Faculty of Resource Science and Technology

MODELLING SPECIES DISTRIBUTION OF FROGS (ANURA: RANIDAE) IN SARAWAK USING SECONDARY DATA

Nurul Nazirah binti Azizan

Bachelor of Science with Honours
(Animal Resource Science and Management)
2012

MODELLING SPECIES DISTRIBUTION OF FROGS (ANURA: RANIDAE) IN SARAWAK USING SECONDARY DATA

NURUL NAZIRAH BINTI AZIZAN

This project is submitted in partial fulfillment of the requirements for a
Bachelor of Science with Honours
(Animal Resource Science and Management)

Faculty of Resource Science and Technology
UNIVERSITY MALAYSIA SARAWAK
2012

Declaration

I hereby declare that the thesis is based on my original work except for the citation which has been acknowledged. I also declared that it has not been previously or concurrently submitted for any other degrees at UNIMAS or any other institutions of higher learning.

____

Nurul Nazirah binti Azizan

Animal Resource Science and Management Programme

Department of Zoology

Faculty of Resource Science and Technique

Universiti Malaysia Sarawak.

Acknowledgement

First of all, I would like to express my gratitude to Allah for His finest will for me to finish my final year project smoothly. To my beloved family, especially my parents, brother and sister. I feel very thankful for all their supports and love. Without their supports, I could not finish my project smoothly.

Special thanks to the Faculty of Resource Science & Technology (FRST) for giving this opportunity, to conduct a project to gain experiences and knowledge. Also thanks to Dr. Ramlah Zainudin on behalf of UNIMAS Zoological Museum and Mrs. Margarita Naming from Sarawak Biodiversity Centre (SBC) for granting us permission to collect data records of frogs.

I would like to thank my supervisor, Miss Siti Nurlydia, and other lecturers, especially Dr. Ramlah Zainuddin and Prof. Dr, Indraneil Das for all their guidance and comments, in improving and finishing my project. I am also indebted to all the staffs and postgraduates of Department of Zoology (DZ), especially Mr. Wahap Marni, and Ms. Elvy Quatrin, for all the help and guidance throughout my project. Other than that, my special gratitude to Mr. Ngumbang from Wildlife Conservation Society (WCS) for teaches and guides us about modeling species distribution.

Lastly, I feel grateful to my entire fellow third year students, especially my lovely housemates and course mates for their helps and supports, despite being busy doing their respective project.

Table of Contents

LIST OF ABBREVIATION

UNIMAS : University Malaysia Sarawak

SBC : Sarawak Biodiversity Centre

GLMs : Generalized Linear Models

AIC : Akaike Information Criterion

GAM : Generalized Additive Models

CART : Classification and Regression Trees

ANNs : Artificial Neural Networks

Ha : hectares

MDF : Mixed Dipterocarp Forest

noind : number of individuals

hbt type : habitat type

disturb : disturbance effect

rainfall : seasonal rainfall variation

LIST OF TABLES

LIST OF FIGURES

Modelling Species Distribution of Frogs (Anura: Ranidae) in Sarawak using Secondary Data

Nurul Nazirah binti Azizan

Animal Resource and Management Programme
Faculty of Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

Modelling species distribution of frogs in Sarawak (family: Ranidae) were conducted using secondary data collected from UNIMAS Zoological museum and Sarawak Biodiversity Centre (SBC). For this study, a total of eight species and 753 individuals of frogs from family Ranidae were detected at 16 surveyed localities in Sarawak. Generalized Linear Models (GLMs) using R software was used to analyze environmental variables that affecting the present distribution of each species detected at the surveyed localities, in which Akaike Information Criterion (AIC) used as model selection. In this study, seasonal rainfall variation has the lowest AIC value for distribution of *Hylarana raniceps* (110.605), while for *Staurois guttatus* and *Hylarana picturata*, disturbance effects of habitat type has affecting their distribution with AIC value of 61.812 and 44.163 respectively. Other than that, habitat type had influenced the distribution of *Meristogenys phaeomerus* (25.543) and *Odorrana hosii* (52.009). Habitat types, seasonal rainfall variation and disturbance effects have affecting the distribution of *Hylarana baramica* (32.904), *Hylarana erythraea* (25.519), and *Meristogenys poecilus* (26.539). Distribution maps of each species detected at surveyed localities were also presented.

Keywords: Modelling species distribution, family Ranidae, R software, environmental variables, AIC value.

ABSTRAK

Pemodelan taburan spesies katak dari famili Ranidae telah dijalankan menggunakan data sekunder dari Muzium Zoologikal UNIMAS dan Pusat Biodiversiti Sarawak. Untuk kajian ini, sebanyak lapan spesies dan 753 individu katak dari famili Ranidae telah dikesan di 16 lokaliti yang dikaji di Sarawak. "Generalised Linear Models" (GLMs) menggunakan perisian R telah digunakan untuk menganalisis pembolehubah persekitaran yang mempengaruhi taburan setiap species yang dikesan di setiap lokaliti yang dikaji, di mana "Akaike Information Criterion" (AIC) digunakan sebagai model pemilihan. Dalam kajian ini, variasi hujan bermusim mempunyai nilai AIC yang terendah untuk taburan Hylarana raniceps (110.605). Manakala untuk Staurois guttatus dan Hylarana picturata, kesan gangguan terhadap jenis habitat telah mempengaruhi taburan mereka, masing-masing dengan nilai AIC 61.812 dan 44.163. Selain itu, jenis-jenis habitat telah mempengaruhi taburan Meristogenys phaeomerus (25.543) dan Odorrana hosii (52.009). Jenis-jenis habitat, variasi hujan bermusim dan kesan gangguan telah mempengaruhi taburan Hylarana baramica (32.904), Hylarana erythraea (25.519), and Meristogenys poecilus (26.539). Peta taburan bagi setiap spesies yang dikesan di lokaliti yang dikaji juga dibentangkan.

Kata kunci: Pemodelan taburan spesies, famili Ranidae, perisian R, pembolehubah persekitaran, nilai AIC.

1.0 INTRODUCTION

Frogs are amphibians in the Order Anura. Inger and Stuebing (2005) stated that their distinctive features include no tail, a short and often stocky body, long hind legs and short front ones, large bulging eyes, and a very wide mouth. Consisting of more than 5,000 species described, they are among the most diverse groups of vertebrates (Inger and Stuebing, 2005). About 155 species of frogs occurs on Borneo; this value is believed to increase due to discoveries of new species every year (Inger and Stuebing, 2005; Haas *et al.*, 2010).

Anurans belong to six families, namely, Bombinatoridae, Megophryidae, Bufonidae, Microhylidae, Ranidae and Rhacophoridae (Inger and Stuebing, 2005). The highest number of species in a family is the tree frogs (Rhacophoridae) with 41 species. Except for the Bombinatoridae, these families are widely distributed in Southeast Asia. Each of the Bornean families is divided into group of related species or genera.

Ranidae, or true frogs, have the widest distribution of any frog family. Ranidae occurs almost everywhere in the world and is the largest family in Borneo. According to Inger and Stuebing (2005), this is probably the most familiar group of frogs. In Borneo, Ranidae consists of 5 genera and 26 species that are *Huia, Meristogenys, Staurois, Odorrana* and *Hylarana* (Table 1).

Table 1: List of species from family Ranidae in Borneo (Sources: Haas *et al.*, 2011)

Scientific Names	**English names**
Huia cavitympanum	Hole-in-the-head Frog
Meristogenys amoropalamus	Mountain Torrent Frog
M. jerboa	Western Torrent Frog
M. kinabaluensis	Montane Torrent Frog
M. macropthalmus	Large-eyed Torrent Frog
M. orphnocnemis	Northern Torrent Frog
M. phaeomerus	Brown Torrent Frog
M. poecilus	-
M. stenocephalus	Narrow-headed Torrent Frog
M. stigmachilus	Spotted-lip Torrent Frog
M. whiteheadi	Whitehead's Torrent frog
M. maryatie	Maryati's Torrent Frog
Hylarana baramica	Brown Marsh Frog
H. raniceps	White-lipped Frog
H. erythraea	Green Paddy Frog
H. glandulosa	Rough-sided Frog
H. laterimaculata	Lesser Rough Sided Frog
H. luctuosa	Mahogany Frog
H. nicobariensis	Cricket Frog
H. picturata	Spotted Stream Frog
H. signata	Striped Stream Frog
H. megalonesa	Large White-lipped Frog
Staurois latopalmatus	Rock Skipper Frog
S. guttatus	Black-spotted Rock Frog
S. tuberilinguis	Green-spotted Rock Frog
Odorrana hosii	Poisonous Rock Frog

Species distribution models attempt to provide detailed predictions of distributions by relating presence or abundance of species to environmental predictors (Elith *et al.*, 2006). According to Pearson (2007), species distribution models commonly utilize associations between environmental variables and known species' occurrence records to identify environmental conditions within which populations can be maintained. Most species distribution modeling efforts are still based on generalised linear models (GLMs; McCullagh and Nelder, 1989) or generalised additive models (GAMs; Hastie and Tibshirani, 1990), but there has been much progress in terms of modeling (Elith *et al.*, 2006).

A generalised linear model (GLMs) is a flexible generalisation of ordinary linear regression that allows for response variables that have other than a normal distribution. GLMs can be used to perform univariate analysis of variance with balanced and unbalanced designs, analysis of covariance, and regression, for each response variable. According to Guissan *et al.* (2002), they are based on an assumed relationship between the mean of the response variable and the linear combination of the explanatory variables.

R is an open software environment for statistical computing and graphics (Hijmans and Elith, 2011). The use of R implies the learning of a command language. R provides a wide variety of statistical, for example linear and non-linear modeling, classical statistical tests, time-series analysis, and classification with clustering.

1.1 Problem Statement

This compilation data from previous studies of surveys conducted throughout selected sites in Sarawak will be re-examined using ecological modelling technique to predicting the distribution of frogs (family: Ranidae). However, with limitation in sampling effort mainly due to sampling cost, time and manpower, most field survey can only cover easily-accessed sites, that usually resulting in bias and incomplete database. Other than that, there are also lacks of studies on modelling species distribution using secondary data in Sarawak. This study also requires long term of effort and research to accomplish it.

1.2 Objectives

1) To compile the data collection throughout area of Sarawak on anurans (family: Ranidae) into a report.
2) To indicate the best environmental predictors or variables that influenced the distribution range of anurans (family: Ranidae) in Sarawak.
3) To mapping current distribution of anurans (family: Ranidae) in Sarawak.

2.0 LITERATURE REVIEW

2.1 Checklist of frogs of Borneo

Recently, significant growth in number of persons working in the amazing herpetofauna and a subsequent increase in the number of species to occur in Borneo has been discovered by thoroughly studies and researches. However, there is a lack of a formal platform for herpetologists working in Borneo to gather and confer current herpetological researches (Zug, 1993). Thus, Inger and Tan (1996) have done checklist and references of the frogs of the Borneo which consisting the original source with type locality, the first use of the present combination of genetic and specific name, the name used in the 1966 monograph and their synonyms. Then the distributions were determined based on political division of Borneo (Inger and Tan, 1996). The checklist is prepared to make the current state of information available for the general community for biogeographic and ecological project proposed or underway.

2.2 Distribution of frogs (family: Ranidae)

Analysis of the distribution of frogs within Borneo during past time is more likely to reflect the distribution of collectors instead of the frogs (Inger, 1966). The distribution of frog ranges from tropic to subarctic regions, but most species are found in tropical rainforests (Inger and Stuebing, 2005). Ranidae, which is the largest family in the Borneo and most familiar group of frogs (Inger and Stuebing, 2005), have the widest distribution of any frog family. In general, Ranidae can be found at river area, stream and dwelling at forest floor. Ranidae also dominated peat swamp forest with 66% of the total individuals captured

(Zainudin, 2006). According to Abdullah *et al.* (2006), the land area in Sarawak is covered by 1.6 million ha or 11.9% of peat swamp forest. The high acidity of the water in peat swamp forests are unsuitable for frogs, since they have aquatic stages in their live cycles and have very permeable skins. Yet, some species from family Ranidae, *Hylarana glandulosa*, could adapt in this condition and presumably have a special tolerance with high acidity (Zainudin, 2002).

Frog habitats were affected by the environment condition such as humidity, food resources, sunlight and predators. Frogs are sensitive and their habitats are very specialised. Therefore, habitats that present with all of these requirements became preferred habitat for most anuran species (Zug, 1993). The study conducted at Crocker Range, Sabah shows that *H. picturata* was abundant near the stream (Zainudin *et al.*, 2002). This species live in the forest litter but when turning to adult stage, they approach stream to breed and remained there for duration of their life (Inger and Stuebing, 2005).

2.3 Modelling species distribution

Predictive modeling of species distribution has become an increasingly important tool to address various issues in ecology, biogeography, evolution and more recently, in conservation biology and climate change (Guisan and Thuiller, 2005). Pearson (2007) stated that this modeling approach has proven valuable for generating biogeographical information that can be applied across a broad range of field. Additionally, the distribution and uniqueness of taxa must be considerately assessed in order to establish a good conservation plan (Peterson *et al.*, 2002), thus area representing high concentrations of rare, endangered or endemic species can be also concerned (Wilson, 1988).

Most research on development of distribution modeling techniques has focused on creating models using presence/absence or abundance data, where regions of interest have been sampled systematically (Cawsey *et al*., 2002; Hirzel and Guisan, 2002; Austin and Cunningham, 1981). However, occurrence data for most species have been recorded without planned schemes, and the great majority of these data consists of presence-only records from museum or herbarium collections that are increasingly accessible electronically (Graham *et al*., 2004; Huettmann, 2005, Soberon and Peterson, 2005; Elith *et al*., 2006).

Ecological modeling techniques are appropriate and useful to associate together the observational data from field and remote sensing, with designated mathematical models and computer stimulations (Gross, 1988) to produce a predictive distribution mapping of targeted species. Guisan and Zimmerman (2000) stated that the modeling techniques have included, but are not limited to, such as multivariate ordination, generalised linear models (GLMs), generalised additive models (GAMs), classification and regression trees (CART), genetic algorithms and also artificial neural networks (ANNs).

A study was conducted by Beger and Possingham (2005) to identify the factors that influence the distribution of coral reef fishes using generalised linear models (GLMs) in species distribution modeling. There are four remotely determined environmental predictor variables used to logistic regression models for 227 fish species which are depth, presence of a land-sea interface, exposure, and the distance to the nearest estuary. Akaike's information criterion (AIC) with the lowest value was used as the model selection for the environmental predictor variables that influencing the distributions of coral reef fish species in regional scale. As the results, depth was most frequently efficient variables for single variable models.

Kery *et al.* (2010) have conducted a study on predicting species distributions of blue hawker, *Aeshna cyanea* (Odonata, Aeshnidae) from checklist data using site-occupancy models in Switzerland. They used checklist data collected by volunteers during 1999 and 2000 to analyze the distributions of blue hawkers. Data from repeated visits to 1-ha pixel is used to derive ‘detection histories’ and apply site-occupancy models to estimate the true species distribution, for example, corrected for imperfect detection. For the results, they compared the site-occupancy model with a conventional distribution model based on a generalised linear model, which assumes perfect detectability (p=1). The conventional model undervalue the species distribution by 60%, and the slope parameter of the occurrence-elevation relationship were also undervalue when assuming p=1. Thus, it showed that site-occupancy model has offer a powerful framework for making inferences about species distributions corrected for imperfect detection.

Gray *et al.* (2010) conducted a study using statistical species distribution models to predict habitat suitability of yellow-cheeked crested gibbon in Cambodia based on relationships between environmental variables and species incidence or abundance. Other than that, the study was conducted to examine the species’ habitat associations and tolerance of habitat fragmentation within wildlife sanctuary. Akaike’ information criterion (AIC) was used for corrected small sample sizes and as results, logistic regression models shows that gibbon occupancy was higher in evergreen than in semi-evergreen forests.

3.0 METHODOLOGY

3.1 Species data collection

Eight out of 26 species of frogs from family Ranidae were selected for the distribution prediction (Table 2). The available database of species count and distribution records was obtained from UNIMAS Zoological Museum and Sarawak Biodiversity Centre (SBC), which included data from year 1995 until present. From UNIMAS Zoological Museum, the data record is obtained from previous sampling done from selected sampling sites in various places in Sarawak, which were Padawan, Batang Ai, Santubong, Gunung Gading National Park, Sedilo Forest Reserves, Simunjan, Kota Samarahan, Ranchan, Lambir National Park, Similajau National Park, Sebangkoi and Stutong Recreational Park. At the Sarawak Biodiversity Centre (SBC), the data were collected from previous sampling that covered the Bau district area.

The data collected were assigned using Microsoft Excel 2010, which were arranged into year, locality, sample occasions, species collected, number of species and total individuals (Appendix 1).

Table 2: List of selected species from family Ranidae

Scientific Names	English Names
M. phaeomerus	Brown Torrent Frog
M. poecilus	-
Hylarana baramica	Brown Marsh Frog
H. raniceps	White-lipped Frog
H. erythraea	Green Paddy Frog
H. picturata	Spotted Stream Frog
S. guttatus	Black-spotted Rock Frog
Odorrana hosii	Poisonous Rock Frog

3.2 Data Analysis

3.2.1 R software

R is an open software environment for statistical computing and graphics. The use of R implies the learning of a command language (R Development Core Team, 2005). R provides a wide variety of statistical, for example linear and non-linear modeling, classical statistical tests, time-series analysis, classification and clustering. Tinn-R is probably the best text file editor to use in conjunction with the R programme. It is specifically designed for working with R script files. In addition to syntax highlighting of R code, Tinn-R can interact with R using specific menus and tool bars. This means that sections of commands can be highlighted and sent to the R console (sourced) with a single button click. Tinn-R can be downloaded from the internet at www.sciviews.org/Tinn-R.

3.2.2 Predictor Variables

For this study, the predictor variables were identified and used in the preparation of distribution models (Table 3). Each predictor were typically organised as raster or grid type files representing a variable of interest. Predictor variables were divided into two which are environmental variables and spatial variables. Environmental variables included the habitat types, seasonal rainfall variation, and disturbance effects on the area studied. For the spatial variables, the latitude and longitude is determined by the coordinates. For environmental variables, habitat types were classified into mixed dipterocarp forest, limestone forest, peat swamp forest, and secondary forest. The habitat types were obtained from Forest Department Sarawak (Forest Department Sarawak, 2010) and published journal (Figure 1). These forest types were assigned to value 1, 2, 3, or 4. For the seasonal rainfall variation, 1 indicated the least rainfall while 2 indicated most rainfall respectively. The information of seasonal rainfall in Sarawak was obtained from Malaysia Meteorological Department (Malaysian Meteorological Department, 2010). In Sarawak, the seasonal rainfall variation were categorised into two types of monsoon, which Southwest Monsoon and Northeast Monsoon. Southwest Monsoon occurred in between May until September, which least rainfall and dry season. While Northeast Monsoon occurred from November until March and there was major rainfall and wet season during the months. The disturbance effects indicated whether the forest or habitat types are disturbed or less disturbed and were noted as 1 or 2 respectively.

Table 3: Explanatory variables used to generate the distribution models

Predictor variables (Environmental)	**Maximum-minimum values**
Habitat types	
- peat swamp forest	1
- mixed dipterocarp forest (MDF)	2
- limestone forest	3
- secondary forest	4
- mixed forest (more than 1 forest)	5
Seasonal rainfall variation	
- least rainfall (May-September)	1
- most rainfall (November-March)	2
Disturbance effects	
- disturbed	1
- less disturbed	2

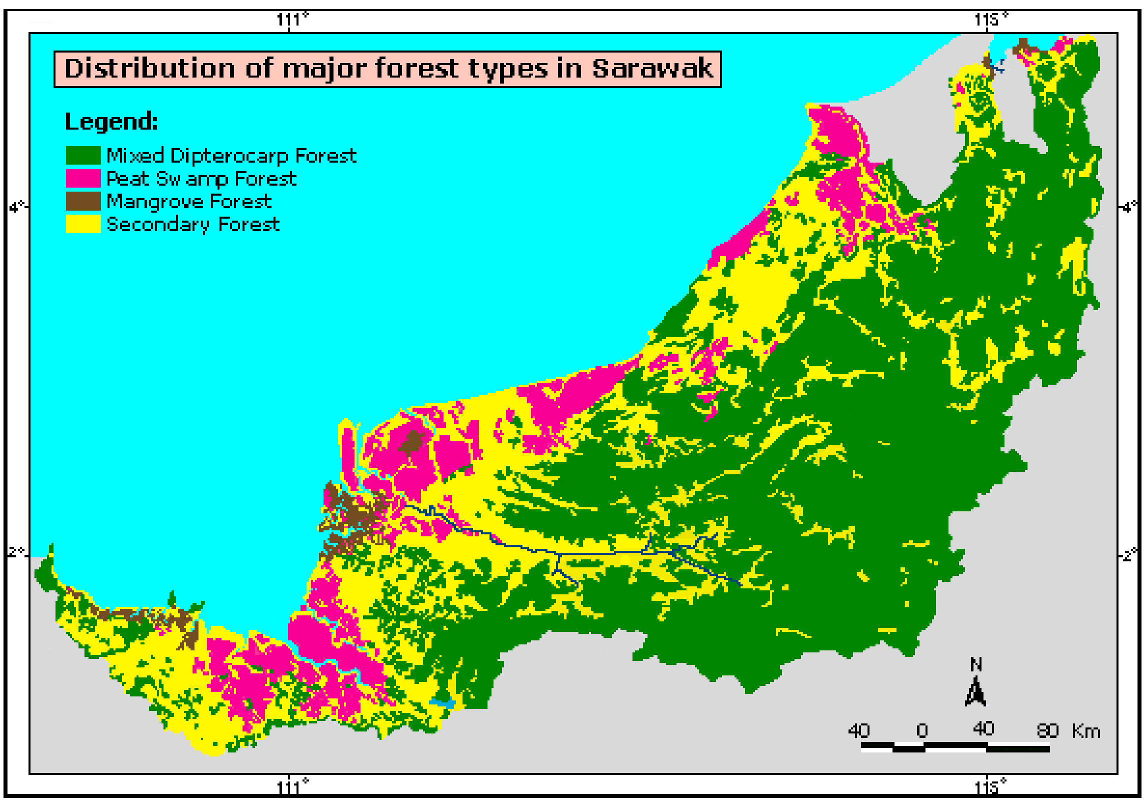

Figure 1: Distribution of major forest types in Sarawak (Sources: Forest Department Sarawak, 2010)

3.2.3 Generalized Linear Models (GLM)

A generalised linear model (GLM) is a generalisation of ordinary least squares regression. Logistic regression is a commonly used technique that performs well in comparison with other modelling techniques (McPerson and Jetz, 2007; Elith et al. 2006). Models are fit using maximum likelihood and by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. All GLM were fitted within the R software. In R, GLM is implemented in the 'glm' function, and the link function and error distribution are specified with the 'family' argument. Examples are:

family = binomial (link = "logit")

family = Gaussian (link = "identity")

family = poisson (link = "log")

These models were generated from presence data only and the information from 3 environmental variables (Table 3). In order to determine whether different classes of environmental variables affecting the current potential distributions than others, current potential distributions for each of the species were run using (a) habitat of forest type variables only, (b) seasonal rainfall variation variables only, (c) disturbance effects variables only, (d) all variables except disturbance effects, (e) all variables except seasonal rainfall variation, (f) all variables except habitat or forest type, and all possible pair-wise combinations of variables classes. Thus, the generally appropriate environmental conditions for the species were established according to the environmental conditions of the observed presence points.

The Akaike information criterion (AIC) was used for model selection, which a high or good fit will result in a low AIC value. In the general case, the AIC is

$$AIC = 2k - 2 \ln (L)$$

where:

k= number is the number of parameters in the statistical model,

L= maximized value of the likelihood function for estimated model.

.