Nurtay Maksat

MKM – 203M

application: [egistic.kz](https://egistic.kz/)

**NDVI index forecasting**

***Abstract***

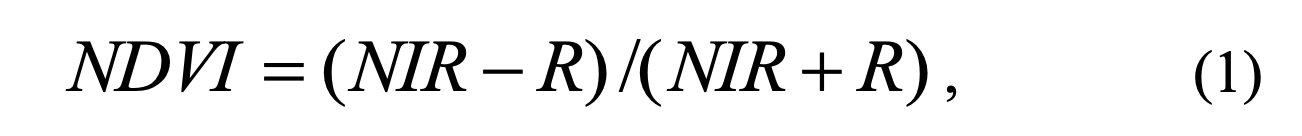
Earth observation time series dependent vegetation forecasts are information on vegetation changes. NDVI is a key parameter for predicting short-term vegetation and managing various issues such as forest fire and forest disease propagation. The computing model and universal approximators are Artificial Neural Networks (ANN), commonly used to model and predict nonlinear, non-stationary and complex processes.

***Introduction***

Ecosystems, like natural plants, are the subjects of human activity. The shift in coverage of vegetation is essential to the state and operation of the ecosystem. Changing the coverage of plants will have a long-term impact on sustainable food supply, the environment and human health of freshwater and timber resources. To provide knowledge about vegetation stability, monitoring and predicting shifts in vegetation coverage at periodic intervals are very significant.

In the development of land coverage in vast geographical areas, the use of remote satellite sensing data as an economical strategy has been used extensively. A significant part of ground cover is vegetation cover. Due to repeated broad range, short revision cycles and good image quality, the identification of changes has become an outdated application of distant sensed data. Change detection is the mechanism by which an entity or event identifies changes in the status by analyzing it at various times. The key requirement for remote sensing data for identification of changes in vegetation is that changes in land cover result in changes in radiance values and in the radiance attributed to changes in the land cover are considerable in terms of radiance variations induced by other elements, such as atmospheric variations, soil moisture differences and sun angle differences.

For tracking time shift related to vegetation, vegetation indexes determined from the satellite images may be used. Vegetation indices (VI's) are surface reflectance combinations that are designed to derive a particular vegetation property. Each of the VI's is designed to accent a specific vegetation property. Vegetation analysis using distant sensed data involves knowledge of the vegetation structure and operation and its properties of reflection. This expertise allows vegetative structures and their state to be associated with their reflectance in an environmentally friendly scheme. The NDVI is structured to estimate cover from the reflective satellite data bands. The NDVI is the normalized difference vegetation index. The NDVI is an indication for the quantity of green vegetation. Previous experiments have shown the ability to use NDVI to explore dynamics of vegetation. The index of NDVI is:



where NIR reflects the near-infraround reflectance and where R represents the red band of the satellite pictures. Low red reflexity and high infraroad reflectance and thus high NDVI values are seen in greener and denser vegetation. The NDVI values are normalized between -1 and +1, with rising positive values indicating increasing green vegetation and decreasing positive values indicating non-vegetated surface features such as water, barren soil, rock, ice, snow, clouds, or artificial materials. The NDVI can also suppress natural noise causes, such as topographical effects and changes in the sun's angle.

***Study Area and Character of the Data***

*Study Area*: Esil district, Akmola region, the region is located in the northern part of Kazakhstan, on a total area of 9649.18 hectares. Approximately the test area fully is covered by agricultural lands.

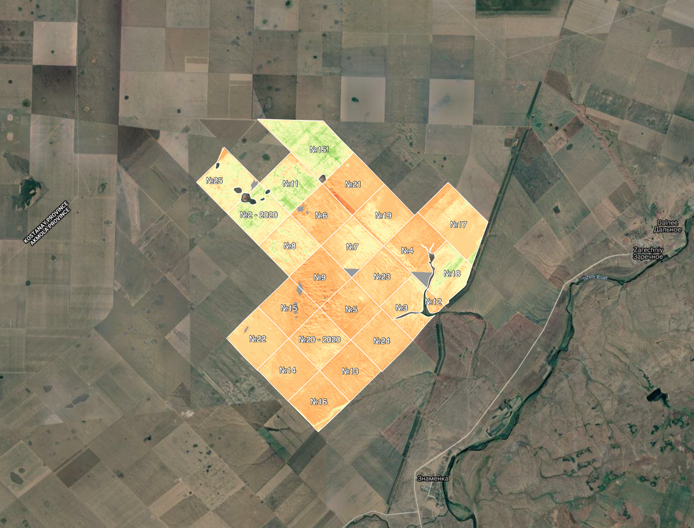
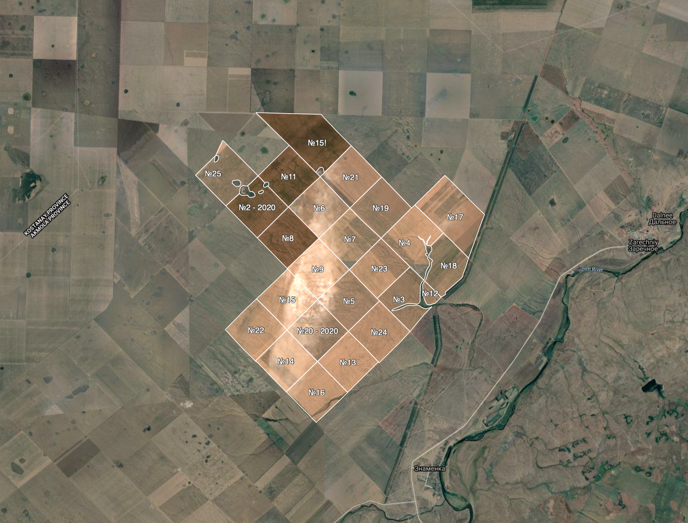
 

Figure 1. Study area

*The train Data Set*: Used multi-temporal composite NDVI information obtained from a 250-m-space-resolution MODIS Terra (NASA research satellite) and generated at 7-day intervals. Data from MODIS Vegetation Indices time series computing data service network is collected. Use of the Whittaker lightening algorithm with smoothing λ=15 parameter and two filtering iterations for smooth and gap filling used results. Iterative filtering has been used, so the NDVI observed is decreased by undetected clouds and bad air quality.

The NDVI data collection consists of 814 NDVI images which have been collected over 20 years every 7 days. For corresponding test site, the NDVI values of these images were collected and used as time series for NDVI.

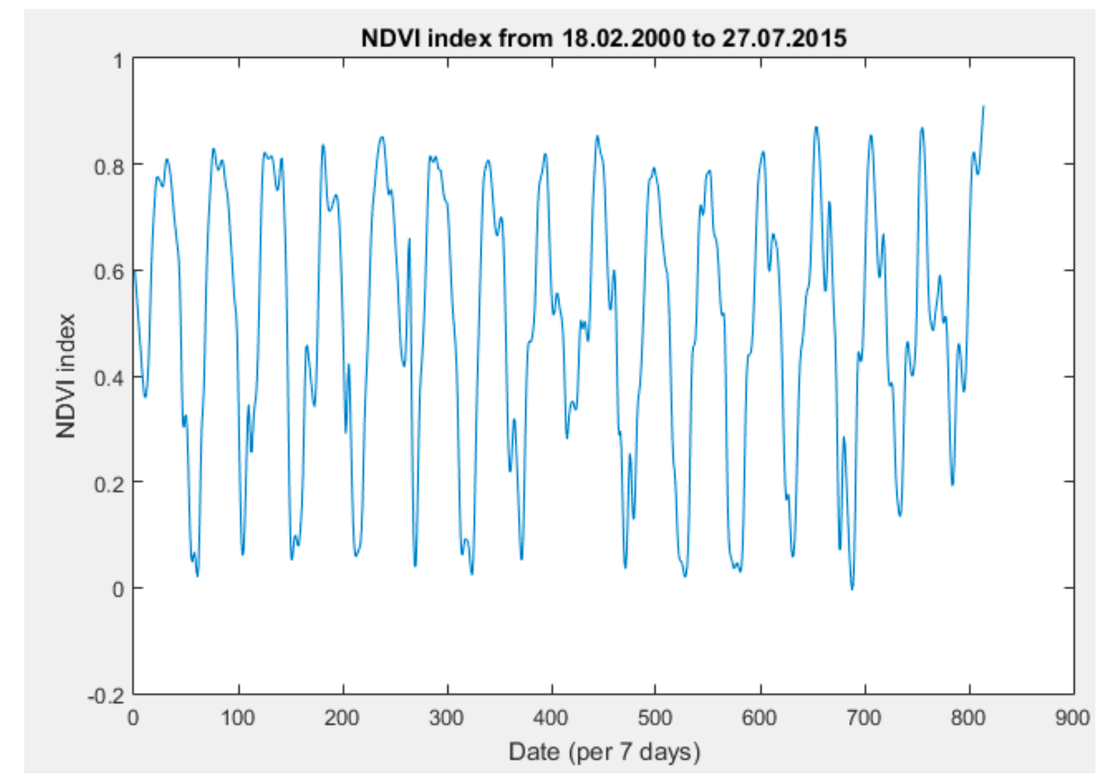


Figure 2. NDVI index from 2000 to 2015.

The data from NDVI time series include a seasonal pathway - time series indicate clear seasonal oscillations that represent the phenological vegetation cycles in which maximum values from NDVI can be observed from May to August. The NDVI values vary from -0.0050 to 0.9109 units. Trends in NDVI can, but may not always be monotonic. For instance, a positive pattern will turn into a negative and the other way around.

***Artificial Neural Networks (ANN)***

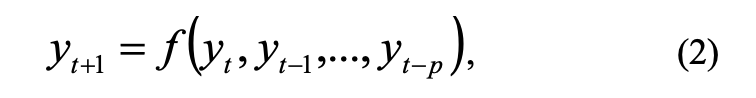
Artificial neural networks (ANNs) are an intelligence type that attempts to imitate the role of real neurons located in the brain of the human being. ANN's are one of the most reliable and used predictive models used in social, economic, industry, technology, currency, stock and other predictions. The artificial neural network structures make them useful for a successful predictive role.

In contrast to conventional model-based mathematical and computational methods like regression and Box-Jenkins’s techniques, which include prior knowledge of the structure of the data relationships, artificial neuronal networks are self-adaptive methods that learn from data, and few assumptions of the problem are necessary in advance.

Neural networks learn from examples and can find functional associations between data even though relationships are uncertain. ANNs are thus suitable for issues whose solutions require information that is difficult to specify, but for which sufficient data or observations are available.

Artificial neural networks may widespread. ANN will always accurately process an early unsightly result after learning the input data, even though the sample data is sounding. Neural networks are less vulnerable than most other approaches to error term predictions and more able to handle noise and noisy elements. Even universal function approximators are artificial neural networks. A neural network has been shown to be able to accurately approximate any continuous function.

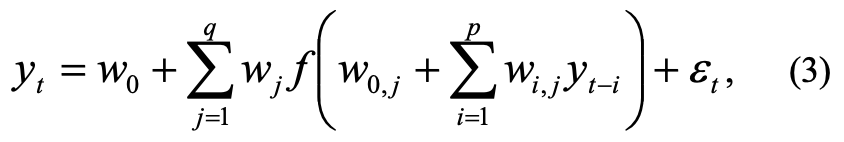
For a time series forecasting problem, a training patterns consists of a history data with fixed number of observations. If time series contains N observations y1, y2, ..., yN, then using an ANN with n input nodes, we have N-n training patterns than can be used for short-term forecasting – one value ahead. The first training pattern will be contain y1, y2, ..., yn as inputs and yn+1 as the output. The second training pattern will contain y2, y3, ..., yn+1 as inputs and yn+2 as the output. The last training pattern will be contain yN-n, yN-n+1, ..., yN-1 inputs and yN as the output. Then pattern yN-n+1, yN-n+2, ..., yN will be used to get forecasting value yN+1. The unknown purpose mapping is carried out by the ANN, where yt is the observation at time t:



ANN's structure include input data and artificial neurons that are known as “units”. Input layer, output layer and one or two intermediate layers called hidden layers are used in the multilevel perceptron. The size and nature of the data set affect the number of hidden layers and neurons within each layer. ANN usually performs better than neural networks with a wide number of hidden layers with one or two hidden layers.

The scalar weight and network architecture store the information and power of the interconnected neurons in a qualified network. If the weight value is zero, then the interaction between the two neurons is prohibitive and if the weight value is negative. The weighted inputs from previous layers are received by an individual processing element and added to the combined operation in each node, a bi-neuron attached to each hidden or output device.

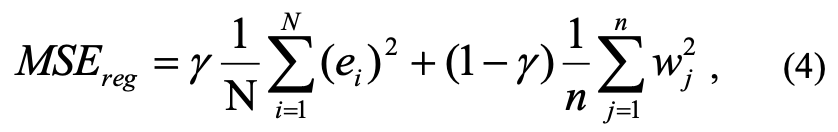
This cumulative total is transferred to the processing element's nodal output, which is weighed and transferred to the processing element of the next layer. The functions of mixture and transition are the activation function together. In most instances, neurons in the input layer don't have an activation mechanism, so the inputs are sent to the hidden layer. As non-linear activation function, the most common activation function for the output layer is linear, which can lead to distortion of the predicted output. Sometimes used as the function of hidden layer transition are the sigmoid (logistic), exponential (hyperbolic), quadratic or linear functions. The relationship between output (yt) and inputs – past time series observations (yt-1, ..., yt-p) is:



where wj are weights between hidden and output layer, wi,j are weights between input and hidden layer, f is an activation function, q is the number of hidden nodes, p is the number of input nodes, ɛt is random error at time t.

The back propagation algorithm of Levenberg-Marqardt with Bayesian regularization is a neural network training tool, which updates weight and weight values according to optimization of Levenberg-Marquardt. The combination of square errors and weights is minimized and the right combination is established to create a well-generating network.

The objective of neural network training is to reduce the global error determined by performance function. The following performance (cost) function is used for Bayesian regularization:



where γ is the performance ratio, e is the error vector, w is the weight and bias variable vector. Minimizing performance function (4) will cause the network to have smaller weights and biases, and this will force the network response to be smoother and less likely to overfit.

***A Layer Recurrent Neural Network***

A recurrent neural network (RNN) is a class of artificial neural networks, where a direct cycle is formed by interactions between units. This provides an internal network state that allows the network to display complex temporal activity. Recurrent neural networks can use their internal memory to process arbitrary sequences of inputs. Therefore, recurrent neural networks are powerful sequence learners.

The layer recurrent network (LRN) is a recurrent dynamic neural network constructed from the previously established Elman neural network. There are feedback loops on the recurring layer network on each layer except the output layer. Feedback on the recurring network of the layer is the relation between the outputs of the hidden layer of neurons and neurons in the context layer which stores hidden delayed layer outputs. The main benefit of the LRN is the strong ability to retrieve functionality, and in the past store valuable data points in the background layer.

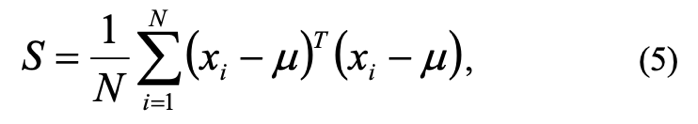
The LRN extends the Elman network to include an infinite number of layers and arbitrary transfer functions on any layer. Exact copies of the basic back propagation algorithm are available to train LRN. An approach to the back propagation algorithm has been used for the original Elman network.

***Stepwise Regression***

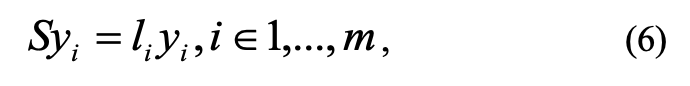
Stepwise regression is a sequential functional sorting process primarily designed for smaller-field fitting where quantitative variables are selected using an automated algorithm. A single separate variable that has the biggest absolute t-test value is the initial stepwise model of regression. T-tests are used to decide whether two data sets differ substantially. A second variable is introduced in the following step, and a new model is developed. If the new model's t-test values are higher than the first model's, the first model is retained, and a third variable is introduced. If the new model outperforms the first, the first variable is discarded, the second variable is retained, and the next model with the second and third variables is formed. This process is repeated until all two variable combinations have been checked, at which point the best performing two variable combination is chosen as the final model before a third variable is introduced. When all significant variables are included in the model, the process is complete.

***Principal Component Analysis (PCA)***

Principal component analysis (PCA) is a statistical attribute extraction approach that employs an orthogonal transformation to turn a series of potentially clustered observations into a collection of values of linearly uncorrelated variables known as principal components. The PCA technique is essentially concerned with locating linear transformations, y1,y2,y3,...,ym, of the original components, x1,x2,x3,...,xp, that have the property of being linearly uncorrelated. The dimension of the initial data set is denoted by p. The y components are selected such that the first principal component y1 has the most variance, the second principal component y2 has the second most variance and is uncorrelated with y1, and so on. As a result, the aim of PCA is to identify a collection of orthogonal components that minimize error in the reconstructed results. The PCA algorithm's first step is to normalize the components such that they have a zero mean and a unity variance. The principal components of the normalized components are then computed using an orthogonalization process. The principal components are orthogonal since they are the eigenvectors of the symmetric sample covariance matrix. The sample covariance matrix is calculated as follows:



where xi is i-th original observation vector (component), μ is the sample mean and N is the number of samples, so that:



where li is i-th largest eigenvalue of S and m is the number of principal components. The PCA method also can be used for input data dimensionality reduction.

***Results***

In several experiments were found that optimal number of hidden nodes is 22. Optimal LRN topology is shown in Fig. 3.

LRN convergence for best model was obtained after 42 epochs (Fig. 4).

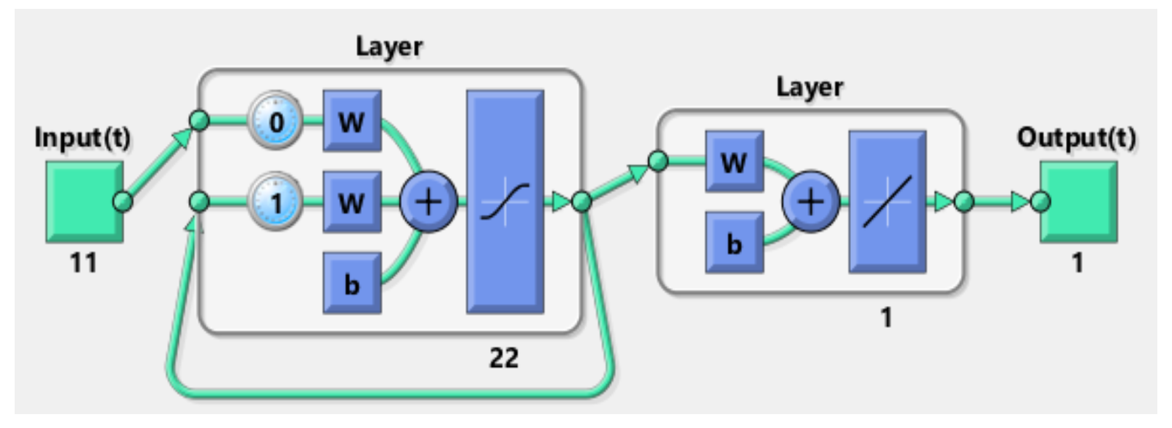


Figure 3. Optimal LRN topology.

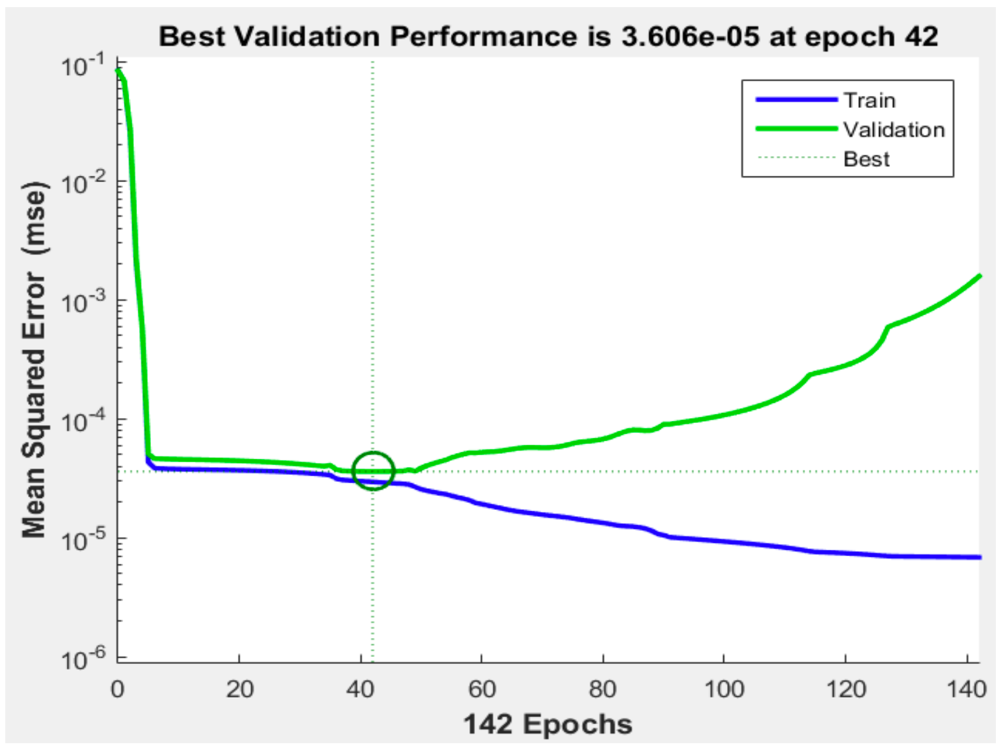


Figure 4. LRN convergence.

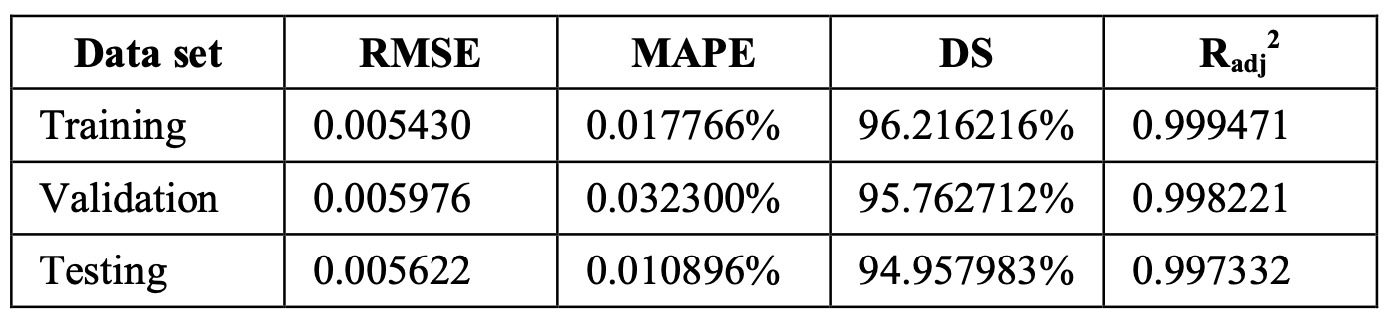


Table 1. FORECASTING PERFORMANCE

The trial data set had the smallest RMSE and MAPE defects, the validation data set had the best directional symmetry, and the training data set had the best modified coefficient of multiple determination. These findings demonstrated the good performance of a regularized layer recurrent neural network because the train set errors, validation set errors, and test set errors all have common characteristics, and no significant overfitting seems to have occurred. Figure 5 depicts the actual and forecast values of the NDVI time series on the training data set.

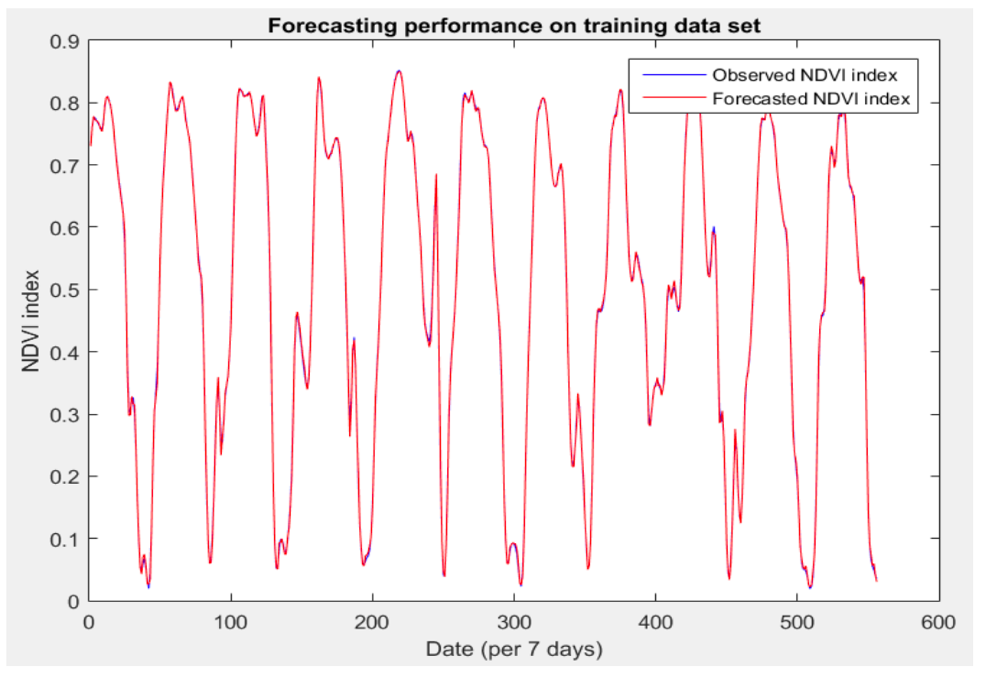


Figure 5. Forecasting performance on training data set

A layer recurrent neural network is used to make one-step-ahead predictions of the normalized gap vegetation index (NDVI) (LRN). The role of recurrent feedback in a neural network is a positive influence in NDVI time series forecasting. This is due to the recurrent neural network having a "deeper memory" than other types of neural networks. The study concludes that the forecasting abilities of a regularized LRN combined with stepwise regression and principal component analysis (PCA) provide a theoretically very useful tool for forecasting NDVI time series.