Development of prediction method for agricultural land using Time series analysis in Dornod, Mongolia

Davaajargal Jargalsaikhan *[1], Bayanjargal Darkhijav[2], Tsolmon Rentsen[3], Enkhjargal Natsagdorj[3] and Ochirhuyag Lkhamjav[4]

[1]*Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences*
[2]*Department of Applied Mathematics, National University of Mongolia*
[3]*Department of Physics, National University of Mongolia*
[4] *Mongolian Geospatial Association*

Abstract

The purpose of this study, we use vegetation index with time series analysis and determine the predicted future prospects and forecasting for an agricultural area.
The study area is situated in the steppe region Dornod province, north-eastern part of Mongolia. In this research, we choose the ARIMA (Autoregressive integrated moving average) model, one of the best-known methods of time series analysis using the NDVI (Normalized Difference Vegetation Index) vegetation index from MODIS remote sensing satellite data between 2010 to 2020. The analysis was performed by Python Jupyter Notebook, ArcGIS, and Envi classic. Validations of this model were issued every four season and the average of agreement is 62 percent.

Key words: *The Normalized Difference Vegetation Index, Time Series Analysis, ARIMA Model,Time series analysis*

1 Introduction

Mongolia is located on the plateau of Central Asia with a 1,565 million square kilometers area, has a harsh continental climate due to its being far from the ocean and most of the Mongolian territory is characterized by arid and semiarid climate, and over 70 % of Mongolia is covered by high-quality steppe grasslands[1]. In Mongolia, approximately 80 % of the total area could be used for agricultural activities (especially pasture) but only 1 % of the total area is used for crop production[2]. Agriculture is an essential economic sector in Mongolia, contributing to more than 20 % of the annual GDP and representing 14 % of the currency revenues[3]. Science-based agricultural production has been developing intensively in Mongolia since 1960[4]. Between 1960 and 1989 the total sown area increased from 267.1 to 846.1 thousand hectares. In 1989, a peak year, the total sown area fell down, reaching 165.0 thousand hectares in 2006[5]. Between 2006 to 2016, however, it roses steadily by 440.6 thousand hectares, but cannot reach 1989. Some agricultural products are growing slowly as a result of the national government program, but the dominant vegetables' are imported from China[6]. That means that Mongolian food security strongly depends on the neighboring countries. Mongolia needs to develop its agricultural sector in agricultural management, especially in main crop production[7]. At present, a widely used GIS application makes land assessments more flexible, with scientific analysis[8]. The Geographic

*davaajargal.j@mas.ac.mn

Information Systems (GIS) is most suited to handle broad extensive data on multiple (spatial, temporal, and scale) from different sources for cost-effective and productive analysis over time[9]. For land crop suitability, some cognitive factors should be carried out which as moisture, digital elevation model, vegetation cover, and effect of topography [10], and demonstrated by means of a GIS-based multi-criteria study to a cropland suitability estimation[7]. Several types of research have been done on the land suitability analysis based on GIS, e.g.[6], [8], [9],[11],[12], [13], and [14], etc.
In this research we used time series method to identify potential agricultural areas based remote sensing satellite data. Remote sensing data efficiently record reflected energy over extended and inaccessible areas using sensors on aircraft or spacecraft, while also providing information on the spatial variability of reflectance periodically[15]. Over the past decade, users have had opportunities to access data acquired from a variety of Earth-observing sensors. The Moderate Resolution Imaging Spectroradiometer (MODIS) is one of the most widely used remote sensing instruments aboard the Terra and Aqua satellites[16]. Two MODIS sensors have sun-synchronous, near-polar circular orbits and observe the same regions 3 hours apart (Terra and Aqua are timed to cross the equator at 10:30 (from north to south) and 13:30 (from south to north), respectively. With a 2330 km swath, MODIS can observe the Earth's surface every 1 to 2 days, using an enhanced spectral resolution sensor in 36 spectral bands. Along with low-level MODIS products, diverse higher-level products for land, atmosphere, cryosphere, and ocean color applications have been distributed to the science and applications community[17]. The Remote sensing indices provide a quantitative proxy of environmental phenomena as an alternative to in-situ measurements. Indices such as NDVI (Normalized Difference Vegetation Index) that emphasize spectrally unique characteristics of the targets of interest have been developed for the application of remote sensing data to vegetation dynamics. NDVI quantifies vegetation by utilizing the spectral characteristics of green vegetation, which strongly reflects in the near-infrared range but absorbs in the red or blue wavelength region, and the value always falls between -1 and 1. (note: a negative NDVI indicates dead plants or inorganic objects, while a positive value indicates live plants, with those close to 1 being healthier)[18]. The NDVI is retrieved from daily atmosphere-corrected bidirectional surface reflectance using a specific compositing method to remove low-quality data, and is widely used in all ecosystems and climates, and in natural resource management studies[19].

2 Data

In this study, we used MODIS NDVI products from remotely sensed images of TERRA/AQUA satellites between 2010 to 2020 with 253 images, which was described in Table 1. The MYD13A1 is the vegetation index of the MODIS land products, its values at a per pixel basis at 500 meters (m) spatial resolution usually produced on 16-day intervals [19]. .

Table 1: Spatial and temporal resolution of the MODIS product

Product	**Name**	**Spatial resolution**	**Temporal resolution**	**Number of images**
MYD13A1v006	NDVI	500m	16 days	253

3 Study area

Mongolia is located in the northern hemisphere, far from the ocean and it belongs to a semi-arid and arid climate region characterized by a continental climate and vulnerable environment. The study area (Figure 1) is Dashbalbar soum of Dornod province which is located in the northeastern part of Mongolia. Geographically, it is mostly steppe,650-700 m above sea level and a small part of the area belongs to the easternmost edge of Khentii suburban mountainous region. The total area of Dashbalbar soum is 8834 thousand square kilometers. Almost 96 percent of the entire area is pasture land. Harsh winter and cool summer, January average temperature, is -30.3C in July average temperature + 26.0C, average annual wind speed 4-5 m/s and a total of annual precipitation 264-300 mm. The sunny day per year is 251-260 days, and the average annual rainfall is 150-300 mm. The average temperature is 27 degrees below zero in the winter, and 21 degrees in the summer. Most of the area is covered by kastanozem soil.

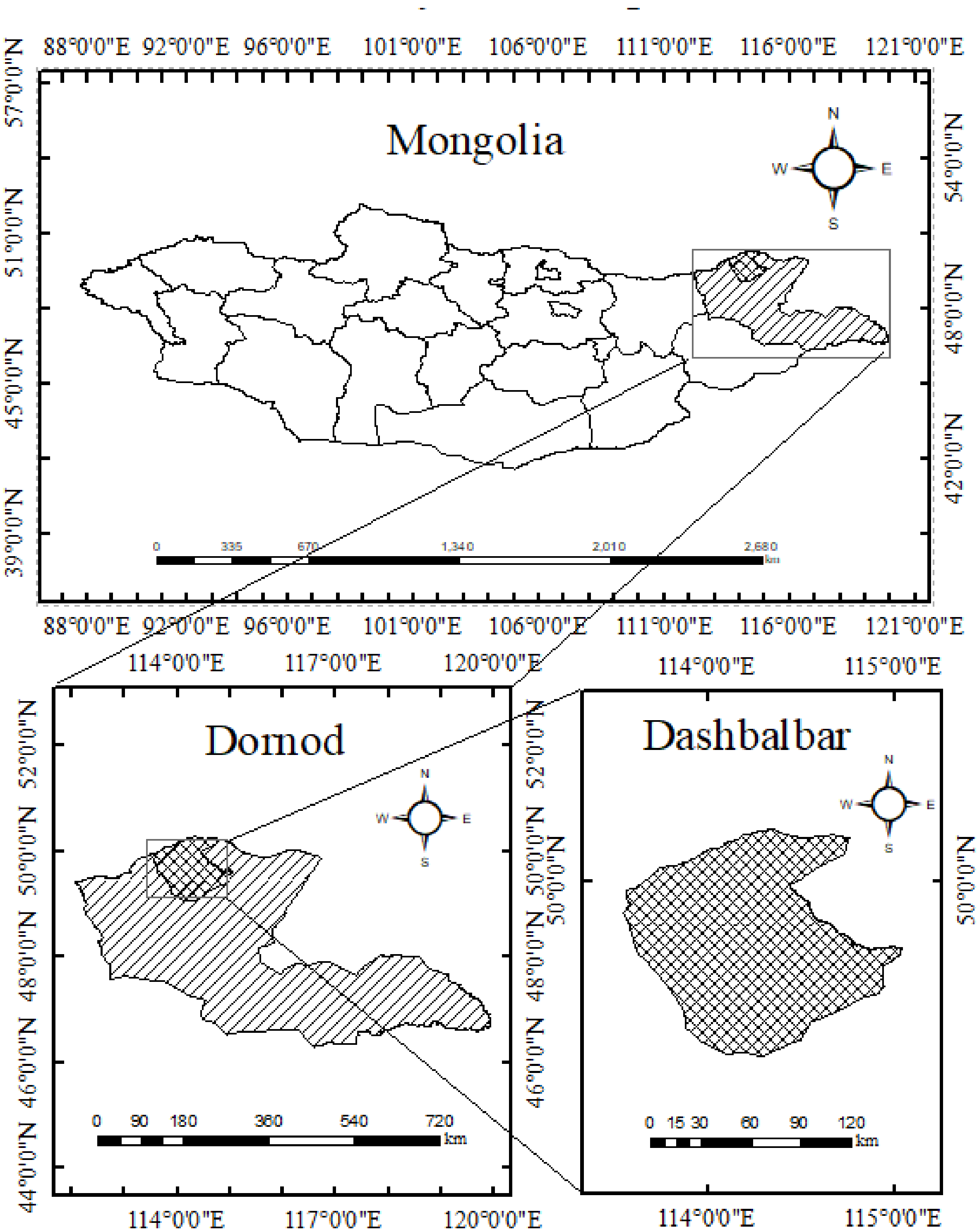


Figure 1: Study Area:Dashbalbar soum, Dornod province, north-east Mongolia

4 Methodology

Time series analysis is a method for analyzing a sequence of data points collected over an interval of time. The most well-known method in time series analysis is ARIMA (Auto-Regressive Integrated Moving Average) which Auto-Regressive (AR) and Moving Average (MA) models are combined A simple form of an AR model of order p, i.e., $AR(p)$, can be written as a linear process given by:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \tag{1}$$

Where x_t is the stationary variable,c is constant, the terms in ϕ_i are autocorrelation coefficients at lags 1,2, , p and ε_t, the residuals, are the Gaussian white noise series with mean zero and variance σ^2.An MA model of order q, i.e., $MA(q)$, can be written in the form:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \tag{2}$$

Where μ is the expectation of x_t (usually assumed equal to zero), the θ_i terms are the weights applied to the current and prior values of a stochastic term in the time series, and $\theta_0 = 1$. We assume that ε_t is a Gaussian white noise series with mean zero and variance σ^2. We can combine these two models by adding them together and form an ARIMA model of order (p, q):

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t + \mu + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \tag{3}$$

Where $\phi_i \neq 0$,$\theta_i \neq 0$, and $\sigma^2 > 0$. The parameters p and q are called the AR and MA orders, respectively. ARMA is a method first introduced by [20] and until now become the most popular model for forecasting univariate time series data. Seasonality is a special property that is frequently observed in many economics, and financial time series[21]) and in our study case. These are fluctuations within a year that repeat in each season. Some of them are the Seasonal autoregressive Integrated Moving Average(SARIMA) model, the Holt-Winters(HW) model [22], the Periodic autoregressive Moving Average(PARMA) model [23], etc.

5 Analysis

We analyzed our model with a Python Jupyter notebook and the process took the following form the best result is $ARIMA(3, 0, 1)(0, 0, 0)[0]$which means seasonal ARIMA is one of the extended models of ARIMA. That is $SARIMAX(3, 0, 1)$ (Figure 2). SARIMAX (Seasonal Autoregressive Integrated Moving Average Exogenous) is comfortable when used on data sets that have seasonal cycles. From the result $SARIMAX(3, 0, 1)$, our time series model for vegetation index as follows:

$$x_t = 0.0231 + 1.7233x_{t-1} - 0.652x_{t-2} - 1.5871x_{t-3} - 0.9188\varepsilon_{t-1} + \varepsilon_t \tag{4}$$

Using ArcGIS and ENVI Classics, we processed in (4) model and gave certain results with created prediction map.

. To analyze our derived model, we need to consider RMSE which means The Root-Mean-Square Error.RMSE is a measure frequently used for assessing the accuracy of prediction obtained by a model. It is the standard

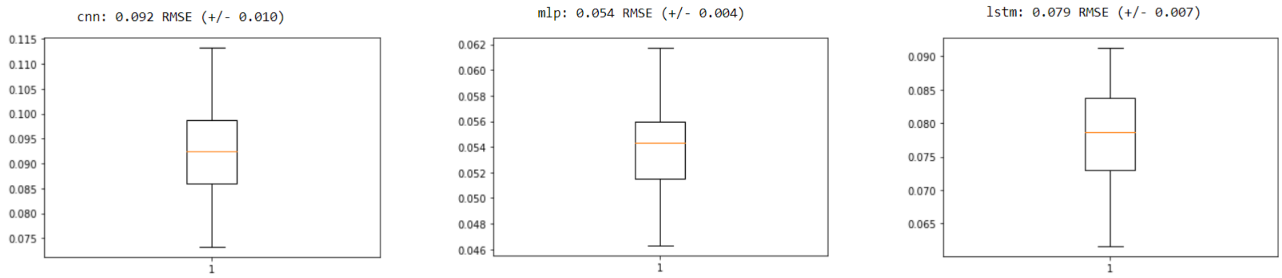
```
Performing stepwise search to minimize aic
 ARIMA(1,0,1)(0,0,0)[0]             : AIC=-494.343, Time=0.07 sec
 ARIMA(0,0,0)(0,0,0)[0]             : AIC=194.249, Time=0.02 sec
 ARIMA(1,0,0)(0,0,0)[0]             : AIC=-479.154, Time=0.04 sec
 ARIMA(0,0,1)(0,0,0)[0]             : AIC=-75.364, Time=0.04 sec
 ARIMA(2,0,1)(0,0,0)[0]             : AIC=-509.398, Time=0.14 sec
 ARIMA(2,0,0)(0,0,0)[0]             : AIC=-502.767, Time=0.07 sec
 ARIMA(3,0,1)(0,0,0)[0]             : AIC=-509.719, Time=0.18 sec
 ARIMA(3,0,0)(0,0,0)[0]             : AIC=-511.719, Time=0.16 sec
 ARIMA(3,0,0)(0,0,0)[0] intercept   : AIC=-536.428, Time=0.13 sec
 ARIMA(2,0,0)(0,0,0)[0] intercept   : AIC=-517.051, Time=0.10 sec
 ARIMA(3,0,1)(0,0,0)[0] intercept   : AIC=-559.821, Time=0.32 sec
 ARIMA(2,0,1)(0,0,0)[0] intercept   : AIC=-521.149, Time=0.36 sec
 ARIMA(3,0,2)(0,0,0)[0] intercept   : AIC=-558.544, Time=0.47 sec
 ARIMA(2,0,2)(0,0,0)[0] intercept   : AIC=-550.176, Time=0.23 sec

Best model:  ARIMA(3,0,1)(0,0,0)[0] intercept
Total fit time: 2.324 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  230
Model:               SARIMAX(3, 0, 1)   Log Likelihood                 261.322
Date:                Mon, 01 Nov 2021   AIC                           -510.643
Time:                        15:49:14   BIC                           -490.015
Sample:                             0   HQIC                          -502.322
                                - 230
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0231      0.003      7.591      0.000       0.017       0.029
ar.L1          1.7233      0.075     23.005      0.000       1.576       1.870
ar.L2         -0.6520      0.136     -4.791      0.000      -0.919      -0.385
ar.L3         -0.1587      0.072     -2.210      0.027      -0.299      -0.018
ma.L1         -0.9188      0.044    -20.744      0.000      -1.006      -0.832
sigma2         0.0059      0.000     12.669      0.000       0.005       0.007
===================================================================================
Ljung-Box (L1) (Q):                   0.07   Jarque-Bera (JB):                23.12
Prob(Q):                              0.79   Prob(JB):                         0.00
Heteroskedasticity (H):               1.06   Skew:                            -0.22
Prob(H) (two-sided):                  0.78   Kurtosis:                         4.49
===================================================================================

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Figure 2: Analysing process

deviation of the residuals (prediction errors) and measures how to spread out these residuals. Residuals are a



(a) CNN (b) MLP (c) LSTM

Figure 3: Score of three methods

measure of how far from the regression line data points are. RMSE is is formulated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x_i})^2} \tag{5}$$

Where N is the total number of observations, x_i is the actual value; whereas, $\hat{x_i}$ is the predicted value. The main benefit of using RMSE is that it penalizes large errors. In the examination, we used CNN(Convolutional neural network), MLP (Multiplayer Perception), and LSTM (Long Short Term Memory) method from Deep learning models[24]. Summarize scores of these methods are shown in Figure 3.
Grid Search of these models is shown in Table 2. .

Table 2: A Grid search methods

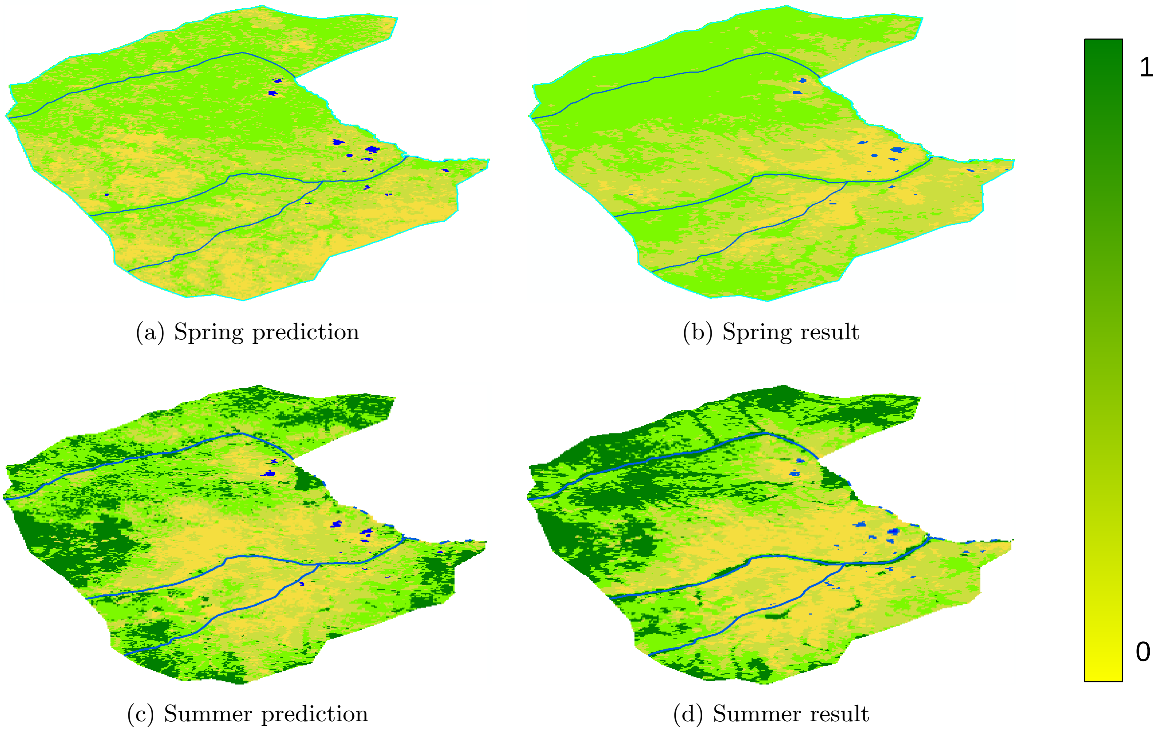
	CNN	MLP	LSTM
n_input	12	12	12
n_filters	64	–	–
n_kernel	3	–	–
n_nodes	–	100	100
n_epochs	100	100	50
n_batch	1	150	1
n_diff	12	12	12
RMSE	0.0484	0.0616	0.0557

.

.

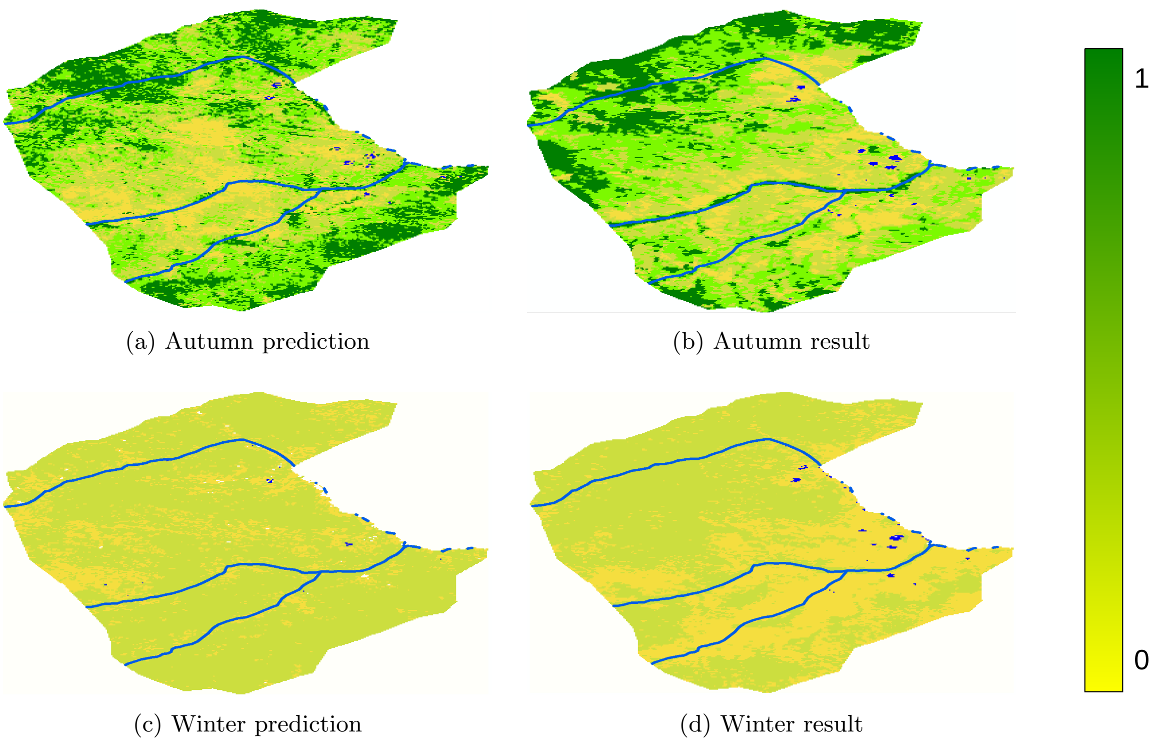
6 Result

We selected MODIS NDVI satellite data for 2010-2020 with 253 images, using a seasonal ARIMA time series model, extension of the ARIMA, and prediction of following years with four seasons. As a result of the processing with Python Jupyter Notebook and obtained the formula (4). Based on processing using this formula (4), we created a prediction map per season(Figure 4 and Figure 5) using ArcGIS 10.4.1 and ENVI Classic 5.2. The results were compared with 2021, with average determination coefficients are 62%(Figure 6). Future forecasting shows in Figure 7.



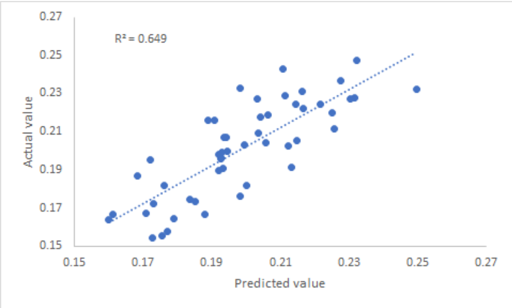
(a) Spring prediction
(b) Spring result
(c) Summer prediction
(d) Summer result

Figure 4: The result map of the proposed model

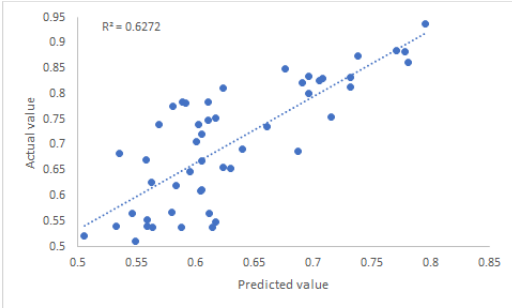


(a) Autumn prediction
(b) Autumn result
(c) Winter prediction
(d) Winter result

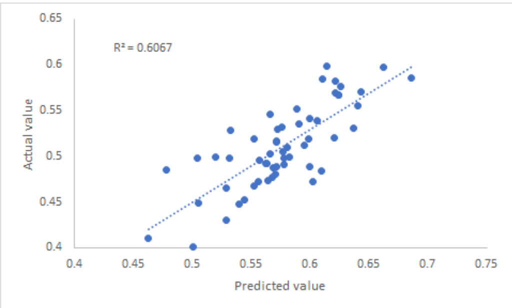
Figure 5: The result map of the proposed model



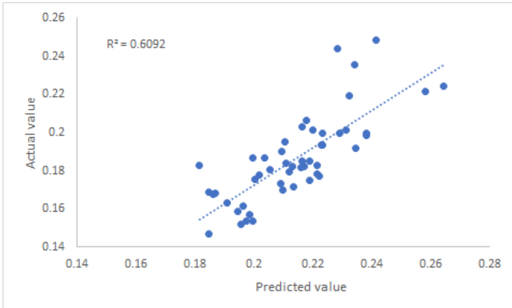
(a) Spring



(b) Summer



(c) Autumn



(d) Winter

Figure 6: Validation the proposed model

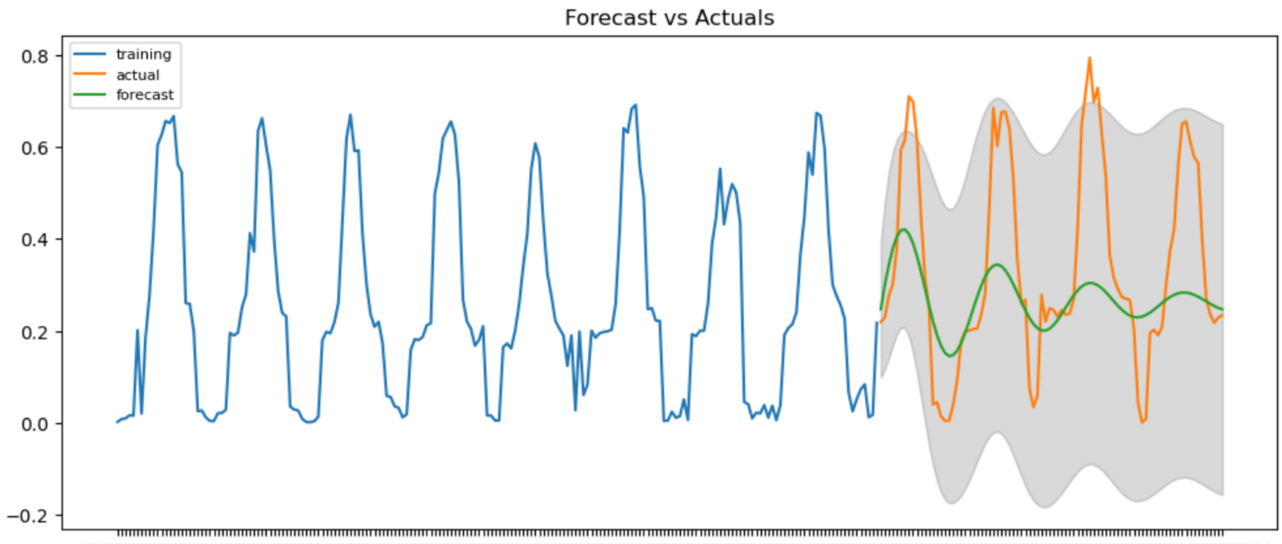


Figure 7: Future forecasting.

7 Conclusion

In this paper, we perform the time series analysis on the vegetation index NDVI and the most convenient model is SARIMAX, one of the ARIMA-type models. From the result, we predicted next year's results, and the comparison was 62%. Also estimating RMSE, analyzing deep learning methods such as CNN, MLP, and LSTM and average RMSE is 0.055. In the future, we will study this method using extensively remote sensing vegetation data such as the Leaf area index (LAI) and Enhanced Vegetation Index (EVI).

References

[1] Fernandez-Gimenez. Maria, and Barbara. Allen-Diaz, ”Vegetation change along gradients from water sources in three grazed Mongolian ecosystems”, *Plant ecology*, 157: 101–118. doi:10.1023/A:1014519206041, 2001.

[2] Hofmann. Jürgen, Dooshin Tuul,and Bazarradnaa Enkhtuya, ”Agriculture in Mongolia Under Pressure of Agronomic Nutrient Imbalances and Food Security Demands: A Case:Study of Stakeholder Participation for Future Nutrient and Water Resource Management”, *Integrated Water Resources Management: Concept, Research and Implementation*, pp.471-514. DOI:10.1007/978-3-319-25071-7_19 (2016).

[3] FAO., 2011. Mongolia and FAO Achievements and Success Stories. Report, Rome, Italy.

[4] Gungaadorj. Sh, ”Brought Virgin Lands under Cultivation and It Developed Independent Agricultural Cropland Sector in Mongolia”, *Proceeding of the National Conference*, ATAR-50, Ulaanbaatar, 23 November 2009, 4-19 (2009).

[5] National Statistics Office of Mongolia (2016) Total Sown Area Data between 1960-2016 Years. *Mongolian Statistical Information Service*, Ulaanbaatar. http://www.1212.mn

[6] Otgonbayar. M, Atzberger. C, Chambers. J, Amarsaikhan. D, Böck. S, and Tsogtbayar. J ”Land Suitability Evaluation for Agricultural Crop-Land in Mongolia Using the Spatial MCDM Method and AHP Based GIS”, *Journal of Geoscience and Environment Protection*, 5(09) 238-263., 4-19 (2016).

[7] E.Natsagdorj,T.Renchin,P.De Maeyer,R.Goossens,T.Van de Voorde, and B. Darkhijav,”A GIS-Based Multi-criteria analysis on cropland suitability in Bornuur soum, Mongolia”, *Remote Sensing and Spatial Information Sciences*, Volume XLIII-B4-2020, XXIV ISPRS Congress (2020).

[8] Pan. G, Pan. J, ”Research in Crop Land Suitability Analysis Based on GIS”. *IFIP Advances in Information and Communication Technology* 369. doi.org/10.1007/978-3-642- 27278-3_33. (2012)

[9] Chen.Y, Yu.J and Khan.S, ”Spatial Sensitivity Analysis of Multi-Criteria Weights in GIS-Based Land Suitability Evaluation” *Environmental Modelling and Software*, 25(2010), 1582-1591.

[10] Eric. K.F,and Abrefa. K.N, ”Digital Soil Mapping in GIS Environment for CropLand Suitability Analysis”, *International Journal of Geomatics and Geosciences*, 2(1), 133-146.(2011)

[11] Perveen. M.F, Nagasawa.R, Uddin. M.I, Delowar. H.K, ”Crop-Land Suitability Analysis Using a Multi-Criteria Evaluation and GIS Approach”, *Symposium on Digital Earth.*(2007)

[12] Kamau.S.W, Kuria.D, and Gachari.M.K,”Crop-Land Suitability Analysis Using GIS and Remote Sensing in Nyandarua County”, Kenya. *Journal of Environment and Earth Science*, 5(6), 121-131(2015).

[13] Quan.B, Zhu.H.J, Chen.S.L, Romkens.M.J.M,and Li. B.C, ”Land Suitability Assessment and Land Use Change in Fujian Province, China”,*Pedosphere*, 17(4), 493-504(2007)

[14] Sarkar.A, Ghosh.A,and Banik.P, ”Multi-Criteria Land Evaluation for Suitability Analysis of Wheat: A Case Study of a Watershed in Eastern Plateau Region, India”. *Geo-Spatial Information Science*, 17(2), 119-128.(2014).

[15] Choe, Hyeyeong, Junhwa Chi. and James H. Thorne, ”Mapping Potential Plant Species Richness over Large Areas with Deep Learning, MODIS, and Species Distribution Models” 1st Edn., *Remote Sensing 13*, no. 13: 2490. (2021) https://doi.org/10.3390/rs13132490

[16] https://modis.gsfc.nasa.gov

[17] Justice, C, Townshend, J. and others, ”An overview of MODIS land data processing and product status” 1st Edn., *Remote. Sens.Environ.*, 2002, 83, 3–15.

[18] Jensen, J.R. "Remote Sensing of the Environment: An Earth Resource Perspective", *Pearson Prentice Hall: Upper Saddle River*, NJ, USA, 2007.

[19] Didan, K. "MYD13A1 MODIS/Aqua Vegetation Indices 16-day L3 Global 500m SIN Grid V006", *NASA EOSDIS Land Processes DAAC*, 2015. doi.org/10.5067/MODIS/MYD13A1.006

[20] Box, G.E.P. and G.M. Jenkins, "Time Series Analysis: Forecasting and Control." 1st Edn., *Holden-Day*, San Fransisco, SBN: 0816211043, pp: 575 (1976)

[21] Ratnadip Adhikari. and R.K.Agrawal, "Forecasting strong seasonal time series with artificial neural network" *Journal of Scientific* & *Industrial Research*, Vol.71, pp:657-666 (2012)

[22] Chatfield.C and Yar.M "Holt-winters forecasting: some practional issues" *J The Statistician*, 37. pages 129-140, 1988

[23] Hipel.K W and McLeod.A.I "Time series modeling of water resource and environmental systems" *Elsevier Science Publishing Company*, Amsterdam, 1994.

[24] Jason Brownlee."Deep Learning for Time Series Forecasting-Predict the Future with MLPs, CNNs and LSTMs in Python" *Machine Learning Mastery*, 2018