Climate based model in determining the distribution pattern of *Cecropia peltata* L across global landscape

A Yudaputra and P Hutabarat

Research Center for Plant Conservation and Botanic Gardens, Indonesian Institute of Sciences. Jl. Ir. H. Djuanda No.13, Paledang, Bogor 16112, West Java, Indonesia

E-mail: angg020@lipi.go.id

**Abstract.** Climate change becomes a major threat to the global biodiversity. It alters the ecological niche of species, even small change in temperature could have a significant impact to the distribution pattern of biodiversity. *Cecropia peltata* is an invasive species with wide range geographic distribution. The aim of this study is to understand the impact of climate change to the current and future distribution of invasive plant *C. peltata*. The Support Vector Machine (SVM) and Boosted Regression Tree (BRT) algorithm of machine learning were used to predict the current and future distribution. The occurrence records of *C. peltata* was obtained from Global Biodiversity Information Facility (GBIF). There were 2691 occurrence records in GBIF database. The global climatic variables with the resolution 2.5 km were used as predictors of model. VIF was used to select the multicollinearity among those variables using threshold of 0.7. The CIMP5 of Global Circulation Model (GCM) was used to understand the impact of climate change to the distribution of the plant. The future projection on year 2070 with the worst climate scenario RCP 8.5 was used on these predictive models. The SVM and BRT models were actually relevant to be used as predictive models with AUC >0.90 and categorised as excellent predictive models. The future distribution pattern was likely to be shifted compared to the current distribution prediction. The output of this study as predictive current and future distribution maps would be useful to provide an information about the potential area where the species might be invading based on the training data (observation data). Furthermore, the prediction of future distribution would be necessary to understand how the climate change literally affects the range of distribution of the invasive plant species.

1. Introduction

Climate change was supposed to be one of factors that affecting the geographic distribution of species across the globe [1,2,3]. The shifting of distribution pattern might be threatening their persistence and viability through the reduction of their geographic range which causes the declining of their population [4]. The impact of climate change to the geographic distribution of species can be assessed using many approaches. Species Distribution Model (SDM) was widely used to understand the potential distribution pattern of the species based on the statistical relationship between occurrence records of the species and their associated climate. It is used to project the potential distribution to the new environmental.

Several assumptions are made in the modelling of species distribution, these are the species response to the environmental variable remain unchanged and the inter-specific interactions are ignored [5,6,7] may not capture the correct relationship species-environment due to correlative process not mechanistic [8] and do not consider long term population viability [9]. SDM was used to address several problems in ecology such as rare species, invasive species, and conservation planning. On
recent study, SDM was used to understand the potential current distribution of invasive species (*Caliantra calothyrsus*) [10].

Many methodologies on species distribution modelling have been rapidly developed in recent decade [11,12]. These applications could be used for conservation planning in a dynamic landscape [13] and disease modelling [1]. There were numerous attempts to find the well-tested model for predicting the species range shifts due to climate change [2]. One of algorithms that w (1) ed in species distribution model was Support Vector Machine (SVM). It was also used in classification of remotely sensed data [14] Basically, SVM used a kernel to map data onto a new hyperspace to simplify the complicated pattern [15]. SVM works by projecting vectors into a high dimension feature space, then kernel will optimize the hyperplane that separates two or more classes

$$\hat{y} = \text{sign}[f(x)] = \left( \sum_{i=1}^{N} a_i y_i K(x, x) + b \right)$$

In which, N is the number of training points, $y_i$ represents the label of $i$th sample, $a_i$ denotes the Lagrangian multipliers, $K(x, x)$ represents the chosen kernel function, and b is a bias parameter [16]. It successfully implements in data with a large set of variables [17], has no theoretical requirements that usual in statistic like the data should be independent, overcomes the autocorrelation problem, requires few model tunning, more stable and has fewer parameters than other computation methods like neural network [18]. SVM often performed better in predicting potential current distribution of *Guettarda speciosa* [19]. SVM showed the AUC>0.80 which is categorized as good predictive model in forecasting the distribution of *Eusideroxylon zwageri* [19].

The basic principle of Booster Regression Tree (BRT) was a combination between two techniques including decision tree algorithm (a model that used to find out the relationship between predictors by recursive binary splits) and boosting method (a model that combines several simple models to improve the performance of predictive model) [20]. The principle work of BRT algorithm was determined by repeatedly fitting the decision trees to improve the model accuracy. BRT algorithm was able to accurately select relevant variables and sometimes provides substantial predictive advantages compared to GAM and GLM [21, 22, 23].

The invasive plant *C. peltata* is a fast-growing, pioneer, short-lived tree, one of members the genus of *Cecropia*, family of Urticaceae [24,25]. It was commonly named as trumpet tree and snake wood. It was listed as one of 100 the world’s worst invasive alien species by the invasive species specialist group [26]. It relatively required much of light and rapidly invaded the disturbed areas such as forest gap canopies, agricultural sites, abandoned land, urban areas, roadside and lava flows [27,28,29,30]. It has a wide range of distribution, originally from southern Mexico, to northern South America and Caribbean Islands [27]. It has been introduced to Malesia, The Pacific and Africa [31,32,30], and was often replacing native plant species, and reducing species richness in a certain landscape. It was commonly found over wide elevation gradient from 50-2700 m. It merely required much rainfall with the 990 mm to 3,810 mm of annual precipitation. It can be found in warm climate from montane region to tropical with mean annual temperatures of 12-24°C [33].

This study would try to provide the valuable insight pertain to how climate change will affect the distribution of *C. peltata*. It was relatively not easy work to survey all locations across landscape to collect the occurrence of the species. For surveying all location will spend lots of sources such as money, time and energy. Species Distribution Modelling (SDM) was now available to provide a relevant predictive map about the potential distribution of species in unsurvey locations. This study trying to understand the prediction of current and future distribution of *C. peltata* under projected climate scenario. The predictive map of current distribution was necessary to figuring out another region of *C. peltata* that having a similar climate condition in which the species occurrences are recorded. The future predictive map will provide an information about how the future distribution might be shifted under future climate projection. Furthermore, the future distribution prediction was quite useful to understand the predictive invasion area for management and control of this invasive plant.
2. Experimental details

Occurrence records were collected from Global Biodiversity Information Facility (GBIF) with the 2,726 total occurrence records around the world (GBIF.org). The package of “rgbif” was used to load the occurrence data from GBIF database to R environment programming [34,35]. The current climate data was extracted from WorldClim 1.4 database [36]. The 19 climatic variables with 2.5 minutes (2.5 km) resolution were then selected to avoid the multicollinearity among those climatic variables. The pairwise correlation (vifcor) with threshold default of 0.7 was used in multicollinearity test. The variables that having value >0.7 will be eliminated in further analysis. Twelve climatic variables which showing the multicollinearity problem were removed. The remaining 7 climatic variables were used as predictors of the model. The package of “usdm” was used in the multicollinearity test for those climatic variables. For future climate scenario, Representative Concentration Pathway (RCP) 8.5 was chosen over the future distribution modelling. RCP 8.5 represented high gas emission (the worst case of climate model simulations in the Fifth Assessment IPCC report) [37]. Boosted Regression Tree (BRT) and Support Vector Machine (SVM) were two species distribution models that used to understand the species-physical environment relationship in spatial geographic landscape. The
presence records and pseudo-absence data of *C. peltata* were used in the model prediction. The 1000 pseudo-absence data were generated to fulfil the requirement of SDM algorithm. We used “sdm” package in R to run these predictive models [38].

3. Results and discussion

![Figure 2. Occurrence records of *Cecropia peltata* L. The occurrence map was obtained from GBIF database [34].](image)

| Algorithm | AUC   | TSS   | Threshold |
|-----------|-------|-------|-----------|
| BRT       | 0.979 | 0.88  | 0.279     |
| SVM       | 0.98  | 0.859 | 0.261     |

![Figure 3. The predictive map of potential current distribution using Boosted Regression Tree (BRT) model](image)
Figure 4. The predictive map of potential future distribution using Boosted Regression Tree (BRT) model.

Figure 5. The predictive map of potential current distribution using Support Vector Machine (SVM) model.
Figure 6. The predictive map of potential future distribution using Support Vector Machine (SVM) model.

Figure 7. The BRT’s predictive distribution changes of *Cecropia peltata* L by subtracting the future distribution to current distribution.
Figure 8. The SVM’s predictive distribution changes of *Cecropia peltata* L by subtracting the future distribution to current distribution.

Figure 9. The future predictive map at year 2070. It representing the new colonization and extinction area of *Cecropia peltata* L built upon BRT model.
**Figure 10.** The future predictive map at year 2070. It representing the new colonization and extinction area of *Cecropia peltata* L built upon SVM model.

**Figure 11.** Response curve of BRT model with seven climatic variables.
Both two predictive models, Support Vector Machine (SVM) and Boosted Regression Tree (BRT) are having high evaluation value using AUC, Kappa, and TSS. These models have the AUC value >0.90 that indicating those are excellent models in predicting the potential current and future distribution of invasive plants *Cecropia peltata*. Not all climatic variables used in the model, only seven climatic variables were involved in the model. All climatic variables that having collinearity problem will be eliminated.

According to the occurrence records database, the plant was mostly found in South America because it originally found in this area, then spreading in many countries in Africa, Asia and small part of Europe. This plant has been successfully adapting with various range of climate condition from tropic to sub tropic so it can be easily found in those locations. Based on the climatic condition, BRT and SVM models provide almost similar predictive map output for the current distribution. The predictive current distribution giving an information about the suitable habitat of the plant based on the climatic condition. It was not possible to survey all the location, we could predict the unsurvey location by calculating the species – climate relationship in the training model.

Our result showed the two models predicted almost the similar predictive maps. The current and future distribution for both models shows difference distribution range. The future distribution of the invasive plant seems to be shifted in some regions. The potential future distribution was expanded to both southern and Northern hemisphere of America, central part of Africa, western Malesia and Philippine, East Asia, and most of Europe. More regions likely changed in term of climatic variables and had potential to be suitable climatic habitat of *C. peltata*. It indicated that those regions have a high risk of the species invasions in the future.

In 2070, the plant will invade prominently some parts of China, Japan, Eastern Malesia, South eastern of Australia and New Zealand, Europe, central and eastern of South America, and South eastern of North America. New colonization area occurs if there is a new territory which is predicted to become the potential distribution area of *Cecropia peltata* by 2070. Extinction was happened if the future prediction was 0 and predicted current distribution was 1 which means the predicted current distribution will loss in predicted future distribution at 2070. Both models predicted new colonization will be seen in all parts of world. It was likely spreading over different areas of the world. Response curve analysis was used to understand the response of the species and their tolerance to the environment gradient changing. In general, the change of temperature and precipitation were two
important climatic variables need to be considered in predicting potential distribution of the species. The BRT models predict the species will responsively adapt to the precipitation of the coldest quarter, followed then precipitation seasonality and mean temperature of wettest quarter. While, the SVM predicts the species will most responsive to the changing of precipitation of the coldest quarter, followed then mean diurnal range, and mean temperature of wettest quarter. In fact, these models predict almost the similar response curve of the species. These important climatic variables supported the result of MaxEnt modelling that found temperature seasonality and annual precipitation would affect the potential distribution of *C. peltata* [39].

The changes of temperature and precipitation due to climate change affected plant invasion and the distribution [40,41,42]. For instance, from our study, high distributional probability of *C. peltata* was driven by precipitation in a certain period of time and the temperature range during that period. Moisture, light and temperature have a correlation on invasive potential of *Cecropia* species [27]. Understanding the biology of invasive plants is important to address a correlation of potential invasion and climatic variables. However, the biology and life history of many *Cecropia* species remains questionable particularly in the invaded areas [43]. *Cecropia* peltata is generally wind-pollinated and have the special adaptation of shedding dry pollen by motion of the anthers [27]. The highest percentage contribution of pollen of *C. peltata* in the air occurred from April to May in Venezuela, before the rainy season runs from May to December [44]. In Costa Rica, the male flower production peaked in April to June (1980 – 1981), followed by rainy season which started in mid-May [45]. Fleming & Williams [46] also reported that fruit and flower peaked earlier at moist sites than at dry sites. In the Neotropics, the million seeds of Cecropia dispersed by helping of frugivorous animals, such as birds, mammals, and fish [27]. However, the seed of *C. peltata* which growing on riverbanks, or near waterways are usually dispersed by water by the ability of floating downstream for a while [46,29].The ability of seed to dormant or having long viability was expected supporting the potential invasion of the species. It was proved by the dominance of the seed of *C. peltata* in the soil seed bank of a karst ecosystem in Indonesia [47]. Those indicated that precipitation is not only having a major role in the flower initiation and pollen dispersal, but also the long-lived seed dispersal of the species through the intensity of water flow during rainy season.

4. Conclusion
The future geographic distribution of the invasive plant *Cecropia peltata* L shifted to different regions resulting from both BRT and SVM predictive models. These models are categorized to excellent predictive models in forecasting the geographic distribution shift of *C. peltata* L under future climate projection. Precipitation of the coldest quarter was considered as the most important climatic variable that determining the future geographic distribution of *C. peltata* under climate change.

Acknowledgement
I would like to express my great thanks to the research coordinator of P2KTKR-LIPI for his supports in our research activities.

References
[1] Peterson A T, Ortega-Huerta M A, Bartley J, Sanchez-Cordero V, Soberon J, Buddemeier R H and Stockwell D R B 2002 Future projections for Mexican faunas under global climate change scenarios Nature 416 626–629.
[2] Thomas C D, Cameron A, Green R E, Bakkenes M, Beaumont L J, Collingham Y C, Erasmus B F N, de Siqueira M F, Grainger A, Hannah L, Hughes L, Huntley B, van Jaarsveld A S, Midgley G F, Miles L, Ortega-Huerta M A, Peterson A T, Phillips O L and Williams S E 2004 Extinction risk from climate change Nature 427 145–148.
[3] Loarie S R, Carter B E, Hayhoe K, McMahon S, Moe R, Knight C A and Ackerly D D 2008 Climate change and the future of California’s endemic flora PLoS One 3 e2502.
[4] Gaston K J 1994 Rarity (London: Chapman and Hall)
[5] Araújo M B and Luoto M 2007 The importance of biotic interactions for modelling species distributions under climate change Glob Ecol. Biogeogr 16 743–753.
[6] Davis A J, Jenkinson L S, Lawton J H, Shorrock B and Wood S 1998 Making mistakes when predicting shifts in species range in response to global warming Nature 391 783–786
[7] Pearson R G and Dawson T P 2003 Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob Ecol Biogeogr 12 361–371.
[8] Kearney M, Porter W 2009 Mechanistic niche modelling: combining physiological and spatial data to predict species’ ranges Ecol Lett 12 334–350.
[9] Keith D A, Akecayaka H R, Thullier W, Midgley G F, Pearson R G, Phillips S J, Regan H M, Araujo M B and Rebelo T G 2008 Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models Biol Lett 4 560–563.
[10] Yudaputra A 2020 Modelling potential current distribution and future dispersal of an invasive species Calliandra calothyrsus in Bali Island, Indonesia Biodiversitas 21 674-682.
[11] Guisan A and Zimmermann N E 2000 Predictive habitat distribution models in ecology Ecological Modelling 135 147–186.
[12] Scott J M, Heglund P J, Morrison M, Raphael M, Hauffer J, Wall B and Samson F B 2002 Predicting species occurrences: Issues of accuracy and scale (Covello, CA: Island Press).
[13] Ferrier S 2002 Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Systematic Biology 51 331–363.
[14] Foody G M and Mathur A 2006 The use of small training sets containing mixed pixel for accurate hard image classification: training on mixed spectral responses for classification by a SVM Remote Sensing of Environment 103 179–189.
[15] Müller K R, Mika S, Rätsch G, Tsuda K and Schölkopf B 2001 An introduction to kernel-based learning algorithms IEEE Transactions on Neural Networks 12 181–202.
[16] Vapnik V and Chervonenkis A 1998 Statistical learning theory. In: Support Vector Machines for Pattern Recognition (John Wiley & Sons, New York).
[17] Tax D M J and Müller K R 2004 A consistency-based model selection for one-class classification. Proceedings 17th International Conference on Pattern Recognition (22–26 August 2004, Cambridge UK) (eds J. Kittler, M. Petrou & M. Nixon) (IEEE Computer Society, Los Alamitos, CA, 363–366.).
[18] Lusk J J, Guthery F S and DeMaso SJ 2002 A neural network model for predicting bobwhite quail abundance in the Rolling Red Plains of Oklahoma. Predicting Species Occurrences: Issues of Accuracy and Scale (eds J.M. Scott, P.J. Heglund, M. Morrison, M. Raphael, J. Hauffer, B. Wall & F.B. Samson), pp. 345–355. Island Press, Covello, CA.
[19] Yudaputra A, Pujaistuti I, Cropper Jr W P 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.
[20] Elith J, Leathwick J R and Hastie T 2008 A working guide to boosted regression trees Journal of Animal Ecology 77 802-813.
[21] Elith J, Graham C H and Anderson R P et al 2006 Novel methods improve prediction of species’ distributions from occurrence data. Ecography, 29, 129–151. De‘Ath G (2007) Boosted trees for ecological modeling and prediction Ecology 88 243-251.
[22] Leathwick J R, Elith J, Chadderton W L, Rowe D and Hastie T 2008 Dispersal, disturbance, and the contrasting biogeographies of New Zealand’s diadromous and non-diadromous fish species Journal of Biogeography 35 1481-1497
[23] Moisen G G, Freeman E A, Blackard J A, Frescino T S, Zimmermann N E and Edwards T C 2006 Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods Ecological Modelling 199 176–187.
[24] Hartshorn G 1980 Neotropical forest dynamics Biotropia 12 23-30.
[25] Coley and Phyllis D 1986 Costs and benefits of defense by tannins in a neotropical tree Oecologia. 70 238–241.
[26] Lowe S, Browne M, Boudjelas S and De Poorter M 2000 100 of the World’s Worst Invasive Alien Species A selection from the Global Invasive Species Database.
[27] Berg C C and Franco-Rosselli P F 2005 Cecropia Flora Neotropica Monograph 94 1–230
[28] Binggeli P 1999 Cecropia peltata L. (Urticaceae). http://members.lycos.co.uk/WoodyPlantEcology/docs/web-sp3.htm?
[29] Sheil D, Padmanaba M 2011 Innocent invaders? A preliminary assessment of Cecropia, an American tree, in Java Plant Ecology & Diversity 4 231–240.
[30] Conn B J, Hadijah J T and Webber B L 2012 The status of Cecropia (Urticaceae) introductions in Malesia: addressing the confusion Blumea 57 136–142.
[31] Assi A I 1979 Cecropia peltata L. (Moraceae): ses origines, introduction et expansion dans l’Est de la Cote D’Ivoire. In: Kunkel G, editor Taxonomic Aspects of African Economic Botany 140–145.
[32] Putz F E and Holbrook N M 1988 Further observations on the dissolution of mutualism between Cecropia and its ants: the Malaysian case Oikos 53 121–125.
[33] Global Invasive Species Database 2020 Species profile: Cecropia peltata. Downloaded from http://www.iucngisd.org/gisd/speciesname/Cecropia+peltata on 26-04-2020.
[34] GBIF.org 2020 GBIF Occurrence .acceded from https://doi.org/10.15468/dl4gig26
[35] Chamberlain S, Boettiger C, Ram K, Barve V and Mcglinn D 2016 rgbif: Interface to the Global Biodiversity Information Facility API. R package version 0.9.3 https://github.com/ropensci/rgbif
[36] Hijmans R J, Cameron S E, Parra J L, Jones P G and Jarvis A 2005 Very high resolution interpolated climate surfaces for global land areas International Journal of Climatology: A Journal of the Royal Meteorological Society 25 1965-1978.
[37] Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N and Rafaj P 2011 RCP 8.5—A scenario of comparatively high greenhouse gas emissions Climatic Change 109 33.
[38] Naimi B and Araújo M B 2020 Package ‘sdm’. https://cran.r-project.org/web/packages/sdm/sdm.pdf
[39] Wan J, Zhang Z, Wang C 2018 Identifying potential distributions of 10 invasive alien trees: implications for conservation management of protected areas Environmental Monitoring and Assessment 190 739 (2018).
[40] Thuiller W, Richardson D M, Pyšek P, Midgley G F, Hughes G O and Rouget M 2005 Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale Global Change Biology 11 2234–2250.
[41] Hellmann J J, Byers J E, Bierwagen B G and Dukes J S 2008 Five potential consequences of climate change for invasive species Conservation Biology 22 534–543.
[42] Bradley B A, Blumenthal D M, Wilcove D S and Ziska L H 2010 Predicting plant invasions in an era of global change Trends in Ecology & Evolution 25 310–318.
[43] Webber B L, Born C, Conn B J, Hadijah J T, and Zalamea, P-C 2011 What is in a name? That which we call Cecropia peltata by any other name would be as invasive? Plant Ecology and Diversity 4, Nos 289 – 293.
[44] Hurtado I and Alson J 1995 Airborne dispersal of Cecropiapeltata L. (Moraceae) pollen, a Neotropical species, in the vicinity of Caracas, Venezuela Aerobiologia 11 101-103.
[45] Fleming T H and Williams C F 1990 Phenology,seed dispersal,andrecruitment in Cecropia peltata (Moraceae) in Costa Rican tropical dry forest J.Trop.Ecol. 6 163-178.
[46] Lok A F S I, Tan K-x, Chong KY, Nghiem T P L and Tan H T W 2010 The distribution and ecology of Cecropia species (Urticaceae) in Singapore Nature in Singapore 3 199–209.
[47] Putri W U, Qayim I and Qadir A 2017 Soil Seed Bank of Two Karst Ecosystems in Bogor, Indonesia: Similarity with the Aboveground Vegetation and Its Restoration Potential The Journal of Tropical Life Science 7 224 – 236.