

AUTOMATED PREDICTION SYSTEM FOR VEGETATION COVER BASED ON MODIS- NDVI SATELLITE DATA AND NEURAL NETWORKS

Sohaib K. M. Abujayyab^{1*}, İsmail Rakıp Karaş²

^{1*} Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk, Turkey - s.jayyab@hotmail.com

² Dept. of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk, Turkey - irkaras@gmail.com

Commission IV

KEY WORDS: Automated Systems, Prediction, NDVI, MODIS, Neural Networks, Early Warning.

ABSTRACT:

Around the world, vegetation cover functioning as shelter to wildlife, clean water, food security as well as treat large part of air pollution problem. Accurate predictive data early warn and provide knowledge for decision makers to reduce the effects of changes in vegetation cover. In this paper, an automated prediction system was developed to forecast vegetation cover. Prediction system based on moderate satellite data spatial resolution and global coverage data. The tools of system automate processing Moderate Resolution Imaging Spectroradiometer (MODIS) images and training neural networks (NN) model based on 60,000 observations to forecast future density of Normalized Difference Vegetation Index (NDVI). Zonguldak data, located in north of Turkey as dense vegetation cover area utilized as case study for system application. This system significantly facilitates predictive process for users than previous long and complex models.

1. INTRODUCTION

Predicting and mapping vegetation cover are substantial application area of interest in the last decades. Vegetation cover significantly affects stabilizing of food systems, food security as well (Lesk, Rowhani, and Ramankutty 2016). Vegetation cover changes is vital variable that influence biological system condition and capacity. Changes in vegetation cover may have long haul effects on freshwater, human welfare, sustainable food creation, climate and atmosphere system (Stepchenko and Chizhov 2015). Giving chiefs precise data about vegetation cover wellbeing expands readiness ability of systems to get human ready for and adjustments for resources shortage (Ziervogel et al. 2010). Accurate vegetation cover data can be utilizing to recognize former circumstances as well as guide efforts to increase resilience to future conditions (Burchfield and Gilligan 2018). In addition, to examine vegetation cover productivity and make action plan (Hira and Deshpande 2015; Shrestha et al. 2017).

Over recent years, production of accurate vegetation cover data has been increased in parallel of the increment in the availability of high temporal, wide-coverage remotely sensed data (Hira and Deshpande 2015; Udelhoven 2012). Open and free distributions policy from space agencies for remote sensing data has been provide plenty of data. Satellite images empowers accurate and cost effective monitoring at recurrent time intervals over enormous zones in the earth (Verbesselt, Zeileis, and Herold 2012). For long time, land vegetation cover observing systems has been utilized satellite images as instruments to give timely and concise data about seasonal growth of land vegetation cover (Rembold et al. 2015). At regional and national dimension, satellite images can be exceptionally beneficial in early warning of land vegetation cover seasonal growth and in estimating future land vegetation cover.

The MODIS NDVI product combination, with 250m spatial resolution from Terra sensor, and 16-day (MOD13Q1) was gained massive reputation in mapping vegetation cover area

(Colditz, Ressler, and Bonilla-Moheno 2015). MODIS is moderate resolution imaging spectroradiometer, while NDVI is normalized difference vegetation index. NDVI is a standout amongst most largely recognized indices for vegetation cover. NDVI mathematically display vegetation greenness density (Reddy and Prasad 2018; Wang et al. 2016). NDVI can be extract through: -

$$NDVI = (NIR - R) / (NIR + R)$$

Where: -

(NIR) is the spectral reflectance in near infrared band in MODIS satellite images

(R) is the spectral reflectance in red band.

Usually, NDVI data are standardized among (- 1 to + 1). Occasionally, values standardized between -2000 to +1000. Upper values of NDVI point toward greenery and intensive vegetation cover. Negative values in NDVI images represent non-green areas, for instance barren lands, snow, clouds, urban areas, water, or artificial materials. The extracted knowledge from the annual profile of NDVI data clearly describe the density status of green vegetation cover. Annual profile displays when the plants reach the uppermost green density due to the suitable conditions, for example availability of sunshine and water. Furthermore, annual profile shows lowermost green density due to leaf senescence, plant death, water shortage, temperature and harvest. This process simply represents the seasonality issue in time series analysis.

Recently, instead of performing satellite image analysis and execution prediction at one-time, a dense long NDVI time series images analysis had turned into a significant option in land vegetation cover prediction (Djeddaoui et al. 2017). On a national level, high temporal NDVI time series from MODIS imagery can, providing superior information about vegetation stability (Djeddaoui et al. 2017; Udelhoven 2012).

* Corresponding author

In other hand, NDVI time series analysis has been accomplished using several statistical and other methods. Univariate autoregressive integrated moving average (ARIMA) method is commonly applied for a univariate NDVI time series prediction (Stepchenko and Chizhov 2015). Nevertheless, these methods founded based on supposition that predicted time series has been stationary and linear. But in NDVI time series prediction case, difficulty arises due to data are not linear and not stationary. Various earlier investigations suggest that nonlinear machine learning methodologies, for example, neural networks NN method accomplish superior performance compared with conventional time series linear methods (Li, Wang, and Li 2015; Menezes and Barreto 2008; Reddy and Prasad 2018; Udelhoven 2012). Therefore, neural networks method employed as new possibility in MODIS NDVI time series prediction instead of conventional statistical prediction techniques (Menezes and Barreto 2008; Stepchenko 2016).

In order to implement spatial prediction based on NDVI time series, researchers propose several frameworks and procedures to mapping the future NDVI. Stepchenko applied several processing steps to implement NDVI short-term forecasting using recurrent neural networks (Stepchenko and Chizhov 2015). Reddy applied long short term memory neural network to forecasting vegetation dynamics using NDVI time series data (Reddy and Prasad 2018). Nay engaged machine learning to estimating remotely sensed vegetation health (Burchfield and Gilligan 2018).

However, precise forecast of NDVI time series is as yet confront several challenges, principally in light of the fact that it involves long and complicated geo-processing tasks until producing vegetation cover maps. The complexity coming from the truth that analyzer need to comprehend wide range of tools and methods from geospatial area and machine learning area in order to carry out prediction. In addition, non-computer specialists facing difficulty to apply NN in geospatial area, which is big barrier for researchers. Furthermore, user could be fall under several processing or programing errors and be timely consuming process. Decision makers suffering from additional restrictions during the analysis, for instance reading spatial information, construction learning dataset from spatial information, big data problem and not easy repeatable models.

In this paper, we present an automated prediction system based on MODIS- NDVI satellite data for vegetation cover using neural networks. The aim to produce a user-friendly predictive system that will increase scarce information capacity for vegetation cover systems. The system designed for ArcGIS and Matlab users with basic background.

2. AUTOMATED PREDICTION SYSTEM

The automated system provide platform for future vegetation cover mapping. The system extracts the knowledge from NDVI time series based on the pixel strategy and time first-location after. The created system tools forecasts future NDVI cover at 250-meter spatial resolution at 16-day intervals. System utilizing ArcGIS-Python-Matlab environments. System based on. All the system scripts and tools can be obtained from (<https://github.com/SohaibAbujayyab>). The system employed neural networks to training the global data coverage (MODIS/Terra Vegetation satellite data) that are available in USGS's server (<https://search.earthdata.nasa.gov/data>).

The system (Figure 1) tools can be used manually in ArcGIS-ArcToolbox or through ArcGIS-Model Builder. The tools displayed alphabetically within ArcToolbox. The system consisting nine sub tools, which demonstrated between Figure 2 and Figure 11. The system tools concisely explained as following: -

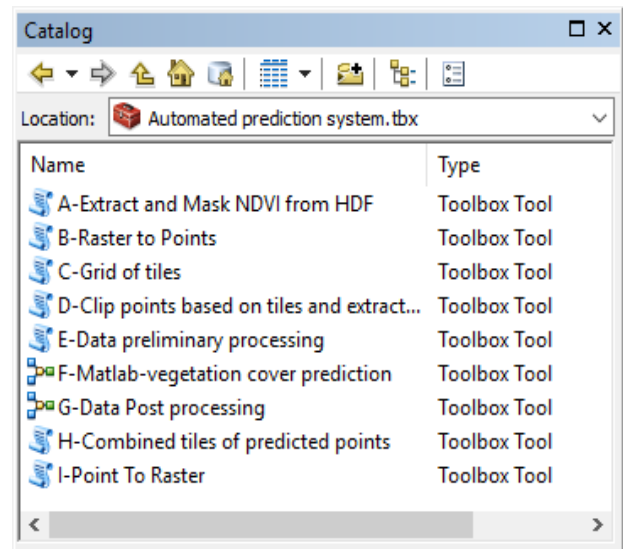


Figure 1 Automated prediction system for vegetation cover based on MODIS- NDVI satellite data and neural networks

(A-Extract NDVI from HDF) Since that prediction system based on pixel time series strategy, then, this work commonly consisting hundreds of images. thus, manually process will be exhausting and time-consuming. The objective of this tool is to automating extracting NDVI images from HDF file combination that can be download from USGS website. In order to manage the big number of HDF files, the tool read the files one by one from input folder without the need to added it together in the tools, which reduce memory consumption. Each historical series of images in same spatial tile must be in one separated folder.

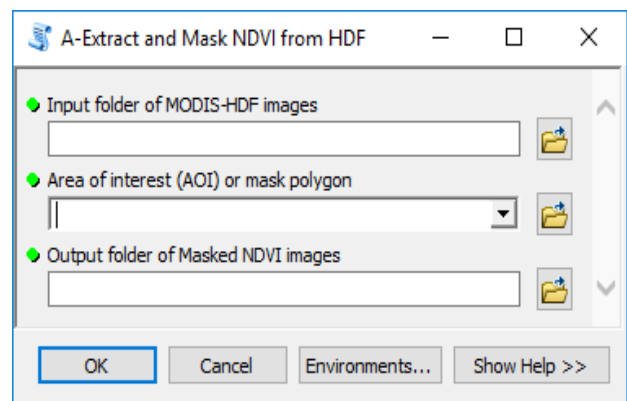


Figure 2 Phase A (Extract and Mask NDVI from HDF)

Additionally, full NDVI images saved temporally in the “in_memory” to reduce processing cost of storing full NDVI images. Once more, due the hundreds of images issue, the full NDVI images in “in_memory” masked to limit time series only in area of interest AOI, which will lead to reductions in processing time. Finally, the tool stores the output masked NDVI images in destination folder. The tool stage shown in Figure 2.

(B-Raster to Point) this tool is normal tools from ArcToolbox. This tool converting any masked raster layer within AOI to point layer. The tool accurately represents the pixel sites by grid of points. Grid of points is required for the next stages in order to construct time series NN datasets. The tool stage demonstrated in Figure 3.

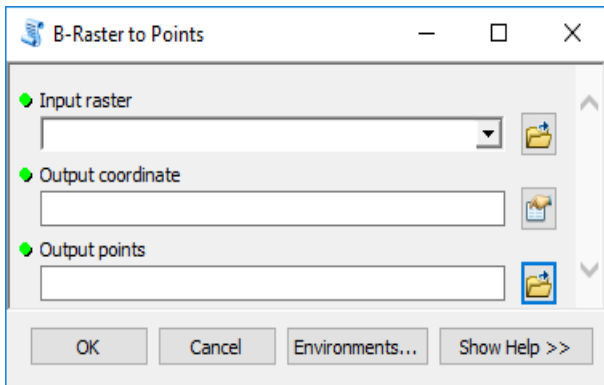


Figure 3 Phase B (Raster to Points)

(C-Grid of tiles) This tool generating polygon feature layer for set of tiles. The purpose is to divided AOI to equal squares to avoid high memory consumption problems in case of low device capability. Tiles here represent the workers in multiprocessing concept. The tools need only to be feed by the AOI layer, input coordinate system, tiles width, tiles height and where the store the output grid. The tool needs to be fed with the study area border for masking as well as the coordinate system. The tool stage displayed in Figure 4.

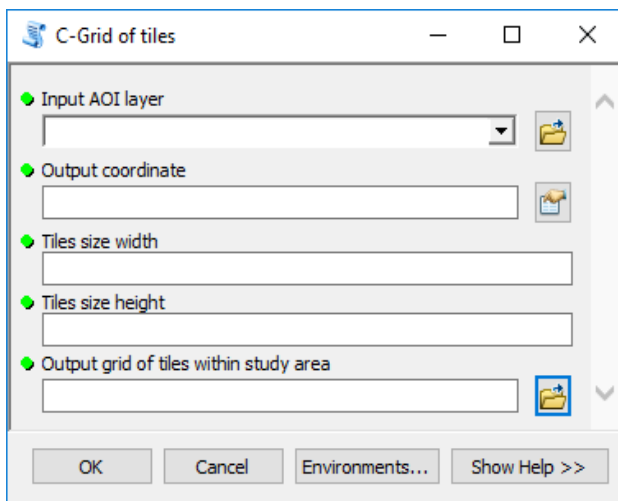


Figure 4 Phase C (Grid of tiles)

(D-Clip points based on tiles and extract NDVI values) This tool shares the same concept of memory management with previous stage. The inputs are grid of points, and grid of tiles. Tool parameters applied filter to input data and allow only for suitable geometric layers. Tools iterate over the tiles in the grid through (arcpy.da.SearchCursor) function to clip points for each tile. Then, to extract the attribute data for each point from the multi NDVI images. By this methodology the tools avoided memory limitation.

Finally, the output of this stage are clipped points tiles files with number of fields that identical with number of input NDVI images. The tool stage illustrated in Figure 5. The output here must save as ESRI-Shapefile format in order to satisfy the Matlab analysis requirements.

(E-Data preliminary processing) this tool generated to remove outlier or missing data points from database. The tool first iterate over the input files to make list of input tile files. Second tool iterate over their fields to generate dynamic SQL expression for selection from all fields. Then tool find all records that out of normal NDVI range values. At this point the

tool remove the outlier points through (arcpy.DeleteFeatures_management) function.

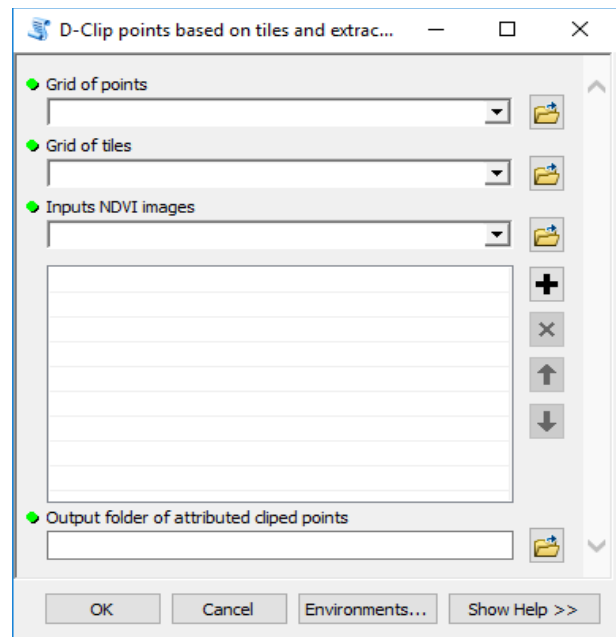


Figure 5 Phase D (Clip points based on tiles and extract NDVI)

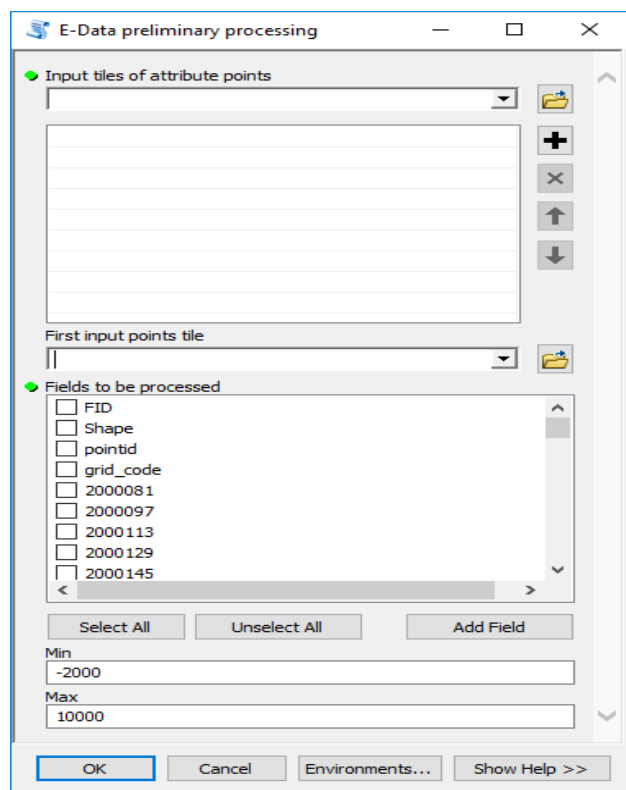


Figure 6 Phase E (Data preliminary processing)

The end-user begins by adding the points files for all tiles, then defining all historical NDVI fields except original required or non-relevant fields in the attribute table. The output here must save as ESRI-Shapefile format in order to fulfil the Matlab analysis requirements. This tool allows the user to migrate the spatial data to Matlab environment with suitable data structure for NN training and testing. The tool stage illustrated in Figure 6.

```

1 Begin
2
3 Make list of input tiles of Shapefiles(SHP) files
4 Find the size of list of SHP files
5
6 for each file out of points tiles SHP files
7   Read SHP of poits tiles
8   Find the size of NDVI matrix
9   for each field in points
10    for each value in field values (Time-Series)
11      finding the max values of field
12      finding the max min of field
13      normalizing NDVI values=(each value in field-minn)/(maxx-minn)
14    end
15  end
16
17 for each pixel point (Time-Series)
18   constructing target data
19   Defining training function = 'trainlm'
20   Create a Nonlinear Autoregressive Network
21   Prepare the Data for Training and Simulation
22   Setup Division of Data for Training, Validation, Testing
23   Train the Network
24   Test the Network
25   creating Closed Loop Network to do multi-step prediction.
26   Multi-step Prediction
27   store performance values
28 end
29
30 Generating new matrix to store predicted NDVI values
31
32 for each record of points in predicted matrix
33   Copy predicted NDVI values to the new matrix
34 end
35
36 for each record of points in predicted matrix
37   Add X and Y coordinates of point to new NDVI matrix from original SHP
38 end
39 Store predicted NDVI matrix to xls file
40 end
    
```

Figure 7 Pseudocode of prediction phase

(F-Matlab vegetation cover prediction) this is the core stage in this system, which responsible to execute NDVI time series training using NN, then accomplish vegetation cover prediction. This is the only tool out of this system tools that need to be implemented in Matlab environment. The system tries to benefit from the powerful tools in Matlab environment. The code simply and concisely described through pseudocode in Figure 7. The tool starts by reading the Shapefiles-tiles that consisting the NDVI historical data and were prepared in the previous phases. The code release simple GUI to select the input Shapefiles-tiles. During this stage, the tool iterates through the Shapefiles-tiles. Sub iteration performed to constructing the time series data set from Shapefiles-tiles in order to achieve NN training. Then numeric values in each time series dataset normalize between (0 and 1). Normalized values in computed utilizing Eq 1.

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad \text{Eq 1}$$

Where;

- x_{norm} =normalized values
- x_{min} = minimum value of x_i
- x_{max} =maximum value of x_i

Again, sub iteration iterates over time series to develops and test NN model, then predict vegetation cover. NN structures

displayed in Figure 8. But the popup windows of neural network training were disabled to reduce the time of processing (net.trainParam.showWindow = 0).

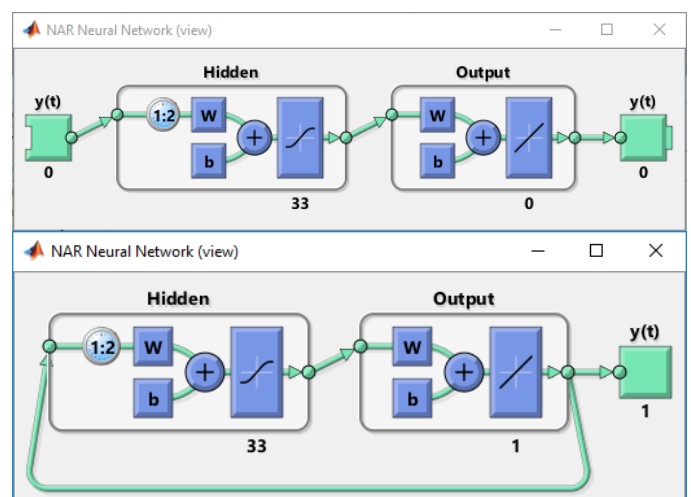


Figure 8 NN structure

The functionality of third and fourth sub iteration are to store predicted values with their

coordinates in XLS files.

The user can define the suitable training algorithm, number of neurons, number of multi prediction step ahead and divide the data for training and testing. In addition, since that the code read the coordinates for predicted vegetation matrix from the original input Shapefiles-tiles, input files to system are flexible and not required to following certain order from beginning. Additionally, the tool stores the MSE for each time series in the output matrix, which can be used to geographically analysis the NN performance.

(G-Data Post processing) this tool as illustrated in Figure 9 reading Excel files of tiles with its predicted data that formerly produced in Matlab. Tool iterate over the Excel files, then positioning them as points layer and store them as separate Shapefiles-tiles. The user here must use (%inline variable substitution%) for file naming. The output here is points layers for tiles with predicted vegetation data.

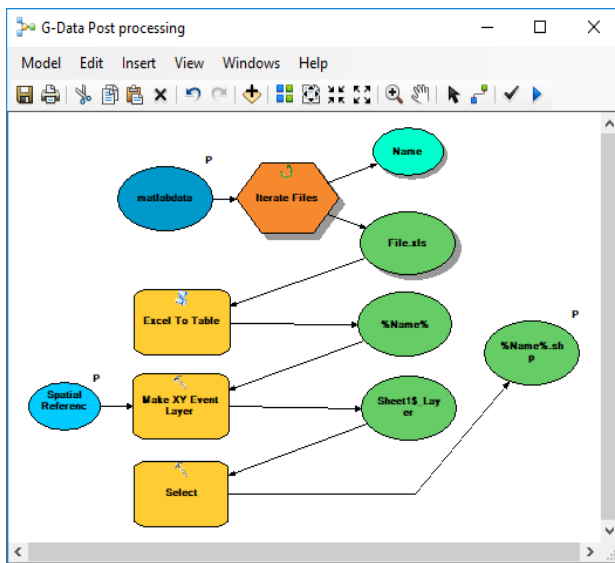


Figure 9 Phase G (Data Post processing)

(H-Combined tiles of predicted points) the purpose of this tool is simply merge points tiles from the previous stage to one points file. The tool combine all tiles to be in one feature class consisting predicted vegetation data. The tool stage illustrated in Figure 10.

(I-Point To Raster) this tool is the latest stage that interpolation raster grid from grid points. Interpolation used method is TopoToRaster method (PointElevation submethod) and based on the predicted vegetation data. The tool flexible to allow the user to accomplish either one or several interpolation and store them as raster grid automatically. The output raster grid files represent the predicted vegetation density in AOI. The tool stage illustrated in Figure 11.

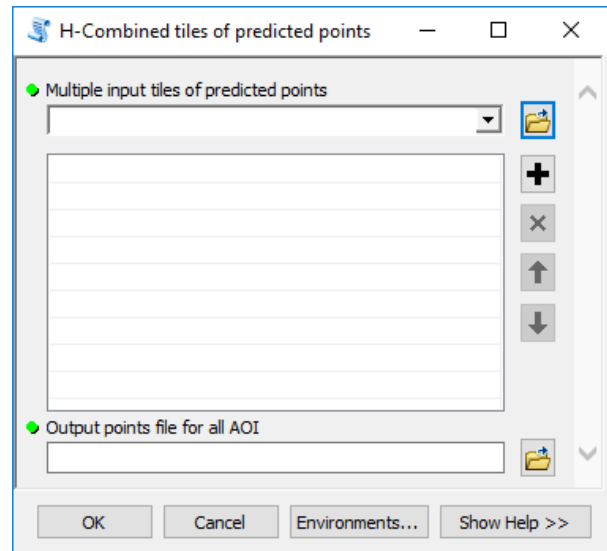


Figure 10 Phase H (Combined tiles of predicted points)

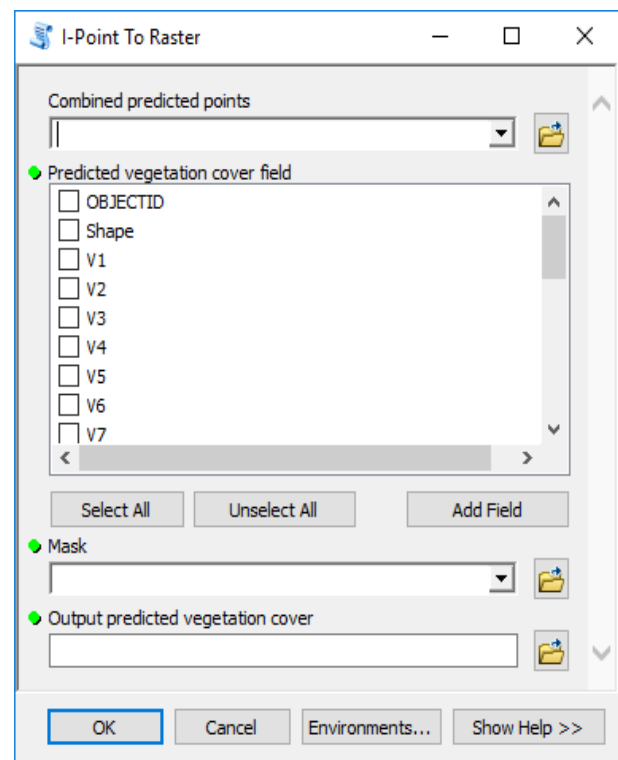


Figure 11 Phase I (Point To Raster)

3. SYSTEM APPLICATION

In this section, the prediction system tools were used to execute an analysis for case study. the case study here is Zonguldak province in north west of Turkey. It is located south of the black sea. Zonguldak is one of forestry province in Turkey. 300 historical MODIS images were downloaded from USGS server. The images data of Zonguldak area fed to the system and the system easily carry out the analysis until producing the vegetation cover map. 60,000 observations were in Zonguldak area fed to the system. The used tiles size were (25 km * 25 km), which effectively processing for all this amount of data. NN model learns potentially complex relationships among past remotely sensed data, then predict future vegetation cover.

Figure 12. illustrate three time step distribution of predicted vegetation cover and tiles of data.

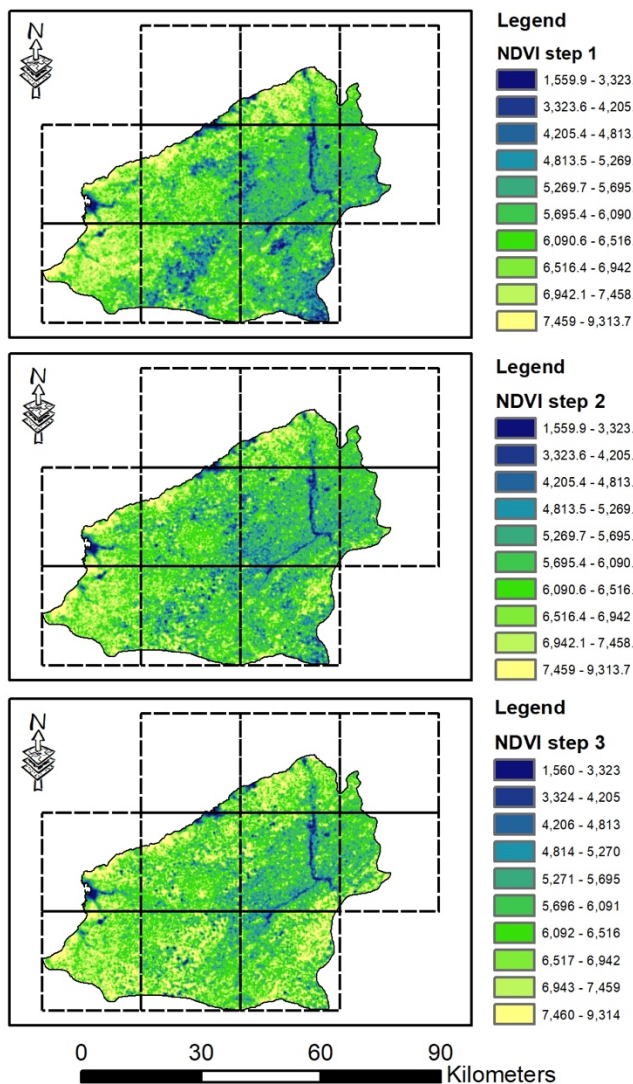


Figure 12 Predicted vegetation cover.

4. CONCLUSION

Vegetation cover has numerous advantages for the society as well as environment. Available vegetation cover information impressively support policy making for early warning, be prepare and respond to abnormal situation in future. speedily increasing awareness and decreasing negative influences on humans, natural and resources.

In this paper, we have created system to generate short-term vegetation cover prediction utilizing global data coverage of MODIS-NDVI. The system automates MODIS-NDVI data processing, training NN models based on historical observations and forecast future values of NDVI.

In conclusion, the system simplify the long and complex process for users. The developed system provide platform for non-GIS experts to easily perform very complex tasks without deep knowledge in geospatial data.

ACKNOWLEDGMENT

This study has been supported by 2221 – Fellowship Program of TUBITAK (The Scientific and Technological Research Council of Turkey). We are indebted for their supports.

REFERENCES

- Burchfield, Emily, and Jonathan Gilligan. 2018. “A Machine-Learning Approach to Forecasting Remotely Sensed Vegetation Health AU - Nay, John.” *International Journal of Remote Sensing* 39(6): 1800–1816. <https://doi.org/10.1080/01431161.2017.1410296>.
- Colditz, Rene R., Rainer A. Ressler, and Martha Bonilla-Moheno. 2015. “Trends in 15-Year MODIS NDVI Time Series for Mexico.” In 2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp), IEEE, 1–4. <http://ieeexplore.ieee.org/document/7245766/> (January 29, 2019).
- Djeddaoui, Farid et al. 2017. “Desertification Susceptibility Mapping Using Logistic Regression Analysis in the Djelfa Area, Algeria.” *Remote Sensing* 9(10): 1031. <http://www.mdpi.com/2072-4292/9/10/1031> (January 28, 2019).
- Hira, Swati, and P.S. Deshpande. 2015. “Data Analysis Using Multidimensional Modeling, Statistical Analysis and Data Mining on Agriculture Parameters.” *Procedia Computer Science* 54: 431–39. <https://www.sciencedirect.com/science/article/pii/S1877050915013745> (January 28, 2019).
- Lesk, Corey, Pedram Rowhani, and Navin Ramankutty. 2016. “Influence of Extreme Weather Disasters on Global Crop Production.” *Nature* 529(7584): 84–87. <http://www.nature.com/articles/nature16467> (January 29, 2019).
- Li, Deren, Shuliang Wang, and Deyi Li. 2015. “Spatial Data Mining.” <http://dx.doi.org/10.1007/978-3-662-48538-5>.
- Menezes, José Maria P., and Guilherme A. Barreto. 2008. “Long-Term Time Series Prediction with the NARX Network: An Empirical Evaluation.” *Neurocomputing* 71(16–18): 3335–43. <https://www.sciencedirect.com/science/article/pii/S0925231208003081> (January 28, 2019).
- Reddy, D. Sushma, and P. Rama Chandra Prasad. 2018. “Prediction of Vegetation Dynamics Using NDVI Time Series Data and LSTM.” *Modeling Earth Systems and Environment* 4(1): 409–19. <http://link.springer.com/10.1007/s40808-018-0431-3> (January 28, 2019).
- Rembold, Felix et al. 2015. “Remote Sensing Time Series Analysis for Crop Monitoring with the SPIRITS Software: New Functionalities and Use Examples.” *Frontiers in Environmental Science* 3: 46. <http://journal.frontiersin.org/Article/10.3389/fenvs.2015.00046/abstract> (January 28, 2019).
- Shrestha, Ranjay et al. 2017. “Regression Model to Estimate Flood Impact on Corn Yield Using MODIS NDVI and USDA Cropland Data Layer.” *Journal of Integrative Agriculture* 16(2): 398–407. <https://www.sciencedirect.com/science/article/pii/S2095311916615022?via%3Dihub> (January 28, 2019).

Stepchenko, Arthur. 2016. "NDVI Index Forecasting Using a Layer Recurrent Neural Network Coupled with Stepwise Regression and the PCA." In , 130–35. <https://www.ictic.sk/archive/?vid=1&aid=2&kid=50501-281> (January 29, 2019).

Stepchenko, Arthur, and Jurij Chizhov. 2015. "NDVI Short-Term Forecasting Using Recurrent Neural Networks." *Environment. Technology. Resources. Proceedings of the International Scientific and Practical Conference 3*: 180. <http://journals.ru.lv/index.php/ETR/article/view/167> (January 29, 2019).

Udelhoven, Thomas. 2012. "Long Term Data Fusion for a Dense Time Series Analysis with MODIS and Landsat Imagery in an Australian Savanna." *Journal of Applied Remote Sensing* 6(1): 063512. <http://remotesensing.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JRS.6.063512> (January 28, 2019).

Verbesselt, Jan, Achim Zeileis, and Martin Herold. 2012. "Near Real-Time Disturbance Detection Using Satellite Image Time Series." *Remote Sensing of Environment* 123: 98–108. <https://www.sciencedirect.com/science/article/abs/pii/S0034425712001150> (January 28, 2019).

Wang, Siyuan et al. 2016. "Temporal Trends and Spatial Variability of Vegetation Phenology over the Northern Hemisphere during 1982-2012" ed. Shijo Joseph. *PLOS ONE* 11(6): e0157134. <https://dx.plos.org/10.1371/journal.pone.0157134> (January 28, 2019).

Ziervogel, Gina, Peter Johnston, Margaret Matthew, and Pierre Mukheibir. 2010. "Using Climate Information for Supporting Climate Change Adaptation in Water Resource Management in South Africa." *Climatic Change* 103(3–4): 537–54. <http://link.springer.com/10.1007/s10584-009-9771-3> (January 29, 2019).