

The potential impact of climate change on the distribution pattern of *Eusideroxylon zwageri* (Bornean Ironwood) in Kalimantan, Indonesia

ANGGA YUDAPUTRA^{1,✉}, IZU FIJRIDIYANTO¹, WENDELL P. CROPPER, JR.²

¹Research Center of Plant Conservation and Botanic Gardens (Bogor Botanic Gardens). Jl. Ir. H. Djuanda No.13, Bogor 16122, West Java, Indonesia. Tel./fax.: +62-251-8311362, 8336871, ✉email: angg020@lipi.go.id

²School of Forest Resources and Conservation, University of Florida. Gainesville, FL 32603, USA

Manuscript received: 1 November 2019. Revision accepted: 29 December 2019.

Abstract. Yudaputra A, Fijridiyanto I, Cropper WP Jr. 2020. The potential impact of climate change on the distribution pattern of *Eusideroxylon zwageri* (Bornean Ironwood) in Kalimantan, Indonesia. *Biodiversitas* 21: 326-333. *Eusideroxylon zwageri* Teijsm & Binn. is a vulnerable tree species with considerable economic value. The high demand for its wood makes this species vulnerable. Population vulnerabilities of *E. zwageri* include habitat loss, land-use change, and forest encroachment. An additional potential risk factor is climate change, with a possible increase in temperature of about 2.5 to 10 degrees Fahrenheit over the next century. Climate is considered to be a principal factor that determines the distribution of many species. This study addresses the potential current and future distribution of *E. zwageri* under climate change. Seven predictor climate variables are selected from 19 climatic variables using VIF (Variance Inflation Factor) to eliminate multicollinearity among variables. The spatial data is prepared using Geographic Information System (GIS). Six species distribution models (RF, SVM, MARS, GAM, GLM) and the ensemble model are applied to understand the potential current geographical distribution of *E. zwageri*. For risk assessment, the potential future distribution is predicted using the ensemble model only. All models are run using R open-source software. In model evaluation, all models have AUC value >0.80, indicating those models are good predictive models. All predictive models have the TSS >0.60 which means those models having a useful agreement between prediction and real observation. Precipitation seasonality, isothermality, and precipitation of the coldest quarter are the most important model variables that influence the current and future distribution of *E. zwageri*. Four models (RF, SVM, GAM, GLM) produce similar predictive maps of potential current distribution. MARS produces a slightly different predictive map. The future projection of ensemble model shows that the distribution area is more likely shifted and decreased from the current to 2050 and 2070.

Keywords: Climate change, *Eusideroxylon zwageri*, Kalimantan, Species Distribution Modelling

INTRODUCTION

Eusideroxylon zwageri Teijsm & Binn. is a rare timber tree and native to Indonesia, Brunei, Malaysia and Philippines region. *E. zwageri* is relatively large and at maturity can reach 50 m in height, with diameters around 120 cm. Moreover, it can reach 1000 years old with a huge crown. In Indonesia, the distribution of *E. zwageri* is widespread across Java, Sumatra, and Kalimantan. It grows well in lowland and secondary forest up to 625 m asl. Areas along rivers and in adjacent hills are suitable habitat for this tree. Average annual rainfall of 2,500-4,000 mm is needed to meet the growth requirement of this tree (Ecocrop 2019). The wood of this tree is highly prized in the international timber industry because it is one of the most durable and heaviest timbers in the world. Due to high timber demands in national or international markets, the population of *E. zwageri* is significantly declining in its natural habitat. The loss of its population was caused by illegal smuggling activity, habitat loss, land-use change and climate change (The Ministry of Forestry Republic of Indonesia 2009). Moreover, the conservation status of *E. zwageri* is vulnerable based on IUCN Redlist (Asian Regional Workshop 1998).

Human disturbances (land-use change, illegal smuggling, habitat lost) might impact the future distribution and

existence of this tree. Climate change is one of factors that drives distribution patterns of biodiversity. Over the past 30 years, climate change has likely caused numerous range shifts and has been implicated in one species-level extinction. Loss of biodiversity and the associated loss of ecosystem services is a grave concern, but better tools are needed to understand risks over large areas into the future. The modeled projection of species distributions for future climate scenarios would be useful to help assess the extinction risk of species throughout the terrestrial landscape (Parmesan and Yohe 2003). Future distribution is estimated using future climate scenarios (Beaumont et al. 2002).

Species Distribution Models (SDMs) are widely used methods in ecology for both theoretical and applied research. SDMs are often used to understand the drivers of global biodiversity patterns and conservation policy (Sinclair et al. 2010; Sinka et al. 2010). SDMs are the implementation of mathematical tools calibrated with observations of occurrence records and associated physical environments (topography, soil, and climate). These tools can be then used to analyze the distribution pattern across landscape (Elith and Leathwick 2009). Many statistical models or machine learning approaches have been adopted for use in SDMs to predict the spatial patterns of global biodiversity. The mathematical approached range from the relatively simple application of methods such as logistic

regression to more complex techniques like machine learning (Elith and Leathwick 2009).

Generalized Additive Model (GAM) is very flexible and therefore useful when the relationship between the variables becomes more complex, in cases not easily fitted with a standard parametric function like the General Linear Model (GLM) (Hastie and Tibshirani 1990). GLM is often used to determine the distribution pattern of species by using the iterative weighted linear regression technique to gain the estimated maximum likelihood of the parameters. The study of predicted areas on eight different species in Australia showed that GLM output produced a better performance in comparison with Boosted Regression Tree (BRT) (Shabani et al. 2016). Random Forest (RF) algorithm belongs to the most powerful machine learning algorithms in common use (Elith et al. 2006; Wisz et al. 2008) and suitable for obtaining powerful ensemble models (Araújo and New 2007). The implementation of RF on SDMs has shown to have a higher accuracy of model than a simple decision tree and other models (Prasad et al. 2006; Cutler et al. 2007). In a recent study, RF is considered to be the best predictive model compared to other predictive models such as GLM, ANN, GAM, and MARS (Yudaputra et al. 2019).

MARS (Multivariate adaptive regression splines) has been applied in a variety of ecological modeling projects (Leathwick et al. 2006; Elith and Leathwick 2007). MARS is relatively simple, it can be thought of as combining classical linear regression, mathematical construction of splines and binary recursive partitioning. MARS is a powerful method to determine the relationship between response and predictor variables either linear or non-linear (Hastie et al. 2001). MARS and MaxEnt are modeling tools that have been found to exhibit good performance (Leathwick et al. 2006). MARS or MaxEnt often give better results than profile techniques (Mateo et al. 2010). SVM (Support Vector Machine) is a machine learning that belongs to a family of generalized linear classifiers. The principle of SVM is a discriminative classifier defined by a separating hyperplane (a line that dividing a plane into two parts where each class lay on either side). The SVM is a reasonable model to use when in 1 sample information is limited (Drake et al. 2006). SVM is widely used in species conservation and academic research (Tsoar et al. 2007). In a recent study modeling the spatial distribution of *Guettarda speciosa* in Indonesia, the mean value of AUC and Kappa (measures of model fit and predictive capacity) from SVM are significantly higher than those from BIOCLIM and DOMAIN (Yudaputra et al. 2019). The ensemble model is relatively new in SDMs, the rationale of ensemble model is that better predictions will be obtained by averaging the weight of all single models. An advantage of using this model is that the prediction for all single models can be retained and documented (Burnham and Anderson 2002). The ensemble model is relatively working well for temporal and spatial projection of SDMs (Thuiller 2004).

The K fold cross-validation is often used to promote the generality and robustness of models. The data can be split into k times yielding k estimates of accuracy that can be averaged (Dormann et al. 2008). The formula of splitting

data should be $1 / (1 + \sqrt[p]{p} - 1)$, where p is the number of predictors. For example, if there are two predictors, the train : test ratio should be 50 : 50, if there are five, it should be 67 : 33, and if there are many (>10), it should be 75 : 25 (Fielding and Bell 1997). In a recent SDM study, splitting the data with the ratio 75 : 25 produced the highest AUC value although it should be noted that the AUC value is only slightly different from other splitting ratios (63 : 33 and 80:20) (Yudaputra et al. 2019).

The evaluation dataset generated by the split-sample approach can be used to assess the accuracy of the model. Model accuracy can be determined using two measures, including Area Under the Curve values (AUC) of receiver operator characteristic (ROC) curves and True Skill Statistics (TSS). The area under the receiver operating characteristic (ROC) curve (AUC) often used for measuring the discriminatory capacity of classification models on species distribution models (SDMs). Recently, the AUC is considered as a standard in measuring the accuracy of species distribution models (Fielding and Bell 1997; Lobo et al. 2008). A model will be considered to discriminate better than chance if the AUC is higher than 0.5 when the curve lies above the diagonal of no discrimination (Krzanowski and Hand 2009). The score of AUC can represent the categories of predictive model performance, model with AUC in range 0.9-1 (excellent), 0.8-0.9 (good), 0.7-0.8 (fair), 0.6-0.7 (poor), and 0.5-0.6 (fail) (Krzanowski and Hand 2009).

True Skill Statistics (TSS) is widely used to evaluate model predictions of Species Distribution Models (SDMs). TSS statistic was often referred to as "Pierce skill score" (Stephenson 2000). The range of TSS from -1 to +1, where +1 indicates perfect agreement and values of zero or less indicates a performance no better than random. TSS statistic has been highly recommended for measuring the performance of predictive maps generated by presence-absence models. TSS statistics can be interpreted in several categories: values < 0.4 were poor, 0.4-0.8 useful, and > 0.8 good to excellent (Allouche et al. 2006).

MATERIALS AND METHODS

The occurrence records of *Eusideroxylon zwageri* are collected from virtual herbarium data (provided by GBIF and Naturalist), Bogoriense' herbarium database, and scientific papers that included information about the occurrence locations of *E. zwageri*. The occurrence records of *E. zwageri* indicate that it is only found in Kalimantan and Sumatra (specifically in Jambi), but not found in other regions in Indonesia. In this study, we only use the occurrence records from Kalimantan due to the poor availability of information about specific occurrence locations in Sumatra. The occurrence records only consist of presence records, simulated absence records are derived by randomly selecting locations where the species is not found (absence-background data); the absence-background is selected using R-GIS programming. Overall, 43 points of occurrence are recorded throughout Kalimantan. Additionally, 100 absence occurrence records are derived from randomly selected in study area (100 pseudo absence data).

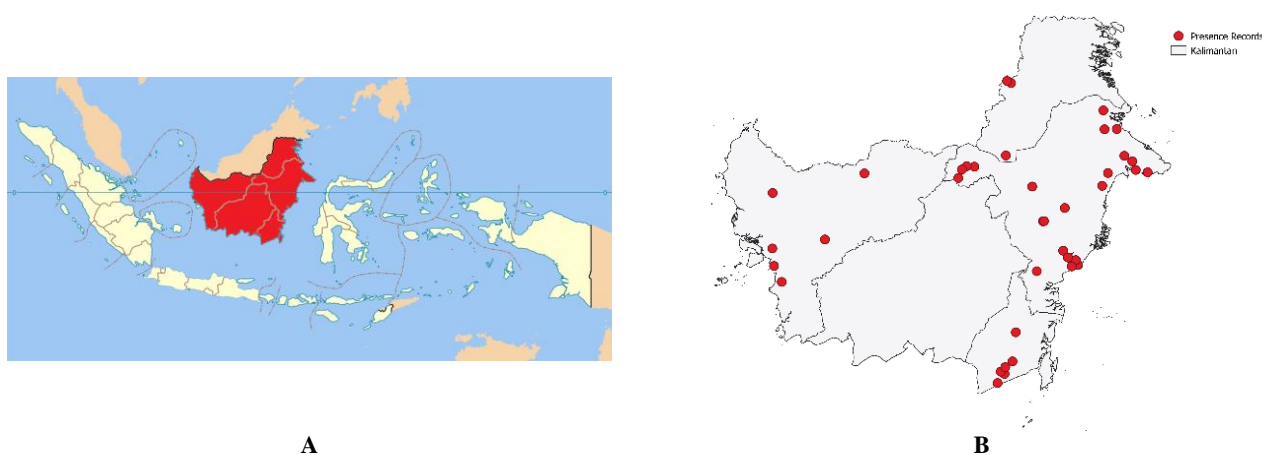


Figure 1. A. Kalimantan of Indonesian Borneo (red color), B. Occurrence records of *Eusideroxylon zwageri* in Kalimantan, Indonesia

The current and future climatic variables are extracted from version 1.4 of the global climate data (worldclim.org). 19 climatic variables are selected, consisting of BIO1 = Annual Mean Temperature, BIO2 = Mean Diurnal Range (Mean of monthly (max temp-min temp)), BIO3 = Isothermality (BIO2/BIO7) (* 100), BIO4 = Temperature Seasonality (standard deviation *100), BIO5 = Max Temperature of Warmest Month, BIO6 = Min Temperature of Coldest Month, BIO7 = Temperature Annual Range (BIO5-BIO6), BIO8 = Mean Temperature of Wettest Quarter, BIO9 = Mean Temperature of Driest Quarter, BIO10 = Mean Temperature of Warmest Quarter, BIO11 = Mean Temperature of Coldest Quarter, BIO12 = Annual Precipitation, BIO13 = Precipitation of Wettest Month, BIO14 = Precipitation of Driest Month, BIO15 = Precipitation Seasonality (Coefficient of Variation), BIO16 = Precipitation of Wettest Quarter, BIO17 = Precipitation of Driest Quarter, BIO18 = Precipitation of Warmest Quarter, BIO19 = Precipitation of Coldest Quarter. Those climatic variables are available in raster format (tiff format) that covers all regions on the globe. The Future climate variables at 2050 (2041-2060) and 2070 (2061-2080) are available at the IPCC5 climate projections from global climate model (GCMs) simulations. The data ACCESS1-0 with Representative Concentration Pathways (RCPs) 4.5 and 8.5 are chosen to understand the future distribution of *E. zwageri*. All of climatic variables have a resolution of about 900 m (30 arc-second resolution). Those data are then clipped into a small part of study area (tiff format). Those clipped-climatic variables (tiff format) are then used for modeling process in R open-source software.

Six different algorithms of Species Distribution Modelling (SDMs) are applied to understand the current and future distribution patterns of *E. zwageri*. Those models are Generalized Linear Model (GLM), Support Vector Machine (SVM), Random Forest (RF), and Generalized Additive Model (GAM), Multivariate Regression Splines (MARS), and ensemble model. The most robust future distribution model is likely derived from the ensemble model. This approach is less constrained by the assumptions of each individual model. The collinearity

test using VIF (Variance Inflation Factor) is applied to understand whether climatic variables are correlated to each other (multicollinearity). The first step, coordinate data frame is converted into coordinate spatial point. There are several R-packages that used in these models including: “dismo” is used to load climate variables, “mapview” used to see the point of occurrences, “usdm” to test collinearity among the climatic variables by providing function “vifstep” and “vifcor”, library “sdm” used to run several algorithms of species distribution models. All algorithms are run using “bootstrap” with two replication.

RESULTS AND DISCUSSION

Results

The 43 presence records are split into training and testing data with percentage of 75: 25. Overall, 100 absence-background data are used and split by 75:25 (same with presence records). Absence-background data is randomly selected throughout the area of study where the presence records are not found (see figure 3). VIF is used to understand the spatial autocorrelation among climatic variables in the area of study. After eliminating the spatial correlation among 19 climatic variables, 7 climatic variables are chosen to be used in the modeling the distribution of *Eusideroxylon zwageri* (see table 1). The AUC and TSS are used to evaluate the model performance of six different algorithms. The size of bootstrap is used to have the best model performance represented by AUC and TSS. The highest AUC and TSS are gained using bootstrap $n=2$ (see table 2). Some predictor variables are likely to influence species distribution much more than others. Assessment of the relative importance of the climatic predictor variables is evaluated using two of the species distribution models (RF and SVM). Those models are chosen due to higher AUC and TSS. Those models yield different outputs of environmental importance. The RF model implicates two of the climatic variables (precipitation of coldest quarter and isothermality) as the most important influencers of the species distribution.

Similarly, analysis of the SVM model identifies isothermality and precipitation seasonality (coefficient of variables) as the two most important variables (see figure 6).

Discussion

Non-climatic variables such as topography, land-use, land cover, dispersal are excluded in this study. Although some of those non-climatic variables probably contribute to determine the distribution of *Eusideroxylon zwageri*, we prefer to focus on bioclimatic variables alone. To some extent, variations in topography and land cover are associated with climatic variation. It is important to keep in mind the assumptions must be made to perform these model analyses. The observed distribution of *E. zwageri* is assumed to be a function of the current climate alone. It is not known if the observed species distribution represents equilibrium or transient state. It is also necessary to assume that there will be no dramatic changes of land-use and illegal harvesting, and the bioclimatic variables are the principal drivers of distribution shifts between the present day and the future. In dynamic species distributions, dispersal is important of non-climatic variables that might influence the distribution of a species. However, the dispersal behavior is quite hard to realistically incorporate in a model projection because it might change over time and tends to be more dynamic with frequency or density dependence (Salas et al. 2017).

Seven climatic predictor variables are chosen by selecting from the 19 climatic variables and then eliminating the high (>0.7) correlated variables. Those variables are assumed to be the climatic variables that drive the distribution of *E. zwageri*. The Variance Inflation Factor (VIF) is used to measure the correlation between one predictor to another predictors in model. The higher values of VIF indicating the greater the correlation of variable with other variables, it means the variable likely giving a less contribution to the model and influencing the performance of model prediction. This parsimonious list of variables is then used to run model prediction using six SDMs algorithms.

Not surprisingly, the six species distribution models, with different sets of assumptions provide different estimates of suitable bioclimatic conditions of *E. zwageri*. Overall, the AUC value indicates fairly good discriminatory ability of the individual model with the AUC >0.80. RF shows the highest AUC value among those models. The RF model performs better than other predictive models (SVM, GLM, DOMAIN, BIOCLIM) in terms of predicting the current distribution of *Guettarda speciosa* (Yudaputra et al. 2019). All models show useful agreement to the observations with a TSS range between 0.6-0.8. RF has the highest TSS value with TSS >0.8, indicating the excellent agreement of the model. Perfect agreement (TSS = 1) means the predicted model and real condition (observation) match 100%.

We used several sizes of bootstrap to evaluate the sensitivity of the model predictions. The bootstrap method is often used to estimate the sampling distribution of test statistics by repeated the observed data. Bootstrap resampling allowing to construct parsimonious high-quality

models (Austin and Tu 2014). The highest AUC and TSS for all predictive models are gained using bootstrap size of 2 ($n = 2$). The trend of graphic shows the bigger size of bootstrap, the lower both AUC and TSS score, but some of models show the decreasing from bootstrap 2 to 20 and increasing at 50, then decreasing to 200. The AUC and TSS value are only slightly different among those predictive models. RF performs better than other predictive models (SVM, GAM, ENSEMBLE, MARS AND GLM) shown by AUC and TSS scores.

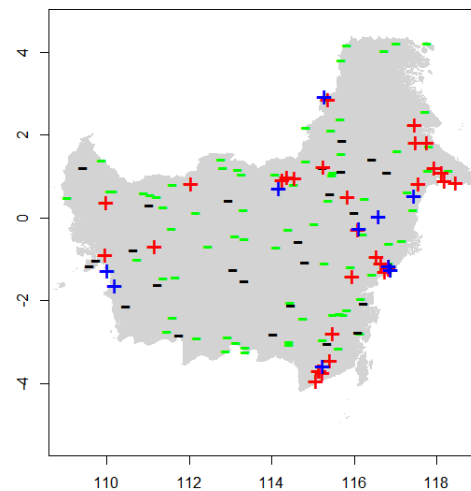


Figure 3. Model design with presence train, presence test, absence-background train, and background test. +: presence train, +: presence test:absence-background train, -: absence-background test (4 fold and 100 absence-background data)

Table 1. Climatic variables selection using VIF

Climatic variables		
1	Bio15	1.813539
2	Bio18	2.353005
3	Bio19	2.113830
4	Bio 2	3.237614
5	Bio 3	3.964943
6	Bio 4	4.722121
7	Bio 8	1.315706

Note:

Bio15: precipitation seasonality (coefficient of variables)

Bio18: precipitation of warmest quarter

Bio19: precipitation of coldest quarter

Bio2: mean diurnal range (mean of monthly (max temp. -min temp.))

Bio3: isothermality (bio2/bio7) (*100)

Bio4: temperature seasonality (standard deviation * 100)

Bio8: mean temperature of wettest quarter

Table 2. The Evaluation of six models represented by AUC and TSS (using bootstrap n=2)

Methods	AUC	TSS
GLM	0.81	0.6
SVM	0.88	0.74
RF	0.92	0.81
MARS	0.86	0.67
GAM	0.85	0.72
ENSEMBLE	0.84	0.66

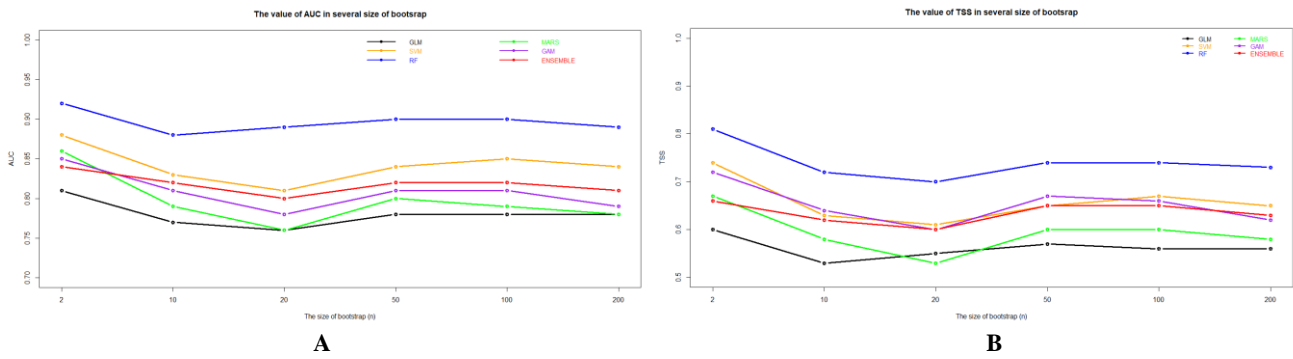


Figure 4. The evaluation of six models in several size of bootstrap: A. The AUC value, B. TSS score

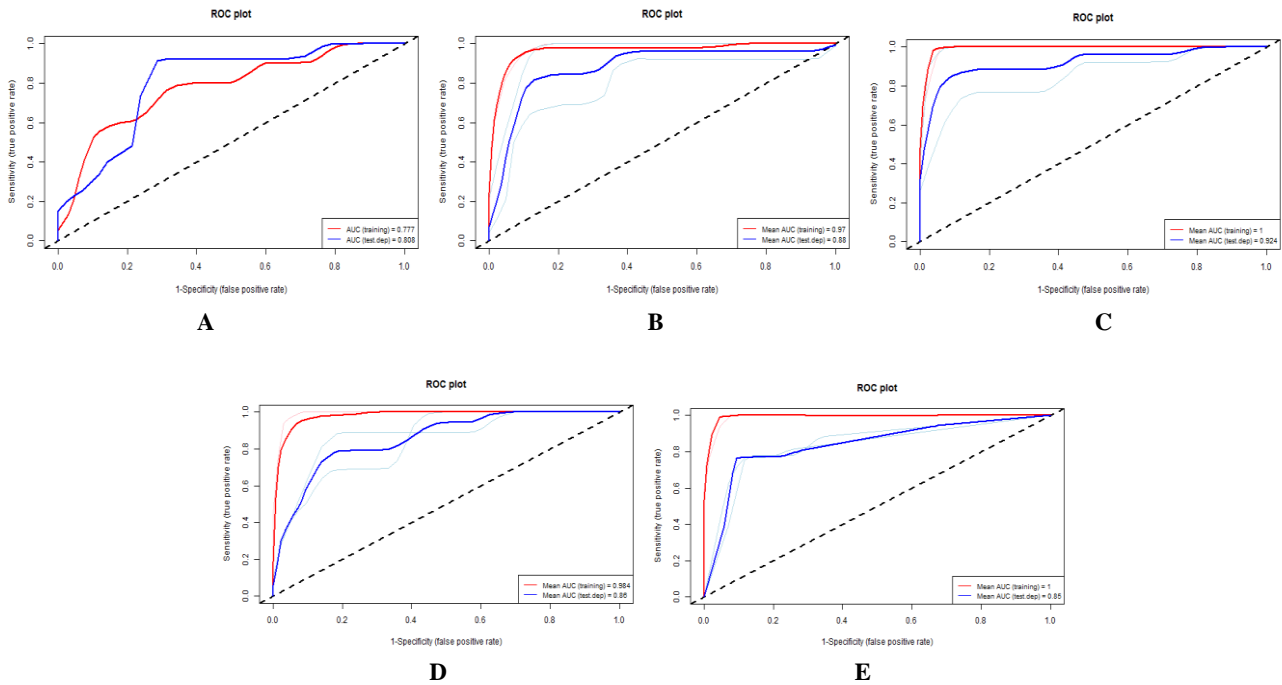


Figure 5. Evaluation of five models that represented by ROC Curve: A. GLM, B. SVM, C. RF, D. MARS, E. GAM

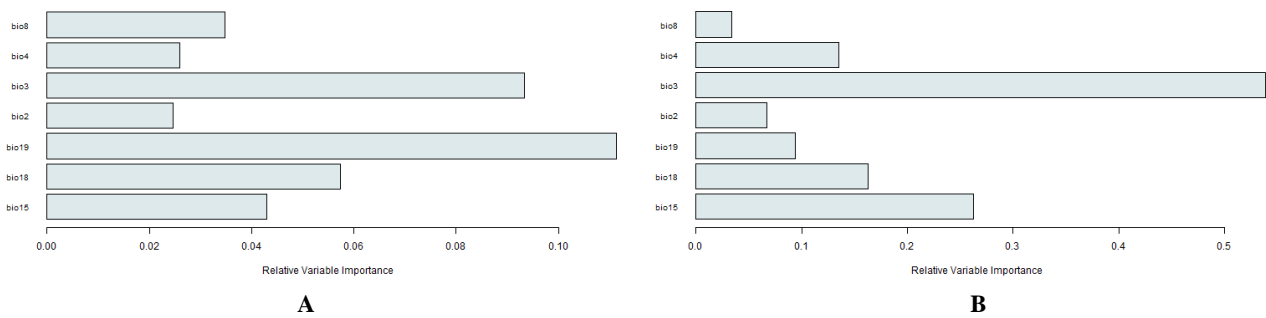


Figure 6. The important variables of predictive models with the highest AUC and TSS score: A. Random Forest (RF), B. Support Vector Machine (SVM)

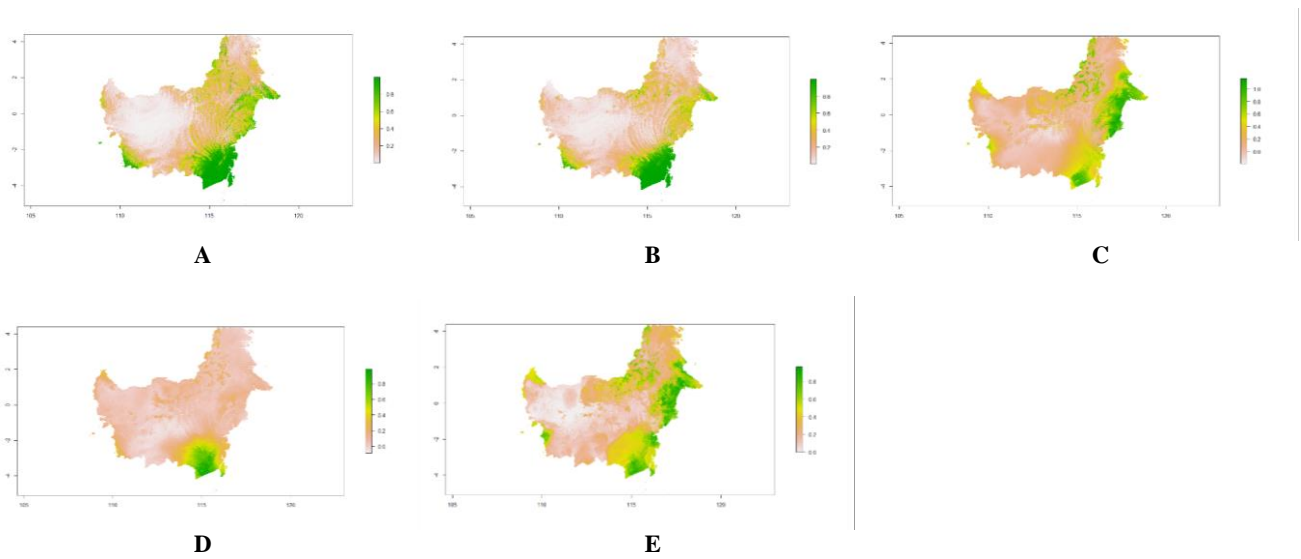


Figure 7. The predictive map of current distribution of *Eusideroxylon zwageri* in Kalimantan, Indonesia; A. GLM, B. SVM, C. RF, D. MARS, E. GAM

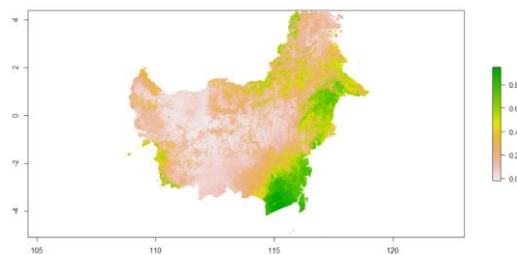


Figure 8. Present-day ensemble model of species distribution modeling for *Eusideroxylon zwageri* in Kalimantan

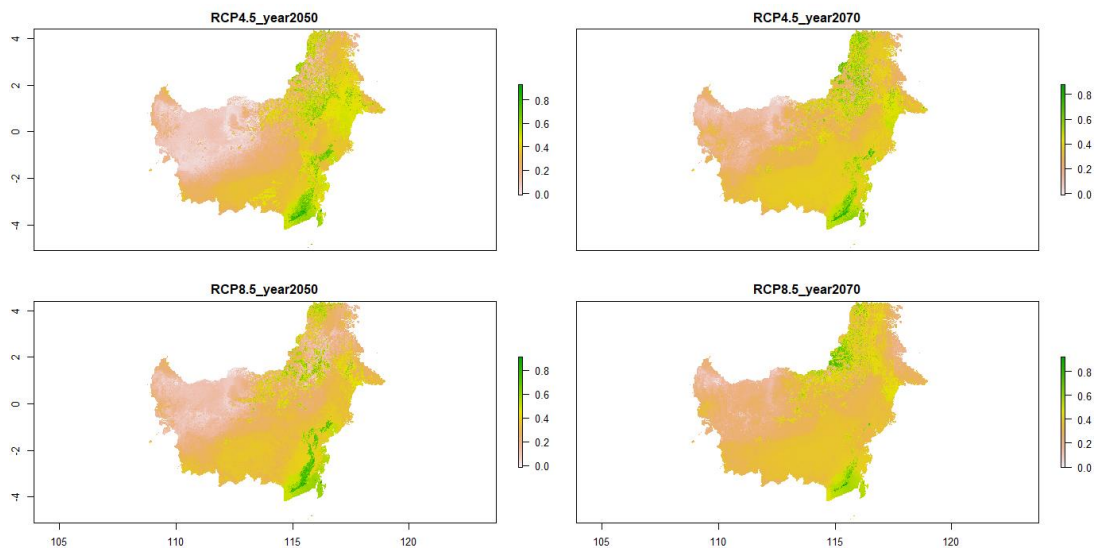


Figure 9. Comparison of ensemble suitable bioclimatic conditions between 2050 and 2070 for *Eusideroxylon zwageri*. Future model ensembles on bioclimatic data from General Circulation Models projected according to: RCP 4.5 to the year 2050, RCP 4.5 to the year 2070, RCP 8.5 to the year 2050, RCP 8.5 to the year 2070

According to the AUC and TSS score, RF and SVM are chosen to determine the most important variables. Precipitation of the coldest quarter and isothermality are the most important variables from RF model. Whereas, SVM gives isothermality and precipitation seasonality as the top two climatic variables. By combining assessments from those two models, three climatic variables including isothermality, precipitation seasonality, and precipitation of coldest quarter are considered as climatic variables that adequately determine the habitat suitability of *E. zwageri*. RF model estimates the importance variables by measuring how prediction error increases. Variable importance is relatively useful to make the output simpler and more readily interpretable (Liaw and Wiener, 2002).

The simulated current distribution for all models has AUC >0.8 that indicating fairly good model performance, but predictive maps from each of the models show a different predictive distribution map of *E. zwageri*. GLM and SVM predict that the largest part of the current distribution lies in south Kalimantan and small part of (Central, East, and North) Kalimantan. RF yields the predictive map that lies mostly in East Kalimantan and North Kalimantan and small parts in (West and South) Kalimantan. MARS predicts mostly in South Kalimantan and small part of Central Kalimantan. GAM predicts current distribution map almost in all parts of Kalimantan, except Central Kalimantan. Mostly the predictive map of GAM lies in East Kalimantan and North Kalimantan. The ensemble model, a result of combining those predictive models, shows the predictive map lies mostly in South and East Kalimantan, the predictive distribution also seen in small parts of North and West Kalimantan.

The ensemble model is likely the best choice to understand future projection. The ensemble model combined with consensus projection generally produces more accurate projections compared to any of the single models (Marmion et al. 2009; Grenouillet et al. 2011). It is quite possible that single models having the best performance on current distribution might not provide the most accurate on future projections (Thuiller 2004; Araújo et al. 2005). The future projection is modeled using two-time scenario 2050 and 2070, we used the climate scenario of RCP 4.5 and RCP 8.5 to project the future distribution shifts. The RCPs is consistent with a wide range of possible changes in future anthropogenic and aims to represent the climatic consequences of projected shifts in atmospheric greenhouse gas concentrations. Emissions in RCP 4.5 peak around 2040, then decline (Meinshausen 2011). In RCP 8.5, emissions continue to rise throughout the 21st century. RCP 4.5 in 2050 the predictive distribution lies in South, East, and North Kalimantan. Meanwhile, in 2070, the predictive distribution still lies in those areas, but the range of distribution becomes relatively narrow based on the distribution of suitable climatic variables. RCP 8.5 in 2050 shows the predictive map lies in South, East, and North Kalimantan. However, in 2070, the range of distribution

decreases in South Kalimantan, and shifted to North Kalimantan.

In conclusion, all of the tested SDMs models produce similar good predictive maps for the current distribution of *Eusideroxylon zwageri*. These analyses indicate that future populations of *E. zwageri* are at risk of declining in geographic distribution due to climate change impacts.

REFERENCES

- Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol* 43: 1223-1232.
- Araújo MB, New M. 2007. Ensemble forecasting of species distributions. *Trends Ecol Evol* 22 (1): 42-47.
- Araújo MB, Whittaker RJ, Ladle RJ, Erhard M. 2005. Reducing uncertainty in projections of extinction risk from climate change. *Global Ecol Biogeogr* 14: 529-538.
- Asian Regional Workshop (Conservation & Sustainable Management of Trees, Viet Nam, August 1996). 1998. *Eusideroxylon zwageri*. The IUCN Red List of Threatened Species 1998: e.T31316A9624725. <http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T31316A9624725.en>. Downloaded on 04 August 2019.
- Austin PC, Tu JV. 2014. Bootstrap methods for developing predictive models. *Amer Stat*. DOI: 10.1198/0003130043277
- Beaumont LJ, Hughes L. 2002. Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Ch Biol* 8: 954-971.
- Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, Springer, Berlin.
- Cutler DR, Edwards Jr TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. *Ecology* 88 (11): 2783-2792.
- Dormann CF, Purschke O, García Márquez JR, Lautenbach S, Schröder B. 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecological Society of America*, Washington, DC.
- Drake JM, Randin C, Guisan A. 2006. Modelling ecological niches with support vector machines. *J Appl Ecol* 43: 424-432.
- Ecocrop. 2019. *Eusideroxylon zwageri*. <http://ecocrop.fao.org/ecocrop/srv/en/cropView?id>. [4 August 2019].
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JMM, Peterson AT, Phillips SJ, Richardson K, Sacchetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29 (2):129-151.
- Elith J, Leathwick JR. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Divers Distrib* 13: 265-275.
- Elith J, Leathwick JR. 2009. Species distribution models: Ecological explanation and prediction across space and time. *Ann Rev Ecol Evol Syst* 40: 677-697.
- Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv* 24: 38-49.
- Grenouillet G., Buisson L, Casajus N, Lek S. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34: 9-17.
- Hastie T, Tibshirani R, Friedman J. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York.

- Hastie TJ, Tibshirani RJ. 1990. Generalized Additive Models. Chapman and Hall, New York.
- Krzanowski WJ, Hand DJ. 2009. ROC Curves for Continuous Data. Chapman and Hall, Boca Raton, FL.
- Leathwick JR, Elith J, Hastie T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol Model* 199: 188-196.
- Liaw A, Wiener M. 2002. Classification and Regression by random Forest. *R News* 2 (3): 18-22.
- Lobo JM, Jiménez-Valverde A, Real R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol Biogeogr* 17: 145-151.
- Marmion M, Parviainen M, Luoto M, Heikkinen RK, Thuiller W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Divers Distrib* 15: 59-69.
- Mateo RG, Croat TB, Felicísimo AM, Muñoz J. 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Divers Distrib* 16 (1): 84-94.
- Meinshausen M, Smith SJ, Calvin K, Daniel JS, Kainuma MLT, Lamarque JF, Matsumoto K, Montzka SA, Raper SCB, Riahi K, Thomson A, Velders GJM, van Vuuren DPP. 2011. "The RCP greenhouse gas concentrations and their extensions from 1765 to 2300 (open access)". *Clim Ch* 109 (1-2): 213-241.
- Parmesan C, Yohe GA. 2003. Globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421: 37-42.
- Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9 (2): 181-199.
- Salas EAL, Harings NM, Seamster VA, and Boykin KG. 2017. Projected Future Bioclimate-Envelope Suitability for Reptile and Amphibian Species of Concern in South Central USA. *Herpetol Conserv Biol* 12 (2): 522-547.
- Shabani F, Kumar L, Ahmadi M. 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecol Evol*. DOI: 10.1002/ece3.2332
- Sinclair SJ, White MD, Newell GR. 2010. How useful are species distribution models for managing biodiversity under future climates? *Ecol Soc* 15: 8-20.
- Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo, CM, Hemingway J. 2010. The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and biometric precis. *Parasites Vectors* 3: 117.
- Stephenson DB. 2000. Use of the "Odds Ratio" for diagnosing forecast skill. *Weather Forecasting* 15: 221-232.
- The Ministry of Forestry Republic of Indonesia. 2009. <https://www.dephut.go.id/>. [4 August 2019].
- Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Ch Biol* 10: 2020-2027.
- Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Divers Distrib* 13: 397-405.
- Wisn MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A. 2008. Effects of sample size on the performance of species distribution models. *Divers Distrib* 14 (5):763-773.
- Yudaputra A, Pujiastuti I, Cropper WP. 2019. Comparing six different species distribution models with several subsets of environmental variables: Predicting the potential current distribution of *Guettarda speciosa* in Indonesia. *Biodiversitas* 20: 2321-2328.
- Yudaputra A, Robiansyah I, Rinandio DS. 2019. The Implementation of Artificial Neural Network and Random Forest in Ecological Research: Species Distribution Modelling with Presence and Absence Dataset. Proceedings the 3rd SATREPS Conference, 22 Nov 2018, Bogor, Indonesia.