PREDICTING ASSAM TEA DISTRIBUTION IN UPPER NORTHERN THAILAND USING SPECIES DISTRIBUTION MODELS

NANGLAE, S.[1] – NILTHONG, R.[1*]

[1]*School of Science, Mae Fah Luang University*
333 moo 1, Thasud, Muang, Chiang Rai 57100 THAILAND
(phone: +66-5-391-6775; fax: +66-5-391-6776)

**Corresponding author*
e-mail: rungrote@mfu.ac.th

(Received 24th Nov 2014; accepted 26th Mar 2015)

Abstract. The aim of this study is to build species distribution models and find factors affecting Assam tea distribution in upper Northern Thailand. The data were analyzed by the logistic regression of 185 records of Assam tea and generated pseudo-absence data using the two-step method. The re-sampling technique used the bootstrapping method to interpolate small sample size problems, and the k-fold cross-validation technique was used to find the best testing data set. Furthermore, statistical values (average value, standard deviation and coefficient of variation) were used to find the best final model. From the effective predictor variables at a significance level of 0.05: aspect, annual mean temperature, mean temperature of the wettest quarter, annual precipitation, precipitation of the wettest month and precipitation of the driest quarter all had positive effects upon model. Conversely: distance from the river, minimum temperature of the warmest month, precipitation of the driest month and precipitation of the wettest quarter all had negative effects upon model. Furthermore, the most suitable areas were found to be predominantly in the north-northwest region of upper Northern Thailand.
Keywords: *species distribution models, logistic regression, Assam tea, resampling technique*

Introduction

Assam tea (*Camillia sinensis* ver. *assamica*) or wild tea is the most commonly used plant in tea production. It originates in India and naturally distributes through highland forests where the climate is rainy and moderately sunny. Assam tea has been continuously maintained in the forests and is recognized as an agroforestry system. The Assam tea forests cover approximately 170,000 km^2 in Thailand, and 45% of these regions consist of highland areas (Pagella et al, 2002). Interestingly, many biodiversified features are found in the Assam tea forests, especially herbal plants. Preechapanya (1996) found that there were 149 plant species from 71 families in the Assam tea forest area of Phi Pan Nam Mountain. Currently, village expansion and farming in and around the Assam tea forests is on the increase, which is resulting in a rapid decrease of numbers of Assam tea forests (Kowsuvon, 2008). A survey of Assam tea distribution is difficult and high cost. Therefore, alternative approaches such as mathematical or statistical models could be tools used to predict the most suitable areas when considering environmental factors. One approach that can be used to identify most suitable areas is ‘Species Distribution Models’ (SDMs).

Species Distribution Models (SDMs) are important tools for predicting how species distribute across all studied areas. They have been used in many fields of biology, such as conservation planning, surveying, evolution studies or assessing the impacts of climate change (Thuiller et al., 2005; Guisan and Thuiller, 2005; Engler et al.,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

2004; Marini et al., 2009). In general, the main objectives for building species distribution models are to indicate appropriate environmental factors, including topography, climate or geology, and to identify areas which generate the most suitable environmental factors (Pearson, 2007). SDMs can be divided into two groups of techniques: The first group is profile techniques, such as environmental envelope (BIOCLIM) or environmental distance (DOMAIN) (Franklin, 2010). These techniques require presence-only data which are based upon the ranges, means, or other values of environmental variables for locations where a species of interest has been in existance. The second group is group discrimination techniques, such as generalized linear models (GLMs), generalized additive models (GAMs), boosted regression trees (BRTs) etc. which need both presence and absence data (Stokland et al., 2011) in order to generate statistical functions or discriminative rules. Generally, group discrimination techniques can produce much better results than profile techniques (Mateo et al., 2010), primarily because these techniques acquire absence data or background data used to improve the relationships between species occurrence and environmental variables (Brotons et al., 2004). However, such group discrimination techniques lack true absence data, and therefore pseudo-absence data or random samples of the background data need to be generated for the inclusive models. Pseudo-absence data are based upon statistical theory, while random samples of background data are based upon machine-learning; e.g. maximum entropy (Maxent) (Mateo et al., 2010). Contemporary statistical analysis has been used extensively, including SDMs. From 1998-2007, approximately 70-80% of articles published in the *Journal of Wildlife Management* and *Landscape Ecology Journal* used regression-based models in predictive modeling (Drew et al., 2011). These methods also need absence or pseudo-absence data in their models. Most pseudo-absence data can be divided into two groups (Lobo and Tognelli, 2011). The first group chooses pseudo-absence data by simple and random selection in the study area (Barbet- Massi et al., 2012; Lobo and Tognelli, 2011). The second group chooses pseudo-absence data using the following two-step techniques (Barbet-Massin et al., 2012; Lobo and Tognelli, 2011; Wisz and Guisan, 2009): First, build suitable areas using profiling techniques, and then randomly select pseudo-absence data outside of those suitable areas. The models with two-step pseudo-absence data produce better results than those with random pseudo-absence. However, the results of these models could lead to weaker predictive power because they lead to over fitting (Wisz and Guisan, 2009).

The objectives of this study were to find the factors affecting the distribution of Assam tea in upper Northern Thailand, and to identify suitable environmental areas for Assam tea production using logistic regression analysis (GLMs technique). The selection of pseudo-absence data using the two step technique and the bootstrap technique were applied to this model, and a resampling technique (bootstrap) was used to remedy the sample size limitations.

Methods

Study area and occurance data

The upper Northern part of Thailand covers nine provinces, namely: Chiang Mai, Chiang Rai, Mae Hong Son, Nan, Phayao, Phrae, Uttaradit, Lampang and Lamphun (shown in *Figure 1*). The area is situated between latitudes 17° 6’ 39” - 20° 29’ 50” N and longitudes 97° 16’ 36” - 101° 27’ 55” E, and covers an area of 96,025 km^2. Most areas are hilly and mountainous, and approximately 64% is covered by forest (Forest Land Management Bureau, 2013).

The data used in this study were obtained from the Tea Institute of Mae Fah Luang University, Thailand. One hundred and eighty five records of Assam tea data (presence data) in the form of geographic coordinates were collected. These data were part of a collection of tea varieties grown a in Northern Thailand project (Sampanvejsobha et al., 2009).

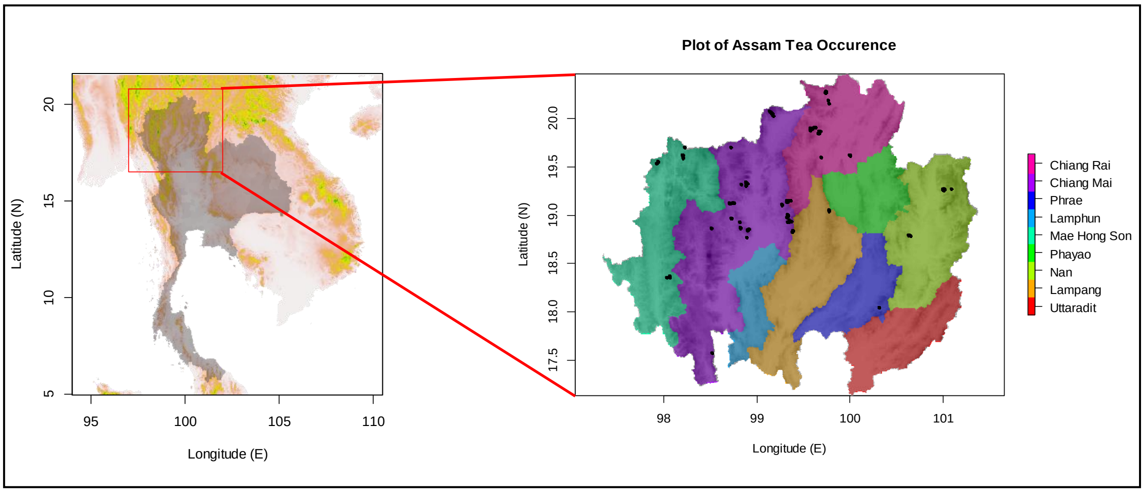

Figure 1. *The demarcation of study areas including 9 provinces in Northern Thailand. The black dots indicate 185 records of Assam tea data.*

Environmental data

Environmental data were separated in two groups: topography factors and climate factors (*Table 1*.). Topography factors were composed of elevation (DEM), slope aspect (Aspect), distance from the river (Distance) and modified groups of rock type with slope (Class1-Class8). The climate factors were composed of relative humidity (Humidity) and bioclimatic variables (Bio1-Bio19). DEM data were obtained from the Land Development Department, Thailand. The distance was calculated from river data that had been obtained from DIVA-GIS (http://www.diva-gis.org). The aspect was calculated from the elevation data, and relative humidity was obtained from the Marc Souris’s page (http://www.rsgis.ait.ac.th/~souris/thailand.htm). Bioclimatic variables were obtained from the WorldClim database, which offered approximately 1 km^2 of spatial resolution (Hijmans et al., 2005; http://www.worldclim.org). The WorldClim data were derived from measurements of altitudes, temperatures and rainfall from weather stations across the globe (Period 1950-2000). Finally, nineteen bioclimatic variables from the WorldClim datasets were used to assess current climatic conditions.

Generating pseudo-absence points

Pseudo-absence points were generated using the two-step method (Barbet-Massin et al., 2012; Lobo and Tognelli, 2011; Wisz and Guisan, 2009): First, a suitable area was built using profiling techniques. In this work, the Bioclim method was used to build a suitable area using the *bioclim* function on dismo package in R programming.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

Second, pseudo-absence points were randomly chosen outside of the suitable area (as shown in *Figure 2*).

Table 1. *The variables used in the species distribution models*

Variable	Unit	Description	Variable	Unit	Description
DEM	m	Elevation	Bio5	℃ × 10	Max Temperature of Warmest Month
Distance	km	Distance from the river	Bio6	℃ × 10	Min Temperature of Coldest Month
Aspect	degree	Direction of the slope facing (0-360 degree)	Bio7	℃ × 10	Temperature Annual Range
Class1	-	Sedimentary and Metamorphic rocks, slope 0-15 degree	Bio8	℃ × 10	Mean Temperature of Wettest Quarter
Class2	-	Sedimentary and Metamorphic rocks, slope 15-30 degree	Bio9	℃ × 10	Mean Temperature of Driest Quarter
Class3	-	Sedimentary and Metamorphic rocks, slope 30-45 degree	Bio10	℃ × 10	Mean Temperature of Warmest Quarter
Class4	-	Sedimentary and Metamorphic rocks, slope >45 degree	Bio11	℃ × 10	Mean Temperature of Coldest Quarter
Class5	-	Igneous rocks, slope 0-15 degree	Bio12	mm/year	Annual Precipitation
Class6	-	Igneous rocks, slope 15-30 degree	Bio13	mm/ month	Precipitation of Wettest Month
Class7	-	Igneous rocks, slope 30-45 degree	Bio14	mm/ month	Precipitation of Driest Month
Class8	-	Igneous rocks, slope >45 degree	Bio15	%	Precipitation Seasonality
Humidity	%	Relative humidity	Bio16	mm/ quarter	Precipitation of Wettest Quarter
Bio1	℃ × 10	Annual Mean Temperature	Bio17	mm/ quarter	Precipitation of Driest Quarter
Bio2	℃ × 10	Mean Diurnal Range	Bio18	mm/ quarter	Precipitation of Warmest Quarter
Bio3	%	Isothermality	Bio19	mm/ quarter	Precipitation of Coldest Quarter
Bio4	%	Temperature Seasonality			

DEM was obtained from the Land Development Department, Thailand.
Distance was calculated from river data downloaded from http://www.diva-gis.org.
Aspect was calculated from Elevation.
Class1to Class8 were dummy variables.
Humidity was downloaded from http://www.rsgis.ait.ac.th/~souris/thailand.htm.
Bio1to Bio19 were downloaded from http://www.worldclim.org.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

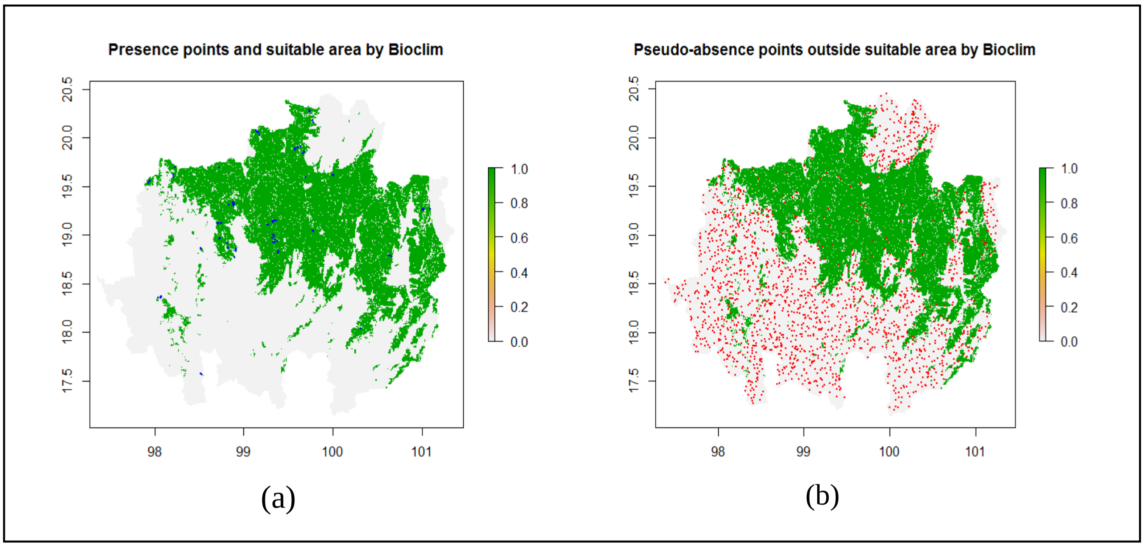

Figure 2. *(a) Suitable areas (green areas) made using the Bioclim method and presence points (blue points); (b) Pseudo-absence points (red points) selected outside the suitable area.*

Logistic regression

Logistic regression is one type of generalized linear model (GLM) that was introduced by Nelder and Wedderburn in 1972. It is suitable for analysis when response data are binary. In this work, the species distribution model was specified by data occurrence (presence data as 1, and absence data as 0). This model used logit link (or logit transformation) to describe the relationships between the response probabilities and the 31 predictor variables (*Table 1*). The logit link is modeled as a linear function:

$$\mathrm{logit}(p) = b_0 + \sum_{i=1}^{n} b_i x_i \qquad \text{(Eq.1)}$$

Where, p is response probabilities to be modeled, n is the number of predictor variables, (xi) and $b_0, b_1, ..., b_n$ are the regression constants. The model was fitted using the maximum likelihood method. The prediction of the model was formed as an exponential function:

$$P(Y) = \frac{e^{\mathrm{logit}(p)}}{1 + e^{\mathrm{logit}(p)}} \qquad \text{(Eq.2)}$$

Where, *P(Y)* is the probability of interested event (probability of presence area). This was transformed into a continuous probability Y ranging from 0 to 1.

In this study, the *glm* function in R programming was used for logistic regression and the *logit* on *glm* function was chosen as the link function.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

Model processing

In *Figure 3*, data including presence points, pseudo-absence points and environmental data were divided into 5 groups labeled with random numbers 1 to 5. Then, group 1 was assigned as testing data and the remaining groups (group 2, 3, 4 and 5) were assigned as training data. The testing data were used to evaluate the models and the training data were used to build the models.

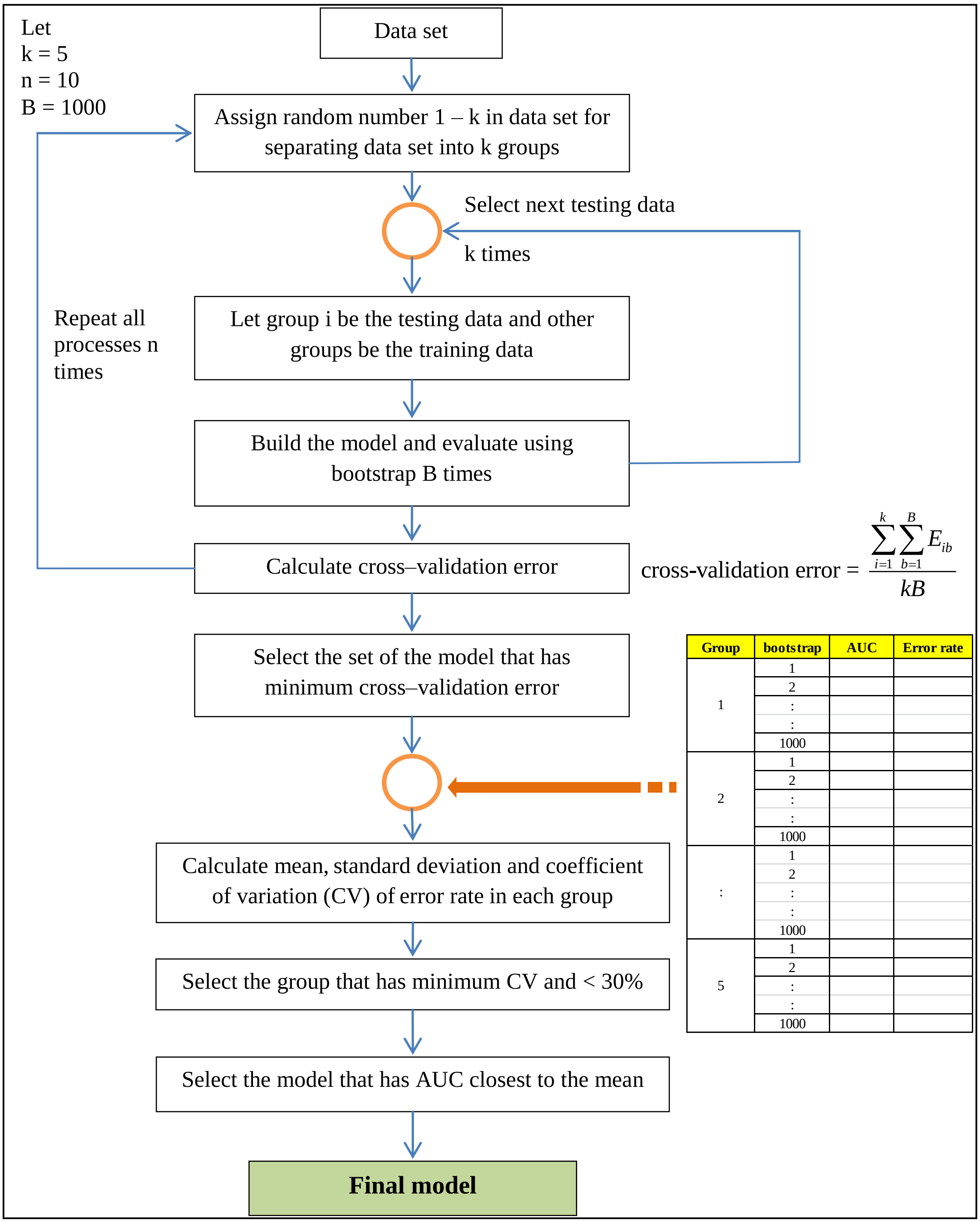

Figure 3. *The model process for selecting the final model for one set of pseudo-absence points with 5 (k)-fold cross-validations, 1000 (B) bootstraps and 10 (n) iterations of all processes*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

We evaluated the models using the evaluate function on dismo package in R programming. Bootstrapping with 1000 replications was needed for the training data to build 1000 models in logistic regression. Next, each of the 1000 models was evaluated using the testing data to provide areas under the receiver operating characteristic curve (AUC) and error rate. After that, the next group (group 2) was used as new testing data and the remaining groups would be new training data. The process was repeated until completing all testing data (5 data groups) and obtaining 5000 models with AUC and error rate values. Cross-validation error was then calculated from the average value of 5000 error rates, and then we reassigned random numbers 1-5 to the data set and repeated all processes 10 times, in order to provide 10 cross-validation error values. The best iteration would be selected from the minimum cross-validation error. Mean, standard deviation and coefficient of variation (CV) of error rates would be calculated within each group that had 5000 models. Next, the group that had a minimum CV at less than 30% would be chosen. CV is normally used to measure dispersion and a CV of less than 30% indicates a normal distribution (Forkman, 2005). Finally, the mean of AUC was calculated and the final model would be selected from the model that had AUC closest to the mean AUC. The final model represents the best logistic regression from one set of pseudo-absence points.

In this study, 5 suitable area maps were generated using the Bioclim method (*Figure 4*). In each map, 20 sets of pseudo-absence points for each suitable area map were generated.

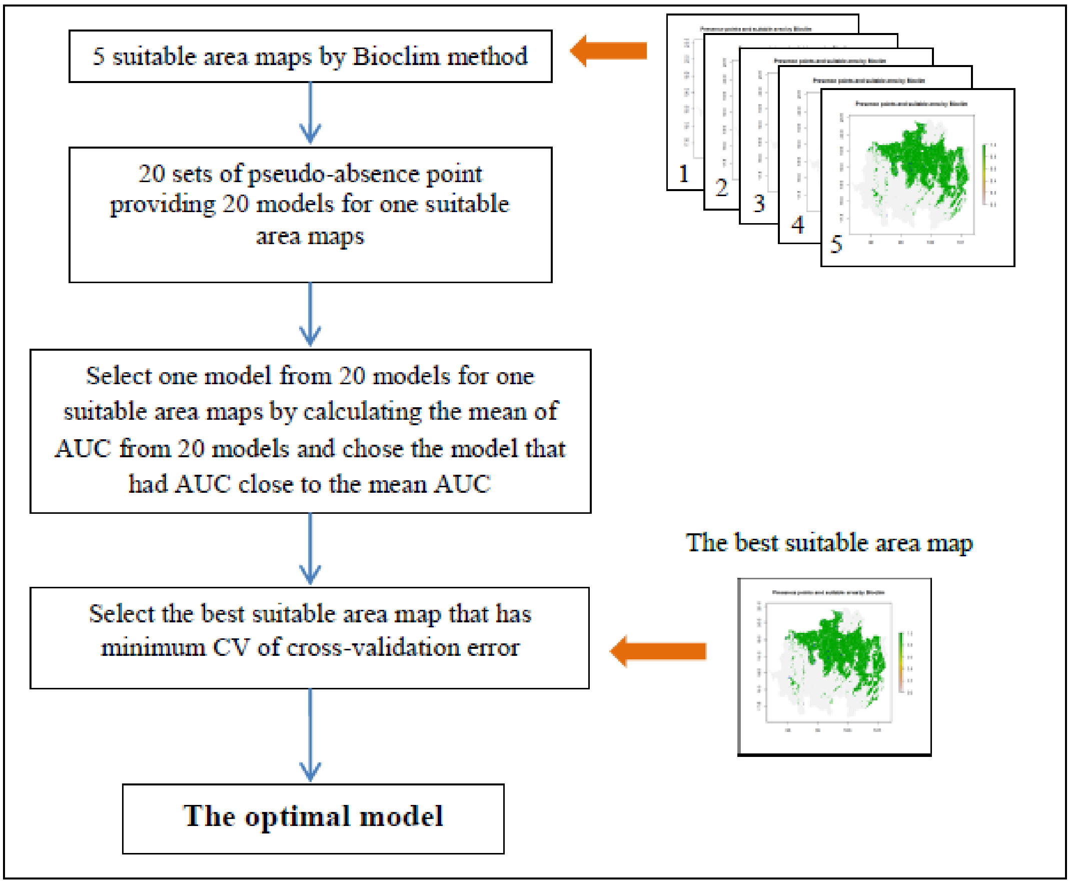

Figure 4. *The process for selecting the optimal model from 5 suitable area maps using the Bioclim method*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

To select the best pseudo-absence set among the 20 sets in each suitable area map, the mean value of AUC was calculated and the model that had AUC closest to the mean was chosen. Next, the best suitable area map was selected for its minimum value of coefficient of variation of cross-validation error. Finally, the optimal model would be processed using the stepwise method to search for effective predictor variables using the step function in R programming along with Akaike's information criterion (AIC). The most suitable model would give the lowest AIC value.

Results and Discussion

Selecting the optimal model

From *Table 2.*, the means of AUC at each suitable area map were similar. Values of 0.93-0.94 were considered most outstanding discrimination (Hosmer and Lemeshow, 2000) and highly successful for an accurate predictive model.

The suitable area map with the minimum coefficient of variation (CV) of cross-validation error was selected in order to reduce bias in the iteration process. Therefore, the group that had minimum CV would have minimum dispersion. Eventually, the most suitable area (map 1) and optimal model in pseudo-absence (set 1) was selected, as seen in the dotted box in *Table 2* below:

Table 2. *The average of AUC and coefficient of variation of cross-validation error at each suitable area map. (The dotted box shows the optimal model from suitable area map set 1.)*

Suitable area map (Bioclim)	Mean of AUC	The best pseudo-absence set	AUC	Coefficient of variation of cross-validation error
1	0.94840	1	0.94838	0.00151
2	0.94027	13	0.93994	0.00307
3	0.93958	16	0.94025	0.00551
4	0.93754	3	0.93768	0.00308
5	0.93978	17	0.93974	0.00326

Factors affecting the model

The stepwise method was used to choose the variables affecting the model, using Akaike information criterion (AIC) after the optimal model had been selected. From *Table 3.*, 18 variables were selected and 10 variables were at a significance level of 0.05. Although some variables (Humidity, Bio2, Bio3, Class1, Class2, Class5, Class6 and Class7) in the model were not apparently significant, they were used in the model due to their presumed involvement, and they were selected in the model using the stepwise method. The most significant variables were topography variables (Aspect and Distance), temperature variables (Bio1, Bio6 and Bio8), and precipitation variables (Bio12, Bio13, Bio14, Bio16 and Bio17), respectively.

Considering topography variables, Aspect, Class1, Class2, Class5 and Class6 had positive effects, while Distance and Class7 had negative effects. Thus, the most suitable area for Assam tea would need more solar radiation because the northwest slope is sunnier and drier than the northeastern slope (Fekedulegn et al., 2003). When

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

considering other topography variables, the most suitable area for Assam tea would need a slope of not more than 30 degree and not be far away from the river or water source.

Table 3. *Factors were selected in GLMs using the stepwise method.*

Factor	Estimate	Sig.	Factor	Estimate	Sig.
Aspect	0.003	*	Bio13	0.064	*
Distance	-0.390	*	Bio14	-0.627	*
Humidity	-0.110		Bio16	-0.107	*
Bio1	0.314	*	Bio17	0.156	*
Bio2	0.260		Class1	17.360	
Bio3	0.538		Class2	17.920	
Bio6	-0.917	*	Class5	17.680	
Bio8	0.532	*	Class6	17.300	
Bio12	0.042	*	Class7	-70.300	
Intercept = -176.800					

“ * ” 0.05 significance level

Considering temperature variables, Bio1, Bio2, Bio3 and Bio8 had positive effects, while only Bio6 had negative effects. The average temperature of Bio1 implicates that the annual average temperature should not be too low for the most suitable area for Assam tea production. Moreover, the fluctuation of temperature between day and night (Bio2) and mean temperature in the wettest quarter (Bio8) should be high, while mean temperature should be sufficiently cold in the coldest month (Bio6). Isothemality (Bio3) is a proportion between the day-to-night temperatures and the summer-to-winter temperatures. Thus, the most suitable area for Assam tea should have high isothemality and/or be far away from the sea.

Considering precipitation variables, Bio12, Bio13 and Bio17 had positive effects, while Bio14 and Bio16 had negative effects. Bio12 indicates that the most suitable area for Assam tea should have sufficient precipitation all year round. It should have high precipitation in the wettest month, but not too high during the whole wettest quarter. Moreover, it should have low precipitation in the driest month, but not too dry during the whole driest quarter. In conclusion, the most suitable area for Assam tea production is possibly in the monsoon regions (Elliott et al, 2006), with two distinct wet seasons and two distinct dry seasons each year.

Predicting map and evaluation

To predict the most suitable area for Assam tea, threshold (or cut-offs) was used to transform the probability map to a presence/absence map by maximizing the sum of the sensitivity and specificity method, and only forest areas were taken into consideration (*Figure 5*). This showed that most Assam tea is greatly distributed in the northwest, towards the North. Most areas are located in Chiang Mai, and in some parts of Chiang Rai and Mae Hong Son provinces.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

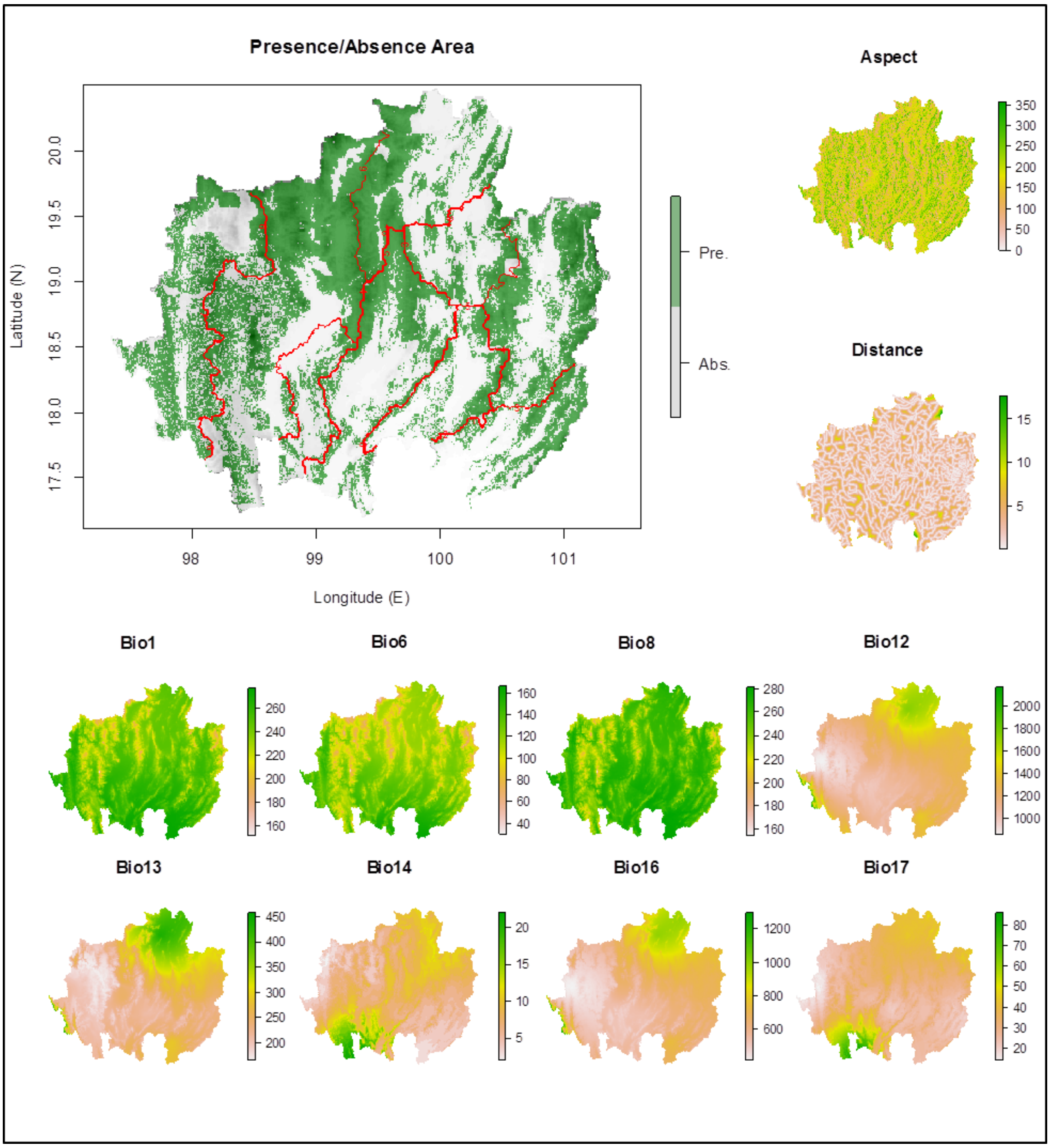

Figure 5. *The most suitable areas for Assam tea forests and effective predictor variables were at a significance level of 0.05*

Conclusion

GLMs depicting the potential distribution of Assam tea with a high AUC value indicates great success as a predicting model. The potential distribution of Assam tea depends upon topographic and climate variables, with Assam tea being dispersed in a forest that has a slope of not more than 30 degrees and is not too far away from a river source. 'Aspect and Distance' were the two most important topography predictors of Assam tea distribution, while Bio1, Bio6 and Bio8 were the most important temperature predictors, along with Bio12, Bio13, Bio14, Bio16 and Bio17 being the most important precipitation predictors. These data implicate that Assam tea grows well in tropical forests and highland areas because the average temperature in the wettest quarter is high

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

and the minimum temperature is low in the coldest month. Thus, potential areas must have sufficient precipitation, but not too much precipitation during the monsoon period. The predictive maps showed taht the most suitable areas were predominantly located in the northwest, towards the North. Climate change is currently a key factor in environmental change. It affects organisms and the biodiversity of the ecological system. Therefore, this study of potential distribution can be beneficial for Assam tea conservation and land management in and around surrounding Assam tea forests. More research is needed to determine whether any available defended areas sufficiently cover the most suitable areas for Assam tea production. The methodological models for Assam tea could be applied to other plant species in other areas of study as well.

Acknowledgements: We would like to thank the Tea Institute and Center of Information Technology Services at Mae Fah Luang University for Assam tea data, and some topographical factors concerning upper Northern Thailand. Moreover, we thank the Remote Sensing and Geographic Information Systems (RSandGIS) Field of Study (FoS), Asian Institute of Technology (AIT), Thailand, for humidity data, and the Land Development Department, Thailand, for elevation data in this study. We thank Dr. Yongyut Trisurat (Professor of Forestry and Director of Biodiversity Center, Kasetsart University, Bangkok, Thailand) for his advice about the species distribution models, and also Phisanu Chiawkhun (Associate Professor of Statistics and Deputy Director of the Office of Educational Quality Development at Chiang Mai University, Thailand) for his advice about the statistical techniques used in this study.

REFERENCES

[1] Barbet – Massin, M., Jiguet, F., Albert, C.H., Thuiller, W. (2012): Selecting pseudo – absences for species distribution models: how, where, and how many?. - Methods in Ecology and Evolution 3:327–338.

[2] Brotons, L., Thuiller, W., Araujo, M.B., Hirzel, A.H. (2004): Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. – Ecography 27:437–448.

[3] Drew, C.A., Wiersma, Y.F., Huettmann, F. (2011): Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. - Springer New York Dordrecht Heidelberg, London.

[4] Elliott, S., Baker, P. J., Borchert, R. (2006): Leaf flushing during the dry season: the paradox of Asian monsoon forests. - Global Ecology and Biogeography 15:248-257.

[5] Engler, R., Guisan, A., Rechsteiner, L. (2004): An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. - Journal of Applied Ecology 41:263–274.

[6] Fekedulegn, D., Hicks Jr, R.R., Colbert, J.J. (2003): Influence of topographic aspect, precipitation and drought on radial growth of four major tree species in an Appalachian watershed. - Forest Ecology and Management 177:409-425.

[7] Forest Land Management Bureau (2013): Forest Land Assessment by Province in 2013. - Royal Forest Department, Bangkok, Thailand.

[8] Forkman, F.J. (2005): Coefficients of Variation – an Approximate F-Test. - Department of Biometry and Engineering, The Swedish University of Agricultural Sciences, Sweden.

[9] Franklin, J. (2010): Mapping species distribution: spatial inference and prediction. - Cambridge University Press, New York.

[10] Guisan, A., Thuiller, W. (2005): Predicting species distribution: offering more than simple habitat models. - Ecology Letters 8:993–1009.

[11] Hijmans, R.J., Cameron, S.E., Parra, J.L., et al. (2005): Very high resolution interpolated climate surfaces for global land areas. - International Journal of Climatology 25(15):1965-1978.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096

[12] Hosmer, D.W., Lemeshow, S. (2000): Applied Logistic Regression. - Wiley, New York.

[13] Kowsuvon, K. (2008): A Case Study of Knowledge Development for Product and Ecology System Integration at Pa Pae Sub-District, Mae Taeng District, Chiang Mai Province, Thailand. - Journal of Soc. Sci. and Human. 2(2)

[14] Lobo, J.M., Tognelli, M.F. (2011): Exploring the effects of quantity and location of pseudo -absences and sampling biases on the performance of distribution models with limited point occurrence data. - Journal for Nature Conservation 19:1 – 7.

[15] Marini, M.A., Barbet-Massin, M., Lopes, L.E., Jiguet, F. (2009): Major current and future gaps of Brazilian reserves to protect Neotropical savanna birds. - Biological Conservation 142:3039–3050.

[16] Mateo, R.G., Croat, T.B., Felicísimo, A.M., Mu˜noz, J. (2010): Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. - Diversity and Distributions 16:84–94.

[17] McCullagh, P., Nelder, J. A. (1989): Generalized Linear Models. - Chapman and Hall, Monographs on Statistics and Applied Probability, London.

[18] Pagella, T., Choocharoen, C., Preechapanya, P., Moss, C., Sinclair, F.L. (2002): Local ecological knowledge about watershed functions in northern Thailand: A guide to using the Agroecological Knowledge Toolkit (AKT). - School of Agricultural and Forest Sciences, University of Wales, Bangor

[19] Pearson, R.G. (2007): Species' Distribution Modeling for Conservation Educators and Practitioners. - Retrieved June 23, 2011. From http://biodiversityinformatics.amnh.org/files/SpeciesDistModelingSYN_1-16-08.pdf

[20] Preechapanya, P. (1996): Indigenous ecological knowledge about the sustainability of tea gardens in the hill evergreen forest of northern Thailand. - Doctoral Dissertation, School of Agricultural and forest Science, University of Wales, Bangor, UK.

[21] Sampanvejsobha et al. (2007): Collection and specify on tea variety grown in northern Thailand project. Tea Institute at Mae Fah Luang University.

[22] Stokland, J.N., Halvorsen, R., Støa, B. (2011): Species distribution modelling – Effect of Design and sample size of pseudo – absence observations. - Ecological Modelling 222:1800 – 1809.

[23] Thuiller, W., Lavorel, S., Arau´ jo, M.B., Sykes, M.T., Prentice, I.C. (2005): Climate change threats to plant diversity in Europe. - Proceedings of the National Academy of Sciences of the United States ofAmerica 102:8245–8250.

[24] Wisz, M. S., Guisan, A. (2009): Do pseudo-absence selection strategies influence species distribution models and their predictions? An information – theoretic approach based on simulated data. - BMC Ecology 9(8).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: 10.15666/aeer/1304_10851096