

Machine Learning-Based Prediction of Insect Damage Spread Using Auto-ARIMA Model

Ece Alkan, Abdurrahim Aydin

Abstract

Differentiating areas of insect damage in forests from areas of healthy vegetation and predicting the future spread of damage increase are an important part of forest health monitoring. Thanks to the wide coverage and temporal observation advantage of remote sensing data, predicting the future direction of insect damage spread can enable accurate and uninterrupted management and operational control to minimize damage. However, due to the large amount of remotely sensed data, it is difficult to process the data and to identify damage distinctions. Therefore, this paper proposes a spatio-temporal Autoregressive Integrated Moving-Average (ARIMA) prediction model based on the Machine Learning technique for processing big data by monitoring oak lace bug (*Corythucha arcuata* (Heteroptera: Tingidae)) damage with remote sensing data. The advantage of this model is the automatic selection of optimal parameters to provide better forecasting with univariate time series. Thus, multiple spatio-temporal warning levels are distinguished according to the damage growth trend in the series, and the network is constructed with improved time series to better predict future insect damage spread. In the proposed model, the historical Red (R) – Green (G) – Blue (B) bands of the Sentinel-2 (GSD 10 m) satellite were tested as a dataset for the oak lace bug damage in the oak forest situated in the campus of Düzce University, Turkey. The dataset, which contained 38 images for each of the RGB bands, was modeled using the open source R programming language for the peak damage period in 2021. As a result of the test, significant correlations were found between the synthetic and true images (True and synthetic band 2: $r=0.960$, $p<0.001$; True and synthetic band 3: $r=0.945$, $p<0.001$; True and synthetic band 4: $r=0.962$, $p<0.001$). Then, the 48-month time series bands were modeled, and the band estimates were made to predict the August 2023 spread. Finally, a synthetic composite image was created for future prediction using the predicted bands. The tests showed that the model had a good performance in insect damage monitoring. With open access Sentinel-2 images, the proposed model achieved the highest prediction accuracy with a rate of 96%, and had a small prediction error.

Keywords: remote sensing, insect damage, machine learning, ARIMA model

1. Introduction

Remote sensing data have been used in forest health monitoring programmes to plot forests that are stress exposed for various reasons (Nicholas et al. 2006). In recent years, a growing number of studies have been conducted on advantages of using high resolution remote sensing data for the detection and monitoring of insect damage areas. Remote sensing data have been used in forest health monitoring programs to identify forests stressed by various causes (Nicholas et al. 2006). In recent years, there has been a growing body of work demonstrating the advantages of using high-resolution remote sensing data to detect

and monitor areas of insect damage. High-resolution IKONOS (Wang et al. 2016) and WorldView-2 (Immitzer and Markus 2014, Lottering et al. 2020) satellite imagery has been used to reduce the workload of teams engaged in the detection of insect damage areas in field studies to determine forest health state. Satellite images have been processed and analyzed with statistical methods such as texture analysis (Gray Level Occurrence Matrix – GLCM), Getis Statistics, Random Forest (RF) classification, and ANOVA tests to determine the risk of future infestation (Wang et al. 2016, Immitzer and Markus 2014, Lottering et al. 2020). These studies show that insect damage detection is possible using satellite imagery. On the other hand, it

is very important to determine the impact of insect damage on host tree species and their distinctive characteristics in a short time and at lower cost. Sentinel-2 satellite data is preferred over other open access satellite data due to its free open access and good temporal resolution. Generally, Sentinel-2 data have been used in applications such as near real-time processing, forest monitoring, insect damage area detection and spread monitoring. Sentinel 2 spectral bands have also been used as multiple combined vegetation indices for mapping insect damage and supporting the pest management strategies of forest and infestation control of forest (Zhan et al. 2020, Gärtner et al. 2016, Kumbula et al. 2019). Insect damage studies have been conducted using single-date imagery (Gärtner et al. 2016, Kumbula et al. 2019), as well as time series studies using multi-temporal imagery to monitor the spread of pest (Hornero et al. 2020). Although insect damage studies using single-data Sentinel-2 images have shown promising results, multi-temporal Sentinel 2 images have generally outperformed those using single data images (Zhan et al. 2020, Hornero et al. 2020). Sentinel-2 temporal-imagery has been used to detect colour change, i.e. phenological change, in vegetations caused by insect damage (Rajeev et al. 2021). However, when using remote sensing data, it is crucial to consider the damage patterns of insects in order to separate insect damage from the phenological cycle of vegetation. For example, while the invasive Bark Beetles damage the vegetation by yellowing, wilting, browning of the leaves and finally the death of the branches (Christiansen et al. 1987), some insect pests silky webs on the leaves and eat the parenchyma up to the upper epidermis, leaving only the main veins (Masaki and Umeya 1977). On the other hand, pests such as Oak lace bug (*Corythucha arcuata*, Say 1832 – Heteroptera: Tingidae) complete their lifecycle on the same vegetation and can be very destructive (Neal and Schaefer 2000). Oak lace bug is a pest that causes vegetation loss and potential risk to the health, productivity and stability of natural oak forests (Anikó et al. 2021). The lace bug feeds on the undersides of leaves by piercing the leaf epidermis and sucking the sap, and the empty cells give the leaves a bronze or silvery appearance (Wappler 2003). In this sense, the ability of the remote sensing system to monitor gradual changes in spectral reflection over forest canopy is essential for the detection, monitoring and management of oak lace bug damage (Gašparović et al. 2022). In this type of damage process, it is possible to monitor and model the annual change in damage by processing multi-temporal remote sensing images (Trubin 2022) and the analysis of the change in these areas can be carried out using a dense Sentinel-2 dataset (Bárta et al. 2021). Due

to the large size of the Sentinel-2 multi-temporal dataset subject to analysis, it is difficult to process and analyze the data and predict future spread direction of the insect damage. Therefore, machine learning techniques such as multi-temporal regression-based are used for processing large data sets (Fernandez-Carrillo et al. 2020, Bárta et al. 2021, Hashim et al. 2021). In the literature, there have been many studies on insect damage detection with machine learning techniques using remote sensing data. Bhattarai et al. (2021) used CART, RF and SVM methods on Gaofen-2 (GF2) images to identify tree mortality affected by the red turpentine beetle (RTB) (*Dendroctonus valens* LeConte) pest. They obtained the highest accuracy of 77.7% with SVM. Fernandez-Carrillo et al. (2020) extracted the distribution of trees with spruce budworm (*Choristoneura fumiferana*) from Sentinel-1,2 satellite data by random forest method using spectral plant indices. Hashim et al. (2021) proposed a multi-temporal regression-based change detection method for mapping areas affected by bark beetle at different severity levels. Sentinel-2 images were used in the study. Normalized difference vegetation index (NDVI) and modified soil-adjusted vegetation index (MSAVI) were used as additional data sources, and RF method was used to segment the damaged areas. Harati et al. (2020), applied Multilayer Perceptron and RF methods to ALOS PALSAR-2 radar images to investigate the effects of *Ganoderma boninense*, which causes basal root rot disease in palm plants, and obtained a segmentation accuracy of 92.70% and 95.65%, respectively. Huo et al. (2021) applied generalized linear regression (GLM) and RF algorithms using aerial photographs and LANDSAT satellite imagery to simulate the spatio-temporal dynamics of mountain pine beetle (MPB) infestation in lodgepole pine forests in British Columbia (BC), CANADA. Bárta et al. (2021) used Sentinel-2 bands for early detection of bark beetle infestation in Norway spruce monoculture forests in the Czech Republic. The potential of using selected vegetation indexes based on seasonal trajectories of damage and vegetation was investigated. The RF algorithm was used to classify healthy (i.e., uninfested stands) and different degrees of damaged trees, and the algorithm was applied to the time series of Sentinel-2 observations since 2019 for early detection of damage to trees based on the assessment of seasonal changes. In another study, hyperspectral remote sensing techniques were used to detect changes in the biochemical-biophysical vegetation characteristics of spruce vegetation, analyzing the hypothesis that it was already prone to damage due to factors such as climate change before the invasion. Thus, a trend towards detectability and differentiation with spectral indicators and

index derivatives for early warning was identified (Lausch et al. 2013). Olsson et al. (2012) emphasized the importance of developing methods that enable effective monitoring of insect attacks in forested areas, noting that if defoliation or discoloration is severe enough, some sensors in satellite data will also facilitate the monitoring of migration patterns of invasive insects, as they provide time series that enable monitoring of insect attacks. In their study, they used SPOT and MODIS data to map the damage caused by *Physokermes inopinatus* in Norway spruce (*Picea abies*) and black crusting caused by *Physokermes inopinatus* during an attack in Scania in southernmost Sweden in 2010. With SPOT data, the area of damage was detected with an estimate of 78%, and with MODIS with 250 m resolution, a 16-day composite NDVI index was produced, showing that larger areas of damage could be detected. This study also emphasized the potential of remote sensing studies for early detection and monitoring of invasive insect damage. In addition to early detection and traditional artificial intelligence algorithms, the use of deep learning methods and innovative convolutional architectures has recently started to be used, albeit rarely, in mapping insect damage areas (Sylvain et al. 2019). For example, Zhao et al. (2019) used the VGG16 convolutional neural network (ESA) to map dead forest cover on aerial photographs in Canada. In addition, deep learning-based approaches with UAVs have also started to be used in the identification of damaged forests and trees. Deng et al. (2020), used ResNet101 and VGG16 architectures from UAV and artificial intelligence technology to identify nematode diseases that cause great economic losses in pine forests due to their destructiveness and rapid spread. Hu et al. (2020) proposed a deep learning-based method for dynamic monitoring and control of diseased pine trees from UAV images. Alexnet, VGG and Inception_v3 networks were used in the study. The recall value obtained in the study is 0.957. Kerkech et al. (2020) used two SegNet models for segmentation in UAV images to map diseased areas in vineyards and thus ensure healthy vine protection, which is very important for yield management. Thus, they proposed a method for mold disease detection. The method is based on the combination of visible and infrared images obtained from two different sensors. The proposed method achieved 92% accuracy for vine and 87% accuracy for leaf. Roosjen et al. (2020) conducted a study on monitoring insect traps by using computerized image processing methods and deep learning methods with the images obtained by UAV in the detection of the invasive *Drosophila suzukii* damage originating from South East Asia. ResNet-18 architecture was used in their study. Qin et al. (2021) used

SCANet convolutional neural network from multi-spectral UAV images for early diagnosis of pine nematode disease, achieving an overall accuracy of 79%, with sensitivity and recall values of 0.86 and 0.91. The management of insect-damaged forest areas largely relies on forecasting models, and the ARIMA time series forecasting model is designed to model social, natural, ecological, financial phenomena (Malik and Umar 2021). This model is used to predict future values based on values observed in the past (Box et al. 1991). In this study, since oak lace bug damage is an ecological phenomenon that changes periodically and steadily over time, time series analysis and near future forecasting were analyzed by Automatic-ARIMA model in R studio software, using a dense series of Sentinel-2 data.

2. Materials and Methods

In this section, data sources, time series analysis model, and the model building processes are discussed in detail. Performance measures and selection criteria of the non-seasonal ARIMA model will be discussed together with their mathematical formulation, application to time series analysis and forecasting of synthetic composite image.

2.1 Problem Definition and Study Area

Oak lace bug was detected in oak species such as *Quercus petraea*, *Quercus robur*, *Quercus pubescens* as a result of surveys covering a total of 865 square kilometers in Türkiye-Düzce between 2002 and 2006 (Mutun et al. 2009). Lace bug is an invasive species that causes serious damage to oak species (Neal et al. 2000). Therefore, in order to estimate the future lace bug distribution, fieldwork was carried out in the oak forest of Düzce University Campus in July–August 2021. During field work in July and August 2021, intensive Lace bug damage was observed, and it was found that lace bugs settled on the undersides of the leaves caused serious yellowing and discoloration in the forest. The oak forest covers a total area of 77.30 hectares. The coordinates of the Düzce University Campus oak forest is (40°54'44.14"N, 31°10'12.87"E), (40°54'25.13"N, 31°11'21.07"E) and selected as the study area (Fig. 1) in order to detect the color change in vegetation with remote sensing data and to forecast its spread.

2.2 Model Input Data

Sentinel-2 data for the Oak forest of Duzce University, covering the period June 2018 to June 2022, were retrieved from the Copernicus portal of the European Space Agency (Copernicus 2022). For

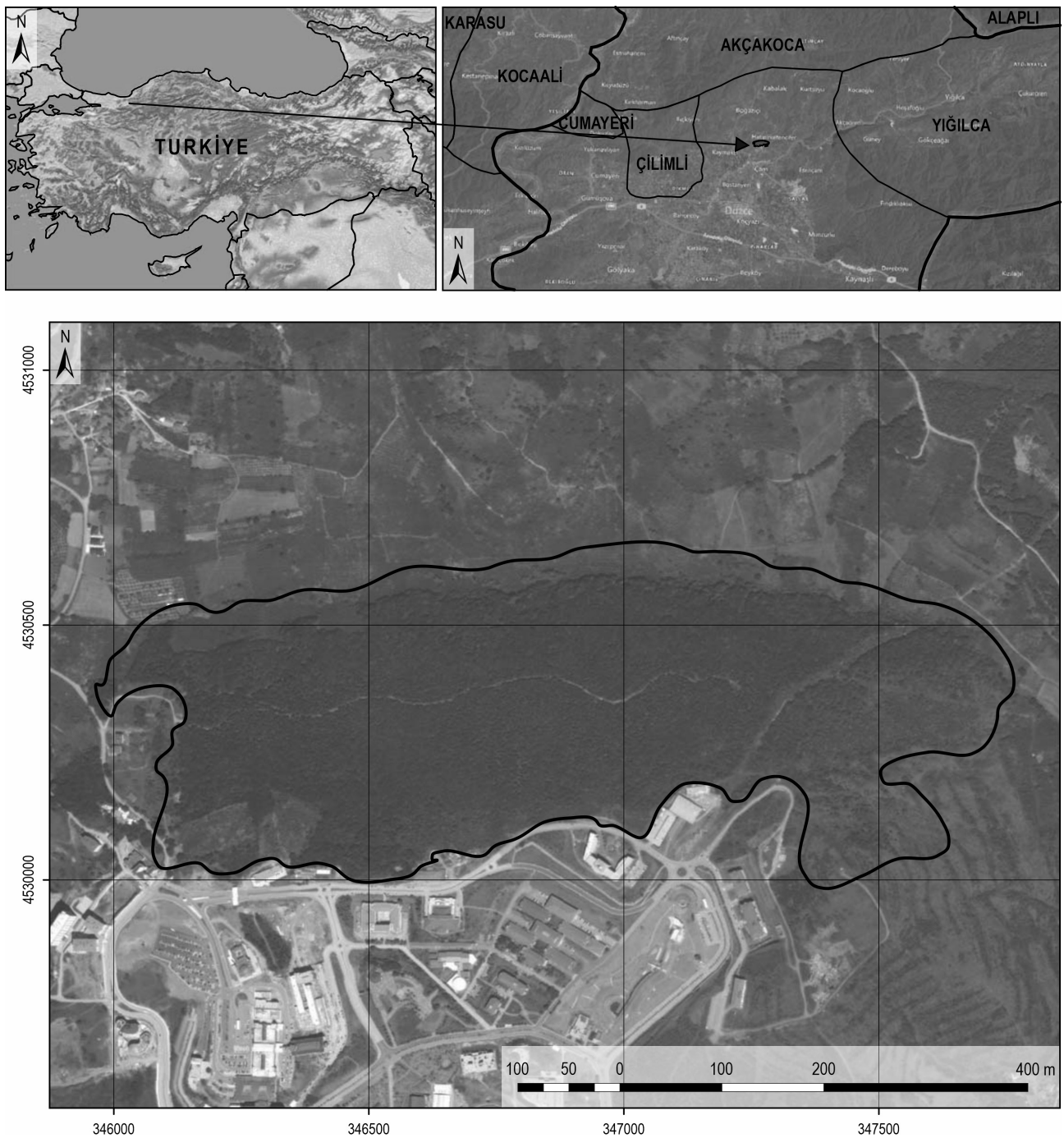


Fig. 1 Study Area

purposes of developing forecasting model, data for the period June 2018 to July 2019 was first used, because of the absence of irregularity components in the time series data as compared to following years, to predict August-2019 (from no damage time series data to predict no damage target month). In this period, there was no lace bug damage in the oak forest and

this period was chosen to test in the ARIMA model. Then data for the period June 2018 to July 2021 were used as input in order to be sure if the model predicted correctly the month of intensive damage of August 2021 (from damage + non damage time series data to predict damage target month). Finally, by using data from June 2018 to May 2022, the spread of lace

bug damage was forecasted for August 2023 (from damage + non damage time series data to future spread forecasting of Lace bug for target month). In the developed model, a total of 14 time series data for model validation, 38 time series data for testing the model performance in the ability of detecting lace bug and 48 time series data for forecasting the future spread were used. Model series quantities are shown in Fig. 2.

2.3 Model Methodology

The model methodology of this study consists of the following steps:

⇒ Data Pre-processing

Atmospheric correction (SNAP software)

Creating Sentinel-2 time series tables (ARCGIS Software)

⇒ Auto-ARIMA and Model Building

Calculating RGB prediction pixels using the Auto-ARIMA Forecast Package in R (R Studio software)

⇒ Model Validation

Calculating Pearson correlations between forecasted pixels and true pixels

⇒ Generating Synthetic Composite Images (ARCGIS Software)

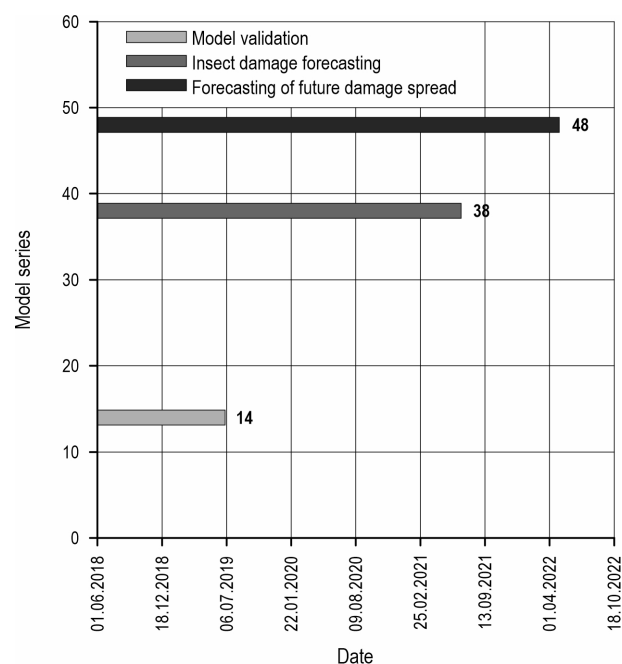


Fig. 2 Model series quantities

Creating true color composite images (TCI) and synthetic composite images (SCI). (ARCGIS software)

The flowchart of the methodology is presented in Fig. 3.

For this study, the Red (R), Green (G), Blue (B) bands of Sentinel-2 multi-temporal imagery, which is open access for 48 months between June 2018 and May 2022, were used. Atmospheric corrections of the bands were performed. Band pixel values were converted from raster-based to vector-based points using ARCGIS software. The point pixel values obtained in ARCGIS software were converted into time series tables with (.csv) file extension in Microsoft Office Excel software. Auto-ARIMA Forecasting Package was used in R studio software to make future damage forecasting with time series tables in (.csv) format. Since Auto-ARIMA Forecasting codes are adapted to »Univariate time series tables« (input data = .csv format), Raster operations were performed in ARCGIS software instead of R package. The output prediction pixels were tabulated, and the correlations between them and the true pixel values were calculated. As a final step, raster processing with the forecasted pixel values was performed in ARCGIS software.

2.3.1 Data Pre-processing

In this section, Red (R), Green (G), Blue (B) bands of open access Sentinel-2 multi-temporal imagery for 48-month between June 2018 and May 2022 were used. Atmospheric Rayleigh Scattering Correction of Sentinel 2 Images is performed using ESA SNAP software. True band (TB) pixel values were extracted from the raster layer based on the vector point layer of RGB (B4, B3, B4) bands by creating a fishnet with ARCGIS sampling, and 14, 38, 48 month series tables were prepared (Table 1).

The tables consist of a series of observations taken in monthly series between 2018 and 2022. They were used to forecast future pixel values based on pixel values observed in the past. Such datasets, where only one variable is observed at a time, are referred to as »Univariate Time Series« (Tichaona et al. 2020). Univariate non-seasonal time series were used for damage detection and future prediction of oak Lace bug, and analyzed with Auto-ARIMA model using open source R studio software.

2.3.2 Auto-ARIMA and Model Building

Once the data set is ready, the model building process begins. First, 14 RGB pixel series were used to validate the ARIMA model and test Sentinel-2 data. The second set of 38 RGB pixel series was used to

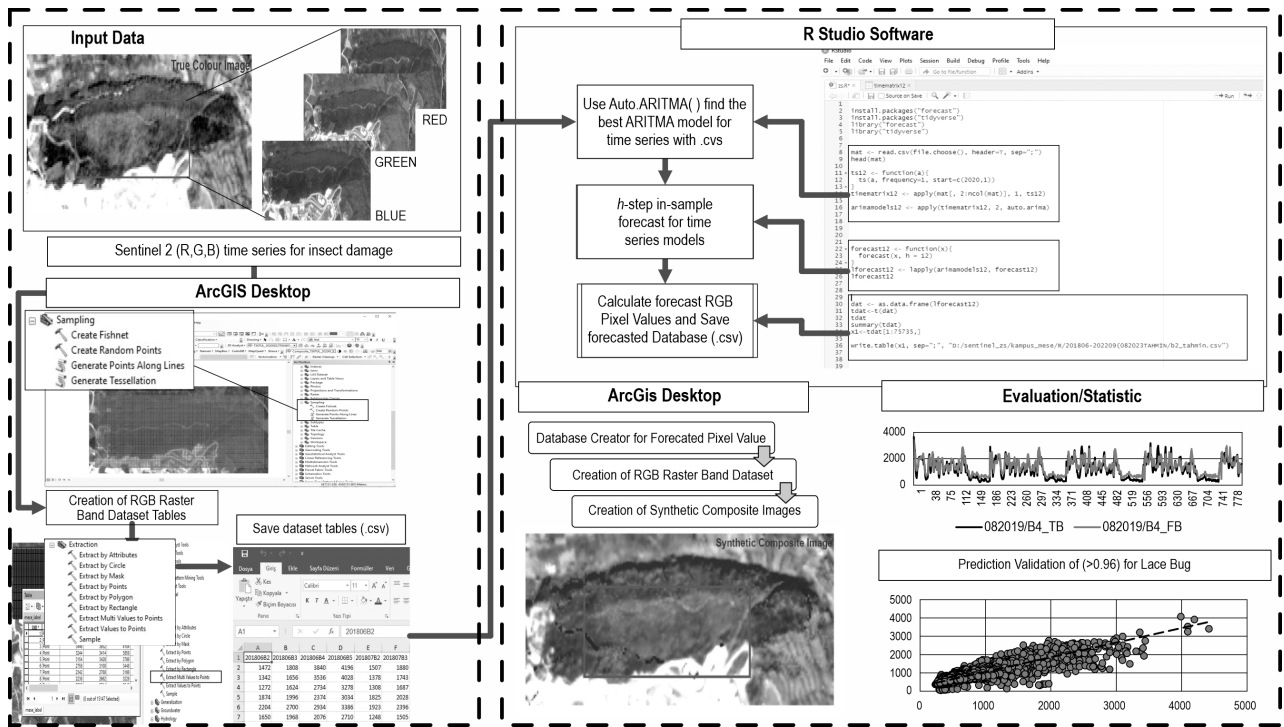


Fig. 3 General process and methodology of Auto-ARIMA model

evaluate the ability of the prediction set to predict lace bug damage. Another data set of 48 RGB pixel series was used to forecast future spread.

The pixel differences between consecutive series are combined with autoregression and average parameters are moved to obtain a non-seasonal ARIMA model with R studio. The number of autoregressive terms (p), the number of non-seasonal differences re-

quired for stationarity (d) and the number of lagged forecast errors (q) in the estimation equation represent the non-seasonal components of the model (Box et al.1991). The model equation is represented by the generalised Eq. 1.

$$y'_t = c + \phi_1 y'_t - 1 + \dots + \phi_p y'_t - p + \theta_1 \epsilon_t - 1 + \dots + \theta_q \epsilon_t - q + \epsilon_t \quad (1)$$

Table 1 Tabular view of R band of June 2018

2018/Jun_B4	2018/Jul_B4	2018/Aug_B4	2018/Sep_B4	2018/Jun_B4
3840	3730	3620	3370	4500
3536	3665	3794	3560	4496
2734	2796	2858	2540	4188
2374	2449	2524	2272	3932
2934	2577	2220	2276	3822
2076	1578	1080	1358	3122
1218	1018	818	972	2904
644	669	695	767	3008
512	587	662	729	3149
:	:	:	:	:	:	:

In this basic ARIMA model, it is necessary to manually provide the optimal p , d and q values. However, the datasets prepared for this study were trained with the »Auto-ARIMA « function of the R package, which updates the parameters (p, d, q) during training to provide better insect pest spread prediction. R is an open source language used worldwide. It has several machine learning packages and advanced implementations for the top machine learning algorithms – which every data scientist must be familiar with, to explore, model and prototype the given data. One of them is the Auto-ARIMA function. Auto-ARIMA has the ability to decide whether or not the data used to train the model needs a seasonal differencing. The function performs a search over the possible models in the dataset (Hyndman and Khandakar 2008). It provides the best model by trying to find the optimal values of p , d and q with different combinations and final values (Box et al. 1991). In Auto-ARIMA, the h -step is applied for sample estimation of time series and only the parameter h is calculated manually. Each value of h represents the date range of the damage area to be estimated. Thus, the RGB pixel value of each band in the selected date range was estimated.

2.3.3 Model Validation

In the next step, Pearson Correlations were calculated to assess the similarity between the forecasted band pixel values of B2, B3, B4 (FB) and the true band pixel values of B2, B3, B4 (TB). The root mean square error (RMSE) was used as a performance measure during training.

The RMSE is represented by the generalised Eq. 2.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{Forecasted}(i) - \text{True}(i))^2}{N}} \quad (2)$$

2.3.4 Generating Synthetic Composite Images

The raster bands of the FB pixel values showing high correlation with the TB pixel values were calculated in the ARCGIS software. Thus, RGB band datasets for each predicted date were calculated with the Point to Raster tool and synthetic composite images were created in ARCGIS software, and a comparison was made between true color composite images (TCI) and synthetic composite images (SCI).

3. Results

First, the results of the Auto-ARIMA prediction models are presented, followed by SFI models of the best estimate of future spread damage. The models

consist of three different sets, one with 14 series and the others with 38 and 48 series, respectively. The number of inputs in the model was chosen according to the objective of each model. First, for model validation, 14 series of inputs were used up to August 2019, when there were no damages.

In the first model, RGB pixel values were forecasted with Auto-ARIMA for August 2019 and the following four months. That is, h -step = 5 was chosen for the prediction model. The FB pixel values were compared with the TB pixel values and time series graphs were created (Fig. 4).

The FB obtained by Auto-ARIMA prediction in time series and their true observation data TB are compared in the graph shown in Fig. 5. For the two datasets (FB and TB), random point cross-sections were taken and their values were reflected (Fig. 5). The Y-axis shows the random pixel point numbers and the X-axis shows the pixel values corresponding to the points. The similarity of the FB values to the TB values in the series revealed that the points were forecasted in good agreement. Then, RMSE was used as the final performance metrics for selecting a suitable model for Auto-ARIMA. Correlations between TB and FB were calculated. The graph of the pixels in August 2019 revealed a high correlation between the TB (B2, B3, B4) pixels and the FB (B2, B3, B4) pixels (Fig. 5).

For 2019, the calculated correlations are $r=0.9221$ between TB pixel values and FB pixel values for B2, $r=0.8427$ between B3 band true pixel values and forecasted pixel values, and $r=0.8568$ between B4 band true pixel values and forecasted pixel values. This confirms that TB and FB pixels should be dependent. In other words, the TBs and FBs of 2019 showed high similarity between each other. These results show that the Auto-ARIMA prediction model can effectively forecast Sentinel-2 TB pixel values in the 14-month time series. Therefore, in order to compare the forecasted values with the true composite image, synthetic (predicted) raster bands (B2, B3, B4) were created from the forecasted pixel values. Then, these predicted synthetic raster bands were combined to create a Synthetic (Forecast) Composite Image (SCI) for August 2019 (Fig. 6).

It is concluded that the SCI is largely representative of the true composite image. That is, the Auto-ARIMA model predicted the study area close to the reality (Fig 6). In the second stage of the model, a 38-month univariate time series data was used to determine the known lace bug damage in August 2021. The dataset was analyzed in R studio software to evaluate the prediction performance of the model. The parameter h was chosen as 1 for the model results. Pearson Correlation Analysis was performed between the FB

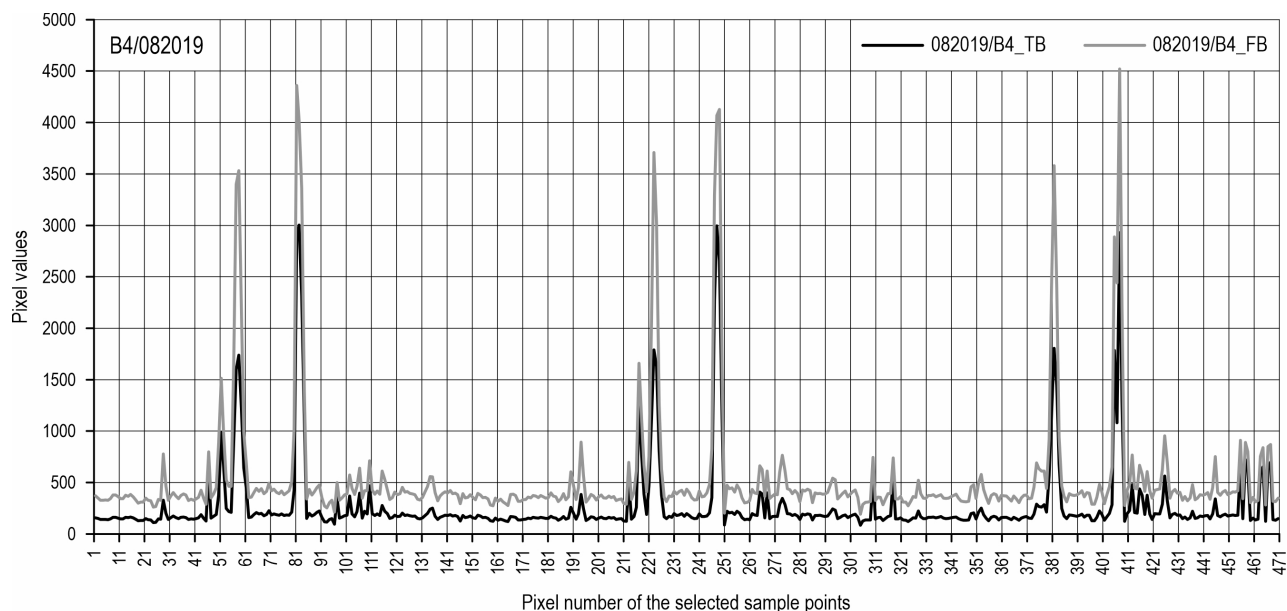


Fig. 4 To Compare Time Series Graphics for true color composite bands (black) and forecasted bands (gray)

pixel values and the TB pixel values. The test revealed significant correlations between the FB and TB ($rB2=0.960$, $p<0.001$; $rB3=0.945$, $p<0.001$; $rB4=0.962$, $p<0.001$).

It is concluded that FB pixel value is largely representative of the TB pixel value based on correlation results. Then, a SCI was created using the forecast raster bands for August 2021. A comparison was made between the TCI and SCI and image processing techniques, such as histogram equalization, were used to reveal the forecast damage area more clearly (Fig 7).

These results showed that the Synthetic (Forecast) Composite Image (SCI) predicted by the Auto.ARIMA model has effectively reflected the current lace bug damage on the Sentinel-2 TCI for August 2021. It was also concluded that the correlations between the FB and TB pixel values of the 38-month time series model were higher than those of the 14-month time series model. In the last model, the future propagation trend of the damage area was estimated. To predict August 2023, a 48-month univariate time series data was used and analyzed with Auto.ARIMA. The parameter h was chosen as 12 for the model results. A SCI was created by calculating raster bands with the FB pixel values (Fig 8).

It was concluded that the SCI created for 2023 does not represent the damage area as seen in 2021. Considering the pixel reflectance values, it was revealed that the SCI in 2023 was more similar to 2018. Pearson Correlation results are shown in Fig. 9.

As a result, this model predicts a decrease in lace bug damage in the future.

4. Discussion

Remote sensing data are important resources that provide solutions to managerial problems by monitoring earth resources. With this data monitoring, insect damage, mapping and forecasting epidemic models provide supplementary inputs for decision support systems. ARIMA is a forecasting model with successful performance in modeling ecological phenomena (Slavia et al. 2019). There are many studies on ecological phenomena such as land surface temperature (LST), urban heat island (Kesavan et al. 2021), drought prediction (Amin et al. 2022), malaria case prediction (Adeola et al. 2019), wheat yield prediction (Deng et al. 2022) with the ARIMA model. However, no studies can be found in the literature on insect damage distribution by using this model. This study used the Auto.ARIMA algorithm to model lace bug damage, and the prediction results showed that the lace bug damage would sharply decrease. In general, the Auto.ARIMA model (without assigned values) shows more acceptable validation values than the classical ARIMA model (with assigned parameters p, q, d) (Choudhary et al. 2022). On the other hand, the highest prediction value of 0.962 obtained with the Auto.ARIMA model algorithm in this study showed that this model can be used to predict insect damage distributions. Moreover, Auto.ARIMA does

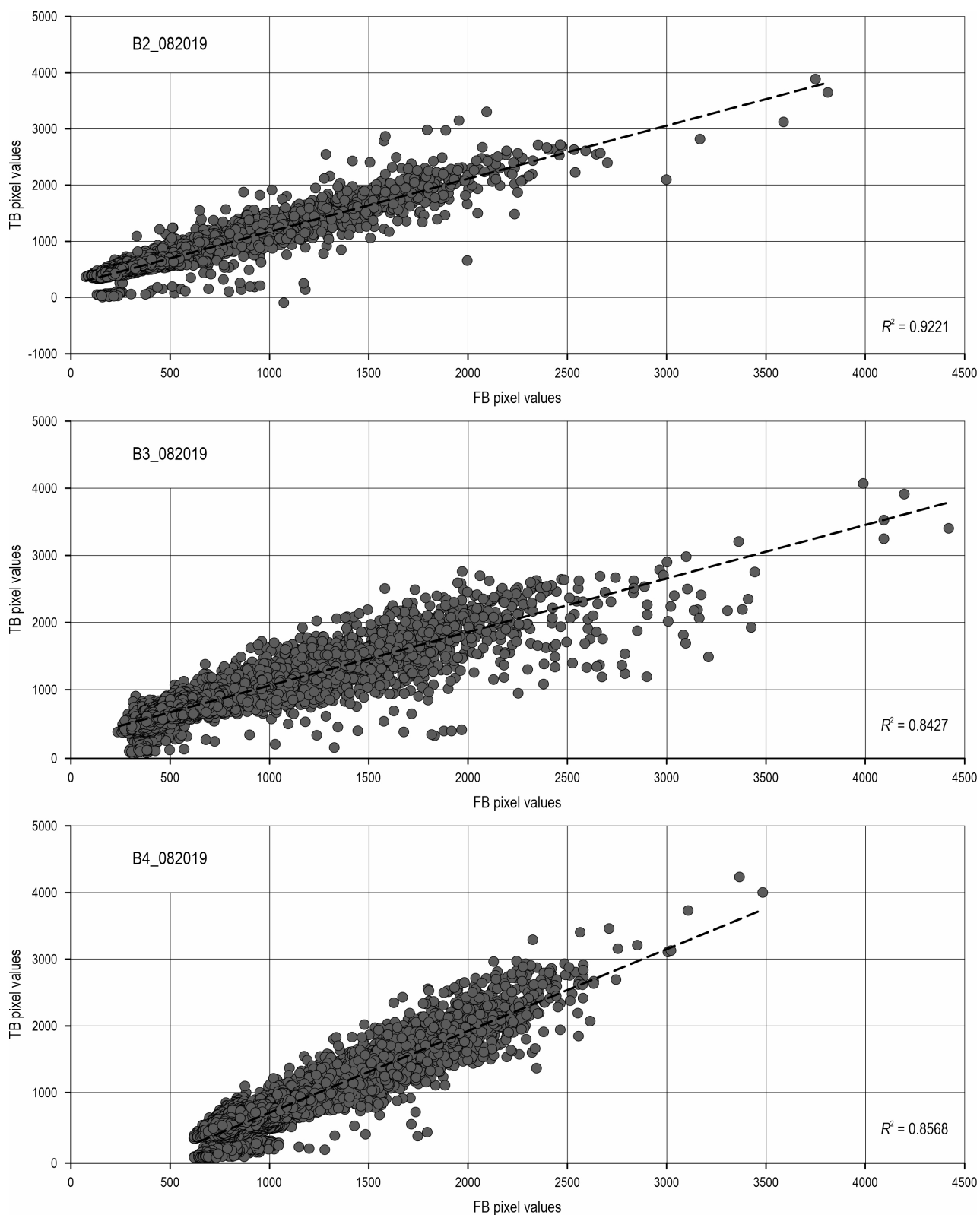


Fig. 5 Correlation Plots Between TB Pixel Values and FB Pixel Values

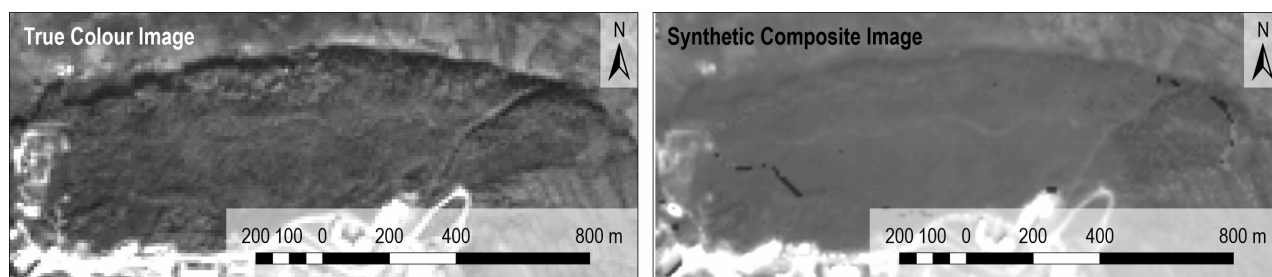


Fig. 6 Forecast Composite Image_201908

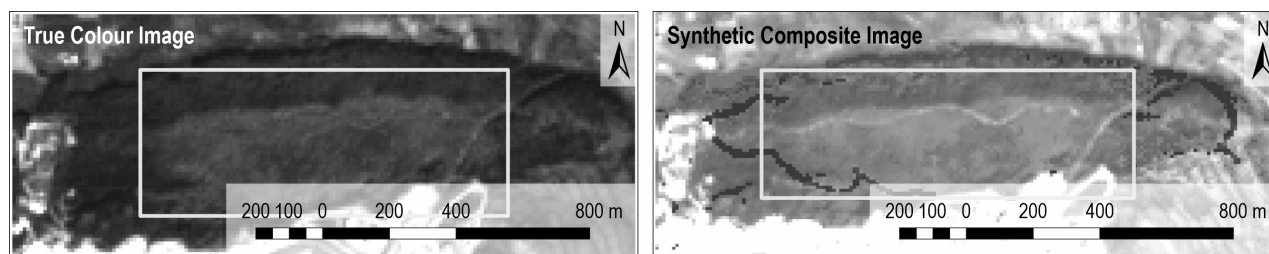


Fig. 7 Forecast Composite Image_202108

not require extensive, long time series like traditional deep learning-based models and does not restrict model accuracy (Deng et al. 2022). In this study, among the three datasets analyzed with Auto.ARIMA using R studio software, prediction validation was obtained in the short-term 14-month dataset (>90), and it was concluded that the dataset constraint does not negatively affect the model validation even in the shortest-term time series. Although the non-seasonal Auto.ARIMA model was used due to temporal data limitations during the period of lace bug damage in the study area, the high model validation obtained with this model after solving the insufficient data problem showed that the Auto.ARIMA model is suitable for future insect damage distribution prediction studies. However, in view of the fact that insect damage is a seasonal ecological phenomenon, it is thought that it would be useful to use seasonal

ARIMA models with a temporally long-term damage data set in future studies. In addition to the limited number of studies on forecasting insect pest distributions, there are no many algorithm-based decision support forecasting models. Zakariyyaa and Onisimo (2013) used an algorithm-based multivariate partial least squares (PLS) regression to estimate the spread of insect pest infestation in forest plantations. Although the highest model validation of 0.65 was achieved, in our study a prediction validation of (>0.96) for lace bug propagation was obtained using the R studio software based Auto.ARIMA model. These results suggest that Auto.ARIMA model can provide high performance for other insect pest spread studies. The use of open access remote sensing data (Sentinel-2) and open source software for predicting insect pest spread has provided advantages in terms of cost and time. The ARIMA model in this study achieved good model accuracy for lace bug damage spread because the predicted model accuracies for the year 2021 (B2=0.960, B3=0.945, B4=0.962) were high and the synthetic composite images reflected the actual insect damage images. In conclusion, in our opinion, in future studies, predicting the direction of spread of other types of insect damage in forests with Auto.ARIMA model will support forest health monitoring studies and management planning related to insect damage.

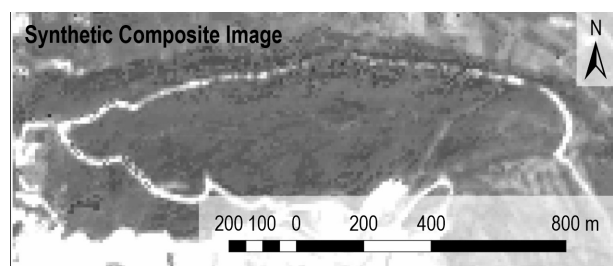


Fig. 8 Forecast Composite Image_202308

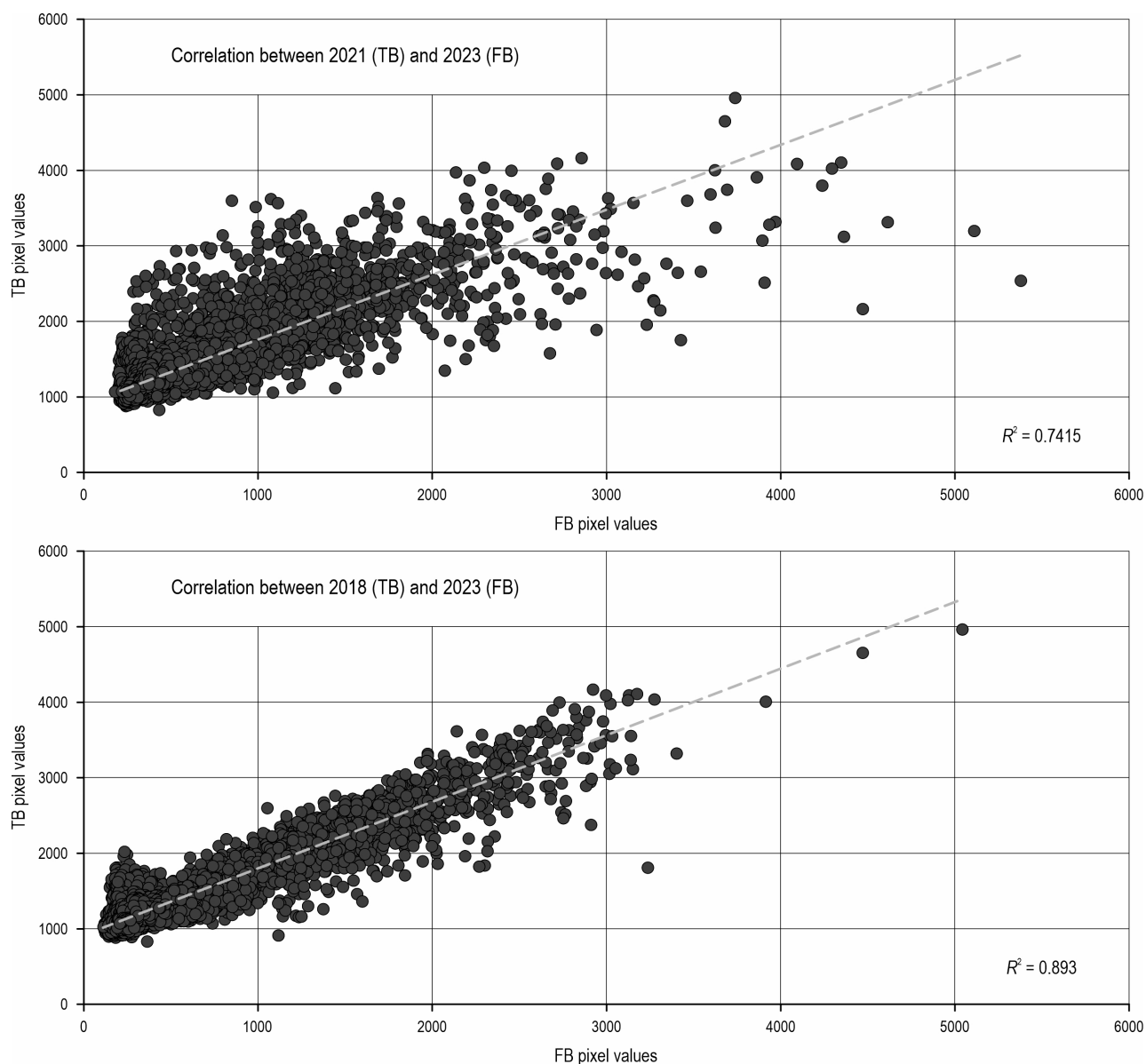


Fig. 9 Pearson Correlation Analyses for 2023

5. Conclusions

In this research, Sentinel 2 time series bands were used as three different data inputs for validation, detection of insect damage area and forecasting future lace bug spread. Reflectance values were predicted by ARIMA Model using machine learning (R package program). Correlations were calculated by comparing the band pixel values predicted as a result of ARIMA model analysis with the actual bands. In the first stage, correlations were calculated between True Band pixel values and False Band pixel values ($rB2=0.9221$,

$rB3=0.8427$, $rB4=r=0.8568$) for 14 time series analysis. With this result, it was concluded that the forecast for 2019, when there was no damage, was highly representative of the actual study area. In the second stage, ARIMA model performed well in predicting insect damage with the correlation results between 38 time series actual and prediction bands ($rB2=0.960$; $rB3=0.945$; $rB4=0.962$). It was also concluded that the prediction accuracy of the ARIMA model increased as the number of time series increased. The pixel values estimated as a result of 48 time series used in the future propagation prediction were very close to the

values in the years when there was no damage. They were compared with the band pixel values of 2018 when there was no damage, and a correlation of $r=0.8930$ was calculated. The results obtained indicated that the future damage spread will decrease. Synthetic composite images were created with the three main results obtained. It was observed that the synthetic images exactly represent the real-time images.

According to these conclusions, the performance measures, PC and RMSE test show that the series models analyzed by ARIMA perform good for short-term forecasting of lace bug damage in oak forest. The ARIMA model building time series used 14, 38 and 48 observations. It is concluded that, despite such short time series, the damage patterns in the time series are good forecasts. However, using the non-seasonal ARIMA model negatively affects the model unless the phenological spectral change of vegetation in winter months is separated from the spectral values of insect damage. In this study, the time series was limited by the fact that damage was clearly observed in the field for only one season. For this reason, the series quantities were increased by using non-seasonal ARIMA to make more accurate predictions. Since insect damage is a seasonal ecological phenomenon, it may be preferable to use a seasonal ARIMA model in future studies with more observations.

Acknowledgements

This article is derived from a PhD dissertation conducted by the lead author under the supervision of the second author at Düzce University, Institute of Graduate Studies, Department of Forest Engineering. Also, the thesis work was supported by DÜBAP project No. 2022.02.02.1351 »Artificial Intelligence Based Detection of Insect Damage in Forests with Remote Sensing Data«. For this reason, we would like to thank Düzce University Scientific Research Projects Coordinatorship for providing financial support for the thesis study. Special thanks to Prof. Ahmet MERT who supported the development of this article idea.

6. References

- Adeola, A.M., Botai, J.O., Mukarugwiza Olwoch, J., de W. Rautenbach, H.C., Adisa, O.M., de Jager, C., Botai, C.M., Aaron, M., 2019: Predicting malaria cases using remotely sensed environmental variables in Nkomazi, South Africa. *Geospatial Health* 14(1): 81–91. <https://doi.org/10.4081/gh.2019.676>
- Zeynolabedin, A., Olyaei, M.A., Zahmatkesh, Z., 2022: Application of meteorological, hydrological and remote sensing data to develop a hybrid index for drought assessment. *Hydrol. Sci. J.* 67(5): 703–724. <https://doi.org/10.1080/02626667.2022.2043551>
- Kern, A., Marjanović, H., Csóka, G., Móricz, N., Pernek, M., Hirka, A., Matošević, D., Paulin, M., Kovač, G., 2021: Detecting the oak lace bug infestation in oak forests using MODIS and meteorological data. *Agric. For. Meteorol.* 306: 10843. <https://doi.org/10.1016/j.agrformet.2021.108436>
- Bárta, V., Lukeš, P., Homolová, L., 2021: Early Detection of Bark Beetle Infestation in Norway Spruce Forests of Central Europe Using Sentinel-2. *Int. J. Appl. Earth Obs. Geoinf.* 100: 102335. <https://doi.org/10.1016/j.jag.2021.102335>
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015: *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken.
- Choudhary, A., Kumar, S., Sharma, M., Sharma, K.P., 2022: A Framework for Data Prediction and Forecasting in WSN with Auto ARIMA. *Wirel. Pers. Commun.* 123: 2245–2259. <https://doi.org/10.1007/s11277-021-09237-x>
- Christiansen E., Waring, R.H., Berryman A.A., 1987: Resistance of conifers to bark beetle attack: Searching for general relationships. *For. Ecol. Manage.* 22(1–2): 89–106. [https://doi.org/10.1016/0378-1127\(87\)90098-3](https://doi.org/10.1016/0378-1127(87)90098-3)
- Copernicus Open Access Hub, <https://scihub.copernicus.eu/>, Access:14.09.2022
- Deng, Q., Mengxuan, W., Zhang, H., Cui, Y., Li, M., Zhang, Y., 2022: Winter Wheat Yield Estimation Based on Optimal Weighted Vegetation Index and BHT-ARIMA Model. *Remote Sens.* 14(9): 1994. <https://doi.org/10.3390/rs14091994>
- Fernandez-Carrillo, A., Patočka, Z., Dobrovolný, L., Franco-Nieto, A., Revilla-Romero, B., 2020: Monitoring Bark Beetle Forest Damage in Central Europe. A Remote Sensing Approach Validated with Field Data. *Remote Sens.* 12(21): 3634. <https://doi.org/10.3390/rs12213634>
- Gärtner, P., Förster M., Kleinschmit, B., 2016: The benefit of synthetically generated RapidEye and Landsat 8 data fusion time series for riparian forest disturbance monitoring. *Remote Sens. Environ.* 177: 237–247. <https://doi.org/10.1016/j.rse.2016.01.028>
- Gašparović, M., Klobučar, D., Gašparović, I., 2022: Automatic Forest Degradation Monitoring By Remote Sensing Methods And Copernicus Data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2022 XXIV ISPRS Congress, June 6–11, Nice, France, 611–616. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-611-2022>
- Hashim, I.C., Shariff, A.R.M., Bejo, S.K., Muharam, F.M., Ahmad, K., 2021: Machine-Learning Approach Using SAR Data for the Classification of Oil Palm Trees That Are Non-Infected and Infected with the Basal Stem Rot Disease. *Agronomy* 11(3): 532. <https://doi.org/10.3390/agronomy11030532>

- Hornero, A., Hernández-Clemente, R., North, P.R.J., Beck, P.S.A., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020: Monitoring the incidence of *Xylella fastidiosa* infection in olive orchards using ground-based evaluations, airborne imaging spectroscopy and Sentinel-2 time series through 3-D radiative transfer modelling. *Remote Sens. Environ.* 236: 111480. <https://doi.org/10.1016/j.rse.2019.111480>
- Hyndman, R.J., Khandakar, Y., 2008: Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* 27(3): 1–22. <https://doi.org/10.18637/jss.v027.i03>
- Immitzer, M., Atzberger, C., 2014: Early Detection of Bark Beetle Infestation in Norway Spruce (*Picea abies*, L.) using WorldView-2 Data. *PFG* 5: 351–367. <https://doi.org/10.1127/1432-8364/2014/0229>
- Kesavan, R., Muthian, M., Sudalaimuthu, K., Sundarsingh, S., Krishnan, S., 2021: ARIMA modeling for forecasting land surface temperature and determination of urban heat island using remote sensing techniques for Chennai city, India. *Arab. J. Geosci.* 14: 1016. <https://doi.org/10.1007/s12517-021-07351-5>
- Kumbula, S.T., Mafongoya, P., Peerbhay, K.Y., Lottering, R.T., Ismail, R., 2019: Using Sentinel-2 Multispectral Images to Map the Occurrence of the Cossid Moth (*Coryphodema tristis*) in Eucalyptus Nitens Plantations of Mpumalanga, South Africa. *Remote Sens.* 11(3): 278. <https://doi.org/10.3390/rs11030278>
- Lottering, R., Mutanga, O., Peerbhay, K., Lottering, S., 2020: Spatially optimizing vegetation indices integrated with sparse partial least squares regression to detect and map the effects of *Gonipterus scutellatus* on the chlorophyll content of eucalyptus plantations. *Int. J. Remote Sens.* 41(16): 6444–6459. <https://doi.org/10.1080/01431161.2020.1739350>
- Malik, S., Umar, N., 2021: Machine learning-based Prediction of Cotton farming using ARIMA Model. *iRASD. J. Comp. & Info Tech.* 2(1): 26–39. <https://doi.org/10.52131/jcsit.2021.0201.0008>
- Masaki, S., Umeya, K., 1977: Larval life, adaptation and speciation in the fall webworm. In: *Adaptation and Speciation in the Fall Webworm* (Ed. Hidaka, T.), Tokyo Kadan-sha Ltd., 23–27.
- Mutun, S., Ceyhan, Z., Sözen, C., 2009: Invasion by the oak lace bug, *Corythucha arcuata* (Say) (Heteroptera: Tingidae), in Turkey. *Turk. J. Zool.* 33(3): 263–268. <https://doi.org/10.3906/zoo-0806-13>
- Neal, J.W., Schaefer, C.W., 2000: Lace Bugs (Tingidae). In: *Heteroptera of Economic Importance*. (Ed(s): Schaefer, C.W., Panizzi, A.R.). CRC Press., Washington D.C., 85–137.
- Coops, N.C., Johnson, M., Wulder, M.A., White, J.C., 2006: Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sens. Environ.* 103(1): 67–80. <https://doi.org/10.1016/j.rse.2006.03.012>
- Bhattacharai, R., Rahimzadeh-Bajgiran, P., Weiskittel, A., Meneghini, A., MacLean, D.A., 2021: Spruce budworm tree host species distribution and abundance mapping using multi-temporal Sentinel-1 and Sentinel-2 satellite imagery. *ISPRS-J. Photogramm. Remote Sens.* 172: 28–40. <https://doi.org/10.1016/j.isprsjprs.2020.11.023>
- Slavia, A.P., Sutoyo, E., Witarsyah, D., 2019: Hotspots Forecasting Using Autoregressive Integrated Moving Average (ARIMA) for Detecting Forest Fires. In: *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, November 5–7, Bali, Indonesia, 92–97. <https://doi.org/10.1109/IoTaIS47347.2019.8980400>
- Mapuwei, T.W., Bodhlyera, O., Mwambi, H., 2020: Univariate Time Series Analysis of Short-Term Forecasting Horizons Using Artificial Neural Networks: The Case of Public Ambulance Emergency Preparedness. *J. Appl. Math.* 2020: 2408698. <https://doi.org/10.1155/2020/2408698>
- Trubin, A., Mezei, P., Zabihi, K., Surový, P., Jakuš, R., 2022: Northernmost European spruce bark beetle *Ips typographus* outbreak: Modelling tree mortality using remote sensing and climate data. *For. Ecol. Manage.* 505: 119829. <https://doi.org/10.1016/j.foreco.2021.119829>
- Wang, H., Zhao, Y., Pu, R., Zhang, Z., 2016: Mapping Robinia Pseudoacacia forest health conditions by using combined spectral, spatial and textural information extracted from Ikonos imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B8, 1425–1429. <https://doi.org/10.5194/isprs-archives-XLI-B8-1425-2016>
- Wappler, T., 2003: New Fossil Lace Bugs (Heteroptera: Tingidae) from The Middle Eocene of The Grube Messel (Germany), with A Catalog of Fossil Lace Bugs. *Zootaxa* 374: 1–26. <https://doi.org/10.11646/zootaxa.374.1.1>
- Oumar, Z., Mutanga, O., 2013: Using WorldView-2 bands and indices to predict bronze bug (*Thaumastocoris peregrinus*) damage in plantation forests. *Int. J. Remote Sens.* 34(6): 2236–2249. <https://doi.org/10.1080/01431161.2012.743694>
- Zhan, Z., Yu, L., Li, Z., Ren, L., Gao, B., Wang, L., Luo, Y., 2020: Combining GF-2 and Sentinel-2 Images to Detect Tree Mortality Caused by Red Turpentine Beetle during the Early Outbreak Stage in North China. *Forests* 11(2): 172. <https://doi.org/10.3390/f11020172>



Authors' addresses:

Ece Alkan, MSc *
e-mail: ecealkan@duzce.edu.tr
Prof. Abdurrahim Aydin, PhD
e-mail: aaydin@duzce.edu.tr
Duzce University
Faculty of Forestry
Department of Forest Engineering
81620 Duzce
TÜRKIYE

* Corresponding author

Received: March 17, 2023
Accepted: June 21, 2023