

## Prediction of The Spread of *Acacia nilotica* Invasive Species Using Generalized Linear Models

Sahid Agustian Hudjimartsu<sup>1</sup>, Yeni Herdiyeni<sup>1</sup>, Lilik Budi Prasetyo<sup>2</sup>, Agung Siswoyo<sup>3</sup>

<sup>1</sup>Department of Computer Science, Faculty of Math and Natural Sciences, Bogor Agricultural University, Dramaga Campus, Bogor, Indonesia 16680

<sup>2</sup>Department of Forest Conservation and Ecotourism, Faculty of Forestry, Bogor Agricultural University, Dramaga Campus, Bogor, Indonesia, 16680

<sup>3</sup>Bromo Tengger Semeru National Park, The Ministry of Environment and Forestry, East Java, Indonesia, 65125

Received August 30, 2017/Accepted December 23, 2017

### Abstract

*Acacia nilotica* is an invasive species in the Baluran National Park. The spread of *A. nilotica* is a serious threat that can disrupt the function of savanna ecosystems in providing food for herbivores and eventually led to a decrease in wildlife populations of herbivorous. The growth of *A. nilotica* invasive species is influenced by some environmental factors. In order to prevent the spread of *A. nilotica*, a model can be formed to predict the species distribution based on the environmental factors namely generalized linear models (GLM). This research used the data from Siswoyo's research in 2014, which consisted of environmental data and distribution of *A. nilotica*. The results showed that there were five environmental variables that significantly affect on the spread of *A. nilotica* invasive species. The variables included elevation, surface temperature, normalized difference vegetation index, normalized difference moisture index, and the distance from the river, with *p*-value less than 0.05. The accuracy of the prediction model on the spread of *A. nilotica* using GLM was 97.22% and the area under curve (AUC) value was 0.981.

**Keywords:** *Acacia nilotica*, Baluran National Park, GLM, invasive species, savanna

\*Correspondence author, email: shudjimartsu@gmail.com, ph.: +62-81287920066

### Introduction

The Baluran National Park (BNP) is renowned as an illustration of the savanna in Africa and also as a miniature forest because almost all types of Indonesian forests can be found in this national park. The park's natural diversity ranges from tropical rainforest to coral reefs that span from Bama Coast in the eastern region Baluran to Bilik Coast in the northern region Baluran (TN Baluran 2017). This condition makes the park a good habitat for biodiversity, especially for flora.

One of the flora in BNP is *A. nilotica*, which was originally brought in around the year 1969 as a natural firebreak (Sabarno 2002). The characteristics and the widespread distribution of *A. nilotica*, which was originally used as a protective savanna, make *A. nilotica* as an invasive species. The invasion of *A. nilotica* can disrupt the function of savanna ecosystems in providing food for herbivores and eventually led to a decrease in wildlife population of herbivorous (Caesariantika *et al.* 2011).

The spread of *A. nilotica* is influenced by some environmental factors (Kriticos *et al.* 2003). The relationship between the environmental factors and the species referred to as an ecology, which can be analyzed by using species distribution models (SDM). Other scholar use species niche

model, ecological niche model or niche theory model terminology (Franklin 2009). The SDM provides a way to analyze the geographical influence of environmental factors on species (Segurado & Araújo 2004) by statistical or machine learning methods.

There are various techniques to develop SDM such as linear regression (Rahmat *et al.* 2012), multivariate (principal component analysis/PCA) (Prayogo *et al.* 2014; Favata *et al.* 2015; Prasetyo *et al.* 2017), maximum entropy (maxent) (Bowler 2014), or generalized linear models (GLM) (Guisan *et al.* 2002; Ulrich *et al.* 2016). The GLMs are alternative to the normal linear model, particularly when the independent variable does not follow normal distribution. GLM is an extensions of linear models that can cope with non-normal distributions of species (Venables & Ripley 1999). The GLM approach can also analyze how important the relationship between environment factors with the distribution of species (Guisan *et al.* 2002).

Several studies have considered GLMs in the prediction of species distribution. For example, Guisan *et al.* (1998) predict the distribution of plant species *Carex curvula* in the Swiss mountains. In the study, many predictors (environment factors) were used, and the authors then employ the PCA to reduce the dimensions before being

analyzed by GLM. Furthermore, Uguru et al. (2014), also consider GLM to predicte the ingredient of protein in *A. nilotica* pods for goats feeding. Siswoyo (2014) also studies the habitat suitability of the species *A. nilotica* in the BNP using logistic regression.

The spread of invasive *A. nilotica*, which threatens the ecosystems of savannah, needs to be monitored by predicting its spread. The objective of this research is to predict of spread of *A. nilotica* invasive species by analyzing the environment factors and how to know the correlation between the variable in the environment factors using GLM. For future work, this predictive modeling can be used as an initial step in maintaining existing ecosystem in BNP especially savanna ecosystems.

## Methods

Several stages in predicting the spread of *A. nilotica* in BNP consists of: data collection, multicollinearity test, data division, GLM prediction modelling, model, and results testing. The tools used in this research are the application of R for statistical analysis by GLM and Quantum GIS for spatial analysis. The flow chart of this research can be seen in Figure 1.

**Data collection** This research used data of Siswoyo (2014) in BNP Situbondo, East Java Province. The data consists of environment data and distribution data of *A. nilotica*. There are six environmental data: elevation, slope, NDVI, NDMI, temperature, and distance from the river. These six factors are spatial data representing primary environmental regimes for the analysis of species distribution models (Franklin 2009) where as the distribution data are presence and absence of *A. nilotica*.

**Data pre-processing** Data pre-processing is divided into three following phases:

### 1 Data extraction

To extract the value of environmental and distribution data, we use overlay analysis. The illustration of overlay analysis can be seen in Figure 2.

### 2 Multicollinearity testing between predictor variables (environment data)

Multicollinearity is a condition where two or more predictor variables highly correlated with one and another. As a result, it is hard to find good estimates of their distinct effects on dependent variables (Midi et al. 2010). Multicollinearity testing between predictor variables can be solved by examining the variance inflation factor (VIF). The VIF value is based on the calculation of tolerance in which if the tolerance value close to zero the variables have a high multicollinearity. Otherwise, if the tolerance value close to one the variables have a very small multicollinearity (Sunaryo et al. 2015).

$$Tolerance = Tol = 1 - R^2 \quad [1]$$

note:  $R^2$  is the determinant coefficient of the predictor variables.

$$VIF = \frac{1}{Tol} = \frac{1}{1 - R^2} \quad [2]$$

### 3 Data division

The data is divided into training data and testing data with K-fold cross validation. Training data was used to analyze the model equations while the testing data was used to test the model equations. K-fold cross validation is a method used to evaluate the learning algorithm by dividing the data into K-fold, as K-1 fold is used as training data and 1 fold as the testing data (Liu & Özsü 2009). Data division with K-fold cross-validation can be illustrated as in Figure 3.

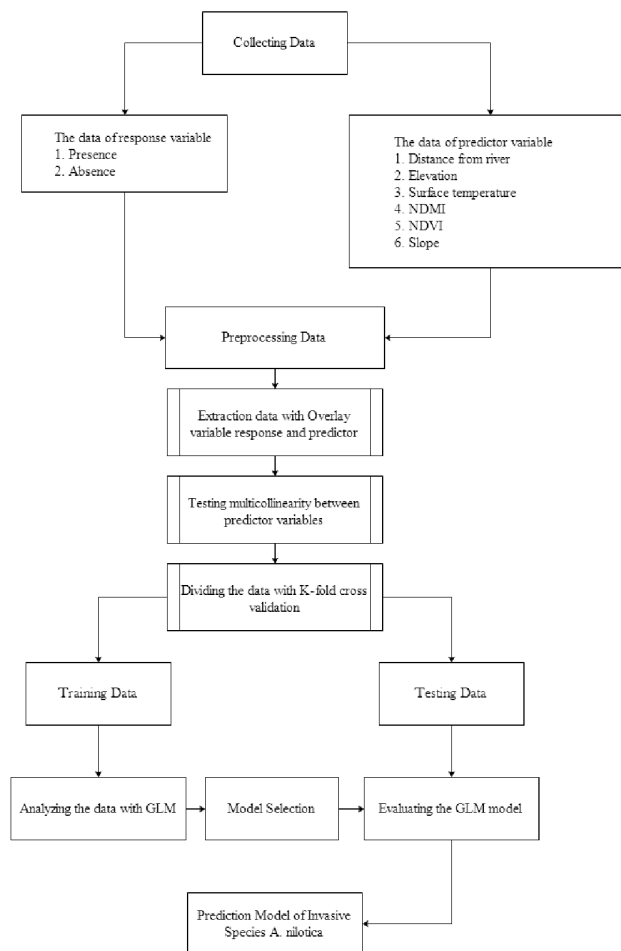


Figure 1 Overlay and extraction data variable response and predictor.

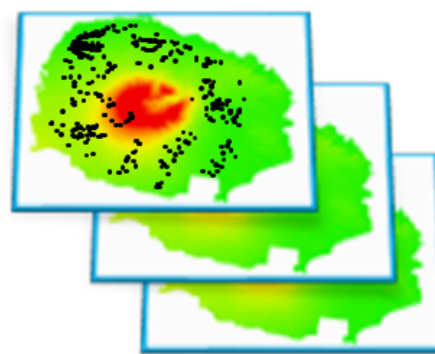


Figure 2 Research flow.

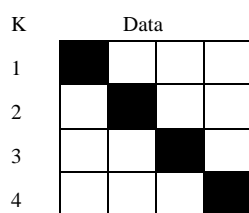


Figure 3 Dividing data with K fold cross validation. Testing data (■), training data (□).

**Generalized linear models (GLMs)** Regression analysis is a modeling with a statistical approach that determines the relationship between variable response (influenced) and variable predictor (cause). The regression approach is theoretically capable of delivering good results when the response variables are normally distributed and the variance of the data is unchanged. One type of regression model is the GLM that can resolve when the response variable is not normally distributed and the variance of data changes (Guisan & Zimmermann 2000).

Ecological data sometimes violates the rules of the linear model assumptions. Venables and Ripley (2013) state that GLM is the development of linear regression, which can overcome non-distributed data on response variables and transform into linearity, which cannot be done by ordinary linear regression. The GLM can be expressed as Equation [3]:

$$g(E(Y)) = LP = \beta_0 + \sum_{j=1}^p x_j \beta_j + \varepsilon \quad [3]$$

note:  $g$  = link function,  $E$  = expected value,  $Y$  = response variable,  $x$  = predictor variable,  $\beta$  = vector of estimated coefficients (slope),  $\beta_0$  = estimated constant (intercept),  $\varepsilon$  = epsilon.

By GLM, the combination of the predictor variable produces linear predictor ( $LP$ ) and the expected value of the  $Y$  variable correlates with the  $LP$  through the  $g()$  as connecting function. The connecting function is described as the mean value of the  $Y$  variable, based on the  $LP$  or the expected exponential link from  $Y$  to the predictor variable. For example, when  $Y$  response is binary (0 and 1), then the distribution response ( $Y$ ) is binomial and logit link is used for the connecting function. This connecting function is often used in species distribution modeling (SDM) because generally data of species emergence is binomial-distributed data. The GLM with binomial distribution is often called logistic regression (Franklin 2009). The connecting function on the binomial-distributed GLM, with the logit link connecting function is shown by Equation [4].

$$LP = \log\left(\frac{\mu}{1-\mu}\right) \quad [4]$$

note:  $LP$  = Linear Predictor,  $\mu = E(Y)$

In logistic regression we are often interested in the predicted probability of a class, where  $\mu = E(Y)$  is the

probability of class 1 (presence), and  $(1 - \mu)$  is the probability of class 0 (absence). so by inverting the above equation we can solve for  $\mu$  as shown as Equation [5].

$$\mu = \left(\frac{e^{LP}}{1+e^{LP}}\right) \quad [5]$$

**Model estimation** In linear regression for estimating unknown parameter the method used is least square when variable  $Y$  normal distributed. Unfortunately, the method cannot be applied to estimate unknown parameter because variable  $Y$  is not normally distributed or dichotomous outcome. The estimation method that leads to the least squares function under the linear regression model is maximum likelihood. The method can provide the foundation for approach to estimate with regression logistic (Hosmer & Lemeshow 2000). It is required to maximizing the log-likelihood for estimating the parameters, because the parameter allegedly not linear and usually solved by an iteratively reweighted least squares (IRLS) algorithm (Franklin 2009).

**Model selection** Model selection can also be referred to as a selection variable is a model estimated to use a subset of different predictor variables and compared with some criteria (Franklin 2009). One of them uses the AIC value (Akaike Information Criterion), which is a statistical approach that is often used in model selection (Venables & Ripley 1999).

The equation of AIC is given by Equation [6]:

$$AIC = -2\text{maximum log likelihood} + 2p \quad [6]$$

Where  $p$  is the number of predictor variables, the smaller AIC value the better the model. There are several ways in the selection model, including:

- 1 backward elimination is removing the insignificant predictor variable in the model.
- 2 forward Selection is adding significant predictor variables to the model.
- 3 stepwise procedures are automatically using forward and backward in the predictor variable selection strategy.

Those ways are important for searching pommonious model that uses an appropriate subset all available predictor variables. Essentially, variety models are estimated using different subsets of predictors and compared with AIC, in terms of their fit to the data (Franklin 2009).

**Model evaluation** The classification table is most proper when classifying a purpose in the analysis (Table 1). Classification table is two-way frequency tables between the actual data and prediction data estimation results of the model based on testing data (Hosmer & Lemeshow 2000).

Receiver operating characteristic (ROC) is a method based on sensitivity and specificity. Sensitivity explains how well the model predicts presence, whereas the specificity explains how well the model predicts absences (Baldwin 2009). ROC curve is a plot between the probability of false positive (1-specificity) and true positive (sensitivity). The area under the ROC curve is called Area Under the Curve (AUC), ranging 0–1 and this value indicates the performance of model to classify the data. The AUC values

of 0.5–0.7 are considered poor (poor model performance), values of 0.7–0.9 are considered to be good enough, and > 0.9 indicates excellent accuracy (Manel *et al.* 2001).

## Results and Discussion

Model prediction for spread of *Acacia nilotica* invasive species In this study, prediction models were analyzed statistically using the GLM. Before the data were analyzed by GLM, we did data pre-processing on the existing data. The results of the preprocessing data such as data extraction between the response variable and predictor variables can be seen on Table 2.

After the extraction data, the next step is multicollinearity testing using VIF to determine whether there is collinearity between variables (Table 3). The results of multicollinearity test showed VIF value was lower than 10. If the VIF value is greater than 10, it indicates a problem multicollinearity between predictor variables (Sunaryo *et al.* 2015).

This research conducted two experiments with K values to obtain better results in the distribution of the training data and testing data. Based on the results of experiment K value on K-fold cross validation between K = 5 and K = 10, it was found that the number of K = 10 has better accuracy compared with the number of K = 5. The results of data division with K = 10 in cross-validation showed that there are 324 training data and 36 testing data.

The training data were used as a data for modeling in GLM. The results GLM modelling are used to analyze the relationship between response (influenced) with the variable predictor (cause) and model selection as shown in Table 4.

Table 4 shows that there are several models in analyzing and selecting models using GLM.

- 1 model 1, analyzing the data using all predictor variables.
- 2 model 2, analyzing the data with backward elimination to eliminate the insignificant predictor variables in the model.
- 3 model 3, analyzing the data with Forward Selection to add significant predictor variables into the model
- 4 model 4, analyzing the data using Stepwise procedures are automatically using the forward and backward in the predictor variable selection strategy.

According to analysis and model selection, there are two models that have the smallest AIC value, namely Model 2 and Model 4. Furthermore, model selection can use Model 2 or Model 4 as the best model because both of these models had the smallest AIC value of 228.85. Model 4 was selected in this study, because the stage process of model selection has been through backward elimination process in Model 2, while Model 4 was implemented all the process in model selection. Model with smaller AIC values has significant predictor variables to response variable. In addition, it was obtained Nagelkerke R<sup>2</sup> value of 0.680 or 68.0% for the 5 selected predictor variables. This indicated that through 5

selected predictor variables, we can illustrate the variance of response variables.

**Data analysis using GLM** The result of predictive model showed that the distribution of *A. nilotica* with Model 4 resulted in significance and coefficient of each variable predictor (Table 5). As seen on Table 5, a total of 5 parameters were significant due to p-values less than 0.05, which means the model used has a confidence level of 95%. According to the result, these 5 predictor variables can affect the invasive distribution of *A. nilotica* species in TN Baluran. Furthermore, the probability equation of predictive model of the invasive species distribution of *A. nilotica* with GLM is shown in Equation [7].

$$P_i = \frac{e^{(-1.853 - 0.024_{x_1} + 17.462_{x_2} - 25.434_{x_3} - 0.002_{x_4} + 0.478_{x_5})}}{1 + e^{(-1.853 - 0.024_{x_1} + 17.462_{x_2} - 25.434_{x_3} - 0.002_{x_4} + 0.478_{x_5})}} \quad [7]$$

note:  $x_1$  = elevation,  $x_2$  = normalized difference moisture index,  $x_3$  = normalized difference vegetation index,  $x_4$  = Distance from river,  $x_5$  = surface temperature, e = exponential (2.718282).

By Equation [7], the coefficients of each predictor variables explain the relationship between predictor variable to the response variable. The negative coefficient has a negative relation to the response variable, thus the less predictor variable value, the higher probability of distribution of *A. nilotica*. Conversely, the positive value coefficient has a positive relationship to the response variable, if the value of the predictor variable increases, the higher probability of distribution of *A. nilotica*.

Distribution of invasive species *A. nilotica* based on elevation. Invasive distribution of *A. nilotica* species is spread at an elevation of 0–100 m above sea level (lowland) and a small portion is at an elevation of 100–200 m above sea level (Figure 4a). Based on the analysis, the high probability of invasive distribution of species *A. nilotica* located in the lowlands. According Sabarno (2002), the lowlands of BNP have alluvial soil and black soil types. Alluvial and black soil types are excellent for growing *A. nilotica* (Bargali & Bargali 2009).

### The invasive spread of *A. nilotica* species based on NDMI

NDMI is the index value derived from remote sensing data related to liquid water. NDMI index value greater than 0.1 is a signal of high humidity (moisture), while a low value (close to -1) explains the low humidity levels (Herbei *et al.* 2012). According to Figure 4b, it can be seen the distribution of *A. nilotica* by NDMI contained of the index value of 0.120–0.211. Its means the area has a moderate humidity. According to Kriticos *et al.* (1999), the growth of *A. nilotica* is influenced by moisture availability.

### The invasive spread of *A. nilotica* species based on NDVI

NDVI is the index value used to measure the greenness of vegetation. Higher NDVI values (close to 1), means a very good green vegetation canopy level, whereas an index value less than 0, is generally a body of water or area without vegetation (Prayogo *et al.* 2014). Map of distribution of *A. nilotica* based on NDVI can be seen in Figure 5a. Figure 5a

Table 1 Classification Table

Predicted	Observed	
	1	0
1	True Positive (sensitivity)	False Positive
0	False Negative	True Negative (Specificity)

Table 2 Result of the extraction data

Variable response (Y)	Elevation (masl)*	Variable predictor (x)				
		NDMI	NDVI	Distance from river (m)	Slope (°)	Surface Temperature (°C)
1	27.00	0.11	0.33	201.25	4.62	24.79
1	9.00	0.11	0.30	494.77	2.92	24.85
1	59.00	0.11	0.32	180.00	9.16	27.20
1	23.00	0.17	0.28	127.28	10.00	22.02
1	17.00	0.21	0.37	757.17	5.40	21.99
0	239.00	0.24	0.39	270.00	5.28	21.52
0	537.00	0.16	0.16	30.00	14.16	19.10
0	582.00	0.16	0.16	42.43	24.10	18.05
0	639.00	0.15	0.16	335.41	19.46	16.13
0	786.00	0.09	0.08	488.36	36.05	13.41

Note: 1= Presence, 0 = absence

Table 3 Collinearity between variable predictor

Variable predictor	Collinearity	
	Tolerance	VIF
Elevation	0.353	2.834
NDMI	0.291	3.438
NDVI	0.221	4.529
Distance from river	0.935	1.069
Surface temperature	0.505	1.980
Slope	0.573	1.745

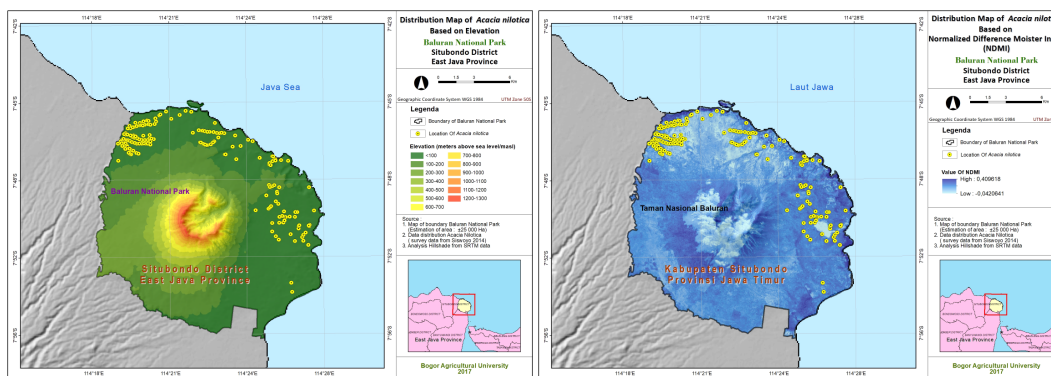
Table 4 Result of GLM analysis and selection model

Model	Total variable	AIC	Nagelkerke R <sup>2</sup>
Model 1	6	230.82	0.680
Model 2	5	228.85	0.680
Model 3	3	231.01	0.667
Model 4	5	228.85	0.680

Table 5 The coefficient and significance of each predictor variable

	Estimate	Std. Error	z value	Pr(> z )	
Intercept	-1.853	2.583	-0.717	0.473	
Elevation	-0.024	-0.004	-6.706	0.000	***
NDMI	17.462	8.593	2.032	0.042	*
NDVI	-25.434	10.069	-2.379	0.017	*
Distance from river	-0.002	0.000	-2.606	0.009	**
Surface temperature	0.478	-0.092	5.200	0.000	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



(a)

(b)

Figure 4 Distribution map of *Acacia nilotica* based on elevation (a) and NDMI (b).

shows that the distribution of *A. nilotica* ranging 0.240–0.400. The NDVI value shows that area are savanna, shrubs, and secondary forest in BNP. The NDVI value indicates the high chance of invasive distribution of *A. nilotica* when grown in savanna and shrubs. In addition, based on the results of the survey (Siswoyo 2014) the distribution of *A. nilotica* is in some spot (spot) of savanna and shrub area.

**The invasive spread of *A. nilotica* species based on distance from rivers** Figure 5b shows *A. nilotica* can be found around riparian zones. Most of the *A. nilotica* were found in distance of 0–400 m from the river. According to (Bargali & Bargali 2009), *A. nilotica* can grow and develop properly while located near water sources. The results showed that the invasive distribution of *A. nilotica* increased

as it grew or was near the border area of the river. This was reiterated by Radford *et al.* (2002) death of all plants, including *A. nilotica* is due to the location is out of riparian zones.

**The invasive spread of *A. nilotica* species based on surface temperature** Invasive species *A. nilotica* are uniformly distributed at surface temperatures of 18–27 °C as shown in Figure 6. The results show that the probability of distribution of *A. nilotica* species increases as they grow or are at higher surface temperatures. According to Kriticos *et al.* (2003) the overall species of *A. nilotica* can adapt and grow well at temperatures from 16–50 °C.

**Model evaluation** Evaluation predictive model of the spread

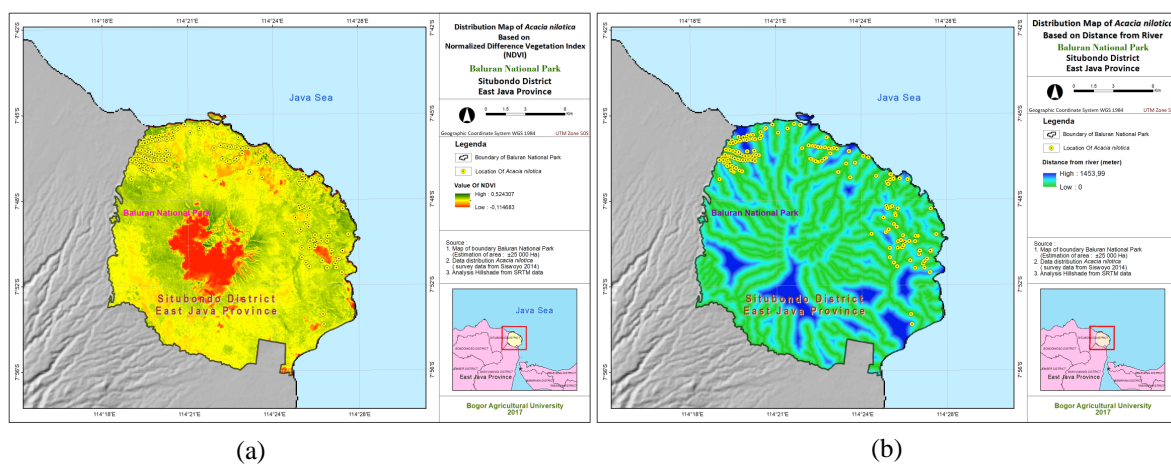


Figure 5 Distribution map of *Acacia nilotica* based on NDVI (a) and distance from river (b).

Table 6 Classification table based on GLM

Prediction	Observed	
	1 (presence)	0 (absence)
1 (presence)	20	1
0 (Absence)	0	15

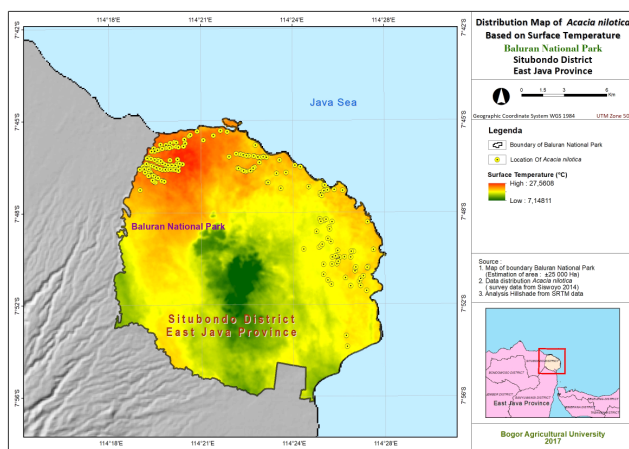


Figure 6 Distribution map of *Acacia nilotica* based on surface temperature.

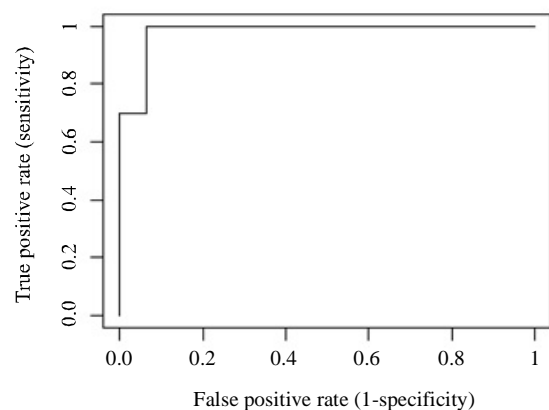


Figure 7 ROC curve.

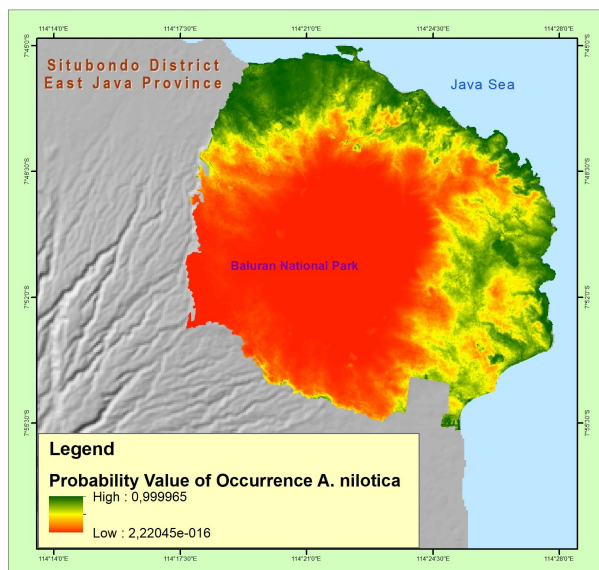


Figure 8 The result predictive modeling with GLM.

of *A. nilotica* species in BNP using GLM was done to find out how proper the model was able to predict the spread of *A. nilotica*. Model evaluation used data testing, started from data division with K-fold cross validation. Test results can be seen in the table of conformity classification (Table 6). based on the classification table, it was obtained accuracy for class occurrence: a presence of 95.24%, an absence of 100%, and overall accuracy of the model was 97.22%. This proves that this model is very good, thus it is can be used as a predictive model of the spread of *A. nilotica* invasive species in BNP. In addition, the classification table can also calculate AUC from the ROC curve (Figure 7).

The area under ROC curve or AUC was used as a measure performance of the classifier and the result AUC value of 0.981. This indicates that the model have a good performance in discriminating between the two categories. The discrimination of two categories is based on thresholding in which this research used spec sens. That was calculated from sum of the highest sensitivity and specificity.

The calculation of the spread *A. nilotica* invasive species using GLM prediction model revealed, that there are about an area of 7.759 ha which is predicted as a site to grow *A. nilotica* invasive species. *A. nilotica* invasive is allegedly spread from northern, eastern, and southern part of BNP. The model prediction map of the spread *A. nilotica* invasive species by GLM is shown in Figure 8.

## Conclusion

Model predictions of the spread *A. nilotica* invasive species in BNP using GLM are very good with the accuracy of the model is 97.22%. Moreover, performance (AUC) value in discriminant between the two categories (presence and absence) of the occurrence of *A. nilotica* invasive species is 0.981. Environmental or predictor variables that significantly affect the response variable are: elevation, surface temperature, NDVI, NDMI, and the distance from the river. These 5 variables have p-value less than 0.05, which means that the model used has a confidence level of 95%.

## References

- Baldwin RA. 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11:854–866. <https://doi.org/10.3390/e11040854>.
- Bargali K, Bargali S. 2009. *Acacia nilotica*: a multipurpose leguminous plant. *Nature and Science* 7(4):11–19.
- Bowler M. 2014. Species abundance distributions, statistical mechanics and the priors of MaxEnt. *Theoretical Population Biology* 92:69–77. <https://doi.org/10.1016/j.tpb.2013.12.002>.
- Caesariantika E, Kondo T, Nakagoshi N. 2011. Impact of *Acacia nilotica* (L.) Willd. ex Del invasion on plant species diversity in the Bekol Savanna, Baluran National Park, East Java, Indonesia. *Tropics* 20:45–53. <https://doi.org/10.3759/tropics.20.45>.
- Favata CA, Christensen DR, Thompson R, McKeown KA, Hanselman JA. 2015. Evaluation of a modified habitat suitability index model for eastern brook trout: Implications for efficient habitat assessment. *Journal of Student Research* 4:90–98.
- Franklin J 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge: Cambridge University Press.
- Guisan A, Edwards TC, Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. *Ecological modelling* 157:89–100. [https://doi.org/10.1016/S0304-3800\(02\)00204-1](https://doi.org/10.1016/S0304-3800(02)00204-1).
- Guisan A, Theurillat JP, Kienast F. 1998. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science* 9:65–74. <https://doi.org/10.2307/3237224>.
- Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological modelling* 135:147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9).
- Herbei MV, Dragomir LO, Oncia S 2012. Using Satellite Images Landsat TM for Calculating Normalized Difference Indexes for The Landscape of Parâng Mountains. University 1 Decembrie 1918 of Alba Iulia.
- Hosmer DW, Lemeshow S. 2000. *Introduction to the Logistic Regression Model: Applied Logistic Regression, Second Edition*. New York: John Wiley & Sons.
- Kriticos D, Sutherst R, Brown J, Adkins S, Maywald G. 2003. Climate change and the potential distribution of an invasive alien plant: *Acacia nilotica* ssp. *indica* in Australia. *Journal of Applied Ecology* 40:111–124.
- Liu L, Özsu MT. 2009. *Encyclopedia of database systems*, Springer Berlin, Heidelberg, Germany.

regression model. *Applied Logistic Regression, Second Edition*. New York: John Wiley & Sons.

Kriticos D, Sutherst R, Brown J, Adkins S, Maywald G. 2003. Climate change and the potential distribution of an invasive alien plant: *Acacia nilotica* ssp. indica in Australia. *Journal of Applied Ecology* 40:111–124.

Liu L , Özsu MT 2009. *Encyclopedia of database systems*. Heidelberg: Springer Berlin, Heidelberg, Germany.

Manel S, Williams HC, Ormerod SJ. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921–931. <https://dx.doi.org/10.1046/j.1365-2664.2001.00647.x>.

Midi H, Sarkar SK, Rana S. 2010. Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics* 13:253–267. <https://doi.org/10.1080/09720502.2010.10700699>.

Prasetyo LB, Supartono T, Kartono AP, Hikmat A, Ramdhoni S. 2017. Habitat Suitability Index (HIS) of Surili (*Presbytis comata* Desmarest, 1822) in mixed forest of Kuningan District, West Java-Indonesia. In: IOP Conference Series: Earth and Environmental Science; Bogor, October 25–26, 2016. IOP Publishing, 012061.

Prayogo H, Thohari AM, Solihin DD, Prasetyo LB , Sugardjito J. 2014. Habitat suitability modeling of bornean orangutan (*Pongo pygmaeus pygmaeus*) in Betung Kerihun National Park, Danau Sentarum and Corridor, West Kalimantan. *Jurnal Manajemen Hutan Tropika* 20(2):112–120. <https://doi.org/10.7226/jtfm.20.2.112-120>.

Radford IJ, Nicholas M, Tiver F, Brown J , Kriticos D. 2002. Seedling establishment, mortality, tree growth rates and vigour of *Acacia nilotica* in different *Astrelba* grassland habitats: Implications for invasion. *Austral Ecology* 27:258–268. <https://doi.org/10.1046/j.1442-9993.2002.01176.x>.

Rahmat UM, Santosa Y, Prasetyo LB, Kartono AP. 2012.

Habitat suitability modeling of javan rhino (*Rhinoceros sondaicus* Desmarest 1822) Ujung Kulon National Park. *Jurnal Manajemen Hutan Tropika* 18:129–137. <https://doi.org/10.7226/jtfm.18.2.129>.

Sabarno M. 2002. Savana Taman Nasional Baluran. *Biodiversitas* 3:207–212. <https://doi.org/10.13057/biodiv/d030107>.

Segurado P, Araújo MB. 2004. An evaluation of methods for modelling species distributions. *Biogeography* 31:1555–1568. <https://doi.org/10.1111/j.1365-2699.2004.01076.x>.

Siswoyo A. 2014. Pemodelan spasial kesesuaian habitat akasia berduri (*Acacia nilotica*) di Taman Nasional Baluran [thesis]. Bogor: Bogor Agricultural University.

Sunaryo S, Setiawan S, Siagian TH. 2015. Mengatasi masalah multikolinearitas dan outlier dengan pendekatan ROBPCA (Studi kasus analisis regresi angka kematian bayi di Jawa Timur). *Jurnal Matematika, Sains, dan Teknologi* 12:1–10.

Taman Nasional Baluran 2017. Informasi Flora dan Fauna di TN Baluran. [7 February 2017].

Uguru C, Lakpini C, Akpa G , Bawa G. 2014. Nutritional potential of acacia (*Acacia nilotica* (L.) Del.) pods for growing red sokoto goats. *IOSR Journal of Agriculture and Veterinary Science* 7:43–49.

Ulrich W, Soliveres S, Thomas AD, Dougill AJ, Maestre FT. 2016. Environmental correlates of species rank-abundance distributions in global drylands. *Perspectives in Plant Ecology, Evolution and Systematics* 20:56–64. <https://doi.org/10.1016/j.ppees.2016.04.004>.

Venables WN, Ripley BD. 1999. *Modern Applied Statistics with S-PLUS*. New York: Springer Science & Business Media.

Venables WN, Ripley BD. 2013. *Modern Applied Statistics with S-PLUS*. New York: Springer Science & Business Media.