A COMPARISON OF SHANNON'S DIVERSITY INDEX ESTIMATORS UNDER UNSEEN SPECIES IN SAMPLE: STUDY ON MACROFUNGI IN KHEK WATERSHED, PHETCHABUN PROVINCE

CHANTARAT HIRUNKITRANGSEE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE (BIOSTATISTICS)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2008

Thesis
Entitled

A COMPARISON OF SHANNON'S DIVERSITY INDEX ESTIMATORS UNDER UNSEEN SPECIES IN SAMPLE : STUDY ON MACROFUNGI IN KHEK WATERSHED, PHETCHABUN PROVINCE

……………………………………………..
Miss Chantarat Hirunkitrangsee
Candidate

……………………………………………..
Assoc. Prof. Piangchan Rojanavipart,
M.H.S. (Biostatistics)
Major-Advisor

……………………………………………..
Assoc. Prof. Dr. Chukiat Viwatwongkasem,
Ph.D. (Statistics)
Co-Advisor

……………………………………………..
Assist. Prof. Dr. Rujirek Boosarawongse,
Ph.D. (Statistics)
Co-Advisor

………………………………………
Prof. Banchong Mahaisavariya,
M.D.
Dean
Faculty of Graduate Studies

……………………………………………..
Assoc. Prof. Dr. Chukiat Viwatwongkasem,
Ph.D. (Statistics)
Chair
Master of Science Program in Biostatistics
Faculty of Public Health

Thesis
Entitled

A COMPARISON OF SHANNON’S DIVERSITY INDEX ESTIMATORS UNDER UNSEEN SPECIES IN SAMPLE : STUDY ON MACROFUNGI IN KHEK WATERSHED, PHETCHABUN PROVINCE

was submitted to the Faculty of Graduate Studies, Mahidol University
For the degree of Master of Science (Biostatistics)
on
30 April, 2008

…………………………………………...
Miss Chantarat Hirunkitrangsee
Candidate

…………………………………………...
Assoc. Prof. Dr. Saran Petpiroon,
Ph.D. (Marine Pollution and Ecology)
Chair

…………………………………………...
Assoc. Prof. Piangchan Rojanavipart,
M.H.S. (Biostatistics)
Member

………………………………………
Assist. Prof. Dr. Rujirek Boosarawongse,
Ph.D. (Statistics)
Member

…………………………………………...
Assoc. Prof. Dr. Chukiat Viwatwongkasem,
Ph.D. (Statistics)
Member

………………………………………
Prof. Banchong Mahaisavariya,
M.D.
Dean
Faculty of Graduate Studies
Mahidol University

…………………………………………
Assoc. Prof. Dr. Phitaya Charupoonphol,
M.D., Dip.Thai Board (Epidemiology)
Dean
Faculty of Public Health
Mahidol University

ACKNOWLEDGEMENT

The success of this thesis can been attributed to the extensive support and assistance from my major advisor, Assoc. Prof. Piangchan Rojanavipart and my co-advisors, Assoc. Prof. Dr. Chukiat Viwatwongkasem and Assist. Prof. Dr. Rujirek Boosarawongse. I deeply thank them for their valuable advice and guidance in this work.

I would like to thank the Forest Microbiology, Forest Entomology and Microbiology Group, Forest and Wild Plant Conservation Research Office, National Park, Wildlife and Plant Conservation Department, Ministry of Natural Resources and Environment, especially thank to Mrs. Kittima Duengkae, Scientist 7 for the valuable macrofungi data which reported in Songkhlanakarin Journal of Science and Technology, volume 28, number 2.

I am grateful to all the lecturers and staff of the Biostatistics Department in Faculty of Public Health for their valuable advice and thanks to my friends in the 30th and 31st biostatistics for their cheerfulness and kind support.

Finally, I am grateful to my family for the financial support, entirely care and love. The usefulness of this thesis, I dedicate to all the teachers who have taught me since my childhood.

Chantarat Hirunkitrangsee

A COMPARISON OF SHANNON’S DIVERSITY INDEX ESTIMATORS UNDER UNSEEN SPECIES IN SAMPLE: STUDY ON MACROFUNGI IN KHEK WATERSHED, PHETCHABUN PROVINCE

CHANTARAT HIRUNKITRANGSEE 4836026 PHBS/M
M.Sc. (BIOSTATISTICS)

THESIS ADVISORS : PIANGCHAN ROJANAVIPART, M.H.S.(Biostatistics), CHUKIAT VIWATWONGKASEM, Ph.D.(Statistics), RUJIREK BOOSARAWONGSE, Ph.D.(Statistics)

ABSTRACT

The purpose of this study was to estimate the Shannon’s diversity index among the maximum likelihood estimator, the correction bias of maximum likelihood estimator and the Chao & Shen estimator, and also compare their properties in terms of bias, variance, and mean squared error.

Simulation study of each estimator at varied number of individuals and number of species, comparison of properties were indicated that the Chao & Shen estimator and the correction bias of maximum likelihood estimator provided the best choice estimator according to its smallest bias and its smallest mean squared error. Regardless of the number of species, the Chao & Shen estimator gave the best diversity index estimator when the number of individuals was less than 200, while the correction bias of maximum likelihood estimator gave the best diversity index estimator when the number of individuals was larger than 500.

The empirical sets of samples on the macrofungi data in Khek watershed, Phetchabun Province, 2002–2004, were illustrated and statistically analysed. Two selected estimators, the correction bias of maximum likelihood estimator and the Chao & Shen estimator, gave similar quantities of diversity index.

The surveillance system along with calculation of proper diversity index estimators especially on macrofungi groups, should play a role in environmental ecology monitoring of the community.

KEY WORDS: SHANNON’S DIVERSITY INDEX ESTIMATORS / UNSEEN SPECIES / MACROFUNGI / SIMULATION STUDY

85 P.

การเปรียบเทียบตัวประมาณค่าดัชนีความหลากหลายของ Shannon ภายใต้ตัวอย่างที่พบไม่ครบทุกชนิด ศึกษาในเห็ดราขนาดใหญ่บริเวณลุ่มน้ำเข็ก จังหวัดเพชรบูรณ์ (A COMPARISON OF SHANNON'S DIVERSITY INDEX ESTIMATORS UNDER UNSEEN SPECIES IN SAMPLE: STUDY ON MACROFUNGI IN KHEK WATERSHED, PHETCHABUN PROVINCE)

จันทรัตน์ หิรัญกิจรังษี 4836026 PHBS/M
วท.ม. (ชีวสถิติ)

คณะกรรมการควบคุมวิทยานิพนธ์ : เพียงจันทร์ โรจนวิภาต, M.H.S.(Biostatistics), ชูเกียรติ วิวัฒน์วงศ์เกษม, Ph.D.(Statistics), รุจิเรข บุศราวงศ์, Ph.D.(Statistics)

บทคัดย่อ

วัตถุประสงค์ของการศึกษานี้คือการประมาณค่าดัชนีความหลากหลายของ Shannon ตามวิธีของตัวประมาณค่าได้แก่ maximum likelihood, correction bias of maximum likelihood และ Chao & Shen โดยพิจารณาคุณลักษณะของค่าความเอนเอียง (bias), ค่าความแปรปรวน (variance) และค่าความแปรปรวนของค่าผิดพลาดระหว่างตัวประมาณค่ากับค่าจริง (mean squared error)

การจำลองสถานการณ์ข้อมูลที่ศึกษาพบว่าตัวประมาณค่าแบบ correction bias of maximum likelihood และตัวประมาณค่าแบบ Chao & Shen ให้ค่าความเอนเอียงน้อยที่สุดและให้ค่าผิดพลาดเฉลี่ยกำลังสองระหว่างตัวประมาณค่ากับค่าจริงต่ำที่สุด ในทุกกรณีของจำนวนชนิดพันธุ์เห็ดราตัวประมาณค่าดัชนีความหลากหลายแบบ Chao & Shen ดีที่สุดเมื่อจำนวนเห็ดรามีน้อยกว่า 200 ต้น กรณีจำนวนเห็ดรามีมากกว่า 500 ต้นตัวประมาณค่าดัชนีความหลากหลายแบบ correction bias of maximum likelihood ให้คุณลักษณะเหมาะสมที่สุด

ข้อมูลเห็ดราขนาดใหญ่ในลุ่มน้ำเข็ก จังหวัดเพชรบูรณ์ พ.ศ. 2545 ถึง พ.ศ. 2547 เป็นตัวอย่างสำหรับการแสดงการวิเคราะห์ทางสถิติ ตัวประมาณค่าดัชนีความหลากหลายที่เหมาะสมคือตัวประมาณค่าแบบ correction bias of maximum likelihood และตัวประมาณค่าแบบ Chao & Shen ซึ่งตัวประมาณค่าทั้งสองนี้ให้ค่าดัชนีความหลากหลายใกล้เคียงกัน

ระบบการเฝ้าระวังรวมถึงการคำนวณค่าดัชนีความหลากหลายด้วยตัวประมาณค่าที่เหมาะสม เป็นประจำทุกปีและฤดูกาล โดยเฉพาะกับเห็ดราขนาดใหญ่ น่าจะเป็นวิธีการติดตามการเปลี่ยนแปลงด้านสิ่งแวดล้อมและนิเวศวิทยา ที่เหมาะสมของชุมชน

85 หน้า

CONTENTS

CONTENTS (Continued)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I
INTRODUCTION

1.1 Background and Problem

A natural environment is one of a major factor influenced in human health. In Thailand, the rapidly economic development and the increase in population have resulted in greater demands for land and goods. These are widely threatened for tropical forest ecosystems and their biodiversity. Especially, the forest is an origin of water sources like a watershed. It is an area of land where all of the water drains into a small stream or wetland, which all living things are inhabitable and where, as human settled, simple logic demanded that they become part of a community. The actions of people who live in a watershed affect the health of the water that run through it. Rainfall washes chemicals, fertilizers, sediment, and other pollutants from the land into water bodies. Fortunately, many parts of Thailand are becoming increasingly active in forest conservation. These reserves are essential not only for prevention of national disasters like floods and droughts, but also storehouses of a vast biodiversity in plants, animals and microorganisms. This may hold the key to our future survival.

Thailand has many forest areas in every region of the country, unfortunately has recently decreased the areas found each year. A whole major forest in Thailand as a tropical forest which is more diverse than temperate woodland, there is a high diversity of organisms. The diversity measurements have an important role in environmental assessment. The major applications of diversity measurement are in nature conservation and environmental monitoring. Diversity indices provide important information about rarity and commonness of species in a community. The ability to quantify diversity in this way is an important tool for biologists or ecologists trying to understand community structure (Magurran, 1988).

A fungus consists of mushrooms, molds and mildews are members of living things. The main role in an overall balance of life is in the recycling of dead organic matter so that it can be reused by plants in the never ending cycle of renewed life. However, there are a few fungi that can cause disease in man such as plant diseases and also create problems to all of us when they attack leather, wood and food products. Some fungi are also used for industrial such as the production of foods, beverages, antibiotics and vitamins.

Statistical techniques have become the analytic methods applied in biomedical science, psychology, education, economics, genetic studies, epidemiology and other areas. A statistical technique can be developed and used in creating estimations for comparison of the diversity index. In biological science field of study, the Shannon diversity index is the most widely used species diversity measure. This index was developed by Claude Shannon in the late 1940’s and has been useful in comparison of diversity between different habitats or communities (Shannon, 1948). However, a biological community usually has some rare species may not be discovered in a sample of individuals because of the existence of many rare species. Studies which concerned with the estimation of Shannon diversity index when the number of species and the species abundances are unknown still of interested. The traditional estimator that ignores the missing species underestimates when there is a non-negligible number of unseen species. There are some estimators under considered to compare the performance. These estimators are studied in order to select the robust estimator of the diversity index putting into biological surveillance system. By statistical theory, these are some suggested estimators:

First, the Maximum likelihood estimation (MLE) is originally developed by R.A. Fisher in the 1920s, states that the desired probability distribution is the one that makes the observed data most likely. MLE has many optimal properties in estimation such as sufficiency (complete information about the parameter of interest contained in its MLE), consistency (true parameter value that generated the data recovered asymptotically, i.e. for data of sufficiently large samples) and efficiency (lowest-possible variance of parameter estimates achieved asymptotically).

Second, the Correction bias of Maximum likelihood estimation (CMLE) is a method which is well known that MLE may be biased when the sample size n or the total number of individuals is small. The bias is usually ignored in practice, the justification being it is small compared with the standard error (SE). In small or moderate-sized samples, however, a bias correction can be applicable.

Third, the Chao & Shen method is a technique for estimation of Shannon's diversity index when the number of species and species abundances are unknown. The method is the concept based on the Horvitz-Thompson (1952), which has been used to adjust the effect of unobserved sampling units in an unequal sampling scheme. The concept of sample coverage is used here for adjustment of the sample fraction of unseen species. So, this method aims to estimate species (S) via the sample coverage estimation and the approach separates the observed frequencies into two groups as abundant and rare species.

Fourth, the Jackknife method, likes the bootstrapping, is a technique of resamplings and also used for estimating the bias and for deriving robust estimates of standard errors and confidence intervals. This technique was also developed as a general method to reduce the bias of a biased estimator. Although the jackknife is a less general technique than the bootstrap, and explores the sample variation differently but the jackknife is easier to apply to complex sampling schemes, such as multi-stage sampling with varying sampling weights, than the bootstrap.

The simulation technique is used for computing the mathematical model in order to evaluate the performance of estimators in terms of bias and efficiency.

1.2 Objectives

This study has the following objectives:

1. To determine the Shannon's diversity index along with the MLE, CMLE, Chao & Shen, and Jackknife estimation on macrofungi in Khek watershed, Phetchabun Province in 2002 to 2004.
2. To compare the performance of Shannon's diversity index estimators among these methods: maximum likelihood estimator (MLE), the correction bias of maximum likelihood estimator (CMLE) and the Chao & Shen estimator.

1.3 Scope of the Study

The data information of macrofungi was conducted in the Khek watershed, Phetchabun Province from 2002 to 2004. The set of information was routine project operation and kindly obtained from the Department of National Park, Wildlife and Plant Conservation, Ministry of Natural Resources and Environment. Original data was collected 100 square metre (m^2) quadrat, located in permanent plot in each 5 different forests i.e. dry evergreen forest 1, secondary dry evergreen forest, dry evergreen forest 2, dry evergreen forest with pine and pine forest. These five forests settle in two areas: Khao Kho Wildlife Captive Breeding Station and Thung Salaeng Luang National Park. Each location was collected the data at three seasons of the year (dry, rain and cool).

1.4 Definition of Terms

Community: An assemblage of population living in a prescribed area or physical habitat, inhabiting some common environment. An organized unit in possessing characteristics additional to its individual and population components, functioning as a unit in terms of flow of energy and matter (Abercrombie et al., 1990).

Fungi: The kingdom which includes mushrooms, mildews, yeasts, etc. They are either unicellular or composed of masses of fine filaments (hyphae) and reproduce by means of spores (Allaby, 1994).

Forest: An extensive area of woodland which is either managed or maintained for the production of timber, etc. The adjectives broadleaved, coniferous, or the name of a particular tree (e.g. pine) are applied to woodlands in which at least 80% of the canopy consists of the tree the name describes, the canopy of mixed forest contains at least 20% of trees other than the dominant species. Rain forest is evergreen forest growing in regions of high rainfall where dry season is short or absent (Allaby, 1994).

Species Richness (S): The oldest and the simplest concept of species diversity that is the number of different organisms present in the community (Krebs, 1999).

Shannon's diversity index (H): A measurement takes into account species richness and proportion of each species within the local community. The index comes from information science. It has also been called the Shannon index or the Shannon-Wiener index in the ecological literature. This diversity index is the most popular measures of species diversity are based on information theory. It should be used only on random samples drawn from a large community in which the total number of species is known (Krebs, 1999).

Simpson's diversity index (D_S): A measurement that accounts for the richness and the percent of each species from a biodiversity sample within a local community. The index assumes that the proportion of individuals in an area indicate their importance to diversity. This diversity index measures the probability that two individuals picked at random from a sample will belong to the same species (Krebs, 1999).

Maximum likelihood estimator (MLE): It is to determine the parameters that maximize the probability (likelihood) of the sample data. From a statistical point of view, the method of maximum likelihood is considered to be more robust (with some exceptions) and yields estimators with good statistical properties. In other words, MLE methods are versatile and apply to most models and to different types of data. In addition, they provide efficient methods for quantifying uncertainty through confidence bounds (ReliaSoft Corporation, c1998-2007).

Simulation: A quantitative technique that utilizes a computerized mathematical model in order to represent actual making under conditions of uncertainty for evaluating alternative courses of action based upon facts and assumptions (Thieraut and Klekamp, 1975).

Performance of Shannon's index estimators in this study can be defined in the following terms:

Bias: The difference between the expected value of a sample statistic and the population parameter (Vogt, 1993).

Variance: A measure of the spread of scores in a distribution of scores, that is, a measure of dispersion. The larger the variance, the further the individual cases are from the mean. The smaller the variance, the closer the individual scores are to the mean. Especially, the population variance is the mean of the sum of squared deviations from the mean score. Taking the square root of the variance gives the standard deviation (Vogt, 1993).

Mean Squared Error (MSE): It is the expected value of the square of the "error". The error is the amount by which the desired estimator differs from the actual observations or population value. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate that is used to assess the accuracy of the estimator or used to determine whether the model does not fit the data. Mean squared error is equal to its variance plus the square of its bias (Vogt, 1993).

CHAPTER II
LITERATURE REVIEW

The study has reviewed literatures based on relevant concepts, theories and researches in the following topics:

- 2.1 Geographical Descriptions
    - 2.1.1 The Khek River
    - 2.1.2 Thung Salaeng Luang National Park
    - 2.1.3 The Ecological Zones of forest in Thung Salaeng Luang National Park
    - 2.1.4 Forest Location of Macrofungal data in Khek watershed, Phetchabun
    - 2.1.5 The five permanent plots of fungi collection
- 2.2 Statistical Analysis of Diversity Measurement
    - 2.2.1 Conventional Index
    - 2.2.2 Estimation of Shannon index and estimators
- 2.3 Relevant Researches

2.1 Geographical Descriptions

2.1.1 The Khek River

The Khek River (or Kaek River see Glaubrecht and Köhler 2004) has its water source in the Phetchabun Mountain Range in Khao Kho District. The river has run across Thung Salaeng Luang National Park and made the famous Sri Dit waterfall in Phetchabun and Kaeng Sopha waterfall in Phitsanulok. The water course passing Wang Thong District and then named Wang Thong River, later joined the Nan river at Bang Krathum District. This river is a tributary of the Nan River in Central Thailand running further south into the Chao Praya, which is the central drainage of Thailand, with its larger part being a large slow river flowing over a broad agriculture area. The

Khek River has cut into its bedrock and transects several small mountain chains composed of Jurassic sandstone and Permian limestone hills. It flows in an east–west direction from the watershed west of Phetchabun to Phitsanulok. Parts of the Khek River are situated within the Thung Salaeng Luang National Park. This river has formed the famous waterfalls through along this river line such as Sakunothayan Fall, Kaeng Song Rapid, Poi Fall, Kaeng Sopha Fall and Sri Dit Fall.

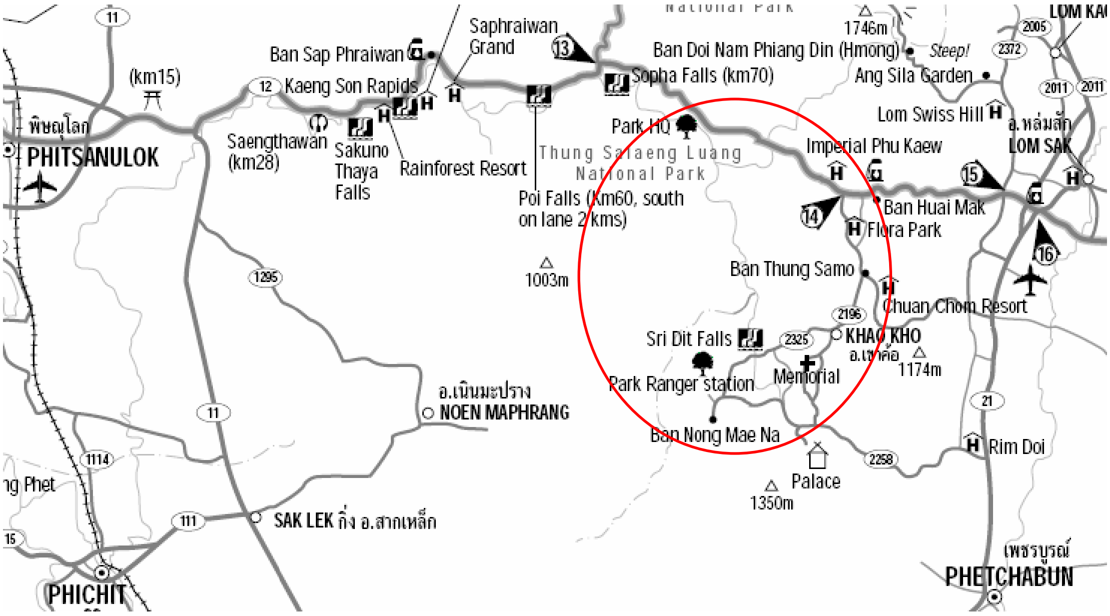

Figure 1. Photography of the Khek watershade

Source: The offices of the Tourism Authority of Thailand (TAT) in *Budget WorldClassDrives*, copyright 2007 by Worldclass Rent a Car Co., Ltd.

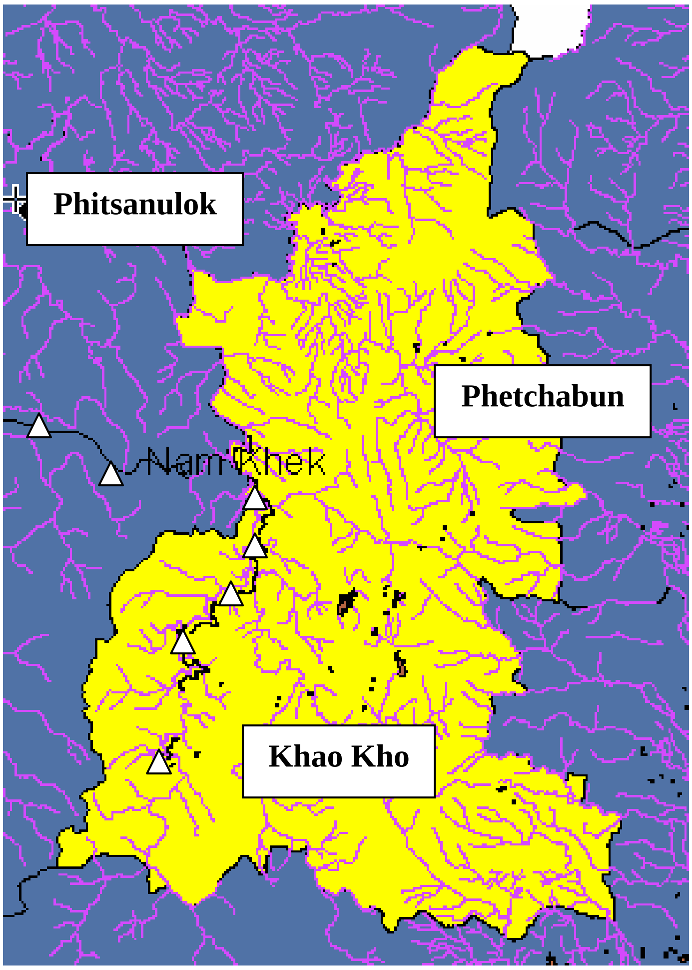

Figure 2. Photography of the river chains in Khao Kho District

Source: The Ministry of Public Health via GIS

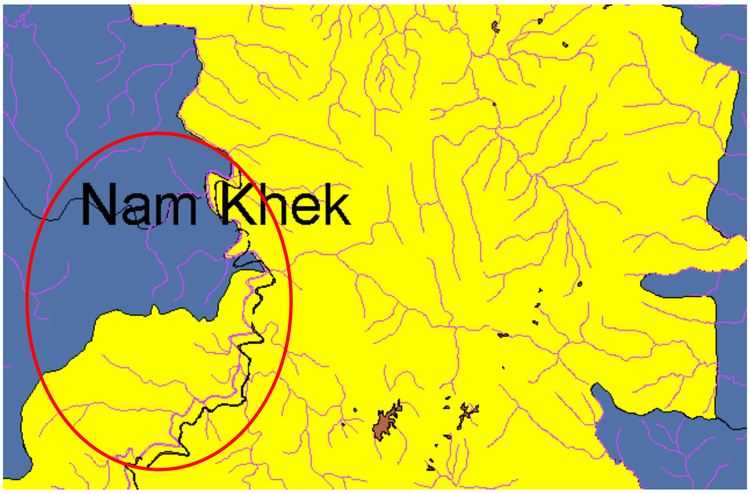

Figure 3. Photography of the shape of Khek watershade

Source: The Ministry of Public Health via GIS

2.1.2 Thung Salaeng Luang National Park

In 1959, 14 important forests were proposed for inclusion in national parks. Eventually, in 1972, the 1,262.55 square kilometers of these areas were declared the Thung Salaeng Luang National Park. The area covered Wang Thong and Nakhon Thai Districts of Phitsanulok and also in Lom Sak, Mueang and Chon Daen Districts of Phetchabun. The park is naturally endowed with pine and bamboo forest as well as mixed species deciduous forest predominate in the area surrounding the river with grassland, lowland scrub and tropical broadleaved evergreen forest covering smaller areas. Most of this area composed with mountain ranges where are watersheds of several streams that are a source of the famous Maenam Nan (National Park Wildlife and Plant Conservation Department, c2005).

2.1.3 The Ecological Zones of forest in Thung Salaeng Luang National Park

Thailand has abundant biodiversity, the variety in genes species and ecosystems, is both rich and naturally abundant. The country consists of a varied geology, very wet and very dry climates. The biodiversity of Thailand has substantial economic values, although they have yet to be fully appreciated or integrated into development planning (The Office of Natural Resources and Environmental Policy and Planning; ONEP, 2004). Thung Salaeng Luang National Park is one of National Parks in Thailand where reserves a vast of biodiversity. This area contains has a lot of hills, with several kinds of forests and wildlife. However, the trees found in Thung Salaeng Luang can be divided into seven kinds altogether as follows:

1) Hill evergreen forest is growing near the mountain virgin forest, which is found in the level of around 1,000 metres above the sea. An example of tree in this forest is the Fagaceae, two-leaf pine tree etc.

2) Tropical rain forest is growing near water sources or in the damp virgin forest, which is found in the level of around 800 metres above the sea. An example of tree in this forest is the Dipterocarpaceae.

3) Dry evergreen forest is almost growing everywhere from the level of 500 metres above the sea. An example of tree in this forest is eaglewood.

4) Pine forest is growing in the level of about 700-900 metres above the sea, which is found near a mountain pinery with some plants like grass and ferns.

5) Mixed deciduous forest is growing in the level of around 400-700 metres above the sea which is found Pterocarpus trees and bamboos in another level.

6) Deciduous dipterocarp forest is found in the level of around 400 metres above the sea, which is dipterocarpaceae trees and wild mango trees are growing.

7) Grassland is an open meadow and consists of some special kind of plant, like grass.

2.1.4 Forest Location of Macrofungal data in Khek watershed, Phetchabun

Phetchabun borders on three regions, the North, the Central and the Northeast. The central part of the province comprised with mountain ranges running along both the western and eastern sectors and this province is administratively divided into the following districts: Muang, Lom Sak, Lom Kao, Chon Daen, Nong Phai, Wichian Buri, Si Thep, Bueng Sam Phan, Wang Pong, Nam Nao and Khao Kho.

Thung Salaeng Luang National Park covers an area of 16 square kilometres, the park is naturally endowed with pine trees and wild flowers which blossom in virgin jungles. The parks mountain ranges are watersheds of several streams which are a source of the famous Maenam Nan. Its area spreads across Phitsanulok and Phetchabun provinces. The park headquarters located at Km. 80 on Highway No. 12 is where information on the park and accommodations can be obtained. Most of the park's attractions are superb for trekking. The route starts from the sub-office at Ban Nong Mae Na (off Highway No.12 at Km. 100 marker) to Khao Kho. The flatland is accessible through a 16 kilometers clay road and a walking route of 15 kilometers from Ban Nong Mae Na. There is also a study trail at the Mae Na park station which is superb for exploring the entire ecological system of the park.

Various macrofungi were collected from 5 permanant plots of 100 m^2 neighbour to the Khek Watershed in Khao Kho District, Phetchabun (Duengkae, 2006).

2.1.5 The five permanent plots of fungi collection.

Description of 5 studied plots:

Plot 1. Dry evergreen forest 1 (T1) located in Khao Kho Wildlife Captive Breeding Station, Khao Kho District, Phetchabun. This area was found at 739 metres (m) above mean sea level (MSL) and has a lot of plant trees are growing such as *Dalbergia fuca* Pierre and *Hopea ferrea* Pierre.

Plot 2. Secondary dry evergreen forest (T2) located in Khao Kho Wildlife Captive Breeding Station, Khao Kho District, Phetchabun. This area was found at 752 m above MSL and has a lot of plant trees are growing such as *Colona flagacarpa* Craib and *Dimocarpus longan* Lour.

Plot 3. Dry evergreen forest 2 (T3) located in Thung Salaeng Luang National Park, Nong Mae Na Sub-district, Khao Kho District, Phetchabun. This area was found at 760 m above MSL and has a lot of plant trees are growing such as *Dipterocarpus costatus* Gaertn.f., *Dalbergia fuca* Pierre, *Hopea ferrea* Pierre and *Dialium cochinchinense* Pierre.

Plot 4. Dry evergreen forest with pine (T4) located in Thung Salaeng Luang National Park, Nong Mae Na Sub-district, Khao Kho District, Phetchabun. This area was found at 873 m above MSL and has a lot of plant trees are growing such as *Pinus merkusii* Jungh. & de Vriese, *Dalbergia fuca Pierre and Lithicarpus* sp.

Plot 5. Pine forest (T5) located in Thung Salaeng Luang National Park, Nong Mae Na Sub-district, Khao Kho District, Phetchabun. This area was found at 915 m above MSL and has a lot of plant trees are growing such as *Pinus merkusii* Jungh. & de Vriese.

2.2 Statistical Analysis of Diversity Measurement

A Mathematical Formulation of Species Diversity, (Magurran 1988 and Krebs 1999)

Suppose there are S species, which label as species 1, 2, …, S.

Suppose that there are X_i individuals in species i^{th}.

Let N denote the total number of individuals, so $N = X_1 + X_2 + \ldots + X_S$ and the proportion of individuals in species i^{th} is $\hat{\pi}_i = \frac{X_i}{N}$.

2.2.1 Conventional Index

The Shannon's diversity index H (Shannon, 1948) is defined by

$$\hat{H} = -\sum_{i=1}^{S} \frac{X_i}{N} \log_e \left(\frac{X_i}{N} \right)$$

or;

$$\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e \left(\hat{\pi}_i \right)$$

where $\hat{H}$ is the estimate of Shannon's diversity index,

S is the number of species,

$\hat{\pi}_i$ is the estimate of proportion of species i^{th} in a community.

Since the theoretical maximum value of H is $\log_e(S)$, and the minimum value (when $N \gg S$) is $\log_e\left[N/(N-S) \right]$.

The variance for H can be estimated as:

$$\widehat{\text{var}\,H} = \frac{\sum_{i=1}^{S} \hat{\pi}_i \left(\log_e \hat{\pi}_i \right)^2 - \left(\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i \right)^2}{N}.$$

Note that the classical Shannon's diversity index is estimated by the MLE method.

The Simpson's diversity index D_S (Simpson, 1949)

Let D denote the probability that two individuals picked at random belong to the same species. For an infinite population, this is given by

$$\hat{D} = \sum_{i=1}^{S} \hat{\pi}_i^2$$

where $\hat{D}$ is the estimate of Simpson index,

$\hat{\pi}_i$ is the estimate of proportion of species i^{th} in a community.

Since D is a measure of concentration, its complementary probability can be considered to be a measure of dispersion. So, let $\hat{D}_S = 1 - \hat{D}$, which we shall call the **Simpson's diversity index.** Also, this index $\left(1-\hat{D}\right)$ ranges from 0 (low diversity) to almost 1 or $1 - 1/S$.

The variance for D_S can be estimated as:

$$\widehat{\operatorname{var} D_S} = \frac{4\left[\sum_{i=1}^{S} \hat{\pi}_i^3 - \left(\sum_{i=1}^{S} \hat{\pi}_i^2\right)^2\right]}{N}.$$

2.2.2 Estimation of Shannon index and estimators

Assume that there are *S* species in a community labeled from 1 to *S*. Denote the probabilities of species discovery (or relative abundance) by $(\pi_1, \pi_2, ..., \pi_S)$ where $\sum_{i=1}^{S} \pi_i = 1$. A widely used measure of biological and ecological diversity is Shannon's index of diversity defined by

$$\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e\left(\hat{\pi}_i\right)$$

This Shannon's diversity index is also referred in the literature. Assumed that random sample size of N is taken with replacement from the community and each individual is classified correctly to species identity.

2.2.2.1 Maximum Likelihood Estimator (MLE)

The maximum likelihood method is a conventional approach to infer population parameters. Let X_i, $i = 1, 2, ..., S$ be the number of individuals of the i^{th} species observed in the sample then $(X_1, X_2, ..., X_S)$ is a multinomial distribution where $\sum_{i=1}^{S} X_i = N$. Remark that the summation is just over all the observed species because any unobserved or missing species $(X_i = 0)$ would not contribute to the sum. The missing species are those with zero frequency in the sample.

In case the number of species is known and relatively small, a widely used estimator of H is the maximum likelihood estimator (MLE) given by

$$\hat{H}_{MLE} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e\left(\hat{\pi}_i\right) = -\sum_{i=1}^{S} \frac{X_i}{N} \log_e\left(\frac{X_i}{N}\right)$$

where $\hat{\pi}_i = X_i / N$ is the estimate of proportion of species i^{th} in a community,

$N = X_1 + X_2 + ... + X_S$ is the total number of individuals.

2.2.2.2 The Correction Bias of Maximum Likelihood Estimator (CMLE)

The bias could be removed by adding $(\hat{S} - 1)/(2N)$ to the estimate of MLE. The formula as:

$$\hat{H}_{CMLE} = \hat{H}_{MLE} + \frac{\hat{S} - 1}{2N},$$

where $\hat{S}$ is the estimate of the number of species that can be calculated by the sum of count number $\left(f_i\right)$ of existing fungi species $\left(S_i\right)$ when $X_i > 0$. Another formula of $\hat{S}$ can be found by Chao and Lee (1992) and Chao *et al.* (1993).

2.2.2.3 The Chao & Shen Estimator

From Horvitz and Thompson (1952) when N individuals have been selected with replacement from the community, the probability of the i^{th} species not being discovered in any individual is $1-\hat{\pi}_i$, thus it is not discovered in these N individuals is $(1-\hat{\pi}_i)^N$. In the traditional maximum likelihood estimator (MLE) $\hat{\pi}_i = X_i/N$ then $\sum_{i=1}^{k}\hat{\pi}_i = \sum_{i=1}^{S}\hat{\pi}_i I(X_i > 0) = 1$. So, this implies that any unseen species has zero maximum likelihood approach. This is un-reasonable in many applications where rare species may exist. Then, use the concept of sample coverage to modify the traditional sample proportion, which represents the fraction of the total abundances of the discovered species. And also, can interpret $1-\hat{C}$ as the conditional (on data) probability of discovering a new species if an additional observation (i.e., individual) were to be taken. Using this notation, the sample coverage *C*, is defined as the sum of the probabilities of the observed classes, that is

$$C = \sum_{i=1}^{S} \pi_i I\left[X_i > 0\right].$$

That *C* varies with the sample and is a random variable. A widely used estimator of *C* is

$$\hat{C} = 1 - f_1/N$$

where f_1 is the sum of count values from X_i when $X_i = 1$,

N is the total number of individuals.

Therefore, the Chao & Shen estimator that combines the Horvitz-Thompson adjustment and the concept of sample coverage is

$$\hat{H}_{Chao} = -\sum_{i=1}^{S} \frac{\hat{C}\hat{\pi}_i \log_e\left(\hat{C}\hat{\pi}_i\right)}{1-\left(1-\hat{C}\hat{\pi}_i\right)^N} I\left(A_i\right)$$

where $I\left[A_i\right]$ is the usual indicator function that equation is

$$I\left[A_i\right] = \begin{cases} 1 & \text{, the event } I\left[X_i > 0\right] \\ 0 & \text{, otherwise} \end{cases}$$

2.2.2.4 The Jackknife Estimator

An alternative method of reducing bias is the jackknife methodology (Zahl, 1977). The jackknife method applied to a general estimation problem is as follows. Let $G_1, G_2, ..., G_n$ be observation groups consist of $X_1, X_2, ..., X_S$ for each group and its can be measurements with distribution depending on a parameter H_{Jack} **,** let $H = H\left(G_1, G_2, ..., G_n\right)$ be an estimate of H_{Jack} defined for all values of n and let H be the estimate of H_{Jack} based on all the observations, in this research consider the Shannon's diversity index and the usual formula is

$$\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i \, .$$

The estimate of partial jackknife sample consists of the data set with omitting the i^{th} observation. Let

$$\hat{H}_i^{(-i)} = H\left(G_1, ..., G_{i-1}, G_{i+1}, ..., G_n\right), \quad i = 1, 2, ..., n \, .$$

Form new values, hereafter called pseudo value of each the i^{th} omitted observation,

$$\hat{H}_i = n\hat{H} - (n-1)\hat{H}_i^{(-i)}, \quad i = 1, 2, ..., n$$

where n is the total number of observation groups.

Then, the jackknife estimator is the expected value of pseudo values and the formula is

$$\hat{H}_{Jack} = \sum_{i=1}^{n} \hat{H}_i \big/ n .$$

2.3 Relevant Researches

Chao and Lee (1992) studied a nonparametric estimation technique is proposed to a new class of estimators based on the statistical concept of sample coverage.

Cowell and Coddington (1994) reviewed the use of several nonparametric methods for estimating species richness from samples and found that Chao and Lee (1992) estimators, which based on abundance data, consistently overestimated species richness, especially when sample numbers were low.

Chao and Shen (2003) developed a new estimation of Shannon’s diversity index when the number of species and species abundances are unknown based on combines the Horvitz-Thompson (1952) adjustment for missing species and concept of sample coverage.

Saha and Paul (2005) studied to obtain and develop a correction bias of maximum likelihood estimator of the intraclass correlation parameter and compare it, by simulation in terms of bias and efficiency, with the MLE. The CMLE estimator has superior bias and efficiency properties in most instances. Analyses of a set of toxicological data from Paul and a set of medical data pertaining to chromosomal abnormalities among survivors of the atomic bomb in Hiroshima from Otake and Prentice show, in general, much improvement in standard errors of the CMLE estimates over the other estimates.

Giachini et al. (2004) studied about the abundance and diversity of fungi was assessed based on the collection of basidiocarps during 12 months comprising the spring of 1995 (November, December), and summer (February, March), autumn (April, May, June), and winter (August, September) of 1996, in *Pinus taeda* and *Eucalyptus dunnii* located in the state of Santa Catarina, southern Brazil. They were inventoried in three blocks of a hundred 4 m^2 plots each, for a total of 300 plots per stand per season (four seasons). Plots were re-sampled at each season. The fungi diversity was obtained via the Shannon diversity index (H) that value 1.22, 1.61, 2.04 and 1.2 of fungi diversity under *E. dunnii* and also that fungi diversity under *P. taeda* value 0.53, 0.24, 0.83 and 1.46 in spring, summer, autumn and winter, respectively. Then, the study in each season is concluded that the winter was the season with the highest diversity of species recorded so that occurrence and diversity of EM species were related to season as well as host species.

Richard et al. (2004) studied about the collection and mapping epigeous fruitbodies of both ectomycorrhizal (EM) and saprobic fungi in an old growth *Quercus ilex* L. Mediterranean forest within a permanent transect of 6400 m^2 over three consecutive fruiting seasons. These results were found that there were 234 species, including 166 EM species and 68 saprobic fungi species. The Shannon diversity index (H) value was high that is 5.48.

Zang et al.(2004) studied the survey and comparison the diversity of arbuscular mycorrhizal (AM) fungi in deforested (Mantoushan) and natural forest (Banruosi) land in the subtropical region of Dujiangyan, southwest China. This study was found that a total of 44 species of AM fungi were isolated, and the same number of AM fungus species (34 species) was found in both forests. There was no significant difference in total spore density of AM fungi between the two sites, but the total species richness of AM fungi was significantly higher in the deforested land than in the natural forest land. The Shannon-Weiner index was higher in the natural forest site (2.67) than in the deforested site (2.15). In suggestion that there was little effect of deforestation on the diversity of AM fungi, and that annual herbaceous plants play a major role in maintaining and increasing AM fungus spore density and species richness in deforested land.

Kindt R. (2002) studied to provide a comprehensive description of methods used in biodiversity and ecological analysis. A survey was conducted in western Kenya involving a complete tree census (trees were defined as all woody or ligneous plants, including shrubs and lianas) and collection of information on 201 farms. To analyze species richness, Shannon’s diversity index, Simpson’s diversity index and Berger-

Parker’s diversity index, and Shannon evenness and equitability. Information was recorded on a recorded form in the field, and was entered and checked for errors in a MS Access database. This database generated the datasets that were used as inputs in the subsequent statistical analyses.

Glaubrecht and Köhler (2004) studied systematic, molecular genetics and morphological differentiation of viviparous freshwater gastropods endemic to the Khek River, central of Thailand. Survey eight locations along a 100-kilometre stretch of the Khek River that found all pachychilid species endemic are herein assigned to the genus *Brotia* which has it distribution in mainland Southeast Asia including Sundaland, but not in the Philippines and Sulawesi, Indonesia. In systematic revision, combining a morphological and molecular genetics approach, reveals the sympatric existence of at least seven species of the genus *Brotia* that is widespread in rivers of South-east Asia where usually only two species at the most coexist. There are *B. binodosa* found in Thung Salaeng Luang, Chattrakan, Poi Falls and Kaeng Song, and *B.armata* found in Thung Salaeng Luang , Sakunothayan Falls and Poi Falls.

CHAPTER III

METHODOLOGY

This chapter emphasizes on the important issues as follows:

3.1 Study Design
  3.1.1 Specimen Collection
  3.1.2 Specimen Classification
3.2 Statistical Procedures
  3.2.1 Dataset preparation
  3.2.2 Preparation of dataset for calculation of the 4 estimators
  3.2.3 Simulation process
  3.2.4 Application of the simulated process to the real situations

3.1 Study Design

The cross-sectional study was designed for monitoring of macrofungi presented in Khek watershed during 2002 to 2004. The macrofungi data was kindly provided by the Forest Microbiology**,** Forest Entomology and Microbiology Group. Forest and Wild Plant Conservation Research Office, National Park, Wildlife and Plant Conservation Department, Ministry of Natural Resources and Environment.

3.1.1 Specimen Collection

The macrofungi species were collected by using the 100 square metres quadrat permanent plots located in each five different forests i.e. Dry Evergreen Forest 1, Secondary Dry Evergreen Forest, Dry Evergreen Forest 2, Dry Evergreen Forest with Pine and Pine Forest. Each location plot was studied and collected the data at three consecutive seasons i.e. dry (February – May), rain (June – October) and cool (November – January).

The collected forest stations of macrofungi were located in two areas that are:

1) Khao Kho Wildlife Captive Breeding Station, Khao Kho District, Phetchabun. This station consists of 2 types of forest;

(1) Dry Evergreen Forest 1
(2) Secondary Dry Evergreen Forest

2) Thung Salang Luang National Park, Nong Mae Na Sub-district, Khao Kho District, Phetchabun. This station consists of 3 types of forest;

(3) Dry Evergreen Forest 2
(4) Dry Evergreen Forest with Pine
(5) Pine Forest

All these 5 forests were collected the macrofungi data in dry season 2002 and 2004, in rainy season 2002, 2003 and 2004, in cool season 2002, 2003 and 2004.

3.1.2 Specimen Classification

All fungi specimens were identified according to mycological taxonomy method. Apart from the formal identification, recorded information of substrates is also important in identification. Standard method begins with a photograph of the specimen in *situ*, or as soon as possible. The technique of Macro-morphology was used illustrated with colour, size, smell, taste, latex, cap, gills, stipe, habit and other characters of fungi, Micro- morphology of mushrooms and fleshy macrofungi, may require examination of a “spore print”. The print is useful for overall spore color and as a source of mature spores for microscopic examination and measurements. The gametangia, sporocarp and the anatomy of ascus are all important features in identification and classification of the orders, families, and genera of the Ascomycetes. The Basidiomycetes are mainly classified on the basis of basidial form, fine structure of hyphal cross walls and nuclei, anatomy of the sporocarp, and other minute details.

Texts that are general keys in identifying most of the fungi detailed in Hanlin 1990, Hawksworth et al. 1995, Jordan 1999, Konemann 1999, Laessoe 1998, Largent 1973, Moser 1973, Pegler 1986, Pegler and Spooner 1994, Ruksawong and Fegel 2001 and The Royal Institute 1996. Details of classification and locations of macrofungi in Khek watershed were shown in Appendix A.

3.2 Statistical Procedures

3.2.1 Dataset preparation

The standard Microsoft Office Suite programs were chosen to handle most of the data due to the predominance of the software, the facility to integrate the different applications (database, spreadsheet, etc.) without altering data. The computer software programme is used for the storage and handling of data. Microsoft Excel is the spreadsheet used for data manipulation and various calculations. The conventional index both Shannon's diversity and Simpson's diversity with their standard deviations $(S.D.)$ were manipulated in Microsoft Excel and classified by forest types, seasons and years.

The management of macrofungal data in Khek watershed was classified in two parts. The first part is raw data (number of individuals) which were stored in Microsoft Access, which can be managed to access according to forest types, scientific names of fungi, seasons and years. The second part, the values of diversity's index and $S.D.$ were calculated with Microsoft Excel and were stored in Microsoft Access prepared for convenience in searching at any station point.

The conventional Shannon's diversity index and its standard deviation $(S.D.)$ are calculated as following:

$$\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e\left(\hat{\pi}_i\right),$$

$$\widehat{\operatorname{var} H} = \frac{\sum_{i=1}^{S} \hat{\pi}_i \left(\log_e \hat{\pi}_i\right)^2 - \left(\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i\right)^2}{N}.$$

For the Simpson's diversity index and its standard deviation (S.D.) can be calculated by

the Simpson index is $\hat{D} = \sum_{i=1}^{S} \hat{\pi}_i^2$,

the Simpson's diversity index is $\hat{D}_S = 1 - \hat{D}$ and

$$\widehat{\operatorname{var} D_S} = \frac{4\left[\sum_{i=1}^{S} \hat{\pi}_i^3 - \left(\sum_{i=1}^{S} \hat{\pi}_i^2\right)^2\right]}{N} .$$

3.2.2 Preparation of dataset for calculation of the 4 estimators

The dataset for all forest types classified by each season and each year was created in order to investigate and compare among the 4 estimators i.e. maximum likelihood estimator (MLE), correction bias of maximum likelihood estimator (CMLE), Chao & Shen and jackknife estimator.

1) Maximum Likelihood Estimator (MLE):

Calculation $\hat{\pi}$ and $\log_e \hat{\pi}$ in the Microsoft Excel, multiply these two terms together. Then, sum all the values from the multiplication of $\hat{\pi}$ and $\log_e \hat{\pi}$, gives

$$\hat{H}_{MLE} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e\left(\hat{\pi}_i\right) .$$

Finally, the result from this formula was the value of the Shannon's diversity index of MLE.

2) The Correction bias of Maximum Likelihood Estimator (CMLE):

Calculation of bias for removing from MLE where $\hat{S}$ is the estimate of the number of species that can be calculated by the sum of count number $\left(f_i\right)$ of existing fungi species $\left(X_i\right)$ when $X_i > 0$. Then, the value of the Shannon's diversity index of CMLE was calculated as

$$\hat{H}_{CMLE} = \hat{H}_{MLE} + \frac{\hat{S}-1}{2N}.$$

3) The Chao & Shen Estimator:

Calculation $\hat{\pi}$ and the sample coverage $\left(\hat{C}\right)$ was estimated by $\hat{C} = 1 - f_1/N$, which was setting each notation in the Microsoft Excel. Then, using $\hat{C}$ multiplier with $\hat{\pi}$ and take the natural logarithm of $\hat{C}\hat{\pi}$ that was equivalent to $\log_e \hat{C}\hat{\pi}$. Finally, the value of the Shannon's diversity index of the Chao & Shen estimator was calculated as

$$\hat{H}_{Chao} = -\sum_{i=1}^{S} \frac{\hat{C}\hat{\pi}_i \log_e\left(\hat{C}\hat{\pi}_i\right)}{1-\left(1-\hat{C}\hat{\pi}_i\right)^N} I\left(A_i\right)$$

where $I\left[A_i\right]$ is the usual indicator function as follow

$$I\left[A_i\right] = \begin{cases} 1, \text{ the event } I\left[X_i > 0\right] \\ 0, \text{ otherwise} \end{cases}$$

4) The Jackknife Estimator:

Calculation $\hat{\pi}$ and $\log_e \hat{\pi}$ in the Microsoft Excel for the Shannon's diversity index of all observations and the usual formula is

$$\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i .$$

Then, the following process will be computed $\hat{\pi}$ and $\log_e \hat{\pi}$ with sequentially deletes observations from data set, one at a time, and recalculated the process with each observation missing once until the end of data set. Here are called partial jackknife sample, which consists of the data set with the i^{th} observation removed and the formula is

$$\hat{H}_i^{(-i)} = g\left(G_1, ..., G_{i-1}, G_{i+1}, ..., G_n\right), \quad i = 1, ..., n$$

where $G_1, G_2, ..., G_n$ is observation groups consist of $X_1, X_2, ..., X_S$ for each group and n is the total number of observation groups.

After that calculation of pseudo values of each the i^{th} omitted observation can be defined by

$$\hat{H}_i = n\hat{H} - (n-1)\hat{H}_i^{(-i)}, \quad i = 1, 2, ..., n$$

Finally, calculate the estimate of the jackknife estimator and the formula is

$$\hat{H}_{Jack} = \sum_{i=1}^{n} \hat{H}_i \big/ n .$$

These 4 estimators were prepared to calculate for all forest types in Khek watershed classified by dry, rainy and cool season in 2002, 2003 and 2004.

3.2.3 Simulation process

In order to compare the performance of each estimator, the bias, variance and mean squared error are to be calculated. The simulation process is performed. In this study, the simulation process was based on random uniform model. The number of species, S, was fixed to be 100 as Chao and Shen (2003) then loss varied to be fixed

80, 60, 40 and 20. The abundance model for $(\pi_1, \pi_2, ..., \pi_S)$ was considered, a normalizing constant (c) defined that $\sum_{i=1}^{S} \pi_i = 1$. Considering, a random uniform model $(\hat{\pi}_i = ca_i)$, where $(a_1, a_2, ..., a_S)$ are a random sample from a uniform $(0,1)$ distribution. The sample sizes of individuals (n) for the simulation were considered as n = 50, 75, 100, 200, 500, 1000 and 2000 for all 5 situations (S = 20, 40, 60, 80 and 100) which each situations was selected the sample sizes (n) to entry in simulated process when $n \geq S$. Then, each combination of abundance model and sample sizes, 1000 simulated data sets were generated. The steps in simulation process were divided into three main parts:

1) Parameter Simulated Process

1.1 Generate random numbers for a_i from a uniform distribution for each data set (fixed 20, 40, 60, 80 and 100 species) where $i = 1, 2, ..., S$.

1.2 Multiply a_i and c for estimating the value of π_i, where $c = 1 \big/ \sum_{i=1}^{S} a_i$.

1.3 Using $\hat{\pi}_i$ from the above to continued calculation of $\log_e \hat{\pi}_i$, then multiple together. Thus, the estimate of Shannon's diversity index of parameter (H) for each data set that can be defined by $\hat{H} = -\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i$.

1.4 The coefficient of variation (CV) is calculated as $CV = \sqrt{\sum_{i=1}^{S} (\hat{\pi}_i - \bar{\pi})^2 / S} \Big/ \bar{\pi}$, where $\bar{\pi} = \sum_{i=1}^{S} \hat{\pi}_i / S$.

2) Statistics Simulated Process

2.1 Generate random numbers for X_i from binomial distribution(n, p) which combination of sample size (n = 50, 75, 100, 200, 500, 1000 and 2000) and the relative abundance model $(\pi_1, \pi_2, ..., \pi_S)$, while using the proportional $\hat{\pi}_i$ from the parameter simulated process.

2.2 Calculation $\hat{\pi}_i$ of statistics process by using the generate X_i from the above, thus, the estimate proportion of Shannon's diversity index of statistics can be defined by $\hat{\pi}_i = X_i / N$ where $N = X_1, X_2, ..., X_N$.

2.3 Multiplication between X_i / n and $\log_e X_i / n$, this calculation is the usual formula of the estimate Shannon's diversity index of MLE (H_{MLE}).

2.4 Computation the estimate of Shannon's diversity index of CMLE (H_{CMLE}) by adding $(\hat{S}-1)/(2N)$ to $\hat{H}_{MLE}$, where $\hat{S}$ is the sum of count number (f_i) of existing fungi species (X_i) when $X_i > 0$.

2.5 Calculation $\hat{C}$ for the estimate of Shannon's diversity index of the Chao & Shen estimator (H_{Chao}) by $\hat{C} = f_1 / N$, where f_1 is the sum of count values from X_i when $X_i = 1$.

3) Properties studies of the estimators

The three estimators were calculated the estimate of Shannon's diversity index (H_i) at 1000 iterations (B) for each model situation of S and n. The performance of the 3 estimators (MLE, CMLE and the Chao & Shen estimator) was investigated properties in term of bias, variance and mean squared error (MSE). All of these terms and expected value were calculated in sequencing.

3.1 The expected value of Shannon's diversity index $\left(E\left(\hat{H}\right)\right)$ was calculated as a mean:

$$E\left(\hat{H}\right) = \frac{1}{B}\sum_{i=1}^{B}\hat{H}_i = \bar{H}, \quad i = 1, 2, ..., B$$

where $\hat{H}_i$ is individual estimate of Shannon's diversity index based on i^{th} of the $B = 1000$ iterations.

3.2 The bias of each estimator is the difference between the expected value of the sample statistic (estimator) and the estimate of population parameter. This was calculated as follows:

$$\text{bias}\left(\hat{H}\right) = E\left(\hat{H}\right) - H = \left(\frac{1}{B}\sum_{i=1}^{B}\hat{H}_i\right) - H = \bar{H} - H\ ,$$

where H is the Shannon's diversity index of population parameter.

3.3 The variance (precision) measures the difference between the individual value of the sample statistic (estimator) and expected value of this sample estimator. This was also calculated for each estimator as follows:

$$\text{var}\left(\hat{H}\right) = E\left[\hat{H} - E\left(\hat{H}\right)\right]^2 = \frac{1}{B}\sum_{i=1}^{B}\left(\hat{H}_i - \bar{H}\right)^2 .$$

3.4 The mean squared error uses to measure the difference between the individual value of the sample statistic (estimator) and the population parameter. This was calculated as

$$\text{MSE}\left(\hat{H}\right) = E\left[\hat{H} - H\right]^2 = \frac{1}{B}\sum_{i=1}^{B}\left(\hat{H}_i - H\right)^2 ,$$

(Note that the mean squared error is equivalent to its variance plus the square of its bias. This can be also calculated as $\text{MSE}\left(\hat{H}\right) = \text{var}\left(\hat{H}\right) + \left(\text{bias}\left(\hat{H}\right)\right)^2$.)

Software programming

The simulation syntax software program, the two parts: calculation for parameter and calculation for statistics were in the main program.

There are 9 subroutines in details for calculations and simulations for each estimator along with their properties. The software programming was shown in Appendix D.

3.2.4 Application of the simulated process to the real situations

The dataset of number of species (S) and number of individuals (n) classified by calendar years and seasons will be applied to the most closed situation of simulation condition in order to point out for the best estimator under study along with the 95% confidence interval of each selected estimators.

Both of the maximum likelihood estimator (MLE) and the correction bias of maximum likelihood estimator (CMLE) were used the variance formula from the traditional Shannon's diversity index (H) that can be defined by

$$\widehat{\operatorname{var} H} = \frac{\sum_{i=1}^{S} \hat{\pi}_i \left(\log_e \hat{\pi}_i\right)^2 - \left(\sum_{i=1}^{S} \hat{\pi}_i \log_e \hat{\pi}_i\right)^2}{N} .$$

Exceptionally, there is not closed form for the variance formula of the Chao & Shen estimator but it can be finding by using bootstrap. However, this study was used the standard deviation $(S.D.)$ from finding in simulation process by using S and n are closed to real data and can be defined by $S.D. = CV \times \bar{H}$, where $\bar{H}$ is the mean value of the sample statistic (estimator) for each situation in simulation process.

Therefore, the calculation of 95% confidence interval for all estimators can be used the normal distribution and defined by $\bar{X} \pm Z\left(S.D./\sqrt{N}\right)$, where $\bar{X}$ is the mean value of each selected estimator and each data set in the real data.

CHAPTER IV
RESULTS

The results are classified into three parts.

1. The descriptive study of macrofungi in term of diversity index including conventional Shannon's diversity indexes and other estimators i.e. the maximum likelihood estimator (MLE), the correction bias of maximum likelihood estimator (CMLE), Chao & Shen estimator and jackknife estimator.

The conventional Shannon indexes were calculated and performed for each 5 forest type, season and calendar year.

The diversity index estimators: MLE, CMLE, Chao & Shen estimator and the Jackknife estimator were calculated and performed for each season in 2002, 2003 and 2004.

2. The simulation study

The performance of those estimators was studied on the bias, variance and also the mean squared error via a simulation study.

The statistic simulation process was 1000 simulation trials that based on parameter estimation earlier. The number of individuals called sample of sizes was classified as 50, 75 and 100, which considered small amount of individuals. The sample size of 200 and 500 was also simulated for each estimator considering as medium size of individuals and sample size of 1000 and 2000 were supposed to be large sample size of individuals under study. The MLE, CMLE and Chao & Shen estimator were comparison studied under statistical properties of bias, variance and mean squared error. At each condition of number of species (S) and number of individuals (n), the suggested estimator was selected.

3. The application of simulated data to real situations.

Looking back to the real raw data of macrofungi in Khek Watershed Phetchabun Province, the number of species (S) and the number of individuals (n) were classified for each season of the calendar year 2002, 2003 and 2004. Best selection of estimator to the nearest possible of simulation process was suggested. The diversity indexes at each season and each calendar year were presented along with the 95% CI of each selected estimator.

4.1 Descriptive study of macrofungi diversity index

According to the mycology taxonomy classification, total of 119 species were isolated and recorded, macrofungi species were classified into the Phylum ascomycota (12 species) and the Phylum basidiomycota (107 species). The most common genera were *Microporus*, *Hexagonia*, *Hymenochaete* and *Stereum* that were found in almost every forest type and the year-round. All macrofungi species at each site and each season were shown in Appendix A.

The maximum number of species was found in secondary dry evergreen forest (28 species 201 individuals) during rainy season 2003 and the minimum number of species was found in pine forest (1 species 2 individuals) during cool season 2003. Details in the number of species (S) and the number of individuals (n) classified by each forest type, each season and each calendar year were shown in Appendix B.

4.1.1 The Shannon's diversity index

The conventional Shannon's diversity indices were calculated for the macrofungi diversity classified by 5 forest types, three tropical seasons and in calendar year 2002, 2003 and 2004. The number of species (S) and number of individuals (n) at each classification sites were also shown in Table 1.

Table 1. Shannon's diversity index and standard deviation of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province, classified by forest type, season and calendar year, along with number of species (*S*) and number of individuals (*n*).

Year	Season	Shannon's diversity index (S.D.)				
		Dry Evergreen Forest 1	**Secondary Dry Evergreen Forest**	**Dry Evergreen Forest 2**	**Dry Evergreen Forest with Pine**	**Pine Forest**
2002	Dry	1.08 (0.04) (S=6, n=711)	1.00 (0.04) (S=6, n=712)	1.60 (0.03) (S=9, n=1177)	1.25 (0.08) (S=6, n=157)	NA
	Rain	1.71 (0.07) (S=14, n=269)	2.11 (0.07) (S=14, n=158)	1.80 (0.05) (S=9, n=223)	1.73 (0.09) (S=8, n=83)	1.54 (0.06) (S=11, n=293)
	Cool	0.81 (0.06) (S=7, n=264)	0.46 (0.06) (S=3, n=238)	1.50 (0.03) (S=5, n=177)	0.19 (0.06) (S=2, n=105)	NA
2003	Dry	-	-	-	-	-
	Rain	2.78 (0.04) (S=27, n=497)	2.62 (0.06) (S=28, n=201)	2.55 (0.05) (S=21, n=306)	2.20 (0.06) (S=13, n=186)	1.59 (0.08) (S=9, n=139)
	Cool	1.42 (0.06) (S=6, n=130)	1.66 (0.03) (S=6, n=237)	1.34 (0.05) (S=5, n=142)	1.83 (0.05) (S=10, n=235)	0.00 (0.00) (S=1, n=2)
2004	Dry	1.15 (0.07) (S=4, n=87)	1.44 (0.06) (S=6, n=143)	1.33 (0.04) (S=5, n=156)	0.60 (0.14) (S=3, n=29)	NA
	Rain	1.77 (0.06) (S=10, n=166)	2.43 (0.06) (S=20, n=235)	1.50 (0.09) (S=12, n=215)	1.60 (0.07) (S=10, n=196)	1.30 (0.11) (S=8, n=77)
	Cool	1.07 (0.05) (S=5, n=214)	1.61 (0.05) (S=7, n=174)	1.50 (0.04) (S=5, n=128)	1.47 (0.07) (S=6, n=103)	NA

Note: The data on dry season in year 2003 were censored.
NA = Not enough data for calculation.

In 45 (3×3×5) combinations of calendar year, season and forest type, approximately 80% of the dataset can be calculated for Shannon's diversity index. The maximum Shannon index was 2.78 (0.04) in Dry Evergreen Forest 1 on rainy season in 2003. The minimum Shannon index was 0.19 (0.06) in Dry Evergreen Forest with Pine on cool season in 2002.

Consider seasons, rain season showed highest diversity index whereas dry season seemed to be the second and cool season had the lowest diversity index among the three seasons.

Among the 5 forest types, even the types not so sharp in differences, the secondary dry evergreen forest and dry evergreen forest 2 seemed to be higher than dry evergreen forest 1 and dry evergreen forest with pine, and the pine forest showed the lowest.

The calendar year 2002 gave the approximate in between estimates of diversity index for each forest type and each season. The calendar year 2003 seemed to have the highest estimate of diversity index, especially in the rainy season. Apart from rainy season, cool season also gave higher index than cool season in 2002 and 2004. Unfortunately the data was not available for calculation in dry season 2003. In the year 2004, pattern of diversity index was the same as in 2002 and 2003.

Details of calculations for Shannon’s diversity index and standard deviation were shown in Appendix B, Table B1.

Another conventional diversity index called Simpson’s diversity index (1948) was also calculated and presented in Appendix C, Table C1.

4.1.2 The diversity index estimators: MLE, CMLE, the Chao & Shen and the Jackknife

The maximum likelihood estimator (MLE), the correction bias of maximum likelihood estimator (CMLE), the Chao & Shen estimator and the jackknife estimator were calculated for the macrofungi diversity in calendar year 2002, 2003 and 2004 that each calendar year was classified by 3 seasons of the year. The dataset for each calculation was the summation of all 5 forest types in corresponding year and season.

Table 2. The four Shannon's diversity index estimators of macrofungi summarized all 5 forest types, classified by season in 2002-2004.

Year	Season	Number of species	Number of individuals	Methods of estimation			
				MLE	CMLE	Chao & Shen	Jackknife
2002	Dry	14	2757	1.51337	1.51696	1.51345	1.70048
	Rain	44	1026	3.09591	3.10651	3.10580	3.80271
	Cool	11	784	1.11884	1.12442	1.12960	1.32799
2003	Dry	-	-	-	-	-	-
	Rain	71	1329	3.63134	3.64479	3.63955	4.20777
	Cool	19	746	2.33177	2.34365	2.34146	2.72949
2004	Dry	11	415	1.98141	1.99319	1.98586	2.36582
	Rain	43	889	2.66849	2.67958	2.69444	3.03080
	Cool	13	619	1.72594	1.73631	1.72918	1.88852

Note: The data on dry season in year 2003 were censored.
MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

Considering the four estimators for comparison, MLE was the baseline of Shannon's diversity index and the jackknife estimator gave the highest magnitude. The CMLE and Chao & Shen estimator were more or less rather showed similarity in magnitude. The MLE, CMLE and Chao & Shen estimator rather showed the average of conventional Shannon's diversity index in Table 1. Worthy mentioned the Chao & Shen estimator gave figure very close to the CMLE and MLE, respectively. The patterns of comparison among these 4 estimators played the same role of action in 2002, 2003 and 2004. These interpretations happened in each dry season, rainy season and cool season. Not surprisingly, the number of macrofungi species was highest in rainy season, about 3 times higher than dry season and cool season. The number of species was also highest in rainy season 2003. Summarizing all the data from 5 forest types made the diversity index estimator was higher and smoother than in separate forests.

In summary, the modified Shannon's diversity index estimator e.g. CMLE, Chao & Shen and jackknife estimator gave more diversity index than the traditional diversity index estimator (MLE–baseline). The Shannon's diversity index from jackknife estimator method gave highest magnitude.

4.2 Simulation Study

The simulation process was performed under the random uniform model, 1000 simulation trials, in order to study estimation of each estimator i.e. MLE, CMLE and Chao & Shen. The properties of estimator i.e bias, variance and MSE were also studied to compare the performance among estimators. These estimators and properties were varied on sample of sizes (n) = 50, 75, 100, 200, 500, 1000 and 2000 for every situation.

Situation 1.

Number of species $(S) = 20$ was fixed, 1000 simulation trials under random uniform model, the Shannon diversity index$(H) = 2.80428$ and CV of H = 0.57205.

Table 3. Comparison of estimators and properties when $S = 20$ at each sample size (n) = 50, 75, 100, 200, 500, 1000 and 2000.

Size *n*	**Method**	**Estimate**	**Bias**	**Variance**	**MSE**
50	MLE	2.60377	-0.20051	**0.01057**	0.05077
	CMLE	2.75762	-0.04666	0.01290	**0.01508**
	Chao & Shen	2.78757	**-0.01671**	0.01584	0.01612
75	MLE	2.67484	-0.12944	**0.00815**	0.02490
	CMLE	2.78583	**-0.01845**	0.00928	**0.00962**
	Chao & Shen	2.77785	-0.02643	0.00959	0.01029
100	MLE	2.70703	-0.09725	0.00739	0.01684
	CMLE	2.79305	**-0.01124**	0.00806	**0.00819**
	Chao & Shen	2.77046	-0.03383	**0.00728**	0.00842
200	MLE	2.75917	-0.04511	0.00505	0.00708
	CMLE	2.80464	**0.00036**	0.00522	0.00522
	Chao & Shen	2.77856	-0.02572	**0.00446**	**0.00513**
500	MLE	2.78691	-0.01737	0.00399	0.00429
	CMLE	2.80562	**0.00134**	0.00401	0.00402
	Chao & Shen	2.79072	-0.01356	**0.00377**	**0.00395**
1000	MLE	2.79477	-0.00952	0.00388	0.00397
	CMLE	2.80421	**-0.00007**	0.00389	0.00389
	Chao & Shen	2.79589	-0.00839	**0.00381**	**0.00388**
2000	MLE	2.79993	-0.00436	0.00369	0.00371
	CMLE	2.80466	**0.00038**	0.00369	**0.00369**
	Chao & Shen	2.80023	-0.00405	**0.00367**	**0.00369**

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

The Chao & Shen estimator gave the smallest bias or the nearest value of estimates to parameter(H) at sample size(n) = 50. At sample size(n) = 75, 100, 200, 500, 1000 and 2000, the CMLEs showed the nearest estimations to parameter(H). As the Chao & Shen estimator and the CMLE gave the nearest estimation the smallest bias of each also performed as well.

Consider variance alone, the MLE gave the smallest among estimators at sample size(n) = 50 and 75. The Chao & Shen estimator gave the smallest variance at n = 100, 200, 500, 1000 and 2000.

Consider MSE alone, the CMLE gave the smallest MSE at sample size(n) = 50, 75, 100 and 2000. The Chao & Shen estimator gave the smallest MSE at n = 200, 500, 1000 and 2000.

In summary, situation $S = 20$, the CMLE gave minimum bias at sample size(n) = 75 and over. The CMLE also gave the lowest MSE at sample size(n) = 50, 75, 100 and 2000. The Chao & Shen estimator gave lowest bias at sample size(n) = 50, and also gave lowest variance at sample size(n) = 100 and over. However, the only at sample size(n) = 200 and over, CMLEs gave minimum bias, whereas the Chao & Shen estimator gave minimum variance and minimum MSE.

Situation 2.

Number of species $(S) = 40$ was fixed, 1000 simulation trials under random uniform model, the Shannon diversity index $(H) = 3.49737$ and CV of H = 0.57381.

Table 4. Comparison of estimators and properties when $S = 40$ at each sample size (n) = 50, 75, 100, 200, 500, 1000 and 2000.

Size *n*	**Method**	**Estimate**	**Bias**	**Variance**	**MSE**
50	MLE	3.09120	-0.40617	**0.01284**	0.17781
	CMLE	3.33892	-0.15845	0.01545	0.04056
	Chao & Shen	3.51133	**0.01396**	0.02599	**0.02619**
75	MLE	3.22008	-0.27729	**0.00799**	0.08487
	CMLE	3.41257	-0.08480	0.00955	0.01674
	Chao & Shen	3.49246	**-0.00490**	0.01327	**0.01330**
100	MLE	3.29472	-0.20265	**0.00574**	0.04680
	CMLE	3.45182	-0.04554	0.00680	0.00887
	Chao & Shen	3.48720	**-0.01017**	0.00871	**0.00881**
200	MLE	3.39779	-0.09957	**0.00363**	0.01354
	CMLE	3.48620	**-0.01117**	0.00397	**0.00410**
	Chao & Shen	3.46577	-0.03159	0.00377	0.00477
500	MLE	3.45890	-0.03846	0.00240	0.00387
	CMLE	3.49651	**-0.00085**	0.00245	**0.00245**
	Chao & Shen	3.47287	-0.02450	**0.00217**	0.00277
1000	MLE	3.47779	-0.01958	0.00209	0.00247
	CMLE	3.49694	**-0.00042**	0.00210	**0.00210**
	Chao & Shen	3.48192	-0.01544	**0.00197**	0.00221
2000	MLE	3.48784	-0.00952	0.00190	0.00199
	CMLE	3.49752	**0.00015**	0.00190	**0.00190**
	Chao & Shen	3.48900	-0.00836	**0.00186**	0.00193

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

The Chao & Shen estimator gave the smallest bias or the nearest value of estimates to parameter (H) at sample size (n) = 50, 75 and 100. At sample size (n) = 200, 500, 1000 and 2000, the CMLEs showed the nearest estimations to parameter(H). As the Chao & Shen estimator and the CMLE gave the nearest estimation the smallest bias of each also performed as well.

Consider variance alone, the MLE gave the smallest among estimators at sample size(n) = 50, 75, 100 and 200. The Chao & Shen estimator gave the smallest variance at n = 500, 1000 and 2000.

Consider MSE alone, the Chao & Shen estimator gave the smallest MSE at sample size(n) = 50, 75 and 100. The CMLE gave the smallest MSE at n = 200, 500, 1000 and 2000.

In summary, situation $S = 40$, the MLE gave smallest variance at sample size(n) = 50, 75, 100 and 200. The CMLE gave smallest bias and smallest MSE at sample size(n) = 200, 500, 1000 and 2000. In the same way, the Chao & Shen estimator gave smallest bias and smallest MSE at sample size(n) = 50, 75 and 100 but also gave smallest variance at sample size(n) = 500, 1000 and 2000.

Situation 3.

Number of species $(S) = 60$ was fixed, 1000 simulation trials under random uniform model, the Shannon diversity index $(H) = 3.90165$ and CV of H = 0.57609.

Table 5. Comparison of estimators and properties when $S = 60$ at each sample size (n) = 75, 100, 200, 500, 1000 and 2000.

Size *n*	Method	Estimate	Bias	Variance	MSE
75	MLE	3.48961	-0.41204	**0.00826**	0.17803
	CMLE	3.73834	-0.16330	0.01010	0.03677
	Chao & Shen	3.91000	**0.00835**	0.01689	**0.01695**
100	MLE	3.59212	-0.30953	**0.00588**	0.10169
	CMLE	3.80245	-0.09920	0.00727	0.01711
	Chao & Shen	3.90986	**0.00821**	0.01125	**0.01131**
200	MLE	3.74651	-0.15514	**0.00309**	0.02715
	CMLE	3.87245	-0.02920	0.00353	**0.00439**
	Chao & Shen	3.88059	**-0.02106**	0.00414	0.00458
500	MLE	3.84203	-0.05961	0.00184	0.00539
	CMLE	3.89774	**-0.00391**	0.00192	**0.00194**
	Chao & Shen	3.87088	-0.03077	**0.00168**	0.00263
1000	MLE	3.87133	-0.03031	0.00145	0.00237
	CMLE	3.89999	**-0.00165**	0.00147	**0.00148**
	Chao & Shen	3.87997	-0.02168	**0.00133**	0.00180
2000	MLE	3.88720	-0.01445	0.00130	0.00151
	CMLE	3.90175	**0.00010**	0.00131	**0.00131**
	Chao & Shen	3.88963	-0.01202	**0.00125**	0.00140

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

The Chao & Shen estimator gave the smallest bias or the nearest value of estimates to parameter (H) at sample size 75, 100 and 200. At sample size 500, 1000 and 2000, the CMLEs showed the nearest estimations to parameter (H). As the Chao & Shen estimator and the CMLE gave the nearest estimation the smallest bias of each also performed as well.

Consider variance alone, the MLE gave the smallest among estimators at sample size 75, 100 and 200. The Chao & Shen estimator gave the smallest variance at n = 500, 1000 and 2000.

Consider MSE alone, the Chao & Shen estimator gave the smallest MSE at sample size 75 and 100. The CMLE gave the smallest MSE at sample size 200, 500, 1000 and 2000.

In summary, situation $S = 60$, the Chao & Shen estimator gave smallest bias and smallest MSE, while the MLE gave smallest variance at sample size 75 and 100. Although the MLE also gave smallest variance at sample size 200, but the Chao & Shen estimator gave smallest bias and the CMLE gave smallest MSE. Moreover, the Chao & Shen estimator gave smallest variance, whereas the CMLE gave smallest bias and smallest MSE at sample size 500 and over.

Situation 4.

Number of species $(S)=80$ was fixed, 1000 simulation trials under random uniform model, the Shannon diversity index $(H)=4.18912$ and CV of H = 0.57640.

Table 6. Comparison of estimators and properties when $S=80$ at each sample size(n) = 100, 200, 500, 1000 and 2000.

Size *n*	**Method**	**Estimate**	**Bias**	**Variance**	**MSE**
100	MLE	3.77940	-0.40972	**0.00656**	0.17443
	CMLE	4.02925	-0.15987	0.00792	0.03348
	Chao & Shen	4.20344	**0.01432**	0.01332	**0.01353**
200	MLE	3.98020	-0.20892	**0.00292**	0.04657
	CMLE	4.13876	-0.05036	0.00343	0.00597
	Chao & Shen	4.18215	**-0.00697**	0.00464	**0.00469**
500	MLE	4.10852	-0.08060	0.00155	0.00804
	CMLE	4.18148	**-0.00764**	0.00166	**0.00172**
	Chao & Shen	4.15692	-0.03220	**0.00152**	0.00256
1000	MLE	4.14793	-0.04119	0.00119	0.00289
	CMLE	4.18588	**-0.00324**	0.00122	**0.00123**
	Chao & Shen	4.16245	-0.02667	**0.00107**	0.00178
2000	MLE	4.16992	-0.01920	0.00100	0.00137
	CMLE	4.18930	**0.00018**	0.00101	**0.00101**
	Chao & Shen	4.17407	-0.01505	**0.00095**	0.00118

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

The Chao & Shen estimator gave the smallest bias or the nearest value of estimates to parameter(H) at sample size(n) = 100 and 200. At sample size(n) = 500, 1000 and 2000, the CMLEs showed the nearest estimations to parameter(H). As the Chao & Shen estimator and the CMLE gave the nearest estimation the smallest bias of each also performed as well.

Consider variance alone, the MLE gave the smallest among estimators at sample size(n) = 100 and 200. The Chao & Shen estimator gave the smallest variance at n = 500, 1000 and 2000.

Consider MSE alone, the Chao & Shen estimator gave the smallest MSE at sample size(n) = 100 and 200. The CMLE gave the smallest MSE at sample size(n) = 500, 1000 and 2000.

In summary, situation $S = 80$, the MLE gave smallest variance at sample size(n) = 100 and 200, while the Chao & Shen estimator gave smallest bias and smallest MSE. Moreover, the Chao & Shen estimator gave smallest variance, whereas the CMLE gave smallest bias and smallest MSE at sample size(n) = 500 and over.

Situation 5.

Number of species $(S) = 100$ was fixed, 1000 simulation trials under random uniform model, the Shannon diversity index$(H) = 4.41290$ and CV of H = 0.57527.

Table 7. Comparison of estimators and properties when $S = 100$ at each sample size(n) = 100, 200, 500, 1000 and 2000.

Size *n*	Method	Estimate	Bias	Variance	MSE
100	MLE	3.90650	-0.50640	**0.00682**	0.26293
	CMLE	4.18725	-0.22565	0.00826	0.05859
	Chao & Shen	4.42556	**0.01267**	0.01519	**0.01416**
200	MLE	4.15097	-0.26193	**0.00295**	0.07156
	CMLE	4.33778	-0.07512	0.00357	0.00921
	Chao & Shen	4.41817	**0.00527**	0.00524	**0.00527**
500	MLE	4.31134	-0.10156	**0.00143**	0.01175
	CMLE	4.40074	**-0.01216**	0.00158	**0.00172**
	Chao & Shen	4.38335	-0.02955	0.00158	0.00245
1000	MLE	4.36080	-0.05210	0.00101	0.00372
	CMLE	4.40783	**-0.00507**	0.00105	**0.00107**
	Chao & Shen	4.38220	-0.03070	**0.00091**	0.00186
2000	MLE	4.38819	-0.02471	0.00082	0.00143
	CMLE	4.41233	**-0.00057**	0.00083	**0.00083**
	Chao & Shen	4.39441	-0.01849	**0.00077**	0.00111

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

The Chao & Shen estimator gave the smallest bias or the nearest value of estimates to parameter$\left(H\right)$ at sample size$\left(n\right)$ = 100 and 200. At sample size$\left(n\right)$ = 500, 1000 and 2000, the CMLEs showed the nearest estimations to parameter$\left(H\right)$. As the Chao & Shen estimator and the CMLE gave the nearest estimation the smallest bias of each also performed as well.

Consider variance alone, the MLE gave the smallest among estimators at sample size$\left(n\right)$ = 100, 200 and 500. The Chao & Shen estimator gave the smallest variance at n = 1000 and 2000.

Consider MSE alone, the Chao & Shen estimator gave the smallest MSE at sample size$\left(n\right)$ = 100 and 200. The CMLE gave the smallest MSE at sample size$\left(n\right)$ = 500, 1000 and 2000.

In summary, situation $S = 100$, the Chao & Shen estimator gave smallest bias and smallest MSE, while the MLE gave smallest variance at sample size 100 and 200. Although the MLE also gave smallest variance at sample size 500, but the CMLE gave smallest bias and smallest MSE. Moreover, the Chao & Shen estimator gave smallest variance, whereas the CMLE gave smallest bias and smallest MSE at sample size 1000 and 2000.

4.2.1 The study on properties of the estimators

After studying each property i.e. bias, variance and mean squared error of estimators (MLE, CMLE and Chao & Shen estimator) in the 5 situations of number of species $\left(S\right)$ i.e. 20, 40, 60, 80 and 100 while were simulated for each number of individual (sample size n) i.e. 50, 75, 100, 200, 500, 1000 and 2000.

The suggested estimator which gave the minimum quantities of properties was presented in Table 8.

Table 8. The suggested estimator at each situation classified by number of species (S) and number of individuals (n).

S	***n***	**Smallest MSE**	**Smallest Bias**	**Smallest Variance**
20	50	CMLE	Chao & Shen	MLE
	75	CMLE	CMLE	MLE
	100	CMLE	CMLE	Chao & Shen
	200	Chao & Shen	CMLE	Chao & Shen
	500	Chao & Shen	CMLE	Chao & Shen
	1000	Chao & Shen	CMLE	Chao & Shen
	2000	CMLE, Chao & Shen	CMLE	Chao & Shen
40	50	Chao & Shen	Chao & Shen	MLE
	75	Chao & Shen	Chao & Shen	MLE
	100	Chao & Shen	Chao & Shen	MLE
	200	CMLE	CMLE	MLE
	500	CMLE	CMLE	Chao & Shen
	1000	CMLE	CMLE	Chao & Shen
	2000	CMLE	CMLE	Chao & Shen
60	75	Chao & Shen	Chao & Shen	MLE
	100	Chao & Shen	Chao & Shen	MLE
	200	CMLE	Chao & Shen	MLE
	500	CMLE	CMLE	Chao & Shen
	1000	CMLE	CMLE	Chao & Shen
	2000	CMLE	CMLE	Chao & Shen
80	100	Chao & Shen	Chao & Shen	MLE
	200	Chao & Shen	Chao & Shen	MLE
	500	CMLE	CMLE	Chao & Shen
	1000	CMLE	CMLE	Chao & Shen
	2000	CMLE	CMLE	Chao & Shen
100	100	Chao & Shen	Chao & Shen	MLE
	200	Chao & Shen	Chao & Shen	MLE
	500	CMLE	CMLE	MLE
	1000	CMLE	CMLE	Chao & Shen
	2000	CMLE	CMLE	Chao & Shen

Note: MLE = Maximum Likelihood Estimator
CMLE = Correction Bias of Maximum Likelihood Estimator

Considering MSE, number of species equals to 20, number of individuals was less than and equal 100, the CMLE was the suggested estimator. The Chao & Shen estimator was the suggested estimator if the number of individuals equals or over 200.

In case number of species higher than 20 ($S = 40, 60, 80$ and 100), the Chao & Shen estimator was the suggested estimator at smaller number of individuals (n = 50, 75 and 100) and CMLE was the suggested estimator at higher number of individuals ($n = 200, 500, 1000$ and 2000).

Considering bias, number of species equals to 20, number of individuals was less than and equal 50, the Chao & Shen estimator was the suggested estimator. The CMLE was the suggested estimator if the number of individuals equals or over 75. In case number of species higher than 20 ($S = 40, 60, 80$ and 100), the Chao & Shen estimator was the suggested estimator at smaller number of individuals (n = 50, 75, 100 and 200) and CMLE was the suggested estimator at higher number of individuals ($n = 500, 1000$ and 2000).

Considering variance, almost number of species, the MLE was the suggested estimator if the number of individuals was less than and equal 200. The Chao & Shen estimator was the suggested estimator at higher number of individuals ($n = 500, 1000$ and 2000).

In summary, two dominant estimators, such as CMLE and Chao & Shen estimator, were suggested to analyze the most of situations with the smallest MSE and smallest bias, regardless of the number of species, S.

Chao & Shen estimator was recommended to use when the number of individuals (n) was less than 200, whereas the CMLE was the most appropriated situations when the number of individuals (n) was greater than 500.

The MLE gave only one property of smallest variance in approximate up to number of individuals $(n = 200)$.

4.3 The application of simulated data to real situations

Looking back to the real data of macrofungi species in Khek Watershed Phetchabun Province, the number of species (S) and the number of individuals (n) were classified for each season of the calendar year 2002, 2003 and 2004. In accordance with the summary suggestion of estimator was shown earlier in Table 8, the best selection of estimator to the nearest possible of situation process was selected. The Shannon's diversity index of each selected estimator was presented along with 95% CI of each selected estimator in Table 9.

Table 9. The quantities and 95% CI of suggested Shannon's diversity index estimator of macrofungi species in Khek Watershed Phetchabun Province, presented at each season and calendar year 2002-2004.

Year	**Season**	**Number of species (S)**	**Number of individuals (n)**	**Suggested estimator**	**Diversity index (95% CI)**	
2002	Dry	14	2757	CMLE Chao & Shen	= 1.51696 = 1.51345	(1.51607 – 1.51786) (1.45365 – 1.57324 ^{a})
	Rain	44	1026	CMLE	= 3.10651	(3.10450 – 3.10852)
	Cool	11	784	Chao & Shen	= 1.12960	(1.01765 – 1.24156 ^{a})
2003	Dry	-	-	-	-	-
	Rain	71	1329	CMLE	= 3.64479	(3.64334 – 3.64623)
	Cool	19	746	Chao & Shen	= 2.34146	(2.22691 – 2.45602 ^{a})
2004	Dry	11	415	Chao & Shen	= 1.98586	(1.83227 – 2.13945 ^{a})
	Rain	43	889	CMLE	= 2.67958	(2.67659 – 2.68257)
	Cool	13	619	Chao & Shen	= 1.72918	(1.60342 – 1.85494 ^{a})

Note: CMLE = Correction Bias of Maximum Likelihood Estimator.
a = Calculation of 95% CI by using variance quantity of parameter in simulation process.
- = no data for calculation.

According to the number of species (S) and number of individuals (n) which naturally larger in rainy season, the CMLE was suggested as the diversity index in the year 2002-2004. The Chao & Shen estimator was suggested for dry season and cool season in the year 2002-2004.

In the year 2002 dry season, 2 estimators: CMLE and Chao & Shen gave rather similar quantity of diversity index (1.51696, 1.51345). The 95% CI for CMLE was smaller (1.51607 – 1.51786) whereas the 95% CI for Chao & Shen estimator was wider (1.45365 – 1.57324).

The 95% CI for Chao & Shen estimator was wider whereas the 95% CI for CMLE was more narrowly. The senses of 95% CI for these two estimators were detected in each situation of selection.

CHAPTER V
DISCUSSION

The discussion for this study is divided into four parts. The first part is concentrating on the general results of Shannon’s diversity index. The second part deals with the comparison of 4 Shannon’s diversity index estimators. The third part is simulation results from 3 Shannon’s diversity index estimators. The fourth part is the applications to the real data situations.

5.1 The Shannon’s diversity index

Diversity is an old and popular concept in ecology. In the middle of 20th century, there have been several indices of diversity introduced. Of those, the ratio of number of species to number of individual (or logarithm of either of these quantities), the Simpson’s index (1949), and the index based on Shannon’s index (1948) were of interested. Shannon’s diversity index is the most widely used diversity index in community ecology (Norris and Pollock 1998).

Studying of the Shannon’s diversity index on macrofungi in Khek Watershed, Phetchabun province, the study sites classified into 5 forest types, 3 seasons during the year and calendar year 2002, 2003 and 2004. The highest diversity index showed in rainy season and followed by dry and cool season. Meanwhile, the macrofungi diversity in 2003 is higher than other two years. Straatsma et al. (2001), Lodge et al. (2004) and Richard et al. (2004) showed that season variation and year annual period are the influential factors that effect to the fungal productivity which some fungal species may be transient because they establish themselves for a certain number of years. Temperature and humidity seem to have effect on fruiting in the development of fruit body initials into mature and visible fruit bodies. In dry season and cool season, there is little humidity or moisture, thus, the number of species in the survey has less than rainy season (The Royal Institute 1996). Other factors may be the structure of the forest stand (Zang 2004), and even the host species or the habitat of fungal species (Giachini 2004). The Shannon’s diversity index on rainy season 2003 has highest diversity index in dry evergreen forest 1 (2.78, *SD* = 0.04) and lowest in pine forest (1.59, *SD* = 0.08).

The macrofungi diversity of all datasets between Shannon’s and Simpson’s diversity index give similar results in term of trend of fluctuation not in quantity numbers. There is evidence that Shannon’s diversity index has slightly greater precision than Simpson’s diversity index. The same opinion were purposed in most debate centre on the statistical merits of Shannon’s and Simpson’s indices (Krebs 1999), suggested that Simpson’s diversity index in its sensitivity to rare species. On the contrary, some previous studies such as Mouillot and Lepretre (1999) compared the root mean squared error (RMSE) of Shannon’s and Simpson’s indices under a variety of simulated conditions, and the Shannon index showed a smaller RMSE than the Simpson index, this is agree with Sherwin et al. (2006) that Shannon’s diversity index is robust estimator than Simpson’s diversity index. In a more comprehensive study, Boxrud et al. (2007) indicate that Shannon's index, although not as easily interpreted as Simpson's index, may provide a better objective measure of the performance characteristics of a species method.

5.2 The MLE, CMLE, Chao & Shen and Jackknife estimator

The accumulated data on macrofungi over 5 forests were classified by each season and each calendar year. It was found that the rainy season of all calendar year in 2002, 2003 and 2004, there are higher diversity than dry season and cool season. Especially, rainy season 2003 gave the highest diversity and followed by rainy season 2002 and rainy season 2004, respectively. The wealthy variation of macrofungi species in natural resources reflected into the highest biodiversity index especially in rainy season in 2003. This rainy season in 2003 gave 3 times higher in diversity index than cool season in 2002 which was referred as the smallest number of species in Khek watershed, Phetchabun Province.

The traditional Shannon’s diversity index (MLE-Maximum Likelihood Estimator) was calculated with the equation proposed by Shannon (1948) and this estimator (MLE) is still biased. The correction bias of maximum likelihood estimator (CMLE) proposed by Lande (1996) removed the bias from the MLE. The CMLE method was also developed in the same way as the Jackknife estimator method proposed by Zahl (1977), which was a general method to reduce the bias of a biased estimator. However, these 3 estimators ignore missing species whereas a nonparametric estimation of the Shannon index of diversity was based on the sample coverage method and on unseen species (Chao and Shen 2003), missing individuals are considered on the basis of rare species.

According to the results, the calculation of Shannon’s diversity index from the traditional maximum likelihood estimator (MLE) gave the baseline quantities of diversity index in all combinations of studied years and seasons. The jackknife estimator gave the diversity highest quantities in all combinations. It can be said that the diversity index values of MLE, CMLE and Chao & Shen estimator showed more or less similar in quantities of the diversity.

By the way of Jackknife estimator, Zahl (1977) said that the advantage of the jackknife method may be simple random sampling of the individuals (plants, animals or whatever). The data collection of macrofungi using 100 square metres (m^2) quadrat can be considered as simple random sampling by geographical area. There are 5 forest types, of which only one permanent quadrat of 100 m^2. The process of calculation for jackknife estimator was not so practical in syntax writing for computer software programming.

5.3 The simulation study

To examine the performance of the 3 estimations of Shannon’s diversity index, the simulation results were based on the simulated data which distributed under random uniform model, which applied from probability sampling by using uniform distribution for random variables when it results from counting take on zero or a positive integer value. For each combination of number of species (S) and number of individual(n), 1000 simulated data set were generated and calculated for the Shannon diversity index(H) parameter along with the CV of parameter H .

Consideration the performance of 3 estimators by according to these following properties:

1) Bias

In every situation, regardless of number of species (S) or number of individuals $(\text{size} = n)$, all bias property showed negative sign to the parameter (H) from simulation study. The negative sign of bias showed underestimate of all estimators. Diversity index in every situation when comparing with the parameter (H), the CMLE and the Chao & Shen estimator showed smaller bias and were selected as suggested estimator.

At number of species $(S = 20)$ and size $(n = 50)$, the Chao & Shen estimator showed minimum negative bias and was selected as suggested estimator. At size ($n \geq 75$ up to 2000), the CMLE showed minimum bias property and was selected as suggested estimator.

At number of species (S) equals to 40, 60, 80 and 100, the most of Chao & Shen estimator gave the smallest bias at sample size 50, 75, 100 and 200. The most of CMLE gave the smallest bias at sample size was higher than and equal 500.

2) Variance

The MLE gave the smaller variance at every number of species (S = 20, 40, 60, 80 and 100) and number of individuals (n) when n was less than or equal 200. At larger number of individuals provides smaller variance for each estimator. The Chao & Shen estimator gave smaller variance at number of individuals was greater or equal to 500.

3) Mean squared error (MSE)

According to the fact that MSE equals variance plus square of its bias, so, the MSE was the most consideration as suggested estimator. At number of species $(S = 20)$, the CMLE gave the smallest MSE when number of individuals (n) was lower than and equal 100. The Chao & Shen estimator gave the smallest MSE at sample size was higher than and equal 200.

At number of species (S) equals to 40, 60, 80 and 100, the Chao & Shen estimator gave the smallest MSE at lower sample size when n was less than or equal 200. The CMLE gave the smallest MSE when n was greater than or equal 500.

5.4 The applications to the real data situations

The most variable macrofungi in Khek watershed of summation 5 forest in each season and each year was represented number of species has trend in between 20 to 80, whereas number of individuals has trend in between 500 to 2000. According to the simulation study, it can be indicated that there are two dominant estimators i.e. CMLE and Chao & Shen estimator are suitable with the dataset. The results from the application with real data was shown that the CMLE is the most suitable to estimate macrofungi diversity in rainy season among 3 years. The Chao & Shen estimator is suitable to estimate biodiversity in dry season and cool season for all calendar years. Number of species in rainy season is between 40 and 80, whereas number of individuals is approximately 1000. In dry season and cool season, the number of species equals to 20, whereas the number of individuals is 500, 1000 and 2000. Regardless number of individuals, it can be seen that number of species affects the variation of biodiversity in natural resources.

The dataset of dry season 2002, there were two suggested estimators; the CMLE and the Chao & Shen estimator which were applied to use in this dataset. Both of these 2 estimators yield similar diversity index for macrofungi data in Khek watershed. The 95% CI ranging of CMLE was narrower than Chao & Shen estimator. The variance of CMLE was calculated by using the variance’s formula from MLE. The variance of Chao & Shen estimator was difficult to calculate analytically. However, the variance of Chao & Shen estimator was applied by using the standard deviation of its estimate from simulation process.

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This study focused on comparing the Shannon’s diversity index estimators from MLE, CMLE and the Chao & Shen Estimator methods. The data are simulated under Random Uniform Model in order to study the properties.

The macrofungi in Khek watershed, Phetchabun Province

The greatest number of species was found in Secondary Dry Evergreen Forest (28 species 201 individuals) during rainy season 2003 and the smallest species was found in Pine Forest (1 species 2 individuals) during cool season 2003, whereas the greatest number of individuals was found in Dry Evergreen Forest 2 (9 species 1177 individuals) during dry season 2002.

The maximum Shannon’s diversity index was 2.78 (0.04) in Dry Evergreen Forest 1 on rainy season 2003, whereas the minimum Shannon’s diversity index was 0.19 (0.06) in Dry Evergreen Forest with Pine on cool season 2002. The rainy season has the biodiversity higher than dry season and cool season. In general pattern of forest types, the Pine Forest seemed to be lowest macrofungi diversity.

The simulated data under random uniform model

It was found that MLE gave the best estimator to estimate the Shannon’s diversity index with respect to its smallest variance. The Chao & Shen estimator and the CMLE provided the best estimator with its smallest bias and smallest MSE for many situations.

Applying the selected estimators to the real situations

CMLE and Chao & Shen estimator were found to be similar in magnitude and CMLE gave the more precise confidence interval. These were well illustrated by macrofungi data.

6.2 Recommendations

The suggestions for the results of the study:

Regardless of the number of species, the Chao & Shen estimator is recommended to estimate the Shannon’s diversity index when the number of individuals is less than 200, while the correction bias of maximum likelihood estimator (CMLE) is recommended to use when the number of individuals is larger than 500.

In simulation study at situation number of species (S) = 20 was found that the Chao & Shen estimator is recommended to use when number of individuals (n) is larger than 200. These can be suggested that maybe based on random uniform model if changing to truncated normal distribution maybe better.

The most bias values of all 5 situations in simulation study are underestimates although Chao & Shen estimator method include with unseen species in its formula. Therefore, it can be suggested the new estimator that not to be included only unseen species but also moreover issue in perused.

The need to monitor trends in biodiversity raises many technical issues underlying composite indices, identification of change-points and estimation of spatially varying time trends (Buckland et al., 2005). However, the limit biodiversity is lost unless biodiversity can be measured and its rate of change quantified. In this paper assume that biodiversity will be monitored by surveying changes in abundance of number of species and number of individuals. The modified diversity estimator (CMLE and Chao & Shen estimator) should be suggested to further use because they improved bias and represented to be more biodiversity than the traditional diversity estimator (MLE).

The proper management of an ecological population is greatly aided by solid information about its species abundance. Solid estimation of the entire probability model allows us develop generator based measure of ecological diversity. This being said that measuring diversity index need to surveillance the changing in biodiversity for biological community in each seasons and calendar years and must be suggested to following what something will has happened in the future.

For Jackknifing using various quadrats, then compute the index for the accumulated total of sample quadrats, adding quadrats one at a time until the index, on visual inspection stops increasing or decreasing except for random fluctuation is recommended.

The suggestions for further study:

This study can be apply to other method such as bootstrap for estimation the Shannon’s diversity index in the further study.

This study focused only on random uniform model. Preston (1948) first proposed the use of truncated log normal distribution for the observed species abundances (Norris and Pollock 1998). The truncated log normal distribution may be as well.

The Computer Assistance Instrument which can provide easy understanding of entering field data should be existed in order to calculate biodiversity and estimator along with the CIs of each estimator.

REFERENCES

Abercrombie M., Hickman M., Johnson M. L. and Thain M. 1990. *The new penguin dictionary of biology*. 8th ed. London, UK: Penguin Group.

Allaby M. 1994. *Macmillan dictionary of the environment*. 4th ed. London and Basingstoke, UK: The Macmillan Press Ltd.

Boxrud D., Pederson-Gulrud K., Wotton J., Medus C., Lyszkowicz E., Besser J. and Bartkus J. M. 2007. Comparison of Multiple-Locus Variable-Number Tandem Repeat Analysis, Pulsed-Field Gel Electrophoresis, and Phage Typing for Subtype Analysis of *Salmonella enterica* Serotype Enteritidis. *J. Clin. Microbiol.*, 45(2), 536-543.

Buckland S. T., Maggurran A. E., Green R. E. and Fewster R. M. 2005. Monitoring change in biodiversity through composite indices. *Phil. Trans. R. Soc. B,* 360, 243-254.

Chao, A. and Lee, S-M. 1992. Estimating the number of classes via sample coverage. *Journal of American Statistical Association*, 87, 210-217.

Chao, A. and Shen, T-J. 2003. Nonparametric Estimation of Shannon's index of diversity when there are unseen species. *Environmental and Ecological Statistics*, 10, 429-443.

Chao, A. Ma, M-C. and Yang, M.C.K. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika*, 80, 193-201.

Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. Roy. Soc. London Ser.*, B 345, 101-118.

Duengkae, K. 2006. Monitoring on species diversity of macrofungi in Khek watershed, Phetchabun Province. *Songklanakarin Journal of Science and Technology*, 28(2), 293-303.

Giachini A. J., Souza L. A. B. and Oliveira V. L. 2004. Species richness and seasonal abundance of ectomycorrhizal fungi in plantations of *Eucalyptus dunnii* and *Pinus taeda* in southern Brazil. *Mycorrhiza*, 14, 375–381.

Glaubrecht, M. and Köhler, F. 2004. Radiating in a river: systematics, molecular genetics and morphological differentiation of viviparous freshwater gastropods endemic to the Kaek River, central Thailand (Cerithioidea, Pachychilidae). *Biological Journal of the Linnean Society*, 82, 275-311.

Hanlin, R. T. 1990. *Illustrated Genera of Ascomycetes*. St. Paul, Minesota, USA: The American Phytopathology Society (APS Press).

Hawksworth, D. L., Kirk, P. M., Sutton, B. C. and Pegler, D. N.. 1995. *Ainsworth & Bisby's Dictionary of the Fungi*. 8th ed. Wallingford, UK: CAB International.

Horvitz, D. G. and Thompson, D. J.. 1952. A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, 47(260), 663-685.

Jordan, P. 1999. *Illustrated Encyclopedia: Mushroom identifier*. London, UK: Lorenz Books.

Kindt, R. 2002. *Methodology for tree species diversification planning for African agroecosystems*. Thesis submitted in fulfillment of the requirement of the degree of doctor (PhD) in applied biological sciences. Faculty of agricultural and applied biological sciences, Ghent University, Gent (Belgium), 332+xi pp.

Konemann, V. 1999. *The Great Encyclopedia of Mushrooms*. Germany.

Krebs, C.J. 1999. Ecological Methedology. Menlo Park, California, USA: Benjamin/Cummings Press.

Laessoe, T. 1998. *Eyewitness Handbooks: Mushrooms*. London, UK: Dorling Kindersley Limited.

Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos*, 76, 5-13.

Largent, D. 1973. *How to Identify Mushroom (to genus) Using Only Macroscopic Features*. Mad River Press.

Lodge D. J., Ammirati J. F., O'Dell T. E. and Mueller G. M. 2004. *Biodiversity of Fungi: Collecting and Describing Macrofungi*. London, UK: Elsevier Academic Press.

Magurran, A. E. (1988). *Ecological Diversity and its Measurement*. Princeton, New Jersey, USA: Princeton University Press.

Moser, M. 1973. *Keys to Agarics and Boleti*. Germany: Whitefriars Press, Ltd.

Mouillot, D. and Lepretre, A. 1999. A comparison of species diversity estimators. *Researches in Population Ecology*, 41, 203-215.

National Park Wildlife and Plant Conservation Department. c2005, n.d. Thung Salaeng Luang National Park. Retrieved 12 November 2007, from URL: http://www.dnp.go.th/nprd/project/tungsalangluang.php

Norris, J.L. and Pollock, K.H. 1998. Non-parametric MLE for Poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics*, 5, 391-402.

Pegler, D. and Spooner, B. 1994. *Mushrooms*. London, UK: The Apple Press.

Pegler, D. N. 1986. Agaric Flora of Sri Lanka. Kew Bull. London, UK: Add. Ser. XII, HMSO, Royal Botanic Gardens, Kew.

Preston, F.W. 1948. The commonness and rarity of species. *Ecology*, 29, 254-283.

ReliaSoft Corporation. c1998-2007, 24 January 2008. MLE (Maximum Likelihood) Parameter Estimation. Retrieved 18 April 2008, from URL: http://www.weibull.com

Richard F., Moreau P. A., Selosse M. A. and Gardes M. 2004. Diversity and fruiting patterns of ectomycorrhizal and saprobic fungi in an old-growth Mediterraneanforest dominated by *Quercus ilex* L. *Canadian Journal of Botany*, 82, 1711–1729.

Ruksawong, P. and Fegel, T. W. 2001. Thai Mushrooms and Other Fungi. Bangkok, Thailand: National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency. [In Thai].

Saha, K. K. and Paul, S. R.. 2005. Bias-corrected maximum likelihood estimator of the intraclass correlation parameter for binary data. *Statistics in Medicine*, 24, 3497–3512.

Shannon, C. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423.

Sherwin W. B., Jabot F., Rush R. and Rossetto M. 2006. Measurement of biological information with applications from genes to landscapes. *Molecular Ecology*, 15, 2857-2869.

Simpson, E. H. 1949. Measurement of diversity. *Nature*, 163, 688.

Straatsma G., Ayer F. and Edli S. 2001. Species richness, abundance, and phenology of fungal fruit bodies over 21 years in a Swiss forest plot. *Mycological Research*, 105(5), 515-523.

The Office of Natural Resources and Environmental Policy and Planning (ONEP). 2004. *Thailand's Biodiversity*. 3rd ed. Bangkok, Thailand.

The offices of the Tourism Authority of Thailand (TAT). c2007, 27 December 2007. Loei and The Middle Mekong. *Budget WorldClass Drives*. Retrieved 20 January 2008, from
URL: http://www.budget.co.th

The Royal Institute. 1996. *Edible mushrooms and toxic mushrooms in Thailand*. 1st ed. Bangkok, Thailand: The Royal Institute Press. [In Thai].

Thieraut, R.J. and Klekamp, R.C. 1975. *Decision making through operations research*. 2nd ed. Singapore: John Wiley &Son.

Vogt, W.P. 1993. *Dictionary of Statistics and Methodology: A non-technical guide for the social sciences*. California, USA: Sage.

Zahl, S. 1977. Jackknifing An Index of Diversity. *Ecology*, 58(4), 907-913.

Zhang Y., Guo L-D. and Liu R-J. 2004. Survey of arbuscular mycorrhizal fungi in deforested and natural forest land in the subtropical region of Dujiangyan, southwest China. *Plant and Soil*, 261, 257–263.

APPENDIX

APPENDIX A
CHECK LIST

Table A1. Checklist of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province during 2002-2004.

No.	Scientific name	Family		Year		Season[1]	Forest Type [2]
			2002	2003	2004		
	Phylum Ascomycota						
1	*Biosporella rinces* (Batsch.) Korf et al.	Leotiaceae	/	/		2	4
2	*Boedijnopeziza institia*	Sarcoscyphaceae		/		2	2
3	*Cookeina sulcipes* (Berk.) Kuntze	Sarcoscyphaceae	/	/	/	2	2
4	*Cookeina tricholoma* (Mont.) Kuntze	Sarcoscyphaceae	/	/	/	2	1, 2, 3, 4
5	*Cordyceps sphecocephala* (Kl.) Sacc.	Clavicipitaceae		/		2	4
6	*Entonaema liquescens* Moell.	Xylariaceae	/	/		2	2, 3
7	*Entonaema splendens* (Berk. Et Curt.) Lloyd	Xylariaceae	/			3	1
8	*Phillipsia domingensis* Berk.	Sarcoscyphaceae	/		/	2	1, 2
9	*Sarcoscypha occidentalis* (Schw.) Sacc.	Sarcoscyphaceae	/			2	4
10	*Xylaria fockeri* Miq.	Xylariaceae		/	/	2, 3	3, 4
11	*Xylaria grammica* (Mont.) Fr.	Xylariaceae			/	2	3
12	*Xylaria polymorpha* (Pers.) Grev.	Xylariaceae		/	/	2	2, 3
	Phylum Basidiomycota						
1	*Amanita* sp.1	Amanitaceae		/		2, 3	5
2	*Amanita* sp.2	Amanitaceae		/		2	2
3	*Amanita* sp.3	Amanitaceae			/	2	1
4	*Amauroderma rugosum* (Blume et Nees ex Fr.) Torr.	Ganodermataceae	/	/		2	1, 2, 3
5	*Auricularia rinces* (Hook.) Underw.	Auriculariaceae		/	/	1, 2	2, 3, 4
6	*Auricularia rinces* (Fr.) P. Henn.	Auriculariaceae	/	/	/	2,3	1, 2
7	*Auricularia polytricha* (Mont.) Sacc.	Auriculariaceae		/		3	4

Table A1. Checklist of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province during 2002-2004 (Continued).

No.	Scientific name	Family		Year		Season[1]	Forest Type [2]
			2002	2003	2004		
8	*Auricularia tenuis* (Lev.) Farlow	Auriculariaceae		/		3	2
9	*Auriscalpium vulgare* S.F. Gray	Auriscalpiaceae	/	/		2	4, 5
10	*Boletellus emodensis* (Berk.) Singer	Strobilomycetaceae		/		2	4
11	*Boletus chromapes* Frost.	Boletaceae		/	/	2	2, 4, 5
12	*Boletus chrysenteron* Bull.	Boletaceae		/		2	1
13	*Campanella junghuhnii* (Mont.) Sing.	Tricholomataceae		/	/	1, 2	1, 2, 3
14	*Cantharellus odoratus* (Schw.) Fr.	Cantharellaceae			/	2	3
15	*Cantharellus subcibarius* Corner	Cantharellaceae	/			2	4
16	*Cladopus repens* Petch	Entolomataceae			/	3	2
17	*Clavaria vermicularis* Swartz : Fr.	Clavariaceae		/		2	3
18	*Clitopilus apalus* (Berk. & Br.) Petch	Entolomataceae		/	/	2	3, 4
19	*Coltricia perennis* (L. ex Fr.) Murr.	Polyporaceae	/			2	5
20	*Coltricia cinnamomea* (Pers.) Murr.	Polyporaceae		/	/	2	1, 2, 5
21	*Coprinus disseminatus* (Pers. : Fr.) S.F. Gray	Coprinaceae		/		2	1, 3
22	*Coprinus* sp.	Coprinaceae	/			2	3
23	*Cortinarius* sp.	Cortinariaceae		/	/	2	5
24	*Crepidotus citrinus* Petch	Crepidotaceae		/		2	4
25	*Crepidotus variabilis* (Pers. Ex Fr.) Kummer	Crepidotaceae	/			2	3
26	*Crinipellis* sp.	Tricholomataceae			/	2	4
27	*Cryptoporus volvatus (*Peck) Shear	Polyporaceae	/			2	5
28	*Cyclomyces fuscus* Fr.	Hymenochaetaceae		/	/	3	3, 4
29	*Cymatoderma dendriticum* (Pers.) Reid	Podoschyphaceae		/		2	2, 4
30	*Deflexula fascicularis* (Bres. Et Pat.) Corner	Clavariaceae		/		2	3
31	*Dicephalospora rufocornea* (Berk. & Br.) Spooner	Sclerotiniaceae		/		2	3
32	*Favolaschia nipponica* Kobayasi	Tricholomataceae	/			2	1
33	*Favolaschia tonkinensis* (Pat.) Singer	Tricholomataceae		/		2	3

Table A1. Checklist of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province during 2002-2004 (Continued).

No.	**Scientific name**	**Family**		**Year**		**Season**[1]	**Forest Type** [2]
			2002	**2003**	**2004**		
34	*Filoboletus manipularis* (Berk.) Sing.	Tricholomataceae		/		2	3
35	*Ganoderma lucidum* (Leyss. Ex Fr.) Karst.	Ganodermataceae	/			1	3
36	*Ganoderma* sp.	Ganodermataceae	/	/		2, 3	1, 2, 4
37	*Geastrum saccatum* (Fr.) Fisch.	Geastraceae	/			2	1
38	*Guepinia spatularia* (Schw.) Fr.	Dacrymycetaceae	/			2	5
39	*Hexagonia tenuis* (Hook.) Fr.	Coriolaceae	/	/	/	1,2,3	1, 2, 3, 4
40	*Hygrocybe cantharellus* (Schw.) Fr.	Hygrophoraceae	/			2	5
41	*Hygrocybe firma* (Berk. Et Broome) Sing.	Hygrophoraceae		/		2	5
42	*Hymenochaete rubiginosa* (Dicks. Ex Fr.) Lev.	Hymenochaetaceae	/	/	/	1,2,3	1, 2, 3, 4
43	*Inocybe asterospora* Quel.	Cortinariaceae		/		2	5
44	*Inocybe calospora* Quel.	Cortinariaceae	/		/	2	3, 4
45	*Inocybe nodulospora* Kobayasi	Cortinariaceae	/			2	5
46	*Inonotus dryadeus* (Pers. Ex Fr.) Murr.	Hymenochaetaceae	/			1	2
47	*Inonotus tomentosus*	Hymenochaetaceae		/		2	1
48	*Laccaria vinaceoavellanea* Hongo	Tricholomataceae	/	/	/	2	5
49	*Laccaria* sp.	Tricholomataceae			/	2	1
50	*Lactarius subpiperatus* Hongo	Russulaceae	/			2	4
51	*Laetiporus sulphureus* (Fr.) Murr.	Polyporaceae	/			2	2
52	*Lentinus polychrous* Lev.	Pleurotaceae		/		2	3
53	*Lentinus* sp.	Pleurotaceae		/	/	2	3
54	*Marasmiellus candidus* (Bolt.) Fr.	Tricholomataceae	/	/	/	2	1, 2, 3
55	*Marasmiellus ramealis*	Tricholomataceae	/	/		2	1
56	*Marasmius conicopapillatus* Henn.	Tricholomataceae		/		2	1
57	*Marasmius crinisequi* F. Mull. Ex Kalchbr.	Tricholomataceae		/		2	1
58	*Marasmius florideus* Berk. & Br.	Tricholomataceae		/		2	1, 2
59	*Marasmius graminum* (Lib.) Berk.	Tricholomataceae		/		2	1
60	*Marasmius maximus* Hongo	Tricholomataceae	/		/	2	3, 4

Table A1. Checklist of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province during 2002-2004 (Continued).

No.	Scientific name	Family	Year 2002	 2003	 2004	Season[1]	Forest Type [2]
61	*Marasmius papyraceus* Massee	Tricholomataceae	/	/	/	2	1, 2, 4
62	*Marasmius pulcherripes* Peck	Tricholomataceae	/	/	/	2	2, 3, 4
63	*Marasmius* sp.1	Tricholomataceae		/	/	2	1, 2
64	*Marasmius* sp.2	Tricholomataceae		/		2	1
65	*Micromphale* sp.	Tricholomataceae			/	2	1, 2
66	*Microporus affinis* (Blume & Nees ex Fr.) Kntz.	Polyporaceae	/	/	/	1,2,3	1, 2, 3, 4
67	*Microporus xanthopus* (Fr.) Kuntze	Polyporaceae	/	/	/	1,2,3	1, 2, 3, 4
68	*Mycena* sp.1	Tricholomataceae		/		2	3
69	*Mycena* sp.2	Tricholomataceae		/	/	2	1, 2
70	*Mycena* sp.3	Tricholomataceae			/	2	2
71	*Oudemansiella radicata* (Rehl. Ex Fr.) Sing.	Tricholomataceae	/			2	1, 2, 3
72	*Panus* sp.	Pleurotaceae	/	/	/	1,3	1, 3
73	*Phellinus gilvus* (Schw. Ex Fr.) Pat.	Hymenochaetaceae	/	/		1, 3	1, 3, 4
74	*Phellinus* sp.	Hymenochaetaceae		/		2	2
75	*Phylloporus bellus* (Mass.) Corner	Boletaceae	/		/	2	5
76	*Pleurocybella porrigens* (Pers. Ex Fr.) Sing.	Tricholomataceae	/		/	2	1, 2, 4
77	*Podoscypha surinamensis.*	Podoschyphaceae	/	/		2	1, 2, 3
78	*Polyporus grammocephalus* Berk.	Polyporaceae	/	/	/	2	1, 2, 4
79	*Polyporus picipes* Fr.	Polyporaceae		/		2	3
80	*Polyporus tenuiculus* (Beauv.) Fr.	Polyporaceae	/	/	/	2	1, 2
81	*Psathyrella candolliana* (Fr.:Fr.) Maire	Coprinaceae		/	/	2, 3	3, 5
82	*Pterula multifida* Fr.	Clavariaceae		/		2	1
83	*Pulveroboletus ravenelii* (Berk. Et Curt.) Murr.	Boletaceae	/			2	5
84	*Pycnoporus coccineus* (Fr.) Bond et Sing.	Coriolaceae	/	/	/	1, 3	1, 2, 3, 4
85	*Ramaria* sp.	Ramariaceae			/	2	2
86	*Rhizopogon* sp.	Rhizopogonaceae		/		2	3
87	*Russula alboareolata* Hongo	Russulaceae			/	2	2
88	*Russula delica* Fr.	Russulaceae		/	/	2	1, 2, 5
89	*Russula emetica* (Schaeff. Ex Fr.) Pers. Ex S.F. Gray	Russulaceae		/		2, 3	1, 4
90	*Schizophyllum commune* Fr.	Polyporaceae		/	/	1, 3	2

Table A1. Checklist of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province during 2002-2004 (Continued).

No.	**Scientific name**	**Family**	**Year** **2002**	**Year** **2003**	**Year** **2004**	**Season**[1]	**Forest Type** [2]
91	*Scleroderma* sp.	Sclerodermataceae			/	2	3
92	*Scytinopogon echinosporus* Berk. Et Broome	Clavariaceae		/		2	2
93	*Stereopsis burtianum* (Peck) Ried	Podoschyphaceae		/	/	1, 2	1
94	*Stereum fasciatum* (Schw.) Fr.	Stereaceae	/	/	/	1, 2, 3	1, 2, 4
95	*Stereum frustulatum* (Pers., Fr.) Fckl.	Stereaceae	/	/	/	1, 2, 3	1, 2
96	*Stereum ostrea* (Bl. Et Nees) Fr.	Stereaceae	/	/	/	1,2,3	1, 2, 3, 4
97	*Termitomyces striatus* (Beeli) Heim	Amanitaceae	/	/		2	1, 2, 3
98	*Thelephora* sp.	Thelephoraceae		/		2	3
99	*Trametes cingulata* Berk.	Polyporaceae	/		/	1, 3	2, 4
100	*Tremella cinnabarina* (Mont.) Lloyd	Tremellaceae		/		2	2
101	*Tremella fuciformis* Berk.	Tremellaceae	/			3	1
102	*Trichaptum elongatum* (Berk.) Imaz.	Polyporaceae	/	/	/	1, 2, 3	2, 3, 5
103	*Trichaptum fuscoviolaceum* (Fr.) Ryv.	Polyporaceae	/	/		2	3, 5
104	*Trogia infundibuliformis* Berk. & Br.	Tricholomataceae	/	/	/	2	1, 2, 3
105	*Tylopilus albo-ater* (Schw.) Murr.	Boletaceae	/			2	5
106	*Xylobolus rinces* (Jungh.) Boiden	Stereaceae	/		/	1, 3	1
107	*Xylobolus spectabilis* (Klotz.) Boiden	Stereaceae	/	/		2	2

Note [1] 1 = Dry season (Feb – May)
2 = Rainy season (Jun – Oct)
3 = Cool Season (Nov – Jan)

[2] 1 = Dry Evergreen Forest 1 (In Khao Kho Wildlife Captive Breeding Station)
2 = Secondary Dry Evergreen Forest
3 = Dry Evergreen Forest 2 (In Thung Salang Luang National Park)
4 = Dry Evergreen Forest with Pine
5 = Pine Forest

APPENDIX B

SHANNON INDEX AND S.D. CALCULATION

Table B1. Shannon’s diversity index (1948) of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province on three seasons during 2002-2004.

Forest	Season	Year	No of species	Total no of individuals	Shannon index	S.D.
Dry Evergreen Forest 1	Dry	2002	6	711	1.08	0.04
Dry Evergreen Forest 1	Rain	2002	14	269	1.71	0.07
Dry Evergreen Forest 1	Cool	2002	7	264	0.81	0.06
Dry Evergreen Forest 1	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest 1	Rain	2003	27	497	2.78	0.04
Dry Evergreen Forest 1	Cool	2003	6	130	1.42	0.06
Dry Evergreen Forest 1	Dry	2004	4	87	1.15	0.07
Dry Evergreen Forest 1	Rain	2004	10	166	1.77	0.06
Dry Evergreen Forest 1	Cool	2004	5	214	1.07	0.05
Secondary Dry Evergreen Forest	Dry	2002	6	712	1.00	0.04
Secondary Dry Evergreen Forest	Rain	2002	14	158	2.11	0.07
Secondary Dry Evergreen Forest	Cool	2002	3	238	0.46	0.06
Secondary Dry Evergreen Forest	Dry	2003	Missing	Missing	-	-
Secondary Dry Evergreen Forest	Rain	2003	28	201	2.62	0.06
Secondary Dry Evergreen Forest	Cool	2003	6	237	1.66	0.03
Secondary Dry Evergreen Forest	Dry	2004	6	143	1.44	0.06
Secondary Dry Evergreen Forest	Rain	2004	20	235	2.43	0.06
Secondary Dry Evergreen Forest	Cool	2004	7	174	1.61	0.05
Dry Evergreen Forest 2	Dry	2002	9	1177	1.60	0.03
Dry Evergreen Forest 2	Rain	2002	9	223	1.80	0.05
Dry Evergreen Forest 2	Cool	2002	5	177	1.50	0.03
Dry Evergreen Forest 2	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest 2	Rain	2003	21	306	2.55	0.05
Dry Evergreen Forest 2	Cool	2003	5	142	1.34	0.05
Dry Evergreen Forest 2	Dry	2004	5	156	1.33	0.04
Dry Evergreen Forest 2	Rain	2004	12	215	1.50	0.09
Dry Evergreen Forest 2	Cool	2004	5	128	1.50	0.04
Dry Evergreen Forest with Pine	Dry	2002	6	157	1.25	0.08
Dry Evergreen Forest with Pine	Rain	2002	8	83	1.73	0.09
Dry Evergreen Forest with Pine	Cool	2002	2	105	0.19	0.06
Dry Evergreen Forest with Pine	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest with Pine	Rain	2003	13	186	2.20	0.06
Dry Evergreen Forest with Pine	Cool	2003	10	235	1.83	0.05
Dry Evergreen Forest with Pine	Dry	2004	3	29	0.60	0.14

Table B1. Shannon's diversity index (1948) of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province on three seasons during 2002-2004 (Continued).

Forest	Season	Year	No of species	Total no of individuals	Shannon index	S.D.
Dry Evergreen Forest with Pine	Rain	2004	10	196	1.60	0.07
Dry Evergreen Forest with Pine	Cool	2004	6	103	1.47	0.07
Pine Forest	Dry	2002	0	0	NA	NA
Pine Forest	Rain	2002	11	293	1.54	0.06
Pine Forest	Cool	2002	0	0	NA	NA
Pine Forest	Dry	2003	Missing	Missing	-	-
Pine Forest	Rain	2003	9	139	1.59	0.08
Pine Forest	Cool	2003	1	2	0.00	0.00
Pine Forest	Dry	2004	0	0	NA	NA
Pine Forest	Rain	2004	8	77	1.30	0.11
Pine Forest	Cool	2004	0	0	NA	NA

Note: The data on dry season in year 2003 were censored.

APPENDIX C
SIMPSON INDEX AND S.D. CALCULATION

Table C1. Simpson's diversity index and standard deviation of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province.

Year	**Season**	**Simpson's diversity index (S.D.)**				
		Dry Evergreen Forest 1	**Secondary Dry Evergreen Forest**	**Dry Evergreen Forest 2**	**Dry Evergreen Forest with Pine**	**Pine Forest**
2002	Dry	0.52 (0.02)	0.48 (0.02)	0.73 (0.01)	0.60 (0.04)	NA
	Rain	0.71 (0.02)	0.85 (0.01)	0.80 (0.01)	0.76 (0.04)	0.70 (0.02)
	Cool	0.42 (0.03)	0.23 (0.03)	0.76 (0.01)	0.09 (0.04)	NA
2003	Dry	-	-	-	-	-
	Rain	0.91 (0.01)	0.89 (0.01)	0.90 (0.01)	0.85 (0.02)	0.72 (0.03)
	Cool	0.72 (0.02)	0.80 (0.01)	0.71 (0.02)	0.80 (0.01)	0.00 (0.00)
2004	Dry	0.63 (0.04)	0.71 (0.02)	0.71 (0.02)	0.34 (0.10)	NA
	Rain	0.79 (0.01)	0.88 (0.01)	0.61 (0.04)	0.73 (0.02)	0.45 (0.04)
	Cool	0.59 (0.02)	0.76 (0.02)	0.76 (0.01)	0.73 (0.02)	NA

Note: The data on dry season in year 2003 were censored.
NA = No data available (Not enough data for calculation).

Table C2. Simpson diversity index (1949) of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province on three seasons during 2002-2004.

Forest	**Season**	**Year**	**No of species**	**Total no of individuals**	**Simpson index**	**S.D.**
Dry Evergreen Forest 1	Dry	2002	6	711	0.52	0.02
Dry Evergreen Forest 1	Rain	2002	14	269	0.71	0.02
Dry Evergreen Forest 1	Cool	2002	7	264	0.42	0.03
Dry Evergreen Forest 1	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest 1	Rain	2003	27	497	0.91	0.01
Dry Evergreen Forest 1	Cool	2003	6	130	0.72	0.02
Dry Evergreen Forest 1	Dry	2004	4	87	0.63	0.04
Dry Evergreen Forest 1	Rain	2004	10	166	0.79	0.01
Dry Evergreen Forest 1	Cool	2004	5	214	0.59	0.02
Secondary Dry Evergreen Forest	Dry	2002	6	712	0.48	0.02
Secondary Dry Evergreen Forest	Rain	2002	14	158	0.85	0.01

Table C2. Simpson diversity index (1949) of Macrofungi in Khek Watershed, Khaokho District, Phetchabun Province on three seasons during 2002-2004 (Continued).

Forest	Season	Year	No of species	Total no of individuals	Simpson index	S.D.
Secondary Dry Evergreen Forest	Cool	2002	3	238	0.23	0.03
Secondary Dry Evergreen Forest	Dry	2003	Missing	Missing	-	-
Secondary Dry Evergreen Forest	Rain	2003	28	201	0.89	0.01
Secondary Dry Evergreen Forest	Cool	2003	6	237	0.80	0.01
Secondary Dry Evergreen Forest	Dry	2004	6	143	0.71	0.02
Secondary Dry Evergreen Forest	Rain	2004	20	235	0.88	0.01
Secondary Dry Evergreen Forest	Cool	2004	7	174	0.76	0.02
Dry Evergreen Forest 2	Dry	2002	9	1177	0.73	0.01
Dry Evergreen Forest 2	Rain	2002	9	223	0.80	0.01
Dry Evergreen Forest 2	Cool	2002	5	177	0.76	0.01
Dry Evergreen Forest 2	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest 2	Rain	2003	21	306	0.90	0.01
Dry Evergreen Forest 2	Cool	2003	5	142	0.71	0.02
Dry Evergreen Forest 2	Dry	2004	5	156	0.71	0.02
Dry Evergreen Forest 2	Rain	2004	12	215	0.61	0.04
Dry Evergreen Forest 2	Cool	2004	5	128	0.76	0.01
Dry Evergreen Forest with Pine	Dry	2002	6	157	0.60	0.04
Dry Evergreen Forest with Pine	Rain	2002	8	83	0.76	0.04
Dry Evergreen Forest with Pine	Cool	2002	2	105	0.09	0.04
Dry Evergreen Forest with Pine	Dry	2003	Missing	Missing	-	-
Dry Evergreen Forest with Pine	Rain	2003	13	186	0.85	0.02
Dry Evergreen Forest with Pine	Cool	2003	10	235	0.80	0.01
Dry Evergreen Forest with Pine	Dry	2004	3	29	0.34	0.10
Dry Evergreen Forest with Pine	Rain	2004	10	196	0.73	0.02
Dry Evergreen Forest with Pine	Cool	2004	6	103	0.73	0.02
Pine Forest	Dry	2002	0	0	NA	NA
Pine Forest	Rain	2002	11	293	0.70	0.02
Pine Forest	Cool	2002	0	0	NA	NA
Pine Forest	Dry	2003	Missing	Missing	-	-
Pine Forest	Rain	2003	9	139	0.72	0.03
Pine Forest	Cool	2003	1	2	0.00	0.00
Pine Forest	Dry	2004	0	0	NA	NA
Pine Forest	Rain	2004	8	77	0.45	0.04
Pine Forest	Cool	2004	0	0	NA	NA

Note: The data on dry season in year 2003 were censored.

APPENDIX D

SIMULATION SYNTAX

A SOFTWARE PROGRAM I

1. Main Program

PART 1: FOR PARAMETERS

Fixed species (s = 20, 40, 60, 80 and 100)

```
NEW FILE.                                                        (1)
SET SEED = 9999.
INPUT PROGRAM.
+ VECTOR a(1000,F8.7).
+ LOOP #U=1 TO s.
+  COMPUTE id=1.
+  COMPUTE tag=$CASENUM.
+  LOOP #V=1 TO 1000.
+   COMPUTE a(#V)=UNIFORM(1).
+  END LOOP.
+  END CASE.
+ END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

AGGREGATE OUTFILE=* MODE=ADDVARIABLES                              (2)
 /PRESORTED
```

```
 /BREAK=id
 /b1 TO b1000=SUM(a1 TO a1000).
EXECUTE.
```

(3)

```
VECTOR c(1000,F8.7).
VECTOR d(1000,F8.7).
VECTOR a = a1 TO a1000.
VECTOR b = b1TO b1000.
LOOP #U=1 TO 1000.
+ COMPUTE c(#U) = (a(#U))*(1/(b(#U))).
+ COMPUTE d(#U) = (c(#U))*LN(c(#U)).
END LOOP.
EXECUTE.
```

(4)

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES
 /PRESORTED
 /BREAK=id
 /e1 TO e1000=MEAN(c1 TO c1000)
 /f1 TO f1000=SUM(d1 TO d1000).
EXECUTE.
```

(5)

```
VECTOR f(1000).
VECTOR c = c1 TO c1000.
VECTOR e = e1 TO e1000.
LOOP #U=1 TO 1000.
+ COMPUTE g(#U)=(c(#U)-e(#U))**2.
END LOOP.
EXECUTE.
```

(6)

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES
 /PRESORTED
 /BREAK=id
 /h1 TO h1000=SUM(g1 TO g1000).
```

```
EXECUTE.

VECTOR i(1000).                                                          (7)
VECTOR e = e1 TO e1000.
VECTOR h = h1 TO h1000.
LOOP #U=1 TO 1000.
+ COMPUTE i(#U) = (SQRT(h(#U)/s))/e(#U).
END LOOP.
EXECUTE.

SAVE OUTFILE='c:\spsstemp\SIMU_PARA_FULL.sav'.

SAVE OUTFILE='c:\spsstemp\SIMU_PARA_Pi.sav'                               (8)
 /KEEP c1 TO c1000
 /RENAME (c1 TO c1000=pi1 TO pi1000).


GET FILE='c:\spsstemp\SIMU_PARA_FULL.sav'.                                (9)
FLIP
  VARIABLES=f1 TO f1000.
COMPUTE v_h=ABS(var001).
VARIABLE LABELS v_h 'H'.
EXECUTE.
SAVE OUTFILE='c:\spsstemp\temp1.sav'
  /KEEP v_h.


GET FILE='c:\spsstemp\SIMU_PARA_FULL.sav'.                               (10)
FLIP
  VARIABLES=i1 TO i1000.
COMPUTE v_cv=var001.
VARIABLE LABELS v_cv 'CV'.
EXECUTE.
SAVE OUTFILE='c:\spsstemp\temp2.sav'
  /KEEP v_cv.


MATCH FILES FILE='c:\spsstemp\temp1.sav'                                 (11)
  /FILE='c:\spsstemp\temp2.sav'.
```

```
FORMATS v_h v_cv (F8.7).
EXECUTE.

SAVE OUTFILE='c:\spsstemp\SIMU_PARA_SUMMARY.sav'.

SAVE TRANSLATE OUTFILE='c:\spsstemp\SIMU_PARA_SUMMARY.xls'    (12)
 /FIELDNAMES
 /REPLACE.
GET FILE='c:\spsstemp\SIMU_PARA_SUMMARY.sav'.
```

PART 2: FOR STATISTICS

Random binomial (n = 50, 75, 100, 200, 500, 1000 and 2000)

```
NEW FILE.                                                      (13)
SET SEED = 9999.
INPUT PROGRAM.
+ LOOP #U=1 TO s.
+  COMPUTE id=1.
+  END CASE.
+ END LOOP.
END FILE.
END INPUT PROGRAM.
EXECUTE.

MATCH FILES FILE=*                                             (14)
  /FILE='c:\spsstemp\SIMU_PARA_Pi.sav'.
EXECUTE.

VECTOR a(1000,F4.2).                                           (15)
VECTOR pi = pi1 TO pi1000.
LOOP #U=1 TO 1000.
+ COMPUTE a(#U)=RV.BINOM(n,pi(#U)).
END LOOP.
EXECUTE.
```

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES                    (16)
 /PRESORTED
 /BREAK=id
 /b1 TO b1000=SUM(a1 TO a1000).
EXECUTE.

VECTOR c(1000).                                          (17)
VECTOR a = a1 TO a1000.
VECTOR b = b1 TO b1000.
LOOP #U=1 TO 1000.
+ COMPUTE c(#U) = (a(#U)/b(#U))*LN(a(#U)/b(#U)).
END LOOP.
EXECUTE.

AGGREGATE OUTFILE=* MODE=ADDVARIABLES                    (18)
 /PRESORTED
 /BREAK=id
 /d1 TO d1000=SUM(c1 TO c1000).
EXECUTE.

VECTOR e(1000).                                          (19)
VECTOR d = d1 TO d1000.
LOOP #U=1 TO 1000.
+ COMPUTE e(#U) = ABS(d(#U)).
END LOOP.
EXECUTE.

VECTOR f(1000).                                          (20)
VECTOR a = a1 TO a1000.
LOOP #U=1 TO 1000.
+ DO IF a(#U)>0.
+ COMPUTE f(#U)=1.
```

```
+ ELSE.
+ COMPUTE f(#U)=0.
+ END IF.
END LOOP.
EXECUTE.
```

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES                              (21)
 /PRESORTED
 /BREAK=id
 /g1 TO g1000=SUM(f1 TO f1000).
EXECUTE.
```

```
VECTOR h(1000).                                                    (22)
VECTOR b = b1 TO b1000.
VECTOR e = e1 TO e1000.
VECTOR g = g1 TO g1000.
LOOP #U=1 TO 1000.
+ COMPUTE h(#U) = e(#U)+((g(#U)-1)/(2*b(#U))).
END LOOP.
EXECUTE.
```

```
VECTOR i(1000).                                                    (23)
VECTOR a = a1 TO a1000.
LOOP #U=1 TO 1000.
+ DO IF a(#U)=1.
+ COMPUTE i(#U)=1.
+ ELSE.
+ COMPUTE i(#U)=0.
+ END IF.
END LOOP.
EXECUTE.
```

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES                              (24)
```

```
 /PRESORTED
 /BREAK=id
 /j1 TO j1000=SUM(i1 TO i1000).
EXECUTE.
```

(25)

```
VECTOR k(1000).
VECTOR b = b1 TO b1000.
VECTOR j = j1 TO j1000.
LOOP #U=1 TO 1000.
+ COMPUTE k(#U)=1-(j(#U)/b(#U)).
END LOOP.
EXECUTE.
```

(26)

```
VECTOR l(1000).
VECTOR a = a1 TO a1000.
VECTOR b = b1 TO b1000.
VECTOR k = k1 TO k1000.
LOOP #U=1 TO 1000.
+ COMPUTE l(#U)=((k(#U)*a(#U)/b(#U))*LN(k(#U)*a(#U)/b(#U)))/(1-((1-
(k(#U)*a(#U)/b(#U)))**b(#U))).
END LOOP.
EXECUTE.
```

(27)

```
AGGREGATE OUTFILE=* MODE=ADDVARIABLES
 /PRESORTED
 /BREAK=id
 /m1 TO m1000=SUM(l1 TO l1000).
EXECUTE.
```

(28)

```
VECTOR n(1000).
VECTOR m = m1 TO m1000.
LOOP #U=1 TO 1000.
+ COMPUTE n(#U)=ABS(m(#U)).
```

```
END LOOP.
EXECUTE.

SAVE OUTFILE='c:\spsstemp\SIMU_STAT_FULL.sav'.
```

(29)

```
GET FILE='c:\spsstemp\SIMU_STAT_FULL.sav'.
FLIP
  VARIABLES=e1 TO e1000.
COMPUTE v_hmle=var001.
VARIABLE LABELS v_hmle 'Hmle'.
EXECUTE.
SAVE OUTFILE='c:\spsstemp\temp1.sav'
  /KEEP v_hmle.
```

(30)

```
GET FILE='c:\spsstemp\SIMU_STAT_FULL.sav'.
FLIP
  VARIABLES=h1 TO h1000.
COMPUTE v_hcmle=var001.
VARIABLE LABELS v_hcmle 'Hcmle'.
EXECUTE.
SAVE OUTFILE='c:\spsstemp\temp2.sav'
  /KEEP v_hcmle.
```

(31)

```
GET FILE='c:\spsstemp\SIMU_STAT_FULL.sav'.
FLIP
  VARIABLES=n1 TO n1000.
COMPUTE v_hchao=var001.
VARIABLE LABELS v_hchao 'Hchao'.
EXECUTE.
SAVE OUTFILE='c:\spsstemp\temp3.sav'
  /KEEP v_hchao.
```

(32)

```
MATCH FILES FILE='c:\spsstemp\temp1.sav'
  /FILE='c:\spsstemp\temp2.sav'
  /FILE='c:\spsstemp\temp3.sav'.
FORMATS v_hmle v_hcmle v_hchao (F8.7).
```

```
SAVE OUTFILE='c:\spsstemp\SIMU_STAT_SUMMARY.sav'.

SAVE TRANSLATE OUTFILE='c:\spsstemp\SIMU_STAT_SUMMARY.xls'     (33)
 /FIELDNAMES
 /REPLACE.
```

2. Subroutine for simulate random uniform distribution:

The syntax shows in lines between (1) and (2).

3. Subroutine for computing the Shannon's diversity index for parameter:

The syntax shows in lines between (3) and (4).

4. Subroutine for computing the coefficient variation (CV) for parameter:

The syntax shows in lines from (5) to (8).

5. Subroutine for transfer the results of parameter to show in Microsoft Excel:

The syntax shows in lines from (9) to (12).

6. Subroutine for computing X_i with random binomial and n individuals in the sample:

The syntax shows in lines from (13) to (15) for computing X_i and in line (16) for computing n.

7. Subroutine for MLE method:

The syntax shows in lines from (17) to (19).

8. Subroutine for CMLE method:

The syntax shows in lines between (20) and (21) for computing $\hat{S}$ and in line (22) for computing CMLE.

9. Subroutine for the Chao & Shen Estimator method:

The syntax shows in lines between (23) and (24) for computing f_1 and in line (25) for computing $\hat{C}$ and in lines from (26) to (28) for computing the Chao & Shen Estimator method.

10. Subroutine for transfer the results of three estimators to show in Microsoft Excel:

The syntax shows in lines from (29) to (33).

A SOFTWARE PROGRAM II

(R: A Programming Environment for Data Analysis and Graphics Version 2.6.0)

1. Main Program

```
### SIMULATION
# START PROGRAM
allH <- numeric()
allCV <- numeric()
allHmle <- numeric()
allHcmle <- numeric()
allHchao <- numeric()
IAi <- numeric()
i <- 1
set.seed <- 9999
# Begin loop
while (i<=1000){
# Parameter section; Number of species (s) = 20, 40, 60, 80, 100
        ai <- runif(s , min=0, max=1)
        b <- sum(ai)
        c <- 1/b
        pi <- ai*c
        H <- abs(sum(pi*(log(pi))))
        CV <- sqrt(sum((pi-mean(pi))^2)/s)/mean(pi)
        allH <- c(allH, H)
        allCV <- c(allCV, CV)
# Statistics section; N = 50, 75, 100, 200, 500, 1000, 2000
        N <- c(50, 75, 100, 200, 500, 1000, 2000)
        j <- 1
        while (j<=7){
            n <- N[j]
            xi <- rbinom(s , n, pi)
```

```
        sumxi <- sum(xi)
        temp <- (xi/sumxi)*(log(xi/sumxi))
        Hmle <- abs(sum(temp[!is.nan(temp)]))
        S <- length(xi[xi>0])
        Hcmle <- Hmle+((S-1)/(2*sumxi))
        f1 <- length(xi[xi==1])
        C <- 1-(f1/sumxi)
        IAi <- ifelse(xi > 0, 1, 0)
        temp<-(C*xi/sumxi)*(log(C*xi/sumxi))/(1-(1-C*xi/sumxi))^sumxi)*IAi
        Hchao <- abs(sum(temp[!is.nan(temp)]))
        allHmle <- c(allHmle, Hmle)
        allHcmle <- c(allHcmle, Hcmle)
        allHchao <- c(allHchao, Hchao)
        j <- j+1}
        # End loop lvl2
    i <- i+1}
    # End loop lvl1
# Export outputs to files
parasec=data.frame(H=allH, CV=allCV)
statsec50=data.frame(Hmle=allHmle[seq(1, 7000, by=7)], Hcmle=allHcmle[seq(1,
7000, by=7)], Hchao=allHchao[seq(1, 7000, by=7)])
statsec75=data.frame(Hmle=allHmle[seq(2, 7000, by=7)], Hcmle=allHcmle[seq(2,
7000, by=7)], Hchao=allHchao[seq(2, 7000, by=7)])
statsec100=data.frame(Hmle=allHmle[seq(3, 7000, by=7)], Hcmle=allHcmle[seq(3,
7000, by=7)], Hchao=allHchao[seq(3, 7000, by=7)])
statsec200=data.frame(Hmle=allHmle[seq(4, 7000, by=7)], Hcmle=allHcmle[seq(4,
7000, by=7)], Hchao=allHchao[seq(4, 7000, by=7)])
statsec500=data.frame(Hmle=allHmle[seq(5, 7000, by=7)], Hcmle=allHcmle[seq(5,
7000, by=7)], Hchao=allHchao[seq(5, 7000, by=7)])
statsec1000=data.frame(Hmle=allHmle[seq(6, 7000, by=7)], Hcmle=allHcmle[seq(6,
7000, by=7)], Hchao=allHchao[seq(6, 7000, by=7)])
```

```
statsec2000=data.frame(Hmle=allHmle[seq(7, 7000, by=7)], Hcmle=allHcmle[seq(7, 7000, by=7)], Hchao=allHchao[seq(7, 7000, by=7)])
allsec50=data.frame(parasec, statsec50)
allsec75=data.frame(parasec, statsec75)
allsec100=data.frame(parasec, statsec100)
allsec200=data.frame(parasec, statsec200)
allsec500=data.frame(parasec, statsec500)
allsec1000=data.frame(parasec, statsec1000)
allsec2000=data.frame(parasec, statsec2000)
write.csv(parasec, file="SIM_PARA.csv")
write.csv(allsec50, file="SIM_ALL_50.csv")
write.csv(allsec75, file="SIM_ALL_75.csv")
write.csv(allsec100, file="SIM_ALL_100.csv")
write.csv(allsec200, file="SIM_ALL_200.csv")
write.csv(allsec500, file="SIM_ALL_500.csv")
write.csv(allsec1000, file="SIM_ALL_1000.csv")
write.csv(allsec2000, file="SIM_ALL_2000.csv")
# END PROGRAM
```

BIOGRAPHY

NAME	Miss Chantarat Hirunkitrangsee
DATE OF BIRTH	2 November 1981
PLACE OF BIRTH	Phattalung, Thailand
INSTITUTIONS ATTENDED	Mahidol University, 2000: Bachelor of Science (Plant Science) Mahidol University, 2005: Master of Science (Biostatistics)
SCHOLARSHIP	This thesis is partially supported by Graduate Studies of Mahidol University Alumni Association in 2007.
HOME ADDRESS	35 Ramate Road, Mueang District, Phattalung, Thailand 93000.