2020
Turkish Journal of
Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS

Türk Uzaktan Algılama ve CBS Dergisi

Volume/Cilt: 1 Issue/Sayı: 2 Year/Yıl: 2020 E-ISSN:2717-7165

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi
Turk J Remote Sens GIS
Journal homepage: http://www.dergipark.gov.tr/en/pub/rsgis

Turk J Remote Sens GIS, (2020), Cilt/Volume:1 Sayı/Issue:2 Eylül/September 2020

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi
Turk J Remote Sens GIS
Journal homepage: http://www.dergipark.gov.tr/en/pub/rsgis

Turk J Remote Sens GIS, (2020), Cilt/Volume:1 Sayı/Issue:1 Eylül/September 2020

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi
Turk J Remote Sens GIS
Journal homepage: http://www.dergipark.gov.tr/en/pub/rsgis

Turk J Remote Sens GIS, (2020), Cilt/Volume:1 Sayı/Issue:1 Eylül/September 2020

İÇİNDEKİLER / CONTENTS

ARAŞTIRMA MAKALELERİ
RESEARCH ARTICLES

2020
Turkish Journal of Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, September 2020, 1(2): 61-77
Journal homepage: http://www.dergipark.org.tr/en/pub/rsgis

Landslide Susceptibility Mapping Using Shallow Neural Networks Model at Refahiye District in Turkey

Sığ Sinir Ağları Modeli Yardımıyla Türkiye'de Refahiye İlçesinin Heyelan Duyarlılığının Haritalanması

Sohaib K. M. Abujayyab[1]*, Ismail R. Karas[2]

[1]*Department of Geography, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey.*
[2]*Department of Computer Engineering, Karabuk University, Demir Celik Campus, 78050 Karabuk/Turkey.*

ORIGINAL PAPER

***Corresponding author:**
Sohaib K. M. Abujayyab
sjayyab@karabuk.edu.tr

doi:

Article history
Received: 27.01.2020
Accepted: 20.08.2020
Published: 30.08.2020

Abstract

Landslides represent a continuous hazard for population and infrastructure. Mapping the landslide susceptibility is an essential issue to avoid the landslides risks. The aim of this paper is to produce a high-accuracy model for landslide susceptibility mapping in Refahiye district in Turkey. The model employed shallow neural networks for landslide susceptibility mapping, while bivariate spearman correlation test was utilized to select the related factors to extract the appropriate data and reduce the computation time of training and mapping. 12 out of 21 spatial factors were selected as relevant factors using Spearman correlation test. Relevant factors are geology, distance from roads, distance from geological faults, distance from water streams, flow direction, aspect, hillshade, heat load index, slope/aspect transformation, site exposure index, compound topographic index, and elevation. The generated dataset was divided into training, validation, and testing datasets using 10-folds cross-validation method. The TrainIm was found to be the best training function with an overall accuracy of 86.3%. The developed NN model was tested using IRIS benchmark dataset and showed higher performance against the logistic regression algorithm. As a result, shallow neural networks method was successfully applied in landslide susceptibility mapping in this study and the method is recommended for future studies.

***Keywords**: GIS, Landslide susceptibility mapping, Shallow neural networks*

Özet

Heyelanlar nüfus ve altyapı için sürekli bir tehlike oluşturmaktadır. Heyelan duyarlılığının haritalanması heyelan risklerini önlemek için önemli bir konudur. Bu çalışmanın amacı, Türkiye'nin Refahiye ilçesinde heyelan duyarlılık haritalaması için yüksek doğruluklu model üretmektir. Modelde heyelan duyarlılık haritalaması için sığ sinir ağlarını kullanılırken, uygun veriden gerekli faktörleri çıkarmak ve haritalama ve eğitim hesaplama süresini azaltmak için iki değişkenli Spearman sıra korelasyon testi kullanılmıştır. 21 mekansal faktörden 12'si, Spearman korelasyon testi kullanılarak ilgili faktörler olarak seçilmiştir. İlgili faktörler jeoloji, yollara uzaklık, jeolojik faylara olan uzaklık, su yollarına olan uzaklık, akış yönü, bakı, arazi kabartı, ısı yük endeksi, eğim / bakı dönüşümü, alan maruziyet indeksi, bileşik topografik indeks ve yüksekliktir. Oluşturulan veri kümesi, 10 katlı çapraz geçerlilik yöntemini kullanarak eğitim, doğrulama ve test veri kümelerine bölünmüştür. %86,3'lük genel doğruluk performansı elde edilen en iyi eğitim fonksiyonu (TrainIm)'dir. Geliştirilen NN modeli, IRIS kıyaslama veri seti kullanılarak test edildi ve lojistik regresyon algoritmasına göre daha yüksek performans gösterdi. Sonuç olarak, bu çalışmada heyelan duyarlılık haritalamasında sığ sinir ağları yöntemi başarıyla uygulanmış ve yöntem gelecekteki çalışmalar için önerilmiştir.

***Anahtar kelimeler**: CBS, Heyelan duyarlılık haritalaması, Sığ sinir ağları*

1. Introduction

Refahiye is a Turkish district located in the east Black Sea region, which has been influenced by landslides hazards. In the Black Sea region, infrastructure destruction caused by landslide occurrence is worse than the destruction caused by earthquakes (Dağ and Bulut, 2012). Since the main and active geological fault line in Turkey is crossing through the mountainous area of the Refahiye district, this area is continually at risk of landslides. Landslide risk problem is expected to continue for long years as a consequence of deforestation, global warming, climate change, and urban growth (Yilmaz, 2009). The possibility of landslides occurrence depends on several complex conditions such as topographic structure, soil types, geology structure, land use, land cover LULC activists, precipitation, and several geomorphometrical factors (Nefeslioglu et al. 2012).

During the last two decades, several pieces of research have been conducted to develop methods and frameworks in order to mapping the landslides susceptibility (Chae et al. 2017; Felicísimo et al. 2013; Pradhan and Lee, 2010; Song et al. 2012; Vakhshoori et al. 2019; Yalcin et al. 2011; Yalcin, 2008; Yıldırım and Güler, 2016). An example of the employed methods for landslides susceptibility mapping were random forest (Hamad et al. 2018), neural networks (NN) (Valencia Ortiz and Martínez-Graña, 2018), logistic regression (Nefeslioglu et al. 2008), support vector machine (Wang et al. 2019) and (Tso and Yau, 2007).

The generalization accuracy of NN algorithm is high among the former methods for the landslides susceptibility mapping (Can et al. 2019; Chae et al. 2017; Vakhshoori et al. 2019; Zhang et al. 2018). The NN method exploits the data acquired from the landslide inventories layer to forecast the possibility of landslides that will occur in the future. NN algorithm can generate a weighted model for mapping landslides susceptibility. The fundamental benefits of employing NN in landslide susceptibility mapping are its capability to handle several types of spatial data at different scales as well as ordinal and nominal data. Although the NN model can achieve high generalization in several areas in the world, mapping landslide susceptibility in new geographical areas need further investigation. The change in the area relatively affects the model performance due to several issues. The change in the analysis area is influenced by the quality of the data, data availability of factors, reasons of landslides, and types of landslides. In addition, when the analysis conduct based on regional scale, the model performance highly influenced by the low data quality and low spatial accuracy. Generally, the former studies focused on increasing the prediction accuracy using more data and more computation. The increment in the data and computation lead to several difficulties for conducting LSM analysis. On the other side, several types of research applied deep neural networks to develop prediction models. Although high model generalization is achieved by deep learning, it is not time efficient. Thus, producing a high accuracy model using a simple algorithm with less computation and minimum number of input data becomes a new aspect in the research area of landslide susceptibility mapping (Lee et al. 2020). Using the bivariate spearman correlation test to select the factors and employing time-efficient shallow NN are the original contributions of this study.

The aim of this study is to produce a highly-generalized model for landslide susceptibility mapping in the Refahiye district of Turkey. The model used shallow neural networks for landslide susceptibility mapping, while bivariate spearman correlation test was employed to select the relevant factors that ensured less data collection and less computation time. This paper consists of four sections as the introduction, methodology (containing study location, data collection, and development of the shallow neural networks model), results and discussion, and conclusion.

2. Methodology

2.1 Study area and data collection

Refahiye is a district of Erzincan in the Eastern Anatolia region of Turkey. It is located between 39°04'15"N - 40°04'16"N latitudes and 38°23'33"E - 39°13'18"E longitudes. The administration area of Refahiye covers an area of approximately 1816 km^2, and the elevation is 1589 m. The total population of the Refahiye district is around 10,569, where 3730 live in the town center. The area has usually a high humidity rate in the winter. In Refahiye, the average annual temperature is 9.0 °C. The average precipitation is 537 mm. Winter months are much rainy compared to the summer months. In Refahiye area, several water streams lows in the area. The main active geological fault line is crossing through the study area. Refahiye district is a mountainous area. Figure 1 shows the study area. Considering its geographical location and geological characteristics, Refahiye district can be considered an active area for landslides occurrence.

To develop the NN model and produce the susceptibility map, the data collection stage has to be performed. In this stage, the landslide inventory map and factors map were prepared. Firstly, the landslide inventory map of the study area was generated based on the investigation from the satellite images and published maps. Satellite images were extracted from Google Earth Pro. In spite of the fact that Google Earth offers images since 1984, only images from 2001 to 2019 were used to achieve suitable spatial resolution for landslides.

Moreover, the landslides map published by General Directorate of Mineral Research and Exploration (GDMRE) (http://yerbilimleri.mta.gov.tr) was utilized to enrich the inventory map (Duman et al. 2011). According to the available landslide data, only the active landslides were considered during the mapping stage. Based on the landslide inventory map of Refahiye, there are 237 landslides in the study area (Duman et al. 2011). The total area of landslides is 51.37 km^2, which represents 2.82% of the total area of Refahiye (1816 km^2). The landslide locations are shown in Figure 1. According to the investigation from the satellite images, the landslides in the study area could be classified as shallow translational slides (Pitasi, 2016; Turner and Schuster, 1996). Example photos of landslides from the study area are given in Figure 2. Secondly, the explanatory factors were collected from several sources such as topographic maps and digital elevation model (JAXA, 2019). The list of collected factors was classified and a list of all 21 factors are given in Table 1 in four groups. In addition, the landslide training data set for NN were generated based on random sample points. The samples are randomly generated. The number of samples set equally between the landslides and non-landslides areas. Thus, 5000 points were generated. Then, the dataset was extracted from 21 spatial layers. Afterwards, 45 missing values were removed from the dataset, leading to a total of 4955 samples remaining. Moreover, the bivariate test named 'spearman correlation' was utilized to test the correlation between the factors. Matlab software was utilized to develop the model and test the correlation among the explanatory factors.

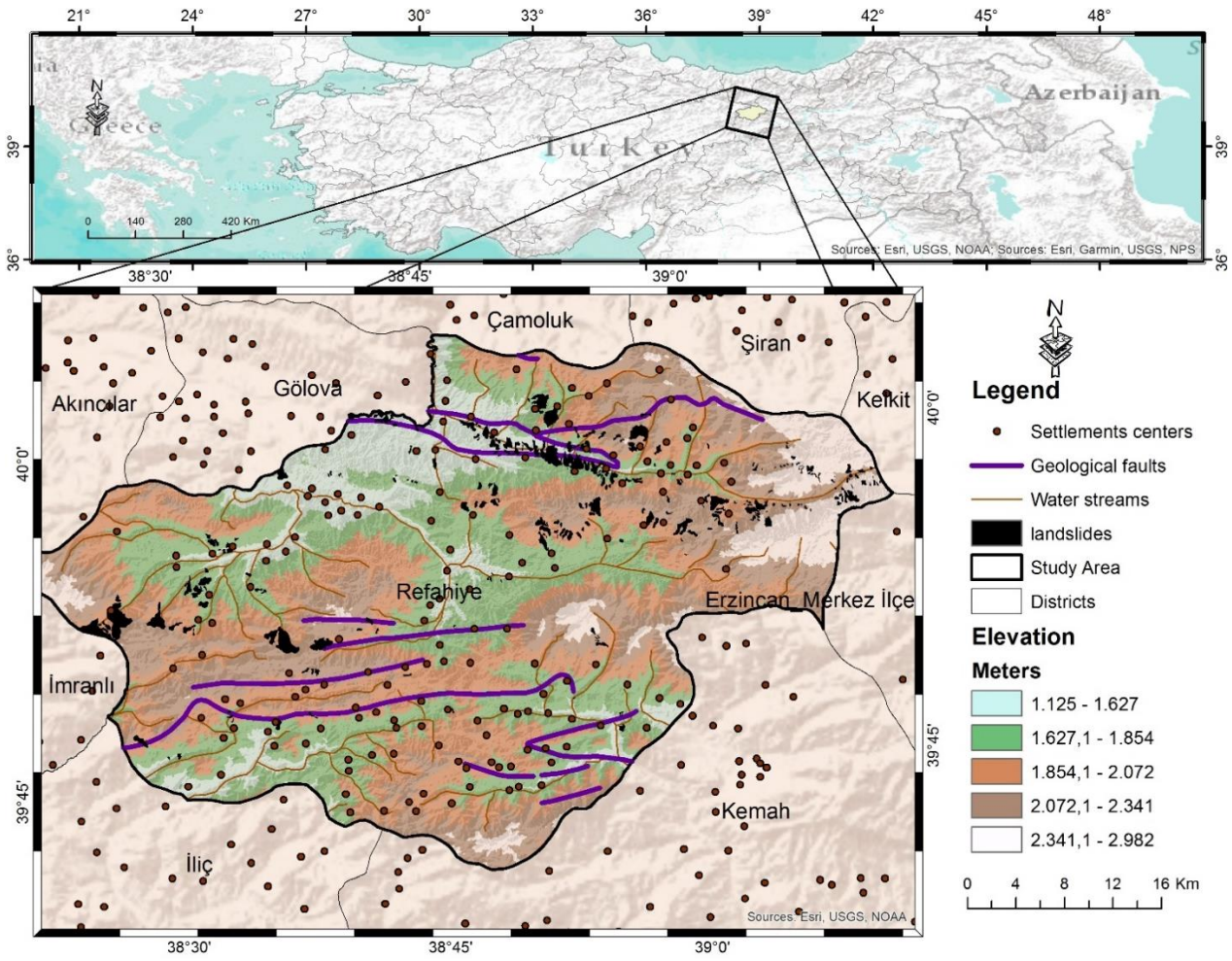


Figure 1. Location map of the study area

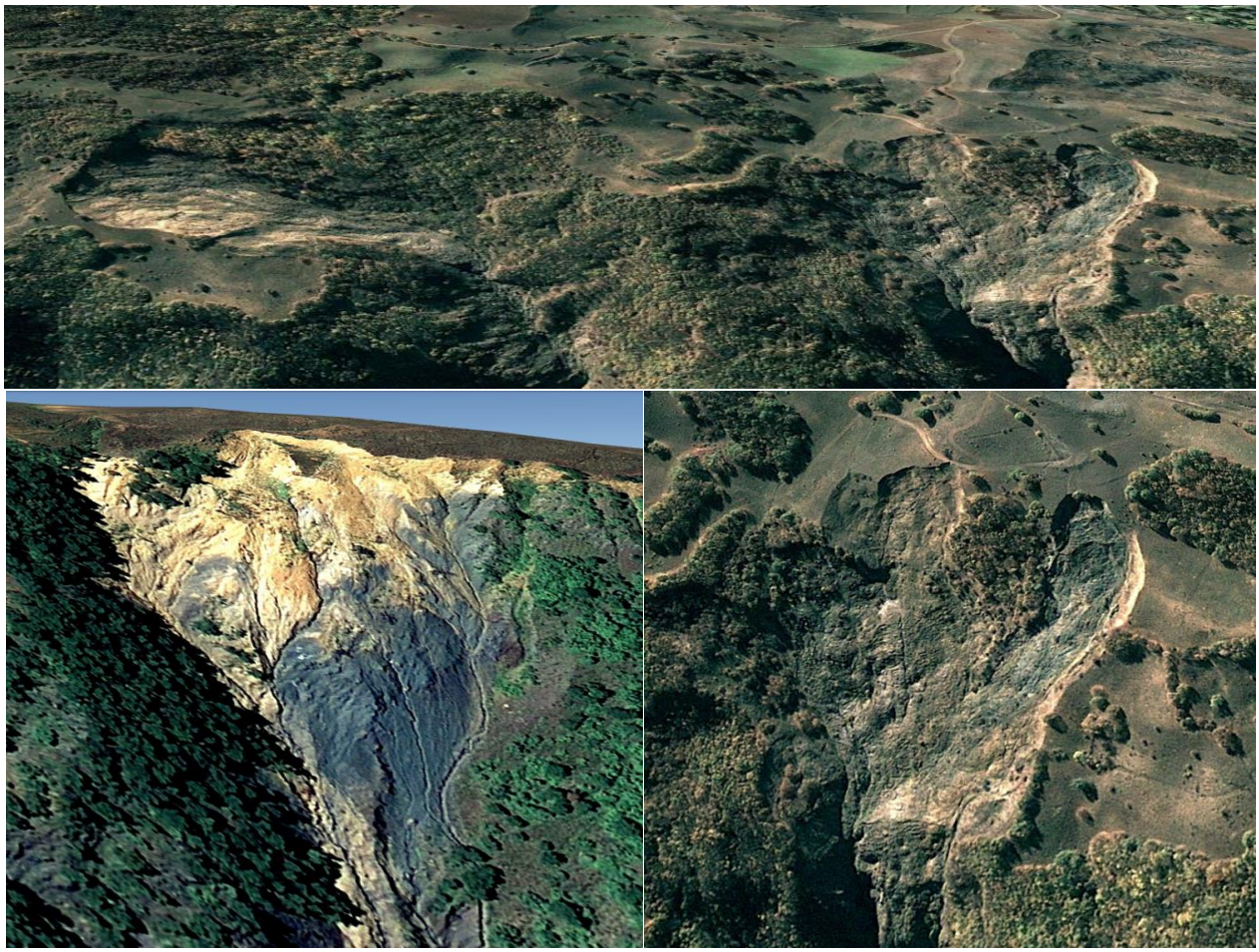


Figure 2. Example photos of landslides from the study area

Table 1. Landslides explanatory factors classified in four groups

Group	Factors
Topographic surfaces and Geomorphology texture	1. Elevation 2. General Curvature 3. Dissection 4. Landform 5. Slope Position 6. Surface Area Ratio 7. Surface Relief Ratio 8. Aspect 9. Hillshade 10. Slope 11. Geology 12. Geological Faults
Geomorphometry	13. Flow Direction 14. Water streams
Temperature and moisture	15. Heat Load Index 16. Site Exposure Index 17. Compound Topographic Index 18. Slope/Aspect Transformation 19. 2nd Derivative Slope
Human activities	20. Land use 21. Roads

Spearman correlation test is a non-parametric experiment used to specify the level of association among two factors. The test was employed to select the important factors that enabled less data collection and less computation time. Spearman correlation can be computed as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

Where;
ρ is the Spearman rank correlation
d is the difference between the ranks of corresponding variables
n is number of observations

2.2 Development of the Shallow Neural Networks model

The shallow neural network model is one of two types of neural networks (Kim and Gofman, 2018; X.-D. Zhang, 2020). The second type is the deep neural networks. Shallow NN is only suitable for landslide susceptibility mapping problem. The deep neural networks are suitable for landslide detection problem, allowing for the detection of the location of the landslides after their occurrence, while the problem of this paper is to find the possibility of future landslides. Neural networks imitate the physical structure of the biological cell (dendrites, nucleus and axons). In the artificial neuron, dendrites are represented by the input factors, Nucleus is represented by the artificial neurons and axons are represented by the connection between the neurons and forward or output layers. A group of artificial neurons arranged in input, hidden and output layers construct the shallow NN model. The neural networks method as a machine learning algorithm is able to mapping the landslide susceptibility based on the former landslides. By extracting the geographical and geological parameters of each landslide site, the NN algorithm is able to generate a weighted statistical model that can map the landslide susceptibility for new areas. The chief benefits of employing NN in landslides susceptibility mapping are its capability to handle several types of spatial data at different scales as well as ordinal and nominal data.

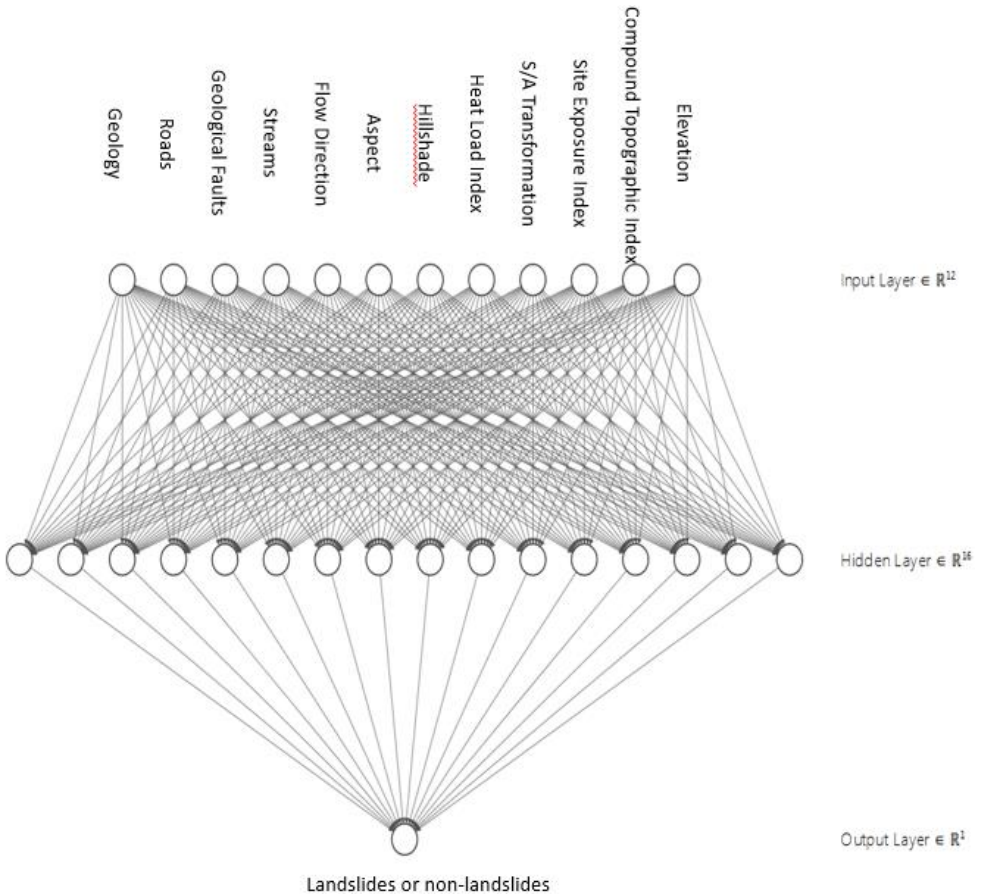


Figure 3. Structure of shallow neural networks

The structure of shallow NN algorithm is displayed in Figure 3. The algorithm structure contains three layers. The first one is the input layer, which represents the input factors. The input factors are the relevant spatial data of the historical landslides. The second layer is the hidden layer, which represents the layer of neurons. This is the most important layer where all operations are performed. The hidden layer contains the artificial neurons that specifying the ability of the model to extract the knowledge from the landslides dataset and build the model. The last layer represents the output layer. The output layer of the current algorithm is scaled between 0 and 1. The values represent the level of susceptibility. The developed model was generated in Matlab software. The Pseudocode of the developed model is given in algorithm 1.

Performance accuracy was calculated for testing dataset by comparing actual state and computed reactions. Performance accuracy of the developed models was measured using Overall Accuracy (confusion matrix), Cohen Kappa Accuracy, and AUC Accuracy (Liang et al. 2015). McCormick (2016) stated that the Cross-validation method effectively assesses the network accuracy through splitting the datasets into training and testing sets (McCormick, 2016). Cross-validation protects the network from overfitting. Wise reported that there are several techniques utilized in former works (Wise, 2011). The 10-fold technique shown in Figure 4 was applied here.

Algorithm 1. Pseudocode of the shallow NN model for landslide susceptibility mapping

```
##Reading the training dataset

## Training step
Separate the dataset into training and testing sets
Normalizing the datasets
Building algorithms for the shallow NN model
Define the NN model (12 inputs, hidden layers, and outputs)

for k=1: number of neurons do
   Training the NN classifier using (k) neuron in the hidden layer
   Calculating the accuracy of the NN model
   If the accuracy is sufficient:
      Stop and store the NN model
   Else:
      Hold on the training process

## Testing step
for i=1: number of chunks do
     read the chunk dataset
     using the trained NN model to test (i) chunk dataset
     Store the output values in TXT format.
```

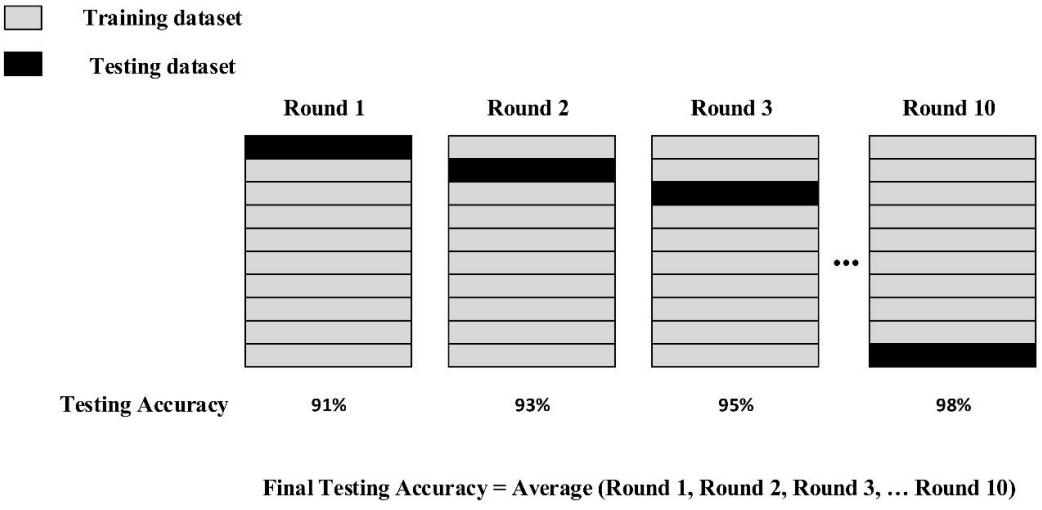


Figure 4. 10-fold scheme for cross-validation

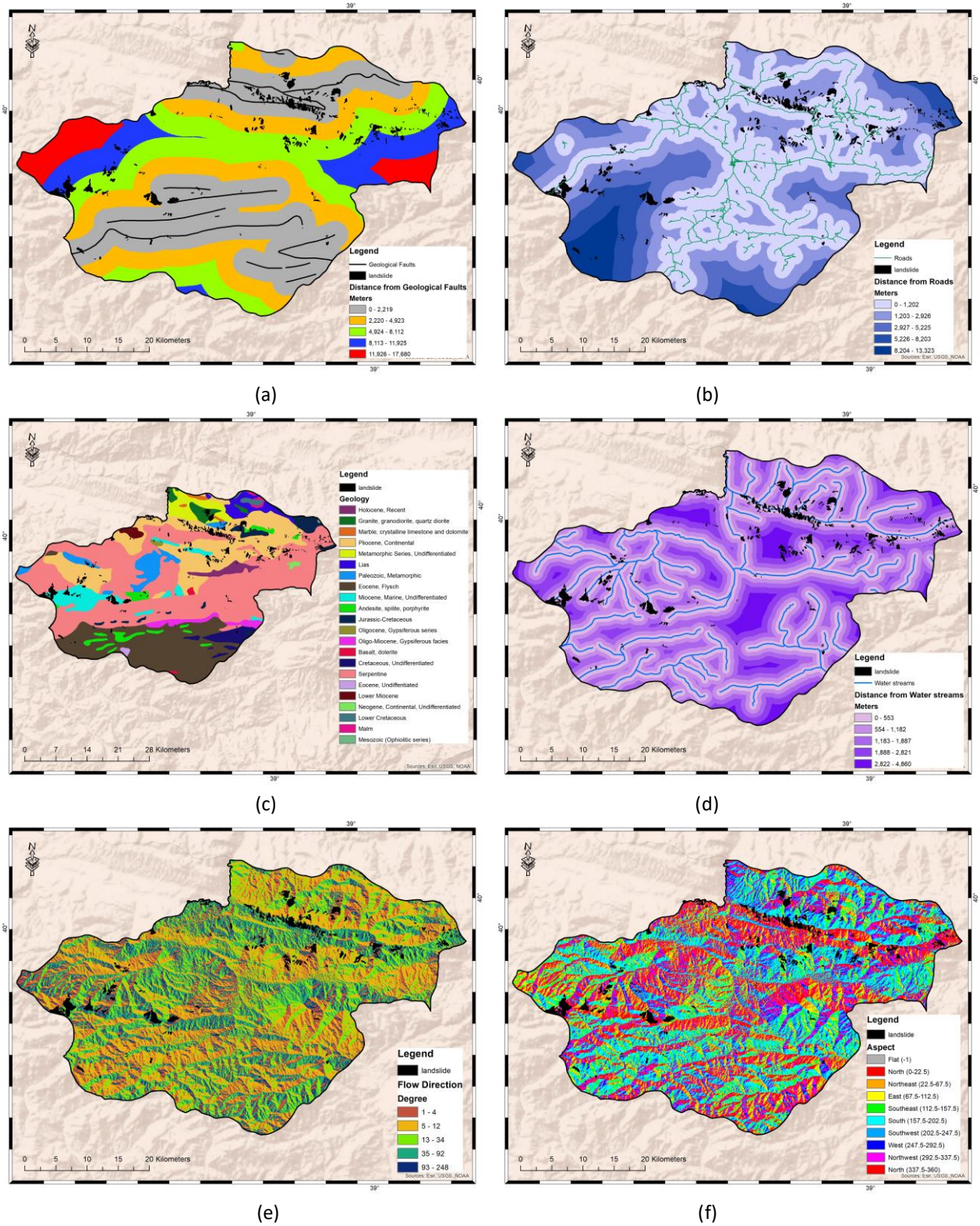


Figure 5. Landslide conditioning factor maps: (a) distance from faults; (b) distance from roads; (c) geology; (d) distance from water streams; (e) flow direction; (f) land aspect

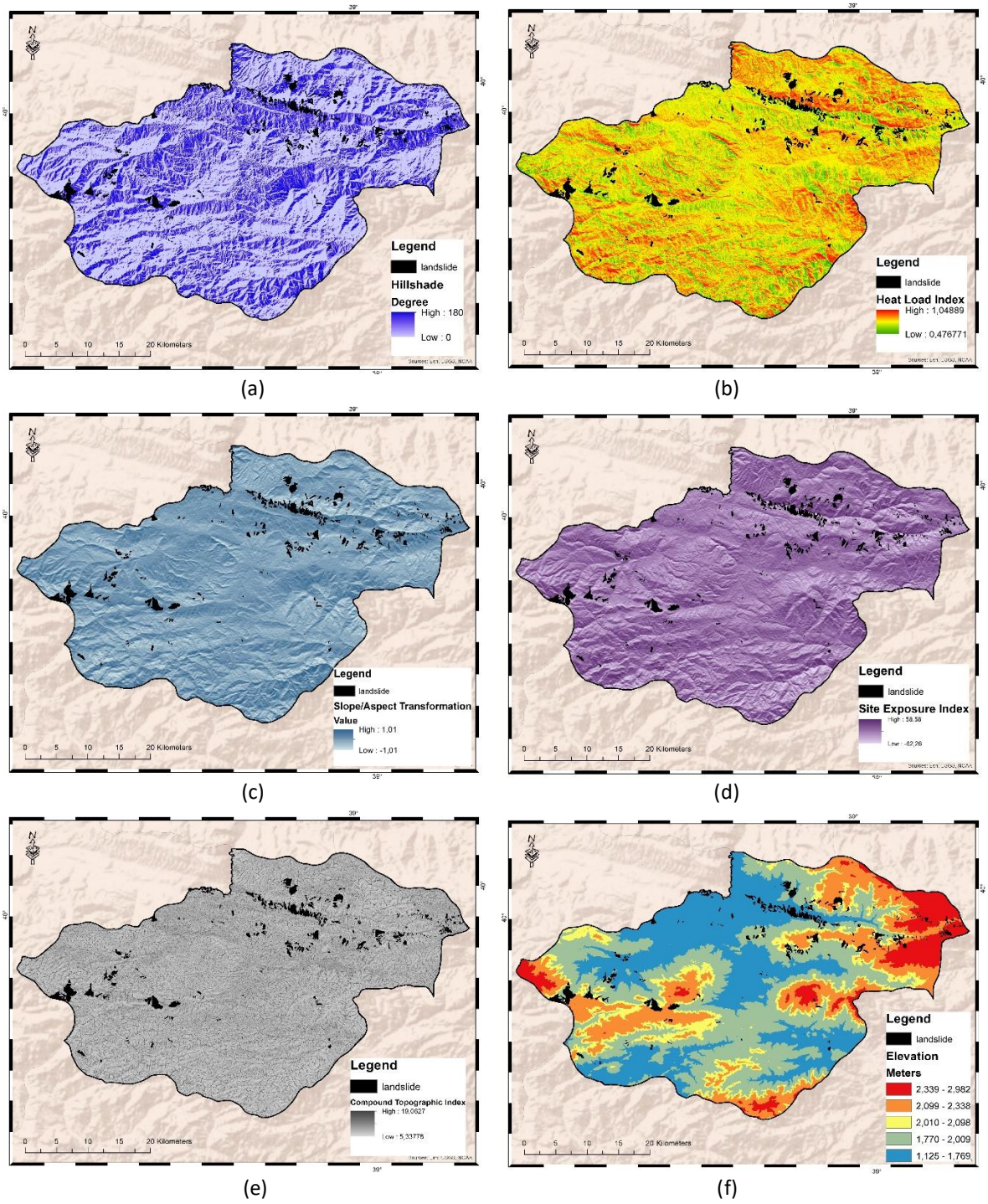


Figure 6. Landslide conditioning factor maps: (a) hillshade; (b) heat load index; (c) slope/aspect transformation; (d) site exposure index; (e) compound topographic index; (f) elevation

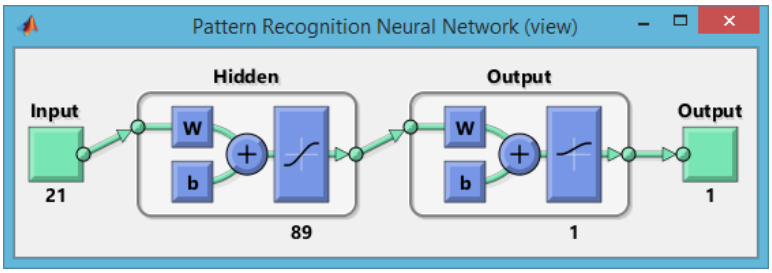
3. Result Analysis and Discussion

In this section, the results of the analysis are given and evaluated. According to the Spearman correlation test, 12 out of 21 factors were selected as relevant factors based on the correlation value of each factor. A factor was selected as relevant if it received a correlation greater than 2%. These 12 factors are geology, distance from roads, distance from geological faults, distance from water streams, flow direction, aspect, hillshade, heat load index, slope/aspect transformation, site exposure index, compound topographic index, and elevation. Some of these factors have a positive correlation and some have negative correlation. For example, if the correlation between the landslides and Compound Topographic Index is positive, that means as long as the Compound Topographic Index increases the landslides are more likely to occur. In addition, if the correlation between the landslides and distance from roads is negative, that means as long as the distance from roads increases the landslides are less likely to occur. The correlation values among the factors are presented in Table 2 and also the maps of selected factors are presented in Figures 5 and 6.

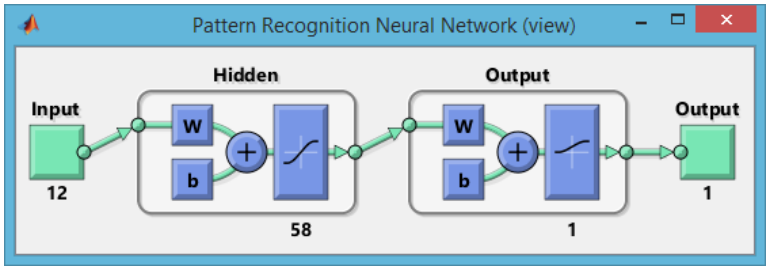
Table 2. The correlation between the occurred landslides and the collected factors

Factors (1-7)	Land use	Geology	Roads	Geological Faults	Water streams	Flow Direction	Slope
Spearman correlation	-0.00376	**-0.2124**	**-0.1764**	**-0.0654**	**-0.2354**	**0.0950**	-0.0062
P_value	0.0081	0.0000	0.0000	0.0000	0.0000	0.0000	0.6621
Factors (8-14)	Aspect	Hillshade	General Curvature	Heat Load Index	Slope/Aspect Transformation	Site Exposure Index	Compound Topographic Index
Spearman correlation	**-0.0432**	**0.0688**	0.0192	**-0.0938**	**0.0926**	**-0.0926**	**0.0519**
P_value	0.0024	0.0000	0.1769	0.0000	0.0000	0.0000	0.0003
Factors (15-21)	2nd Derivative Slope	Surface Relief Ratio	Surface Area Ratio	Slope Position	Landform	Dissection	Elevation
Spearman correlation	-0.0188	0.0290	-0.0062	0.0215	0.0215	0.0332	**-0.0824**
P_value	0.1855	0.0414	0.6620	0.1304	0.1304	0.0195	0.0000

Based on the selected 12 factors and all 21 factors, a new training dataset was generated and used to train the shallow NN model. The developed models are presented in Figure 7. Several training functions were employed to train the NN model and improve their performances. These functions are Levenberg-Marquardt backpropagation (Trainlm), Scaled Conjugate Gradient (Trainscg) and Bayesian regularization backpropagation (Trainbr). The best performance accuracy was obtained by the Trainlm function. In addition, further experiments were conducted to find the effects of the number of neurons on the performance accuracy. Based on the best training function (Trainlm), an optimizer developed to find the accuracy of the developed model using range of number of neurons in the hidden layer. Figure 8 demonstrates the curves of the accuracy over the number of neurons from one to 60 based on the Trainlm training function using the three-evaluation metrics (Overall Accuracy (confusion matrix), Cohen Kappa Accuracy, and ROC-AUC Accuracy). The outcomes illustrate that the best performance accuracies are (Overall Accuracy 88%, ROC-AUC 91.5%, and Cohen Kappa 75.6%) obtained when using 50 neurons in the hidden layer. Thus, the final shallow NN model contained three-layers. The input layer consisted of 12 variables. The hidden layer consisted of 50 neurons and the output layer consisted of one output, which represented the degree of landslide susceptibility.



(a)



(b)

Figure 7. An example of the developed model structure based on all 21 factors (a) and relevant 12 factors (b)

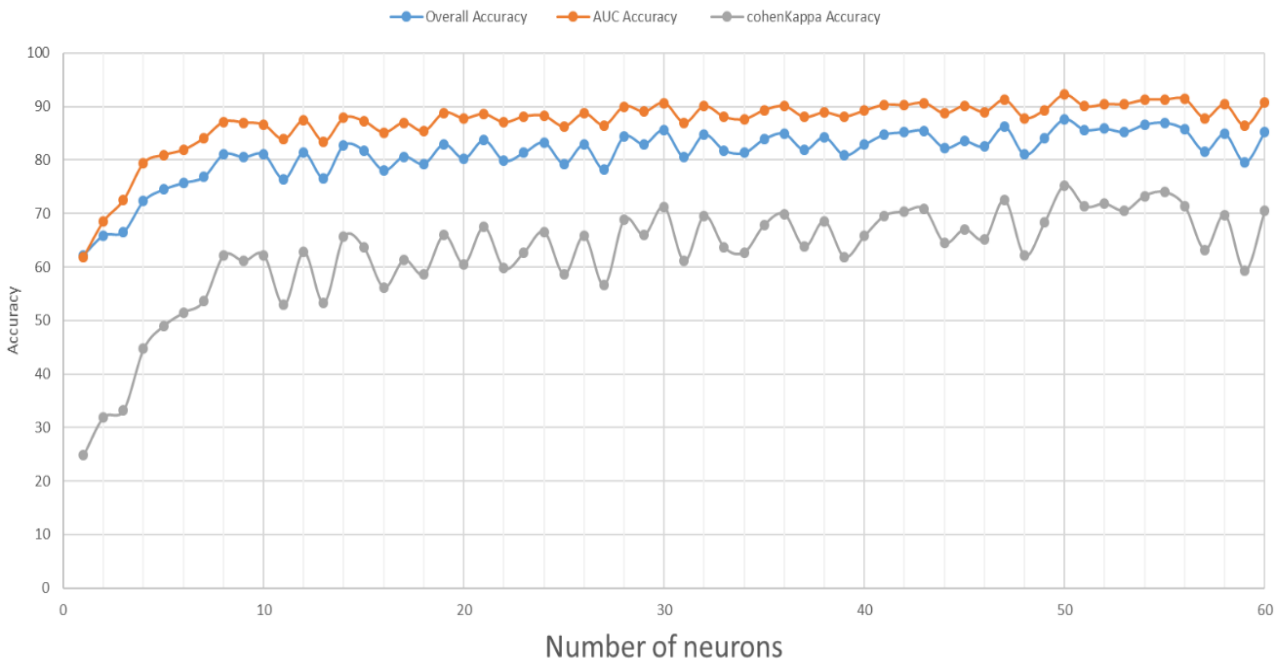
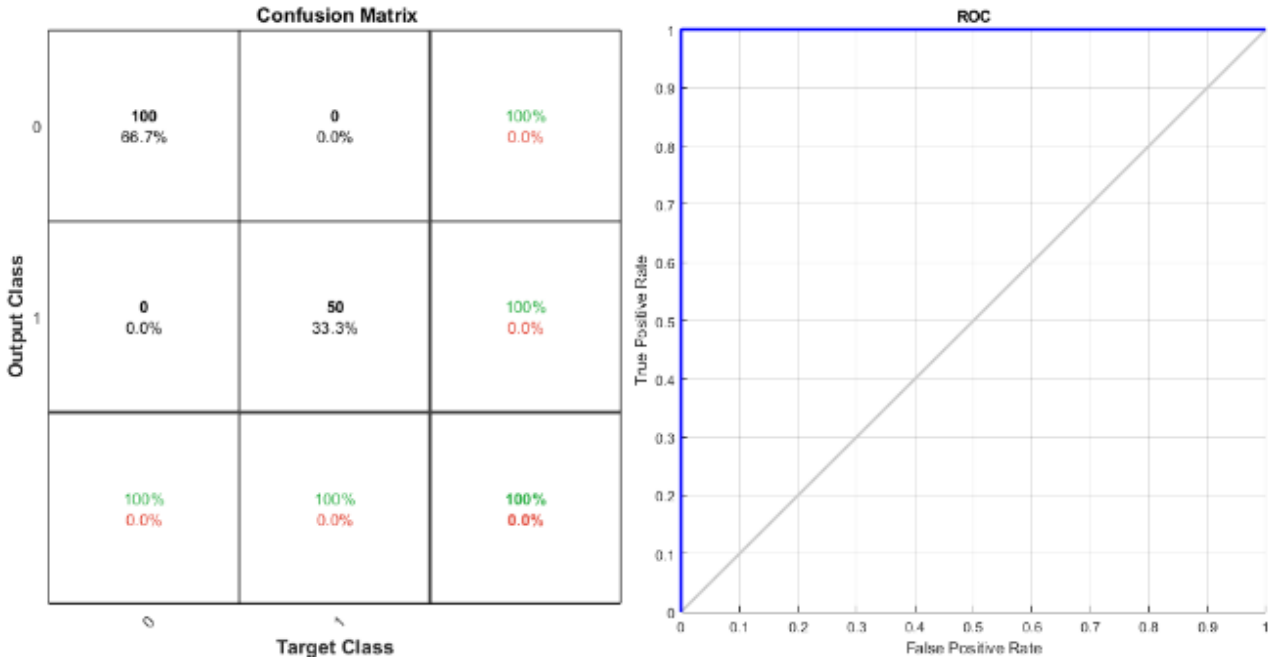


Figure 8. An optimizer to find the accuracy of the developed model across the number of neurons in the hidden layer. Three metrics were used to assess the accuracy: Overall Accuracy (confusion matrix), Cohen Kappa Accuracy, and ROC-AUC Accuracy

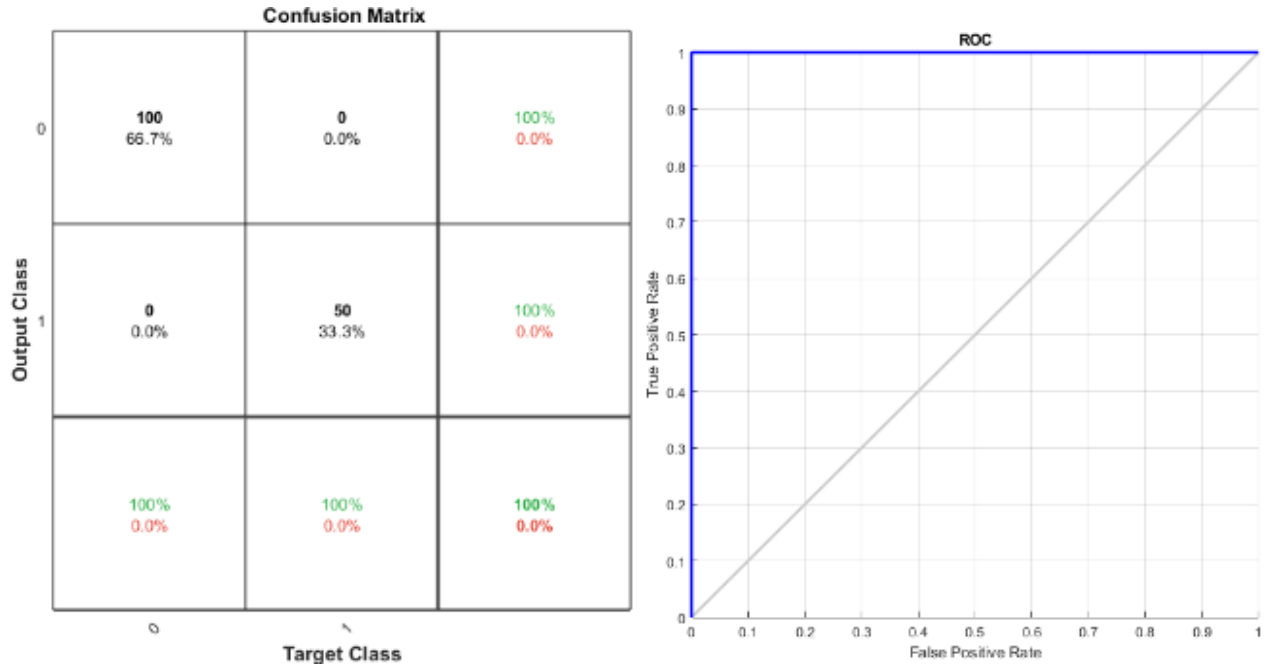
Furthermore, for further investigation of the developed NN model, the model was tested again based on the best structure. NN model was tested using the Confusion matrix metric. The experiment was conducted after dividing the dataset into training, validation, and testing datasets using the 10-folds cross-validation method. Thus, the model accuracy using confusion matrix for training, validation, and testing dataset as well as the overall model accuracy illustrated in Figure 9. The performance accuracy based on the training sub dataset was found as 88.6%, while the performance accuracy based on the testing sub dataset was found as 81.0%. The accuracy between training dataset and testing dataset showed that there was no over-fitting problem in the developed NN model since the testing accuracy was less than the training accuracy.

Figure 9. NN model accuracy based on the confusion matrix metric. The overall accuracy is 86.3%

Additionally, the developed NN model was compared to other benchmark data and another convolution method. The NN model was compared with IRIS benchmark data. The IRIS data was tested using the developed NN model and another model was developed using the logistic regression algorithm. Figure 10 illustrates the performance accuracy of each model based on IRIS benchmark dataset. Then, the developed model using logistic regression algorithm applied based on the 21 factors and based on the 12 selected factors. Figure 11 illustrates the performance accuracy of each model with three evaluation metrics. Lastly, the evaluation metrics, which were calculated for the NN model based on the landslide data, are presented in Figure 12.

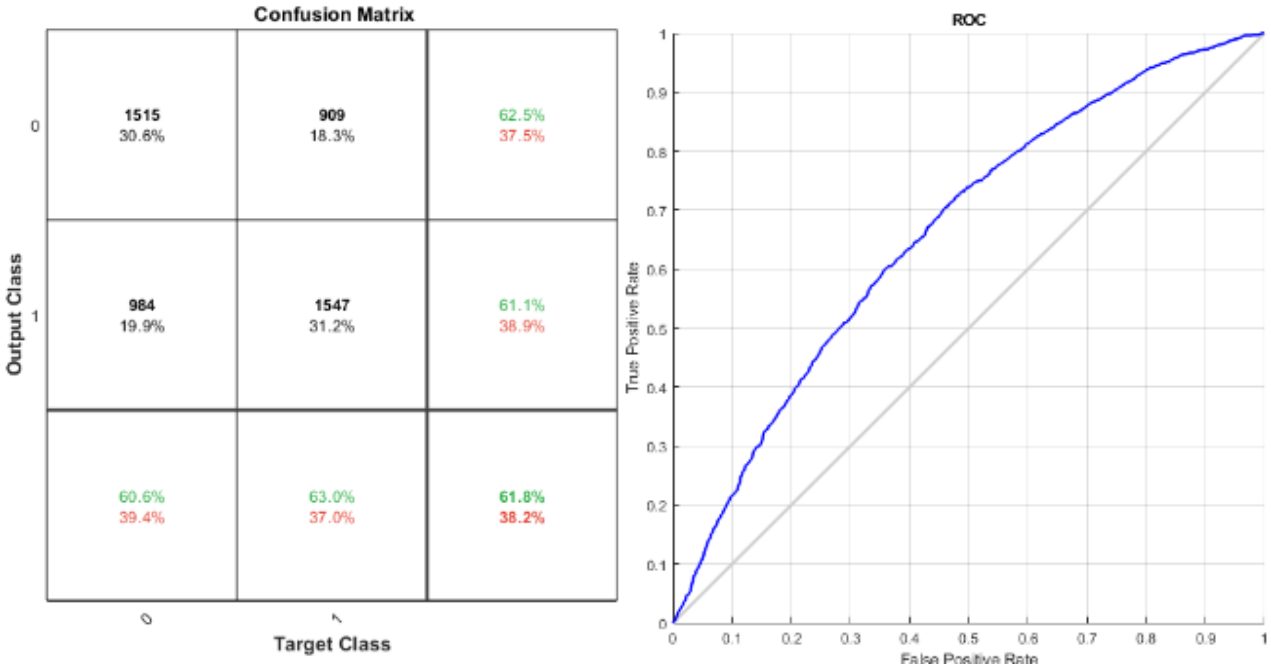


(a)

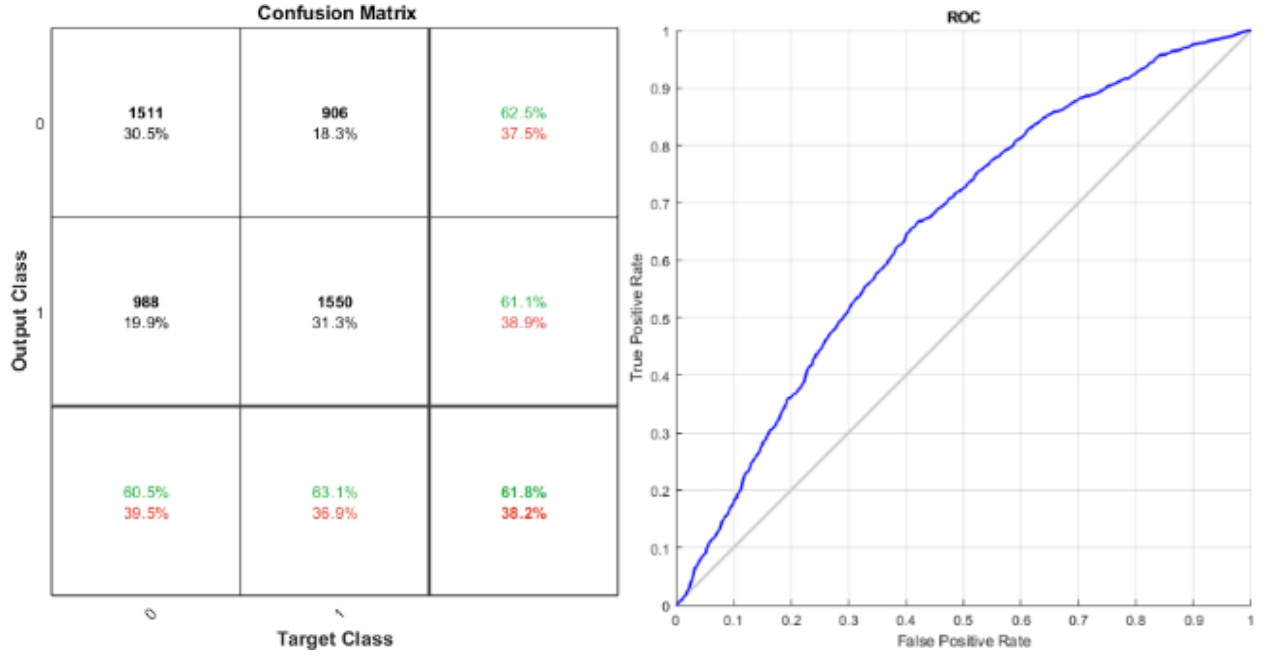


(b)

Figure 10. Performance accuracy of developed models based on IRIS data using (a Logistic Regression = Overall Accuracy 100%, Cohen Kappa Accuracy 100%, and ROC-AUC Accuracy is 100%) and (b neural networks = Overall Accuracy 100%, Cohen Kappa Accuracy 100%, and ROC-AUC Accuracy is 100%)).

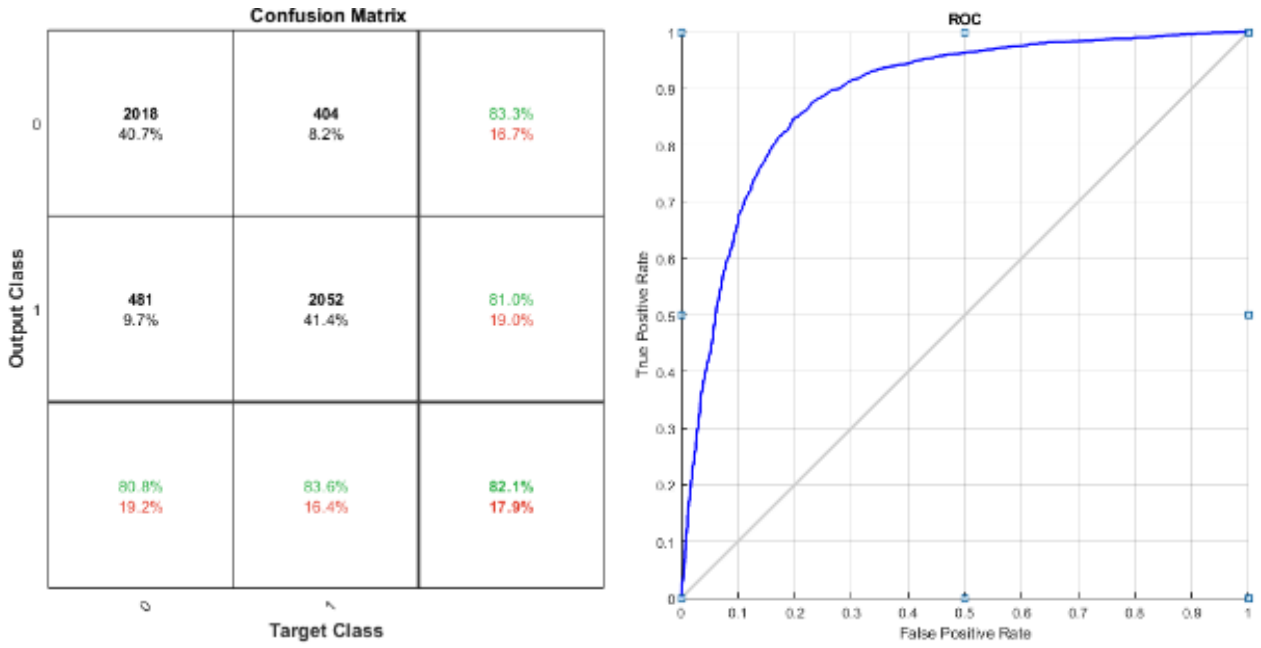


(a)

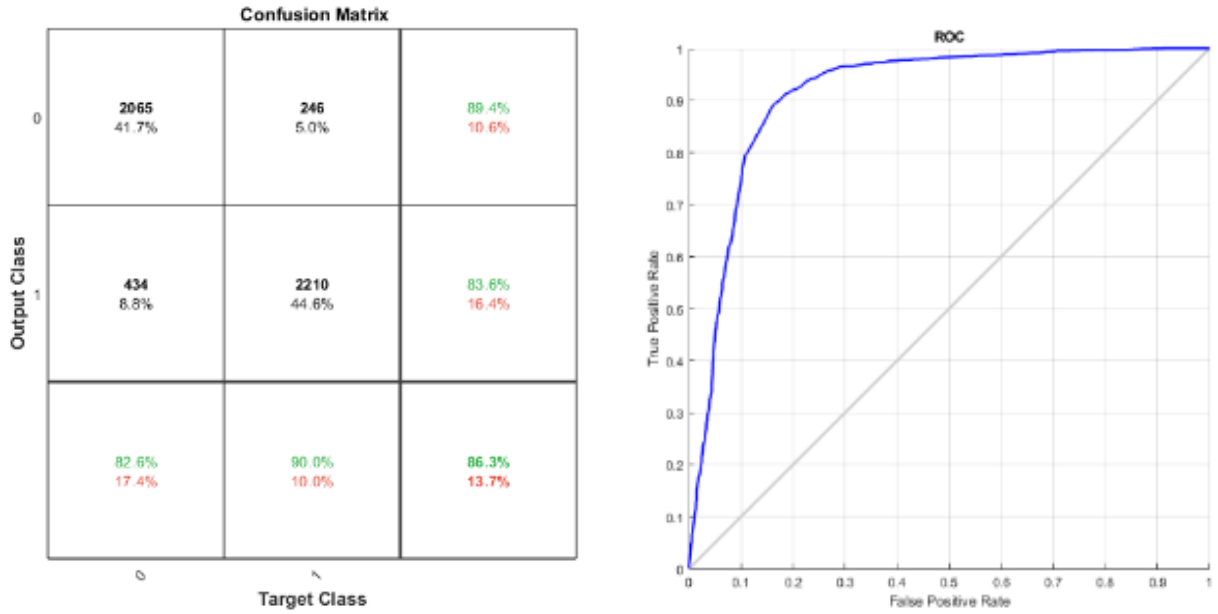


(b)

Figure 11. Performance accuracy of developed models based on landslides data from the study area using Logistic Regression (a 21 factors = Overall Accuracy 61.80%, Cohen Kappa Accuracy 23.60%, and ROC-AUC Accuracy is 66.20%) and (b 12 factors = Overall Accuracy 61.78%, Cohen Kappa Accuracy 23.56%, and ROC-AUC Accuracy is 65.31%))



(a)



(b)

Figure 12. Performance accuracy of developed models based on landslides data from the study area using NN (a 21 factors = Overall Accuracy 82.14%, Cohen Kappa Accuracy 64.28%, and ROC-AUC Accuracy is 88.82%) and (b 12 factors = Overall Accuracy 86.3%, Cohen Kappa Accuracy 72.56%, and ROC-AUC Accuracy is 91.39%))

The final landslide susceptibility map was produced using the developed NN model. The full maps of input factors were processed and extracted from ArcGIS 10.5 software to MATLAB R2019b as text files. The text files were processed using the chunk processing in Python programming language. The study area was divided into 15 chunks. After processing the chunks datasets using the trained NN model, the predicted susceptibility values were extracted to ArcGIS and reshaped to generate the landslide susceptibility map. The landslide susceptibility map is illustrated in Figure 13. The susceptibility levels vary between 0 and 1.

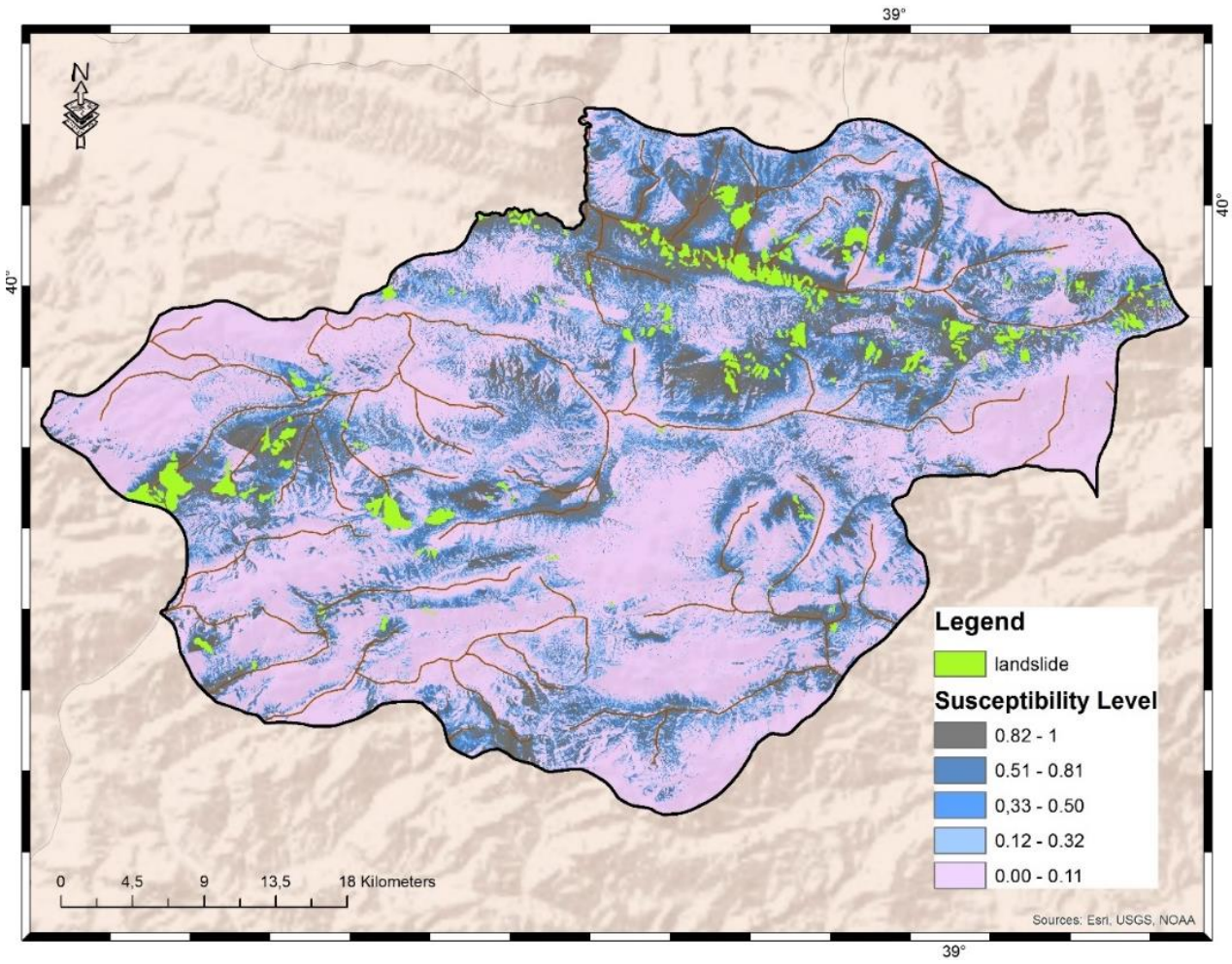


Figure 13. Landslide susceptibility map of Refahiye district in Turkey

4. Conclusion

Landslides are among the most common negative natural phenomena that cause loss of lives and property. In some cases, the damage caused by landslides are greater than that caused by earthquakes. Mapping the landslide susceptibility is an essential issue to avoid or manage the landslides risks. The aim of this study is to produce a highly-generalized model for landslide susceptibility mapping in the Refahiye district of Turkey. The model shallow neural network was used for landslide susceptibility mapping, while bivariate spearman correlation test was employed to select the important factors that enabled less data collection and less computation time. Using the Spearman correlation test, 12 out of 21 factors were selected as relevant factors. Important factors are geology, distance from roads, distance from geological faults, distance from water streams, flow direction, aspect, hillshade, heat load index, slope/aspect transformation, site exposure index, compound topographic index and elevation. The generated dataset was divided using a cross-validation method to test the developed model. The Levenberg-Marquardt backpropagation (Trainlm) was found to be the best training function with a performance accuracy of 86.3% when using 50 neurons in the hidden layer. The generated NN model was tested using the IRIS benchmark dataset and compared against the logistic regression algorithm. The NN model presented a better performance than the logistic regression and shows higher accuracy with the IRIS benchmark dataset. As a result, shallow neural networks method was successfully applied in landslide susceptibility mapping in this study and the method is recommended for future studies.

References

Can, A., Dagdelenler, G., Ercanoglu, M., & Sonmez, H. (2019). Landslide susceptibility mapping at Ovacık-Karabük (Turkey) using different artificial neural network models: comparison of training algorithms. *Bulletin of Engineering Geology and the Environment*, *78*(1), 89–102.

Chae, B.-G., Park, H.-J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, *21*(6), 1033–1070.

McCormick, C. (2016, February 24). *K-fold cross-validation, with Matlab code*. Retrieved from https://chrisjmccormick.wordpress.com/2013/07/31/k-fold-cross-validation-with-matlab-code/#

Dağ, S., & Bulut, F. (2012). Coğrafi Bilgi Sistemleri Tabanlı Heyelan Duyarlılık Haritalarının Hazırlanmasına Bir Örnek: Çayeli (Rize, KD Türkiye). Jeoloji Mühendisliği Dergisi, 36(1), 35–62.

Felicísimo, Á., Cuartero, A., Remondo, J., & Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides*, *10*(2), 175–189.

Hamad, R., Balzter, H., & Kolo, K. (2018). Predicting Land Use/Land Cover Changes Using a CA-Markov Model under Two Different Scenarios. *Sustainability*, *10*(10), 3421, doi: 10.3390/su10103421

JAXA. (2019, April 9). *About ALOS - Overview and Objectives*. Retrieved from https://www.eorc.jaxa.jp/ALOS/en/about/about_index.htm

Kim, D. E., & Gofman, M. (2018, January). Comparison of shallow and deep neural networks for network intrusion detection. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), 2018*. pp. 204–208. IEEE.

Lee, D.-H., Kim, Y.-T., & Lee, S.-R. (2020). Shallow Landslide Susceptibility Models Based on Artificial Neural Networks Considering the Factor Selection Method and Various Non-Linear Activation Functions. *Remote Sensing, 12*(7), 1194. doi: 10.3390/rs12071194

Liang, D., Tsai, C.-F., & Wu, H.-T. (2015). The effect of feature selection on financial distress prediction. *Knowledge-Based Systems*, *73*, 289–297.

Nefeslioglu, H. A., Gokceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology*, *97*(3), 171–191.

Nefeslioglu, H. A., San, B. T., Gokceoglu, C., & Duman, T. Y. (2012). An assessment on the use of Terra ASTER L3A data in landslide susceptibility mapping. *International Journal of Applied Earth Observation and Geoinformation*, *14*(1), 40–60.

Pitasi, A. (2016). *Susceptibility analysis to identify the zones potentially exposed to rapid flowslide risk* (Doctoral dissertation). Mediterranea University of Reggio Calabria, Italy. https://doi.org/10.13140/RG.2.2.11505.28004

Pradhan, B., & Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software*, *25*(6), 747–759.

Song, K.-Y., Oh, H.-J., Choi, J., Park, I., Lee, C., & Lee, S. (2012). Prediction of landslides using ASTER imagery and data mining models. *Advances in Space Research*, *49*(5), 978–993.

Duman, T. Y., Çan, T., & Emre, Ö. (2011). 1:1,500,000 Scale Landslide Inventory Map of Turkey. *General Directorate of Mineral Research and Exploration*, Special Publication Series-27, Ankara, Turkey.

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768.

Turner, A. K., & Schuster, R. L. (1996). *Landslides: investigation and mitigation*. Washington: National academy Press.

Vakhshoori, V., Pourghasemi, R. H., Zare, M., & Blaschke, T. (2019). Landslide Susceptibility Mapping Using GIS-Based Data Mining Algorithms. *Water, 11*(11), 2292, doi: 10.3390/w11112292.

Valencia Ortiz, J. A., & Martínez-Graña, A. M. (2018). A neural network model applied to landslide susceptibility analysis (Capitanejo, Colombia). *Geomatics, Natural Hazards and Risk*, *9*(1), 1106–1128.

Wang, L., Jia, Y., Yao, Y., & Xu, D. (2019). Accuracy Assessment of Land Use Classification Using Support Vector Machine and Neural Network for Coal Mining Area of Hegang City, China. *Nature Environment and Pollution Technology*, *18*(1), 335–341.

Wise, S. (2011). Cross-validation as a means of investigating DEM interpolation error. *Computers & Geosciences*, *37*(8), 978–991.

Yalcin, A, Reis, S., Aydinoglu, A. C., & Yomralioglu, T. (2011). A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena*, *85*(3), 274–287.

Yalcin, Ali. (2008). GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena*, *72*(1), 1–12.

Yilmaz, I. (2009). A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. *Bulletin of Engineering Geology and the Environment*, *68*(3), 297–306.

Yıldırım, Ü., & Güler, C. (2016). Identification of suitable future municipal solid waste disposal sites for the Metropolitan Mersin (SE Turkey) using AHP and GIS techniques. *Environmental Earth Sciences*, *75*(2), 101, doi: 10.1007/s12665-

015-4948-8.

Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., & Zhang, S. (2018). Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. *Sensors*, *18*(11), 3717, doi: 10.3390/s18113717

Zhang, X.-D. (2020). *A Matrix Algebra Approach to Artificial Intelligence*. Springer, Singapore.

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, September 2020, 1(2): 78-84
Journal homepage: http://www.dergipark.org.tr/en/pub/rsgis

NDVI Analysis of Australian Bushfires with Cloud Computing

Bulut Tabanlı Hesaplama ile Avustralya Orman Yangınlarının NDVI Analizi

Nusret Demir[1*]

[1] *Akdeniz University, Faculty of Science, Department of Space Science and Technologies, 07058, Antalya/Turkey.*

ORIGINAL PAPER

***Corresponding author:**
Nusret Demir
nusretdemir@akdeniz.edu.tr

doi:

Article history
Received: 05.04.2020
Accepted: 20.08.2020
Published: 30.09.2020

Abstract

Our planet is constantly exposed to massive forest fires, which threaten the natural ecosystem. Determining damages of forest fires has been a subject of extensive research for many years. Remote sensing is one of the effective technologies used for monitoring forest fires. However, accessing and processing data are both costly and time consuming. Therefore, the use of cloud technologies for this purpose is beneficial for rapid response. Australia experienced a series of wildfires from June 2019 to February 2020. These fires are considered as one of the biggest disasters of our age. In our study, Landsat data was used to track the trend of fires across the entire timeline during forest fire events. The Google cloud platform Google Earth Engine was used to obtain the results. Landsat 8 images were processed for each month from June 2019 to March 2020. Landsat 5 images were used to eliminate the clouds. Thus, from June-2019 to March-2020, all images were processed and the damaged areas were determined by NDVI and vegetation analysis. The forest cover reference data of previous years were used for the NDVI threshold value in the study.

Keywords: *Remote sensing, Forest fire, Google Earth Engine*

Özet

Gezegenimiz sürekli olarak büyük orman yangınlarına maruz kalmakta ve bu da doğal ekosistemi tehdit etmektedir. Orman yangınlarının zararlarının belirlenmesi uzun yıllardır geniş bir araştırma konusu olmuştur. Uzaktan algılama, orman yangınlarını izlemek için etkili teknolojilerden biridir. Fakat veriye ulaşma, işleme hem maliyetli hem de zaman alıcı işlemlerdir. Bu yüzden bulut teknolojilerinin bu amaçla kullanımı hızlı müdahale için faydalı olmaktadır. Avustralya, Haziran 2019'dan Şubat 2020'ye kadar bir dizi orman yangını yaşamıştır. Bu yangın dizisi çağımızın en büyük felaketlerinden birisi olarak gösterilmektedir. Çalışmamızda, orman yangını olayları sırasında tüm zaman çizgisi boyunca yangının eğilimini izlemek için Landsat verileri kullanılmıştır. Sonuçları elde etmek için Google bulut platformu Google Earth Engine kullanılmıştır. Haziran 2019'dan Mart 2020'ye kadar olan görüntüler her ay için Landsat 8 TOA görüntüleri işlenmiştir. Bulutların elemine edilmesi için de Landsat 5 görüntülerinden faydalanılmıştır. Böylece, Haziran-2019'dan Mart-2020'ye kadar tüm görüntüler işlenerek NDVI ve bitki örtüsü analizi ile hasar gören bölgeler belirlenmiştir. Çalışmada NDVI eşik değeri için geçmiş yıllara ait orman örtüsü referans verisi kullanılmıştır.

Anahtar kelimeler: *Uzaktan algılama, Orman yangını, Google Earth Engine*

1. Introduction

Australia suffered a series of bushfires from June 2019 up to February 2020, and still continuing. Australia has been facing with such disasters starting from the first cases reported in the 17th century. The recent fire caused the destruction of ca. 110,000 sq.km forest area with ca.6000 buildings, killing at least 34 people (Wikipedia, 2020). The forest cover of Australia includes various types of vegetation, including acacia, callitris, casuarina, eucalypt mallee, woodland, mangrove, melaleuca, rainforest and other native forests (Australian Government, 2019). Remote sensing is one of the efficient technologies used to monitor the wildfires.

The literature reports studies that use active or passive remote sensing datasets to identify the effects of the fires by reporting the total amount of the burned areas, forest types etc. Allison et al. (2016) presented a detailed review on optical and thermal remote sensing for wildfire detection and monitoring. They point the importance of the remote sensing with use of several types of sensors, such as Hyper-spectral cameras, image intensifiers and thermal cameras. Leblon (2001) summarized the studies that used remote sensing to monitor forest wildfire. His review is based on active and passive remote sensing methods, including NOAA-AVHRR NDVI images. Chuvieco and Congalton (1989) derived a forest fire hazard map using Landsat Thematic Mapper datasets. Sunar and Özkan (2001) analyzed the forest fire occurred in the Marmaris province of Muğla in July 1996 using Landsat Thematic Mapper dataset. Akther and Hassan (2011) evaluated several sensors for predicting the fire danger conditions over boreal forest regions of Alberta during the period 2006-2008 with use of surface temperature (TS), normalized multiband drought index (NMDI) and temperature vegetation wetness index (TVWI) measures.

The burned and fire-affected forest areas are identified with several measures, including the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), normalized burn ratio (NBR), integrated forest index (IFI) (Chen et al., 2011) etc. The enhanced vegetation index (EVI) is a measure that uses the atmospheric resistance coefficients, red bands, infrared bands and canopy background parameter to quantify vegetation greenness (USGS, 2020). This index was previously used to identify vegetation health (Amalo et al. 2018; Ambika and Mishra, 2019), environmental impacts (Xiao et al. 2003; Zhang et al. 2016) and also damage caused by forest fires (Caccamo et al. 2015; Chen et al. 2011; Wittenberg et al. 2007). Another measure, normalized burn ratio (NBR), is also highly mentioned and has been used in the previous literature to investigate damages caused by the forest fires.

Illera et al. (1996) analyzed temporal Normalized Difference Vegetation Index (NDVI) values derived from the Advanced Very High Resolution Radiometer (AVHRR) to assess the fire risk over forest areas. Matricardi et al. (2010) identified the impact of forest fires on natural forests from the southern Brazilian Amazon state of Mato Grosso by using a 13-year series of annual Landsat images. Zhu et al. (2012) analyzed Landsat 7 datasets to monitor the damages caused by the forest fires on a Savannah River site.

Detecting the vegetation from NDVI needs some thresholds and researchers apply different variants to achieve the most reliable results (Remmel and Perera, 2001). Escuin et al. (2008) utilized statistical analysis to determine the thresholds from the pixel values of the group of fires. Fernández et al. (1997) emphasized the importance of the study area and the used kernel sizes in the determination of the thresholds. Parto et al. (2020) proposed a series of thresholds to detect the level of fire damage based on MODIS dataset. Enhanced vegetation index is another measure, which is sensitive to the leaf area index (Chen et al. 2011).

Previous studies showed that damage detection mostly relied on NDVI changes and risk modeling was conducted with other indices. In this study, Landsat datasets acquired during the bushfire events is used to monitor the fire trends through a specific timeline. We use Google Earth Engine, a Google Cloud Computing Platform, to derive the results. Because Google Earth Engine provides quick access to its rich geospatial data catalog, and tools for algorithm implementations in JavaScript that saves the hardware and software costs for the users. The literature reported studies that used GEE for forest fire monitoring, including fire severity metrics (Daldegan et al. 2019; Parks et al. 2018) and assessment of fire damages (Liu et al. 2020; Long et al. 2019; Quintero et al. 2019). We calculate NDVI along the given timeline without any empirically selected thresholds. Our method estimates the threshold used to derive the vegetation areas by using the data collected from the field before the fire event. The objective of our study is to investigate the change in the vegetation cover using the NDVI values of Landsat 8 Top of Atmosphere dataset with the use of Google Earth Engine platform.

2. Materials and Methods

The study area is defined as the whole Australia. The used data are obtained from Google Earth Engine USGS Landsat 8 Collection 1 Tier 1 TOA Reflectance catalog. The NDVI values are calculated using the red and near-infrared bands. The images are directly loaded with JavaScript codes and analyzed accordingly. Normalized difference vegetation index is a common measure used to identify the vegetation cover and also their health. In this study, the change in the vegetation cover during the bushfire period is used to assess the effect of the fire. The images acquired from June 2019 to March 2020 (a total of 10 images) are processed accordingly for each month, and respective NDVI and vegetation land cover regions, which are mainly bushes, are detected. The methodology is shown in Figure 1.

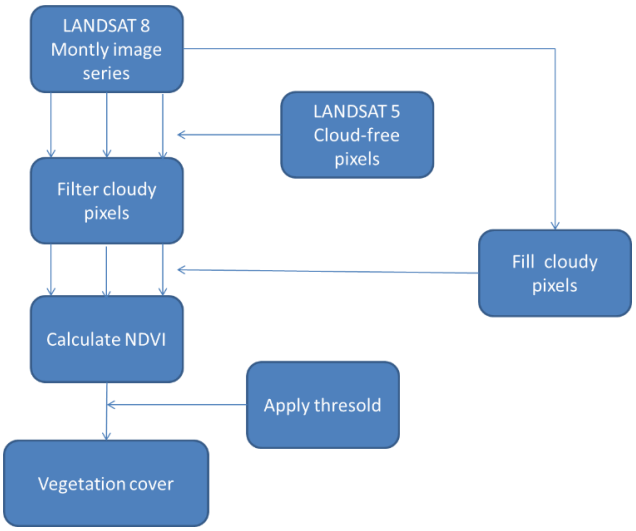


Figure 1. Method workflow

The method starts by loading images from Google Earth Engine Landsat 8 Catalog. The cloudy pixels are filtered out using Landsat 5 images acquired in 2012. For the elimination of the pixels with clouds, the cloud-free Landsat 5 datasets were used. The difference value between Landsat 5 and Landsat 8 images was calculated. The difference value which is larger than a certain threshold in the blue band gave the pixels which contain the clouds (Karpov, 2017). The cloudy pixels are filled with three different methods, including mosaicking, calculation of the mean and calculation of the median values of all images within the analyzed month. Vegetation cover is sensitive to NDVI values. Index values above zero indicates any type of vegetation. The aim of this study is to identify the changes in main forest cover, not any other low vegetation like grass. A reference dataset provided from the Australian Government is used to estimate the threshold value. This dataset includes all information about the forest types in Australia. This dataset is in grid format and shown in Figure 2.

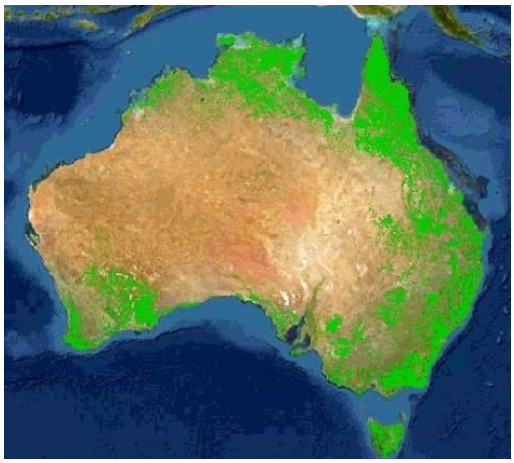


Figure 2. The forest cover of Australia in 2018 (Australian Government, 2019)

According to the Australian Government Report, the forests covered an area of 134,037,200 ha in 2018. Here, we calculate the NDVI of the June 2018, and estimated the threshold value to identify the vegetation cover. The threshold value of 0.150884999 is found to give the most accurate value (i.e. 134,037,436 ha) for the total forest covered area of June 2018. Then, we use this selected threshold to monitor and calculate the vegetation coverage throughout the selected timeline, from June 2019 to February 2020.

3. Results

The change in the forest area is monitored through the NDVI values calculated from June 2019 to February 2020. The calculated NDVI images are shown in Figure 3. In this figure, the green color represents the vegetation cover.

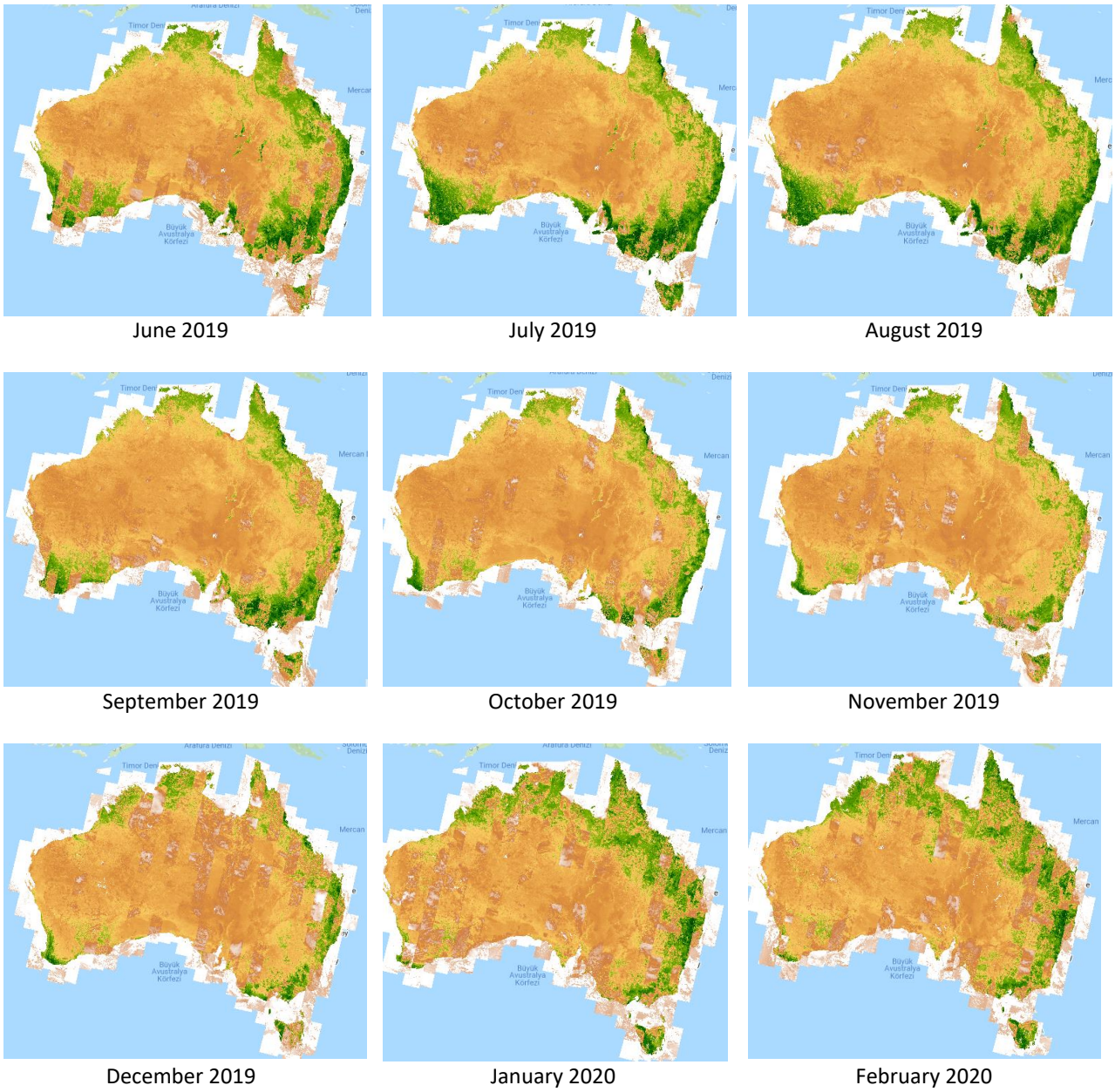


Figure 3. The NDVI images for the timeline between June 2019 and February 2020, green: vegetation cover

After applying the selected threshold, the areas covered by forests are calculated and shown in Table 1. The graph of the change in forested areas is shown in Figure 4.

Table 1. Total forest areas estimated from NDVI Analysis (1000 ha)

Month/Method	Mosaicing	Use of Mean	Use of Median
June 2019	171301	146042	156419
July 2019	182697	167964	175030
August 2019	198711	179303	187481
September 2019	195057	195314	202000
October 2019	187259	177212	186352
November 2019	170370	152604	166417
December 2019	148678	137358	149730
January 2020	114755	92216	104631
February 2020	162376	186341	202826

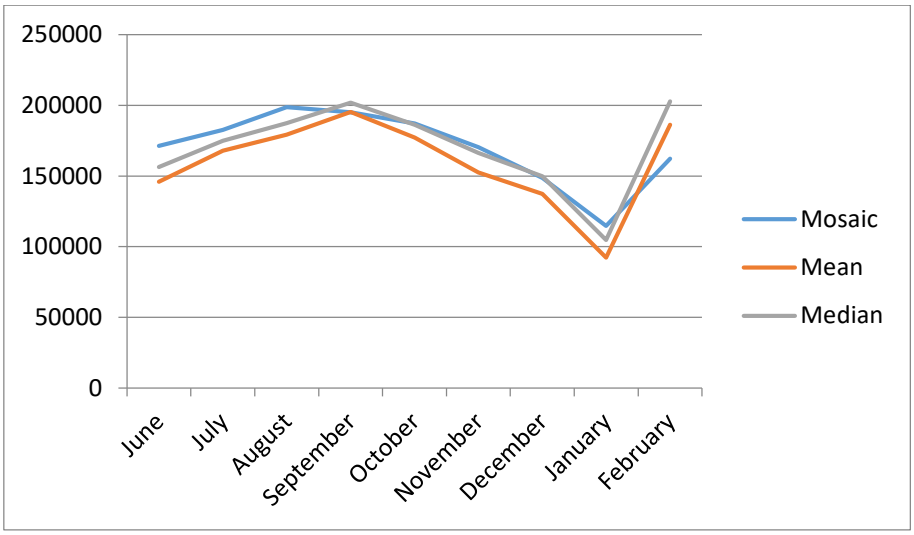


Figure 4. Total areas of estimated high vegetation cover (1000 ha) from three different criteria

As shown in Figures 3 and 4, the forestry cover changed dramatically from June to January and the highest peak observed on December 2019. he results show that new vegetation regions are increased from January to February 2020, which is needed to be confirmed with ground-truth validation since it is not discussed in this paper.

4. Discussion

Vegetation cover change studies have already a large portion in the previous remote sensing literature. The state-of-art is already significant, and provides large scale information about the outcomes of the wildfires over forestry. This study investigates the advantages of the Google's Cloud Computing Earth Observation Platform in producing the results in a short span of time with high performance.

The use of GEE allows for

- downloading any satellite image dataset
- processing data without the need of an external software
- the visualization of the data online, and sharing the results with third-parties.

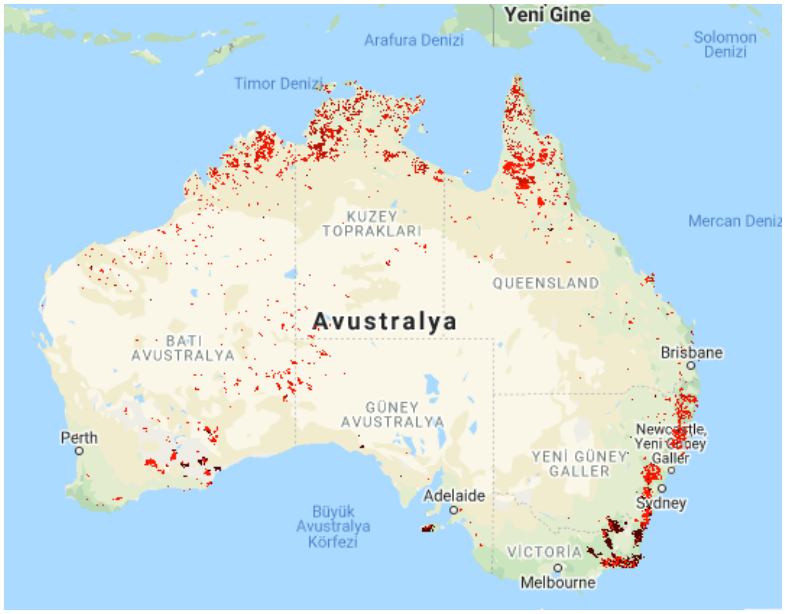


Figure 5. Damaged areas obtained by MODIS-based results

According to the report published by BBC (2020), the total burned area is ca. 110,000 sq.km, the area calculated by subtracting the forest cover binary image of the January 2020 from that of the September 2019 is 103,098 sq.km in our study. This shows the use of GEE gives efficient, quick, and reliable response for forest fire damage assessment. On the other hand, there is a real time bushfire monitoring system provided by Australian Government called 'Australia Sentinel Hotspots', which provides live fire monitoring system with use of AVHRR, MODIS (Figure 5) and VIIRS datasets.

5. Conclusion

Australia has been facing with bushfires regularly happen from past to present. The first fire was reported in 1642 and a huge fire, which is called 'Black Thursday', was reported in 1851 around the region of Virginia. During the 20. Century, Australia also faced with the high-extent forest fires in the years of 1925, 1938 and 1966 (Wikipedia, 2020). A series of bushfire events occurred during the second half of 2019 and beginning of 2020, affecting the whole country. This was reported as one of the biggest disasters that humanity faced in 2020. As presented in this paper, the climate change and other external factors can be investigated with enough available data related to the occurred natural hazards. Remote sensing methodologies and satellite datasets provide powerful functionalities to assess the damages caused by the wildfires. Two important factors for efficient monitoring is the availability of the datasets and the software packages. The methodology used in this study is based on open-access datasets and Google Earth Engine, an open-access cloud-based remote sensing data analysis platform. The future studies may focus on the identification of the damages of the wildfires in global scale.

References

Akther, M. S., & Hassan, Q. K. (2011). Remote sensing-based assessment of fire danger conditions over boreal forest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *4*(4), 992-999.

Allison, R. S., Johnston, J. M., Craig, G., & Jennings, S. (2016). Airborne optical and thermal remote sensing for wildfire detection and monitoring. *Sensors, 18*(8), 1310, doi: 10.3390/s16081310.

Amalo, L. F., Hidayat, R., & Sulma, S. (2018). Analysis of agricultural drought in east java using vegetation health index. *Agrivita*, *40*(1), 63-73.

Ambika, A. K., & Mishra, V. (2019). Observational Evidence of Irrigation Influence on Vegetation Health and Land Surface Temperature in India. *Geophysical Research Letters*, *46*, 13441-13451.

BBC. (2020, January 21). Australia fires: A visual guide to the bushfire crisis. Retrieved from https://www.bbc.com/news/world-australia-50951043.

Caccamo, G., Bradstock, R., Collins, L., Penman, T., & Watson, P. (2015). Using MODIS data to analyse post-fire vegetation recovery in Australian eucalypt forests. *Journal of Spatial Science*, *60*(2), 341–352.

Chen, X., Vogelmann, J. E., Rollins, M., Ohlen, D., Key, C. H., Yang, L., … Shi, H. (2011). Detecting post-fire burn severity and vegetation recovery using multitemporal remote sensing spectral indices and field-collected composite burn index data in a ponderosa pine forest. *International Journal of Remote Sensing*, *32*(23), 7905–7927.

Chuvieco, E., & Congalton, R. G. (1989). Application of remote sensing and geographic information systems to forest fire hazard mapping. *Remote sensing of Environment*, *29*(2), 147-159.

Daldegan, G. A., Roberts, D. A., & de Figueiredo Ribeiro, F. (2019). Spectral mixture analysis in Google Earth Engine to model and delineate fire scars over a large extent and a long time-series in a rainforest-savanna transition zone. *Remote Sensing of Environment*, *232*, 111340, doi: 10.1016/j.rse.2019.111340.

Escuin, S., Navarro, R., & Fernández, P. (2008). Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images. *International Journal of Remote Sensing*, *29*(4), 1053–1073.

Fernández, A., Illera, P., & Casanova, J. L. (1997). Automatic mapping of surfaces affected by forest fires in Spain using AVHRR NDVI composite image data. *Remote Sensing of Environment*, *60*(2), 153–162.

Australian Government. (2019, February 16). Forests of Australia. Retrieved from https://www.agriculture.gov.au/abares/forestsaustralia/forest-data-maps-and-tools/spatial-data/forest-cover.

Illera, P., Fernandez, A., & Delgado, J. A. (1996). Temporal evolution of the NDVI as an indicator of forest fire danger. *International Journal of remote sensing*, *17*(6), 1093-1105.

Karpov, A. (2017, August 8). Description code for article "Using the Google Earth Engine (GEE) for Detection of Burned Areas. Retrieved from https://digital-geography.com/description-code-article-using-google-earth-engine-gee-detection-burned-areas/.

Leblon, B. (2001). Forest wildfire hazard monitoring using remote sensing: A review. *Remote Sensing Reviews*, *20*(1), 1-43.

Liu, D., Chen, N., Zhang, X., Wang, C., & Du, W. (2020). Annual large-scale urban land mapping based on Landsat time series in Google Earth Engine and OpenStreetMap data: A case study in the middle Yangtze River basin. *ISPRS Journal of Photogrammetry and Remote Sensing*, *159*, 337-351.

Long, T., Zhang, Z., He, G., Jiao, W., Tang, C., Wu, B., ... & Yin, R. (2019). 30 m Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine. *Remote Sensing*, *11*(5), 489, doi: 10.3390/rs11050489.

Matricardi, E. A., Skole, D. L., Pedlowski, M. A., Chomentowski, W., & Fernandes, L. C. (2010). Assessment of tropical forest degradation by selective logging and fire using Landsat imagery. *Remote Sensing of Environment*, *114*(5), 1117-1129.

Parks, S. A., Holsinger, L. M., Voss, M. A., Loehman, R. A., & Robinson, N. P. (2018). Mean composite fire severity metrics computed with google earth engine offer improved accuracy and expanded mapping potential. *Remote Sensing*, *10*(6), 879, doi: 10.3390/rs10060879.

Parto, F., Saradjian, M., & Homayouni, S. (2020). MODIS Brightness Temperature Change-Based Forest Fire Monitoring. *Journal of the Indian Society of Remote Sensing*, *48(1)*, 163–169.

Quintero, N., Viedma, O., Urbieta, I. R., & Moreno, J. M. (2019). Assessing Landscape Fire Hazard by Multitemporal Automatic Classification of Landsat Time Series Using the Google Earth Engine in West-Central Spain. *Forests*, *10*(6), 518, doi: 10.3390/f10060518.

Remmel, T. K., & Perera, A. H. (2001). Fire mapping in a northern boreal forest: assessing AVHRR/NDVI methods of change detection. *Forest Ecology and Management*, *152*(1-3), 119-129.

Sunar, F., & Özkan, C. (2001). Forest fire analysis with remote sensing data. *International Journal of Remote Sensing*, *22*(12), 2265-2277.

USGS. (2020, August 8). Landsat Surface Reflectance-Derived Spectral Indices. Retrieved from https://www.usgs.gov/land-resources/nli/landsat/landsat-enhanced-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con.

Wittenberg, L., Malkinson, D., Beeri, O., Halutzy, A., & Tesler, N. (2007). Spatial and temporal patterns of vegetation recovery following sequences of forest fires in a Mediterranean landscape, Mt. Carmel Israel. *Catena*, *71*(1), 76-83.

Xiao, X., Braswell, B., Zhang, Q., Boles, S., Frolking, S., & Moore III, B. (2003). Sensitivity of vegetation indices to atmospheric aerosols: continental-scale observations in Northern Asia. *Remote Sensing of Environment*, *84*(3), 385-392.

Wikipedia. (2020, April 5). Bushfires in Australia. Retrieved from https://en.wikipedia.org/wiki/Bushfires_in_Australia.

Zhang, T., Gong, W., Wang, W., Ji, Y., Zhu, Z., & Huang, Y. (2016). Ground level PM2. 5 estimates over China using satellite-based geographically weighted regression (GWR) models are improved by including NO2 and enhanced vegetation index (EVI). *International journal of environmental research and public health*, *13*(12), 1215, doi: 10.3390/ijerph13121215.

Zhu, Z., Woodcock, C. E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote sensing of environment*, *122*, 75-91.

2020
Turkish Journal of Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, Eylül 2020, 1(2): 85-96
Dergi Anasayfası: http://www.dergipark.org.tr/en/pub/rsgis

Meteosat LSA SAF DIDSSF Ürününün Türkiye İçin Tutarlılığının Değerlendirilmesi

Evaluation of the Accuracy of Meteosat LSA SAF DIDSSF Product for Turkey

Kazım Kaba[1]*, Derya Öztürk Çetni[2], H. Mustafa Kandırmaz[3]

[1] *Atatürk Üniversitesi, Astrofizik Araştırma ve Uygulama Merkezi, 25240, Erzurum/Türkiye.*
[2] *Atatürk Üniversitesi, Fen Fakültesi, Astronomi ve Uzay Bilimleri Bölümü, 25240, Erzurum/Türkiye.*
[3] *Çukurova Üniversitesi, Fen-Edebiyat Fakültesi, Fizik Bölümü, 01330, Adana/Türkiye.*

ARAŞTIRMA MAKALESİ

***Sorumlu yazar:**
Kazım Kaba
kkaba46@gmail.com

doi:

Yayın süreci
Geliş tarihi: 08.04.2020
Kabul tarihi: 12.08.2020
Basım tarihi: 30.09.2020

Özet

Yeryüzüne gelen güneş radyasyonu, iklim, tarım, hidroloji ve enerji uygulamaları için büyük öneme sahiptir. Bu konularda yapılan çalışmalar için enerji değerlerinin zamansal ve alansal dağılımının bilinmesi dünya genelinde ihtiyaç duyulan bir bilgidir. Genel olarak güneş enerjisi ölçümleri meteoroloji istasyonlarında noktasal olarak gerçekleştirilmektedir. Son yıllarda yapılan çalışmalar göstermiştirki yüzeye gelen güneş enerjisi değerleri uydu verileri kullanılarak başarıyla tahmin edilebilmektedir. Meteosat uyduları Türkiye'yi de kapsayacak şekilde Avrupa ve Afrika'ya ait görüntüler kaydetmektedir. Bu kayıtlardan LSA SAF birimi tarafından yeryüzüne ait çeşitli parametreler tahmin edilmektedir. Bu parametrelerden biri olan yüzeye gelen günlük kısa dalga enerjinin tahmin edildiği DIDSSF verisidir. Bu çalışmanın amacı Türkiye için DIDSSF ürünün yer ölçümleri ile kıyaslayarak doğruluğunu belirlemektir. DIDSSF ürününün doğruluğu ülke coğrafyasına olabildiğince homojen dağılmış olan 46 adet MGM istasyonu ile ölçülen global güneş radyasyonu değerleri ile test edilmiştir. İstasyonlardan gelen değerler tek tek incelendiğinde, belirleme katsayısı olan R^2'nin 0,7129 ila 0,9585 arasında değişim gösterdiği tespit edilmiştir. 46 adet istasyondan elde edilen ortalama R^2 değeri ise 0,9058 olarak bulunmuştur.

***Anahtar kelimeler**: Güneş radyasyonu, DIDSSF, Meteosat, Türkiye, Validasyon*

Abstract

The solar radiation that reaches to the surface of the Earth is of great importance for many applications such as climate, agriculture, hydrology and energy. Having knowledge on the temporal and spatial distribution of energy values is a worldwide need for the studies on these topics. Measurement of the solar radiation is generally carried out in a pointwise manner in meteorological stations. Recent studies show that solar radiation over large areas can be estimated successfully using satellite images. The Meteosat satellites record images of Europe and Africa, covering also Turkey. Using these records, the LSA SAF unit predicts various parameters that belongs to Earth. One of these parameters is the DIDSSF (daily downward surface short-wave flux), which predicts the daily short-wave radiation reaching to the Earth surface. This study aimed to evaluate the accuracy of the DIDSSF estimation for Turkey by comparing it against terrestrial measurements. The accuracy of the DIDSSF product was tested through the use of global solar radiation values measured at 46 MGM stations which were homogeneously distributed over the whole country as much as possible. Examination of the stations one by one revealed that the determination coefficient R^2 varied between 0.7129 and 0.9585. The average R^2 value obtained from 46 stations was found to be 0.9058.

***Keywords**: Solar radiation, DIDSSF, Meteosat, Turkey, Validation*

1. Giriş

Yeryüzüne gelen güneş radyasyonu, iklim, tarım, hidroloji ve enerji uygulamaları için büyük öneme sahiptir. Bu konularda yapılan çalışmalar için enerji değerlerinin zamansal ve alansal bakımdan bilinmesi dünya genelinde ihtiyaç duyulan bir bilgidir. Bu kapsamda herhangi bir yüzeye ait enerji değerlerinin ölçümü dünyada noktasal ve alansal olmak üzere farklı yöntem ve aletler ile yapılmaktadır. Noktasal ölçümlere, meteorolojik yer istasyonlarında yapılan ölçümler ve alansal ölçümlere de uydu görüntülerinden hesaplanan değerler örnek verilebilir. Güneş enerjisi değerleri doğrudan ölçüm yapılmasının yanı sıra interpolasyon, regresyon ve yapay öğrenme gibi çeşitli tahmin yöntemleriyle de meteorolojik parametrelerden türetilmektedir (Gautier vd. 1980; Behrang vd. 2010; Rusen vd. 2013; Kaba vd. 2018).

Türkiye güneş enerjisi değerleri bakımından oldukça yüksek potansiyele sahip bir coğrafyadadır. Son yıllarda bu potansiyelin kullanımına yönelik yatırımlar ile çeşitli bölgelerimizde güneş panelleri kurulmaktadır. Bunun gibi sistemlerin kurulumu için en uygun yerlerin belirlenmesi amacıyla başta enerji değerleri olmak üzere uzun dönem iklim verilerinin kullanılması gerekmektedir. Başarılı sonuçlar elde etmek için ise girdi olarak kullanılan verilerin alansal ve zamansal sürekliliğin yanı sıra yüksek doğrulukta olması gerekmektedir. Ülkemizde güneş ışınım şiddeti meteorolojik istasyonlarda noktasal olarak ölçülmektedir. Çeşitli nedenlerden dolayı her ölçümün sistematik ve/veya beklenmeyen hata barındırması ise olası bir durumdur. Bu nedenle ölçüm verilerinin kullanılmadan önce kalibrasyonu yapılmalıdır. Alansal sürekliliğe sahip veriye ihtiyaç duyulduğunda ise mevcut noktasal ölçüm değerleri kullanılarak veri olmayan yerler için interpolasyon yöntemleriyle tahminler yapılmaktadır. Bu yöntemle yapılan tahminler özellikle ölçüm yapılan istasyonlardan uzaklaştıkça yüksek hatalar içermektedir. İnterpolasyon yöntemi ile üretilen enerji haritaları hassas ve kritik önem taşıyan çalışmalar için yetersiz kalabilmektedir. Buna karşın, günümüzde uydu kayıtlarının kullanılmasıyla düzenli ve kararlı yapıda olan bu veri setleri ile daha başarılı sonuçlar elde edilebilir. Uydu verilerinin kapsadığı alan, zamansal ve alansal sürekliliğinden, maliyeti ve sayısal olması bakımından kullanılması büyük avantajlar sağlamaktadır.

Güneş enerjisi tahminlerine yönelik çalışmalar dünyada ve ülkemizde uzun yıllardır yapılmaktadır. Araştırmacılar, geçmişten günümüze çeşitli yöntemlerle yeryüzüne gelen güneş enerjisini tahmin etmişlerdir. Geleneksel yöntem diye bilinen ve öncülüğünü Angström-Prescott metodunun yaptığı, güneşlenme süresi, bulutluluk ve sıcaklık gibi meteorolojik değişkenleri kullanan lineer ve lineer olmayan istatistiksel çalışmalar bir hayli fazladır (Angstrom, 1924). Bir diğer yöntem ise süper bilgisayarların kullanıldığı atmosferik ve meteorolojik verilerin simülasyonunu yapan fiziksel yöntemlerdir (Gautier vd. 1980; Cogliani vd. 2007). Çalışmalar da yer ölçümleri ve uydu verileri veya her ikisi de birlikte girdi olarak kullanılabilmektedir (Rusen vd. 2013). Kullanılan veri tipine göre modeller yer, uydu ya da hibrit olarak tanımlanmaktadır. Yöntemlerin birbirine göre avantaj ve dezavantajları olmakla beraber hepsi de gerçekte bir eksikliği gidermekte veya yer ve uydu kayıtlarına alternatif veriler üretmektedir. Uydular yörüngelerine göre yer sabit ve kutupsal yörüngeli olarak ikiye ayrılmaktadır. Yer sabit yörüngeli uydular daima yeryüzünün aynı bölgesini kayıt ederek yüksek zamansal çözünürlükte veri üretirken yeryüzünün diğer bölümlerine ait kayıt yapılmaması ve yersel çözünürlüğünün nispeten düşük olması bu uyduların dezavantajıdır. Kutupsal yörüngeli uydular ise yeryüzünün her bölgesine ait yüksek yersel çözünürlükte görüntü alırlar fakat zamansal çözünürlük bakımından (günde 1-2 kayıt) dezavantajlı durumdadırlar. Yer ölçümlerinin ise noktasal olup sadece ölçüm yapılan istasyonların çevresini temsil etmesi, istasyon ve cihazlara bağlı olarak ölçümlerin homojen olmayışı ve dağlık alanlar ile denizler gibi istasyon kurulamayan bölgeler için veri sağlanmaması dezavantaj oluşturmaktadır.

Son dönemlerde makine öğrenme algoritmaları, veri madenciliği ve görüntü işleme gibi birçok disiplinde olduğu gibi güneş ışınım değerlerini tahmin etmek için de yaygın bir şekilde kullanılmaktadır (Behrang vd. 2010; Quesada-Ruiz vd. 2015; Shamshirband vd. 2016; Kaba vd. 2018; Wu vd. 2019). Güneş enerjisi tahminleri için geliştirilen makine öğrenme algoritmalarından Yapay Sinir Ağları (YSA) ve Destek Vektör Makineleri (DVM) ve son zamanlarda derin öğrenme algoritmaları yaygın kullanılan yöntemlerdir. Yine bu yöntemlerde de yer ve uydu verileri kullanılarak enerji tahminleri yapılabilmektedir. Araştırmacılar çalışmalarında makine öğrenme algoritmaları ile yapılan regresyon sonuçlarının geleneksel regresyon analizine dayanan çalışmalardan daha iyi sonuçlar verdiğini göstermişlerdir (Behrang vd. 2010; Quesada-Ruiz vd. 2015; Shamshirband vd. 2016; Kaba vd. 2018; Wu vd. 2019; Feng vd. 2020).

Türkiye'nin de kurucu üyesi olduğu EUMETSAT (European Organization fort he Exploitation of Meteorological Satellites) tarafından işletilen yer sabit yörüngeli Meteosat uyduları Türkiye'yi de kapsayacak şekilde Avrupa ve Afrika'ya ait görüntüler kaydetmektedir. Bu kayıtlardan LSA SAF (Satellite Application Facility on Land Surface Analysis) tarafından yeryüzüne ait çeşitli ürünler üretilmektedir. Bu ürünlerden biri olan yüzeye gelen günlük kısa dalga enerji (DIDSSF, Daily Downward Surface Shortwave Flux) verisi bu çalışmanın konusunu oluşturmaktadır (Trigo vd. 2011). Çalışmada DIDSSF ürününün doğruluğu ülke coğrafyasına olabildiğince homojen dağılmış 46 adet MGM (Meteoroloji Genel Müdürlüğü) istasyonlarında ölçülen global güneş radyasyonu değerleri ile test edilmiştir.

LSA SAF DSSF ürünü için validasyon raporu (https://nextcloud.lsasvcs.ipma.pt/s/b6fkBGcBYxKi9fo#pdfviewer) ile İber yarımadası ve İspanya için yapılan çalışmalar bu ürün hakkında bilgi vermektedir (Cristóbal ve Anderson, 2013; Moreno vd. 2013). Cristóbal ve Anderson (2013) çalışmalarında İber yarımadasının kuzeydoğusu için DIDSSF ürününün validasyonunu yapmışlardır.

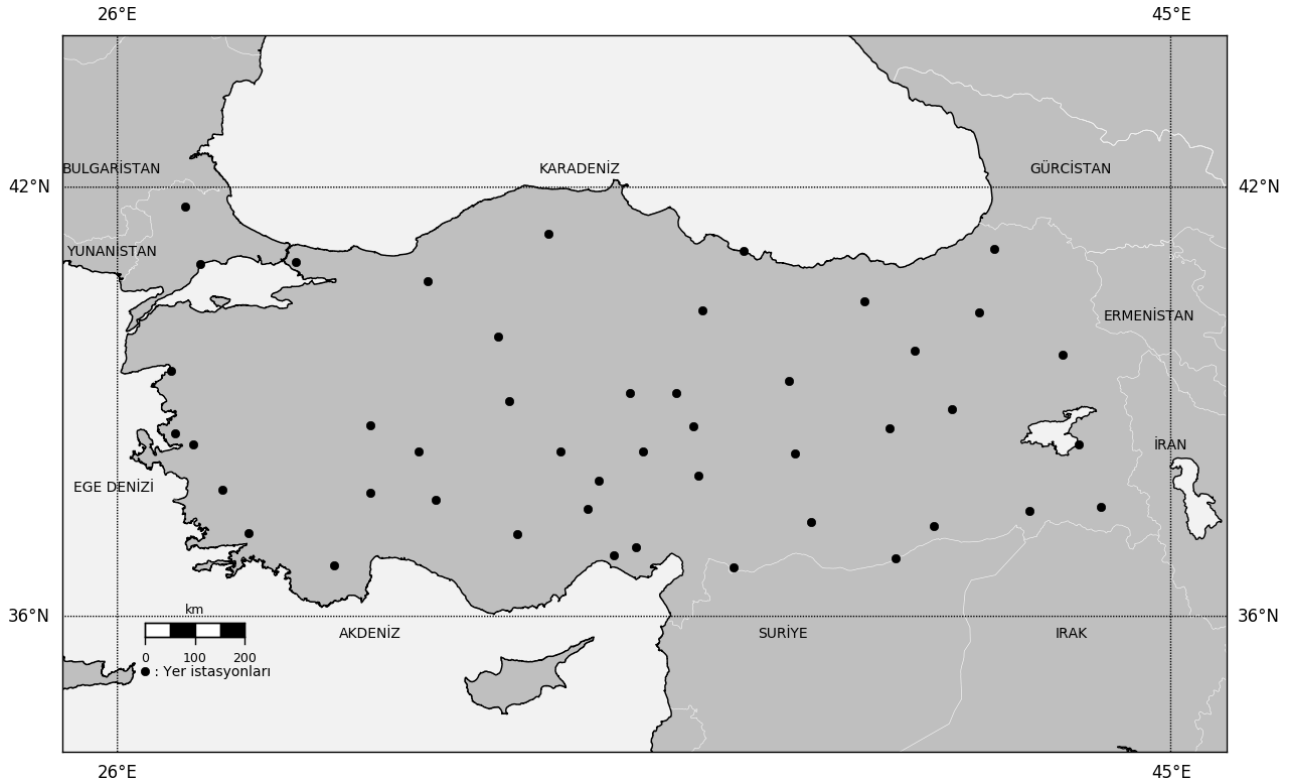
Çalışmada arazi türüne (düz, engebeli) ve hava şartlarına (Açık, bulutlu, Karlı) göre 2008-2010 verileriyle değerlendirme yapmışlardır. Tüm hava şartlarında MBE (Mean bias error), MAE (Mean absolute error), RMSE (Root mean square error) ve R^2 (Coefficient of determination) sırasıyla düz arazi için, -6 W/m^2, 24 W/m^2, 34 W/m^2 ve 0.96 ve engebeli arazi için -5 W/m^2, 35 W/m^2, 48 W/m^2, 0.93 olarak verilmiştir. Moreno vd. (2013) İspanya için yaptıkları çalışmada 2008 yıllı için DIDSSF ürününün analizini ve validasyonunu ele almışlardır. Çalışmada MBE, MAE, RMSE, rRMSE (Percent/relative root mean square error) ve R (R^2 ifadesinin karekökü) indekslerini -0.12 MJ/m^2, 1 MJ/m^2, 1.3 MJ/m^2, %8 ve 0.99 olarak vermişlerdir. Bu sonuçlar dikkate alındığında DIDSSF ürünü İber yarımadası ve İspanya için Türkiyeye göre daha yüksek doğruluk sağlamaktadır.

2. Materyal ve Yöntem

2.1 Çalışma Alanı ve Veri

Bu çalışmada DIDSSF (Daily Downward Surface Shortwave Flux) ürününün Türkiye için doğruluğunu tespit etmek amacıyla, 26-45^o doğu boylamları ile 36-42^o kuzey enlemleri arasında bulunan Türkiye bölgesi çalışma alanı olarak seçilmiştir. Uydudan elde edilen enerji değerlerinin yer ölçümleri ile tutarlılığını değerlendirmek için Ocak 2013 ile Ekim 2015 aralığını kapsayan yaklaşık üç yıllık (34 ay) zaman dilimine ait MGM tarafından işletilen 46 adet yer istasyonunda ölçülen günlük global güneş ışınımı değerleri ile Meteosat uydularından elde edilen ve yüzeye ulaşan günlük enerji miktarını veren DIDSSF ürünü çalışmada kullanılmıştır. Şekil 1'de çalışma alanını ve kullanılan istasyonlarının konumlarını gösteren harita, Tablo 1'de ise kullanılan yer istasyonlarının enlem, boylam, rakım ve şehir bilgileri verilmiştir.

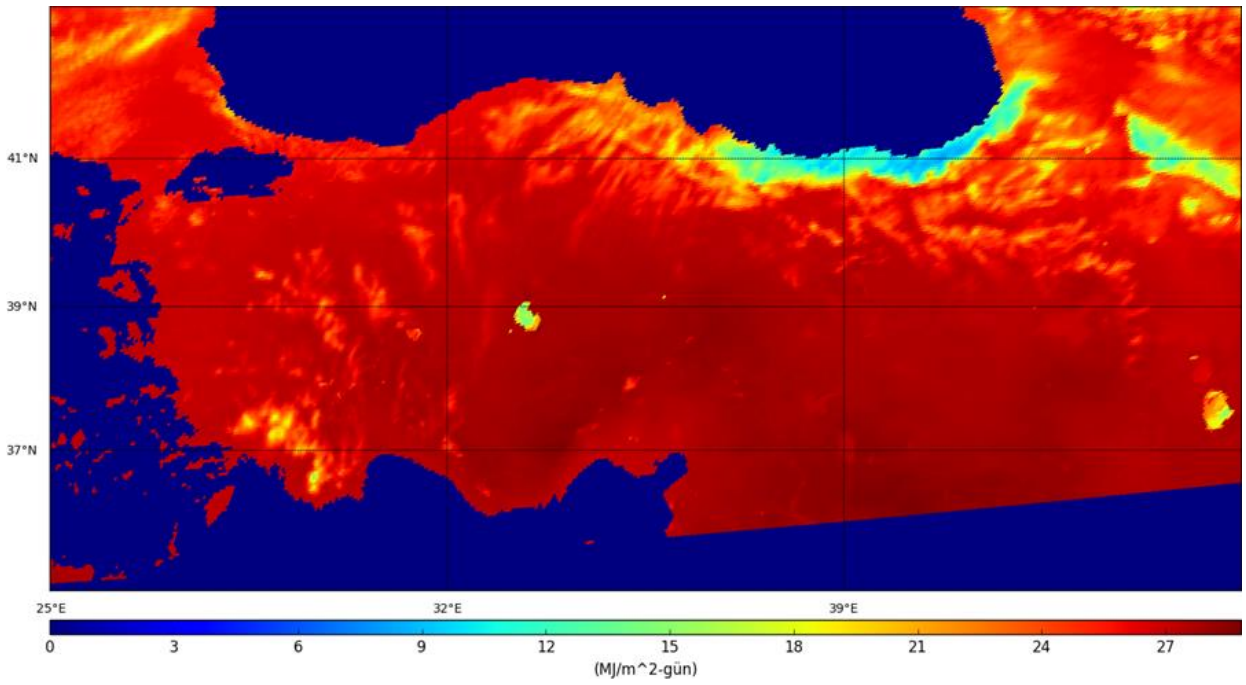
Türkiye deniz seviyesinden ortalama 1130 m yüksekliğe sahip olup yıllık ortalama sıcaklık değerleri 3.6 ^{o}C ile 20.1 ^{o}C arasında değişmektedir (Deniz vd. 2011). Yıllık ortalama yağış miktarı 648 mm civarında iken, ortalama güneşlenme süresi 7.2 saat/gün ve ortalama global güneş radyasyonu 4.2 kWh/m^2 değerindedir. Türkiye'nin Akdeniz ve Ege kıyıları ile Marmara bölgesinde Akdeniz iklimi, iç kısımlarında karasal iklim ve Karadeniz bölgesinde ise Karadeniz iklimi özellikleri görülür (Sahin 2012; Sahin ve Cigizoglu 2012; Iyigun vd. 2013; Kaba vd. 2018).



Şekil 1. Çalışmada kullanılan MGM istasyonlarının konumları

Tablo 1. Çalışmada kullanılan yer istasyonlarının bilgileri (Enlem ve Boylam derece biriminde, İN: İstasyon numarası)

İN	Enlem	Boylam	Rakım (m)	Şehir	İN	Enlem	Boylam	Rakım (m)	Şehir
17045	41,1752	41,8187	625	Artvin	**17285**	37,5745	43,7388	1727	Hakkari
17052	41,7382	27,2178	232	Kırklareli	**17287**	37,5209	42,4523	1350	Şırnak
17056	40,9585	27,4965	4	Tekirdağ	**17292**	37,2095	28,3668	646	Muğla
17065	40,9866	29,2135	123	İstanbul	**17351**	37,0041	35,3443	23	Adana
17070	40,7329	31,6022	743	Bolu	**17624**	41,143	37,293	16	Ordu
17074	41,371	33,7756	800	Kastamonu	**17688**	40,3013	41,5378	1576	Erzurum
17086	40,3312	36,5577	611	Tokat	**17718**	39,7728	40,3843	1429	Erzincan
17088	40,4598	39,4653	1216	Gümüşhane	**17722**	39,4983	26,9755	20	Balıkesir
17099	39,7253	43,0522	1646	Ağrı	**17734**	39,3618	38,1142	1121	Sivas/Divriği
17130	39,9727	32,8637	891	Ankara/Keçiören	**17749**	38,4639	27,3705	208	İzmir/Kemalpaşa
17162	39,185	36,0805	1182	Sivas/Gemerek	**17754**	39,0788	33,0657	1005	Konya/Kulu
17172	38,4693	43,346	1675	Van	**17760**	39,1897	35,2532	1070	Yozgat
17190	38,738	30,5604	1034	Afyonkarahisar	**17776**	38,9597	41,0503	1366	Bingöl
17192	38,3705	33,9987	970	Aksaray	**17789**	38,6237	27,0433	10	İzmir/Menemen
17199	38,3367	38,2173	950	Malatya	**17802**	38,7251	36,3904	1542	Kayseri/Pınarbaşı
17227	37,8167	27,8873	32	Aydın	**17806**	38,6907	39,926	1000	Elazığ
17239	38,3688	31,4297	1002	Konya/Akşehir	**17836**	38,3713	35,4745	1204	Kayseri/Develi
17240	37,7848	30,5679	997	Isparta	**17866**	38,024	36,4823	1344	Kahramanmaraş
17242	37,6777	31,7463	1141	Konya/Beyşehir	**17906**	37,548	34,4867	1453	Niğde/Ulukışla
17246	37,1932	33,2202	1018	Karaman	**17944**	37,3651	38,5134	622	Şanlıurfa/Bozova
17250	37,9587	34,6795	1195	Niğde/Merkez	**17952**	36,7372	29,9125	1095	Antalya
17262	36,7085	37,1123	640	Kilis	**17968**	36,8406	40,0307	360	Şanlıurfa/Ceylanpınar
17275	37,3063	40,7235	1040	Mardin	**17978**	36,8898	34,9558	12	Mersin



Şekil 2. Çalışma alanını kapsayan DIDSSF ürünü

LSA SAF birimi Meteosat uydu verilerinden bitki örtüsü, kar örtüsü, yüzey sıcaklığı, yangın ve radyasyon gibi yer yüzeyini karakterize eden uydu ürünlerinin geliştirilmesi ve işlenmesinden sorumlu olan bir EUMETSAT birimidir. LSA SAF ürünlerden biri olan DIDSSF ürünü 0.3-4 µm dalga boyu aralığında birim yüzeye ulaşan enerji değerlerini vermektedir. Yüzeye ulaşan enerji değeri daha çok güneş zenit açısına ve bulut kapalılığı ile nispeten azda olsa atmosferik soğurma ve yüzey albedosuna bağlıdır. Şekil 2'de çalışma alanını kapsayan 15-08-2013 tarihli örnek bir DIDSSF ürününden elde edilen harita verilmiştir. DIDSSF ürünü, MSG (Meteosat Second Generation) SEVIRI (Spinning Enhanced Visible and Infared Imager) sensörünün 0.6, 0.8 ve 1.6 µm solar kanal verileri kullanılarak havanın açık veya kapalı durumuna göre iki farklı algoritma ile üretilmektedir. Açık hava modelinde DSSF (Downward Surface Shortwave Flux) doğrudan hesaplanırken, kapalı hava durumunda atmosfer dışı albedo gözlemlenen yansıma değerlerinden hesaplanarak DSSF ürününün hesaplanmasında girdi olarak kullanılmaktadır. DSSF ürünü hava tahmini, hidroloji, iklim, tarım ve çevreyle ilgili çalışmalar gibi pek çok uygulama için kritik öneme sahiptir. 15 dakikalık zamansal çözünürlükte üretilen DSSF ürününden günlük DIDSSF ürünü türetilmektedir (EUMETSAT LSA SAF, 2020). MSG SEVIRI sensörü 12 farklı kanalda (başka bir ifadeyle spektral çözünürlüğü 12'dir), 3 km yersel çözünürlükte (HRV kanalı 1km), normal taramada 15 dakikalık zamansal çözünürlükte (hızlı taramada 5 dakika), 10 bitlik radyometrik çözünürlükte ve Avrupa ile Afrika'yı kapsayan görüntüler sağlamaktadır. LSA SAF biriminin ürettiği pek çok ürün SEVIRI sensörünün bu temel özelliklerine dayanmaktadır. Bu çalışmada kullanılan DIDSSF ürünü ise 3 km yersel çözünürlükte, 32 bit veri formatında ve günlük ortalama enerji değerlerini (J/m^2 biriminde) içermekte olup veriler Ocak-2018'de online olarak elde edilmiştir (EUMETSAT LSA SAF, 2011). Çalışma zaman dilimini kapsayan toplam 1034 gün için 1034 görüntü olması beklenirken 35 adet görüntü eksik olup toplam 999 görüntü elde edilmiştir.

2.2 Regresyon Analizi ve İstatistiksel İndeksler

İki ya da daha fazla değişken arasındaki ilişkiyi ölçmek ve tahminler yapmak için kullanılan istatistiksel analize regresyon analizi denir. Regresyon analizinde açıklanan ya da tahmin edilen değişken bağımlı değişken (Y), açıklayıcı değişken bağımsız değişken (X) olmak üzere regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi kullanarak bir eşitlik geliştirir. A ve B katsayılar olmak üzere, bağımlı değişken ile bağımsız değişken arasındaki doğrusal ilişkiyi açıklayan tek değişkenli regresyon modeli basit regresyon olarak adlandırılır ve matematiksel ifadesi Denklem 1'de verilmiştir. Bu çalışmada da uydu verileri ile yer verileri arasındaki ilişki değerlendirilmek amacıyla basit doğrusal regresyon analizi kullanılmıştır. X değişkeni yer verilerine, Y değişkeni ise uydu verilerine karşılık gelmektedir.

Uydudan elde edilen enerji değerlerinin yer ölçümleriyle ilişkisini değerlendirmek için MBE (Ortalama bias hata), MAE (Ortalama mutlak hata), RMSE (Ortalama karesel hatanın karekökü), rRMSE (Bağıl RMSE) ve R^2 (Belirleme katsayısı) istatistiksel indeksleri kullanılmıştır. Bu yöntemlerin matematiksel ifadeleri aşağıdaki denklemlerde verilmiştir. Denklemlerde Y_i yer istasyonlarında ölçülen ve U_i uydudan elde edilen enerji değerleridir. $\overline{Y}$ ve $\overline{U}$ yer ve uydu değerlerinin ortalamasını, n ise toplam veri sayısını ifade etmektedir.

$$Y = AX + B \tag{1}$$

$$MBE = \frac{\sum_{i=1}^{n}(Y_i - U_i)}{n} \tag{2}$$

$$MAE = \frac{\sum_{i=1}^{n}|Y_i - U_i|}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - U_i)^2}{n}} \tag{4}$$

$$rRMSE = \frac{100 * n}{\sum_{i=1}^{n}(Y_i)} * \sqrt{\frac{\sum_{i=1}^{n}(Y_i - U_i)^2}{n}} \tag{5}$$

$$R^2 = \frac{[\sum_{i=1}^{n}(Y_i - \bar{Y})\,(U_i - \bar{U})]^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(U_i - \bar{U})^2} \tag{6}$$

3. Bulgular ve Tartışma

Çalışmada Türkiye'yi kapsayacak şekilde HDF (h5) formatında alınan DIDSSF ürünü, coğrafik koordinat sisteminde, WGS84 elipsoidinde ve geotiff dosya formatında hazırlanmıştır. 32 bit veri formatında olan piksel değerleri veri sağlayıcının belirttiği üzere 10'a bölünerek enerji değerleri J/m^2 boyutunda elde edilmiştir. Enerji değerleri için MJ/m^2 birimi literatürde daha yaygın kullanılması nedeniyle J/m^2 birimi MJ/m^2birimine dönüştürülmüştür. Uydu verisi (DIDSSF) MJ/m^2 biriminde hazırlanması nedeniyle Meteoroloji Genel Müdürlüğünden alınan ve birimi "cal/cm^2" olan enerji değerleri MJ/m^2 ortak birimine dönüştürülmüştür. (1 cal/cm^2 = 0.041868 MJ/m^2). 46 adet yer istasyonunun verileri ile bu istasyon konumlarına karşılık gelen DIDSSF ürünündeki piksel değerlerinden Ocak-2013 ve Ekim-2015 arasını kapsayan her günün enerji değerlerini içeren bir veri seti oluşturulmuştur. Bu veri seti Ocak-2013 ve Ekim-2015 tarihleri arasındaki uydu ve yer verilerinin aynı günlere ve aynı alanlara denk gelen değerlerini içermektedir. Çünkü bu çalışma uydudan elde edilen enerji verisinin (DIDSSF) yer verileriyle ne kadar uyumlu olduğunu göstermeyi amaçlamaktadır. Bu nedenle böyle bir karşılaştırılma yapılabilmesi için her iki veride aynı zamanda ve aynı yer için alınmış olmalıdır. Çalışmada DIDSSF ürününün hazırlanması, veri setinin oluşturulması, analiz ve grafiklerin üretilmesi aşamasında Python programlama dili ve onun datetime, gdal, numpy, scit-learn, basemap ve matplotlib kütüphaneleri kullanılmıştır. Hazırlanan veri setinin temel istatistik değerleri Tablo 2'de verilmiştir. Tablo 2 incelendiğinde 46 istasyonun yer ve uydu verisi için minimum, maksimum, ortalama, standart sapma ve varyans değerleri görülür. Yer ve uydu ölçümü için maksimum değerler karşılaştırıldığında uydu verisinin yer istasyonlarından açık bir şekilde daha küçük (~2.3 MJ/m^2gün) ölçtüğü görülür. Standart sapma ve ortalama değerler yer ve uydu ölçümü için nispeten benzer iken varyans değerlerinin uydu için daha yüksek olduğu görülür. Basit lineer regresyon analizi kullanılarak her bir istasyon için yer ve uydu ölçümü arasındaki lineer denklemler tespit edilmiştir. Bu lineer denklemlerin katsayıları Tablo 3'te verilmiştir. Bu istasyonlar için bulunan katsayılar lineer denklemlerde kullanılarak bu iki verinin birbirine dönüştürülmesi yapılabilir veya bu iki veriden biri eksik olduğunda diğerinin değeri kullanılarak o eksik veri bulunabilir.

Tablo 2. Çalışmada kullanılan 46 adet yer istasyonu için temel istatistik değerler

	Yer verisi (MJ/m^2gün)	**Uydu verisi (MJ/m^2gün)**
Minimum	0	0
Maksimum	34,6056	32,3132
Ortalama	16,9835	17,4488
Standart sapma	8,53388	8,87203
Varyans	72,8271	78,7129

Tablo 3. Yer verileri ile uydu verileri arasındaki lineer denklemlerin katsayıları (İN: İstasyon numarası)

İN	A	B	İN	A	B
17045	0,8869	1,1416	**17285**	1,0426	1,1640
17052	1,0280	0,4434	**17287**	1,0378	0,4182
17056	1,0195	0,2153	**17292**	0,9657	1,9084
17065	1,0065	1,2387	**17351**	1,0764	0,6325
17070	0,9904	0,6081	**17624**	0,9623	1,0836
17074	0,9688	0,6017	**17688**	0,9864	-0,3315
17086	0,8202	3,8753	**17718**	1,0258	-0,4605
17088	0,9956	1,7467	**17722**	1,0152	0,1706
17099	0,9689	0,7179	**17734**	1,0158	0,2823
17130	0,9828	0,6330	**17749**	1,0013	0,6584
17162	0,9922	0,2421	**17754**	0,9929	0,1502
17172	1,0302	-0,3310	**17760**	1,0029	0,0970
17190	0,9923	0,3136	**17776**	0,9477	1,8267
17192	1,0103	-0,1068	**17789**	1,0175	0,7291
17199	1,0432	-0,0079	**17802**	0,8725	1,9856
17227	1,0577	-0,4672	**17806**	1,1344	-0,3846
17239	0,9787	-0,2826	**17836**	0,9954	0,0116
17240	1,0006	-0,2595	**17866**	0,9544	0,8239
17242	1,0248	-0,3556	**17906**	1,0259	-0,4580
17246	0,9764	-0,5889	**17944**	1,0266	-0,2808
17250	1,0259	-0,5144	**17952**	0,9942	0,3987
17262	1,0267	0,0757	**17968**	1,0381	0,4317
17275	0,9925	0,6421	**17978**	1,0840	-0,8485

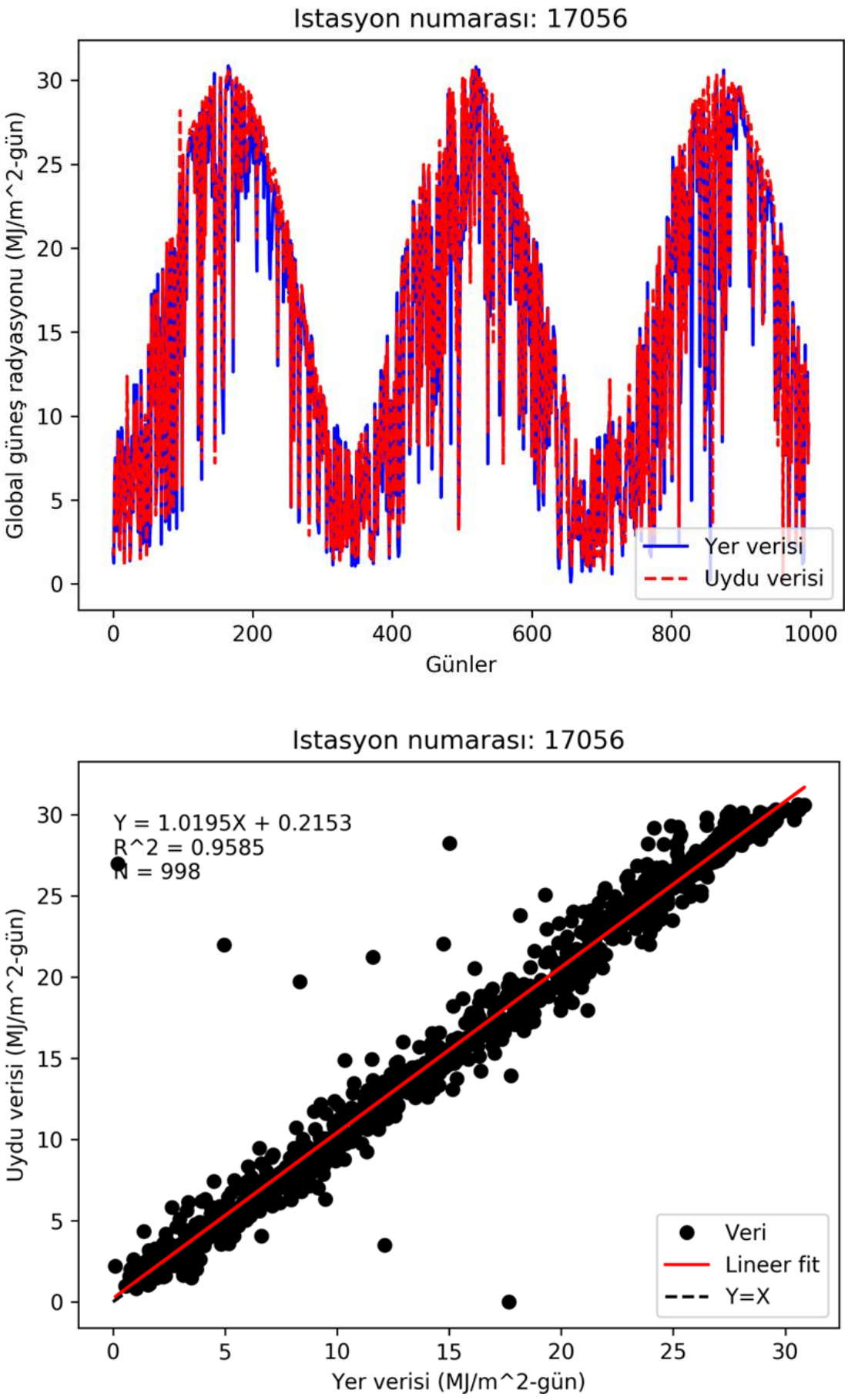
Tablo 4'te DIDSSF verisi ile yer verileri arasındaki ilişkiyi ifade eden istatistiksel indekslerin değerleri verilmiştir. Tablodan MBE değerleri incelendiğinde, bütün istasyonlar için minimum MBE -1,9168 MJ/m^{2}gün değeri ile Adana (17351) istasyonunda ve maksimum MBE 1,0509 MJ/m^{2}gün değeri ile Karaman (17246) istasyonunda olduğu görülür. Bütün istasyonların ortalama MBE değeri ise -0,4661 MJ/m^{2}gün olarak bulunmuştur. Ortalama MBE değerinin negatif çıkması genelde, uydu verisinin yer ölçümlerinden daha büyük ölçüldüğü anlamına gelir. Tablo 4'ten MAE değerleri incelendiğinde, 1,0018 ile 2,6787 MJ/m^{2}gün arasında değiştiği görülür. En düşük (en iyi) MAE değeri Tekirdağ (17056) istasyonunda ve en yüksek (en kötü) MAE değeri Kayseri/Pınarbaşı (17802) istasyonunda bulunmuştur. Bütün istasyonların ortalama MAE değeri ise 1,5541 MJ/m^{2}gün olarak bulunmuştur. Yine Tablo 4'te RMSE değerlerine bakıldığında, minimum ve maksimum değerlerin 1,7157 ve 4,7644 MJ/m^{2}gün olduğu görülür ve diğer istasyonların RMSE değerleri bu minimum ve maksimum değerler aralığında değişmektedir. Minimum (en iyi) RMSE değeri Kastamonu (17074) istasyonunda ve maksimum (en kötü) RMSE değeri Tokat (17086) istasyonunda gerçekleşmiştir. Bütün istasyonların ortalama RMSE değeri ise 2,5484 MJ/m^{2}gün olarak bulunmuştur. Tablodan bağıl RMSE (rRMSE) değerleri incelendiğinde %10,7126, ile %33,9319 aralığında değiştiği görülür ve bütün istasyonların ortalama rRMSE değerleri %15,0791 olarak bulunmuştur. Tablo 4'te gösterilen R^2 değerleri, 0,7129 ila 0,9585 arasında değişim göstermektedir. 46 adet istasyonun ortalama R^2 değeri ise 0,9025 olarak bulunmuştur. Tablodan 46 adet istasyon içerisinden 31 tanesinin R^2 değeri ortalama değerin (0,9025) üzerinde ve 15 tanesinin de ortalama değerin altında olduğu görülür. R^2 değeri 0,8'den küçük olan dört istasyon bulunmaktadır. Bunlar sırasıyla Tokat (17086), Hakkari (17285), Adana (17351) ve Kayseri/Pınarbaşı (17802) istasyonlarıdır. R^2 değeri en büyük 10 istasyon ise sırasıyla Kırklareli (17052), Tekirdağ (17056), Bolu (17070), Kastamonu (17074), Ankara/Keçiören (17130), Afyonkarahisar (17190), Aksaray (17192), Isparta (17240), Konya/Beyşehir (17242) ve Karaman (17246) olarak sıralanmıştır. Bu istasyonların diğer istatistiksel değerlerinin de beklendiği gibi düşük olduğu Tablo 4'ten görülebilir. Yukarıda sıralanan istasyonların ortak özellikleri, çoğunlukla Türkiye'nin iç kesimlerinde bulunması ve deniz etkisinin doğrudan görülmediği yerler (Kırklareli ve Tekirdağ hariç) olarak göze çarpmaktadır.

Her ölçümde olduğu gibi yer veya uydu bazlı olsun enerji ölçümünde de hata bulunmaktadır. Buna bu çalışmada kullanılan DIDSSF ürünü ve meteorolojik yer istasyonunda ölçülen enerji değerleri de dahildir. MGM ölçümlerinde istasyonun bulunduğu konum (istasyon bina veya ağaç gölgesinde kalabilir), cihazın hassasiyeti, bakımı (kalibrasyonu), ve insan faktörü gibi birçok nedenden dolayı hata meydana gelmektedir. Öte yandan uydu kayıtlarında ise uydu görüş açısı arttıkça görüntülerin kenar kısımlarına doğru hassasiyet düşebilir. Bu hata kaynakları dikkate alındığında her istasyon için aynı doğrulukta (hassasiyette) ölçüm yapılması beklenemez. Bu nedenle validasyon sonuçları her bir istasyon için farklı çıkmaktadır.

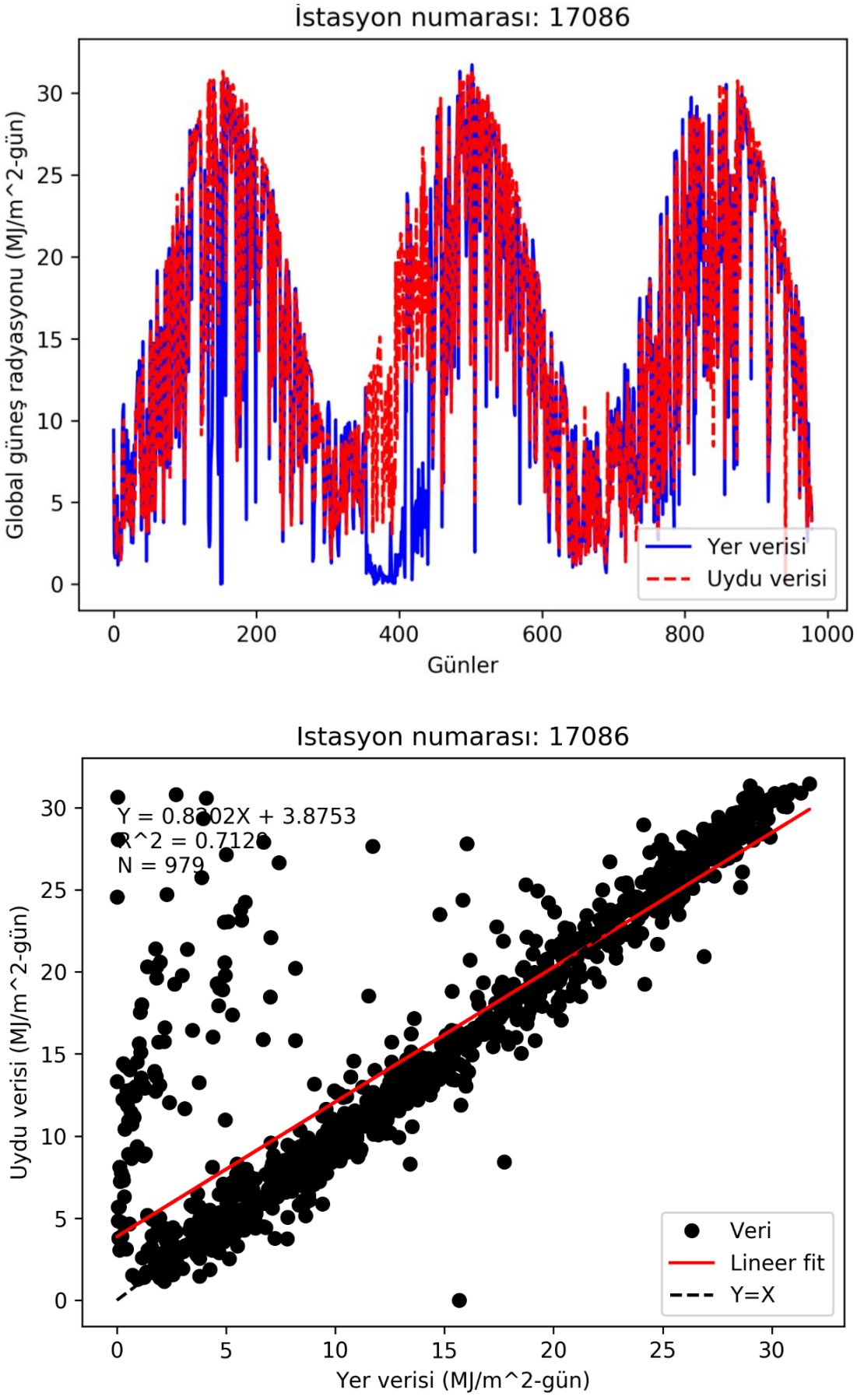
Tablo 4. Yer verileri ile uydu verileri arasındaki hataları veren istatistiksel indeks değerleri (MBE, MAE ve RMSE MJ/m^{2}gün biriminde, İN: İstasyon numarası)

İN	MBE	MAE	RMSE	rRMSE	R^2	İN	MBE	MAE	RMSE	rRMSE	R^2
17045	0,4144	1,8703	2,7773	20,1832	0,8600	**17285**	-1,8886	2,5690	3,6888	21,6724	0,7912
17052	-0,8564	1,1921	1,9510	13,2349	0,9479	**17287**	-1,0922	1,6581	2,4664	13,8373	0,9181
17056	-0,5136	1,0018	1,8186	11,9131	0,9585	**17292**	-1,3396	1,6215	2,4649	14,8466	0,9235
17065	-1,3334	1,6941	3,1792	21,7745	0,8681	**17351**	-1,9168	2,2906	3,5376	21,0394	0,7621
17070	-0,4673	1,2144	1,9971	13,6479	0,9422	**17624**	-0,5818	1,3822	2,3394	17,5915	0,9250
17074	-0,1432	1,0165	1,7157	11,6609	0,9576	**17688**	0,5595	1,9490	2,7539	16,3920	0,8727
17086	-1,3508	2,2505	4,7644	33,9319	0,7129	**17718**	0,0141	1,5695	2,4029	13,8754	0,9196
17088	-1,6848	2,1329	2,9071	20,5328	0,8609	**17722**	-0,4329	1,1650	2,4694	14,2863	0,9190
17099	-0,1980	1,6664	2,7661	16,5687	0,8989	**17734**	-0,5434	1,5574	2,6172	15,7967	0,9027
17130	-0,3515	1,0477	1,9851	12,1067	0,9474	**17749**	-0,6805	1,3164	2,1677	12,7695	0,9376
17162	-0,1090	1,5429	2,5964	15,2192	0,9037	**17754**	-0,0258	1,3123	2,2702	12,9739	0,9281
17172	-0,2216	1,6166	2,6846	14,6690	0,8920	**17760**	-0,1468	1,4159	2,6085	15,0844	0,9042
17190	-0,1840	1,2127	1,9612	11,6399	0,9462	**17776**	-0,9226	1,8941	3,1257	18,0870	0,8752
17192	-0,0815	1,2073	2,0174	11,0784	0,9464	**17789**	-1,0292	1,4262	2,2832	13,3491	0,9283
17199	-0,7382	1,5244	2,5251	14,6027	0,9101	**17802**	0,2531	2,6787	4,3691	24,8818	0,7404
17227	-0,5705	1,2200	2,0815	11,5761	0,9349	**17806**	-1,8109	2,2217	2,9532	18,0791	0,8661
17239	0,6515	1,6014	2,3876	13,7714	0,9214	**17836**	0,0730	1,2199	2,3330	12,5755	0,9273
17240	0,2490	1,1947	1,9409	10,7729	0,9463	**17866**	-0,0231	1,6199	2,7951	15,9255	0,8956
17242	-0,0896	1,2670	1,9250	10,7126	0,9457	**17906**	-0,0068	1,4087	2,2241	12,3806	0,9252
17246	1,0509	1,6144	2,2632	11,5615	0,9380	**17944**	-0,2222	1,2495	2,2032	11,6694	0,9330
17250	0,0298	1,3870	2,1720	11,5866	0,9271	**17952**	-0,2902	1,1642	2,0920	11,2089	0,9354
17262	-0,5698	1,4981	2,6516	14,3375	0,9013	**17968**	-1,1372	1,6145	2,4632	13,3189	0,9092
17275	-0,5024	1,7293	3,3025	17,7668	0,8558	**17978**	-0,6801	1,4803	2,2280	12,2506	0,9102

Şekil 3'te Tekirdağ (17056) ve Şekil 4'te Tokat (17086) istasyonları için yer ölçümü ile DIDSSF verisi arasındaki ilişkiyi veren çizgi (zaman serisi) ve dağılım grafikleri verilmiştir. Zaman serisi grafiklerinde yatay eksen zamanı, dikey eksen ise enerji değerlerini temsil etmektedir. Zaman serisi grafiklerde mavi renk yer verilerine, kırmızı renk ise uydu verilerine karşılık gelmektedir. Bu grafikler R^2 değerleri baz alındığında yer ölçümü ile uydu verisi arasındaki en yüksek (en iyi) ve en düşük (en kötü) uyumu gösteren grafiklerdir. Grafiklerde yüksek değere sahip tepe noktaları 2013, 2014 ve 2015 yıllarının yaz günlerini (Haziran, Temmuz ve Ağustos), düşük değerler ise bu yılların kış günlerini (Aralık, Ocak ve Şubat) ifade etmektedir. Grafiklerde değerlerin üst üste geldiği günlerde, yer ve uydu verileri arasında yüksek uyumun olduğu değerlerin üst üste düşmediği zamanlarda ise uyumun düşük olduğu anlaşılır. Zaman serisi grafiklerinde olduğu gibi dağılım grafiklerinden de bu iki istasyon için uyum ve uyumsuzluk durumları görülebilir.

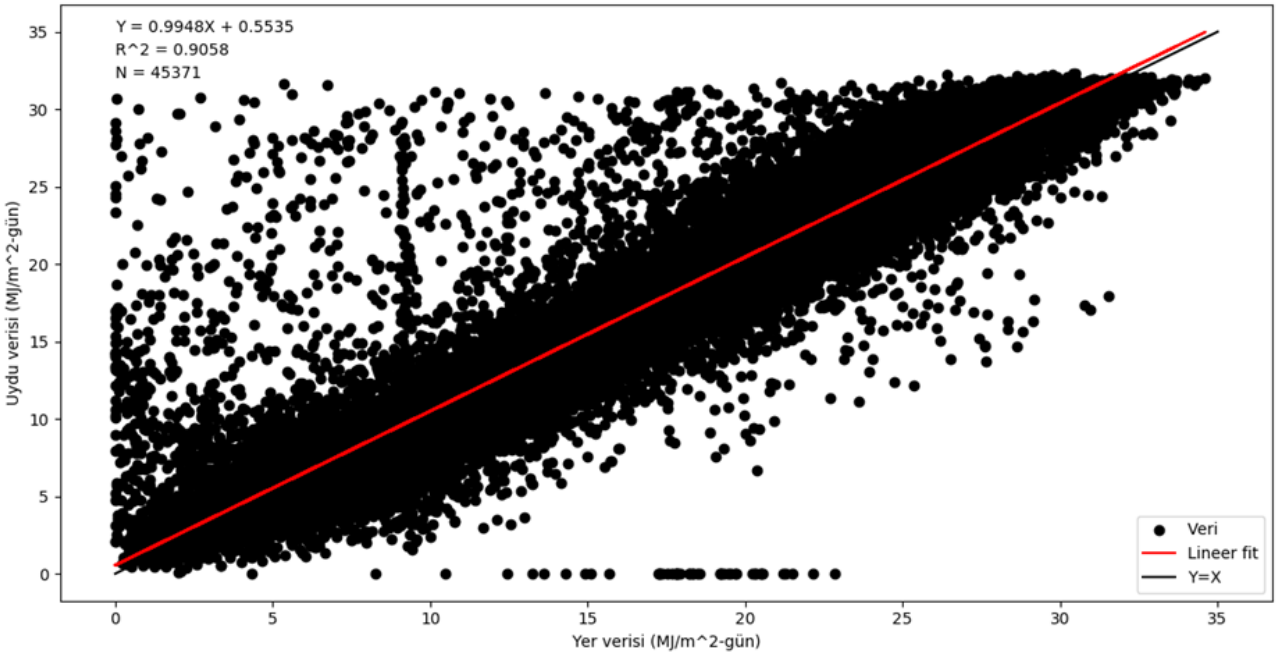


Şekil 3. En yüksek R^2 değerine sahip Tekirdağ istasyonuna ait verilerin çizgi ve dağılım grafikleri



Şekil 4. En düşük R^2 değerine sahip Tokat istasyonuna ait verilerin çizgi ve dağılım grafikleri

Çalışma zaman dilimine ait 999 adet günlük enerji değerlerini veren uydu görüntüsü ve 46 validasyon noktası için 45954 (999*46) veri örneği olması beklenirken bazı günler için kayıp yer verisi olması nedeniyle toplam 45371 adet verinin dağılım grafiği Şekil 5'te verilmiştir. Grafik incelendiğinde R^2 değerinin yüksek (0,9058) olduğu ve DIDSSF ürününün yer verileriyle yüksek uyum gösterdiği görülür. Şekil 5'te uydu verilerinin yüksek olduğu (25 MJ/m^2gün) değerlere karşılık, yer verilerinin küçük (0-15 MJ/m^2gün) oluşunun, iki veri arasındaki ilişkiyi düşürdüğü görülebilir. Aynı zamanda uydu değerlerinin sıfır ölçümüne karşılık yer ölçümünün 13-23 MJ/m^2gün aralığında değerler alması da iki veri arasındaki tutarlılığı azaltmaktadır.



Şekil 5. Tüm istasyonların (46 adet) verileri ile elde edilen dağılım grafiği

4. Sonuçlar

Çalışmada DIDSSF ürününün doğruluğu, ülke coğrafyasına olabildiğince homojen olarak dağılmış MGM (Meteoroloji Genel Müdürlüğü) tarafından işletilen 46 adet yer istasyonlarında ölçülen günlük global güneş radyasyonu değerleri ile test edilmiştir. Her ölçümde olduğu gibi yer verileri de sistematik ve/veya beklenmedik hatalar içerebilmesine karşın uydu verileriyle yüksek uyum gösterdiği tespit edilmiştir (~%90, bkz. Şekil 5). Cristóbal ve Anderson (2013) ile Moreno vd. (2013) kendi bölgeleri için yaptıkları çalışmalarda DIDSSF ürününün yüksek doğruluğa sahip olduğunu göstermişlerdir. Sonuçlar değerlendirildiğinde DIDSSF ürünü yüksek doğruluğu, alansal ve zamansal sürekliliği bakımından yer ölçümleri yerine kullanımı tercih edilebilir. Güneş enerjisi ile güneşlenme süresi arasında yüksek korelasyon olması sebebiyle, DIDSSF ürünü güneşlenme süresine ihtiyaç duyulan çalışmalarda da kullanılabilir. DIDSSF ürünü geçmişe dönük geniş arşiv verisi sağlaması nedeniyle uzun yıllara dayalı iklim değişikliği gibi, uzun dönemli verilerin kullanıldığı çalışmalara da olumlu katkı sağlayacaktır. Bu ürün, kendi içinde homojen değerlere sahip olması nedeniyle bölge kıyaslamalarının yapılacağı, örneğin güneş panelleri için yer seçimi gibi çalışmalarda da kullanılması, yer verilerine göre daha doğru sonuçlar verecektir.

Kaynaklar

Angstrom, A. (1924). Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Quarterly Journal of the Royal Meteorological Society*, *50*(210), 121-126.

Behrang, M. A., Assareh, E., Ghanbarzadeh, A., & Noghrehabadi, A. R. (2010). The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy*, *84*(8), 1468-1480.

Cogliani, E., Ricchiazzi, P., & Maccari, A. (2007). Physical model SOLARMET for determinating total and direct solar radiation by meteosat satellite images. *Solar Energy*, *81*(6), 791-798.

Cristóbal, J., & Anderson, M. C. (2013). Validation of a Meteosat Second Generation solar radiation dataset over the northeastern Iberian Peninsula. *Hydrol. Earth Syst. Sci., 17*, 163–175, 2013.

Deniz, A., Toros, H., & Incecik, S. (2011). Spatial variations of climate indices in Turkey. *International Journal of climatology*, *31*(3), 394-403.

EUMETSAT LSA SAF, (2011). Product User Manual – Down-welling Surface Shortwave Flux (DSSF). Retrieved from https://nextcloud.lsasvcs.ipma.pt/s/bXEkdAKRJJGSn3S#pdfviewer.

EUMETSAT LSA SAF, (2020). Daily Downward Surface Shortwave Flux (DIDSSF, LSA-203). Retrieved from https://landsaf.ipma.pt/en/products/longwave-shortwave-radiation/didssf/.

Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., & Gao, L. (2020). Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renewable and Sustainable Energy Reviews*, *118*, 109393, doi: 10.1016/j.rser.2019.109393

Gautier, C., Diak, G., & Masse, S. (1980). A simple physical model to estimate incident solar radiation at the surface from GOES satellite data. *Journal of Applied Meteorology*, *19*(8), 1005-1012.

Iyigun, C., Türkeş, M., Batmaz, İ., Yozgatligil, C., Purutçuoğlu, V., Koç, E. K., & Öztürk, M. Z. (2013). Clustering current climate regions of Turkey by using a multivariate statistical method. *Theoretical and applied climatology*, *114*(1-2), 95-106.

Kaba, K., Sarıgül, M., Avcı, M., & Kandırmaz, H. M. (2018). Estimation of daily global solar radiation using deep learning model. *Energy*, *162*, 126-135.

Moreno, A., Gilabert, M. A., Camacho, F., & Martínez, B. (2013). Validation of daily global solar irradiation images from MSG over Spain. *Renewable Energy*, *60*, 332-342.

Quesada-Ruiz, S., Linares-Rodríguez, A., Ruiz-Arias, J. A., Pozo-Vázquez, D., & Tovar-Pescador, J. (2015). An advanced ANN-based method to estimate hourly solar radiation from multi-spectral MSG imagery. *Solar Energy*, *115*, 494-504.

Rusen, S. E., Hammer, A., & Akinoglu, B. G. (2013). Estimation of daily global solar irradiation by coupling ground measurements of bright sunshine hours to satellite imagery. *Energy*, *58*, 417-425.

Sahin, S. (2012). An aridity index defined by precipitation and specific humidity. *Journal of Hydrology*, *444*, 199-208.

Sahin, S., & Cigizoglu, H. K. (2012). The sub-climate regions and the sub-precipitation regime regions in Turkey. *Journal of Hydrology*, *450*, 180-189.

Shamshirband, S., Mohammadi, K., Tong, C. W., Zamani, M., Motamedi, S., & Ch, S. (2016). A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. *Theoretical and Applied Climatology*, *125*(1-2), 53-65.

Trigo, I. F., Dacamara, C. C., Viterbo, P., Roujean, J. L., Olesen, F., Barroso, C., ... & Geiger, B. (2011). The satellite application facility for land surface analysis. *International Journal of Remote Sensing*, *32*(10), 2725-2744.

Wu, L., Huang, G., Fan, J., Zhang, F., Wang, X., & Zeng, W. (2019). Potential of kernel-based nonlinear extension of Arps decline model and gradient boosting with categorical features support for predicting daily global solar radiation in humid regions. *Energy Conversion and Management*, *183*, 280-295.

2020
Turkish Journal of Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, Eylül 2020, 1(2): 97-105
Dergi Anasayfası: http://www.dergipark.org.tr/en/pub/rsgis

Hafif Gradyan Artırma Makineleri ile Tarımsal Ürünlerin Sınıflandırılması

Crop Classification Using Light Gradient Boosting Machines

Mustafa Üstüner[1]*, Saygın Abdikan[2], Gökhan Bilgin[3], Füsun Balık Şanlı[4]

[1]*Artvin Çoruh Üniversitesi, Harita Mühendisliği Bölümü, 08100, Artvin/Türkiye.*
[2]*Hacettepe Üniversitesi, Geomatik Mühendisliği Bölümü, 06800, Ankara/Türkiye.*
[3]*Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul/Türkiye.*
[4]*Yıldız Teknik Üniversitesi, Harita Mühendisliği Bölümü, 34220, İstanbul/Türkiye.*

ARAŞTIRMA MAKALESİ

***Sorumlu yazar:**
Mustafa Üstüner
mustuner@artvin.edu.tr

doi:

Yayın süreci
Geliş tarihi: 20.05.2020
Kabul tarihi: 13.08.2020
Basım tarihi: 30.09.2020

Özet

Son yıllarda, makine öğrenmesi ve veri bilimi alanındaki araştırmacılar özgün ve de özellikle topluluk öğrenmesi alanında yeni algoritmalar geliştirmeye başlamışlardır. Bu yeni nesil topluluk öğrenme algoritmalarından olan aşırı gradyan artırma (XGBoost) ve hafif gradyan artırma makineleri (LightGBM) yöntemleri, mevcut ve aynı zamanda sık kullanılan makine öğrenme algoritmalarına kıyasla daha yüksek performans gösterdiklerinden dolayı veri bilimindeki araştırmacıların ilgisini çekmiştir. Ancak uzaktan algılama görüntülerinin sınıflandırılması amacıyla henüz yeterli düzeyde test edilmemiştir. Bu çalışma kapsamında, XGBoost ve LightGBM algoritmalarının çok zamanlı polarimetrik sentetik açıklıklı radar (PolSAR) görüntüleri kullanılarak tarımsal ürünlerin sınıflandırılmasındaki performansları, hem işlem hızı hem de elde edilen sınıflandırma doğrulukları açısından karşılaştırılmıştır. Çalışma kapsamında, tam polarimetrik RADARSAT-2 uydu görüntülerine ait doğrusal geri saçılım değerleri kullanılmıştır. Tarımsal ürünlerin zamana bağlı dinamik olarak değişen yapılarından dolayı çalışmamızda çok zamanlı bir veri seti kullanılmıştır. Deneysel sonuçlarımız, LightGBM yönteminin hem işlem hızı hem de sınıflandırma performansı açısından XGBoost yöntemine göre daha üstün olduğunu göstermiştir, bu iki yöntem için elde edilen doğruluklar sırası ile 0.860 ve 0.845'dir. Sınıflandırma sonuçlarının elde edilmesinde k-katlamalı (k=5) çapraz doğrulama tekniği kullanılmıştır.

***Anahtar kelimeler**: Uzaktan algılama, Sentetik açıklıklı radar, Hafif gradyan artırma makineleri*

Abstract

In recent years, machine learning and data science communities have started to develop novel algorithms, especially in the area of ensemble learning. The new generation ensemble learning algorithms such as extreme gradient boosting (XGBoost) and light gradient boosting machines (LightGBM) have gained great attention in data science because of their greater performance compared to the state-of-art machine learning algorithms. However, they have not yet been fully tested for the classification of remotely-sensed images. This paper compares the performance of the XGBoost and LightGBM in terms of classification accuracy and computation time for crop classification using multi-temporal polarimetric SAR (PolSAR) data. The linear backscatter of full-polarimetric RADARSAT-2 were used as the polarimetric feature in this research. A multi-temporal dataset was used in our study because of the time-dynamic structure of crops. Our experimental results demonstrate that LightGBM yielded greater performance compared to XGBoost in terms of classification accuracy (0.860 vs 0.845) and computation cost. K-fold (k=5) cross validation was used to assess the classification results.

***Keywords**: Remote sensing, Synthetic aperture radar, Light gradient boosting machine*

1. Giriş

Tarımsal ürünlerin sınıflandırılması, uzaktan algılama tabanlı tarım uygulamalarının başında gelmektedir ve çoğu uygulama için de önemli olan altlık veriler sağlamaktadır (Atzberger, 2013). Ürün rekoltesinin tahmin edilmesi ve tarımsal ürünlerin izlenmesinde, uzaktan algılama verileri sıklıkla kullanılmaktadır ve bu aşamada ürün türlerinin doğru tespit edilmesi büyük önem arz etmektedir (McNairn vd. 2002). Uzaktan algılama verileri, yer gözlem uydularının sinoptik görüş özelliği ve aynı bölgeyi farklı zaman aralıklarında tekrar görüntüleyebilme yeteneğinden dolayı tarım ile ilgili çalışmalarda, klasik yöntemlere nazaran, tercih edilmektedir (Villa vd. 2015; Khosravi ve Alavipanah, 2019). Tarımsal ürünlerin sınıflandırılması, arazi örtüsünün sınıflandırılması işlemine nazaran daha zorlu bir süreçtir. Bunun nedeni tarımsal ürünlerin yılın bazı periyodlarında birbirine yakın yansıma (reflektans) değerlerine sahip olmasıdır (Waldhoff vd. 2012; Hütt vd. 2016). Çok zamanlı (multi-temporal) veriler kullanıldığında ise, her bir tarım ürünü farklı fenolojik özellik göstereceğinden tarımsal ürünlerin birbirlerinden ayırt edilmesi daha kolay olacaktır. Bu nedenle, tarımsal ürünlerin yüksek doğrulukla tespit edilmesinde çok zamanlı uydu görüntülerine ihtiyaç duyulmaktadır (Gómez-Chova vd. 2015; Hütt vd. 2016).

Optik uzaktan algılama görüntüleri (RapidEye, Sentinel-2, Landsat-8, Worldview-2 vb.) yüksek/orta çözünürlükte spektral ve mekânsal bilgi içerdiklerinden dolayı tarımsal ürünlerin sınıflandırılmasında başarılı sonuçlar vermektedir. Bu başarının altında yatan ana nedenler, tarımsal ürünlerin farklı spektral yansıma değerlerine sahip olmaları ve zaman bazlı analizlerde ürünlerin yansıma değerindeki değişimlerin (fenolojik değişimler) birbirinden farklı olmasıdır (Conrad vd. 2014; Inglada vd. 2015). Optik uzaktan algılama sistemleri pasif sistemlerdir ve görüntü alımında bazı kısıtlamalara sahiptir. Atmosferik koşullardan kaynaklanan hatalar (bulut ve sis) ve güneş enerjisine duyulan ihtiyaç nedeniyle bazı durumlarda istenilen kalitede görüntü elde edilememektedir. Özellikle katma değeri yüksek tarımsal ürünlerin belli zaman aralıklarında izlenmesinde, optik uydu görüntüleri yukarıda bahsedilen kısıtlamalardan kaynaklı yetersiz kalmaktadır. Bu tarz görüntü alımındaki kısıtlamaların önüne SAR (Sentetik Açıklıklı Radar) görüntüleri kullanılarak geçmek mümkündür. SAR görüntüleri, optik görüntülere nazaran bulut, sis vb. durumlardan etkilenmemektedir ve zamansal izlemenin kritik öneme sahip olduğu tarım projelerinde sıklıkla tercih edilmektedir. Bu avantajının yanı sıra, karmaşık SAR sinyali (geri saçılan sinyalin faz ve genlik bilgisinin birlikte kaydedilmesi nedeni ile) hedef objenin fiziksel ve geometrik yapısına karşı duyarlıdır ve ürünlerin birbirinden ayrılmasında önemli rol oynamaktadır (Jiao vd. 2014; McNairn ve Shang, 2016; Skakun vd. 2016). Bu nedenle, özellikle son yıllarda SAR görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı giderek artmaktadır. Diğer bir önemli etken ise, Avrupa Uzay Ajansı (ESA) kopernik programı kapsamında fırlatılan Sentinel-1 (A/B) SAR uydu görüntülerinin ücretsiz olarak elde edilebiliyor olmasıdır. Optik görüntüler ile kıyaslandığında, SAR ya da tam polarimetrik SAR (PolSAR) görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı daha az sayıdadır (Khosravi ve Alavipanah, 2019). Yapılan çalışmalar incelendiğinde, SAR/PolSAR görüntülerinin sınıflandırılmasında en sık kullanılan algoritmalar maksimum olabilirlik (maximum likelihood), yapay sinir ağları, karar ağaçları, rastgele orman (RO) ve destek vektör makineleri (DVM) yöntemleridir (Skakun vd. 2016). DVM ve RO algoritmaları parametrik olmayan yapıları ve üstün sınıflandırma kabiliyetleri nedeniyle diğer yöntemlere kıyasla uzaktan algılamada daha sık kullanılmıştır (Pal, 2012; Petropoulos vd. 2012). İki sınıfı birbirinden ayıran en uygun hiperdüzlemin belirlenmesi esasına dayanan DVM, sınıfların doğrusal düzlemlerle ayrılamaması durumunda çekirdek fonksiyonlarından faydalanmaktadır. Bu noktada, çekirdek türünün seçimi ve çekirdeğe ait en uygun parametrelerin belirlenmesi önem arz etmektedir (Melgani ve Bruzzone, 2004; Kavzoglu ve Colkesen, 2009). Karar ağaçları tabanlı bir yöntem olan RO algoritması ise, destek vektör makinelerine kıyasla daha az sayıda parametreye ihtiyaç duyması ve daha hızlı olması nedeniyle tercih edilmektedir (Belgiu ve Drăguţ, 2016).

Rastgele orman algoritması uzaktan algılama görüntülerinin sınıflandırılmasında kullanılan en yaygın topluluk öğrenme algoritmasıdır. Diğer yaygın kullanılan topluluk öğrenme algoritmaları ise torbalama (bagging) ve hızlandırma (boosting) algoritmalarıdır (Akar ve Güngör, 2012; Belgiu ve Drăguţ, 2016). Örüntü tanıma ve makine öğrenmesi alanlarındaki son gelişmeler ile birlikte, yeni nesil topluluk öğrenme algoritmaları ortaya çıkmış ve başarılı bir şekilde kullanılmaya başlanmıştır. Uzaktan algılama alanında da özellikle son yıllarda kullanılmaya başlanan yeni nesil topluluk öğrenme algoritmalarına aşırı gradyan artırma (extreme gradient boosting, XGBoost) ve hafif gradyan artırma makineleri (light gradient boosting machines, LightGBM) yöntemleri örnek olarak verilebilir (Georganos vd. 2018; Ustuner ve Balik Sanli, 2019; Abdi, 2020). Bu algoritmaları uzaktan algılama görüntülerinin sınıflandırılması amacıyla kullanan çalışmalar sınırlı sayıdadır. Georganos vd. (2018) yüksek çözünürlüklü uydu görüntülerinden faydalanarak arazi örtüsü ve kullanımının sınıflandırılmasında XGBoost, DVM, RO, k-en yakın komşuluk ve özyinelemeli bölümleme (recursive partitioning) yöntemlerini kullanmıştır. Shi vd. (2019) çalışmalarında, LiDAR nokta bulutu verisinin sınıflandırılması işleminde LightGBM yönteminden faydalanmışlardır. Jiang vd. (2019) şeker kamışının tespit edilmesi işleminde XGBoost ve RO algoritmalarının performansını karşılaştırmıştır. Moorthy vd. (2020) ise, LiDAR nokta bulutu verisinden yaprak ve ağaçlık alanların sınıflandırılması işleminde RO, XGBoost ve LightGBM yöntemlerinin performanslarını karşılaştırmıştır.

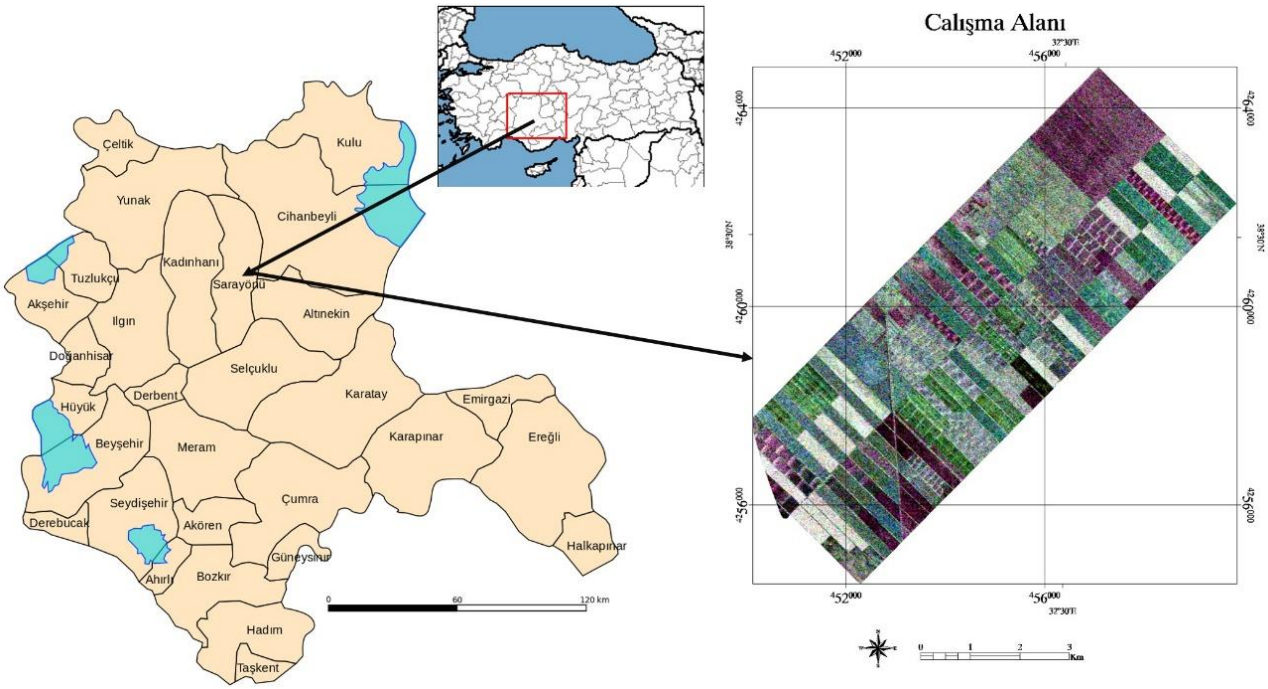
Tarım ürünlerinin uzaktan algılama görüntülerinden sınıflandırılmasında ise, en sık kullanılan algoritmalar maksimum olabilirlik, RO, DVM, derin öğrenme ve XGBoost yöntemleridir. Zhong vd. (2019) çalışmalarında, tarım ürünlerinin sınıflandırılması amacıyla derin öğrenme, XGBoost, DVM ve RO yöntemlerinden faydalanmışlardır. Saini ve Ghosh (2019) çalışmalarında Sentinel-2 görüntülerinden faydalanmış ve tarımsal ürünlerin sınıflandırılması işleminde XGBoost, RO ve DVM algoritmalarının performansını karşılaştırmıştır. Dey vd. (2020) ise çalışmalarında, tam polarimetrik Radarsat-2 PolSAR uydu görüntülerinden faydalanılmış ve çalışmalarında XGBoost ve RO algoritmalarının tarımsal ürünlerin sınıflandırılmasındaki performansını karşılaştırmıştır.

Yukarıda listelenen çalışmalardan sadece Moorthy vd. (2020) LightGBM ve XGBoost yöntemlerini karşılaştırmıştır fakat o çalışma ise tarımsal ürünlerin sınıflandırılması ile ilişkili değildir. Bu iki yöntemin (LightGBM ve XGBoost yöntemleri) tarımsal ürünlerin sınıflandırılmasındaki karşılaştırmalı analizini yapan bir çalışmaya henüz literatürde rastlanmamıştır.

Bu çalışmada, çok zamanlı PolSAR uydu görüntülerinin tarımsal ürünlerin sınıflandırılmasındaki kullanımı XGBoost ve LightGBM yöntemleri ile araştırılmıştır. Kullanılan yöntemlerin hem işlem hızı hem de sınıflandırma performansı açısından karşılaştırmalı analizi yapılmıştır. Sınıflandırma doğruluklarının (başarımlarının) belirlenmesinde ise k-katlamalı çapraz doğrulama yöntemi kullanılmıştır.

2. Çalışma Alanı ve Kullanılan Veriler

Konya ilinin Sarayönü ilçesine bağlı Gözlü köyü sınırları içerisinde yer alan çalışma alanı, Konya şehir merkezinin 65 km kuzeyinde yer almakta ve yaklaşık olarak 42 km^{2}'lik bir alanı kapsamaktadır (Şekil 1). Tarımsal üretim ve hayvancılık faaliyetlerinin sürdürüldüğü ilçede buğday, arpa, mısır, patates ve ayçiçeği üretimleri yapılmaktadır. Geniş düzlüklere sahip olan Konya ovası, ülke ölçeğinde tarımsal üretimin yapılmasında önemli bir yere sahiptir. Tamamen tarım ürünleri ile kaplı olan çalışma alanında, yem bitkisi, mısır, patates, ayçiçeği ve buğday olmak üzere beş farklı ürün türü bulunmaktadır. Tarımsal ürünlerin zamana bağlı dinamik olarak değişen yapılarından dolayı çalışmamızda çok zamanlı bir veri seti kullanılmıştır.



Şekil 1. Çalışma Alanı

Çalışma kapsamında, tek bakışlı karmaşık (single look complex) görüntü formatındaki tam polarimetrik Radarsat-2 PolSAR uydu görüntülerinden faydalanılmıştır. Kullanılan PolSAR görüntüsünün özellikleri ve görüntülerin alım tarihleri Tablo 1'de görülmektedir.

Tablo 1. Radarsat-2 görüntüsünün özellikleri

Algılayıcı	Radarsat-2
Dalga Boyu	C-Bant (5.6 cm)
Çözünürlük	4.7 m x 5.1 m (Menzil x Azimut)
Geliş Açısı	40°
Polarizasyon (Alım Modu)	Tam Polarimetrik
Görüntü Alım Tarihleri	13.06.2016
	07.07.2016
	31.07.2016
	24.08.2016

Çalışma kapsamında, uydu görüntülerinin alım tarihleri ile eş zamanlı arazi çalışması gerçekleştirilmiş ve ürünlere ilişkin örnek noktalar el GPS'i yardımıyla toplanmıştır. Arazi çalışmasının ardından, örnek noktalardan faydalanarak tarla bazında (poligon olarak) yer gerçekliği verisi oluşturulmuş ve bu veriler hem sınıflandırma hem de doğruluk analizi işlemlerinde kullanılmıştır (Tablo 2).

Tablo 2. Yer gerçekliği verisi

Sınıf	Yer Gerçekliği Verisi (Piksel Sayısı)
Yem Bitkisi	5460
Mısır	5581
Patates	6093
Ayçiçeği	5361
Buğday	5481

Tek bakışlı karmaşık görüntü formatında elde edilen görüntülerden, doğrusal geri saçılım değerlerini içeren polarimetrik özelliklerin oluşturulması için bazı önişleme adımları gerekmektedir. Bu önişleme adımları Avrupa Uzay Ajansı'nın ücretsiz olarak sağladığı SNAP (Sentinel Application Platform) yazılımı (versiyon 6.0) ile gerçekleştirilmiştir. Bu önişleme adımları sırasıyla, görüntünün radyometrik (sigma nought) kalibrasyonu, benek (speckle) filtreleme, ortorektifikasyon ve piksel değerlerinin radar geri saçılım (dB) değerine dönüştürülmesi işlemleridir. Bu temel önişleme adımları dört farklı tarih için de uygulanıp çok zamanlı bir veriseti (time-stack) oluşturulmuştur.

3. Sınıflandırma Yöntemleri

3.1 Aşırı Gradyan Artırma (XGBoost)

Aşırı Gradyan Artırma (XGBoost), yüksek sınıflandırma performansı elde etmesi ile hem veri bilimi hem de uzaktan algılama alanında gittikçe popüler hale gelen ve gradyanı artırılmış (gradient boosted) karar ağaçları temeline dayalı çalışan bir makine öğrenme algoritmasıdır (Chen ve Guestrin, 2016; Georganos vd. 2018). Yeni nesil topluluk öğrenme algoritmalarından olan XGBoost, algoritmanın eğitilmesi sürecinde aşırı uyumlama (overfitting) sorununun önüne geçerek modelin genel doğruluğunu (performansını) artırmaktadır. Bu yöntemin başarısının altında yatan ana neden, öğrenme sürecinde kullandığı amaç fonksiyonudur. Amaç fonksiyonu, kayıp/yitim fonksiyonu (Loss function) ve düzenlileştirme (regularization) teriminden oluşmaktadır. Kayıp/yitim fonksiyonu, modelin yaptığı her bir tahmin değerinin (predicted class) gerçek değerinden olan farkını hesaplamaktadır. Düzenlileştirme terimi ise, modelin karmaşıklığını kontrol etmekte ve bu da modeldeki aşırı uyumlama sorununu ortadan kaldırmaktadır (Chen ve Guestrin, 2016; Rumora vd. 2020; Abdi, 2020). Bu çalışma kapsamında amaç fonksiyonu olarak, softmax çok sınıflı amaç fonksiyonu kullanılmıştır. XGBoost sınıflandırma işlemi, açık kaynaklı Python XGBoost paketinden (versiyon 1.0.2) faydalanarak gerçekleştirilmiştir (XGBoost, 2020) ve çalışma kapsamında belirlenen parametreler Tablo 3'te verilmiştir.

En uygun parametre değerleri (yaprak sayısı ve öğrenme oranı parametreleri için) ızgara arama (grid search) algoritması ile bulunmuştur. En uygun parametreler, yaprak sayısı için [100-1000] ve öğrenme oranı için ise [0.1-1] değerleri arasında aratılmıştır. Izgara arama algoritması, parametre çifti (yaprak sayısı ve öğrenme oranı parametreleri) için tüm kombinasyonları eğitir ve çapraz doğrulama tekniği ile model doğruluğunu hesaplar. En yüksek model doğruluğunu veren parametre çifti, en uygun parametre değeri olarak kabul edilir (Ayhan ve Erdoğmuş, 2014; Maxwell vd. 2018).

Tablo 3. XGBoost parametreleri

Parametre	Değer
Hızlandırma Türü	DART (Dropouts meet Multiple Additive Regression Trees)
Maksimum Derinlik	5
Özellik Alt-Örnekleme	1
Yaprak Sayısı	100
Öğrenme Oranı	0.1

3.2 Hafif Gradyan Artırma Makineleri

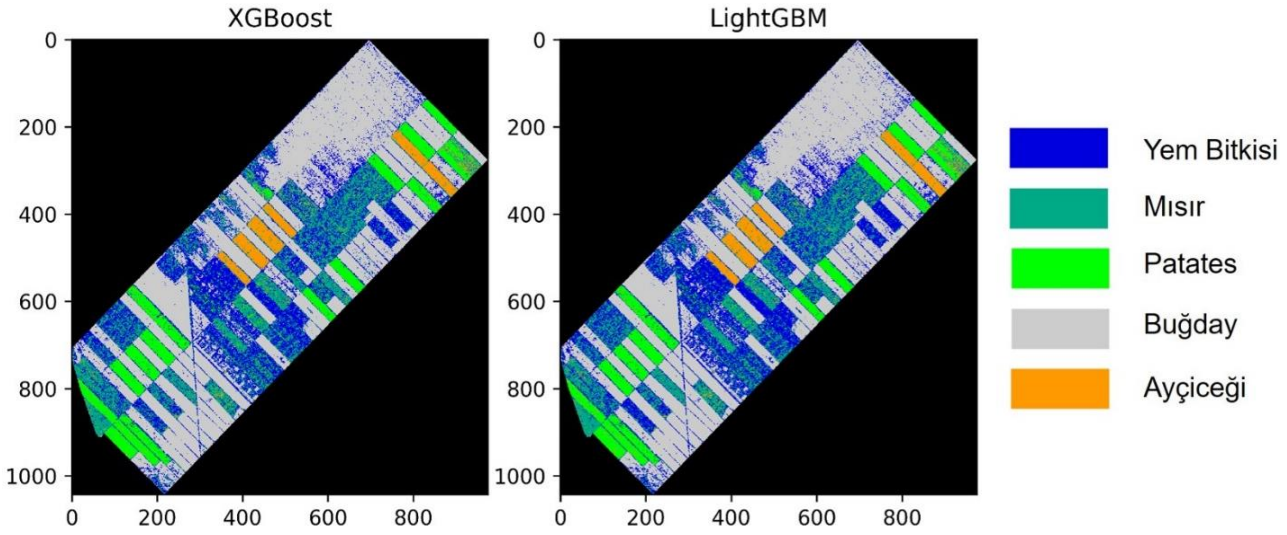
Hafif gradyan artırma makineleri (LightGBM), gradyan artırma çatısı altında çalışan karar ağaçları tabanlı yeni nesil topluluk öğrenme algoritmalarındandır. 2017 yılında Microsoft tarafından geliştirilen LightGBM, özellikle son yıllarda makine öğrenmesi yarışmalarında gösterdiği üstün başarıları ile hem veri bilimi hem de uzaktan algılama alanındaki araştırmacıların dikkatini çekmiştir (Ke vd. 2017; Liu vd. 2017; Ustuner ve Balik Sanli, 2019; Moorthy vd. 2020). Yöntemin isminden de anlaşılacağı üzere yüksek işlem hızına sahip bir algoritma olduğu için "Light" ön takısını almıştır. Bu yöntemi, diğer gradyan artırma algoritmalarından ayıran bir özelliği ise karar ağaçlarının eğitilmesi aşamasında kullandığı büyüme stratejisidir. LightGBM düşey yönde büyüme stratejisini (yapraksal büyüme) kullanırken, diğer gradyan artırma yöntemleri ise seviyesel (level-wise) büyüme stratejisini kullanmaktadır (Ke vd. 2017; Li vd. 2019; Ustuner ve Balik Sanli, 2019). LightGBM'i ayrıcalıklı ve özgün kılan diğer önemli bir detay ise, bünyesinde barındırdığı ve işlem hızını artırmasını sağlayan iki yeni algoritmadır. Bu algoritmalar, gradyan tabanlı tek taraflı örnekleme (Gradient-based One-Side Sampling, GOSS) ve ayrıcalıklı öznitelik desteleme (Exclusive Feature Bundling, EFB) yöntemleridir. Gradyan tabanlı tek taraflı örnekleme ile verinin tamamını kullanmak yerine veriden ürettiği alt örneklenmiş (subsampled) veri kümesini kullanır. Ayrıcalıklı özellik desteleme ile de seyrek yapıdaki özellikleri daha sık/yoğun özelliklere dönüştürerek işlem karmaşıklığını azaltmaktadır (Ke vd. 2017; Li vd. 2019; Ustuner ve Balik Sanli, 2019). LightGBM sınıflandırma işlemi için, Microsoft tarafından ücretsiz olarak erişilebilen açık kaynaklı Python lightgbm paketi (versiyon 2.3.1) kullanılmıştır (LightGBM, 2020). Sınıflandırma işleminde kullanılan parametreler aşağıdaki tabloda görülmektedir. Parametrelerin seçimi LightGBM'in ana sayfasındaki "parametre ayarlama" sayfasındaki öneriler dikkate alınarak seçilmiştir.

Tablo 4. LightGBM Parametreleri

Parametre	Değer
Hızlandırma Türü	GOSS
Maksimum Derinlik	5
Yaprak Sayısı	100
Öğrenme Oranı	0.1
Amaç Fonksiyonu	Çok-sınıflı
Özellik Alt-Örnekleme	1.0
Yapraktaki Minimum Veri Sayısı	20
İterasyon Sayısı	100

4. Uygulama

Sınıflandırılmış görüntülerin doğruluk analizinde k-katlamalı çapraz doğrulama tekniği kullanılmıştır. Bu yöntemde, k sayıda parçaya bölünen orijinal veri (yer gerçekliği verisi) setinin (k-1) sayıdaki parçası modeli eğitmek için kullanılırken, geriye kalan tek parça ise modelin doğruluğunu test etmek amacıyla kullanılır. Bu işlem k defa tekrar edilir ve her bir seferde farklı bir parça test/eğitim verisi olarak kullanılır. Model doğruluğu ise, elde edilen doğrulukların ortalamasına eşittir (Kohavi, 1995; Toosi vd., 2019). Çalışmamızda k değeri 5 olarak seçilmiştir. Sınıflandırılmış görüntüler Şekil 2'de sunulmuştur.



Şekil 2. Sınıflandırılmış Görüntüler

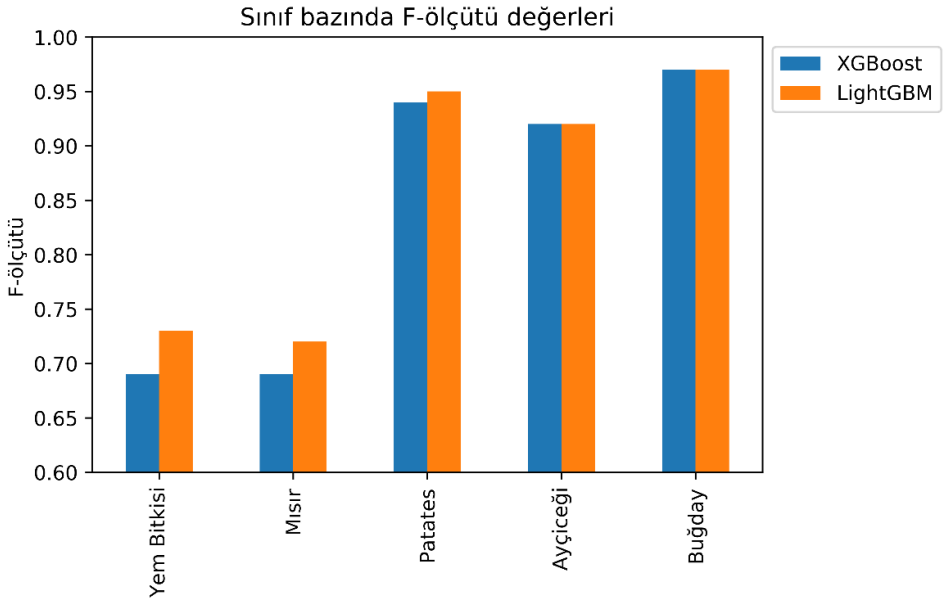
Çok zamanlı bir veri setinin kullanıldığı çalışmamızda, LightGBM yönteminin, XGBoost yöntemine göre hem işlem hızı (7.11 sn karşın 49.25 sn) hem de sınıflandırma performansı (0.860 karşın 0.845) açısından daha üstün olduğu sonucuna varılmıştır. Verilerin sınıflandırılması ve doğruluk analizi işlemleri Intel Core İ7 işlemci ve 8GB belleğe (RAM) sahip bir bilgisayarda Windows 10 işletim sistemi altında gerçekleştirilmiştir. Sınıflandırma ve doğruluk analizi işlemlerinde Python açık kaynak kodlu XGBoost (versiyon 1.0.2), LightGBM (versiyon 2.3.1) ve Scikit-learn (versiyon 0.19) (Pedregosa vd. 2011) kütüphaneleri kullanılmıştır.

Tablo 5. Doğruluk değerleri ve işlem süreleri

Yöntem	Doğruluk (+/- Standart Sapma)	İşlem Süresi (sn)
XGBoost	0.845 (+/- 0.035)	49.25
LightGBM	0.860 (+/- 0.027)	7.11

Sınıflandırma doğrulukları arasındaki farkın istatistiksel olarak anlamlı olup olmadığı McNemar's testi kullanılarak analiz edilmiştir (Abdikan vd. 2015). Ki-kare dağılımını esas alan McNemar's testine göre, çalışmamızda χ^2 test istatistiği %95 güven aralığındaki kritik değerden (χ^2=3.84) büyük olduğundan sınıflandırma sonuçları arasındaki farkın anlamlı olduğu sonucuna varılmıştır. Sınıf bazındaki doğrulukların analiz edilmesinde ise, kullanıcı doğruluğu (KD) ve üretici doğruluğu (ÜD) değerlerinin harmonik ortalamasına eşit olan F-ölçütü (F-score) değerleri kullanılmıştır. Sınıf bazında F-ölçütü değerleri Şekil 3'te görülmektedir.

$$F - \text{ölçütü} = 2 \times \frac{(\text{ÜD} \times KD)}{(\text{ÜD} + KD)} \quad (1)$$



Şekil 3. Sınıf bazında F-ölçütü değerleri

Ayçiçeği ve buğday sınıfları için, her iki yöntem de eşit F-ölçütü değerleri elde etmiştir fakat diğer üç sınıf türü (yem bitkisi, mısır ve patates) için LightGBM daha yüksek değerler elde etmiştir. Bu farklılığın ana nedeni, aynı eğitim verisi kullanılmasına rağmen algoritmaların öğrenme sürecindeki yapılarının birbirinden farklı olmasıdır. Patates, ayçiçeği ve buğday sınıfları için F-ölçütü değerleri 0.90'nın üzerinde elde edilirken, mısır ve yem bitkisi için bu değer 0.75 değerinin altında kalmıştır. Sınıf bazında F-ölçütü değerleri incelendiğinde ise, en yüksek değer (0.97) buğday sınıfına ait olurken, en düşük değer (0.69) ise XGBoost yöntemi ile yem bitkisi sınıfına aittir.

5. Sonuç ve Öneriler

Bu çalışmada, yeni nesil topluluk öğrenme algoritmalarından XGBoost ve LightGBM yöntemlerinin tarımsal ürünlerin sınıflandırılmasındaki performansları karşılaştırılmıştır. Bu performans karşılaştırması, hem model doğruluğu hem de işlem süresi baz alınarak yapılmıştır. Deneysel sonuçlarımız, tarımsal ürünlerin sınıflandırılması işleminde LightGBM yönteminin XGBoost yöntemine kıyasla daha üstün olduğu sonucuna varmıştır. LightGBM yönteminin işlem süresi olarak XGBoost yönteminden yedi kat daha hızlı olduğu tespit edilmiştir. Fakat iki yöntem için de en uygun parametrelerin bulunması hem yorucu hem de vakit alan bir süreçtir. Kullanılan girdi verinin boyutu ve sınıf türlerinin sayısına/karmaşıklığına bağlı olarak değişkenlik gösterebilen en uygun parametreler, bu iki yöntem için de üzerinde durulması gereken önemli konulardır. Bu çalışma ile ayrıca, çok zamanlı PolSAR verisinin tarımsal ürünlerin sınıflandırılmasındaki kullanılabilirliği test edilmiştir.

Gelecek çalışmalarımızda, daha çok sayıda sınıfa sahip verilerle bu kıyaslamanın tekrar yapılması ve özellikle kullanılan parametrelerin sınıflandırma doğruluğuna olan etkilerinin araştırılması hedeflenmektedir. PolSAR verisi kullanılarak yapılan bu çalışmanın, optik veriler üzerinde de test edilmesi planlanmaktadır.

Teşekkür

Bu çalışma, Yıldız Teknik Üniversitesi (YTÜ) Bilimsel Araştırma Projeleri Koordinatörlüğü tarafından FBA-2017-3062 kodlu proje kapsamında desteklenmiştir. Destekleri için Yıldız Teknik Üniversitesi (YTÜ) Bilimsel Araştırma Projeleri Koordinatörlüğü'ne teşekkür ederiz.

Kaynaklar

Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, *57*(1), 1-20, doi: 10.1080/15481603.2019.1650447.

Abdikan, S., Bilgin, G., Sanli, F. B., Uslu, E., & Ustuner, M. (2015). Enhancing land use classification with fusing dual-polarized TerraSAR-X and multispectral RapidEye data. *Journal of Applied Remote Sensing*, *9*(1), 096054, doi: 10.1117/1.JRS.9.096054.

Akar, Ö., & Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation, 1*(2), 105-112.

Atzberger, C. (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing, 5*(2), 949-981.

Ayhan, S., & Erdoğmuş, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *9*(1), 175-201.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24-31.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016. Proceedings.* (pp. 785–794).

Conrad, C., Dech, S., Dubovyk, O., Fritsch, S., Klein, D., Löw, F., ... & Zeidler, J. (2014). Derivation of temporal windows for accurate crop discrimination in heterogeneous croplands of Uzbekistan using multitemporal RapidEye images. *Computers and Electronics in Agriculture*, *103*, 63-74.

Dey, S., Mandal, D., Robertson, L. D., Banerjee, B., Kumar, V., McNairn, H., Bhattacharya, A. & Rao, Y. S. (2020). In-season crop classification using elements of the Kennaugh matrix derived from polarimetric RADARSAT-2 SAR data. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102059, doi: 10.1016/j.jag.2020.102059.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., & Wolff, E. (2018). Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience & Remote Sensing*, *55*(2), 221-242.

Gómez-Chova, L., Tuia, D., Moser, G., & Camps-Valls, G. (2015). Multimodal classification of remote sensing images: A review and future directions. In *Proceedings of the IEEE*, 103(9), (pp. 1560-1584). IEEE.

Hütt, C., Koppe, W., Miao, Y., & Bareth, G. (2016). Best accuracy land use/land cover (LULC) classification to derive crop types using multitemporal, multisensor, and multi-polarization SAR satellite images. *Remote Sensing*, *8*(8), 684, doi: 10.3390/rs8080684.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., ... & Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing, 7*(9), 12356-12379.

Jiang, H., Li, D., Jing, W., Xu, J., Huang, J., Yang, J., & Chen, S. (2019). Early season mapping of sugarcane by applying machine learning algorithms to Sentinel-1A/2 time series data: a case study in Zhanjiang City, China. *Remote Sensing*, *11*(7), 861, doi: 10.3390/rs11070861.

Jiao, X., Kovacs, J. M., Shang, J., McNairn, H., Walters, D., Ma, B., & Geng, X. (2014). Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *96*, 38-46.

Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, *11*(5), 352-359.

Khosravi, I., & Alavipanah, S. K. (2019). A random forest-based framework for crop mapping using temporal, spectral, textural and polarimetric observations. *International Journal of Remote Sensing*, *40*(18), 7221-7251.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems (NIPS 2017), 2017. Proceedings.* (pp. 3146-3154).

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *14th international joint conference on Artificial intelligence (IJCAI'95)*, *1995*. *Proceedings*. (pp. 1137-1145).

Li, W., Ding, S., Chen, Y., Wang, H., & Yang, S. (2019). Transfer learning-based default prediction model for consumer credit in China. *The Journal of Supercomputing*, *75*(2), 862-884.

LightGBM, (2020, June 15). *LightGBM Python API,* Retrieved from https://lightgbm.readthedocs.io/en/latest/Python-API.html.

Liu, L., Ji, M., & Buchroithner, M. (2017). Combining partial least squares and the gradient-boosting method for soil property retrieval using visible near-infrared shortwave infrared spectra. *Remote Sensing*, *9*(12), 1299, doi: 10.3390/rs9121299.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784-2817.

McNairn, H., & Shang, J. (2016). A review of multitemporal synthetic aperture radar (SAR) for crop monitoring. In Y. Ban (Eds.), *Multitemporal Remote Sensing: Methods and Applications,* (pp. 317-340), Cham, Switzerland: Springer International Publishing AG.

McNairn, H., Ellis, J., Van Der Sanden, J. J., Hirose, T., & Brown, R. J. (2002). Providing crop information using RADARSAT-1 and satellite optical imagery. *International Journal of Remote Sensing*, *23*(5), 851-870.

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(8), 1778-1790.

Moorthy, S. M. K., Calders, K., Vicari, M. B., & Verbeeck, H. (2019). Improved supervised learning-based approach for leaf and wood classification from LiDAR point clouds of forests. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(5), 3057-3070.

Pal, M. (2012). Advanced algorithms for land use and cover classification. In X. Yang & J. Li (Eds.), *Advances in Mapping from Remote Sensor Imagery: Techniques and Applications*, (pp. 70-82), Boca Raton: CRC Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Petropoulos, G. P., Kalaitzidis, C., & Vadrevu, K. P. (2012). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences*, *41*, 99-107.

Rumora, L., Miler, M., & Medak, D. (2020). Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers. *ISPRS International Journal of Geo-Information, 9*(4), 277, doi: 10.3390/ijgi9040277.

Saini, R., & Ghosh, S. K. (2019). Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date Sentinel-2A imagery. *Geocarto International*, 1-19, doi: 10.1080/10106049.2019.1700556.

Shi, X., Cheng, Y., & Xue, D. (2019, October). Classification Algorithm of Urban Point Cloud Data based on LightGBM. In *IOP Conference Series: Materials Science and Engineering* (Vol. 631, No. 5, p. 052041). IOP Publishing.

Skakun, S., Kussul, N., Shelestov, A. Y., Lavreniuk, M., & Kussul, O. (2016). Efficiency assessment of multitemporal C-band Radarsat-2 intensity and Landsat-8 surface reflectance satellite imagery for crop classification in Ukraine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(8), 3712-3719.

Toosi, N. B., Soffianian, A. R., Fakheran, S., Pourmanafi, S., Ginzler, C., & Waser, L. T. (2019). Comparing different classification algorithms for monitoring mangrove cover changes in southern Iran. *Global Ecology and Conservation*, *19,* e00662, doi: 10.1016/j.gecco.2019.e00662.

Ustuner, M., & Balik Sanli, F. (2019). Polarimetric Target Decompositions and Light Gradient Boosting Machine for Crop Classification: A Comparative Evaluation. *ISPRS International Journal of Geo-Information, 8*(2), 97, doi: 10.3390/ijgi8020097.

Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., & Brivio, P. A. (2015). In-season mapping of crop type with optical and X-band SAR data: A classification tree approach using synoptic seasonal features. *Remote Sensing*, *7*(10), 12859-12886.

Waldhoff, G., Curdt, C., Hoffmeister, D., & Bareth, G. (2012). Analysis of multitemporal and multisensor remote sensing data for crop rotation mapping. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1*, 177-182.

XGBoost, (2020, June 15). *XGBoost Python Package*, Retrieved from https://xgboost.readthedocs.io/en/latest/python/index.html.

Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, 430-443.

2020
Turkish Journal of Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, September 2020, 1(2): 106-124
Journal homepage: http://www.dergipark.org.tr/en/pub/rsgis

Simulating Future Ecosystem Services of the Sokoto-Rima Basin as Influenced by Geo-Environmental Factors

Jeo-Çevresel Faktörlerden Etkilenen Sokoto-Rima Havzasının Gelecekteki Ekosistem Servislerinin Simüle Edilmesi

Saheed Adekunle Raji[1,2]*, Mayowa Fasona[2], Shakirudeen Odunuga[2]

[1]*Federal University of Petroleum Resources, College of Science, Department of Env. Management and Toxicology Effurun/Nigeria.*
[2]*University of Lagos, Faculty of Social Sciences, Department of Geography, Akoka-Lagos/Nigeria.*

ORIGINAL PAPER

***Corresponding author:**
Saheed Adekunle Raji
raji.saheed@fupre.edu.ng

doi:

Article history:
Received: 11.07.2020
Accepted: 07.09.2020
Published: 30.09.2020

Abstract

Understanding the possible pattern of future ecosystem services can provide a systematic foundation for environmental resource management. This is vital towards science-based policy design, formulation, implementation, and review for the ecologically sensitive regions of the world such as the semi-arid West Africa. Therefore, the overarching objective of this study was to uncover the future ecosystem services of the Sokoto-Rima basin, particularly crop production (CP), seasonal water yield (SWY), habitat quality (HQ), and nutrient retention ratio (NRR) using InVEST software. Future (2050) land cover was constructed based on 2002, 2012 and 2015 Climate Change Initiative (CCI) remotely sensed data of the European Space Agency (ESA) using the artificial neural network algorithm of QGIS software. Results showed that CP averaged at 1.5 to 2.6 tons/km^2, indicating a low crop yield; annual SWY averaged at 464.64 mm, showing a low flow status; roughly 73% of the basin have low HQ; and NRR remains in the low category throughout the period of study. This suggests that scenario-based further assessment of ecosystem service interactions could be vital for sustainable land use policy.

Keywords: *GIS, Semi-arid, Ecosystem services, Sokoto-Rima*

Özet

Gelecekteki ekosistem servislerinin olası modelini anlamak, çevresel kaynak yönetimi için sistematik bir temel sağlayabilir. Bu, yarı kurak Batı Afrika gibi dünyanın ekolojik olarak hassas bölgeleri için bilime dayalı politika tasarımı, formülasyonu, uygulanması ve gözden geçirilmesi için hayati önem taşımaktadır. Bu nedenle, bu çalışmanın genel amacı, InVEST yazılımını kullanarak Sokoto-Rima havzasının gelecekteki ekosistem servislerini, özellikle mahsul üretimi (MÜ), mevsimsel su verimi (MSV), habitat kalitesi (HK) ve besin tutma oranını (BTO) ortaya çıkarmaktır. Gelecekteki arazi örtüsü (2050), QGIS yazılımının yapay sinir ağı algoritması kullanılarak 2002, 2012 ve 2015 İklim Değişikliği Girişimi'nin (CCI) uzaktan algılanmış Avrupa Uzay Ajansı (ESA) verilerine dayanarak oluşturulmuştur. Sonuçlar, MÜ'nin ortalama 1,5 ila 2,6 ton/km^2 olduğunu ve düşük mahsul verimine işaret ettiğini, yıllık MSV ortalamasının 464,64 mm olduğunu ve bunun düşük akış durumunu gösterdiğini; havzanın yaklaşık %73'ünün düşük habitat kalitesine sahip olduğunu ve BTO'nun çalışma süresi boyunca düşük kategoride kaldığını göstermektedir. Bu durum, ekosistem servis etkileşimlerinin senaryoya dayalı olarak daha fazla değerlendirilmesinin, sürdürülebilir arazi kullanım politikası için hayati önem taşıyabileceğini göstermektedir.

Anahtar kelimeler: *CBS, Yarı kurak, Ekosistem servisleri, Sokoto-Rima*

1. Introduction

Global attention on ecosystem services became intensified sequel to the publication of the Millennium Ecosystem Assessment (MEA) in 2005, which showed that roughly 70% of the 25 major ecosystem services of the Earth have either declined or are currently declining, thereby calling for a multi-scale and integrated approach towards reversing the trend (MEA, 2005a; MEA, 2005b; Setten et al. 2012; Wangai et al. 2016; Hølleland et al. 2017; Haines-Young & Potschin, 2018). The import of this call is borne out of the fact that vital human needs of food, shelter, water, medicine, climate regulation, pollution control and general wellbeing are directly and indirectly supplied by ecosystem services (MEA, 2005b). Therefore, natural supply of ecosystem services is directly proportional to socio-ecological and environmental dynamics. One key derivative from the MA report is the need to engage emerging scientific approaches in measuring, mapping and modeling ecosystem services dynamics towards sustainable use of ecological resources (Hølleland et al. 2017; Haines-Young and Potschin, 2018). The achievement of this goal however is a function of geography, sociocultural, economic and environmental value which is attached to the specific ecosystem service (Setten et al. 2012; Yang et al. 2018).

Given that the ability of an ecosystem to supply good and services is directly proportional to the functioning and integrity of that particular ecosystem, an altered environment such as dam construction for water, nutrient enhancement for soil fertility, and agricultural land expansion will proportionally impair the delivery of some ecosystem services (MEA, 2005b; Wangai et al. 2016; Haines-Young & Potschin, 2018). These activities could create rapid merits in terms of improved crop yield, increased water flow and enhance other consumption benefits in the short term; however, the long term consequences for biodiversity and other ecological losses could be severe. Natural responses to these dynamics is a function of ecological integrity and resilience of the particular ecological biome (MEA, 2005b; Roche & Campaigne 2017). Within semi-arid contexts, these scenarios have become normalized as land cover change (also referred to as land use change) remains the main driving force altering ecosystem services mostly driven by mono-directional policies such as agricultural improvement programmes, water supply enhancement schemes etc. (Yang et al. 2018). The manifestation of these events within a fragile ecosystem with brittle ecological balance presents a herculean challenge for environmental management (Yang et al. 2018; Wangai et al. 2016). It is therefore vital to spatially disentangle these ecosystem services and examine the nature and dynamics in the future for sustainable environmental management.

The Sokoto-Rima basin, a vital semi-arid ecological unit of northern Nigeria, is not immune to these realities. In this area, livelihood is tied to land where agriculture remains the lifeblood of the local economy. Despite this merit, natural dynamics of climate variability and change (Oguntunde et al. 2011), acute natural land degradation emanating from southward migration of the Sahara Desert (Oladipo, 1995; Olagunju, 2015), and multifaceted droughts of the 19th century (Oyebande and Odunuga, 2010) have altered the delivery of vital ecosystem services. These coupled with unsustainable agricultural practices such as overgrazing have led to ecosystem resource conflicts between crop farmers and migrant herdsmen (Fasona et al. 2007).

To guarantee the sustainability of the ecosystem services, management approaches that guarantee and synthesize status monitoring through mapping and future projections are vital (Huber et al. 2013). Such management approaches, driven by policy, will examine the dynamic factors that stimulate changes in the ecosystem services, track their transition pathways, and project future scenarios with respect to socioeconomic, ecological and environmental effects. Literature on the use the spatial tools of Remote Sensing and Geographic Information Systems (GIS) for simulating land cover and other geo-environmental factors for ecosystem services is on the rise. In a recent study, Satya et al. (2020) utilised GIS-based spatial factors to simulate future land use land cover scenarios of city of Warangal India and results showed that biophysical and socio-economic factors contributed to rise of built-up areas while agricultural land use will decline in the future. Fu et al. (2017) claimed that land use change influences variations in ecosystem services in mountainous areas of China. The study simulated units of provisioning ecosystem services to indicate heterogeneous nature of land use decisions on management of ecological resources. Yang et al. (2018) predicted that future ecosystem services are directly influenced by land use policy and climate in Yanhe watershed of northern Shaanxi Province of China.

The aim of this study was therefore to simulate the future ecosystem services of the Sokoto-Rima basin focusing on crop production, seasonal water yield, habitat quality and nutrient retention ratio. Specifically, the study was set to: (1) Examine the key geo-environmental factors that drive changes in land cover; (2) Quantify the current ecosystem services and; (3) model the future ecosystem services of the Sokoto-Rima basin using InVEST software package. This is with a view to examine the dynamics of changes in ecosystem services. The result of the study is expected to provide a suitable background for the assessment of ecological resources of the Sokoto-Rima basin.

2. Materials and Methods

2.1 The Study Area

Located in north-western end of Nigeria, Sokoto-Rima basin is a transnational catchment that extends beyond the boundary of Nigeria into neighboring countries, particularly Niger in the north and Benin Republic in the west. It is geographically confined to Latitudes 10˚32'35'' N to 13˚32'55''N and Longitudes 3˚30'30'' E to 8˚1'15''E with a total land area of 94,026.5 km^2 (see Figure 1). The ecosystem of the study area is greatly influenced by tropical savanna climate which is characterized by considerable seasonal variations. Annually, rainfall averages 350 mm in the north and 895 mm in the south while temperature is relatively high throughout the year, exceeding 30^0 C. Climate dynamics reflects on river flow characterization and ecosystem functioning with series of minor ephemeral streams that feed the major permanent rivers and lakes (Figure 1). Agriculture, particularly crop production and animal husbandry, is the lifeblood of the basin, hence, its settlement pattern is mainly rural. Crops such as rice, sorghum, millet, corn, beans, and groundnut are extensively cultivated on a small-scale and climate-reliant fashion.

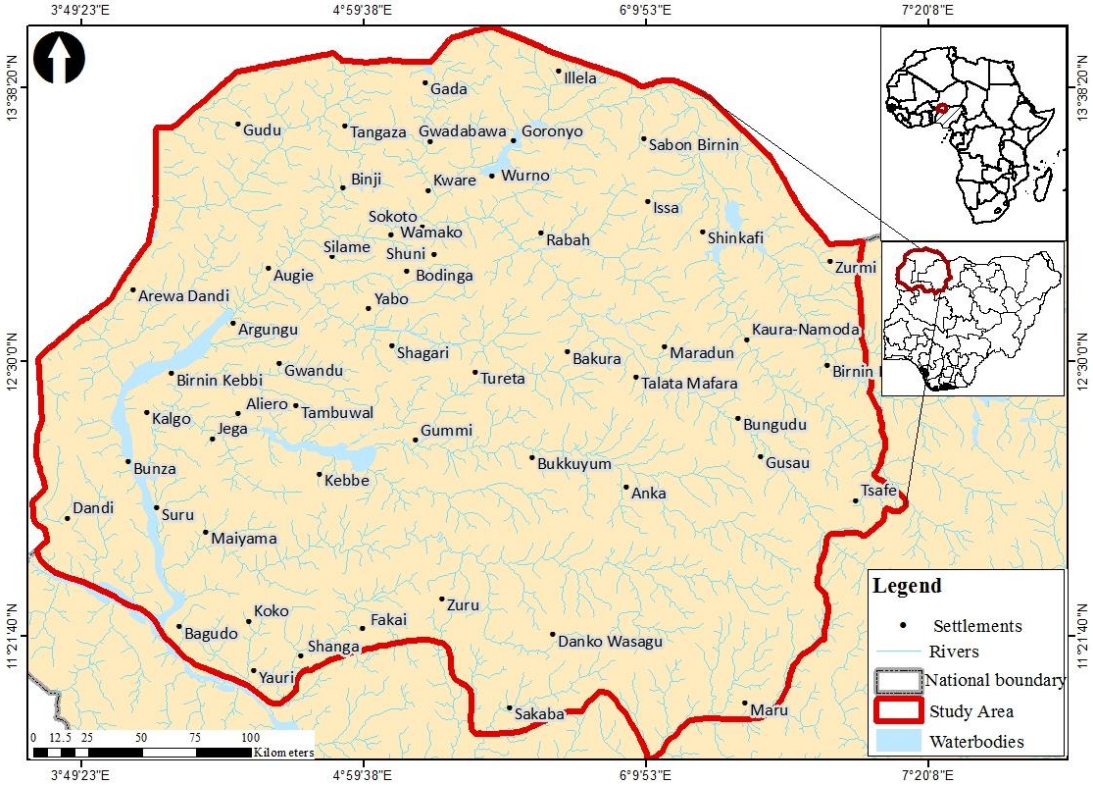


Figure 1. The study area in context of Africa and north-western part of Nigeria

2.2 Data Characteristics and Sources

The CCI (Climate Change Initiative) land cover data, a multi-sourced pre-classified remotely sensed data, was acquired from the portal of the European Space Agency (ESA). The data was processed from previously existing satellite data sources mainly: AVHRR (Advanced Very High Resolution Radiometer), Envisat MERIS (Medium Resolution Imaging Spectrometer), Sentinel-3 Ocean and Land Colour Instrument (OLCI), and Sea and Land Surface Temperature Radiometer (SLSTR). It has a spatial resolution of 300 meters and a dynamic range of 32 bits, which is suitable for an extensive basin such as the Sokoto-Rima. The pre-processed data was classified from source using a hybrid classification algorithm, which combines the strengths of supervised and unsupervised techniques (Defourny et al. 2016). It has 75% accuracy pre-tested during data validation task with GlobCover dataset (Defourny et al. 2016). Pérez-Hoyos et al. (2017) demonstrated the fitness of the data for cropland monitoring from continental to country level in Africa. However, data for the years 2002, 2012 and 2015 were acquired for the study.

Elevation, slope and aspect datasets were derived from ALOS World 3D Digital Surface Model (DSM) data acquired from the archives of JAXA (Japan Aerospace Exploration Agency). It has a horizontal resolution of one arc-second and 32-bit quantization depth. The eMODIS NDVI and evapotranspiration (ET) datasets were acquired from the USGS (United States Geological Survey) Famine Early Warning System (FEWS) data portal for West Africa. The datasets have a spatial resolution of 250 meters and a quantization level of 16 bits. Monthly rainfall and maximum temperature data covering climatic stations (Sokoto, Yelwa, Birnin Kebbi, Argungu, Gusau, Goronyo, Wurno, Kano, and Kaduna) within Nigeria were sourced from the Nigeria Meteorological Agency (NIMET) while supplementary climate data was sourced from Princeton University's station-based data from locations in Benin (Malanville) and Niger (Dabnou). The data spanned the years 1951 to 2015. Soil data (HYSOG Soil Group) was acquired from the portal of United States Oak Ridge National Library Distributed Active Archive Centre (ORNL DAAC). The data, which was developed in 2015, has a spatial resolution of 250 meters and 16-bit quantization level. Geology data was sourced from the defunct Federal Department of Agricultural Land Resources (FDAL) Kaduna Nigeria. It has a scale of 1:650,000. Population values and local government land areas were extracted from the archives of National Bureau of Statistics (NBS) 2010 version. These datasets were used to generate population density for each of the local government areas within the Sokoto-Rima basin.

2.3 Land cover Simulation

Simulation of the future land cover is based on the assumption that future land cover scenarios are governed by coupled human and natural systems (Lambin et al. 2003; Fasona et al. 2014). An algorithm that integrates top-down system dynamics with bottom-up cellular automata becomes the most apt due to its system coupling merits. MOLUSCE (Model for Land Use Change Evaluation), which is a spatially explicit QGIS plugin that permits the utilization of machine learning algorithm known as artificial neural network (ANN) with cellular automata (CA) to model future land cover patterns, was therefore adopted for the study. ANN facilitates the execution of fuzzy logic such that its algorithm works on multiple data structures and the output is binary-like. Its core element is designed to interact with connected neurons and the attached weight adjustment between the neurons. This process facilitates its ability to generate transition probabilities which is the chances of one data transforming into another based on the neuron connections within the multispectral data space. ANN thus permits the incorporation of series of geo-environmental variables as determinants of land cover change. The entire process was thus used to simulate 2015 and 2050 land cover data and validate with acquired land cover.

To set the model in MOLUSCE, land cover data of 2002 and 2012 alongside 13 geo-environmental variables (rainfall, temperature, elevation, slope, aspect, soil, geology, distance to road, distance to water body and distance to settlements, population density, and NDVI for 2002 and 2012) were set as input variables in the software. Pearson's correlation coefficient, which is a part of the MOLUSCE QGIS plugin, was used to check their level of multi-collinearity amongst these datasets.

2.4 Simulation Validation

Four model validation methods (Kappa for histogram information (K_{hist}), Kappa for pixel level location (K_{loc}) and overall kappa K_{std} with percentage of correctness) as enunciated by Geri et al. (2011) were adopted. These are defined in equations (1), (2) and (3) as:

$$K_{hist} = \frac{P_{max} - P(Y)}{1 - P(Y)} \quad (1)$$

$$K_{loc} = \frac{P(X) - P(Y)}{P_{max} - P(Y)} \quad (2)$$

$$K_{std} = \frac{P(X) - P(Y)}{1 - P(Y)} \quad (3)$$

where: $P(X) = \sum_{i=1}^{c} p_{ij} = \sum_{i=1}^{c} p_{iT} p_{Tj}$, $P_{max} = \sum_{1=i}^{c} \min(P_{iT} P_{Tj})$, P_{ij} is the *i,j*th cell of the land cover contingency table, P_{iT} is the total of all cells in the *i*th row, P_{Tj} is the addition of all cells in the *j*th column, and *c* is the number of land cover categories.

2.5 Measuring Ecosystem Services

Sequel to the nature of the study area in terms of socioeconomics, dominant livelihood, biodiversity as well as data availability on ecosystem functioning, four ecosystem services were selected to be quantified, namely, crop production (CP), seasonal water yield (SWY), habitat quality (HQ), and nutrient retention ratio (NRR). Each of these were computed using the models enshrined in the InVEST (Integrated Valuation of Ecosystem Services and Trade-offs) and expressed as follows:

Crop production; The CP Regression Model was used for the study. It was based on separate computation of average yield of a given crop per land area within the multispectral space. Data requirements of the model include land cover, and crop fertilization rates identified with agricultural land use with inherent local climate. Mathematically, the model was defined by Mueller et al. (2012) as stated in equation (4):

$$Y_{modGC} = min\left(Y_{max}(1 - b_{NP}\, exp\,(-c_N N_{GC})), Y_{max}\,(1 - b_{NP}\, exp\,(-c_P P_{GC})), Y_{max}\,(1 - b_K\, exp\,(-c_K K_{GC}))\right) \quad (4)$$

where: Y_{modGC} is yield for a particular crop (tonnes/km^2); Y_{max} is maximum yield; b_{NP} and b_K are the y-intercepts for each nutrient-yield response curve; c_N, c_P, and c_K are response coefficients that describe the percentage of Y_{max} attained at a given nutrient level; N_{GC} P_{GC} K_{GC} are the crop-based fertilization rates for rice, sorghum, millet and groundnut stated in Table 1. Modeled yield for each crop was added using raster math in ArcGIS 10.5 to produce the overall CP value (tonnes/km^2).

Table 1. Specific crop fertilization rates used for CP model (Tarfa et al. 2017)

S/N	Crop name	Nitrogen rate	Phosphorous rate	Potassium rate
1	Rice	100.0449	22.03867	33.0196
2	Sorghum	64.0113	14.0059	25.2265
3	Millet	60.3244	13.0096	12.1201
4	Groundnut	58.0119	22.1017	35.0112

Seasonal water yield; The SWY Model was used to quantify water yield in the Sokoto-Rima basin. It calculates the average amount of available water within a parcel of land as a function of the land cover and represents the quickflow of the basin. The Sokoto-Rima basin has a tropical continental climate where water availability is restricted to the wet season which spans June to August, an evidence of clear seasonality. Data required by the model includes monthly evapotranspiration, monthly rainfall, land cover, basin outlay, and digital elevation model (DEM). The SWY model is mathematically described in equations (5) and (6):

$$QF_{i,m} = n_m * \left(\left(a_{i,m} - S_i\right) exp\,\left(-\frac{0.2S_i}{a_{i,m}}\right) + \frac{S_i^2}{a_{i,m}} exp\,\left(\frac{0.8S_i}{a_{i,m}}\right) E_1\left(\frac{S_i}{a_{i,m}}\right)\right) * (25.4\left[\frac{mm}{in}\right] \quad (5)$$

where: $S_i = \frac{100}{CN_i} - 10\ [in]$, CN_I is the curve number for pixel *i*, a function of the local land cover and soil type; $a_{i,m}$ is the mean rain depth on a rainy day at pixel *i* on month *m*, and E_1 is the exponential integral function, $E_1(t) = \int_1^\infty \frac{e^{-t}}{t} dt$ in which *t* is the time study period.

The yearly quickflow which is the aggregate of the monthly (*m*) is given as:

$$QF_i = \sum_{m=1}^{12} QF_{i,m} \quad (6)$$

Habitat quality; The Habitat Quality model stated in InVEST was used to quantify HQ. The model hypothesizes that spatial patterns of biodiversity can be estimated using land cover distribution with respect to localized spatial threats. This approach was adopted as a measure of ecosystem biodiversity and its capacity to estimate spatial habitat degradation across the study area. Based on the level of socio-economic development of the study area, cultivated lands, settlements, railways lines, main roads and proximity to waterbody were identified as spatial threats and thus integrated to the model. Sharp et al. (2018) described the HQ model, which is mathematically expressed as:

$$Q_{xj} = H_j\left(1 - \left(\frac{D_{xj}^z}{D_{xj}^z + k^z}\right)\right) \quad (7)$$

where Q_{xj} is habitat quality of cell *x* in land cover *j*, D_{xj}^{z} is associated total threat level within the multispectral space *x* and *j*, *z* (hard coded =2.5) and *k* (half-saturation constant was set as 0.5) are scaling parameters.

Nutrient retention ratio; The Nutrient Delivery Ratio Model was used to quantify NRR. The model hypothesizes a nutrient cycling mass balance approach, which explains the spatial movement of nutrients under the influence of vegetation, rainfall and crop utilization of nutrients. DEM, land cover, nutrient load parameter, retention efficiency and length and nutrient runoff proxy (rainfall) were the datasets required by the model. The derived equation for NRR is given as:

$$NRR = 1 - \frac{NRE}{NT} \tag{8}$$

where: *NRR* (nutrient retention ratio) is the fraction of nutrients retained at a pixel position. *NT* (nutrient transport) is a measure of nutrient flow across from a pixel position to another. NRR ranges between 0 and 1 indicating no and maximum retention ratio, respectively.

3. Result and Discussion

3.1 Multicollinearity Test Analysis of the Geo-environmental Variables

The test for multicollinearity as revealed by Pearson's correlation measured at α= 0.05 level of significance is presented in Table 2. Although the returned correlation coefficient showed direct (positive) and indirect (negative) linear association correlation, no degree of multicollinearity was detected. Overall, the relationship between the geo-environmental variables that could trigger or indicate the presence of multiple collinearity linkages was low (the computed *r* was less than 0.7). This shows that causally irrelevant variables have been eliminated and the variables are therefore suitable for future land cover simulation.

Table 2. Pearson's Correlation matrix of the geo-environmental variables

	Aspect	NDVI (2012)	Rain	Pop den	Elev.	Temp.	Soil	Dist. to set	NDVI (2002)	Geol	Dist. to road	Slope	Dist. to water
Aspect	-	0.02	0.02	-0.01	0.01	-0.02	0.02	0.02	0.02	0.01	0.02	0.22	-0.01
NDVI (2012)		-	0.07	-0.12	0.16	-0.19	0.02	0.12	0.48	0.03	0.13	0.02	-0.02
Rainfall			-	-0.27	-0.14	-0.16	0.11	0.13	0.25	0.20	-0.06	0.08	-0.04
Pop. den.				-	-0.15	0.24	-0.04	-0.16	-0.19	-0.05	-0.06	-0.02	0.10
Elev.					-	-0.61	-0.13	0.06	0.30	0.17	0.13	0.15	-0.25
Temp.						-	-0.01	-0.11	-0.36	-0.25	-0.14	-0.06	-0.07
Soil							-	-0.06	0.05	0.14	-0.07	-0.07	-0.11
Dist. to set								-	0.24	-0.19	0.36	0.01	0.01
NDVI (2002)									-	0.14	0.19	0.07	-0.09
Geol										-	-0.09	0.05	-0.08
Dist. to road											-	0.02	0.04
Slope												-	-0.02
Dist. to water													-

3.2 Spatiotemporal Dynamics of Land cover

Since the functioning of ecosystems and the capacity to provide goods and services are analogous to dynamics on land, it is essential to scrutinize the nexus between the various natural and anthropogenic activities on land as indicated the dynamics of land cover characterization. Details of spatiotemporal change in the mapped land cover classes for the Sokoto-Rima basin are presented in Table 3 for the period 2002 to 2012.

As an agrarian landscape, cropland dominated the land cover in 2002 with land area of 68,470.94 km^2 which is almost three-quarters of the land proportion at 72.82%. Other land cover themes include agroforestry (11,843.30 km^2), shrubland (1,838.41 km^2), grassland (8,487.87 km^2), waterbody (3,043.92 km^2), settlement (81.90 km^2), bare surface (73.53 km^2), and woodland (186.64 km^2) occupying 12.6%, 1.96%, 9.02%, 3.24%, 0.09%, 0.08% and 0.20% of the land area, respectively. A decade after, in 2012, the land cover proportions returned similar distributions compared to 2002 with minuscule differences in some land cover themes. Specifically, cropland, waterbody, settlement, bare surface and woodland increased to 68,797.03 km^2 (73.17%), 3,056.43 km^2, (3.25%), 147.62 km^2, (0.16%), 90.55 km^2, (0.10%), 187.02 km^2, (0.20%) respectively. Agroforestry, shrubland and grassland reduced to 11,808.13 km^2 (12.56%), 1,692.01 km^2 (1.80%), and (8,247.72 km^2, 8.77%), correspondingly.

Pattern of change indicated that that cropland, waterbody, settlement, bare surface and woodland increased spatiotemporally by 0.35% (326.08 km^2), 0.01% (12.51 km^2), 0.07% (65.73 km^2), 0.02% (17.019 km^2), and 0.0004% (0.38 km^2), respectively (Table 3). Equally, agroforestry, shrubland and grassland contracted by -0.04% (-35.17 km^2), -0.16% (-146.4 km^2) and -0.26% (-240.14 km^2), respectively. It can be summed that land cover losses in the Sokoto-Rima basin is mainly dominated by natural landscapes which were replaced by anthropogenic activities particularly crop cultivation. The observed trend is anticipated to have consequences on the provisioning and regulating ecosystem goods and services which remain the prime source of sustenance in the Sokoto-Rima basin.

Table 3. Land cover change statistics for the Sokoto-Rima basin from 2002 to 2012

Land cover classes	Land cover (2002) (km^2)	Land cover (2012) (km^2)	Change in land area (Km^2)	2002 (%)	2012 (%)	Proportion of change (%)
Cropland	68,470.94	68,797.03	326.08	72.82	73.17	0.35
Agroforestry	11,843.30	11,808.13	-35.17	12.60	12.56	-0.04
Shrubland	1,838.41	1,692.01	-146.40	1.96	1.80	-0.16
Grassland	8,487.87	8,247.72	-240.144	9.03	8.77	-0.26
Waterbody	3,043.92	3,056.43	12.51	3.24	3.25	0.01
Settlement	81.90	147.62	65.73	0.09	0.16	0.07
Bare surface	73.53	90.55	17.02	0.08	0.10	0.02
Woodland	186.64	187.02	0.38	0.20	0.20	0.00
TOTAL	**94,026.51**	**94,026.51**		**100**	**100**	

3.3 Transition Potential Modeling based on Artificial Neural Network (ANN)

Transition potential analysis provides the level of performance of the ANN algorithm to generate simulated land cover data. Figure 2 shows the neural network learning curve with the red and green curves depicting the behavior of the neural training and validation, respectively. The spike in the graph shows the largest error possible at the 1,000th training sample level, which was further corrected as the training progresses thus eliminating both issues of unrepresentative train dataset and unrepresentative validation dataset. The representative test of good fit of the neural training is presented in Table 4. It shows that training of dataset returned 69.019% and 68.165% level of accuracy for the years 2015 and 2050, respectively. Meanwhile, the validation accuracy was found as 76.203% and 75.877% for the years 2015 and 2050, respectively.

Table 4. ANN training validation analysis

Parameters	2015 simulation value	2050 simulation value
Scale for training	80%	80%
Scale for validation	20%	20%
Hidden neurons	10	10
Learning rate	0.05	0.05
iterations	5,000	5,000
Training of dataset	69.019%	68.165%
Accuracy of validation dataset	76.203%	75.877%

Transition potential describes the foundation and conversion pathway with probability matrix which shows the transfer direction of land cover classes from a particular period to another. For this study, Table 5 describes the Markovian transition probability matrix for the eight land cover classes for the period 2002-2012. Overall, settlement (1.000), woodland (0.9995), waterbody (0.9987) and cropland (0.9965) had the highest Markovian probabilities while shrubland (0.7957) had the least indicating the range of possibilities to remain unchanged in the future. In terms of transformation to other land cover classes, cropland had the highest Markovian probabilities, thus, it can transform to cropland, shrubland, grassland, settlement, bare surface and woodland with values of 0.9965, 0.1705, 0.0293, 0.0012 and 0.0005, respectively. Woodland was found to have the least dynamic probability to change to other land cover classes apart from self-transformation. This makes woodland less likely to change overtime.

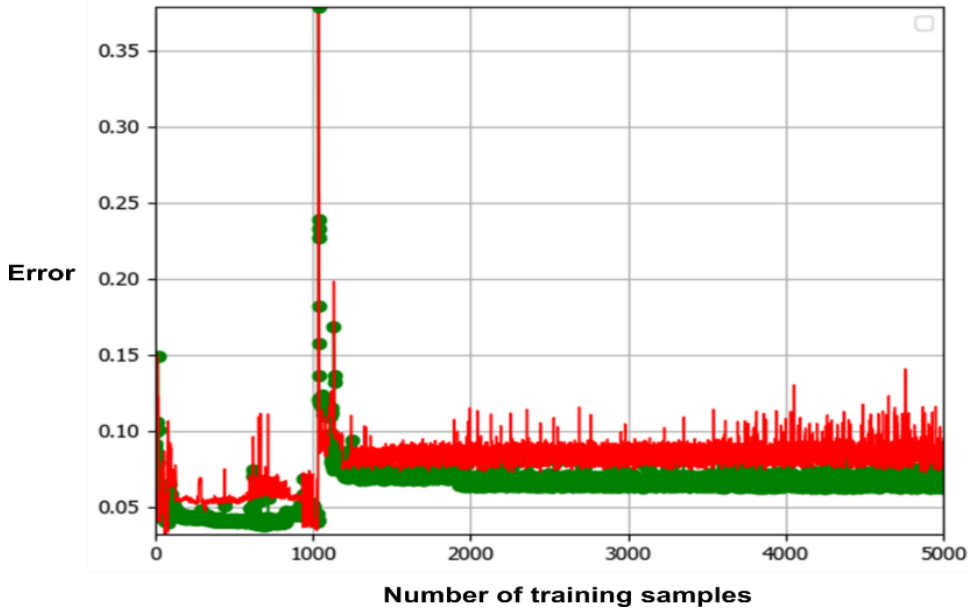


Figure 2. Neural network learning curve for the study with 5,000 training samples. The green curve shows the neural train generated from the training sample and the red curve shows the associated neural train validation

Table 5. Markovian transition probability matrix of land cover classes from 2002 to 2012

Land cover	Cropland	Agroforestry	Shrubland	Grassland	Waterbody	Settlement	Bare surface	Woodland
Cropland	0.9965	0.0000	0.0012	0.0017	0.0000	0.0006	0.0000	0.0000
Agroforestry	0.0000	0.9840	0.0106	0.0036	0.0000	0.0018	0.0000	0.0000
Shrubland	0.1705	0.0331	0.7957	0.0000	0.0007	0.0000	0.0000	0.0000
Grassland	0.0293	0.011	0.0024	0.953	0.0018	0.0004	0.0021	0.0000
Waterbody	0.0000	0.0000	0.0003	0.0005	0.9987	0.0005	0.0000	0.0000
Settlement	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Bare surface	0.0012	0.0000	0.0000	0.0098	0.0025	0.0000	0.9865	0.0000
Woodland	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9995

3.4 Cellular Automata (CA) based Dynamics of Land Cover Simulation

The result of the Markovian transitions stated in Table 5 generates three types of land cover results as determined by CA. The first is transition potential maps, which spatially display the probabilistic potential of one land cover class to change to another with values ranging from 0 to 100 indicating levels of land cover transition. A total of 41 maps were generated with cropland having strongest connection to all other 7 land cover classes. Second is the certainty raster, which is the difference between large transition potential maps of the two periods 2002 and 2012. The largeness of the transition potential maps creates room for the degree of occurrence of the land cover classes in the predicted and simulated land cover classes.

Thus, Figure 3 depicts the certainty raster for the years simulated 2015 and the predicted 2050. Transitional certainty for the simulated year 2015 depicts some varying ranges while that of the predicted 2050 returned proportions that is quasi-equal showing the pattern of anticipated land cover changes for the two periods.

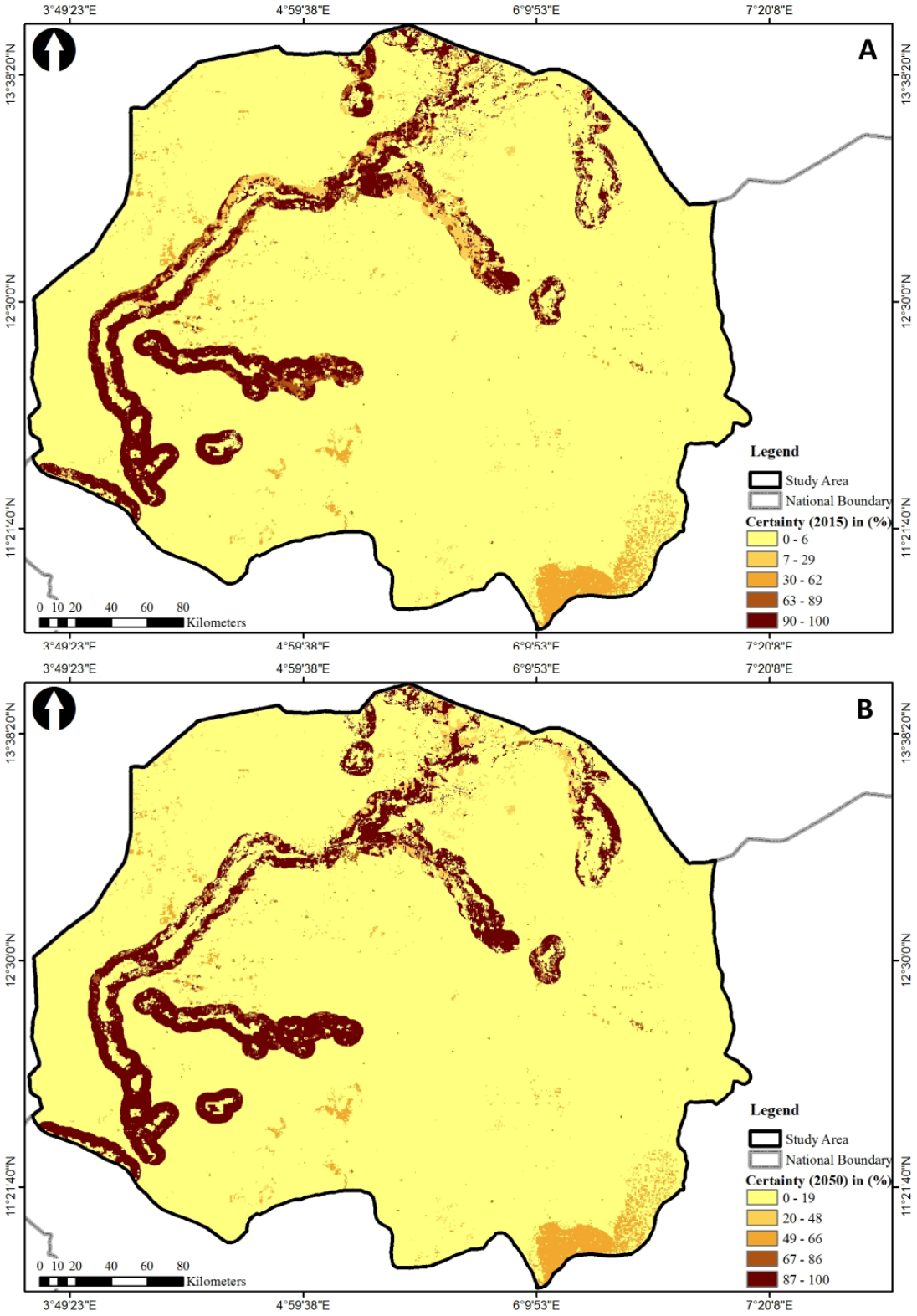


Figure 3. Certainty raster showing the spatial differentials in transition potentials for 2015 (a) and 2050 (b)

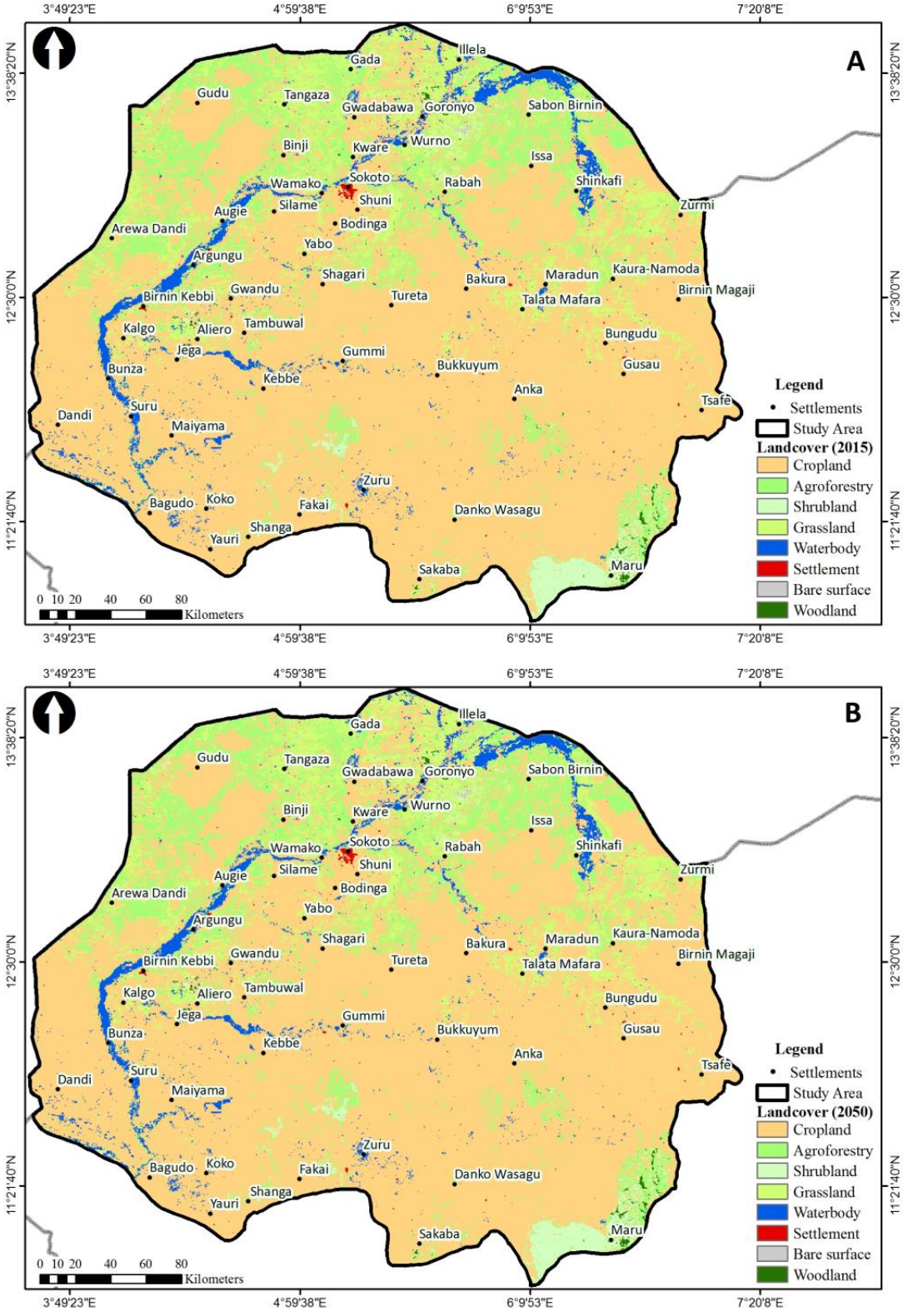


Figure 4. Simulated land cover for the years 2015 (a) and 2050 (b)

The third output is the 2015 simulated land cover and the predicted 2050 land cover (Figure 4). The corresponding land area and proportion for the simulated and predicted land cover datasets are specified in Table 7. In terms of proportionality, there exists some quasi-symmetric similarities. Cropland dominated the area with 73.02% (68,829.272 km^2) as simulated for 2015 and by 2050 as predicted, it will decrease to 73.126% (68,758.191 km^2) while woodland occupied the least proportion and land area with 0.198% (185.945 km^2) simulated for 2015 and as predicted for 2050, it will occupy 0.197% (185.315 km^2) of the Sokoto-Rima basin.

In terms of predicted changes, by 2050, it is expected that reductions will be observed in cropland, agroforestry, waterbody, bare surface and woodland while grassland and settlement will expand spatiotemporally.

Table 6. Proportion of the simulated land cover (2015) and predicted land cover (2050)

Land cover class	Simulated 2015 land area (km^2)	Proportion (%)	Predicted 2050 land area (km^2)	Proportion (%)	Land Change (km^2)	Change (%)
Cropland	68,829.272	73.202	68,758.191	73.126	-71.081	-0.076
Agroforestry	11,705.373	12.449	11,653.121	12.393	-52.252	-0.056
Shrubland	1,690.986	1.798	1,690.536	1.798	-0.45	0
Grassland	8,324.121	8.853	8,451.328	8.988	127.207	0.135
Waterbody	3,053.235	3.247	3,050.172	3.244	-3.063	-0.003
Settlement	147.837	0.157	148.288	0.158	0.451	0.001
Bare surface	89.729	0.095	89.549	0.095	-0.18	0
Woodland	185.945	0.198	185.315	0.197	-0.63	-0.001
TOTAL	**94,026.50**	**100**	**94,026.50**	**100**		

3.5 Validation of Land Cover Simulation

Table 7 showed the model performance of both simulated 2015 land cover and predicted 2050 datasets. The statistics for kappa location, histogram and overall showed closeness to 1, which indicates a strong agreement. Importantly, the overall percentage of correctness showed that the 2015 simulated land cover dataset returned 88.01%, while the 2050 predicted land cover data returned 87.87%, showing strong performance of the simulation and prediction tasks (Geri et al. 2011). The 2050 dataset is therefore suitable to be used to model future ecosystem services.

Table 7. Kappa index of the model validation for the simulated and predicted land cover datasets

Model validation parameter	Simulated 2015	Predicted 2050
Kappa (location)	0.9378	0.9279
Kappa (histogram)	0.9606	0.9674
Kappa (overall)	0.9142	0.9116
% of correctness	88.0124	87.8714

3.6 Nature of Baseline Ecosystem Services

Crop Production (2015); Five levels of crop production indicated by crop yields were detected: very low (0-0.58 tonnes/km^2), low (0.58-1.59 tonnes/km^2), moderate (1.59-2.63 tonnes/km^2), high (2.63-3.87 tonnes/km^2) and very high (3.87-6.16 tonnes/km^2) (Table 8). The aggregate of low and very low yield statuses constitute 56.64% while moderate status comprises 21.88%, showing that the crop yield scale in the Sokoto-Rima basin is average. This productivity level affirmed the assertion that crop production in the area is small-scale, which is producing at low level while areas with high crop yield status occupy roughly 21% of the land area. Spatially, these productivity levels can be observed in Figure 5, indicating low productivity are scattered throughout the study area while high yields can be identified in some location at the western, eastern and southern axis of the Sokoto-Rima basin.

Table 8. Areal extent of aggregate crop yields status for the static year 2015

Status of aggregate crop yield (tonnes/km^2)	Land area (km^2)	Proportion (%)
Very low (0-0.57)	27,751.64	30.42
Low (0.57-1.59)	24,742.45	26.22
Moderate (1.59-2.63)	20,749.85	21.88
High (2.63-3.87)	13,943.02	14.51
Very high (3.87-6.16)	6,839.53	6.97
Total	**94,026.50**	**100.00**

Seasonal Water Yield (2015); Hydrological processes of quick flow define the nature of water provision in a particular basin as defined by the SWY model. As shown in Table 9, seasonal water ranges from very low (0-88.69 mm) to very high (813.24-1155 mm) in the Sokoto-Rima basin. As expected of a typical dryland, 61% of the area has low SWY status while 23.18% fall under areas with high SWY and 15.83% moderate SWY. The lower section of Figure 5 shows the hydrological network of streams and rivers feeding main water bodies (Sokoto and Rıma Rivers) which defines the level of seasonal water in the basin. However, there exists a clearly marked spatial delineation between the north and the south in the former is characterized by low SWY and the latter high SWY.

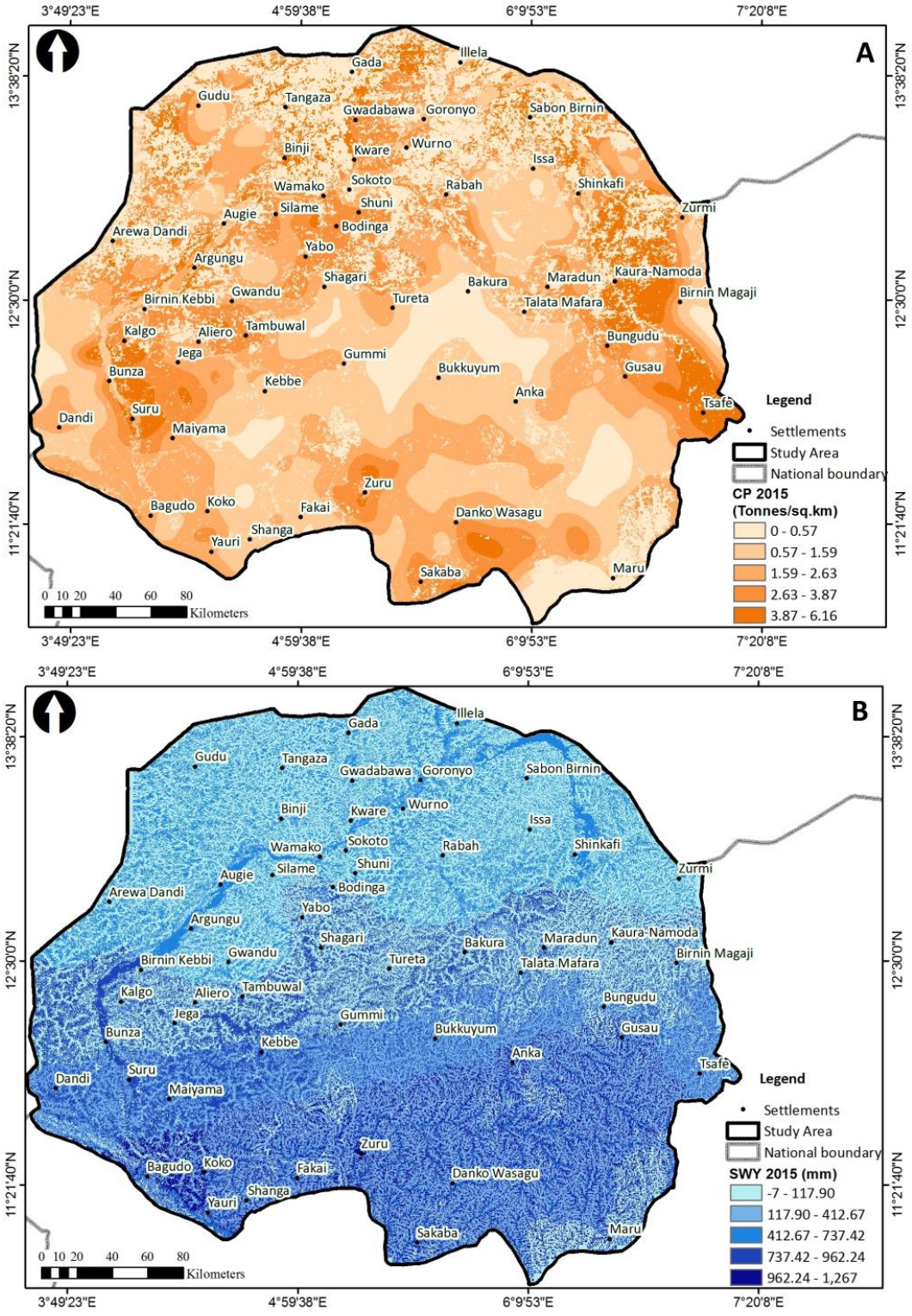


Figure 5. Spatial distribution of crop production (a) and seasonal water yield (b) for the Sokoto-Rima basin in 2015

Table 9: Proportion of SWY status in the study area for the static year 2015

Status/modelled SWY (mm)	Land area (km^2)	Proportion (%)
Very low (0-88.69)	34,127.55	36.30
Low (88.69-320.21)	23,220.20	24.70
Moderate (320.21-599.06)	14,889.05	15.83
High (599.06-813.24)	13,472.80	14.33
Very high (813.24-1155)	8,316.90	8.85
Total	**94,026.5**	**100.00**

Habitat Quality (2015); HQ expresses ecosystem integrity in terms of its intactness and capacity to supply ecosystem services within a specific area coupled with ability to withstand degradation (Terrado et al. 2016). The 2015 HQ of the Sokoto-Rima basin showed that very high status comprises 3.25% (3,057.32 km^2), high status comprises 10.40% (9,778.23 km^2), moderate status comprises 12.97% (12,195.08 km^2) and very low status comprises 73.38% (68,995.87 km^2) (Table 10). This outcome is spatially expressed in Figure 6a, which is dominated by low HQ status. Also, high status was restricted to spots of protected vegetation particularly woodland, showing that level of biodiversity is very low and the possibility of degradation of natural habitats is very high in the study area. Although, the semi-arid ecosystem is often characterized by low grass and tree density, issues of high land clearance for cultivation in the Sokoto-Rima basin has been prevalent for years (Olagunju, 2015), which is a pointer to high level of degradation in HQ detected. Furthermore, this was stated by Oladipo (1995) to be responsible for increasing instances of southward migration of the Sahara Desert which has hampered the level of biodiversity and quality of the protected areas of the northern part of Nigeria including the Sokoto-Rima basin.

Table 10: Proportion of HQ status in the study area for 2015

Status/HQ level	Land Area (km^2)	Proportion (%)
Very high (0.799-1.0)	3,057.32	3.25
High (0.342-0.799)	9,778.23	10.40
Moderate (0.175-0.342)	12,195.08	12.97
Very low (0-0.124)	68,995.87	73.38
Total	**94,026.50**	**100.00**

Nutrient retention ratio (2015); NRR directly measures the extent of nutrient cycling in space and it is a vital approach in explaining the spatial flow of nutrient within a given ecosystem. The nature of NRR as shown in Table 11 states that very high constitutes 6,196.88 km^2 (6.59%), high constitutes 6,104.74 km^2 (6.49%), moderate constitutes 12,744.18 km^2 (13.55%), low constitutes 24,777.04 km^2 (26.35%) and very low constitutes 44,203.66 km^2 (47.01%) of the Sokoto-Rima basin. The low distribution of NRR over space as shown in Figure 6 showed that the quick uptake of nutrient is directly proportional to the dominant typology of land cover in the study area. Spatially nutrients retention observed at edges of river systems showed high potentials while those at the cultivated areas such as wetlands and crop cultivation complexes showed moderate to low statuses thus establishing the classification posited by Salata et al. (2017).

Table 11: Proportion of NRR status in the study area for 2015

NRR Level	Land Area (km^2)	Proportion (%)
Very high (0.8-1.0)	6,196.88	6.59
High (0.6-0.8)	6,104.74	6.49
Moderate (0.4-0.6)	12,744.18	13.55
Low (0.2-0.4)	24,777.04	26.35
Very low (0-0.2)	44,203.66	47.01
Total	**94,026.50**	**100.00**

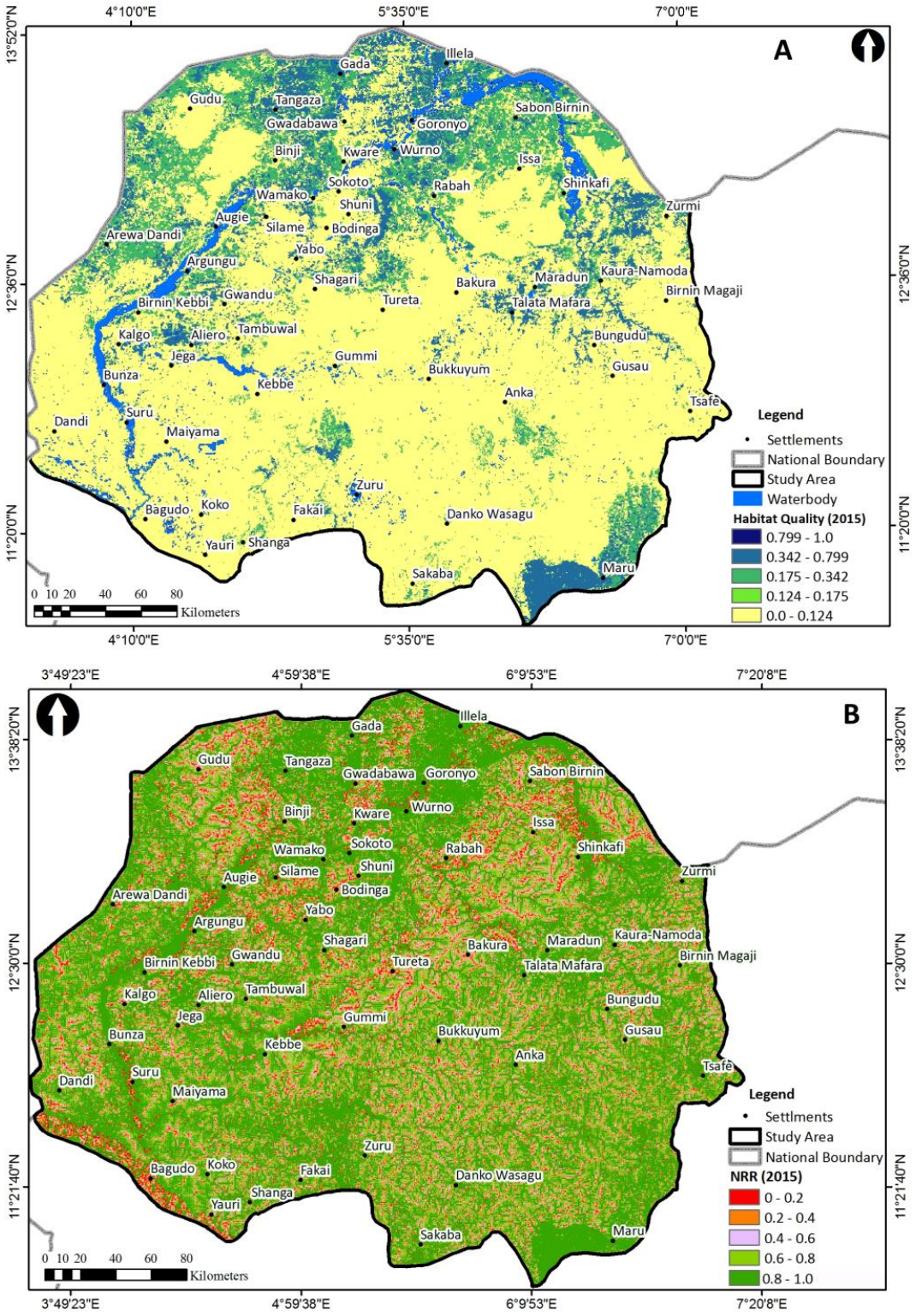


Figure 6. Spatial distribution of habitat quality (a) and nutrient retention ratio (b) for the Sokoto-Rima basin in 2015

3.7 Future Ecosystem Services (2050)

Crop production (2050); By 2050, CP status will remain unchanged, sustaining the average crop yield level with the prior observed five levels of crop yield. Very low status will occupy 23.49% (22,089.40 km^2), low status 32.99% (31,019.28 km^2), moderate 22.03% (20,713.01 km^2), high 14.59% (13,715.10 km^2) and very high status 6.9% (6,489.71 km^2) in the study area. In comparison with the 2015 crop production, crop yield in terms of tonnage per land area will remain averagely low, signifying no systematic change in productivity levels in the Sokoto-Rima basin.

Spatially, Figure 7a shows that CP status in 2050 will show some localized differences compared with 2015 with improvements in the degree of low status in areas such as Anka, Bukkuyum and Shanga in the central, and southern zones of the study area.

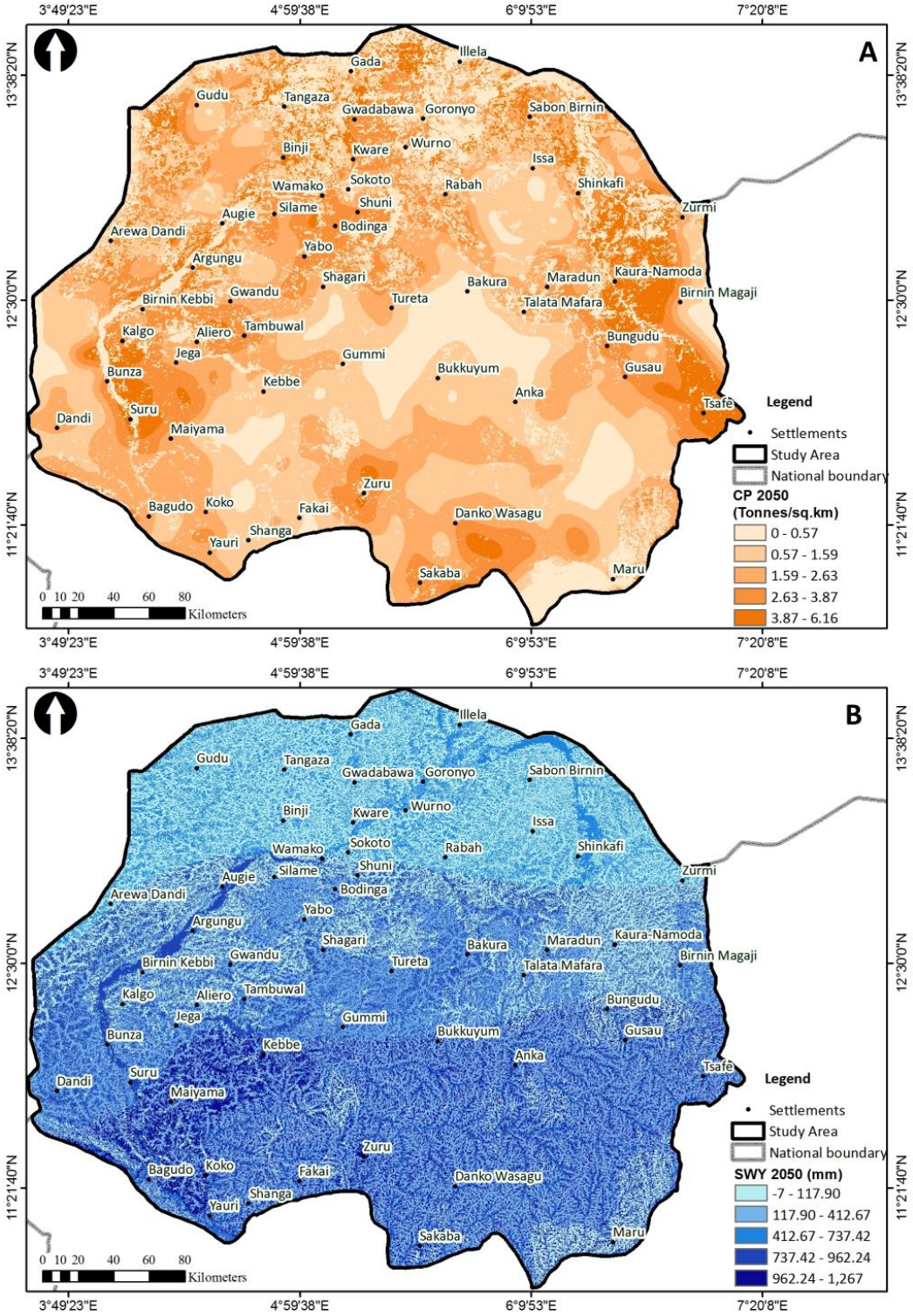


Figure 7. Spatial distribution of simulated crop production (a) and seasonal water yield (b) for the year 2050

Seasonal water yield (2050); The status levels observed in 2015 will be maintained by 2050 with some differences in land area and proportions. As shown in Figure 7, Very low yield will occupy 28.42% (26,718.17 km^2), low 31.61% (29,721.67 km^2), moderate 11.96% (11,243.45 km^2), high 14.43% (13,570.08 km^2) and very high statuses will occupy, 13.58% (12,773.13 km^2) of the landscape correspondingly.

Natural water flow from hydrological sources specifically quick flow (runoff) from rainfall will slightly improve the seasonally available water in the Sokoto-Rima basin. This could be stimulated variations in climatic parameters of the study area. As shown in Figure 7b, improvements will be observed in low, high and very statuses in close to the major rivers, the water reservoirs, and the hilly areas of the eastern axis of the study area. Areas such as Suru, Maiyama, Kebbe, Bunza, Bagudo and Yauri in the southern axis are expected to experience high SWY by 2050.

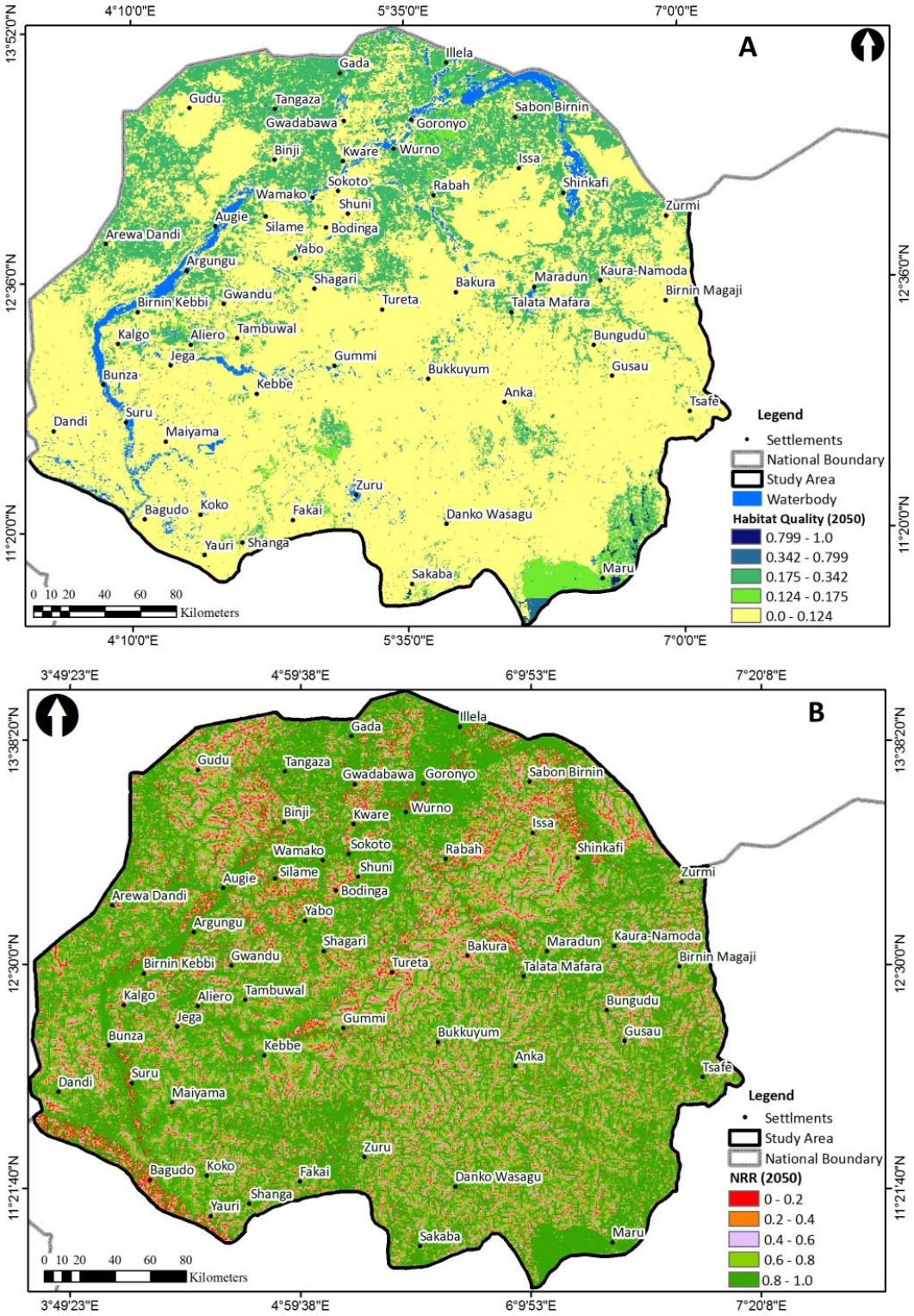


Figure 8. Spatial distribution of simulated habitat quality (a) and nutrient retention ratio (b) for the year 2050

Habitat quality (2050); Slight improvements in habitat quality will surface in 2050 as the HQ levels will increase to five from the four levels observed in 2015. Very high HQ status will occupy 0.12% (114.61 km2), high HQ 0.25% (231.48 km2), moderate HQ 21% (19,741.73 km^2), low HQ 2.15% (2,018.09 km^2) and very low HQ 76.49% (71,920.59 km^2) of the landscape. This shows that some of the high levels will be lost to lower HQ status, an observation of reduction in biodiversity. Figure 8a testifies spatially to this assertion with high biodiversity areas projected in localized communities such as northern areas of Zuru and areas surrounding Maru at the southeastern axis of the Sokoto-Rima basin.

Nutrient retention ratio (2050); Retention of nutrient will be proportionally similar but slightly different in terms of land area, thus sustaining the agrarian cum semi-arid ecosystem pattern. As shown in Figure 8b, very low NRR will amount to 6.58% (6,190.64 km^2) of the landscape, low NRR will constitute 6.51% (6,125.49 km^2), moderate will occupy 13.76% (12,935.24 km^2), high will constitute 26.8% (25,200.32 km^2) and very high 46.34% (43,574.8 km^2) of the study area. Compared to 2015 status, a meagre change is anticipated by the year 2050. This could be related to the linkage of nutrient retention to the natural land surface configuration such as soil characteristics and terrain features which usually remain largely unchanged. However, 1.7% increase in high status could be related to possibility of more water within the basin as depicted as affirmed by the 2050 SWY represents an improvement in available nutrient flow within the Sokoto-Rima basin by the year 2050.

4. Conclusion

Quantifying the status and dynamic patterns of ecosystem services is dependent on spatially explicit models which are inherently reliant on land cover and its associated spatial data. A key aspect of this is predicting future land cover in order to produce future ecosystem services. To simulate and predict future land cover of any location, key spatial drivers spanning natural and anthropogenic activities are essential. In this study, an attempt was made to predict and simulate future land cover of the Sokoto-Rima basin of north-western Nigeria. These predicted and simulated datasets were used as principal data for the simulation of future ecosystem services using ecosystem service models embedded in InVEST software. The result showed that proximate and underlying drivers are important in future land cover prediction. Also, provisioning and regulating ecosystem services drive the functioning of the semi-arid ecosystem of the Sokoto-Rima basin. Correlation analysis of the geo-environmental variables showed that multi-collinearity effects have been eliminated from the data and CA-ANN is a suitable land cover prediction method for simulating future land cover of the study area. Crop production has shown that the Sokoto-Rima basin is indeed a low-producing agrarian ecosystem with average tonnage of crop produced in 2015 and similar pattern will be replicated by 2050. Seasonal water yields are highly variable, fluctuating between low to moderate water quantity required for proper functioning which is a typical trait of a dryland such as the semi-arid Sokoto-Rima basin. Habitat quality however showed that biodiversity will reduce as the years go by, requiring measures to resuscitate the dwindling natural ecological resources of the basin. The crop production model of InVEST provides a coarse global crop output within an area of less heterogeneous landscape and land management. Land management based crop fertilization rates provided by Tarfa et al. (2017) has provided a gap filling merit to address this drawback however engagement of localized crop yields and influence of land tenure systems can be assessed under a climate change scenario to check this drawback. In addition, the quick flow approach adopted from the InVEST's seasonal water yield model is based on surface water which is a determinant factor for sustenance of life for the dryland region. Basin scale hydrology is required to fully capture the pattern of water resource of the Sokoto-Rima basin.

Anthropogenic influence on the Sokoto-Rima basin will increase in the future and the persistence of this trend require checks and balances that will enshrine proper land use policy and sustainable environmental development. Woodland development through the Great Green Wall program can be up-scaled to the level of communities in a synergistic fashion with crop production (Davies, 2017). When adopted, this approach will also revive the dwindling level of the existing protected areas. This will secure existing agrarian system, improve ecological integrity, ensure regulated water availability, and nutrient flow within the natural and anthropogenic ecosystems will improve overtime. Moreover, ecosystem service interaction and integration to biodiversity management within the Sokoto-Rima basin can be pursued both at the watershed and local government levels in the study area.

Acknowledgements

The authors appreciate the kind permission of the following organizations for the granting data access; European Space Agency (ESA) for the land cover data, Japanese Aerospace Exploration Agency (JAXA) for the datasets on terrain, Princeton Climate Analytics (PCA) for the climate data, United States Geological Survey (USGS) for soil, NDVI and evapotranspiration datasets and National Bureau of Statistics of Nigeria (NBS) for the population data. Finally, we greatly acknowledge the efforts of the three anonymous reviewers in improving the quality of paper.

References

Davies, J. (2017). Biodiversity and the great green wall: Managing nature for sustainable development in the sahel. Ouagadougou: IUCN, xiv + 66 pp.

Defourny, P., Bontemps, S., Martin, B., Brockman, C., Fomferra, N., Grit, K., & Krüger, O. (2011). Landcover CCI: Product Specification Document. Year 2 Version 1.3.

Fasona, M. J., Soneye, A. S., Ogunkunle, O. J., Adeaga, O. A., Fashae, O. A., & Abbas, I. I. (2014). Simulating Land-Cover and Land-Use Change in the Savanna Under Present Day and Future Climate Scenarios-A GIS-Based Approach. *Earth Science Research*, *3*(1), 25.

Fasona, M., Omojola, A., Adeaga, O., & Dabi, D. (2007). Aspects of Climate Change and Resource Conflicts in the Nigeria Savannah. In *Report of IPCC/TGICA Expert Meeting on Integrating Analysis of Regional Climate Change and Response Options* (pp. 45-55).

Fu, Q., Li, B., Hou, Y., Bi, X., & Zhang, X. (2017). Effects of land use and climate change on ecosystem services in Central Asia's arid regions: a case study in Altay Prefecture, China. *Science of the Total Environment*, *607*, 633-646.

Geri, F., Amici, V., & Rocchini, D. (2011). Spatially-based accuracy assessment of forestation prediction in a complex Mediterranean landscape. *Applied Geography*, *31*(3), 881-890.

Haines-Young R., & Potschin M. (2018). *Common International Classification of Ecosystem Services* (CICES) V5.1.

Hølleland, H., Skrede, J., & Holmgaard, S. B. (2017). Cultural heritage and ecosystem services: A literature review. *Conservation and ManageMent of arChaeologiCal sites*, *19*(3), 210-237.

Huber, R., Rigling, A., Bebi, P., Brand, F. S., Briner, S., Buttler, A., ... & Lischke, H. (2013). Sustainable land use in mountain regions under global change: synthesis across scales and disciplines. *Ecology and Society*, *18*(3), 36, doi:10.5751/es-05499-180336.

Lambin, E. F., Geist, H. J., & Lepers, E. (2003). Dynamics of land-use and land-cover change in tropical regions. *Annual review of environment and resources*, *28*(1), 205-241.

MEA. (2005a). Ecosystem and Human Well-Being: Synthesis. Millennium Ecosystem Assessment, Island Press, Washington, DC.

MEA. (2005b). Ecosystem and Human Well-Being: Biodiversity Synthesis. Millennium Ecosystem Assessment, World Resources Institute, Washington, DC.

Mueller, N. D., Gerber, J. S., Johnston, M., Ray, D. K., Ramankutty, N., & Foley, J. A. (2012). Closing yield gaps through nutrient and water management. *Nature*, *490*(7419), 254-257.

Oguntunde, P. G., Abiodun, B. J., & Lischeid, G. (2011). Rainfall trends in Nigeria, 1901–2000. *Journal of Hydrology*, *411*(3-4), 207-218.

Oladipo, E. O. (1995). Some statistical characteristics of drought area variations in the savanna region of Nigeria. *Theoretical and Applied climatology*, *50*(3-4), 147-155.

Olagunju, T. E. (2015). Drought, desertification and the Nigerian environment: A review. *Journal of Ecology and the Natural Environment*, *7*(7), 196-209.

Oyebande, L., & Odunuga, S. (2010). Climate change impact on water resources at the transboundary level in West Africa: the cases of the Senegal, Niger and Volta Basins. *The Open Hydrology Journal*, *4*(1), 163-172.

Pérez-Hoyos, A., Rembold, F., Kerdiles, H., & Gallego, J. (2017). Comparison of global land cover datasets for cropland monitoring. *Remote Sensing*, *9*(11), 1118, doi:10.3390/rs9111118.

Roche, P. K., & Campagne, C. S. (2017). From ecosystem integrity to ecosystem condition: a continuity of concepts supporting different aspects of ecosystem sustainability. *Current opinion in environmental sustainability*, *29*, 63-68.

Salata, S., Garnero, G., Barbieri, C. A., & Giaimo, C. (2017). The integration of ecosystem services in planning: An evaluation of the nutrient retention model using InVEST software. *Land*, *6*(3), 48, doi:10.3390/land6030048.

Satya, B. A., Shashi, M., & Deva, P. (2020). Future land use land cover scenario simulation using open source GIS for the city of Warangal, Telangana, India. *Applied Geomatics*, 1-10.

Setten, G., Stenseke, M., & Moen, J. (2012). Ecosystem services and landscape management: three challenges and one plea. *International Journal of Biodiversity Science, Ecosystem Services & Management*, *8*(4), 305-312.

Sharp, R., Tallis, H. T., Ricketts, T., Guerry, A. D., Wood, S. A., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., … & Douglass, J. (2018). InVEST 3.5.0. User's Guide. The Natural Capital Project, Stanford University. University of Minnesota, The Nature Conservancy, and World Wildlife Fund.

Tarfa, B. D., Amapu, I. Y., Daudu, C. K., Arunah, U. L., Shero, I. A., Isah, A. S., Yakubu, A. A., Abdu, N., … & Ugbabe, O. O. (2017). Optimizing fertilizer use within the context of integrated soil fertility management in Nigeria. In C. S. Wortmann, & K. Sones (eds.), Fertilizer Use Optimization in Sub-Saharan Africa (pp. 148–163). Nairobi: CAB International.

Terrado, M., Sabater, S., Chaplin-Kramer, B., Mandle, L., Ziv, G., & Acuña, V. (2016). Model development for the assessment of terrestrial and aquatic habitat quality in conservation planning. *Science of the total environment*, *540*(1), 63-70.

Wangai, P. W., Burkhard, B., & Müller, F. (2016). A review of studies on ecosystem services in Africa. *International journal of sustainable built environment*, *5*(2), 225-245.

Yang, S., Zhao, W., Liu, Y., Wang, S., Wang, J., & Zhai, R. (2018). Influence of land use change on the ecosystem service trade-offs in the ecological restoration area: Dynamics and scenarios in the Yanhe watershed, China. *Science of the total environment*, *644*, 556-566.

2020
Turkish Journal of Remote Sensing and GIS

Turkish Journal of Remote Sensing and GIS
Türk Uzaktan Algılama ve CBS Dergisi

Turk J Remote Sens GIS, September 2020, 1(2): 125-136
Journal homepage: http://www.dergipark.org.tr/en/pub/rsgis

Modelling of Temporal Query of 3D Legal Objects in Turkey

Türkiye'de 3B Yasal Nesnelerin Zamansal Sorgulamasının Modellenmesi

Mehmet Alkan[1]*, Hicret Gürsoy Sürmeneli[1]

[1]*Yıldız Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, 34349, Istanbul/Turkey.*

ORIGINAL PAPER

***Corresponding author:**
Mehmet Alkan
alkan@yıldız.edu.tr

doi:

Article history:
Received: 27.04.2020
Accepted: 04.09.2020
Published: 30.09.2020

Abstract

Turkey's existing cadastral system consists of two primary components. These are title recordings and cadastral recordings. Immovables are registered to the title in three ways: land, independent section and permanent rights. The third dimension on land and real estate is registered in title with right, restriction and responsibility (RRR is as permanent rights). On the other hand, the Turkish cadastral system allows real estate to be registered temporarily. A real estate property is registered to the existing cadastral system with transaction date and time. Thus, any transaction can be monitored over time temporarily. Therefore, the results of the study contribute to the literature by examining the scope of the Turkish legal system with a 3D cadastre. Besides, the RRR legal system with time aspect in the Turkish cadastral system is modelled based on the ISO 19152:2012 LADM. In this context, the paper consists of five main chapters. In section 1, the literature on 3D cadastre and LADM is given. The methodology of the study is described in section 2. Chapter 3 discusses 3D RRR scope of the evaluations regarding the legal objects in Turkey. Besides, time data and RRR are incorporated into Chapter 3. Sections 4 and 5 include discussion and recommendations.

***Keywords**: 3D RRR, Temporal query, LADM, Land title*

Özet

Türkiye'nin mevcut kadastro sistemi iki ana bileşenden oluşmaktadır. Bunlar tapu kayıtları ve kadastro kayıtlarıdır. Taşınmazlar tapuya üç şekilde kaydedilir: arazi, bağımsız bölüm ve sürekli haklar. Arazi ve gayrimenkul üzerine üçüncü boyut, hak, sınırlama ve sorumluluk ile tapuda tescil edilir (RRR, kalıcı haklardır). Öte yandan, Türk kadastro sistemi gayrimenkulün geçici olarak kayıt altına alınmasına izin vermektedir. Bir gayrimenkul, işlem tarihi ve saati ile mevcut kadastro sistemine kaydedilir. Böylelikle herhangi bir işlemin zaman içinde gerçekleşmesi geçici olarak izlenebilir. Dolayısıyla, çalışma sonuçları Türk hukuk sistemi kapsamını 3 boyutlu kadastro ile inceleyerek literatüre katkı sağlamaktadır. Ayrıca, Türk kadastro sisteminde zaman yönü olan RRR hukuk sistemi, ISO 19152: 2012 AİTM'ye göre modellenmiştir. Bu bağlamda makale beş ana bölümden oluşmaktadır. Bölüm 1'de 3 boyutlu kadastro ve AİTM ile ilgili literatür verilmektedir. Çalışmanın metodolojisi 2. bölümde açıklanmıştır. Bölüm 3 Türkiye yasal nesnelerine ilişkin değerlendirmelerin 3D RRR kapsamı tartışılmıştır. Ayrıca, zaman verileri ve RRR bölüm 3'e entegre edilmiştir. Bölüm 4 ve 5, tartışma ve tavsiyeleri içermektedir.

***Anahtar kelimeler**: 3D RRR, Zamansal sorgulama, AİTM, Tapu kaydı*

1. Introduction

Cadastral studies are seen as the cornerstone of the land administration system. According to many researchers, cadastral maps should provide complete information to record RRR on the cadastral parcel within the scope of 3D cadastre (Kaufmann and Steudler, 1998; Stoter and van Oosterom, 2006). However, most of the countries create cadastral maps based on 2D land parcels for land administration systems (Döner et al. 2011; Ho et al. 2015; Rajabifard et al. 2018; Alkan et al. 2018; Surmeneli et al. 2020). Thus, RRR on the land cannot be adequately represented.

Therefore, the need for three dimensional cadastral data increases. As a result of this need, the demand for the 3D cadastre representing the real world has come up. International studies on future cadastral systems have been investigated e.g Cadastre 2014 (Steudler, 2014) and 2034 (ICSM, 2015). Within the scope of Cadastre 2034 vision, the cadastral system of the future should include all RRR related to real estates with advanced policies, standards and models.

Over the last decade 3D Cadastral systems and immovable properties have intensively studied by institutions and scientists (Karki et al. 2010; Aien et al. 2011; Döner et al. 2011; Guo et al. 2011; van Oosterom, 2013; Paasch and Paulsson, 2014; Kitsakis et al. 2016; Alkan et al. 2020; Surmeneli et al. 2020). It depends on these studies, 3D cadastral studied could listed in the five main headings in FIG publication (van Oosterom, 2018).

- Legal foundations for 3D Cadastre
- Initial Registration of 3D Parcels
- 3D Cadastral Information Modelling
- 3D Spatial DBMS for 3D Cadastres
- Visualisation and New Opportunities

Exploring the examples from 3D Cadastre worldwide, it has been found many similarities. The proposed solutions are similar, even different aspects of 3D features are taken into account, based on the structure of their cadastral systems, the types of recorded objects, etc. There are regulations in Australia (Aien, 2013) and Canada (Stoter and van Oosterom, 2006), which contain detailed explanations on how to measure and register a parcel in 3D. The Netherlands changed the Civil Law on Cadastre in 2007, thus redefining the ownership of land, which lead to a registration of 3D objects determined by the law independently from the parcel (Wakker et al. 2003). In the Netherlands and Switzerland, there is a separate Line cadastre for the recording of underground networks (Wakker et al. 2003; Steudler, 2015). In Norway, the registration of underground structures is not mandatory and is made optional (Herdlevær, 2018). In Switzerland, studies have been carried out to convert the existing cadastral database into a 3D state. For the Swiss Cadastre, a comprehensive plan covering the transition to 3D cadastre was prepared (Steudler, 2015). In some countries, the whole cadastre is already digital. However, there are still countries that are continuously working in this direction, such as Greece (van Oosterom et al. 2018).

In this paper, our motivation is to support the regulation and analysis of 3D land rights, restrictions, and responsibilities for the Turkish Cadastre System based on the international standards with the LADM within ISO 19152 aspect of time. In the current research, we share our main output: design 3D legal model with time aspect using LADM for the Turkish cadastral system.

1.1 3D RRR and LADM

Within the Cadastre 2034 (ICSM, 2015) vision is recommended that the cadastre should be considered and registered in 3 dimensions by emphasising that the ideal cadastral system should show the entire legal status of the land including public rights, responsibilities and restrictions. Lemmen and van Oosterom (2013) explain three specialisation classes of right, restriction and responsibility: 'A "right" is an action, activity or class of actions that a system participant may perform on or using an associated resource. Examples are ownership right, tenancy right, possession, customary right or an informal right. A "restriction" is a formal or informal entitlement to refrain from doing something; e.g. it is not allowed to build within 200 meters of a fuel station, or servitude or a mortgage as a restriction to the ownership right. A "responsibility" is a formal or informal obligation to do something; e.g. the responsibility to clean a ditch, to keep a snow-free pavement or to remove icicles from the roof during winter or to maintain a monument.

The three-dimensional cadastre is a cadastre that provides information on the rights, responsibilities and restrictions on registration and not only on the parcel but also the 3D possessive units (Alkan et al. 2020). In this context, the Basic Model of Land Administration constitutes a basic class in order to define the rights, responsibilities and constraints concerning the 3rd dimension of the real estate. With this class structure, the management of the rights, responsibilities and restrictions that may occur on the spatial unit will be ensured. The main starting point of the Land Administration Domain Model (LADM) is to establish a common ontology for RRR affecting the land administration and its geometric components. Thus, it will enable communication between related parties within a country or between different countries (van Oosterom et al. 2006; Lemmen et al. 2015). The LADM is developed in line with the Cadastre 2014 (Stuedler, 2014) vision and complies with international ISO and OGC standards (Lemmen et al. 2009; Lemmen et al. 2011; Tjia and Coetzee, 2013). Besides, it has been conducted in the studies showing the compatibility of LADM with INSPIRE (Alkan and Polat, 2017; Alkan et al. 2020; Surmeneli et al. 2020).

LADM has three main packages and one sub-package. These are LA_Party (Party package), LA_AdministrativePackage (Management package) and LA_SpatialUnitPackage (Spatial Unit package) and LA_SurveyingAndRepresentation (Figure 1).

LA_VersionedObject is a superclass. Classes LA_Party, LA_GroupParty, LA_PartyMember, LA_Mortgage, LA_RRR, LA_BAUnit, LA_SpatialUnit, LA_SpatialUnitGroup, LA_RequiredRelationshipSpatialUnit, LA_RequiredRelationshipBAUnit, LA_Level, LA_- BoundaryFaceString, LA_BoundaryFace, and LA_Point are all subclasses from LA_VersionedObject. LA_VersionedObject has attributes for the management of versions of objects in LADM and for the management of data quality.

It also supports the time component of the Land Administration Basic Model. The most important feature of the model is a flexible model and can be expanded within specified standards. It is possible to associate with external classes such as Valuation, Address, and the Land cover as required by the model feature.

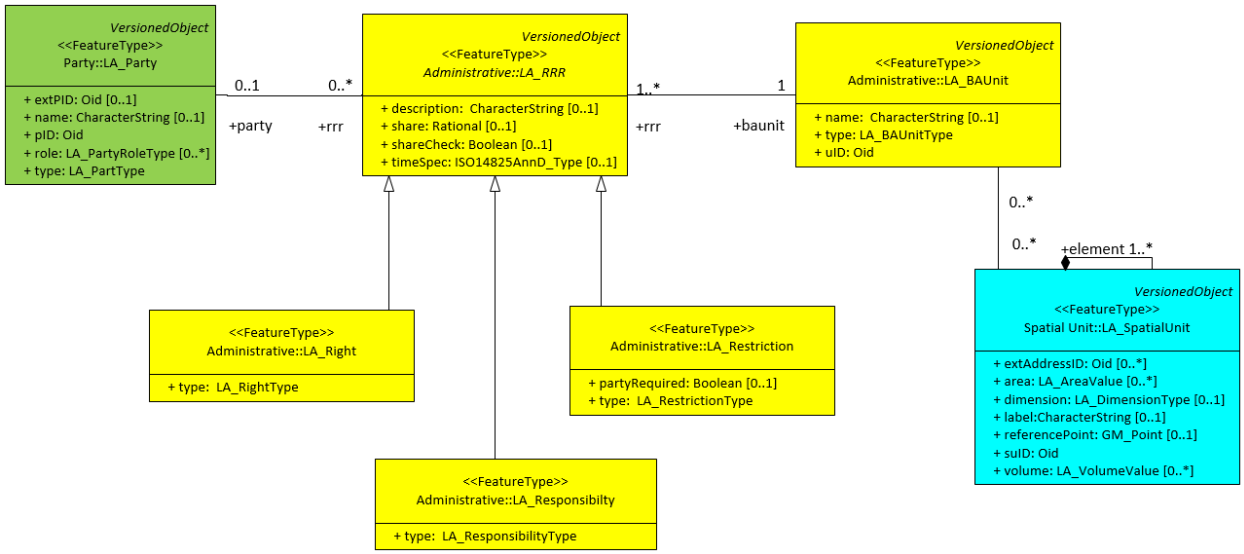


Figure 1. LADM core classes (Lemmen et al. 2015)

2. Methodology

This article primarily explores how existing Turkish cadastral data models address 3D RRR problems. Then he proposes a conceptual 3D cadastral data model as a solution that can provide 3D data. The 3D RRR model represents 3D legal objects. In this context, the 3D RRR data model is equipped with the Turkish cadastral system data model concepts. First, it facilitates the modelling of all existing interests as legal objects to develop an experimental 3D RRR data model for Turkish cadastral systems. The system methodology that invokes the design and development approach as shown in Figure 2. Within the scope of the methodology, first of all, 3D RRRs in the Turkish cadastral system were determined.

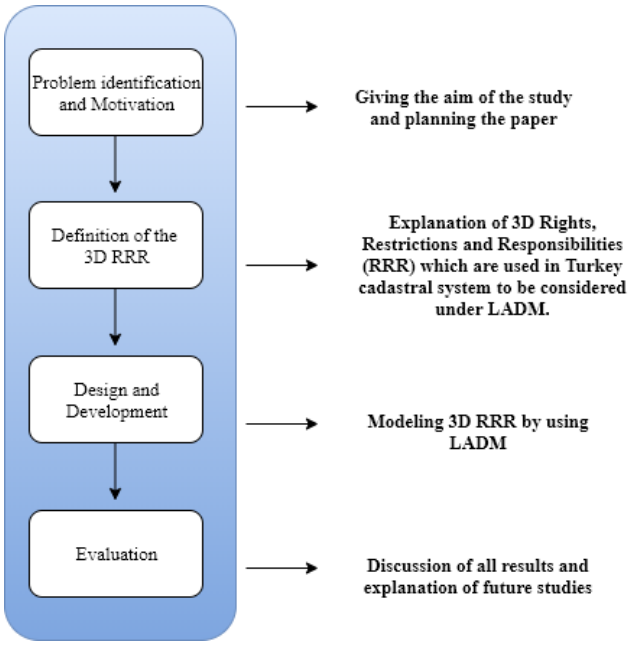


Figure 2. Methodology

3. Evaluation of Turkey Legal Objects Scope of 3D RRR

Today, in many cadastral systems, 2D cadastral parcels are the basis for registration of rights, which are the legal status of the buildings and the rights of the objects registered. Besides, above and below the rights of the objects and buildings are determined by rights related to registrations. The most comprehensive right a person can have is the right to property. Within the scope of the right to property, the person is free to use what anybody owns while observing the rights of others and taking into account the limitations of laws, rules, and non-written laws (Alkan et al. 2020).

The third dimension in the constitution is not sufficiently detailed in the explanation of the vertical rights. There are several rights set out in our constitution regarding the implementation of these rights regarding the third dimension. For example, the easement right is a right. Also, the easement right restricts the rights of the immovable in favour of another immovable. It is also a type of right that determines the limits on how to exercise the immovable rights. The types of easement in kind are listed as the right of access, superficies (right of construction), resource rights, and other easements. In the current cadastral system, the 3D spatial data related to the position where the rights are applied are not included in the cadastral register.

Furthermore, since the data of the objects located under the parcel are not available in the cadastre, the relationship of these objects with the parcels cannot be analysed in the current system. Another obstacle is the lack of representation of the 3D position within the condominium in the cadastral system. Another problem in the present system is that the parcels affected by any object positioned below or above the land surface cannot be questioned as a whole. Besides, the parcels that are transacted for a certain period (for instance, during a sale) cannot be queried as a whole. For example, parcels that are processed between two dates cannot be queried from the system. For each parcel, that transaction within a certain time interval can be displayed in the existing cadastral system. When the studies that were done worldwide and in Turkey were examined, the necessity for 2D cadastral operations to be replaced by 3D parcel object became clear. Besides, the cadastral operations on the 3D parcel object are emphasised. Cadastre 2014 and 2034 vision approaches, which are accepted internationally and constitute the basis for the cadastral systems, bring a different perspective to the cadastral system and become the main component of the land administration system (van Oosterom et al. 2018).

3.1. 3D RRR in Turkey

3D cadastre should have all the rights available in 2D cadastre. All rights and interests in the land should be recorded in 3D cadastre with the integration of temporal data. In this respect, according to the Turkish Civil Code, the real rights are divided into two as property rights and limited real rights.

The right to property is the right of real or legal persons to make all kinds of savings such as using, buying, selling, renting and lending that property on a real estate. According to the owner of the right to property:

- Single-person ownership
- Association ownership
- Cooperation ownership
- Common ownership

It is the state of being able to make all kinds of transactions on the real estate. The person has the Property right and Limited Real Rights on the real estate. The limited real rights are divided into two as Mortgage and Easement.

The right to the easement is a limited real right, which allows using an immovable property for a special reason. According to civil law, easement rights, personal easement, usufruct, residence right, superficies, resource right and other easement rights can be listed.

The mortgage is the right of the right holder to guarantee a receivable and to sell the property through forced execution and to ensure that it will receive it if the debtor does not fulfil his debt. The mortgage has three categories; mortgage, mortgage certificate, land charge note.

It is the part where the information is restricting the use of limited real rights in the land register. These restrictions can be listed as representations, right and liabilities, annotations, mortgage and easement. There is a “representations” column on the land register page for each property. The information in this column is called a “representation.” In general terms, a representation is a process of formalizing some issues regarding the actual and legal status of the real estate. The representations are recorded in the land register with the subject, which is a transaction, the page number of the land register, and the document number. The RealEstateID, Share and Area related to real estate are included in the representations class. To query temporal data are RegisterDate, RegisterTime, DeleteDate, and DeleteTime in the representation class.

Rights and Liability class is the part where rights such as easement, usufruct, right of access, and timeshare property rights are registered to the land register. It is registered in its field as representations.

The annotation process is part of the restrictions for any real estate. The annotation process is registered in the area of the land register as in the representations process.

The mortgage is both a type of right and restriction. Some rights and restrictions may overlap, such as the mortgage.

These are the obligations that an interest holder must fulfil on the real estate. These obligations include the tax on the real estate, maintenance, and repair according to the type of real estate, or regular payments and payment of the easement rights related to the real estate. There may be at least one or more types of responsibility on an immovable.

A logical data model has been created in line with all this information. It has been determined which data will be used in the designed model. Tables 1 and 2 have been created accordingly.

Table 1 shows the comparative representation of the data and classes to be used in the model based on LADM.

Table 1. Comparison of 3D RRR used in Turkish land registration system with LADM

Turkey Legislation Profile	LADM	Turkey Conceptual Data Model Class
Undefined	LA_RRR	TR_RRR
Right	LA_Right	TR_Right
LimitedRealRight	LA_Right	TR_LimitedRealRight
RealRight	LA_Right	TR_RealRight
Mortgage	LA_Mortgage	TR_Mortgage
Easement	LA_Right	TR_Easement
RightofLandChanges	LA_Right	TR_RightofLandChanges
Restriction	LA_Restriction	TR_Restriction
RightAndLiability	LA_Restriction	TR_RightAndLiability
Annotations	LA_Restriction	TR_Annotations
Representation	LA_Restriction	TR_Representation
Responsibility	LA_Responsibility	TR_Responsibility

Legal procedures used in current cadastral systems in Turkey are classified under RRR that are classes of LADM (Table 2).

Table 2. The legal process used in Turkish cadastral system is classified under RRR

Right	Restriction	Responsibility
Ownership	ZoningImplementation	RealEstateTax
Coownership	BanOnConstruction	IrrigationCanalsMaintenance
UsageRight	ProhibitedMilitaryZone	HistoricalArtifactsMaintenance
RentalRight	FillingArea	IncumbranceOnRealEstate
RightOfAccess	ZoningStatus	
Pasture	Mortgage	
Usufruct	Easement	
ResidanceRight	Coastline	
Superficies	CulturalProperty	
TimesharePropertyRight	HistoricalMonumentMaintenance	
Preemtive	Forest	
ChargeonLand	RiskArea	
InformalResidance	ArchaeologicalSite	
ShootingDrillRight	Glebe	
PlayFieldsRight	ConstructionFacilityBan	
WaterWayRight	CultivatedArea	
AgriculturalActiviteRight	MiningAnnotations	
TradinationalRight	22/a Implementation	
Mortgage	LandConsolidation	
MortgageCertificate	Expropriation	
Resale	UrbanTransformation	
RentalAndSaleCondition	Mortgage	
RetrievalAndSaleCondition	Annotations	
	Representation	
	Easement	
	RightOfAccess	
	TimesharePropertyRight	
	Usufruct	
	Lien	
	Sequestration	
	LandRental	
	ContractToSell	
	RefusalContract	
	PersonalRight	
	ConstructionRightInReturnForFlat	
	PreeptionContract	
	RestrictionOfRightAlienation	
	BanningOfRightAlienation	
	Mortgage	
	MortgageCertificate	
	Resale	
	RentalAndSaleCondition	
	RetrievalAndSaleCondition	

3.2. The time aspect of 3D RRR in Turkey

The boundaries, geometry, and owner information of cadastral data change over time. In this case, changes in the cadastral systems should be followed in temporal (Alkan, 2005). The integration of time data into cadastral systems will make it easier to keep up with changes that occur over time. The Turkish cadastral system allows real estate to be recorded temporarily. A real estate property is registered in the existing cadastral system with originating right in the type of registration date and registration time. Thus, the realisation of any transaction over time can be monitored in temporarily. In the Turkish cadastral system, cadastral data is recorded temporarily in the title. Cadastral operations performed on a cadastral object are seen in the paper-based cadastral system. Then, the cadastral objects were transformed from a paper-based system to a CAD-based system.

So, the cadastral processes performed on a cadastral object can be questioned temporarily in passing the cadastral data to the CAD system. Thus, the whole process can be displayed for only one cadastral object. However, the deficiency of the system is its inability to query for more than one cadastral object within a certain period of time. Figure 4 shows how the transactions occurring in real estate are recorded temporarily. There may be more than one transaction on immovable property. The temporal expressions of these processes were explained by van Oostreom et al. (2006). Temporal values can be disjoint, touch, overlap, inclusive and equal (Figure 3).

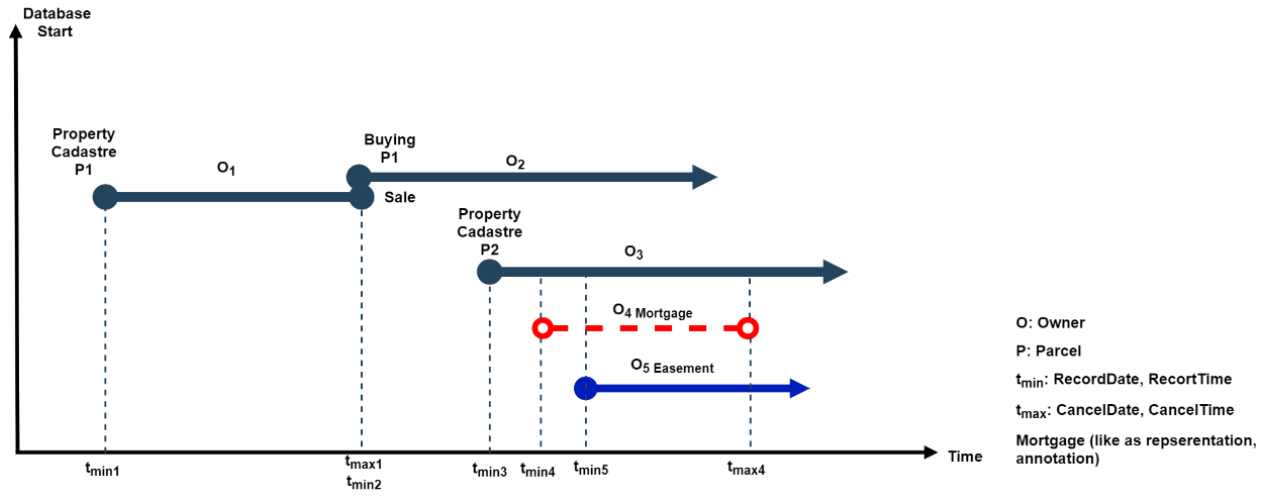


Figure 3. Time registration of immovable properties in Turkish cadastral system

Thus, it is possible to examine the status of the cadastral object or group for a certain period. It will monitor these changes in the cadastral system, including where, when, and how the changes occurred. What kind of changes occur and the reasons for the changes. Figure 4 shows in which operations the cadastral objects occur over time and their final status.

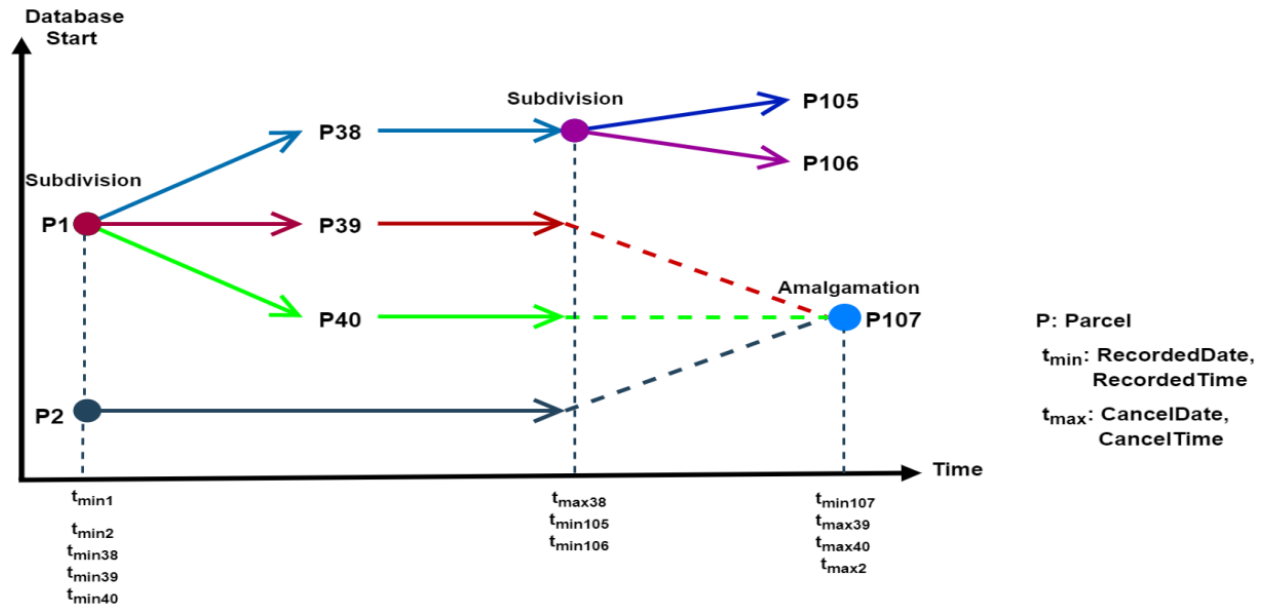


Figure 4. Temporal monitoring of transactions on a real estate

4. 3D RRR Model Scope of LADM

The TR_RRR package is based on the LADM. TR_RRR is an abstract class in which the rights, restrictions and responsibilities set out in the Civil Code are represented to represent the 3rd dimension. The TR_RRR package has three sub-classes TR_Restriction, TR_Responsibility and TR_Right.

Figure 5 shows the TR_Right class. Freehold rights according to civil law, including applicable in Turkey and limited real rights are divided into two classes, and it is divided into two subcategories, namely easement and mortgage. Information regarding the real estate that is a mortgage is provided with the RealEstateID attribute.

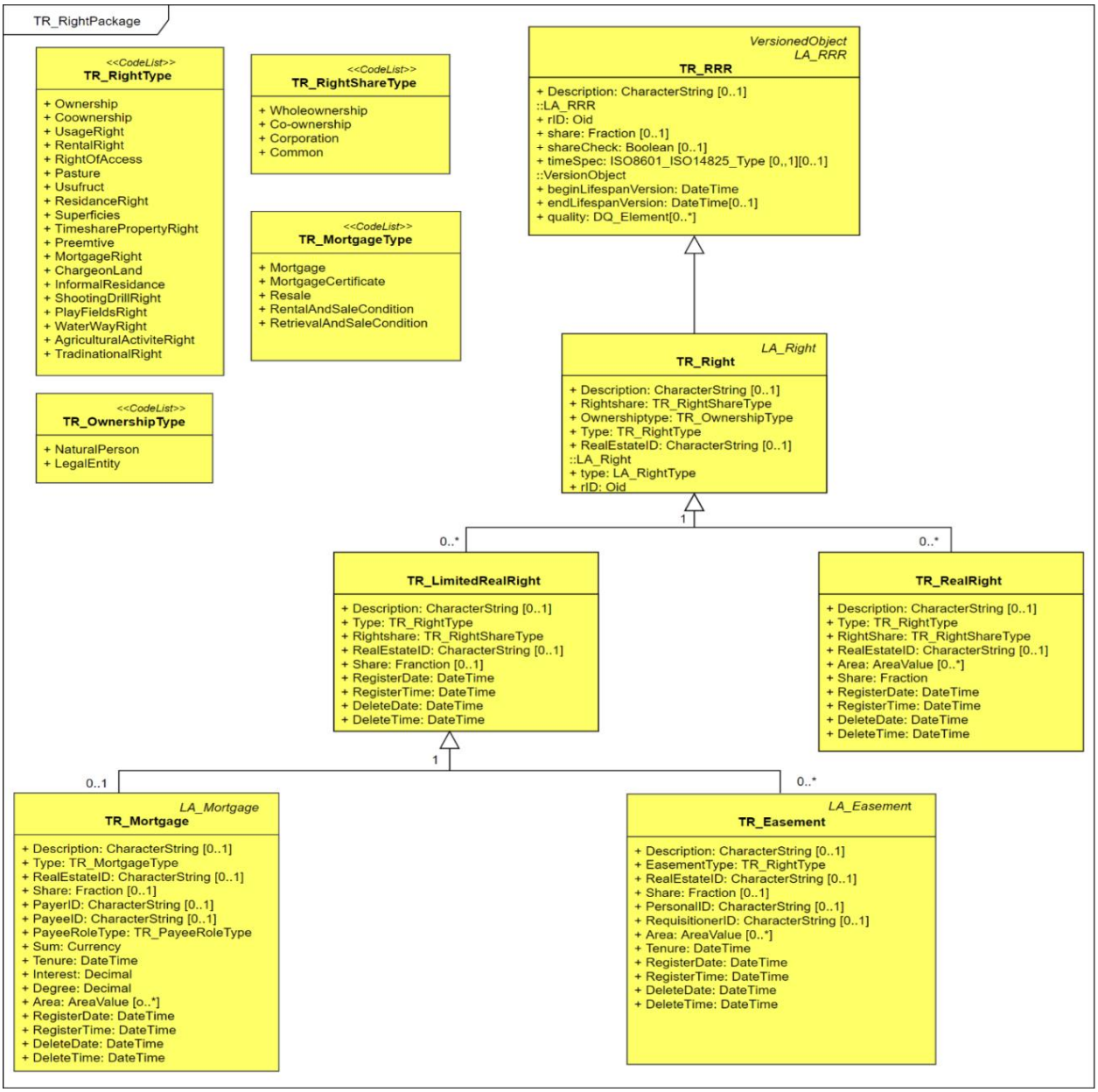


Figure 5. TR_Right class

The information of the debtor and the creditor can be queried with the attributes PayerID and PayeeID. According to the civil law, the amount of mortgage that is not certain or can be changed, an individual mortgage degree is placed and maintains its order regardless of the changes that will occur after the registration. According to this statement, the attribute of the Degree and Sum are in the mortgage class. At the same time, the amount of mortgage, duration, and interest information can be queried from the system. Lastly, information on the RegistrationDate, RegistrationTime, DeleteDate, and DeleteTime attributes are available when the mortgage entitlement starts and ends. The easement right is a type of right that gives the right holder the right to use and benefit from that real estate. The type of easement right can be one of the rights types specified in the RightType code list. Information about the real estate established in the easement can be queried with the RealEstateID, Share and Area attribute. The information of the real estate's owner can be determined by the PersonalID, the information of the person requesting the easement to be established by the RequestonerID attributes.

Finally, the RegistrationDate, RegistrationTime, DeleteDate and DeleteTime attributes are available for information on when the right easement entitlement starts and ends. General information about the right of the easement can be found in the description attribute. Since there may be no or more than one rights type related to a real estate, the relationship types 0 .. * (0-many) is determined.

Figure 6 shows the TR_Restriction class. The TR_Restriction class consists of four subclasses. These are TR_RightAndLiability, TR_Mortgage, TR_Annotations, and TR_Representation. The representations are recorded in the land registry with the subjects which are a transaction, the page number of the land registry and the document number. The RealEstateID, Share and Area related to real estate are included in the TR_Representations class. The Representation CodeList shows that what types of representations might be. Since there may be no or more than one representation related to a real estate, the type of relationship is determined as 0 .. * (0-many). The mortgage is both a type of right and restriction. Some rights and restrictions may overlap, such as the mortgage. If the mortgage expires, the information in the relevant line is overwritten, and the date and document number are written. As a mortgage can be established on real estate, the type of relationship is selected as 0..1 (0-one). Fig. 10 shows that all classes attribute in the Restrictions package.

TR_RightsAndLiability class is the section where the rights such as easement rights, usufruct rights, right of passage, and right of ownership are registered to the land registry (Alkan et al. 2020). The attributes required for registration to the land registry (such as date, document number, transaction description) are defined in the classes. Type of Right and Liability is in the Restrictions CodeList. 0 .. * (0-many) relationship type is selected for Rights and Liability.

TR_Annotation class is the part of the restrictions for any real estate. The annotation process is registered in the area of the land registry as in the representations process. Annotation types can be listed in Annotation CodeList. Relationship type is defined as 0 .. * (0-many).

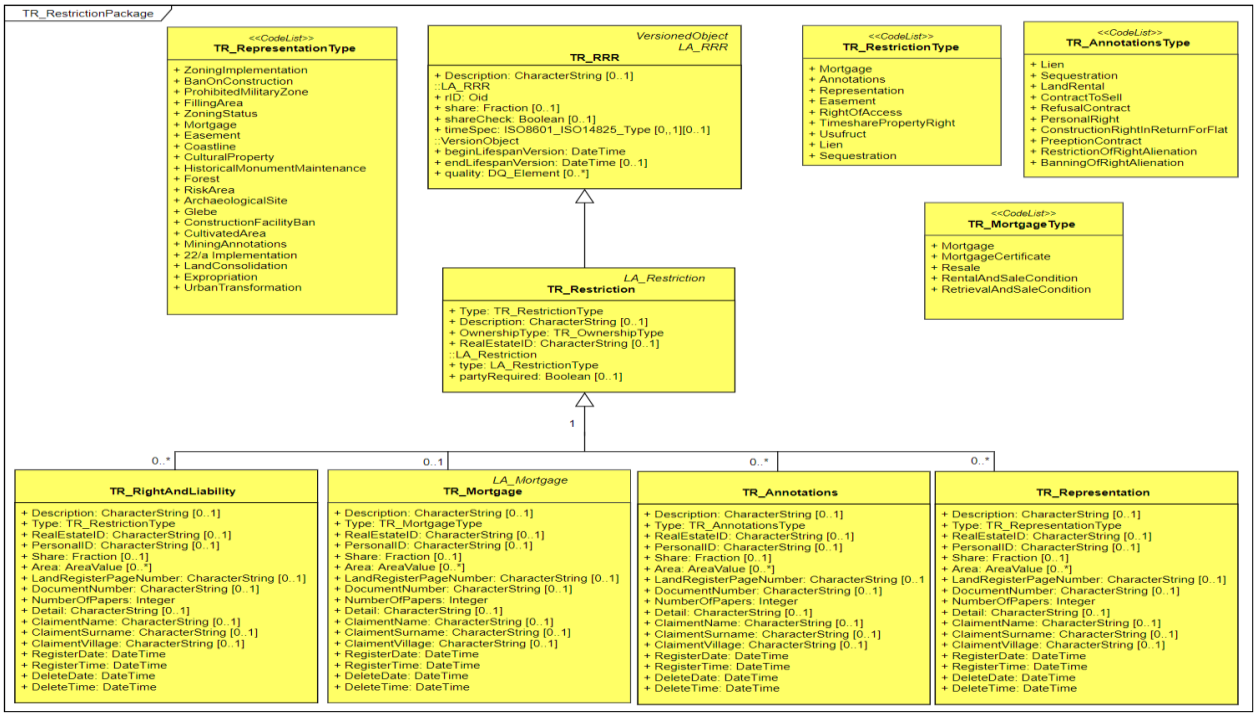


Figure 6. TR_Restriction class

In figure 7, the TR_Responsibility class is represented. It is the class in which a person's obligations to be fulfilled are represented. These obligations include the tax on the real estate, maintenance, and repair according to the type of real estate, or regular payments and payment of the easement rights related to the real estate. There may be at least one or more types of responsibility on an immovable. Therefore, the relationship type is determined as 1 .. * (one to many).

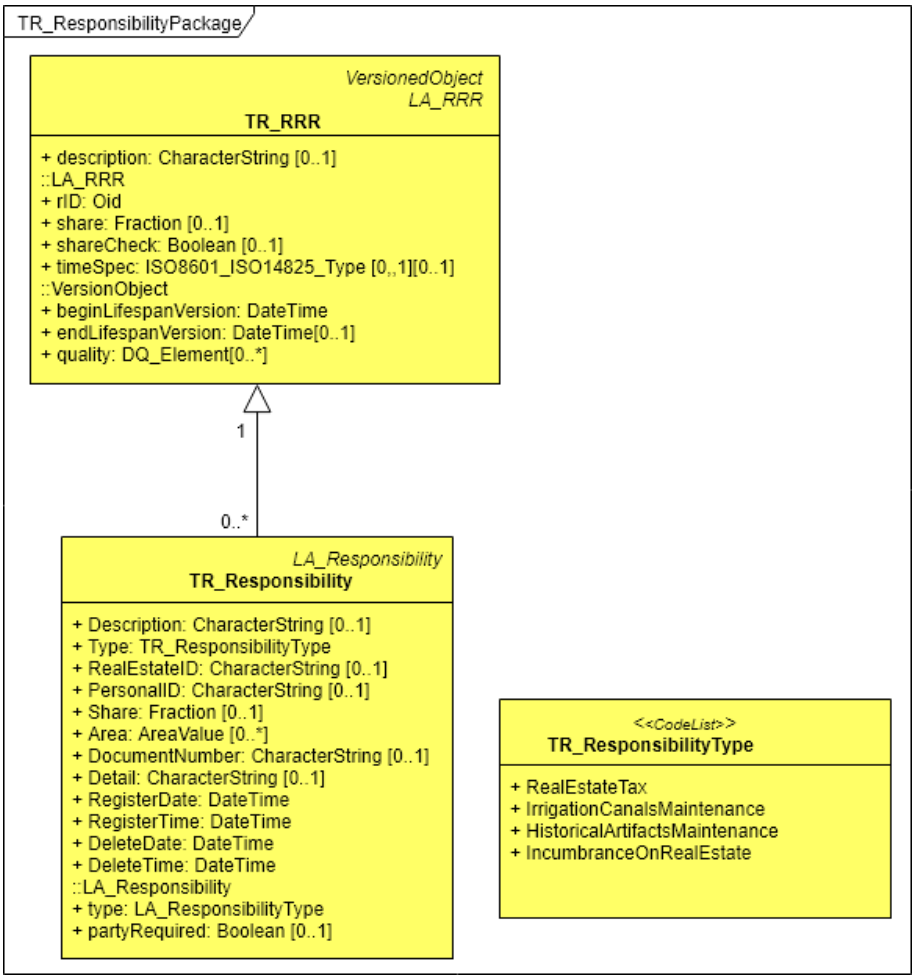


Figure 7. TR_Responsibility class

5. Results and Discussion

In the last two decades, modern technology has been developing with the 2D cadastre no longer responds adequately to the property needs. In this context, international standards as ISO 19152, LADM have been defined in order to develop 3D cadastre. Besides, international standards have been carried out in many countries for 3D cadastral systems. It is observed that in the context of an insufficient number of scientific studies on 3D cadastre in Turkey. When our study examines it, these studies are considered to be insufficient new cadastral models based on international standards for Turkey. From these academic studies, it can be categorised as direct cadastral studies and indirectly as cadastral studies. The studies carried out within the scope of 3D cadastre are generally aimed at analysing the cadastral situation of the country, examining the use of easement rights and legal regulations arising from the third dimension and presenting sample database designs for local cadastral installation.

In this study, the main objective is to contribute to the country profile of the 3D cadastre in Turkey. On the other hand, they are matching the title of the land registry with the LADM. Besides, the existing land registration system in Turkey was discussed in this paper. 3D RRR following the legal regulations in the title system has been determined with which attributes. After determining the required data and data types, 3D RRR is modelled based on LADM.

ISO 19152 LADM is an important step in terms of standardisation and communication of cadastre and land management systems all over the world. With this regards, it is essential to use of ISO standards for Turkish cadastral systems which the management of temporal data and 3D data/rights/boundaries will be significantly facilitated. For this reason, it is assumed that LADM is considered as a suitable template for analysing data requirements and international standards to design a model that will meet the requirements in the land registry and cadastre data. Also, the scope of the paper, RRR package, classes, properties, relations and methodology of LADM model elements were analysed and adapted to our model. Finally, in this study explains 3D cadastral systems with the international standards using compatible or not for Turkey.

6. Conclusion and Recommendations

Firstly, within the scope of this research, the 3D cadastral systems carried out in some of the countries are examined in the introduction. These studies are in line with that recorded in Turkey. Later, cadastral objects and 3D rights, restrictions, and responsibilities registered under legal regulations in the cadastral system and time data allowing for the temporal inquiry were identified. LADM was selected for the study as it is an ISO standard and suitable for the representative of the Turkish cadastral system. Then, logical and conceptual data models for 3D cadastral data were developed. The similarities and differences between the designed model and LADM have been identified and showed Table 1, and 3D RRR are modelled based on LADM.

The 3D cadastral systems have advantages and disadvantages that are determined in this study. It is also could be listed as follows. Firstly, to be held in Turkey in the future will contribute to academic studies related to 3D cadastre. Secondly, it will contribute to the cadastral studies which are carried out institutionally. International contributions are the introduction of the national profile of the Turkish cadastral system in the process of 3D cadastral transformation. At the same time, the use of LADM will not only provide interoperability but will also provide a common language for the promotion of the Turkish cadastral system on international platforms. The disadvantages are that the study remains only as a conceptual model.

As a result, this study determined the three-dimensional representation RRR based on LADM and international standards. Also, it is determined that the current Turkish cadastral system. On the other hand, the paper designed data organisation are compatible with a data model based on LADM. Finally, this study contributes to developing a national standard for 3D cadastral system. The result is a developed model which is of value for General Directorate of Land Registry and Cadastre and can be used as a basis of a 3D national data standard. Furthermore, it can contribute to the establishment of spatial data infrastructure in Turkey and can be internationally recognised.

References

Aien, A., Rajabifard, A., Kalantari, M., & Williamson, I. (2011, May). Aspects of 3D cadastre: a case study in Victoria. In *FIG Working Week 2011* (pp. 1-15). Marrakech, Morocco. FIG.

Aien, A. (2013). *3D cadastral data modelling* (Doctoral dissertation), University of Melbourne, Victoria, Australia.

Alkan, M. (2005). *Temporal geographic information systems design for land registry and cadastral data* (PhD Thesis), Karadeniz Technical University, Trabzon, Turkey.

Alkan, M., & Polat, Z. A. (2017). Design and development of LADM-based infrastructure for Turkey. *Survey review*, *49*(356), 370-385.

Alkan, M., Gursoy Surmeneli, H., & Polat, Z. A. (2018, May). Design and determine cadastral and land management performance of Turkey with cadastre 2034 vision. In *FIG Congress 2018*. Istanbul, Turkey. FIG.

Alkan, M., Gürsoy Sürmeneli, H., & Polat, Z. A. (2020). Design and development 3D RRR model for Turkish cadastral system using international standards. *Survey Review*, 1-13, doi: 10.1080/00396265.2020.1758386.

Döner, F., Thompson, R., Stoter, J., Lemmen, C., Ploeger, H., van Oosterom, P., & Zlatanova, S. (2011). Solutions for 4D cadastre–with a case study on utility networks. *International journal of geographical information science*, *25*(7), 1173-1189.

Guo, R., Ying, S., Li, L., Luo, P., & van Oosterom, P. (2011). A multi-jurisdiction case study of 3D cadastre in Shenzhen, China as experiment using the LADM. In *2nd international workshop on 3D Cadastres* (pp. 31–50). Delft, The Netherlands.

Herdlevær, H. (2018). Cadastral template. Norwegian Mapping and Cadaster Authority Director Cadastral Department.

Ho, S., Rajabifard, A., & Kalantari, M. (2015). 'Invisible'constraints on 3D innovation in land administration: A case study on the city of Melbourne. *Land Use Policy*, *42*, 412-425.

Kaufmann, J. & Steudler, D. (1998). Cadastre 2014, A Vision for a Future Cadastral System. Copenhagen: International Federation of Surveyors.

Karki, S., McDougall, K., & Thompson, R. (2010, April). An overview of 3D Cadastre from a physical land parcel and a legal property object perspective. In *24th International Federation of Surveyors International Congress (FIG 2010): Facing the Challenges - Building the Capacity* (pp. 1-13). Sydney, Australia. FIG.

Kitsakis, D., Paasch, J. M., Paulsson, J., Navratil, G., Vučić, N., Karabin, M., ... & El-Mekawy, M. (2016, October). 3D real property legal concepts and cadastre: a comparative study of selected countries to propose a way forward. In *5th International FIG 3D Cadastre Workshop* (pp. 1-24). 18-20 October, Athens, Greece. FIG.

ICSM. (2015). Cadastre 2034: Powering Land and Real Property. Intergovernmental Committee on Surveying and Mapping – ICSM. Canberra, Australia.

Lemmen, C., van Oosterom, P., Uitermark, H., Thompson, R., & Hespanha, J. P. (2009, May). Transforming the Land Administration Domain Model (LADM) into an ISO standard (ISO19152). In *FIG Working Week Eilat*, Israel. FIG.

Lemmen, C. H. J., van Oosterom, P. J. M., Uitermark, H. T., Zevenbergen, J. A. & Cooper, A. K. (2011, September). Interoperable domain models: The ISO Land Administration Domain Model LADM and Its External Classes. In *28th Urban Data Management Symposium (UDMS 2011)*. Delft, The Netherlands.

Lemmen, C., van Oosterom, P., & Bennett, R. (2015). The land administration domain model. *Land use policy*, *49*, 535-545.

Lemmen, C. H. J., & van Oosterom, P. J. M. (2013, September). The land administration domain model standard. In 5th Land Administration Domain Model Workshop. Kuala Lumpur, Malaysia, FIG.

Paasch, J.M. & Paulsson, J. (2014, November). Legal Framework 3D Cadastres-Position Paper 1. In *4th International Workshop on 3D Cadastres, 2014*. *Proceedings*. (pp. 411-416). Dubai, UAE.

Rajabifard, A., Atazadeh, B., & Kalantari, M. (2018). A critical evaluation of 3D spatial information models for managing legal arrangements of multi-owned developments in Victoria, Australia. *International Journal of Geographical Information Science*, *32*(10), 2098-2122.

Surmeneli, H. G., Koeva, M. N., Zevenbergen, J. A., & Alkan, M. (2020). Towards Integration of Ladm And Citygml For The Cadastral System of Turkey. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *43*, 691-698.

Steudler, D. (2015, May). Dimension cadastre–stepping beyond limits. In *FIG working Week 2015: From the Wisdom of the Ages to the Challenges of the Modern World, 2015.* Sofia, Bulgaria. FIG.

Steudler, D. (2014). *CADASTRE 2014 and Beyond*. FIG Publication No:61. Copenhagen, Denmark: International Federation of Surveyors.

Stoter, J. E., & van Oosterom, P. (2006). *3D cadastre in an international context: legal, organizational, and technological aspects*. Boca Raton, FL: CRC Press.

Tjia, D., & Coetzee, S. (2013). Application of the Land Administration Domain Model to the city of Johannesburg land information system. *South African Journal of Geomatics*, *2*(3), 260-279.

van Oosterom, P., Ploeger, H., Stoter, J., Thompson, R., & Lemmen, C. (2006, October). Aspects of a 4D cadastre: a first exploration. In *Peoceedings of Shaping the Change, XXIII international FIG congress*. Munich, Germany.

van Oosterom, P. (2013). Research and development in 3D cadastres. *Computers, Environment and Urban Systems*, *40*, 1-6. doi: 10.1016/j.compenvurbsys.2013.01.002.

van Oosterom, P. J. M. (2018). Best practices 3D cadastres – extended version. FIG Publication. International Federation of Surveyors. *Copenhagen, Denmark.*

Wakker, W. J., van der Molen, P., & Lemmen, C. (2003). Land registration and cadastre in the Netherlands, and the role of cadastral boundaries: the application of GPS technology in the survey of cadastral boundaries. *Journal of geospatial engineering*, *5*(1), 3-10.