$

where $\hat{y}_i$ – the forecasted value, $y_i$ – the observed value, $N$ – the number of data set observations. The MAPE (mean absolute percentage error) is a relative percentage error measure that uses absolute values and is given by:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| 100. \quad (12)$$

Directional symmetry (DS) is a statistical measure that provides a numerical value of the correctness of the forecasted time series directions in percentage terms and is given by:

$$DS = \frac{100}{N-1}\sum_{i=2}^{N} d_i, \quad (13)$$

where

$$d_i = \begin{cases} 1, \text{if}\,(y_i - y_{i-1})(\hat{y}_i - \hat{y}_{i-1}) \geq 0 \\ 0, \text{else} \end{cases}. \quad (14)$$

Directional symmetry statistic gives the percentage of observations when the sign of the change in value from one period to the next is the same for both the actual and forecasted time series. The adjusted coefficient of multiple determination ($R_{adj}^2$) is a measure of covariation between the observed and predicted time series data and the values lie within the interval [0, 1]. The adjusted coefficient of multiple determination is given by:

$$R_{adj}^2 = 1 - \left(\frac{N-1}{N-p}\right)\frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (15)$$

where $p$ is the number of features in the input data set and $\bar{y}$ is the mean value of the observed time series values.

## VIII. Results

Table II shows the performance of the continuous state second order Markov chain model on the NDVI data set.

TABLE II

Forecasting Performance

| Data set | RMSE | MAPE | DS | $R_{adj}^2$ |
|---|---|---|---|---|
| **Validation** | 0.029109 | 27.236117 % | 80.991736 % | 0.988336 |
| **Testing** | 0.027724 | 4.937291 % | 87.603306 % | 0.980318 |

The RMSE, MAPE and DS errors were the smallest ones on a testing data set, but the adjusted coefficient of multiple determination was the best one on a validation data set. The results of forecasting performance on both data sets showed acceptable accuracy of a continuous state second order Markov chain. The observed and forecasted values of the NDVI time series on a validation data set are shown in Fig. 6.

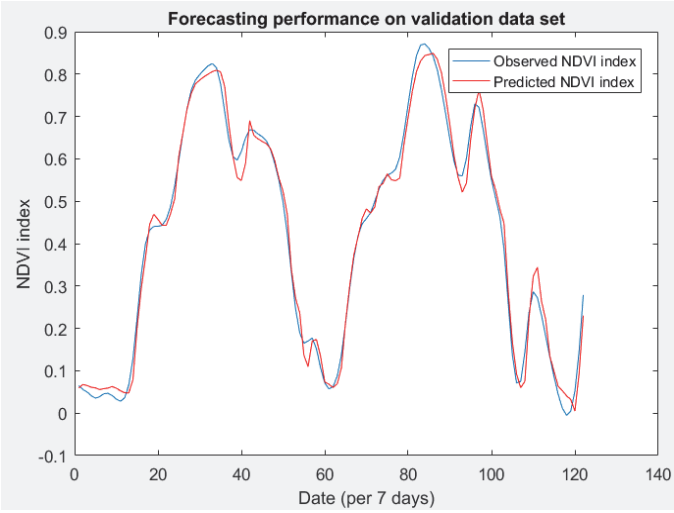


Fig. 6. The observed and forecasted time series data on a validation data set.

The observed and forecasted values of the NDVI time series on a testing data set are shown in Fig. 7.

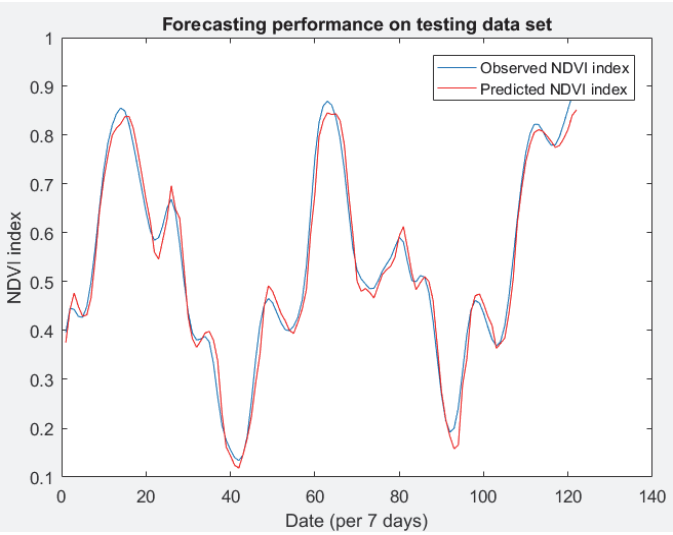


Fig. 7. The observed and forecasted time series data on a testing data set.

From Figs. 6 and 7 it is seen that there is an adequate reproducibility between the forecasted and observed time series.

## IX. Conclusion

In this paper, one-step-ahead prediction of the normalised difference vegetation index (NDVI) data has been obtained using a discrete time, continuous state second order Markov chain that is proposed as an alternative to the discrete time, discrete state first order Markov chain that was used in the authors' previous research. In order to improve the forecasting accuracy, according to the aim of the experiment, stepwise regression as a feature selection method and principal component analysis as a feature extraction method have been used in this study.

Markov chain states have been formed in this study by real numbers, and the transition probabilities evaluated from the data in online regime for every given actual state separately and not using some previously calculated probability distribution, such as a probability transition matrix. It helps make Markov chain adaptive. 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. However, these probabilistic forecasts have shown an acceptable forecasting accuracy with RMSE error 0.0277, MAPE error 4.9372 %, directional symmetry 87.6033 % and the adjusted coefficient of multiple determination 0.9803 on a testing data set.

## References

[1] Y. Wang (Ed.), *Remote Sensing of Coastal Environments* (Remote Sensing Applications Series), CRC Press, 2009, p. 41. https://doi.org/10.1201/9781420094428

[2] E. Sahebjalal and K. Dashtekian, "Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods," *African Journal of Agricultural Research*, vol. 8, no. 37, pp. 4614–4622, Sept. 2013. https://doi.org/10.5897/AJAR11.1825

[3] A. Stepchenko and J. Chizhov, "Applying Markov Chains for NDVI Time Series Forecasting of Latvian Regions," *Information Technology and Management Science*, vol. 8, no. 1, pp. 57–61, Dec. 2016. https://doi.org/10.1515/itms-2015-0009
[4] D. R Anderson, D. J. Sweeney, T. A. Williams, J. D. Camm and J. J. Cochran, *Quantitative Methods for Business*, 13th ed. South-Western College Pub, 2015, p. 771.
[5] Z. Peng, C. Bao, Y. Zhao, H. Yi, L. Xia, H. Shen and F. Chen, "Weighted Markov chains for forecasting and analysis in Incidence of infectious diseases in jiangsu Province, China,", *Journal of Biomedical Research*, vol. 24, no. 3, pp. 207–214, May 2010. https://doi.org/10.1016/S1674-8301(10)60030-9
[6] 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. https://doi.org/10.5539/mas.v4n5p162
[7] University of Natural Resources and Life Sciences, Vienna. Data service platform for MODIS Vegetation Indices time series processing at BOKU, Vienna. [Online]. Available: http://ivfl-info.boku.ac.at/ [Accessed: Aug. 25, 2015].
[8] 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," in *Proc. SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III*, 85380A, 2012. https://doi.org/10.1117/12.974857
[9] W. Watthayu, "Loopy Belief Propagation: Bayesian Networks for Multi-Criteria Decision Making (MCDM)," *International Journal of Hybrid Information Technology*, vol. 2, no. 2, Apr. 2009.
[10] H. Kantz, D. Holstein, M. Ragwitz and N. K. Vitanov, "Markov chain model for turbulent wind speed data," *Physicsa A: Statistical Mechanics and its Applications*, vol. 342, issue 1–2, pp. 315–321, Oct. 2004. https://doi.org/10.1016/S0378-4371(04)00488-1
[11] 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.
[12] B. Klikova and A. Raidl, "Reconstruction of Phase Space of Dynamical Systems Using Method of Time Delay," *WDS'11 Proceedings of Contributed Papers*, Part III, pp. 83–87, 2011.
[13] M. Templ, A. Kowarik and P. Filzmoser, "Iterative stepwise regression imputation using standard and robust methods," *Journal of Computational Statistics and Data Analysis*, vol. 55, no. 10, pp. 2793–2806, Oct. 2013. https://doi.org/10.1016/j.csda.2011.04.012
[14] J. M. Saleh and B. S. Hoyle, "Improved Neural Network Performance Using Principal Component Analysis on Matlab," *International Journal of The Computer, the Internet and Management*, vol. 16, no. 2, pp. 1–8, 2008.
[15] P. Ng'andwe, J. Mwitwa and A. Muimba-Kankolongo, *Forest Policy, Economics, and Markets in Zambia*, Academic Press, 2015, pp. 6–7.

**Artūrs Stepčenko** is a third-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

**Jurijs Čižovs** 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