



แบบจำลองที่เหมาะสมต่อการกระจายของชาป่า ในเขตพื้นที่ภาคเหนือตอนบน

THE OPTIMAL MODEL FOR THE DISTRIBUTION OF WILD TEA IN
UPPER NORTHERN OF THAILAND

สุรัสวดี นางแล

ปรัชญาดุษฎีบัณฑิต

สาขาวิชาวิทยาศาสตร์เชิงคำนวณ

สำนักวิชาวิทยาศาสตร์

มหาวิทยาลัยแม่ฟ้าหลวง

2557

©ลิขสิทธิ์ของมหาวิทยาลัยแม่ฟ้าหลวง

แบบจำลองที่เหมาะสมต่อการกระจายของชาป่า ในเขตพื้นที่ภาคเหนือตอนบน
THE OPTIMAL MODEL FOR THE DISTRIBUTION OF WILD TEA IN
UPPER NORTHERN OF THAILAND

สุรัสวดี นางแล

คุณูปการนี้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิทยาศาสตร์เชิงคำนวณ

สำนักวิชาวิทยาศาสตร์
มหาวิทยาลัยแม่ฟ้าหลวง

2557

ลิขสิทธิ์ของมหาวิทยาลัยแม่ฟ้าหลวง

แบบจำลองที่เหมาะสมต่อการกระจายของชาป่า ในเขตพื้นที่ภาคเหนือตอนบน

THE OPTIMAL MODEL FOR THE DISTRIBUTION OF WILD TEA IN
UPPER NORTHERN OF THAILAND

สุรัสวดี นางแล

คุณูปนิพนธ์นี้ได้รับการพิจารณาอนุมัติให้นับเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาปรัชญาดุษฎีบัณฑิต
สาขาวิชาวิทยาศาสตร์เชิงคำนวณ
2557

คณะกรรมการสอบคุณูปนิพนธ์

.....ประธานกรรมการ
(ดร. อนันต์ อึ้งวิเศษพันธ์)

.....อาจารย์ที่ปรึกษา
(ดร. รุ่งโรจน์ นิลทอง)

.....กรรมการภายนอกมหาวิทยาลัย
(ศาสตราจารย์ ดร. ขงยุทธ ไตรสุรัตน์)

.....กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ พิษณุ เจริญคุณ)

ลิขสิทธิ์ของมหาวิทยาลัยแม่ฟ้าหลวง

กิตติกรรมประกาศ

งานวิจัยครั้งนี้สำเร็จลงได้ด้วยความกรุณาจาก ดร. รุ่งโรจน์ นิลทอง และ ดร.อนันต์ อึ้งวิเศษพันธ์ อาจารย์ประจำสำนักวิทยาศาสตร์ มหาวิทยาลัยแม่ฟ้าหลวง ศาสตราจารย์ ดร. ยงยุทธ ไตรสุรรัตน์ อาจารย์ประจำภาควิชาชีววิทยาป่าไม้ คณะวนศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ รองศาสตราจารย์พิษณุ เจียวคุณ อาจารย์ประจำภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ผู้ซึ่งกรุณาให้ความรู้ คำแนะนำ คำปรึกษา และตรวจทานแก้ไขงานวิจัย สำเร็จสมบูรณ์ กรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์ สถาบันฯและศูนย์บริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่ฟ้าหลวงที่อนุเคราะห์ข้อมูลในการทำวิจัยในครั้งนี้ ตลอดจนโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) ที่สนับสนุนทุนการศึกษา ผู้เขียนขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ท้ายที่สุดนี้ขอกราบขอบพระคุณบิดามารดา ที่ได้ให้กำลังใจและให้การสนับสนุนด้านการศึกษามาโดยตลอด และขอบคุณพี่ ๆ นักศึกษาสาขาวิชาวิทยาศาสตร์เชิงคำนวณ มหาวิทยาลัยแม่ฟ้าหลวงทุกคนที่คอยให้กำลังใจและให้คำปรึกษามาโดยตลอด หากมีสิ่งขาดตกบกพร่องหรือผิดพลาดประการใด ผู้เขียนขออภัยเป็นอย่างสูง และหวังว่างานวิจัยนี้คงเป็นประโยชน์สำหรับผู้สนใจต่อไป

สุรัสวดี นางแล

ชื่อเรื่องคุณิพนธ์	แบบจำลองที่เหมาะสมต่อการกระจายของชาป่า ในเขตพื้นที่ภาคเหนือตอนบน
ชื่อผู้เขียน	สุรัสวดี นางแล
หลักสูตร	ปรัชญาดุษฎีบัณฑิต (วิทยาศาสตร์เชิงคำนวณ)
อาจารย์ที่ปรึกษา	ดร. รุ่งโรจน์ นิลทอง

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อหาปัจจัยที่เหมาะสมต่อการเจริญเติบโตของชาป่า โดยพิจารณาปัจจัยทางด้านภูมิศาสตร์กายภาพ อาทิ ระดับความสูงจากระดับน้ำทะเล ระยะห่างจากแม่น้ำ, ทิศทางการหันเข้าหาแสงของพื้นที่, ข้อมูลชั้นดินและความลาดชัน เป็นต้น และปัจจัยทางด้านสภาพภูมิอากาศ อาทิ อุณหภูมิ ปริมาณน้ำฝน และความชื้นสัมพัทธ์ เป็นต้น โดยการสร้างแบบจำลองการกระจายของชาป่าในพื้นที่ภาคเหนือตอนบนด้วยแบบจำลอง Generalized Linear Models (GLMs), แบบจำลอง Generalized Additive Models (GAMs) และแบบจำลอง Maximum Entropy (MAXENT) โดยมีการนำเทคนิค Resampling มาใช้ในการแก้ไขปัญหาคำนวณกลุ่มตัวอย่างที่มีจำนวนน้อย และการใช้ข้อมูลเสมือนของจุดที่ไม่พบชาป่า (Pseudo-Absence Points) ซึ่งได้มาจากการสุ่มในวิธีที่แตกต่างกัน นอกจากนี้ยังมีการศึกษาผลกระทบของการเปลี่ยนแปลงของสภาพภูมิอากาศ (Climate Change) ต่อการกระจายของต้นชาป่า

จากผลการศึกษาในกรณีแบบจำลอง GLMs ที่มีการนำข้อมูล Pseudo-absence มาใช้ในการคำนวณร่วมพบว่า ตัวแปรที่สำคัญได้แก่ ตัวแปรความชื้นสัมพัทธ์, Isothermality, อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด, ปริมาณน้ำฝนรายปี, ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุดและกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดดินและความลาดชัน ส่วนในการศึกษาการสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี GAMs พบว่าแบบจำลอง GAMs ที่ได้จากการนำตัวแปรที่มีนัยสำคัญทางสถิติจากแบบจำลอง GLMs มาปรับปรุงมีค่า AUC สูงขึ้น สำหรับการสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี MAXENT จะทำการวิเคราะห์องค์ประกอบ (Factors

Analysis) ก่อน ซึ่งสามารถจัดกลุ่มตัวแปรได้ 7 กลุ่มปัจจัย โดยกลุ่มปัจจัยซึ่งมีค่าการมีส่วนร่วมของตัวแปร (Variable Contribution) มากที่สุดเป็นปัจจัยที่เกี่ยวข้องกับตัวแปรอุณหภูมิกับความสูงจากระดับน้ำทะเล ประกอบไปด้วยตัวแปรความสูงจากระดับน้ำทะเล, อุณหภูมิเฉลี่ยรายปี, อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด, อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด, อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด, อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด, อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุดและตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด

เมื่อเปรียบเทียบแบบจำลอง GLMs, GAMs และ MAXENT พบว่า แผนที่ยพยากรณ์ของแบบจำลอง MAXENT ซึ่งใช้ข้อมูลจากการค้นพบต้นขาเพียงอย่างเดียว ให้ค่าที่มีความแตกต่างจากแผนที่ยพยากรณ์ของแบบจำลอง GLMs และ GAMs ซึ่งใช้ข้อมูลจากการค้นพบขาและข้อมูลเสมือนของจุดที่ไม่พบต้นขา (Pseudo-Absence Points) อย่างเห็นได้ชัดเจนน ซึ่งเมื่อพิจารณาค่า AUC พบว่าการสร้างแบบจำลอง GAMs จะให้ค่า AUC มากที่สุด รองลงมาคือการสร้างแบบจำลอง MAXENT และ GLMs ตามลำดับ

ในการศึกษาผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นขาป่า โดยเปรียบเทียบระหว่างสภาพภูมิอากาศในปีปัจจุบัน (ค.ศ. 2000) และปีอนาคต (ค.ศ. 2050) พื้นที่การกระจายของขาป่าส่วนใหญ่ไม่มีการเปลี่ยนแปลงมากนัก โดยเฉลี่ยอยู่ที่ประมาณร้อยละ 75-90 ของพื้นที่ภาคเหนือตอนบน โดยพื้นที่ที่คาดว่าจะมีการลดลงของการกระจายของต้นขาป่า โดยเฉลี่ยอยู่ที่ร้อยละ 6-23 และเมื่อเปรียบเทียบกันในแต่ละแบบจำลองพบว่าแบบจำลอง GLMs และ GAMs ให้ผลอย่างมีนัยสำคัญต่อการประเมินความเสี่ยงต่อการสูญพันธุ์ของขาป่า

คำสำคัญ: แบบจำลองการกระจายของสิ่งมีชีวิต/ Generalized Linear Models/ Generalized Additive Models/ Maximum Entropy/ การเปลี่ยนแปลงสภาพภูมิอากาศ

Dissertation Title	The Optimal Model for the Distribution of Wild Tea in Upper Northern of Thailand
Author	Suruswadee Nanglae
Degree	Doctor of Philosophy (Computational Science)
Advisor	Dr. Rungrote Nilthong

ABSTRACT

Wild tea has been found widely in northern part of Thailand. However, determination of wild tea distribution is also complexity that makes difficult to perceive the implications of environmental information. Here, this study was to investigate the key environmental parameters including geological factors (e.g. elevation, distance from rivers, aspect, rock and slope) and climate factors (e.g. Temperature, precipitation and humidity) which are considered that affected the distribution of wild tea. To generate species distribution models (SDMs), the Generalized Linear Models (GLMs), Generalized Additive Models (GAMs) and Maximum Entropy (MAXENT) were used. Resampling technique was applied due to limited sample sizes. In case of GLMs and GAMs, needed pseudo-absence points had been generated from various methods to pursuit an optimal model. Moreover, the results of the predictions could illustrate the effect of climate change on the distribution of wild tea.

The distribution models were developed and key factors were analyzed in this study. GLMs was carried out to generate model derived from presence and pseudo-absence data. The model suggested humidity, isothermality, average temperature of wettest quarter, annual precipitation, precipitation in driest month, rock type and slope which were linked to distribution of wild tea. Promisingly, GAMs improved by using key factors from GLMs showed higher value

in the area under the curve (AUC). Lastly, MAXENT was determined through factor analysis and categorized the factors into 7 distinct groups. The key variable factors were temperature and elevation such as annual average temperature, maximum temperature of warmest month, minimum temperature of coldest month, average temperature of wettest quarter, average temperature of driest quarter, average temperature of warmest quarter and average temperature of coldest quarter.

In comparison of these models, GAMs which used key factors from GLMs had significantly demonstrated higher AUC compared to the two models, MAXENT and GLMs. In addition, the effect of climate change on distribution of wild tea has been studied to predict for next 50 years (from 2000 to 2050). The model suggested that wild tea tree will still grow and distribute mainly in northern part of Thailand approximately 75-90% but only 6-23% will be lost. Importantly, both models derived from GLMs and GAMs showed as substantial models to foresee the extinction of wild tea in future.

Keywords: Species Distribution Models (SDMs)/Generalized Linear Models (GLMs)/
Generalized Additive Model (GAMs)/ Maximum Entropy (MAXENT) / Climate
Change

สารบัญ

	หน้า
กิตติกรรมประกาศ	(3)
บทคัดย่อภาษาไทย	(4)
บทคัดย่อภาษาอังกฤษ	(6)
สารบัญตาราง	(11)
สารบัญภาพ	(15)
บทที่	
1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการศึกษา	3
1.3 กรอบแนวความคิด	4
1.4 ประโยชน์ที่คาดว่าจะได้รับจากการศึกษา	5
2 ทฤษฎีที่เกี่ยวข้องและแนวคิดในการศึกษา	6
2.1 ชาป่า (Wild Tea)	6
2.2 แบบจำลองการกระจายของสิ่งมีชีวิต (Species Distribution Models: SDMs)	14
2.3 เทคนิควิธีการสร้างแบบจำลองการกระจายตัวของสิ่งมีชีวิต	24
3 ระเบียบและวิธีการศึกษา	39
3.1 ขอบเขตของการศึกษา	39
3.2 การเก็บรวบรวมข้อมูล	39
3.3 การสร้างชุดข้อมูลสำหรับการสร้างแบบจำลอง	54
3.4 กระบวนการในการสร้างแบบจำลอง การตรวจสอบ และการพยากรณ์	59

สารบัญ (ต่อ)

บทที่	หน้า
3.5 การเปรียบเทียบผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change)	65
3.6 การวิเคราะห์ข้อมูลโดยวิธีทางสถิติ	68
4 ผลการวิเคราะห์ข้อมูล	71
4.1 การวิเคราะห์ข้อมูลทั่วไป	71
4.2 การสร้างแบบจำลองการกระจายของสิ่งมีชีวิต	79
4.3 การเปรียบเทียบความแตกต่างของแบบจำลองการกระจายของสิ่งมีชีวิต	157
4.4 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นชาป่า	174
5 สรุปผลการศึกษาและข้อเสนอแนะ	184
5.1 สรุปผลการศึกษาและอภิปรายผลการศึกษา	184
5.2 ข้อเสนอแนะ	198
รายการอ้างอิง	200
ภาคผนวก	208
ภาคผนวก ก วิธีคัดเลือกรัศมีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจาก การสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา	209
ภาคผนวก ข วิธีคัดเลือกรัศมีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจาก การสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence	231
ภาคผนวก ค วิธีคัดเลือกรัศมีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจาก การสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน	257

สารบัญ (ต่อ)

บทที่	หน้า
ภาคผนวก ง แบบจำลอง GLMs โดยแยกวิเคราะห์ระหว่างตัวแปรทางด้าน ภูมิศาสตร์และตัวแปรทางด้านภูมิอากาศ	264
ภาคผนวก จ ข้อมูลการสำรวจต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน	299
ประวัติผู้เขียน	308



สารบัญตาราง

ตาราง	หน้า
2.1 ตัวอย่างของฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function) ที่ถูกจัดให้อยู่ในกลุ่มของ Exponential Family	25
2.2 Canonical Link Function ของแต่ละฟังก์ชันการแจกแจงความน่าจะเป็น	26
3.1 ตัวอย่างข้อมูลที่นำมาทดสอบโดยวิธี Cochran's Q Test	69
4.1 ค่าสถิติของความสูงจากระดับน้ำทะเล ความลาดชันและระยะห่างจากแม่น้ำของพื้นที่ที่มีการค้นพบต้นชาป่าจำนวน 185 จุด	72
4.2 จำนวนและร้อยละของจุดพิกัดของการค้นพบต้นชาป่าจำนวน 185 จุด จำแนกตามโครงสร้างทางธรณีวิทยาของชนิดหิน ประเภทของความลาดชันและการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน	74
4.3 ค่าสถิติของความชื้นสัมพัทธ์ (Humidity) ในพื้นที่ที่มีการค้นพบต้นชาป่า	75
4.4 ค่าสถิติของตัวแปรด้าน Bioclimatic (Bio1-Bio11) ที่เป็นตัวแปรที่เกี่ยวข้องกับอุณหภูมิ จำนวน 185 จุด	76
4.5 ค่าสถิติของตัวแปรด้าน Bioclimatic (Bio12-Bio19) ที่เป็นตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน (หน่วย : มิลลิเมตร) จำนวน 185 จุด	78
4.6 ค่า Cross-Validation Error ของชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตร	80
4.7 ค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 10 ชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตร	81
4.8 ค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3	83
4.9 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา โดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว	84
4.10 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา	88

สารบัญตาราง (ต่อ)

ตาราง	หน้า
4.11 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว	91
4.12 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence	97
4.13 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอนซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว	100
4.14 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน	105
4.15 ค่าสถิติทดสอบในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา	108
4.16 แบบจำลอง GAMs และค่าองค์ความเป็นอิสระของ Smooth Function ในแบบจำลองที่มีการปรับปรุงทั้ง 2 ครั้ง	109
4.17 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	112
4.18 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	114
4.19 ค่าสถิติทดสอบ Chi-square และค่าองค์ความเป็นอิสระของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence	117
4.20 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs	120
4.21 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GAMs	122
4.22 ค่าสถิติทดสอบของตัวแบบในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน	125

สารบัญตาราง (ต่อ)

ตาราง	หน้า
4.23 ค่าสถิติทดสอบ Chi-Square และค่าองค์ความเป็นอิสระของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-absence แบบ 2 ขั้นตอน	127
4.24 ค่าสถิติทดสอบของแบบจำลอง GAMs หลังจากเปลี่ยน Smooth Function ของตัวแปร Bio17 เป็นฟังก์ชันเชิงเส้นแทน	129
4.25 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	131
4.26 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	133
4.27 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง MAXENT	136
4.28 Correlation Matrix ของแต่ละตัวแปร	140
4.29 ค่า Loading Factors ของแต่ละตัวแปรในการสร้างปัจจัย 7 ตัว	146
4.30 ค่า Weight Score ของแต่ละตัวแปรในการสร้างปัจจัย 7 ตัว	150
4.31 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบ	154
4.32 ค่า Coefficient Value ของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี	158
4.33 ผลการตรวจสอบแบบจำลอง GLMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธีโดยแยกตามประเภทของตัวแปร	160
4.34 ผลของการทดสอบโดยวิธี Cochran's Q Test	164
4.35 ผลการตรวจสอบแบบจำลอง GAMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธี	166
4.36 ผลของการทดสอบโดยวิธี Cochran's Q Test	169

สารบัญตาราง (ต่อ)

ตาราง	หน้า
4.37 ผลการตรวจสอบแบบจำลอง GLMs, GAMs และ MAXENT	170
4.38 ผลของการทดสอบโดยวิธี Cochran's Q Test	173
4.39 ค่าสถิติของปัจจัยอุณหภูมิในปัจจุบันเปรียบเทียบกับอนาคต	174
4.40 ค่าสถิติของปัจจัยปริมาณน้ำฝนในปัจจุบันเปรียบเทียบกับอนาคต	175
5.1 ค่าเฉลี่ยของตัวแปรที่มีผลต่อแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-absence ทั้ง 3 วิธี โดยจำแนกตามจังหวัดที่มีการค้นพบต้นชาป่า	188
5.3 ร้อยละของจุดพิกัดของการค้นพบต้นชาป่าตามการจัดกลุ่มของโครงสร้างทาง ธรณีวิทยาของชนิดหินและประเภทของความลาดชัน โดยจำแนกตามจังหวัดที่มี การค้นพบต้นชาป่า	189
5.3 ร้อยละของพื้นที่พยากรณ์ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า และการเปลี่ยนแปลงของพื้นที่พยากรณ์ระหว่างเวลาปัจจุบันและอนาคต อันสืบ เนื่องมาจากการเปลี่ยนแปลงสภาพภูมิอากาศ	196

สารบัญภาพ

ภาพ	หน้า
1.1 กรอบแนวความคิดของตัวแปร	4
2.1 ต้นชาป่าซึ่งถูกจัดอยู่ในกลุ่มพันธุ์ชาอัสสัม ที่หมู่บ้านป่าเมี่ยง ตำบลป่าตึง อำเภอแม่จัน จังหวัดเชียงราย	7
2.2 ลักษณะใบและเมล็ดของต้นชาป่า	8
2.3 ความสัมพันธ์ของ Species Distribution ในรูปแบบของพื้นที่เชิงภูมิศาสตร์ (Geographical Space) และพื้นที่เชิงสิ่งแวดล้อม (Environmental Space)	17
2.4 วิธีการของ SDMs ที่สอดคล้องกับข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่	20
2.5 วิธีการของแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี BIOCLIM	22
2.6 วิธีการของแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี DOMAIN	22
2.7 การจำแนกทางเลือกของความน่าจะเป็นที่เป็นไปได้ 3 ทาง	37
3.1 จุดพิกัด (จุดสีดำ) ที่มีการค้นพบต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบนจำนวน 185 จุด	40
3.2 ข้อมูลแผนที่ลักษณะความสูงของภูมิประเทศแบบดิจิทัล (Digital Elevation Model: DEM) ในเขตพื้นที่ภาคเหนือตอนบน โดยมีหน่วยเป็น เมตร	41
3.3 ความชื้นสัมพัทธ์โดยมีหน่วยเป็น เปอร์เซ็นต์	42
3.4 พื้นที่ป่าในเขตภาคเหนือตอนบน	43
3.5 ขอบเขตของชั้นหินในเขตพื้นที่ภาคเหนือตอนบน โดยพื้นที่ที่สีน้ำตาลเป็นส่วนของหินตะกอนและหินแปร พื้นที่สีเขียวเป็นส่วนของหินอัคนีและพื้นที่สีน้ำเงินเป็นส่วนของพื้นน้ำ	44
3.6 ตัวแปรที่เกี่ยวข้องกับอุณหภูมิ (Bio1-Bio11) ในเขตพื้นที่ภาคเหนือตอนบน	46
3.7 ตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน (Bio12-Bio19) ในเขตพื้นที่ภาคเหนือตอนบน	48
3.8 ประเภทของความลาดชันในเขตพื้นที่ภาคเหนือตอนบน	49

สารบัญภาพ (ต่อ)

ภาพ	หน้า
3.9 ทิศด้านลาดในเขตพื้นที่ภาคเหนือตอนบน	50
3.10 ระยะห่างจากแม่น้ำสายหลักและแม่น้ำสาขาในเขตพื้นที่ภาคเหนือตอนบน	51
3.11 การจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชันในเขตพื้นที่ภาคเหนือตอนบน	52
3.12 ข้อมูลของตัวแปรหุ่น Class1-Class8 ในเขตพื้นที่ภาคเหนือตอนบน	53
3.13 หลักการการสุ่มจุด Pseudo – Absence	55
3.14 ตัวอย่างการสุ่มจุด Pseudo-Absence โดยการสุ่มในพื้นที่ที่กำลังศึกษาที่รัศมี 30 กิโลเมตรและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร	55
3.15 ตัวอย่างการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ที่รัศมี 30 กิโลเมตรและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร	56
3.16 ตัวอย่างการสุ่มจุด Pseudo – Absence แบบสองขั้นตอนและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร	57
3.17 ขั้นตอนการสร้างชุดข้อมูล (Data Set)	58
3.18 ขั้นตอนการสร้างแบบจำลองโดยนำวิธี Bootstrap และ วิธี k-Fold Cross-Validation มาประยุกต์ใช้	60
3.19 ขั้นตอนการคัดเลือกแบบจำลองจากชุดของแบบจำลองที่มีค่า Cross-Validation Error น้อยที่สุด	61
3.20 วิธีการหาแบบจำลองที่เหมาะสมในแต่ละวิธีของการสร้างชุด Pseudo-Absence	64
3.21 ตัวแปรภูมิอากาศที่เกี่ยวข้องกับอุณหภูมิ (Bio1-Bio11) ในปัจจุบันเปรียบเทียบกับปี ค.ศ. 2050	66
3.22 ตัวแปรภูมิอากาศที่เกี่ยวข้องกับปริมาณน้ำฝน (Bio12-Bio19) ในปัจจุบันเปรียบเทียบกับปี ค.ศ. 2050	67

สารบัญภาพ (ต่อ)

ภาพ	หน้า
4.1 จำนวนจุดพิกัดและร้อยละของการค้นพบต้นชาป่าจำนวน 185 จุด จำแนกตามทิศด้านลาด	73
4.2 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า Error Rate ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง	82
4.3 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง	82
4.4 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา	87
4.5 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	89
4.6 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence	96
4.7 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	98
4.8 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน	104
4.9 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	106

สารบัญภาพ (ต่อ)

ภาพ	หน้า
4.10 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	111
4.11 แผนที่มีความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	113
4.12 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	115
4.13 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs	119
4.14 แผนที่มีความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs	121
4.15 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	123
4.16 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Bio17	128
4.17 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	130
4.18 แผนที่มีความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว	132
4.19 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	134
4.20 ค่า Variable Contribution ของทุกตัวแปร	135
4.21 แผนที่มีความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง MAXENT	137

สารบัญภาพ (ต่อ)

ภาพ	หน้า
4.22 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	138
4.23 ผลการวิเคราะห์ของ Parallel Analysis โดย Scree Plots	145
4.24 ผลการจัดกลุ่ม Factors	149
4.25 Factors ทั้ง 7 ตัวในเขตพื้นที่ภาคเหนือตอนบน	152
4.26 ค่า Variable contribution ของตัวแปร Factors และตัวแปร Class	153
4.27 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบ	155
4.28 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบเปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน	156
4.29 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี	162
4.30 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี	163
4.31 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี	167
4.32 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี	168
4.33 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าเปรียบเทียบระหว่างแบบจำลอง GLMs, GAMs และ MAXENT	171

สารบัญภาพ (ต่อ)

ภาพ	หน้า
4.34 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบระหว่างแบบจำลอง GLMs, GAMs และ MAXENT	172
4.35 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1)	176
4.36 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2)	177
4.37 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3)	178
4.38 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1)	179
4.39 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2)	180
4.40 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3)	181
4.41 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง MAXENT ก่อนการวิเคราะห์องค์ประกอบ	182

สารบัญภาพ (ต่อ)

ภาพ

หน้า

4.42 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดย
แบบจำลอง MAXENT หลังการวิเคราะห์หิ้งค์ประกอบ

183



บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ชา (*Camillia sinensis*) จัดเป็นพืชที่สำคัญชนิดหนึ่งในภาคเหนือของประเทศไทย ชาส่วนใหญ่ที่นิยมเพาะปลูกในเชิงเศรษฐกิจได้แก่ชาสายพันธุ์จีน (*Camillia sinensis* ver. *sinensis*) ถึงแม้ว่าชาจีนจะได้รับความนิยมในการเพาะปลูก แต่ในบางพื้นที่ยังคงมีการปลูกชาพันธุ์พื้นเมืองหรือชาป่า ซึ่งจัดเป็นชาสายพันธุ์อัสสัม (*Camillia sinensis* ver. *assamica*) มีแหล่งกำเนิดดั้งเดิมจากประเทศอินเดีย โดยกระจายตัวตามธรรมชาติบนพื้นที่ภูเขาสูง ส่วนใหญ่อยู่ในพื้นที่ป่าดิบเขา และเจริญเติบโตได้ดีตามป่าที่มีร่มไม้ แสงแดดส่องผ่านพุ่มประมาณ ชาส่วนใหญ่เจริญเติบโตได้ดีในเขตร้อนชื้น มีปริมาณน้ำฝนสม่ำเสมอ โดยมีอุณหภูมิที่เหมาะสมอยู่ที่ 18-35 °C (Braber, Sato & Lee, 2011)

ชาป่านอกจากจะผลิตเป็นชาขงแล้ว ตามชนบททางภาคเหนือของไทยยังนิยมนำมาทำเป็นเมี่ยง ซึ่งเป็นใบชาที่นำมาหมักไว้จนใบชาเปลี่ยนสภาพเป็นสีเหลือง ใบชู่ แล้วจึงนำมาบริโภคโดยการเคี้ยวหรืออมเป็นของว่าง และเรียกพื้นที่ที่มีชาป่าขึ้นอยู่ว่า “ป่าเมี่ยง” เนื่องจากลักษณะพื้นที่เป็นป่า ซึ่งป่าเมี่ยงถูกจัดให้เป็นระบบวนเกษตร (Agroforestry) ที่สำคัญ ซึ่งเป็นการผสมผสานระหว่างไม้ยืนต้น พืชเศรษฐกิจ และปศุสัตว์ (สุรจิต ภูภักดิ์, 2549) โดยมีหมู่บ้านเล็ก ๆ กระจายอยู่รอบ ๆ ป่าเมี่ยง โดยทั่วไปจะเรียกหมู่บ้านเหล่านี้ว่า “หมู่บ้านป่าเมี่ยง” โดยมีประชาชนหรือชาวป่าเมี่ยงใช้ประโยชน์จากพื้นที่ป่าเมี่ยง จากการสำรวจระบบวนเกษตรของชาวพื้นเมืองบนพื้นที่สูงทางภาคเหนือของไทย (Preechapanya, 2002) พบว่า ป่าเมี่ยงประกอบไปด้วยความหลากหลายทางชีวภาพมากมาย โดยเฉพาะพืชที่เป็นประโยชน์ต่อการดำรงชีวิตเป็นจำนวนมาก รวมทั้งพืชอาหารที่ขึ้นเอง และต้นไม้ใหญ่ที่ปล่อยให้ขึ้นเพื่อเป็นร่มเงาแก่ต้นชาป่า และเป็นไม้พื้นสำหรับนั่งใบเมี่ยงหรือไว้ใช้ในการสร้างสิ่งปลูกสร้างหรือเครื่องเรือน นอกจากนี้ยังมีพืชที่เป็นพืชสมุนไพร ดังนั้นชาวป่าเมี่ยงจะทำหน้าที่เป็นผู้ปกป้องรักษาผืนป่าเมี่ยง และจากการศึกษาภูมิปัญญาชาวบ้านด้านนิเวศน์ที่เกี่ยวกับความยั่งยืนของป่าเมี่ยงในป่าดิบเขาทางภาคเหนือของประเทศไทย (พรชัย ปรีชาปัญญา, 2536) พบว่า พื้นที่ป่าเมี่ยงส่วนใหญ่อยู่ในเขตลุ่มน้ำ 1A ซึ่งสำนักอนุรักษ์และจัดการ

ต้นน้ำ กรมอุทยานแห่งชาติ สัตว์ป่า และพันธุ์พืช (ม.ป.ป.) อธิบายความหมายของเขตลุ่มน้ำ 1A หมายถึงพื้นที่ต้นน้ำลำธารที่ยังมีสภาพป่าสมบูรณ์ ในปี พ.ศ. 2525 สำหรับลุ่มน้ำปิง วัง ยม น่าน ชี มูล และลุ่มน้ำภาคใต้, ปี พ.ศ. 2528 สำหรับลุ่มน้ำภาคตะวันออก และปี พ.ศ.2531 สำหรับลุ่มน้ำ ตะวันตก ภาคกลาง ลุ่มน้ำป่าสัก ลุ่มน้ำภาคเหนือและตะวันออกเฉียงเหนือ และส่วนอื่น ๆ (ลุ่มน้ำ ชายแดน) ซึ่งห้ามไม่ให้มีการเปลี่ยนแปลงลักษณะพื้นที่ป่าไม้เป็นรูปแบบอย่างอื่นอย่างเด็ดขาดใน ทุกกรณี ดังนั้นพื้นที่ป่าเมืองจึงถือเป็นแหล่งต้นน้ำและทรัพยากรธรรมชาติที่สำคัญ และจากการ สำรวจพบว่า ป่าเมืองมีพื้นที่ครอบคลุมราว 170,000 ตารางกิโลเมตร โดยพื้นที่ประมาณร้อยละ 45 อยู่ในพื้นที่ภูเขาสูง (Pagella, Choocharoen, Preechapanya, Moss & Sinclair, 2002)

จากการสำรวจหมู่บ้านป่าเมืองพบว่า ส่วนใหญ่จะกระจายอยู่ล้อมรอบเทือกเขาฝืนน้ำ ซึ่งเป็นแหล่งต้นน้ำที่สำคัญของไทย อาทิ แม่น้ำวัง แม่น้ำยม แม่น้ำน่าน เป็นต้น โดยอยู่บริเวณรอยต่อ ระหว่างจังหวัดเชียงใหม่ เชียงราย พะเยา ลำปาง และแพร่ (Preechapanya, 2002) ชาวป่าเมืองจะเป็น ผู้ใช้ประโยชน์จากพื้นที่ป่าเมือง ดังนั้นหากพบเห็นการทำลายป่าหรือไฟป่า ก็จะช่วยกันยับยั้งทันที เนื่องจากภัยพิบัติที่เกิดขึ้นจะส่งผลโดยตรงต่อชุมชน ดังนั้นป่าเมืองจึงถือเป็นแหล่งพื้นที่กันชนที่ ป้องกันแหล่งต้นน้ำจากไฟป่า และการบุกรุกทำลายพื้นที่ป่าเพื่อการเกษตร ซึ่งในการศึกษาพื้นที่ที่มีความ เหมาะสมต่อการเจริญเติบโตของต้นชาป่าโดยวิธีการสร้างแบบจำลองการกระจายของ สิ่งมีชีวิต (Species Distribution Models : SDMs) นอกจากสามารถนำไปใช้ประโยชน์เพื่อการ อนุรักษ์ รวมถึงการนำไปใช้ประโยชน์เพื่อการจัดการทรัพยากรธรรมชาติและความหลากหลายทาง ชีวภาพในพื้นที่แล้ว ยังมีประโยชน์ต่อการบริหารและการจัดการแนวป่ากันชนเพื่อป้องกันป่าต้นน้ำ ลำธารที่สำคัญในพื้นที่หรือนำไปประยุกต์ใช้เพื่อการจัดการในพื้นที่ที่มีศักยภาพเพียงพอต่อการ เจริญเติบโตของต้นชาป่า ทั้งทางด้านสิ่งแวดล้อม เศรษฐกิจ และสังคม รวมถึงการพัฒนาพื้นที่ที่มี การเจริญเติบโตของต้นชาป่าให้สอดคล้องกับสภาพแวดล้อมทางธรรมชาติในพื้นที่อีกด้วย

แบบจำลองการกระจายตัวของสิ่งมีชีวิตโดยทั่วไปมีวัตถุประสงค์หลักคือ การบ่งบอก ลักษณะปัจจัยทางสิ่งแวดล้อมที่เหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตและสามารถระบุพื้นที่ซึ่ง มีปัจจัยทางสิ่งแวดล้อมที่เหมาะสมได้ (Pearson, 2007) โดยการนำข้อมูลในส่วนของระบบ สารสนเทศภูมิศาสตร์ (Geographic Information System: GIS) และข้อมูลการค้นพบของสิ่งมีชีวิตที่ ต้องการศึกษในพื้นที่ (Species Occurrence Data หรือ Presence Data) ซึ่งอยู่ในรูปแบบของจุดพิกัด ที่มีการค้นพบของสิ่งมีชีวิตที่ต้องการศึกษาในพื้นที่นั้น หรือข้อมูลการทำงานของลักษณะเฉพาะใน สิ่งมีชีวิตที่กำลังศึกษา (Functional Traits Data) มาประมวลผลโดยวิธีการทางคณิตศาสตร์และสถิติ เพื่อพยากรณ์ขอบเขตพื้นที่การกระจายของสิ่งมีชีวิตที่กำลังศึกษา เป็นแบบจำลองที่มีความสำคัญ

อย่างยิ่ง ในการแก้ไขปัญหาต่าง ๆ ในระบบนิเวศวิทยา (Ecology), ชีวภูมิศาสตร์ (Biogeography), วิวัฒนาการของสิ่งมีชีวิต (Evolution), การวางแผนเพื่อการอนุรักษ์ (Conservation Planning) และงานวิจัยที่เกี่ยวกับการเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change) (Guisan & Thuiller, 2005)

ในการศึกษานี้จึงมีวัตถุประสงค์เพื่อหาปัจจัยที่เหมาะสมต่อการเจริญเติบโตของชาป่า โดยพิจารณาปัจจัยทางด้านภูมิศาสตร์กายภาพ อาทิ ระดับความสูงจากระดับน้ำทะเล ดิน ความลาดชันของพื้นที่ เป็นต้น และปัจจัยทางด้านสภาพภูมิอากาศ อาทิ อุณหภูมิ ปริมาณน้ำฝน และความชื้น เป็นต้น เพื่อนำมาสร้างแบบจำลองการกระจายของสิ่งมีชีวิตของชาป่าในพื้นที่ภาคเหนือตอนบน โดยวิธีการทางคณิตศาสตร์และสถิติ โดยมีการนำเทคนิคทางสถิติมาใช้ในการแก้ไขปัญหาในกลุ่มตัวอย่างหรือจำนวนจุดพิกัดมีขนาดเล็ก ซึ่งเลือกใช้เทคนิคการสุ่มซ้ำของข้อมูล (Resampling Technique) มาใช้ในการแก้ไขปัญหา โดยเลือกใช้วิธีการบูตสเตรป (Bootstrap) ควบคู่กับวิธีการประเมินความแม่นยำด้วยวิธีการไขว้ข้าม k กลุ่ม (k -Fold Cross-Validation) (Fu, Carroll & Wang, 2005) นอกจากนี้ยังศึกษาวิธีการในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต เพื่อหาแบบจำลองที่เหมาะสม รวมถึงการศึกษาผลกระทบต่อการกระจายของต้นชาป่าที่เกิดจากการเปลี่ยนแปลงของสภาพภูมิอากาศ (Climate Change)

1.2 วัตถุประสงค์ของการศึกษา

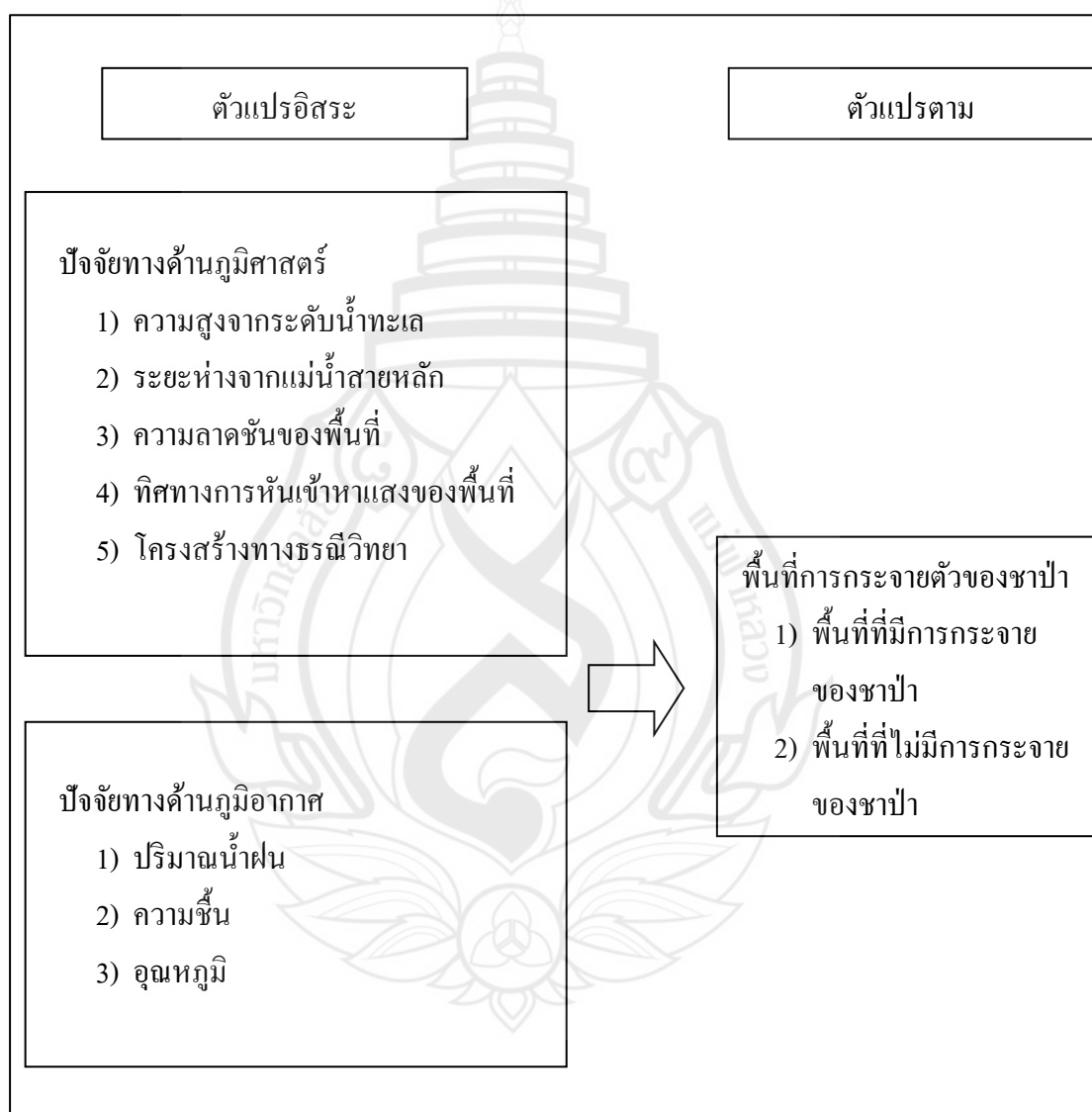
1.2.1 สร้างแบบจำลองเพื่อแสดงการกระจายของต้นชาป่าในพื้นที่ภาคเหนือตอนบนของประเทศไทยโดยวิธีการ Generalized Linear Models (GLMs), Generalized Additive Models (GAMs) และ Maximum Entropy (MAXENT)

1.2.2 ศึกษาปัจจัยทั้งทางด้านภูมิศาสตร์และภูมิอากาศที่มีความเหมาะสมต่อการเจริญเติบโตของชาป่า

1.2.3 ศึกษาผลกระทบต่อการกระจายของต้นชาป่าที่เกิดจากการเปลี่ยนแปลงของสภาพภูมิอากาศ (Climate Change)

1.3 กรอบแนวความคิด

ในการสร้างแบบจำลองเพื่อแสดงการกระจายตัวของชาป่าในเขตพื้นที่ภาคเหนือ ตัวแปรที่ใช้มี 2 ชุด คือ ชุดแรกเป็นตัวแปรตามคือ พื้นที่ที่มีการกระจายตัวของชาป่า ส่วนชุดที่ 2 เป็นตัวแปรอิสระ ซึ่งได้แก่ ปัจจัยทางด้านภูมิศาสตร์กายภาพและปัจจัยทางด้านภูมิอากาศ ดังแสดงในภาพที่ 1.1



ภาพที่ 1.1 กรอบแนวความคิดของตัวแปร

1.4 ประโยชน์ที่คาดว่าจะได้รับจากการศึกษา

1.4.1 สามารถสร้างแบบจำลองการกระจายของต้นชาป่าในพื้นที่ภาคเหนือตอนบนของประเทศไทยได้

1.4.2 สามารถสรุปปัจจัยที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่าในพื้นที่ภาคเหนือของประเทศไทยได้

1.4.3 สามารถนำไปใช้เป็นแนวทางในการพัฒนาและการอนุรักษ์เชิงพื้นที่ ในบริเวณพื้นที่ที่คาดว่าจะมีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

1.4.4 สามารถนำไปใช้เป็นแนวทางในการป้องกันผลกระทบที่อาจเกิดขึ้นจากการเปลี่ยนแปลงสภาพภูมิอากาศ



บทที่ 2

ทฤษฎีที่เกี่ยวข้องและแนวคิดในการศึกษา

2.1 ชาป่า (Wild Tea)

ชา (Tea) มีชื่อวิทยาศาสตร์ว่า *Camellia sinensis* จัดเป็นไม้พุ่มหรือไม้ยืนต้นขนาดเล็ก ไม่ผลัดใบ มีการกระจายทางเหนือสุดตรงเส้นรุ้งที่ 42 องศาเหนือในประเทศจอร์เจีย ใกล้ทะเลดำ และใต้สุดตรงเส้นรุ้งที่ 30 องศาใต้ ในเมืองนาตาล (Natal) ประเทศบราซิล (Harler, 1964) โดยแหล่งกำเนิดชาตามธรรมชาติมีจุดศูนย์กลางอยู่ทางตะวันออกเฉียงใต้ของจีน แล้วแพร่กระจายไปตามแนวชายแดนของรัฐอัสสัม (Assam) และสหภาพพม่า ไปยังมณฑลเจียงซีของจีนทางด้านทิศตะวันออกแล้วลงสู่ทิศใต้ตามเทือกเขาของสหภาพพม่า ตอนเหนือของไทยไปสิ้นสุดที่เวียดนาม (สุคนทรถ เกตุเจริญ และ อัญชลี พัดมีเทศ, ม.ป.ป.) ในประเทศไทยมีแหล่งกำเนิดเดิมอยู่ตามภูเขาสูงทางภาคเหนือของประเทศ โดยกระจายอยู่ในหลายจังหวัดที่สำคัญได้แก่ เชียงใหม่, เชียงราย, แม่ฮ่องสอน, แพร่, น่าน, ลำปาง, และตาก (สายลม สัมพันธ์เวชโสภา และคณะ, 2550)

2.1.1 ลักษณะทั่วไปของชาป่า

สุคนทรถ เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) ได้จำแนกพันธุ์ชาที่ปลูกในประเทศไทยออกเป็น 3 สายพันธุ์ ได้แก่

1. กลุ่มพันธุ์ชาจีน (*Camellia sinensis* ver. *sinensis*) เป็นไม้พุ่มขนาดเล็ก สูงประมาณ 1 ถึง 6 เมตร มีก้านใบสั้น ใบมีความกว้างประมาณ 1.6 ถึง 4 เซนติเมตร ความยาวใบประมาณ 4 ถึง 10 เซนติเมตร ขอบใบหยักเป็นรูปโค้งเล็กน้อย แผ่นใบมีตั้งแต่สีเขียวอ่อนถึงสีเขียวเข้ม
2. กลุ่มพันธุ์ชาอัสสัม (*Camellia sinensis* ver. *assamica*) เป็นไม้พุ่มขนาดกลางถึงใหญ่ บางครั้งอาจพบที่มีความสูงถึง 17 เมตร การเรียงตัวของใบบนกิ่งเป็นแบบสลับและเวียน ใบมีความกว้างประมาณ 3 ถึง 6 เซนติเมตร ความยาวใบประมาณ 7 ถึง 16 เซนติเมตร ขอบใบมีหยักเป็นฟันเลื่อยชัดเจน แผ่นใบมีตั้งแต่สีเขียวอ่อนถึงสีเขียวเข้ม
3. กลุ่มพันธุ์ชาเขมร (*Camellia sinensis* ver. *indo-china*) เป็นไม้พุ่ม ลำต้นเดี่ยวขนาดใหญ่ มีความสูงของทรงพุ่มประมาณ 5 เมตร ขนาดใบใกล้เคียงกับกลุ่มชาจีน ใบมีความกว้างประมาณ 3

ถึง 3.5 เซนติเมตร ความยาวใบประมาณ 7.5 เซนติเมตร ขอบใบมีหยักเป็นฟันเลื่อยชัดเจน แผ่นใบมีตั้งแต่สีเขียวอมแดงถึงสีเขียวเข้ม แต่ก็ยังอมแดงอยู่ ใบแก่ค่อนข้างกรอบ



ภาพที่ 2.1 ต้นชาป่าซึ่งถูกจัดอยู่ในกลุ่มพันธุ์ชาอัสสัม ที่หมู่บ้านป่าเมือง ตำบลป่าตึง อำเภอแม่จัน จังหวัดเชียงราย

ชาป่าหรือชาพันธุ์พื้นเมือง (ภาพที่ 2.1 และภาพที่ 2.2) ถูกจัดอยู่ในกลุ่มสายพันธุ์ชาอัสสัม (*Camellia sinensis* var. *assamica*) มีแหล่งกำเนิดจากประเทศอินเดีย โดยกระจายตัวตามธรรมชาติบนพื้นที่ภูเขาสูง ส่วนใหญ่ขึ้นในพื้นที่ป่าดิบเขา เจริญเติบโตได้ดีตามป่าที่มีร่มไม้และแสงแดดส่องผ่านพอประมาณ ผลิตรากสำคัญที่ได้จากชาป่าในพื้นที่ภาคเหนือของประเทศไทย นอกจากผลิตชาไว้ดื่มแล้ว ยังมีการผลิตเมี่ยง ซึ่งเป็นการนำใบชามามากไ้ระยะหนึ่ง แล้วจึงนำมาบริโภคโดยผสมกับเกลือและน้ำตาล บางท้องถิ่นเรียกว่า “เมี่ยงอม” และเรียกต้นชาป่าว่า “ต้นเมี่ยง” ส่วนพื้นที่ที่

ปลูกต้นเมี่ยง เรียกว่า “ป่าเมี่ยง” เนื่องจากลักษณะสวนมีลักษณะเป็นป่า โดยป่าเมี่ยงจัดเป็นระบบวนเกษตรดั้งเดิมที่สำคัญ โดยมีหมู่บ้านเล็ก ๆ กระจายอยู่รอบ ๆ โดยทั่วไปเรียกหมู่บ้านเหล่านี้ว่า “หมู่บ้านป่าเมี่ยง”



ภาพที่ 2.2 ลักษณะใบและเมล็ดของต้นชาป่า

ชาป่าหรือชาพันธุ์พื้นเมือง (ภาพที่ 2.1 และภาพที่ 2.2) ถูกจัดอยู่ในกลุ่มสายพันธุ์ชาอัสสัม (*Camellia sinensis* ver. *assamica*) มีแหล่งกำเนิดจากประเทศอินเดีย โดยกระจายตัวตามธรรมชาติบนพื้นที่ภูเขาสูง ส่วนใหญ่ขึ้นในพื้นที่ป่าดิบเขา เจริญเติบโตได้ดีตามป่าที่มีร่มไม้และแสงแดดส่องผ่านพุ่มประมาณ ผลิตรากที่สำคัญที่ได้จากชาป่าในพื้นที่ภาคเหนือของประเทศไทย นอกจากผลิตชาไว้ดื่มแล้ว ยังมีการผลิตเมี่ยง ซึ่งเป็นการนำใบชามาคั่วไว้ระยะหนึ่ง แล้วจึงนำมาบริโภคโดยผสมกับเกลือและน้ำตาล บางท้องถิ่นเรียกว่า “เมี่ยงอม” และเรียกต้นชาป่าว่า “ต้นเมี่ยง” ส่วนพื้นที่ที่ปลูกต้นเมี่ยง เรียกว่า “ป่าเมี่ยง” เนื่องจากลักษณะสวนมีลักษณะเป็นป่า โดยป่าเมี่ยงจัดเป็นระบบวนเกษตรดั้งเดิมที่สำคัญ โดยมีหมู่บ้านเล็ก ๆ กระจายอยู่รอบ ๆ โดยทั่วไปเรียกหมู่บ้านเหล่านี้ว่า “หมู่บ้านป่าเมี่ยง”

จากการสำรวจชาป่าบริเวณเทือกเขาฝิ่นน้ำ (พรชัย ปรีชาปัญญา, 2536) พบต้นชาป่ามีความสูงสุดถึง 20 เมตร มีเส้นผ่าศูนย์กลางที่ระดับอกประมาณ 30 เซนติเมตร คาดว่ามีอายุประมาณ 200-300 ปี และจากการสำรวจของคณะทำงานโครงการหลวงวิจัยชา (สุกนารถ เกตุเจริญ และ

อัญชลี พัดมีเทศ, ม.ป.ป.) ได้พบแหล่งชาป่าที่บ้านไม้สูง อำเภอปางมะผ้า จังหวัดแม่ฮ่องสอน บริเวณติดต่อชายแดนประเทศพม่า ต้นชาป่าที่พบมีอายุหลายร้อยปี เส้นผ่าศูนย์กลางลำต้นประมาณ 50 เซนติเมตร ชาวบ้านละแวกนั้นเรียกต้นชาพันธุ์นี้

2.1.2 ความสำคัญของชาป่า

ชาป่ามีความสำคัญในแง่ของการอนุรักษ์พื้นที่และขนบธรรมเนียมการปฏิบัติของประชาชนในท้องถิ่น รวมทั้งการบริหารจัดการชาป่าในรูปแบบของป่าเมือง ซึ่งถือว่าเป็นระบบวนเกษตรที่มีความสำคัญ โดยเป็นระบบที่ผสมผสานระหว่างป่าไม้ เกษตร และเลี้ยงสัตว์ สุรจิต ภูภักดี (2549) อธิบายระบบของวนเกษตรว่า เป็นระบบที่มีการเกื้อกูลกันระหว่างคนกับไม้ยืนต้นอายุยาว ภายใต้สิ่งแวดล้อมในการดำรงชีพเดียวกัน โดยมีไม้ยืนต้นเป็นองค์ประกอบหลัก ซึ่งผลประโยชน์ที่ได้จากระบบนี้อาจออกมาในรูปแบบของรายได้ อาหาร หรือพลังงาน โดยให้มิผลกระทบต่อสิ่งแวดล้อมน้อยที่สุด ดังนั้นการบริหารจัดการชาป่าในรูปแบบของป่าเมืองจึงเป็นการรักษาสมดุลของสิ่งแวดล้อมมากกว่าการผลิตสินค้า ป่าเมืองจึงถูกจัดว่าเป็นระบบที่มีความหลากหลายทางชีวภาพมากที่สุดแห่งหนึ่ง จากการสำรวจตามการใช้ประโยชน์ที่ดินบริเวณป่าเมืองในพื้นที่ลุ่มแม่น้ำตอนหลวง ตำบลเทพสะเค็ด อำเภอดอยสะเก็ด จังหวัดเชียงใหม่ พบว่ามีความหลากหลายทางชีวภาพมากมาย โดยเฉพาะพืชที่เป็นประโยชน์ต่อการดำรงชีวิต รวมทั้งพืชอาหารที่ขึ้นเองหรือเกษตรกรเป็นคนปลูก อีกทั้งไม้ใหญ่ที่ปล่อยให้ขึ้นเป็นร่มเงาแก่ต้นชาป่า และเป็นไม้ฟืน ไม้ฟืนหรือเป็นไม้แปรรูปเพื่อใช้ในการสร้างบ้านและเครื่องเรือน ตลอดจนพืชสมุนไพรต่าง ๆ (Preechapanya, 1996)

จากการสำรวจพื้นที่ป่าเมืองในประเทศไทยพบว่า ป่าเมืองมีพื้นที่ครอบคลุมราว 170,000 ตารางกิโลเมตร โดยพื้นที่ประมาณร้อยละ 45 อยู่ในพื้นที่ภูเขาสูง (Pagella et al., 2002) และจากการสำรวจป่าเมืองในเขตพื้นที่ดอยยังพบว่า พื้นที่ป่าเมืองส่วนใหญ่อยู่ในเขตลุ่มน้ำชั้นที่ 1A (พรชัย ปรีชาปัญญา, 2536) ซึ่งสำนักอนุรักษ์และจัดการต้นน้ำ กรมอุทยานแห่งชาติ สัตว์ป่า และพันธุ์พืช (ม.ป.ป.) อธิบายความหมายของเขตลุ่มน้ำ 1A ว่า เป็นพื้นที่ต้นน้ำลำธารที่ยังมีสภาพป่าสมบูรณ์ ในปี พ.ศ. 2525 สำหรับลุ่มน้ำปิง วัง ยม น่าน ชี มูล และลุ่มน้ำภาคใต้ ปี พ.ศ. 2528 สำหรับลุ่มน้ำภาคตะวันออก และปี พ.ศ. 2531 สำหรับลุ่มน้ำตะวันตก ภาคกลาง ลุ่มน้ำป่าสัก ลุ่มน้ำภาคเหนือและตะวันออกเฉียงเหนือ และส่วนอื่น ๆ (ลุ่มน้ำชายแดน) ซึ่งห้ามไม่ให้มีการเปลี่ยนแปลงลักษณะพื้นที่ป่าไม้เป็นรูปแบบอย่างอื่นอย่างเด็ดขาดในทุกกรณี

การศึกษาโครงสร้างและองค์ประกอบของป่าเมืองที่บ้านปางกำแพงหิน ตำบลเทพสะเค็ด อำเภอดอยสะเก็ด จังหวัดเชียงใหม่ (พรชัย ปรีชาปัญญา, 2536) พบการใช้ที่ดินของหมู่บ้านอยู่ 4 ส่วนด้วยกันคือ ส่วนที่ 1 เป็นส่วนของที่อยู่อาศัย ซึ่งอยู่บริเวณที่ราบเขาแคบ ๆ ตั้งกันเป็นกลุ่มตาม

ลำน้ำ ส่วนที่ 2 เป็น สวนหลังบ้าน ซึ่งอยู่บนพื้นที่ลาดชันหลังบ้าน โดยพืชในสวนหลังบ้านส่วนใหญ่เป็นพืชใช้สอย เช่น ต้นชาป่า ไม้ผล พืชอาหาร เป็นต้น ส่วนที่ 3 เป็นป่าเมี่ยงที่อยู่ในบริเวณที่ลาดชันสูง อันประกอบไปด้วย ต้นชาป่า ไม้พื้น ไม้ใช้สอย และส่วนที่ 4 เป็นส่วนของป่าต้นน้ำ ลำธารและป่าใช้สอย ซึ่งสอดคล้องกับการศึกษาการอนุรักษ์ป่าควบคู่ไปกับป่าเมี่ยงที่หมู่บ้านศรีนาปาน-ตาแวน ตำบลเรือง อำเภอเมือง จังหวัดน่าน (บัวทอง จิณะหล้า, 2551) พบว่า ชาวบ้านมีการแบ่งพื้นที่ป่าออกเป็น 5 ส่วน คือ ส่วนที่ 1 เป็นป่าต้นน้ำ ถือเป็นเขตอนุรักษ์เพราะเป็นต้นกำเนิดลำห้วยสำคัญ ๆ คือลำห้วยหลวง ห้วยตาแวน ห้วยวะ ซึ่งใช้สำหรับการบริโภค อุปโภค และการเกษตร ส่วนห้วยน้ำดังเป็นแหล่งน้ำสำหรับผลิตประปาภูเขา ในเขตป่าต้นน้ำ มีการห้ามเข้าไปใช้ประโยชน์โดยเด็ดขาด ส่วนที่ 2 เป็นป่าเศรษฐกิจหรือป่าเมี่ยง เป็นป่าเมี่ยงที่อยู่ใต้ร่มเงาของต้นไม้ใหญ่ โดยอยู่ระหว่างเขตป่าอนุรักษ์กับป่าใช้สอย ส่วนที่ 3 เป็นป่าใช้สอย ส่วนใหญ่จะเป็นพื้นที่ปฏิรูปที่ดินเพื่อเกษตรกรรม (สปก.) จะเป็นพื้นที่ไม้เบญจพรรณและพื้นที่ทำการเกษตร ส่วนที่ 4 เป็นป่าชุมชนเป็นป่าที่สงวนไว้สำหรับนำไม้มาใช้ในกิจกรรมของหมู่บ้าน และส่วนที่ 5 เป็นป่าพิธีกรรม หรือป่าประเพณี เช่น ป่าสุสาน ป่าวัดร้าง และป่าห่อเจ้าหลวงป่าเมี่ยง ดังนั้นป่าเมี่ยงจึงถือว่าเป็นพื้นที่แหล่งต้นน้ำและทรัพยากรธรรมชาติที่สำคัญของชาวป่าเมี่ยง อีกทั้งยังเป็นพื้นที่ที่เกื้อหนุนต่อการดำรงชีพอีกด้วย

จากการสำรวจหมู่บ้านป่าเมี่ยงพบว่า ส่วนใหญ่จะกระจายอยู่ล้อมรอบเทือกเขาฝิปันน้ำ ซึ่งเป็นแหล่งต้นน้ำที่สำคัญของไทย อาทิเช่น แม่น้ำวัง แม่น้ำยม แม่น้ำน่าน เป็นต้น โดยอยู่บริเวณรอยต่อระหว่างจังหวัดเชียงใหม่ เชียงราย พะเยา ลำปาง และแพร่ (Preechapanya, 2002) และมีชาวป่าเมี่ยงเป็นผู้ใช้ประโยชน์จากพื้นที่ป่าเมี่ยง ดังนั้นหากพบเห็นการทำลายป่าหรือไฟป่า ก็จะช่วยกันยับยั้งทันที เนื่องจากภัยพิบัติที่เกิดขึ้นจะส่งผลโดยตรงต่อชุมชน ชาวป่าเมี่ยงจะคอยถางวัชพืชและทำแนวกันไฟป่าในฤดูแล้ง ทำให้ป่าเมี่ยงเปรียบเสมือนแนวกันไฟไปในตัว เช่นเดียวกับชาวป่าเมี่ยงที่หมู่บ้านศรีนาปาน-ตาแวน ตำบลเรือง อำเภอเมือง จังหวัดน่าน (บัวทอง จิณะหล้า, 2551) ได้มีการรวมกลุ่มกันภายใต้ชื่อ “เรื่องรักษ์ป่า” โดยได้ร่วมกันบวชป่า ทำฝายแม้ว และเฝ้าระวังไฟป่า รวมทั้งตั้งกฎกติการ่วมกันในการใช้ทรัพยากรจากป่า โดยชาวบ้านมีแนวคิดในการอนุรักษ์ที่ว่า “เมี่ยงมี ป่า ยัง คนอยู่” กล่าวคือ เมี่ยงมีความสำคัญกับชีวิต ถ้าไม่รักษาป่า เมี่ยงก็ไม่อยู่ เงินก็ไม่มี และถ้าหากบริเวณเทือกเขาฝิปันน้ำไม่มีหมู่บ้านป่าเมี่ยง ป่าไม้บนเทือกเขาฝิปันน้ำคงถูกบุกรุกเพื่อทำไร่ผืนในช่วงปี พ.ศ. 2510-2520 ไปจนหมด เหมือนที่พบร่องรอยที่เป็นไร่ร้างบริเวณเทือกเขาถนนธงชัย หรืออาจถูกบุกรุกทำไร่ข้าวโพดดังที่พบเห็นในกลุ่มน้ำน่าน (Preechapanya, 2002) ป่าเมี่ยงจึงถือว่าเป็นแหล่งพื้นที่กันชนที่ป้องกันแหล่งต้นน้ำจากไฟป่า และการบุกรุกทำลายพื้นที่ป่าเพื่อการเกษตร

2.1.3 ปัจจัยที่ส่งผลกระทบต่อต้นชา

แยกพิจารณาเป็น 2 กลุ่มใหญ่ ๆ คือ ปัจจัยทางด้านสภาพภูมิอากาศและปัจจัยทางด้านสภาพภูมิศาสตร์กายภาพ โดยปัจจัยทางด้านสภาพภูมิอากาศ ได้แก่ ปริมาณน้ำฝน อุณหภูมิ เป็นต้น ในขณะที่ปัจจัยทางด้านสภาพภูมิศาสตร์กายภาพ ได้แก่ ระดับความสูงจากระดับน้ำทะเล ความลาดชัน และดิน เป็นต้น

2.1.3.1 ปริมาณน้ำฝน (Rainfall)

ปริมาณน้ำฝนเป็นปัจจัยสำคัญอย่างหนึ่งที่มีผลต่อการเจริญเติบโตของต้นชา จากรายงานพืชเศรษฐกิจพิเศษสำหรับวนเกษตรบนเกาะแปซิฟิกโดย Braber et al. (2011) ได้อธิบายว่าต้นชาสามารถเจริญเติบโตได้ดีที่สุดภายใต้ปริมาณน้ำฝนสูงและมีการกระจายอย่างสม่ำเสมอ ในเขตร้อนจะต้องมีปริมาณน้ำฝนไม่น้อยกว่า 1,500 มิลลิเมตรต่อปี มีฤดูแล้งน้อยกว่า 3 เดือน และอาจมีชลประทานเสริม ปริมาณน้ำฝนสูงสุดไม่ควรเกิน 3,000 มิลลิเมตรต่อปี แต่ในบางพื้นที่ของศรีลังกาชาสามารถเจริญเติบโตได้ดีในพื้นที่ที่มีปริมาณน้ำฝนมากกว่า 5,000 มิลลิเมตรต่อปี ซึ่งตรงกับ Wijeratne et al. (2007) ได้ประเมินผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อผลผลิตชาในศรีลังกา กล่าวว่า ปริมาณน้ำฝนที่เหมาะสมสำหรับการปลูกชาอยู่ระหว่าง 223 มิลลิเมตรต่อเดือนถึง 417 มิลลิเมตรต่อเดือน หรือประมาณ 2,676 มิลลิเมตรต่อปีถึง 5,004 มิลลิเมตรต่อปี นอกจากนี้ยังกล่าวเพิ่มเติมอีกว่า ปริมาณน้ำฝนที่ลดลง 100 มิลลิเมตรต่อเดือนส่งผลให้การผลิตชาลดลง 30-80 กิโลกรัมต่อเฮกตาร์ (30-80 กิโลกรัมต่อ 10,000 ตารางเมตร)

นอกจากนี้ Tea Research Association (n.d.) ได้ศึกษาพื้นที่ปลูกชาทางภาคตะวันออกเฉียงเหนือของอินเดียพบว่า มีปริมาณน้ำฝนโดยเฉลี่ย 2,000-4,000 มิลลิเมตรต่อปี ซึ่งถือว่าเพียงพอต่อการปลูกชา แต่อย่างไรก็ตามกลับพบว่า การกระจายของปริมาณน้ำฝนที่เกิดขึ้นทั้งฤดูกาลมีผลต่อผลผลิตชาด้วย โดยปริมาณน้ำฝนส่วนเกินในช่วงมรสุมของเดือนมิถุนายนถึงกันยายนจะทำให้เกิดปัญหาการระบายน้ำ แต่ในช่วงเดือนพฤศจิกายนถึงมีนาคม ปริมาณน้ำฝนกลับน้อยกว่าปริมาณน้ำที่สูญเสียไปจากการระเหย ทำให้ดินขาดความชุ่มชื้นซึ่งจะส่งผลกระทบต่อทรงพุ่มของต้นชา ดังนั้นปริมาณน้ำฝนในช่วงฤดูใบไม้ผลิและฤดูหนาวจึงเป็นตัวแปรสำคัญในการเพิ่มผลผลิต สุคนารถ เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) ได้แนะนำว่าพื้นที่ที่ปลูกชาควรเป็นพื้นที่ที่มีปริมาณน้ำฝนสม่ำเสมอ ปริมาณน้ำฝนอย่างต่ำควรอยู่ในช่วง 1,140-1,270 มิลลิเมตรต่อปี เนื่องจากถ้าต้นชาขาดน้ำแล้วจะทำให้การเจริญเติบโตของต้นชาหยุดชะงัก ไม่มีการแตกยอด ซึ่งส่งผลให้การผลิตชาลดลงด้วย

2.1.3.2 อุณหภูมิ (Temperature)

อุณหภูมิถือว่าเป็นอีกปัจจัยหนึ่งที่มีความสำคัญต่อการการผลิตชา โดยมีผลต่ออัตรา การสังเคราะห์แสง การเจริญเติบโตและการพักตัว โดยอุณหภูมิที่สูงกว่า 32°C ไม่เหมาะสมต่อ กระบวนการสังเคราะห์แสง และอุณหภูมิที่ต่ำกว่า 12°C จะทำให้การเจริญเติบโตของต้นชา หยุดชะงักและทำให้ผลผลิตต่ำ (Tea Research Association, n.d.) ซึ่ง Wijeratne et al. (2007) ได้ รายงานว่า อุณหภูมิที่เหมาะสมที่สุดต่อการปลูกชาในศรีลังกามากที่สุดคือ 22°C ในขณะที่ Braber et al. (2011) ได้อธิบายว่า อุณหภูมิที่เหมาะสมสำหรับการเจริญเติบโตของต้นชาอยู่ในช่วง $18 - 30^{\circ}\text{C}$ และการเจริญเติบโตจะหยุดชะงักเมื่อมีอุณหภูมิสูงสุดสูงกว่า $32-35^{\circ}\text{C}$ และมีอุณหภูมิต่ำกว่า $12-13^{\circ}\text{C}$ ซึ่งสอดคล้องกับรายงานของ Carne and Balerdi (2005) ศุภนารถ เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) ได้กล่าวเพิ่มเติมว่า ชาสามารถเจริญเติบโตได้ในอุณหภูมิที่แตกต่างกัน โดยชาจะ เจริญเติบโตได้ดีที่สุดในช่วง $25-30^{\circ}\text{C}$ ในพื้นที่อุณหภูมิค่อนข้างคงที่ตลอดปี ซึ่งจะทำให้ชาสร้างยอด ใหม่ได้อย่างต่อเนื่อง

2.1.3.5 ระดับความสูงจากระดับน้ำทะเล (Elevation)

ศุภนารถ เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) ได้กล่าวไว้ว่าใบชาที่ปลูกในพื้นที่สูง ตั้งแต่ 1,000 เมตรจากระดับน้ำทะเลขึ้นไป จะมีกลิ่นและรสชาติที่ดีกว่าใบชาที่ปลูกในที่ต่ำ แต่ ในทางกลับกันต้นชาที่ปลูกในพื้นที่ต่ำกลับให้ผลผลิตที่ดีกว่าต้นชาที่ปลูกในพื้นที่สูง เนื่องจาก อุณหภูมิต่ำบนพื้นที่สูงทำให้การเจริญเติบโตชะลอตัว ซึ่ง Braber et al. (2011) ได้อธิบายเพิ่มเติมว่า ชาหลายชนิดที่มีคุณภาพดีจะพบได้ในพื้นที่สูงซึ่งมีปริมาณน้ำฝนน้อยกว่า 2,000 มิลลิเมตรต่อปี ทำให้ต้นชาในพื้นที่เหล่านี้เจริญเติบโตช้าลง แต่ส่งผลให้รสชาติของชาดีขึ้น นอกจากนี้ Wijeratne et al. (2007) ได้รายงานชาที่ปลูกในระดับความสูงต่ำและปานกลางจะอ่อนแอกว่าต้นชาที่ปลูกใน พื้นที่สูง ซึ่งสอดคล้องกับการสำรวจชาป่าโดย พรชัย ปรีชาปัญญา (2536) พบว่า ชาป่าที่ขึ้นเองตาม ธรรมชาติส่วนใหญ่พบอยู่ในระดับความสูงประมาณ 1,000-1,200 เมตรจากระดับน้ำทะเลโดยเป็น พื้นที่ป่าธรรมชาติ จึงสันนิษฐานได้ว่าต้นชาที่อยู่ในพื้นที่สูงมีโอกาสอยู่รอดในธรรมชาติได้ดีกว่า ต้นชาที่ปลูกในพื้นที่ต่ำ

2.1.3.4 ความลาดชัน (Slope)

Braber et al. (2011) กล่าวถึงความลาดชันที่มีผลกระทบต่อต้นชาว่า ต้องมีความชันไม่ เกิน 30 องศา และต้องมีการจัดการเป็นพิเศษเพื่อป้องกันการชะพังทลายของหน้าดินโดยเฉพาะ อย่างยิ่งต้นชาที่ยังไม่แก่ เนื่องจากระบบรากของต้นชาเหล่านี้กำลังอยู่ในช่วงพัฒนา การยึดเกาะของ รากยังไม่เต็มที่ นอกจากนี้ผลการพังทลายของดินอย่างรุนแรงยังส่งผลให้แนวโน้มความเป็นเบส และธาตุฟอสฟอรัสต่ำ รวมทั้งส่งผลถึงความเข้มข้นของไนโตรเจนในดิน ในขณะที่ศุภนารถ

เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) ได้ให้ความเห็นว่า พื้นที่ปลูกชาควรมีความลาดชันไม่เกิน 45 องศา

2.1.3.5 ดิน (Soil)

ดินเป็นปัจจัยหลักที่มีความสำคัญต่อการเจริญเติบโตของพืชทุกชนิด ดันชาก็เช่นเดียวกัน โดย Braber et al. (2011) ได้อธิบายว่า เนื้อดินที่มีอินทรีย์วัตถุสูงเป็นหัวใจหลักต่อการปลูกดันชา โดยดินที่เหมาะสมต่อดันชาอยู่ลึก 1.5-2 เมตร แต่ในบางพื้นที่ชาสามารถเจริญเติบโตได้ดีที่ความลึก 60-100 เซนติเมตร เช่น ชาที่ปลูกในเวียดนาม และมีลักษณะทางเคมีที่สำคัญคือมีคุณสมบัติเป็นกรดโดยมีช่วงค่าความเป็นกรด (pH) ที่เหมาะสมอยู่ระหว่าง 6.0-6.5 แต่โดยทั่วไปจะอยู่ระหว่าง 4.4-5.5 ซึ่งสอดคล้องกับ Tea Research Association (n.d.) ที่กล่าวว่าดินที่เหมาะสมต่อการเจริญเติบโตของดันชาควรมีค่า pH อยู่ระหว่าง 4.5-5.5 และมีอินทรีย์วัตถุในดินมากกว่าร้อยละ 2 ในขณะที่สุภนารถ เกตุเจริญ และ อัญชลี พัดมีเทศ (ม.ป.ป.) กล่าวว่า พื้นที่ที่ไม่มีความเหมาะสมต่อการปลูกชาคือพื้นที่ที่ดินมีหินปูน ซึ่งแสดงว่าดินมีความเป็นด่างหรือพื้นที่ที่ดินมีค่า pH เกินกว่า 6 โดยเสนอว่าค่า pH ของดินที่มีความเหมาะสมควรอยู่ในช่วง 4.5-6.0 ซึ่งสอดคล้องกับคำแนะนำในคู่มือการผลิตชาของกระทรวงเกษตรของประเทศอูกันดา (Ministry of Agriculture, The Republic of Uganda, 2007)

2.1.3.6 ปัจจัยอื่น ๆ

นอกจากปัจจัยหลัก ๆ ที่ได้กล่าวมาข้างต้นแล้ว ชายังมีปัจจัยอื่น ๆ ที่ส่งผลกระทบต่อดันชา โดย Preechapanya (2002) กล่าวว่า บริเวณป่าเมี่ยงที่มีดันชา ชาวป่าเมี่ยงจะทำลายกอไผ่เนื่องจากระบบรากของไผ่มีสารยับยั้งการเจริญเติบโตของรากพืชชนิดอื่น รวมถึงรากของดันชาด้วย ซึ่งพบได้ทั้งหญ้าคาและหญ้าไม้งวด Preechapanya (2002, quoted in McClure, 1966)

นอกจากนี้การปลูกดันชาไม้ใหญ่เพื่อให้ร่มเงาแก่ดันชาซึ่งส่งผลดีต่อการเจริญเติบโตของดันชาด้วย โดย Ranganathan (n.d.) ได้อธิบายว่า ร่มเงาจากต้นไม้อย่างจะช่วยกรองแสงแดดโดยยอมให้แสงแดดผ่าน 60-70% และช่วยลดความเสียหายจากความเข้มแสงที่มากเกินไป ซึ่งสอดคล้องกับคำกล่าวของ Preechapanya (2002) ที่อธิบายว่า ต้นไม้อย่างจะช่วยทำให้ใบของดันชาไม่แก่เร็วเนื่องจากความเข้มของแสงแดด และอาจทำให้ใบเป็นโรคใบจุด อีกทั้งทรงพุ่มของต้นไม้อย่างยังป้องกันน้ำค้างแข็งที่มาจับใบชาในฤดูหนาว ซึ่งเป็นสาเหตุที่ทำให้เกิดโรคใบจุดได้เช่นเดียวกัน นอกจากนี้ร่มเงาจากต้นไม้อย่างยังช่วยป้องกันความชื้นในดิน ทำให้น้ำในดินเพียงพอต่อการเจริญเติบโตของดันชา อีกทั้งยังควบคุมการเจริญเติบโตของหญ้าคาซึ่งทำให้การเจริญเติบโตของดันชาลดลง โดยปลายรากของหญ้าคาจะมีสารยับยั้งการเจริญเติบโตของรากพืชชนิดอื่นเหมือนดันชา และยังยับยั้งการเจริญเติบโตของจุลินทรีย์ที่จำเป็น ทำให้กระบวนการย่อยเศษซากพืชช้าลง

ส่งผลให้แร่ธาตุจากการย่อยสลายอยู่ในสภาพที่ไม่พร้อมใช้งาน และในบริเวณที่มีร่มเงาหนาแน่นยังทำให้เฟิร์นหลายชนิดเจริญเติบโตได้ดี ซึ่งพืชเหล่านี้จะช่วยรักษาความชื้นในดินและปรับปรุงโครงสร้างของดิน ทำให้ดินมีธาตุอาหารเพียงพอซึ่งเกิดจากการทิ้งใบของพืชเหล่านี้ Braber et al. (2011) ยังได้อธิบายเพิ่มเติมว่า นอกจากต้นไม้อายุจะช่วยกรองแสงแล้ว ยังช่วยป้องกันการพังทลายของหน้าดิน ช่วยควบคุมวัชพืช และช่วยเพิ่มธาตุอาหารในดินจากปุ๋ยพืชสดจากใบไม้ที่ร่วงหล่น แต่ต้นไม้อายุต้องเป็นต้นไม้ที่มีระบบรากลึก เพื่อจะได้ไม่แย่งอาหารจากต้นชา เช่นต้น โอ๊คเงิน (Silver Oak: *Grevillea robusta*) ซึ่งใช้กันอย่างแพร่หลายในอินเดียและแอฟริกา

จากการศึกษาบทบาทของไม้พืชร่มต่อการงอกของเมล็ดไม้และการรอดตายของกล้าไม้โดย Preechapanya (2002) พบว่า พืชดั้งเดิมในพื้นที่จะเป็นตัวชี้วัดลักษณะนิเวศ เช่นลักษณะภูมิอากาศ ลักษณะอากาศใกล้ผิวดิน หรือลักษณะดินของพื้นที่ที่เหมาะสมต่อการเจริญเติบโตของต้นชา เช่นที่เคนย่า ชาวสวนชาจะเลือกพื้นที่ปลูกชาที่มี *Pteridium aquilinum* (Preechapanya, 2002, quoted in Brown, 1966) ซึ่งเป็นเฟิร์นชนิดเดียวกับที่พบในพื้นที่ป่าเมี่ยงในประเทศไทย

2.2 แบบจำลองการกระจายของสิ่งมีชีวิต (Species Distribution Models: SDMs)

วัตถุประสงค์หลักโดยทั่วไปของการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต (Species Distribution Models: SDMs) คือการบ่งบอกลักษณะปัจจัยทางสิ่งแวดล้อมที่เหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตเป้าหมายและสามารถระบุพื้นที่ซึ่งมีปัจจัยทางสิ่งแวดล้อมที่เหมาะสมได้ (Pearson, 2007) โดยการนำข้อมูลในส่วนหนึ่งของระบบสารสนเทศภูมิศาสตร์ (Geographic Information System: GIS) และข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่ (Species Occurrence Data) หรือข้อมูลการทำงานของลักษณะเฉพาะในสิ่งมีชีวิตที่กำลังศึกษา (Functional Traits Data) มาประมวลผลโดยวิธีการทางคณิตศาสตร์และสถิติ เพื่อพยากรณ์ขอบเขตพื้นที่การกระจายของสิ่งมีชีวิตที่กำลังศึกษา ลักษณะวิธีการของ SDMs สามารถแยกพิจารณาได้ 2 แบบคือ วิธีการศึกษาเชิงความสัมพันธ์ (Correlative Approach) และ วิธีการศึกษาเชิงกลไก (Mechanistic Approach) (Beerling, Huntley & Bailey, 1995)

วิธีการศึกษาเชิงความสัมพันธ์จะพิจารณาความสัมพันธ์ระหว่างข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่กับข้อมูลในส่วนหนึ่งของระบบสารสนเทศภูมิศาสตร์ ซึ่งเป็นปัจจัยทางสิ่งแวดล้อม โดยใช้วิธีการทางสถิติ เพื่อดูความสัมพันธ์ว่าส่งผลกระทบต่อปรากฏตัวของสิ่งมีชีวิตในพื้นที่อย่างไร Robertson, Peter, Villet and Ripley (2003) ได้อธิบายว่า วิธีการศึกษาเชิงความสัมพันธ์จะใช้ค่าของตัวแปรอิสระ (Independence Variables) หรือตัวแปรพยากรณ์ (Predicting Variables)

ที่เกี่ยวข้องกับข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่ เพื่อจำแนกตัวแปรอิสระลงในพื้นที่ที่ปรากฏ สิ่งมีชีวิตเป้าหมายและพื้นที่ที่ไม่ปรากฏสิ่งมีชีวิตเป้าหมาย ซึ่งอาจอยู่ในรูปแบบของค่าความเหมาะสม (Suitability Values) หรือค่าความน่าจะเป็น (Probability) โดยจะแสดงออกในรูปแบบของแผนที่

ในขณะที่วิธีการศึกษาเชิงกลไกเป็นการนำข้อมูลการทำงานของลักษณะเฉพาะในสิ่งมีชีวิตเป้าหมายเข้ามาเชื่อมกับข้อมูลในส่วนของระบบสารสนเทศภูมิศาสตร์ โดยจะพิจารณาถึงกระบวนการที่สำคัญของสิ่งมีชีวิตที่ตอบสนองต่อสภาพแวดล้อมที่เป็นแหล่งที่อยู่อาศัย Beerling et al. (1995) ได้อธิบายว่า วิธีการศึกษาเชิงกลไกจะพยายามจำลองกลไกพื้นฐานของสิ่งมีชีวิตที่มีความสัมพันธ์กับคุณลักษณะด้านสิ่งแวดล้อม โดยใช้ความรู้เกี่ยวกับการตอบสนองของสิ่งมีชีวิตที่กำลังศึกษาต่อสภาพสิ่งแวดล้อม ซึ่งจะศึกษาจากชีวประวัติของสิ่งมีชีวิตเป้าหมาย

Stepenson (1998) ได้ให้ความคิดเห็นว่าความแตกต่างระหว่างวิธีการศึกษาเชิงความสัมพันธ์และวิธีการศึกษาเชิงกลไกนั้นยังไม่ชัดเจน โดยเขาได้สังเกตว่า การศึกษาทางด้านกลไกของสิ่งมีชีวิต ขึ้นอยู่กับการทดสอบความสัมพันธ์เชิงปริมาณระหว่างปัจจัยที่สำคัญทางชีววิทยาและการกระจายของพืช ในทำนองเดียวกันวิธีการศึกษาเชิงความสัมพันธ์จะใช้พื้นฐานทางกลไกของสิ่งมีชีวิตที่ต่อเมื่อสงสัยว่าตัวแปรพยากรณ์นั้นมีความสำคัญทางสรีระวิทยาของสิ่งมีชีวิต โดยทั่วไปอย่างไร

Robertson et al. (2003) กล่าวว่าการศึกษาการกระจายของสิ่งมีชีวิตโดยทั่วไปเป็นการพยากรณ์การกระจายของสิ่งมีชีวิตที่คาดว่าปรากฏในพื้นที่ (Potential Distribution) ภายใต้สภาพแวดล้อมในปัจจุบันหรือสถานการณ์การเปลี่ยนแปลงของสภาพแวดล้อมโดย Stepenson (1998) กล่าวว่า วิธีการศึกษาเชิงกลไกมีประสิทธิภาพในการพยากรณ์ภายใต้สภาวะแวดล้อมที่เปลี่ยนแปลงได้ดีกว่าวิธีการศึกษาเชิงความสัมพันธ์ ซึ่งวิธีการศึกษาเชิงกลไกจำเป็นต้องอาศัยองค์ความรู้ทางชีววิทยาของสิ่งมีชีวิตที่กำลังศึกษาเป็นอย่างดี ในขณะที่วิธีการศึกษาเชิงความสัมพันธ์เหมาะกับการประมาณพื้นที่ที่คาดว่าจะมีการกระจายของสิ่งมีชีวิตที่ยังไม่เป็นที่รู้จักหรือมีองค์ความรู้ทางชีววิทยาของสิ่งมีชีวิตเป้าหมายไม่มากพอ นอกจากนี้ Robertson et al. (2003) ยังกล่าวเพิ่มเติมว่า วิธีการศึกษาเชิงความสัมพันธ์ยังสามารถนำไปใช้เพื่อให้ได้ข้อมูลเชิงลึกของปัจจัยซึ่งอาจมีส่วนสำคัญในการจำกัดการกระจายตัวของสิ่งมีชีวิตเป้าหมาย เมื่อข้อมูลทางชีววิทยาของสิ่งมีชีวิตที่เป้าหมายนั้นยังไม่เพียงพอ โดยข้อมูลเชิงลึกเหล่านั้นยังสามารถนำไปใช้ร่วมกับตัวแปรสำคัญทางกลไกต่าง ๆ ของสิ่งมีชีวิต เพื่อนำไปใช้ในการการศึกษาเชิงกลไกเพิ่มเติม

2.2.1 ทฤษฎีของแบบจำลองการกระจายของสิ่งมีชีวิต

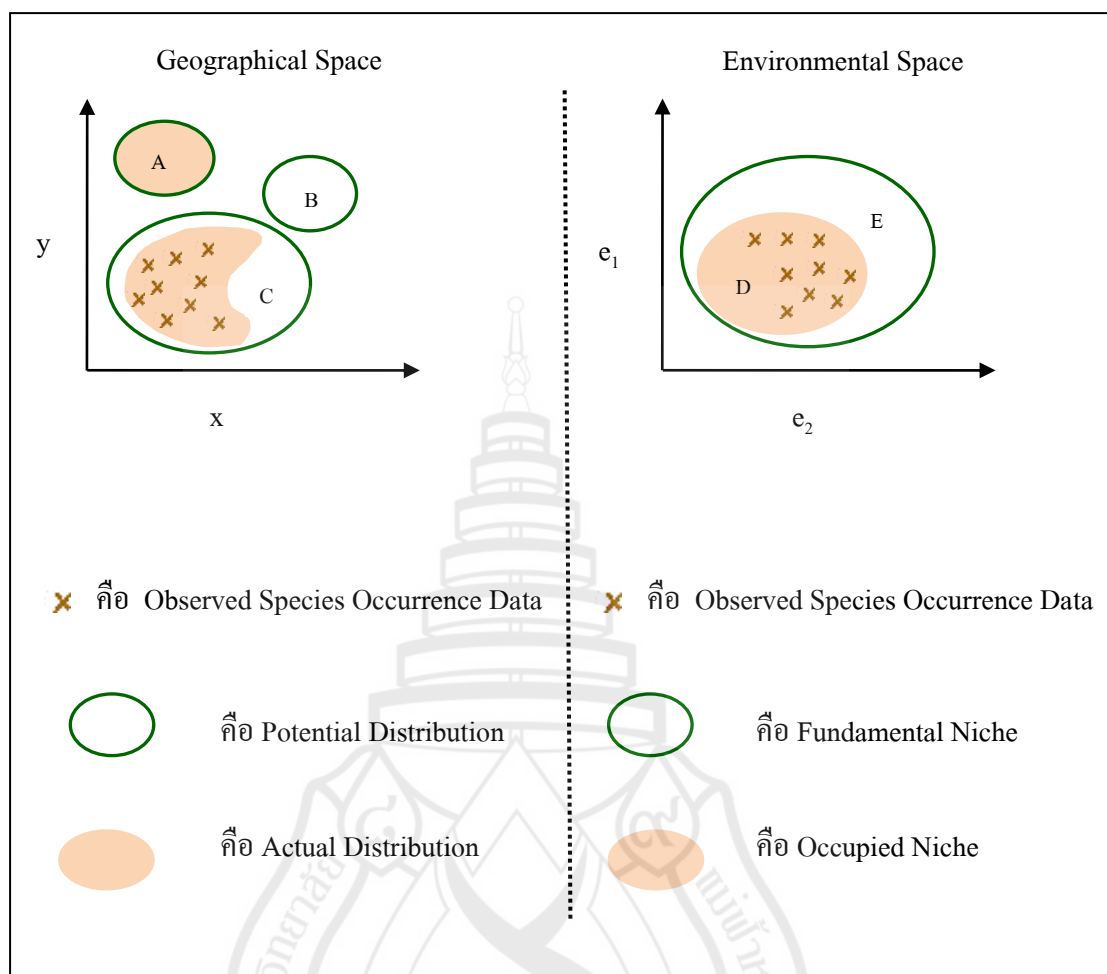
2.2.1.1 พื้นที่เชิงภูมิศาสตร์ (Geographical Space) และพื้นที่เชิงสิ่งแวดล้อม (Environmental Space)

พื้นที่เชิงภูมิศาสตร์ (Geographical Space) จะแสดงการกระจายของสิ่งมีชีวิตที่ปรากฏจริงในพื้นที่ (Actual Distribution) และการกระจายของสิ่งมีชีวิตที่คาดว่าจะปรากฏในพื้นที่ (Potential distribution) โดยใช้แกน XY ในการอ้างอิงการกระจายเชิงพื้นที่ที่ต้องการศึกษา ซึ่งอยู่ในรูปแบบของแผนที่การกระจาย จากภาพที่ 2.3 ในส่วนของพื้นที่เชิงภูมิศาสตร์เส้นทึบจะแสดงถึงขอบเขตของการกระจายของสิ่งมีชีวิตที่คาดว่าจะปรากฏในพื้นที่ ส่วนพื้นที่ที่แรเงา จะแสดงถึงขอบเขตของการกระจายของสิ่งมีชีวิตที่ปรากฏจริงในพื้นที่

จากพื้นที่ B ในภาพที่ 2.3 ถึงแม้ว่าจะเป็นพื้นที่ที่สภาพสิ่งแวดล้อมมีความเหมาะสมต่อสิ่งมีชีวิตเป้าหมาย แต่ก็ไม่ปรากฏสิ่งมีชีวิตอยู่ในพื้นที่ ซึ่งอาจเป็นเพราะสิ่งมีชีวิตไม่สามารถกระจายไปยังพื้นที่นั้นได้ เนื่องจากอาจมีพื้นที่ที่ไม่เหมาะสมต่อการเจริญเติบโตกั้นกลาง

เช่นเดียวกับพื้นที่ที่ไม่ได้แรเงารอบตัว C ซึ่งเป็นพื้นที่ที่เหมาะสมต่อสิ่งมีชีวิตแต่สิ่งมีชีวิตไม่สามารถอยู่อาศัยได้ ซึ่งอาจเป็นเพราะผลกระทบจากสิ่งมีชีวิตชนิดอื่น ๆ ในพื้นที่ ยกตัวอย่างเช่น การศึกษาระบบวนเกษตรชาวพื้นเมืองบนพื้นที่สูงทางภาคเหนือของประเทศไทย (Preechapanya, 2002) ได้อธิบายว่า ในป่าเมืองหรือสวนชาป่าจะไม่พบคันไผ่ ถึงแม้ว่าคันไผ่จะพบได้ในเขตป่าธรรมชาติทั่วไปในพื้นที่ เนื่องจากเกษตรกรหรือผู้ดูแลป่าเมืองมักจะทำลายกอไผ่ที่ขึ้นในสวนเมือง โดยให้เหตุผลว่าระบบรากของไผ่มีสารยับยั้งการเจริญเติบโตของรากของพืชอื่น รวมทั้งต้นชาป่าด้วย ซึ่งสารดังกล่าวยังพบได้ที่ปลายรากของหญ้าคาและหญ้าไม้กวาดอีกด้วย ดังนั้น พื้นที่ที่มีคันไผ่จึงเป็นสาเหตุหนึ่งที่ทำให้ต้นเมืองหรือต้นชาป่าไม่สามารถเจริญเติบโตได้ ถึงแม้ว่าพื้นที่นั้นจะมีความเหมาะสมต่อการเจริญเติบโตของชาป่าก็ตาม

ส่วนพื้นที่ A เป็นพื้นที่ซึ่งมีการปรากฏของสิ่งมีชีวิตอยู่จริงแต่ยังไม่ถูกค้นพบหรืออาจมีการเคลื่อนย้ายหรือถูกทำลายก่อนเข้าไปสำรวจ



ที่มา Pearson (2007)

ภาพที่ 2.3 ความสัมพันธ์ของ Species Distribution ในรูปแบบของพื้นที่เชิงภูมิศาสตร์ (Geographical Space) และพื้นที่เชิงสิ่งแวดล้อม (Environmental Space)

พื้นที่เชิงสิ่งแวดล้อม (Environmental Space) มีแนวคิดโดยการกำหนดตัวแปรทางสิ่งแวดล้อมซึ่งมีผลต่อการตอบสนองของสิ่งมีชีวิต โดยใช้ทฤษฎีลักษณะเฉพาะเชิงนิเวศ (Ecological Niche Theory) ในการอธิบายความสัมพันธ์ ซึ่งจะแสดงความสัมพันธ์เชื่อมโยงของโครงสร้างสิ่งแวดล้อมที่มีลักษณะเฉพาะ โดยจะแสดงลักษณะเฉพาะพื้นฐาน (Fundamental Niche) ที่มีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตและลักษณะเฉพาะของพื้นที่อยู่อาศัย (Occupied Niche) ที่มีการปรากฏสิ่งมีชีวิตนั้น

โดยแนวความคิดเรื่องลักษณะเฉพาะ (Niche) ริเริ่มมาจาก Grinnell (1917) และ Grinnell (1924) ซึ่งเป็นบุคคลแรกอธิบายความสัมพันธ์ของปัจจัยทางสิ่งแวดล้อมที่ส่งผลต่อการดำรงชีวิตอยู่และการเกิดใหม่ของสิ่งมีชีวิต ซึ่งปัจจัยทางสิ่งแวดล้อมมีทั้งปัจจัยที่เกิดจากสิ่งไม่มีชีวิต (Abiotic Factors) เช่น อุณหภูมิ, ปริมาณน้ำฝน และปัจจัยที่เกิดจากปฏิสัมพันธ์ระหว่างกัน (Interaction Factors) ของสิ่งมีชีวิต ในทางกลับกัน Elton (1927) ได้มองว่าลักษณะเฉพาะเป็นบทบาทหนึ่งในระบบนิเวศวิทยาและการเปลี่ยนแปลงภายในระบบนิเวศวิทยา โดยบทบาทดังกล่าวเกิดจากสิ่งมีชีวิตต่างชนิดกันในสถานที่แตกต่างกัน ทฤษฎีลักษณะเฉพาะเชิงนิเวศถูกใช้อย่างเป็นทางการโดย Hutchinson (1957) ซึ่งได้อธิบายความหมายของลักษณะเฉพาะเชิงนิเวศว่าเป็นเซตของปัจจัยทางสิ่งแวดล้อมที่มีผลต่อการดำรงชีวิตอยู่ของสิ่งมีชีวิต โดยกำหนดให้ลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตเป็น ไฮเพอร์วอลุ่มขนาด n มิติ (n -Dimensional Hypervolume) นั่นคือ ทุก ๆ จุดในลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตจะสอดคล้องกับปัจจัยทางสิ่งแวดล้อมที่มีผลต่อการดำรงชีวิตอยู่ของสิ่งมีชีวิต

จากภาพที่ 2.3 ในส่วนของพื้นที่เชิงสิ่งแวดล้อม ถูกกำหนดด้วยตัวแปรทางสิ่งแวดล้อม 2 ตัว คือ e_1 และ e_2 พื้นที่ภายใต้เส้นทึบ จะแสดงถึงขอบเขตของลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโตซึ่งแสดงถึงอาณาเขตของปัจจัยทางสิ่งแวดล้อมที่มีความเหมาะสมต่อสิ่งมีชีวิต ถ้านำเอาเงื่อนไขทางสิ่งแวดล้อมในลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโตมาพล็อตในพื้นที่เชิงภูมิศาสตร์จะเกิดการกระจายของสิ่งมีชีวิตที่คาดว่าปรากฏในพื้นที่นั้น

ส่วนพื้นที่แรเงา D แสดงถึงลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโตโดยมีสิ่งมีชีวิตอาศัยอยู่จริงหรือเป็นสภาวะที่พบว่ามีสิ่งมีชีวิตอาศัยอยู่ ซึ่งเป็นส่วนหนึ่งของลักษณะเฉพาะพื้นฐานที่มีความเหมาะสมต่อการเจริญเติบโต โดยลักษณะเฉพาะที่พบว่ามีสิ่งมีชีวิตอาศัยอยู่จะสะท้อนให้เห็นถึงข้อจำกัดทั้งหมดในการกระจายของสิ่งมีชีวิตที่ปรากฏอยู่จริงในพื้นที่ รวมทั้งข้อจำกัดเชิงพื้นที่อันเนื่องมาจากสิ่งมีชีวิตไม่สามารถกระจายเข้าสู่พื้นที่ได้และปฏิสัมพันธ์กับสิ่งมีชีวิตอื่น ๆ เช่น การปรากฏตัวของคู่แข่งหรือการขาดแหล่งอาหาร สุดท้ายคือการทำลายหรือการเคลื่อนย้ายออกไปโดยฝีมือมนุษย์

2.2.1.2 การพยากรณ์การกระจายของสิ่งมีชีวิต

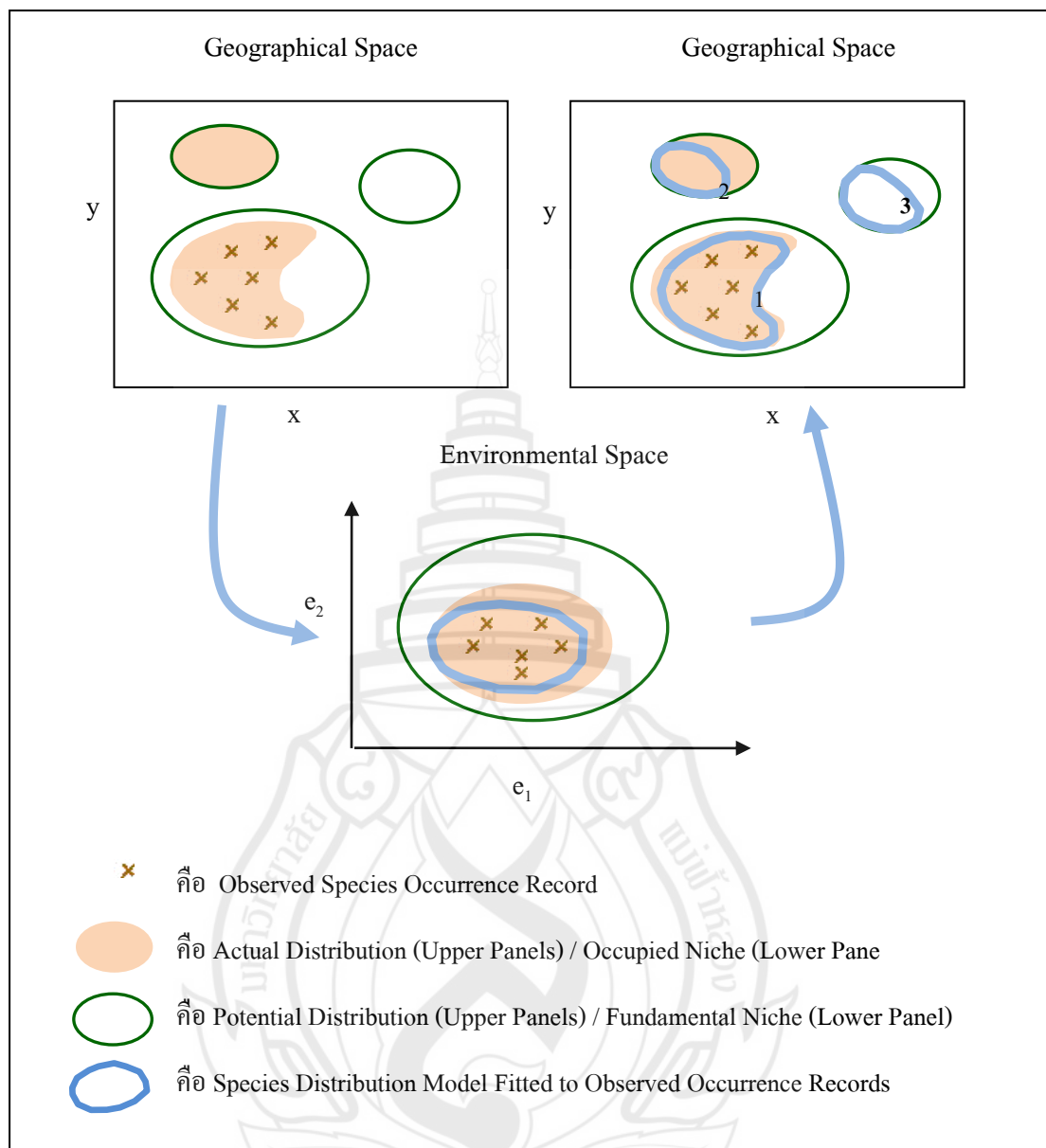
ในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตสิ่งแรกที่ต้องพิจารณาคือวัตถุประสงค์ของการสร้างแบบจำลอง โดยพิจารณาว่าสร้างแบบจำลองเพื่ออะไร และสิ่งที่ต้องการประมาณค่าจากแบบจำลองการกระจายของสิ่งมีชีวิตคืออะไร ถ้าแบบจำลองที่ใช้มีวัตถุประสงค์ในการให้ความสำคัญต่อการอนุรักษ์ในระดับสูง จุดมุ่งหมายหลักที่ต้องการคือการกระจายของสิ่งมีชีวิตที่ปรากฏจริงในพื้นที่ แต่ในทางกลับกันหากวัตถุประสงค์หลักต้องการแสวงหาพื้นที่ที่

เหมาะสมสำหรับการส่งคืนของสิ่งมีชีวิตที่ใกล้สูญพันธุ์ จุดมุ่งหมายหลักที่ต้องการคือการหาการกระจายของสิ่งมีชีวิตที่คาดว่าปรากฏอยู่ในพื้นที่

เทคนิคในการสร้างแบบจำลอง จะถูกนำมาใช้เพื่อต้องการอธิบายลักษณะเฉพาะที่มีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตในพื้นที่เชิงสิ่งแวดล้อม โดยแสดงในรูปแบบของความสัมพันธ์ระหว่างข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่ภายใต้ชุดปัจจัยทางสิ่งแวดล้อม (Pearson, 2007) จากภาพที่ 2.4 จะแสดงวิธีการของแบบจำลองการกระจายตัวของสิ่งมีชีวิตที่สอดคล้องกับข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่ที่ต้องการศึกษา โดยมีจุดที่น่าสังเกตในพื้นที่เชิงสิ่งแวดล้อม คือแบบจำลองไม่สามารถระบุได้ทั้งลักษณะเฉพาะของพื้นที่อยู่อาศัย (Occupied Niche) หรือลักษณะเฉพาะพื้นฐาน (Fundamental Niche) แต่แบบจำลองจะระบุเฉพาะส่วนที่ถูกกำหนดโดยค่าสังเกต เมื่อนำกลับเข้าไปพยากรณ์ในพื้นที่เชิงภูมิศาสตร์ แบบจำลองจะถูกระบุอยู่กับการกระจายทั้ง 2 ประเภทนั่นคือ การกระจายของสิ่งมีชีวิตที่ปรากฏอยู่จริงในพื้นที่และการกระจายของสิ่งมีชีวิตที่คาดว่าปรากฏอยู่ในพื้นที่ ในพื้นที่พยากรณ์ที่ 1 หมายถึงพื้นที่ที่ทราบแน่นอนว่ามีการกระจายของสิ่งมีชีวิตเป้าหมาย

ส่วนในพื้นที่พยากรณ์ที่ 2 หมายถึงพื้นที่บางส่วนของกระจายของสิ่งมีชีวิตที่ปรากฏอยู่จริงในพื้นที่ซึ่งยังไม่ถูกค้นพบในขณะนี้ อันเนื่องมาจากค่าสังเกตไม่ได้ถูกระบุอยู่ในขอบเขตทั้งหมดของสถานะที่พบว่ามีสิ่งมีชีวิตอาศัยอยู่ กล่าวคือ มีการสุ่มตัวอย่างที่ไม่สมบูรณ์ เราได้จากพื้นที่ D ในภาพที่ 2.3 การเพิ่มการสุ่มตัวอย่างในพื้นที่นี้อาจทำให้เกิดการค้นพบสิ่งมีชีวิต ซึ่งจะมีประโยชน์ในการสร้างแนวทางในการสำรวจสิ่งมีชีวิตภาคสนาม นอกจากนี้อาจมีประโยชน์ต่อการวางแผนการอนุรักษ์โดยเฉพาะในพื้นที่ที่ไม่ค่อยรู้จักหรือพื้นที่ที่มีการบุกรุกอย่างรุนแรง

ส่วนในพื้นที่พยากรณ์ที่ 3 เป็นพื้นที่ของการกระจายที่มีศักยภาพแต่ไม่ได้เป็นแหล่งที่อยู่อาศัย ซึ่งอาจเกิดจากขอบเขตทั้งหมดของการกระจายของสิ่งมีชีวิตที่คาดว่ามิอยู่ในพื้นที่ไม่ได้ถูกระบุ เนื่องมาจากการสุ่มตัวอย่างที่ไม่สมบูรณ์ การมีปฏิสัมพันธ์กันของสิ่งมีชีวิตหรือข้อจำกัดในการแพร่กระจายของสิ่งมีชีวิต เราได้จากพื้นที่ D และ E ในรูปที่ 2.3 ขอบเขตนี้อาจระบุถึงพื้นที่ที่เหมาะสมต่อการเกิดใหม่ของสิ่งมีชีวิตหรือเป็นพื้นที่ที่อาจมีการรุกรานของสิ่งมีชีวิต นอกจากนี้ ยังแสดงถึงพื้นที่ที่คาดว่าจะมีการปรากฏสิ่งมีชีวิตใหม่ภายใต้สภาวะสิ่งแวดล้อมที่เหมาะสมต่อการดำรงชีวิต อย่างไรก็ตามการสร้างแบบจำลองในบางครั้งก็ต้องการการพยากรณ์การกระจายของสิ่งมีชีวิตที่ปรากฏอยู่จริงในพื้นที่เช่น การสร้างแบบจำลองที่มีวัตถุประสงค์ในการเลือกพื้นที่ที่มีความสำคัญเป็นอย่างยิ่งต่อการอนุรักษ์ เป็นต้น



ที่มา Pearson (2007)

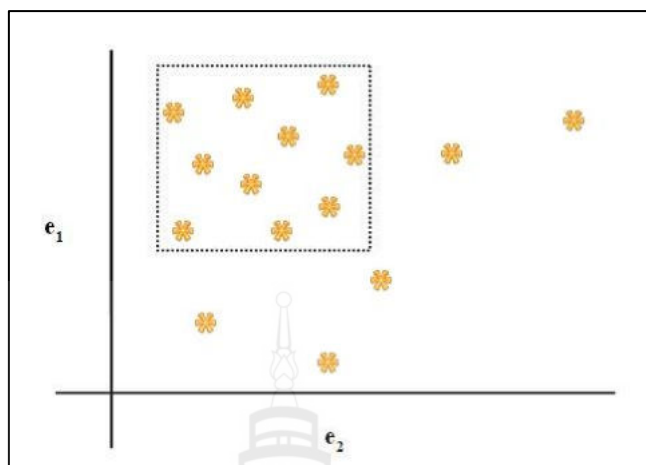
ภาพที่ 2.4 วิธีการของ SDMs ที่สอดคล้องกับข้อมูลการปรากฏของสิ่งมีชีวิตในพื้นที่

2.2.2 การเลือกใช้แบบจำลองการกระจายของสิ่งมีชีวิต

โดยทั่วไปเทคนิคในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตถูกแบ่งออกเป็น 2 กลุ่มใหญ่ ๆ โดยขึ้นอยู่กับข้อมูลการปรากฏของสิ่งมีชีวิต (Species Occurrence Data) นั่นก็คือข้อมูลที่ถูกรวบรวมเก็บไว้ในรูปแบบของจุดพิกัดทางภูมิศาสตร์ (Geographic Coordinates) หรืออาณาเขต (Polygon) ของการปรากฏของสิ่งมีชีวิตในพื้นที่ที่กำลังศึกษา (Drew, Wiersma & Huettmann, 2011) โดยข้อมูลการปรากฏสิ่งมีชีวิตมีอยู่ 2 แบบ คือ ข้อมูลปรากฏ (Presence Data) และข้อมูลไม่ปรากฏ (Absence Data) ซึ่งเทคนิคในการสร้างแบบจำลองที่ต้องการข้อมูลปรากฏเพียงอย่างเดียว (Presence-Only Data) จะเรียกเทคนิคการสร้างแบบจำลองนี้ว่าเทคนิคโปรไฟล์ (Profile Techniques) (Pearce & Boyce, 2006; Stokland, Halvorsen & Støa, 2011; Wisz & Guisan, 2009; Mateo, Croat, Felicísimo & Muñoz, 2009) ในขณะที่เทคนิคในการสร้างแบบจำลองที่ต้องการข้อมูลทั้งข้อมูลปรากฏและข้อมูลไม่ปรากฏ (Presence-Absence Data) จะเรียกเทคนิคการสร้างแบบจำลองนี้ว่าเทคนิคจำแนกกลุ่ม (Group Discrimination Techniques) (Stokland et al., 2011; Wisz & Guisan, 2009; Mateo et al., 2009)

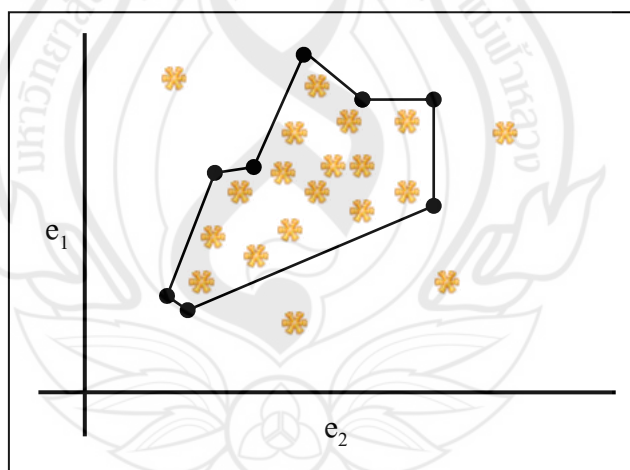
เทคนิคการสร้างแบบจำลองการกระจายสิ่งมีชีวิตโดยเทคนิคโปรไฟล์ จะขึ้นอยู่กับค่าพิสัย (Range), ค่าเฉลี่ย (Means) หรือค่าอื่น ๆ ที่อธิบายตัวแปรสิ่งแวดล้อมในพื้นที่ที่ปรากฏสิ่งมีชีวิตที่ต้องการศึกษา โดยสามารถแยกพิจารณาได้ 2 วิธีคือ วิธี Environmental Envelope และวิธี Environmental Distance (Franklin, 2010)

วิธี Environmental Envelope ยกตัวอย่างเช่น BIOCLIM จะใช้พื้นที่สี่เหลี่ยม (Hyper-Box) ในการจำแนกพื้นที่ที่มีความเหมาะสม ดังภาพที่ 2.5 ซึ่งจำลองเทคนิควิธี BIOCLIM ในการจำลองการกระจายของสิ่งมีชีวิต โดยมีปัจจัยทางสิ่งแวดล้อม 2 ตัว คือ e_1 และ e_2 โดย ✖ จะแสดงว่ามีการปรากฏสิ่งมีชีวิตภายใต้เงื่อนไขของปัจจัยสิ่งแวดล้อม และมีพื้นที่สี่เหลี่ยมแสดงขอบเขตของปัจจัยทางสิ่งแวดล้อมที่เหมาะสม ซึ่งถูกกำหนดโดยค่าสูงสุดและค่าต่ำสุดของปัจจัยทางสิ่งแวดล้อม หรือ 95% ของค่าปัจจัยทางสิ่งแวดล้อม (Franklin, 2010) ในขณะที่วิธี Environmental Distance ยกตัวอย่างเช่น DOMAIN จะใช้ความคล้ายคลึงกันจากจุดหนึ่งไปยังอีกจุดหนึ่งในการจำแนกขอบเขตของปัจจัยทางสิ่งแวดล้อมที่เหมาะสม (Carpenter, Gillison & Winter, 1993) ดังภาพที่ 2.6



หมายเหตุ. * แสดง จุดที่มีการปรากฏสิ่งมีชีวิตภายใต้ปัจจัย
สิ่งแวดล้อม e_1 และ e_2
พื้นที่สี่เหลี่ยม แสดง ขอบเขตของปัจจัยทางสิ่งแวดล้อมที่เหมาะสม

ภาพที่ 2.5 วิธีการของแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี BIOCLIM



หมายเหตุ. * แสดง จุดที่มีการปรากฏสิ่งมีชีวิตภายใต้ปัจจัยสิ่งแวดล้อม e_1 และ e_2
●—● แสดง ขอบเขตของปัจจัยทางสิ่งแวดล้อมที่เหมาะสม

ภาพที่ 2.6 วิธีการของแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี DOMAIN

การสร้างแบบจำลองการกระจายสิ่งมีชีวิตโดยเทคนิคจำแนกกลุ่ม (Group Discrimination Techniques) สามารถแยกพิจารณาได้ 2 แนวทาง คือ แนวทางที่ 1 เป็นเทคนิคที่ใช้การสร้างแบบจำลองทางสถิติ (Statistical Models) ยกตัวอย่างเช่น การวิเคราะห์การถดถอย (Regression Analysis) หรือการวิเคราะห์การถดถอยสมัยใหม่ (Modern Regression Analysis) ซึ่งได้แก่ Generalized Linear Models (GLMs), Generalized Additive Models (GAMs), Boosted Regression Trees (BRTs) หรือ Multivariate Adaptive Regression Splines (MARSs) เป็นต้น โดยใช้ทั้งข้อมูลปรากฏและข้อมูลไม่ปรากฏ (Presence-Absence Data) ในการสร้างแบบจำลอง ส่วนแนวทางที่ 2 เป็นวิธีการที่อยู่บนพื้นฐานของ Machine-Learning โดยใช้ข้อมูลปรากฏ (Presence Data) ร่วมกับข้อมูลลักษณะสิ่งแวดล้อมในพื้นที่ (Background Data) ในการสร้างแบบจำลอง ยกตัวอย่างเช่น Maximum Entropy (MAXENT) (Mateo et al., 2009)

การสร้างแบบจำลองการกระจายสิ่งมีชีวิตโดยเทคนิคจำแนกกลุ่มเป็นเทคนิคในการสร้างแบบจำลองที่ต้องการข้อมูลทั้งข้อมูลปรากฏและข้อมูลไม่ปรากฏ (Presence-Absence Data) ซึ่งโดยปกติแล้ว ข้อมูลไม่ปรากฏ (Absence Data) มักจะไม่มีการบันทึกหรือเก็บรวบรวมไว้ ดังนั้นจึงมีการจำลองข้อมูลไม่ปรากฏขึ้นมา โดยเรียกข้อมูลนี้ว่า “ข้อมูลเสมือนไม่ปรากฏ” (Pseudo-Absence Data) โดยทั่วไปเทคนิคในการเลือกจุด Pseudo-absence สามารถจำแนกได้ 2 กลุ่มใหญ่ๆ (Lobo & Tognelli, 2011) คือ การเลือกจุด Pseudo-Absence โดยการสุ่มในพื้นที่ที่กำลังศึกษา (Barbet-Massin, Jiguet, Albert & Thuiller, 2012; Lobo & Tognelli, 2011) และการเลือกจุด Pseudo-Absence แบบ 2 ขั้นตอน (Barbet-Massiet et al., 2012; Lobo & Tognelli, 2011; Wisz & Guisan, 2009)

การเลือกจุด Pseudo-Absence โดยการสุ่มในพื้นที่ที่กำลังศึกษาเป็นการสุ่มเลือกอย่างง่าย โดยกำหนดขอบเขตของการสุ่มเป็นพื้นที่ที่กำลังศึกษา ในขณะที่การเลือกจุด Pseudo-Absence แบบ 2 ขั้นตอนเป็นการสุ่มเลือกจุด Pseudo-Absence ภายนอกพื้นที่ที่เหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตเป้าหมาย โดยขั้นตอนแรกจะทำการสร้างพื้นที่ที่คาดว่าจะมีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตนั้นโดยใช้เทคนิคโปรไฟล์ จากนั้นในขั้นตอนที่ 2 จะทำการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่ที่คาดว่าจะมีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตนั้น ๆ

2.3 เทคนิควิธีการสร้างแบบจำลองการกระจายตัวของสิ่งมีชีวิต

2.3.1 Generalized Linear Models (GLMs)

GLMs (Wood, 2006) ถูกแนะนำครั้งแรกโดย Nelder และ Wedderburn ในปี ค.ศ. 1972 ซึ่ง GLMs จะใช้ฟังก์ชันในกลุ่ม Exponential Family ทำให้เกิดตัวแบบในรูปเชิงเส้น โดยการประมาณค่า (Estimation) และการอนุมาน (Inference) จะอยู่บนพื้นฐานของทฤษฎี Maximum Likelihood Estimation โดย GLMs จะประกอบด้วย 3 ส่วน คือ

ส่วนที่ 1 ส่วนของฟังก์ชันการแจกแจงความน่าจะเป็น ซึ่งฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function) ของตัวแปรตาม (y) ใน GLMs จะเป็นการแจกแจงที่มาจากกลุ่มฟังก์ชันในกลุ่ม Exponential Family โดยฟังก์ชันการแจกแจงความน่าจะเป็นจะถูกจัดให้อยู่ในกลุ่มของ Exponential Family ได้ถ้าฟังก์ชันการแจกแจงนั้นสามารถเขียน Probability Density Function ในรูปแบบของ

$$f_{\theta}(Y) = \exp \left[\frac{Y\theta - b(\theta)}{a(\phi)} + c(Y, \phi) \right] \quad (1)$$

เมื่อ	a, b และ c	คือ	Arbitrary Function
	ϕ	คือ	Arbitrary Scale Parameter
	θ	คือ	Canonical Parameter ของการแจกแจง

ซึ่งฟังก์ชันการแจกแจงความน่าจะเป็นในรูปแบบอื่นที่ถูกจัดให้อยู่ในกลุ่มของ Exponential Family แสดงในตารางที่ 2.1

ตารางที่ 2.1 ตัวอย่างของฟังก์ชันการแจกแจงความน่าจะเป็น (Probability Distribution Function) ที่ถูกจัดให้อยู่ในกลุ่มของ Exponential Family

ฟังก์ชัน การแจกแจง	$f(y)$	θ	ϕ	$a(\phi)$	$b(\theta)$	$c(y, \phi)$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$; $-\infty < y < \infty$	μ	σ^2	$\phi(=\sigma^2)$	$\frac{\theta^2}{2}$	$-\frac{1}{2}\left[\frac{y^2}{\phi} + \ln(2\pi\phi)\right]$
Poisson	$\frac{\mu^y \exp(-\mu)}{y!}$; $y = 0, 1, 2, \dots$	$\ln(\mu)$	1	$\phi(=1)$	$\exp(\theta)$	$-\ln(y!)$
Binomial	$\binom{n}{y} \left(\frac{\mu}{n}\right)^y \left(1 - \frac{\mu}{n}\right)^{n-y}$; $y = 0, 1, 2, \dots, n$	$\ln\left(\frac{\mu}{n-\mu}\right)$	1	$\phi(=1)$	$n \ln(1 + e^\theta)$	$\ln\binom{n}{y}$

ส่วนที่ 2 ส่วนของ Linear Predictors จะเป็นตัวทำนายเชิงเส้นที่ประกอบไปด้วยพารามิเตอร์ β และตัวแปรอิสระ (Independence Variables) โดยมีรูปแบบดังนี้

$$\eta = X\beta$$

ส่วนที่ 3 ส่วนของ Link function เป็นส่วนของฟังก์ชันเชื่อมโยงเพื่อแสดงความสัมพันธ์ระหว่าง Linear predictors และฟังก์ชันการแจกแจงความน่าจะเป็น เมื่อใช้ฟังก์ชันการแจกแจงของ Canonical Parameter (θ) จะได้ Canonical Link Function ที่อยู่ในเทอมของ μ โดยในแต่ละฟังก์ชันการแจกแจงความน่าจะเป็น จะมีชื่อเรียก Link Function แตกต่างกันไป ดังตารางที่ 2.2

ตารางที่ 2.2 Canonical Link Function ของแต่ละฟังก์ชันการแจกแจงความน่าจะเป็น

Distribution	Name	Link Function	Mean Function
Normal	Identity	$X\beta = \mu$	$\mu = X\beta$
Poisson	Log	$X\beta = \ln(\mu)$	$\mu = \exp(X\beta)$
Binomial	Logit	$X\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$

ตัวแบบทั่วไปของ GLMs คือ

$$g(\mu_i) = X_i\beta \quad ; \quad \mu_i \equiv E(Y_i) \quad (2)$$

โดย

g คือ Link Function หรือ ฟังก์ชันการเชื่อมโยง

X_i คือ Matrix ของตัวแปรอิสระ (Independent Variable)

β คือ Vector ของตัวพารามิเตอร์

Y_i มีฟังก์ชันการแจกแจงอยู่ในกลุ่ม Exponential family ยกตัวอย่างเช่น

Poisson Distribution, Binomial Distribution, Gamma Distribution, Normal Distribution เป็นต้น นั่นคือ $Y_i \sim f_{\theta_i}(y_i)$

ในการประมาณค่าสัมประสิทธิ์โดยใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Method) จะได้ฟังก์ชันความน่าจะเป็น (Likelihood Function) ดังนี้

$$L(\beta) = \prod_{i=1}^n f_{\theta_i}(y_i) \quad (3)$$

เมื่อใส่ค่า $\ln$ ทั้งสองข้าง จะได้

$$\begin{aligned}\ln(L(\beta)) &= \sum_{i=1}^n \ln(f_{\theta_i}(y_i)) \\ &= \sum_{i=1}^n \left[\frac{y_i \theta_i - b_i(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi) \right]\end{aligned}$$

แทน $\ln(L(\beta))$ ด้วย $l(\beta)$ จะเรียกฟังก์ชันนี้ว่า Log Likelihood Function

$$l(\beta) = \sum_{i=1}^n \left[\frac{y_i \theta_i - b_i(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi) \right] \quad (4)$$

สมมติให้ $a_i(\phi) = \frac{\phi}{\omega_i}$ เมื่อ ω_i คือ ค่าคงที่ที่รู้ค่า โดยทั่วไปให้ $\omega_i = 1$ ในกรณีนี้จะได้

$$l(\beta) = \sum_{i=1}^n \left[\omega_i \left(\frac{y_i \theta_i - b_i(\theta_i)}{\phi} \right) + c_i(y_i, \phi) \right] \quad (5)$$

ในการประมาณค่าสัมประสิทธิ์ด้วยวิธีภาวะน่าจะเป็นสูงสุด จะกระทำโดยการหาอนุพันธ์อันดับ 1 ของ Log Likelihood Function เทียบกับ β แต่ละตัวแล้วให้เท่ากับ 0 จากนั้นแก้สมการหาค่า β

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^n \omega_i \left(y_i \frac{\partial \theta_i}{\partial \beta_j} - b'_i(\theta_i) \frac{\partial \theta_i}{\partial \beta_j} \right) \quad (6)$$

จากกฎลูกโซ่

$$\frac{\partial \theta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j} \quad (7)$$

และจากการหาค่าเฉลี่ยทั่วไปของฟังก์ชันการแจกแจงใน Exponential Family จากสมการ (1) ใส่ค่า $\log$ ทั้งสองข้าง

$$\ln[f_\theta(y)] = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

พิจารณาในฟังก์ชันของ θ แทน $\ln[f_\theta(y)]$ ด้วย $l(\theta)$ จะได้

$$l(\theta) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

หาอนุพันธ์อันดับ 1 เทียบ θ

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (8)$$

ให้ l มีคุณสมบัติเป็นตัวแปรสุ่ม (Random Variable) โดยแทนค่าสังเกต y ด้วยตัวแปรสุ่ม Y ดังนั้นค่าคาดหวัง (Expected Value) ของ $\frac{\partial l}{\partial \theta}$ คือ

$$E\left(\frac{\partial l}{\partial \theta}\right) = \frac{E(Y) - b'(\theta)}{a(\phi)}$$

จากคำตอบทั่วไป $E\left(\frac{\partial l}{\partial \theta}\right) = 0$ จะได้

$$E(Y) = b'(\theta) \quad (9)$$

ซึ่ง $E(Y)$ คือค่าเฉลี่ยของฟังก์ชันการแจกแจงความน่าจะเป็นที่มีค่าเท่ากับ $b'(\theta)$ เนื่องจาก $\mu \equiv E(Y)$ จะได้

$$\mu_i = b'_i(\theta_i) \quad (10)$$

ดังนั้น

$$\frac{\partial \mu_i}{\partial \theta_i} = b_i''(\theta_i)$$

จะได้

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b_i''(\theta_i)} \quad (11)$$

จากสมการที่ (7) และสมการที่ (11) นำไปแทนในสมการที่ (6) จะได้

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^n \frac{[y_i - b_i'(\theta_i)]}{b_i''(\theta_i)/w_i} \frac{\partial \mu_i}{\partial \beta_j} \quad (12)$$

จากสมการที่ (8) หาอนุพันธ์อันดับ 2 เทียบ θ จะได้

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}$$

จากคำตอบทั่วไป $E\left(\frac{\partial^2 l}{\partial \theta^2}\right) = -E\left[\left(\frac{\partial l}{\partial \theta}\right)^2\right]$ จะได้

$$\frac{b''(\theta)}{a(\phi)} = \frac{E[(Y - b'(\theta))^2]}{a(\phi)^2}$$

ซึ่ง $E[(Y - b'(\theta))^2]$ คือความแปรปรวนของฟังก์ชันการแจกแจงที่มีค่าเท่ากับ $b''(\theta)a(\phi)$ นั่นคือ

$$\text{var}(Y) = b''(\theta)a(\phi) \quad (13)$$

จาก $a(\phi) = \frac{\phi}{\omega}$ จะได้

$$\text{var}(Y) = \frac{b''(\theta)\phi}{\omega} \quad (14)$$

โดยปกติ จะพิจารณา $\text{var}(Y)$ เป็นฟังก์ชันของ $\mu \equiv E(Y)$ โดยจะกำหนดให้ $V(\mu_i) = \frac{b''(\theta_i)}{\omega_i}$ ดังนั้น

$$\text{var}(Y_i) = V(\mu_i)\phi \quad (15)$$

จากสมการที่ (14) และสมการที่ (15) จะได้

$$b''(\theta_i) = V(\mu_i)\omega \quad (16)$$

แทนสมการที่ (10) และสมการที่ (16) ในสมการที่ (12) จะได้

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\phi} \sum_{i=1}^n \frac{[y_i - \mu_i]}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j}$$

กำหนดให้

$$\frac{\partial l}{\partial \beta_j} = 0$$

จะได้

$$\sum_{i=1}^n \frac{[y_i - \mu_i]}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0, \quad \forall j \quad (17)$$

ในการคำนวณสมการที่ (17) เพื่อหาค่า β สามารถทำได้หลายวิธี ทั้งวิธี Non-Linear Weight Least Squares หรือวิธี Iterative Method

ในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs นั้น เนื่องจากตัวแปรตาม y เป็นค่าของ Presence Data และ Pseudo-Absence Data ซึ่งค่าของ y จะมีค่าเท่ากับ 1 เมื่อค่า y อยู่ในจุดพิกัดของ Presence Data และค่า y จะมีค่าเท่ากับ 0 เมื่อค่า y อยู่ในจุดพิกัดของ Pseudo-Absence Data ดังนั้น y จะมีค่าเพียง 2 ค่า คือ 0 และ 1 จะได้ y มีการแจกแจงแบบ Bernoulli ซึ่งก็คือการแจกแจงแบบ Binomial ที่มีค่า $n = 1$ จะได้ฟังก์ชันการแจกแจงความน่าจะเป็น ดังนี้

$$f(y) = \mu^y (1-\mu)^{1-y} \quad (18)$$

โดยสามารถเขียนให้อยู่ในรูปแบบ Exponential Family ได้ดังนี้

$$\begin{aligned} f(y) &= \mu^y (1-\mu)^{1-y} \\ \ln[f(y)] &= \ln[\mu^y (1-\mu)^{1-y}] \\ &= \ln[\mu^y] + \ln[(1-\mu)^{1-y}] \\ &= y \ln[\mu] + (1-y) \ln[1-\mu] \\ &= y \ln[\mu] + \ln[1-\mu] - y \ln[1-\mu] \\ &= y \ln\left[\frac{\mu}{1-\mu}\right] - \ln\left[\frac{1}{1-\mu}\right] \\ f(y) &= \exp\left[y \ln\left[\frac{\mu}{1-\mu}\right] - \ln\left[\frac{1}{1-\mu}\right]\right] \end{aligned}$$

จะได้ว่า

$$\begin{aligned} \theta &= \ln\left(\frac{\mu}{1-\mu}\right), \quad b(\theta) = \ln(1+e^\theta) \\ \phi &= 1, \quad a(\phi) = \phi, \quad c(y, \phi) = 0 \end{aligned}$$

จากสมการที่ (9)

$$\begin{aligned} E(Y) &= b'(\theta) \\ &= \frac{e^\theta}{1+e^\theta} \\ &= \mu \end{aligned}$$

โดยมี link function คือ

$$g(\mu) = \theta = \ln\left(\frac{\mu}{1-\mu}\right)$$

จะได้

$$\ln\left(\frac{\mu}{1-\mu}\right) = X\beta \quad (19)$$

เมื่อกำหนดให้ μ เป็นความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ นั่นคือ $y=1$ จะได้ว่า $1-\mu$ คือความน่าจะเป็นของการเกิดเหตุการณ์ที่ไม่สนใจ นั่นคือ $y=0$ ดังนั้นอัตราส่วนระหว่างความน่าจะเป็นของการเกิดเหตุการณ์ที่สนใจ (μ) ต่อความน่าจะเป็นของการเกิดเหตุการณ์ที่ไม่สนใจ ($1-\mu$) ก็คือโอกาสของการเกิดเหตุการณ์ที่ $y=1$ มีอัตราส่วนเป็นกี่เท่าของการเกิดเหตุการณ์ที่ $y=0$ จะเรียกอัตราส่วนนี้ว่า Odds ratio

$$\text{Odds ratio} = \frac{\mu}{1-\mu}$$

ซึ่งจะเรียก $\ln\left[\frac{\mu}{1-\mu}\right]$ ว่า Log of the Odds หรือเรียกว่า Logit Transformation หรือ Logit (μ) และจะเรียกสมการพยากรณ์ในลักษณะนี้ว่าการวิเคราะห์สมการถดถอยโลจิสติก (Logistic Regression Analysis)

สรุป การวิเคราะห์สมการถดถอยโลจิสติก จะมีรูปแบบสมการคือ

$$E(Y) = \frac{e^{x\beta}}{1+e^{x\beta}}$$

โดยมีรูปแบบพหุคูณคือ

$$\text{Logit}(\mu) = X\beta$$

2.3.2 Generalized Additive Models (GAMs)

GAMs ถูกแนะนำครั้งแรกโดย Trevor Hastie และ Robert Tibshirani ในปี ค.ศ. 1986 (Hastie & Tibshirani, 1986) ซึ่งถูกจัดเป็นการวิเคราะห์การถดถอยแบบ Non-Parametric (Irwin, 2005) โดยสมมติว่าค่าเฉลี่ยของตัวแปรขึ้นอยู่ด้วย Additive Predictors ผ่านฟังก์ชันการเชื่อมโยง (Link functions) แบบไม่เป็นเชิงเส้น (SAS, n.d.) ซึ่งจะใช้ฟังก์ชันการแจกแจงในกลุ่มของ Exponential Family ในการสร้างแบบจำลอง

กำหนดให้ y เป็นตัวแปรตาม (Dependence Variable) และ $x_1, x_2, \dots, x_p$ เป็นเซตของตัวแปรอิสระ (Independence Variables) โดยทั่วไปรูปแบบมาตรฐานของแบบจำลองการถดถอยเชิงเส้น (Linear Regression) คือ

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (20)$$

เมื่อ $\beta_0, \beta_1, \dots, \beta_p$ เป็นพารามิเตอร์ที่ถูกประมาณค่าโดยวิธี Least-Square ในขณะที่รูปแบบทั่วไปของ Additive Model คือ

$$E(y) = s_0 + s_1(x_1) + s_2(x_2) + \dots + s_p(x_p) \quad (21)$$

เมื่อ $s_1, s_2, \dots, s_p$ เป็น Smooth Functions ซึ่งถูกประมาณค่าในรูปแบบของ Non-Parametric องค์ประกอบโดยทั่วไปของ GAMs ประกอบไปด้วย 3 ส่วนคล้ายๆ กับ GLMs ดังนี้

ส่วนที่ 1 ส่วนของฟังก์ชันการแจกแจงความน่าจะเป็นของตัวแปรตาม (y) จะเป็นการแจกแจงที่มาจากกลุ่มฟังก์ชันในกลุ่ม Exponential Family

ส่วนที่ 2 ส่วนของ Additive Component จะเป็นตัวทำนายที่ประกอบไปด้วย Smooth Functions (s_i) และตัวแปรอิสระ (Independence Variables) โดยมีรูปแบบดังนี้

$$\eta = s_0 + \sum_{i=1}^p s_i(x_i)$$

ส่วนที่ 3 ส่วนของ Link Function เป็นส่วนของฟังก์ชันเชื่อมโยงเพื่อแสดงความสัมพันธ์ระหว่าง Additive Component และฟังก์ชันการแจกแจงความน่าจะเป็น

การประมาณค่า Smooth Functions ในแต่ละฟังก์ชัน $s_j(i)$ จะประมาณค่าผ่าน Scatterplot Smoother เริ่มต้นจากการพิจารณากรณีตัวแปรอิสระ (x) 1 ตัวแปร (Univariate Smooth Functions) แบบจำลองจะมีรูปแบบ ดังนี้

$$E(y|x) = s(x) \quad (22)$$

ซึ่งในกรณีที่ตัวแปรอิสระ 1 ตัวในเทอมของค่าคงที่ s_0 จะละเว้นแต่จะถูกรวมอยู่ใน เทอมของ $s(x)$ วิธีประมาณค่า $s(x)$ มีหลายวิธีด้วยกัน หนึ่งในนั้นคือวิธี Local Average Estimates นั่นคือ

$$\hat{s}(x_i) = \text{Ave}_{j \in N_i} \{y_j\} \quad (23)$$

เมื่อ Ave คือ Averaging Operator
 N_i คือ Neighborhood ของ x_i

ถ้า Ave หมายถึงค่าเฉลี่ย (Mean) จะเรียกวิธีประมาณค่า $s(X)$ ว่า Running Mean Smoother แต่ถ้า Ave หมายถึงค่ามัธยฐาน (Median) จะเรียกวิธีการประมาณค่านี้ว่า Running Median Smoother

เนื่องจากวิธีการ Running Mean หรือ Running Median Smoother ก่อให้เกิดปัญหาการเอนเอียง (Biases) ที่มีขนาดใหญ่โดยเฉพาะจุดสุดท้าย จึงมีการปรับวิธีการเพื่อแก้ปัญหาดังกล่าวนี้ โดยเรียกวิธีการนี้ว่า Running Line Smoother ซึ่งมีรูปแบบการประมาณกันไป ยกตัวอย่างดังต่อไปนี้

1. Running Line Smoother แบบ Linear Regression Splines เป็นการสร้าง Smooth Function แบบเส้นตรง โดยมีรูปแบบการประมาณดังนี้

$$\hat{s}(x_i) = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_i \quad (24)$$

เมื่อ $\hat{\beta}_{0i}$ และ $\hat{\beta}_{1i}$ เกิดจากการประมาณค่าโดยวิธี Least square ของข้อมูลใน N_i

2. Running Line Smoother แบบ Polynomial Spline เป็นการสร้าง Smooth Function ในรูปแบบพหุนามดีกรีต่าง ๆ ยกตัวอย่าง เช่น

รูปแบบการประมาณ Smooth Function พหุนามดีกรี 2 (Quadratic Spline)

$$\hat{s}(x_i) = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_i + \hat{\beta}_{2i}x_i^2 \quad (25)$$

รูปแบบการประมาณ Smooth Function พหุนามดีกรี 3 (Cubic Spline)

$$\hat{s}(x_i) = \hat{\beta}_{0i} + \hat{\beta}_{1i}x_i + \hat{\beta}_{2i}x_i^2 + \hat{\beta}_{3i}x_i^3 \quad (26)$$

พิจารณากรณีตัวแปรอิสระ p ตัว กำหนดโดยเวกเตอร์ $\mathbf{X} = (x_1, x_2, \dots, x_p)$ ในการประมาณค่า $s_0, s_1(\cdot), \dots, s_p(\cdot)$ สำหรับ Additive Regression Model

$$E(y | \mathbf{X}) = s_0 + \sum_{j=1}^p s_j(x_j) \quad (27)$$

สำหรับรูปแบบเชิงเส้น $\eta = \beta_0 + \sum_{i=1}^p \beta_i x_i$ จะประมาณค่า $\beta_0, \beta_1, \dots, \beta_p$ ซึ่งขึ้นอยู่กับตัวแปร $\mathbf{X}$ โดยวิธีการวิเคราะห์การถดถอย ในขณะที่รูปแบบ Smooth Function $\eta = s_0 + \sum_{i=1}^p s_i(x_i)$ จะประมาณค่า $s(\cdot)$ โดยการทำ Smoothing ซ้ำ ซึ่งขึ้นอยู่กับตัวแปร $\mathbf{X}$ โดยทั่วไปในการประมาณค่า Smooth Function ของ Additive Regression Model จะใช้วิธี Backfitting Algorithm ดังนั้นในส่วนการประมาณค่า Smooth Function ของ GAMs จะใช้วิธี Backfitting Algorithm เป็นส่วนสำคัญในการประมาณค่า

2.3.3 Maximum Entropy (MAXENT)

วิธี MAXENT เป็นวิธีการที่มีพื้นฐานมาจาก Statistical Mechanics โดยมีวัตถุประสงค์ทั่วไปคือการพยากรณ์และหาข้อสรุปจากข้อมูลที่ไม่สมบูรณ์ (Phillips, Anderson & Schapire, 2006) การใช้วิธี MAXENT ในการหาการกระจายตัวของสิ่งมีชีวิตเป็นการประมาณค่าความน่าจะเป็นของการกระจายของค่าเอนโทรปีสูงสุด (Maximum Entropy)

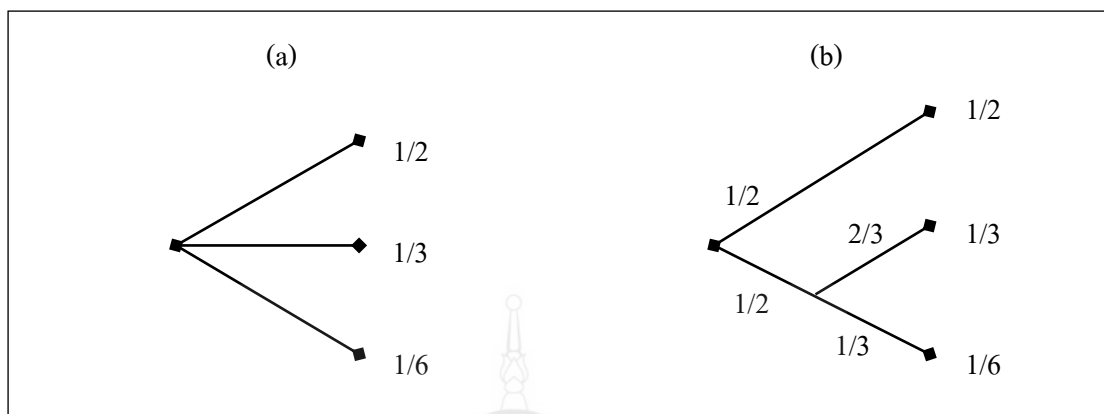
กำหนดให้ π คือความน่าจะเป็นของการกระจายที่ไม่ทราบค่า และเวกเตอร์ X เป็นเซตของจุดในพื้นที่ที่ศึกษา ดังนั้นความน่าจะเป็นของการกระจาย ณ จุดที่ $X = x_i$ คือ $\pi(x_i)$ โดยมีผลรวมของความน่าจะเป็นในทุกจุดเท่ากับ 1 ในการประมาณค่าความน่าจะเป็น จะกำหนด $\hat{\pi}(x_i)$ เป็นค่าความน่าจะเป็นที่ถูกประมาณขึ้น ณ จุดที่ $X = x_i$ ดังนั้นเอนโทรปีของ $\hat{\pi}$ คือ

$$H(\hat{\pi}) = - \sum_{x_i \in X} \hat{\pi}(x_i) \ln \hat{\pi}(x_i) \quad (28)$$

H มีคุณสมบัติที่สำคัญ 3 ข้อ (Shannon, 1984) ดังนี้

1. H มีความต่อเนื่องบน $\hat{\pi}(x_i)$ กล่าวคือ หากเปลี่ยนค่าความน่าจะเป็นอันใดอันหนึ่งเพียงเล็กน้อย ค่าเอนโทรปีก็จะเปลี่ยนแปลงเพียงเล็กน้อยเช่นกัน
2. ถ้าทุก ๆ ค่าของ $\hat{\pi}(x_i)$ มีค่าเท่ากัน นั่นคือ $\hat{\pi}(x_i) = \frac{1}{n}$ แล้ว H จะเป็น Monotonic Increase Function ของ n กล่าวคือ เหตุการณ์มีทางเลือกหรือความไม่แน่นอนมากขึ้น หรือมีเหตุการณ์ที่เป็นไปได้มากขึ้น
3. หากทางเลือกถูกจำกัดโดยแบ่งออกเป็น 2 ทาง H จะถูกถ่วงน้ำหนักด้วยผลรวมของ H แต่ละค่า ดังแสดงในภาพที่ 2.7 แสดงตัวอย่างการจำแนกทางเลือกของความน่าจะเป็นที่เป็นไปได้ 3 ทาง นั่นคือ ความน่าจะเป็นของภาพที่ 2.7 (a) $\pi_1 = \frac{1}{2}$, $\pi_2 = \frac{1}{3}$ และ $\pi_3 = \frac{1}{6}$ ในภาพที่ 2.7 (b) ทางเลือกจะถูกจำกัดโดยแบ่งออกเป็น 2 ทาง โดยใช้ชั้นแรกความน่าจะเป็นในแต่ละทางมีค่าเท่ากับ $\frac{1}{2}$ และในชั้นที่สองของทางเลือกที่ 2 ความน่าจะเป็นจะถูกแยกออกเป็น $\frac{2}{3}$ และ $\frac{1}{3}$ ซึ่งผลลัพธ์สุดท้ายที่ได้ของทั้ง 3 ทางเลือกจะต้องเท่ากับค่าความน่าจะเป็นในภาพที่ 2.7 (a) ซึ่งจะได้ว่า

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right)$$



ที่มา Shannon (1948)

ภาพที่ 2.7 การจำแนกทางเลือกของความน่าจะเป็นที่เป็นไปได้ 3 ทาง

H ถูกนำมาใช้เพื่อแสดงความไม่แน่นอนของระบบ ซึ่งตามทฤษฎีอุณหพลศาสตร์นั้น กล่าวไว้ว่า เมื่อมีการเปลี่ยนแปลงใด เกิดขึ้น เอนโทรปีรวมของสิ่งต่าง ๆ ที่เกี่ยวข้องกับการเปลี่ยนแปลงนั้นก็จะเพิ่มสูงขึ้นด้วย

แนวคิดในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต (Species Distribution Models) โดยวิธี MAXENT คือการหาค่าความน่าจะเป็น π บนพื้นที่ X (Phillips & Dudík, 2008) ซึ่งกำหนดให้ X คือเซตของจำนวนกริดภายในพื้นที่ที่ศึกษานั้นคือ

$$X = \{x_1, x_2, \dots, x_n\}$$

ดังนั้นค่าความน่าจะเป็น π จะมีค่าเป็นบวก โดย $\sum_{i=1}^n \pi(x_i) = 1$ โดยพื้นที่ X จะถูกกำหนดด้วยข้อจำกัดของตัวแปรสิ่งแวดล้อม เช่น ความสูงจากระดับน้ำทะเล ปริมาณน้ำฝน อุณหภูมิ เป็นต้น ซึ่งอยู่ในรูปแบบของฟังก์ชันพื้นฐานโดยกำหนดให้ $f = \{f_1, f_2, \dots, f_m\}$ เป็นเซตของฟังก์ชันที่รู้ค่าจริง (Known Real-Valued Function) บน X จะเรียกข้อจำกัดเหล่านี้ว่า “Feature” ยกตัวอย่าง Feature ในรูปแบบต่างๆ เช่น

1. แบบเชิงเส้น (Linear) : $f_j = v_k$
2. แบบยกกำลังสอง (Quadratic) : $f_j = v_k^2$
3. แบบผลคูณ (Product) : $f_j = v_k v_l$

$$4. \text{แบบ Threshold : } f_j = \begin{cases} 1 & \text{if } v_k \geq a \\ 0 & \text{else} \end{cases}$$

กำหนดให้ $x_1, x_2, \dots, x_n$ แทนข้อมูลการปรากฏสิ่งมีชีวิต (Presence Data) ซึ่งถูกเลือกอย่างเป็นอิสระจาก X ด้วยความน่าจะเป็น π ซึ่งเป็นความน่าจะเป็นที่ไม่ทราบค่า ดังนั้นสิ่งที่ต้องการคือการประมาณค่าความน่าจะเป็น ($\hat{\pi}$)

โดยแต่ละ Feature f_j จะถูกกำหนดในทุกๆ จุดของ x ที่อยู่ใน X (Schapire, Phillips, & Dudík, n.d.) โดยทั่วไปแล้ว สำหรับการแจกแจงความน่าจะเป็น p ใดๆ และฟังก์ชัน f จะใช้สัญลักษณ์ $p[f]$ แทนค่าคาดหวัง (Expected Value) ของฟังก์ชัน f ภายใต้ความน่าจะเป็น p ดังนั้น กำหนดให้ $\pi[f_j]$ เป็นค่าคาดหวังของ Feature f_j ภายใต้ความน่าจะเป็น π จะได้ว่า

$$\pi[f_j] = \sum_{x \in X} \pi(x) f_j(x) \quad (29)$$

ซึ่ง $\pi[f_j]$ สามารถประมาณค่าได้โดยใช้เซตของกลุ่มตัวอย่าง $x_1, x_2, \dots, x_n$ บน X โดยมีค่าเฉลี่ยของ f_j (Empirical Average) คือ

$$\tilde{\pi}[f_j] = \frac{\sum_{i=1}^n f_j(x_i)}{n} \quad (30)$$

ซึ่ง $\tilde{\pi}[f_j]$ จะมีการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) จากหลักการของ Maximum Entropy จะพยายามหาความน่าจะเป็นที่มีเอนโทรปีสูงสุด ซึ่งจะมีลักษณะการแจกแจงเข้าสู่การแจกแจงแบบยูนิฟอร์ม โดยมีค่าเอนโทรปีดังสมการ (28) ภายใต้ข้อจำกัดของแต่ละ Feature (Phillips, Anderson & Schapire, 2006) ดังนั้น $\tilde{\pi}[f_j]$ จะถูกนำมาใช้เป็นตัวประมาณของ $\pi[f_j]$ นั่นคือ

$$\hat{\pi}[f_j] = \tilde{\pi}[f_j] \quad \text{สำหรับทุกๆ ค่าของ } f_j \quad (31)$$

บทที่ 3

ระเบียบและวิธีการศึกษา

ในบทนี้เป็นการกล่าวถึงระเบียบและวิธีในการศึกษา ซึ่งได้แก่ขอบเขตของการศึกษา การเก็บรวบรวมข้อมูล การสร้างตัวแปร การสร้างชุดข้อมูล รวมถึงวิธีการสร้างแบบจำลองและการหาแบบจำลองที่เหมาะสม

3.1 ขอบเขตของการศึกษา

การศึกษานี้เป็นการสร้างแบบจำลองการกระจายตัวของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน ประกอบด้วย 9 จังหวัด ดังต่อไปนี้ เชียงราย เชียงใหม่ พะเยา แพร่ น่าน ลำพูน ลำปาง อุตรดิตถ์และแม่ฮ่องสอน ดังแสดงในภาพที่ 3.1 โดยชาป่าที่ศึกษาเป็นชาสายพันธุ์อัสสัม (*Camillia sinensis* ver. *assamica*) ซึ่งเป็นชาที่เกิดขึ้นเองตามธรรมชาติ ในพื้นที่ป่าเขตภาคเหนือตอนบน

3.2 การเก็บรวบรวมข้อมูล

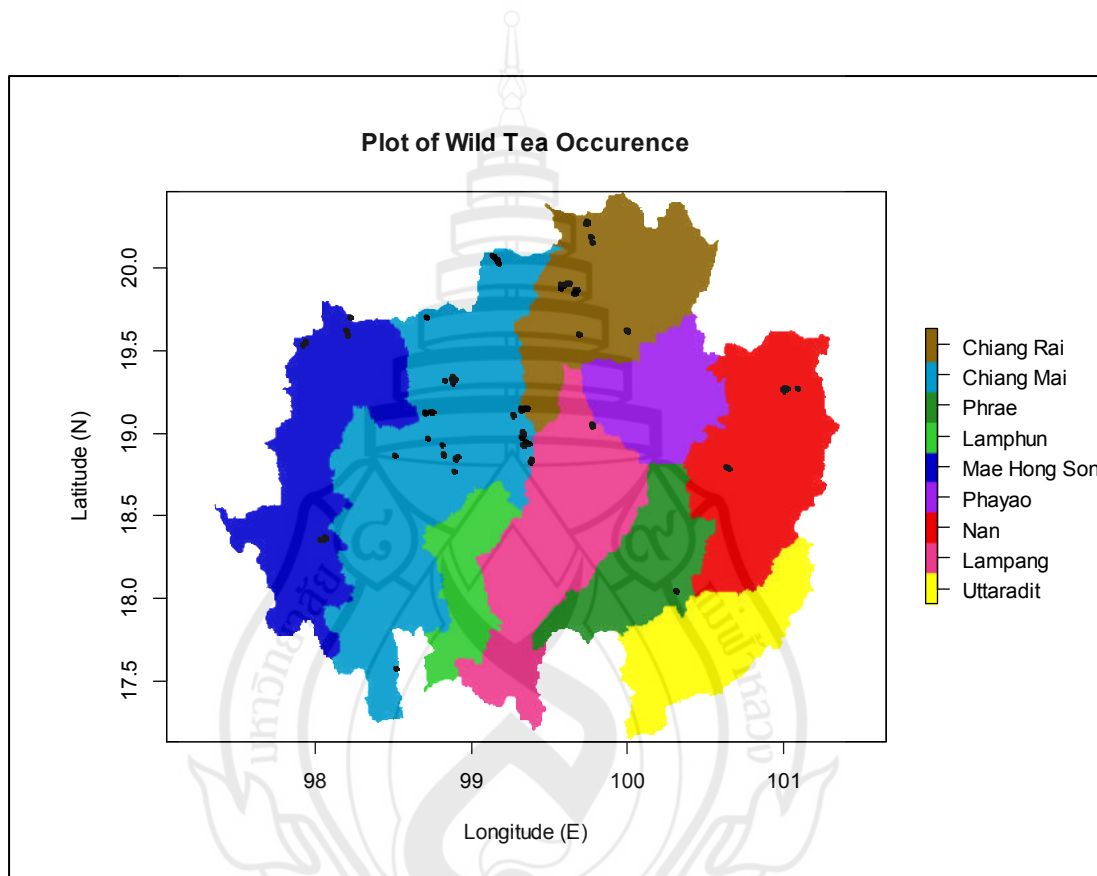
ข้อมูลที่ใช้ในการศึกษานี้จะใช้ในการสร้างแบบจำลอง โดยข้อมูลที่ใช้ในการสร้างแบบจำลองแบ่งออกเป็น 2 ประเภทคือข้อมูลที่ได้มาจากการรวบรวมจากหน่วยงานต่าง ๆ และข้อมูลที่ได้จากการสร้างปัจจัยเพิ่มเติมจากข้อมูลเดิมที่มีอยู่ โดยมีรายละเอียดดังต่อไปนี้

3.2.1 ข้อมูลที่ได้จากการรวบรวมจากหน่วยงาน

ข้อมูลที่ได้จากการรวบรวมจากหน่วยงานต่าง ๆ ได้แก่ ข้อมูลชาป่า ข้อมูลทางภูมิศาสตร์ และข้อมูลทางภูมิอากาศ ดังต่อไปนี้

3.2.1.1 ข้อมูลชาป่า

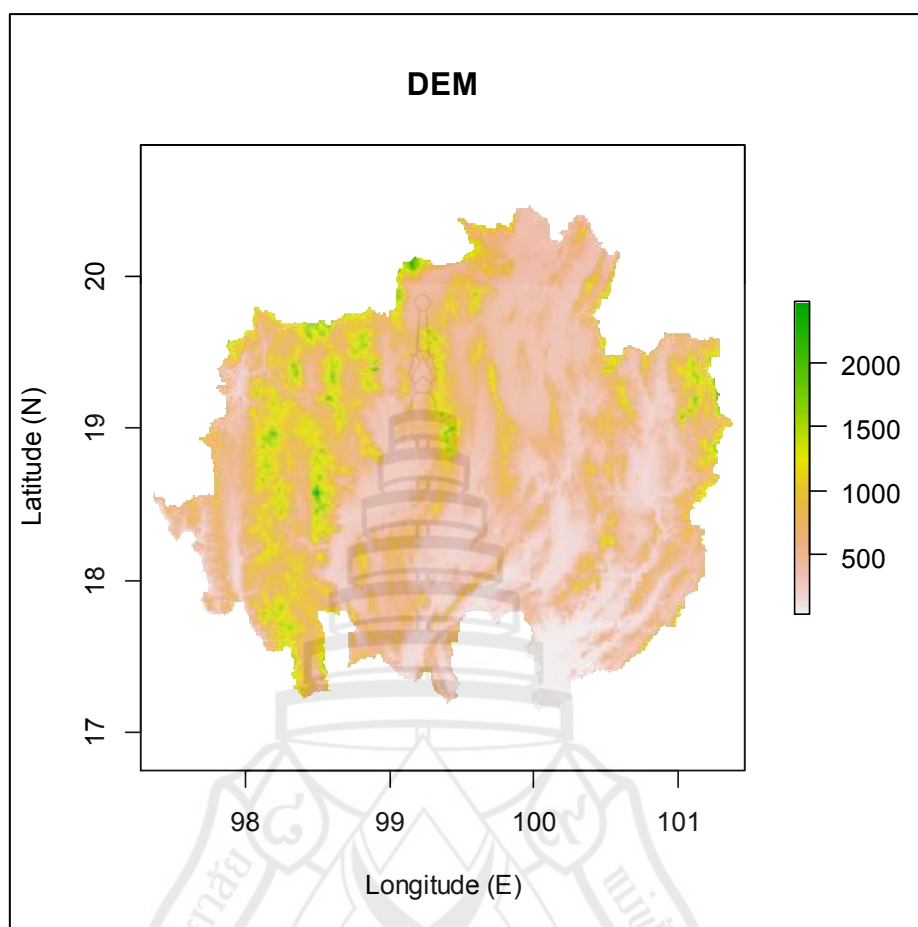
ข้อมูลชาป่าในเขตภาคเหนือเป็นข้อมูลในลักษณะของจุดพิกัด ซึ่งเก็บรวบรวมโดยสถาบันชา มหาวิทยาลัยแม่ฟ้าหลวง เป็นส่วนหนึ่งของโครงการเก็บและกำหนดพันธุ์ชาที่เหมาะสมในพื้นที่ปลูกภาคเหนือของประเทศไทย (สายลม สัมพันธ์เวชโสภา และคณะ, 2550) ดังแสดงในภาพที่ 3.1



ภาพที่ 3.1 จุดพิกัด (จุดสีดำ) ที่มีการค้นพบต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบนจำนวน 185 จุด

3.2.1.2 ความสูงจากระดับน้ำทะเล

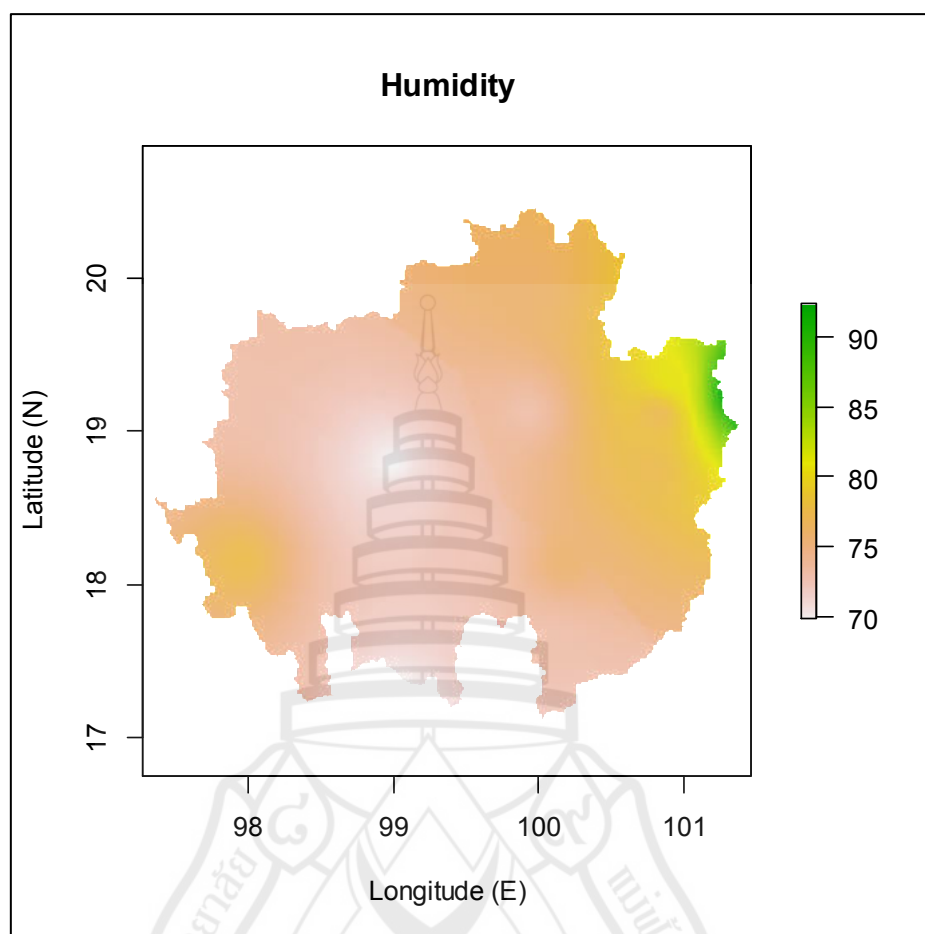
ข้อมูลความสูงจากระดับน้ำทะเลใช้ข้อมูลแผนที่ลักษณะความสูงของภูมิประเทศแบบดิจิตอล (Digital Elevation Model: DEM) ซึ่งมีความละเอียดทางราบ 30 เมตรของกรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์ โดยมีหน่วยเป็นเมตร (m) ดังแสดงในภาพที่ 3.2



ภาพที่ 3.2 ข้อมูลแผนที่ลักษณะความสูงของภูมิประเทศแบบดิจิทัล (Digital Elevation Model: DEM) ในเขตพื้นที่ภาคเหนือตอนบน โดยมีหน่วยเป็น เมตร

3.2.1.3 ความชื้นสัมพัทธ์

เป็นข้อมูลออนไลน์จาก <http://www.rsgis.ait.ac.th/~souris/thailand.htm> ของสาขาวิชา Remote Sensing and Geographic Information Systems สถาบันเทคโนโลยีแห่งเอเชีย (Asian Institute of Technology: AIT) ซึ่งเก็บรวบรวมข้อมูลเป็นรายเดือนระหว่างปี 2524 ถึงปี 2544 โดย Dr. Marc Souris นักวิจัยวิทยาศาสตร์คอมพิวเตอร์และนักพัฒนาระบบสารสนเทศทางภูมิศาสตร์ ดังแสดงในภาพที่ 3.3



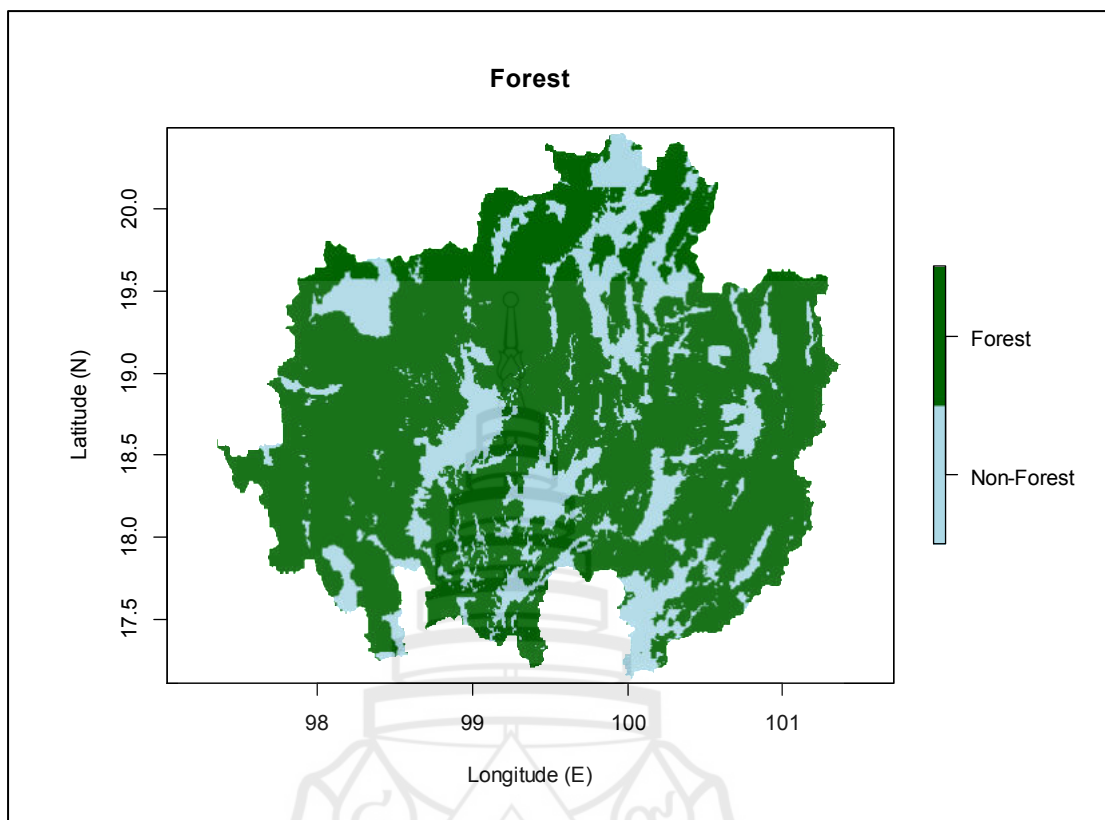
ภาพที่ 3.3 ความชื้นสัมพัทธ์โดยมีหน่วยเป็น เปอร์เซนต์

3.2.1.4 แม่น้ำสายหลักและแม่น้ำสาขา

เป็นข้อมูลออนไลน์จาก <http://www.diva-gis.org/> ซึ่งแสดงเส้นแม่น้ำสายหลักและแม่น้ำสาขา ซึ่งเป็นเว็บไซต์ที่บริการโปรแกรมคอมพิวเตอร์สำหรับการทำแผนที่และวิเคราะห์ทางภูมิศาสตร์โดยไม่คิดค่าใช้จ่าย

3.2.1.5 พื้นที่ป่า

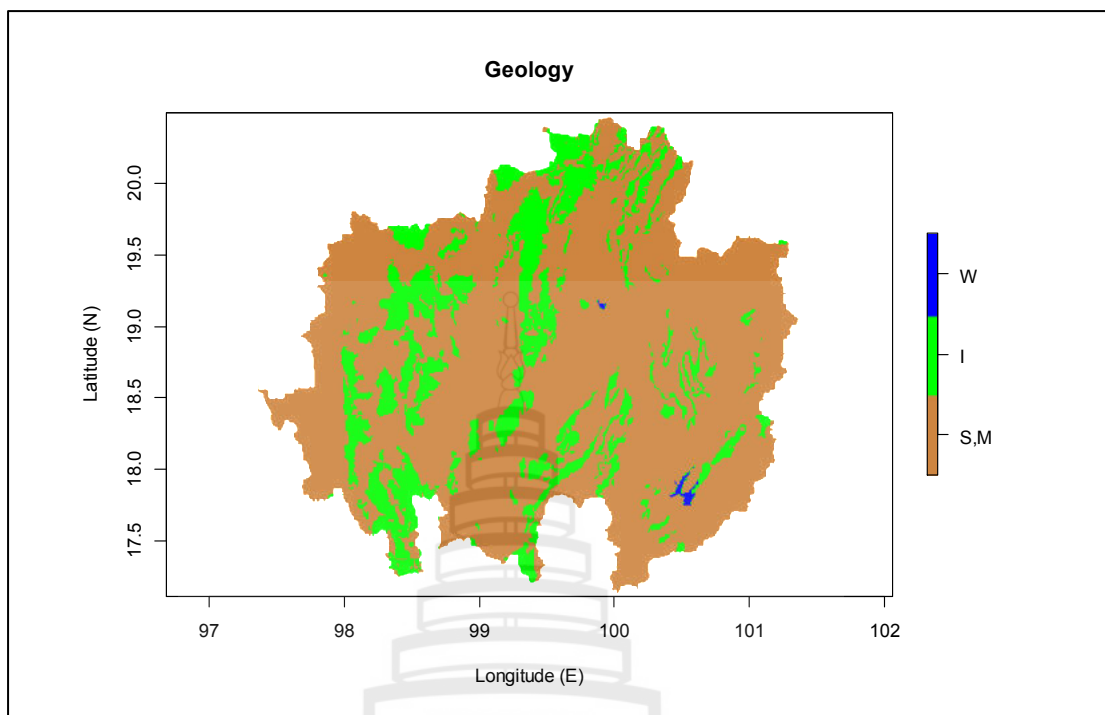
เป็นข้อมูลแสดงเขตพื้นที่ป่าไม้ในเขตพื้นที่ภาคเหนือโดยศูนย์บริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่ฟ้าหลวง ซึ่งเป็นข้อมูลป่าไม้ปี พ.ศ. 2545 จากกรมป่าไม้ กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม ดังแสดงในภาพที่ 3.4



ภาพที่ 3.4 พื้นที่ป่าในเขตภาคเหนือตอนบน

3.2.1.6 โครงสร้างทางธรณีวิทยาของชนิดหิน

ข้อมูลโครงสร้างทางธรณีวิทยาของชนิดหินที่โผล่ที่ผิวดินหรือถูกดินปกคลุมอยู่ไม่เกิน 1 ถึง 1.5 เมตร โดยศูนย์บริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยแม่ฟ้าหลวง เป็นโครงสร้างทางธรณีวิทยาของชนิดหินปี พ.ศ. 2545 จากกรมทรัพยากรธรณี กระทรวงทรัพยากรธรรมชาติและสิ่งแวดล้อม ซึ่งแบ่งออกได้เป็น 3 ประเภท ประเภทที่หนึ่งเป็นเขตของหินตะกอนและหินแปร (Sedimentary and Metamorphic Rocks : S,M) ประเภทที่สองเป็นเขตของหินอัคนี (Igneous Rocks : I) และประเภทที่สามเป็นเขตของพื้นที่ที่เป็นน้ำ (Water : W) ดังแสดงในภาพที่ 3.5



ภาพที่ 3.5 ขอบเขตของชั้นหินในเขตพื้นที่ภาคเหนือตอนบน โดยพื้นที่สีน้ำตาลเป็นส่วนของหินตะกอนและหินแปร พื้นที่สีเขียวเป็นส่วนของหินอัคนีและพื้นที่สีน้ำเงินเป็นส่วน of พื้นที่น้ำ

3.2.1.7 สภาพภูมิอากาศ (Bioclimatic Variables)

เป็นข้อมูลออนไลน์จาก <http://www.worldclim.org> ซึ่งสร้างมาจากอุณหภูมิและปริมาณน้ำฝน เก็บข้อมูลระหว่างปี ค.ศ. 1950-2000 มีทั้งหมด 19 ตัวแปรด้วยกัน โดยแบ่งเป็นตัวแปรที่เกี่ยวข้องกับอุณหภูมิ (Bio1-Bio11) ซึ่งมีหน่วยเป็นองศาเซลเซียส ($^{\circ}\text{C}$) คูณด้วย 10 นั้นหมายความว่า ถ้าค่าของตัวแปรอุณหภูมิมีค่าเท่ากับ 231 จะหมายความว่าตัวแปรนั้นมีอุณหภูมิ 23.1 องศาเซลเซียส ดังแสดงในภาพที่ 3.6 ซึ่งมีรายละเอียดดังนี้

1. Annual Mean Temperature (Bio1) คืออุณหภูมิเฉลี่ยรายปี มีหน่วยเป็นองศาเซลเซียส
2. Mean Diurnal Range (Bio2) คือช่วงอุณหภูมิระหว่างวันโดยเฉลี่ย หาได้จากค่าเฉลี่ยของผลต่างระหว่างอุณหภูมิสูงสุดและอุณหภูมิต่ำสุดในแต่ละวัน
3. Isothermality (Bio3) เป็นดัชนีที่ใช้อธิบายการผันผวนของอุณหภูมิรายวันโดยเปรียบเทียบกับช่วงของอุณหภูมิรายปี ซึ่งอาจมีผลต่อการกระจายของสิ่งมีชีวิตที่แตกต่างกันคำนวณได้จากช่วงอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) หารด้วยผลต่างของอุณหภูมิสูงสุดในเดือน

ที่ร้อนที่สุด (Bio5) และอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6) ซึ่งก็คือช่วงของอุณหภูมิรายปี (Bio7)

4. Temperature Seasonality (Bio4) คือช่วงฤดูกาลของอุณหภูมิ เป็นดัชนีที่ใช้สำหรับวัดการเปลี่ยนแปลงของอุณหภูมิในช่วงปีที่กำหนดหรือเฉลี่ยรายปี ซึ่งขึ้นอยู่กับส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของอุณหภูมิเฉลี่ยรายเดือน

5. Max Temperature of Warmest Month (Bio5) คืออุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด

6. Min Temperature of Coldest Month (Bio6) คืออุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด

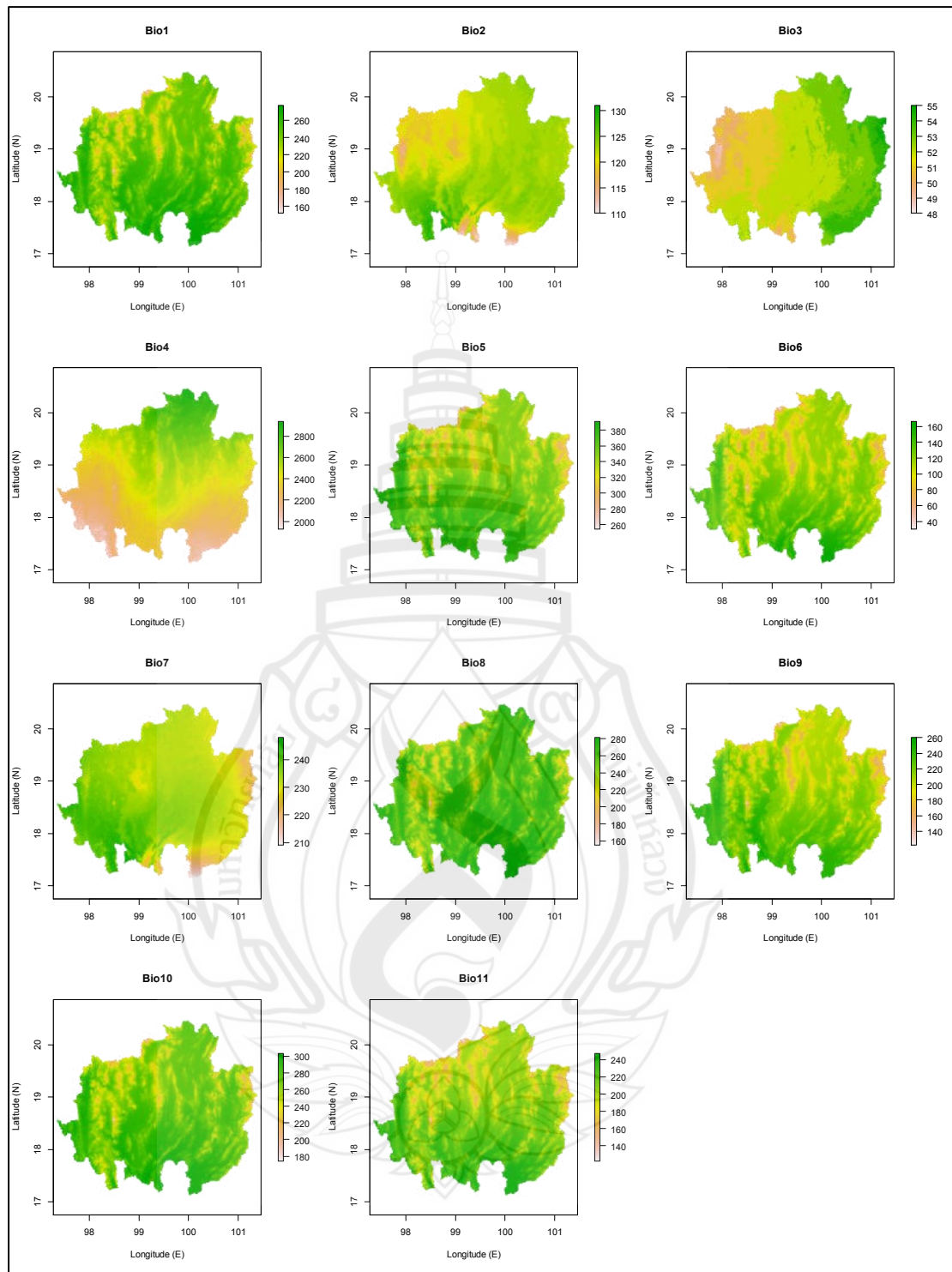
7. Temperature Annual Range (Bio7) คือช่วงของอุณหภูมิรายปี คำนวณได้จากผลต่างระหว่างอุณหภูมิสูงสุดในเดือนที่ร้อนที่สุดและอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด

8. Mean Temperature of Wettest Quarter (Bio8) คืออุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด

9. Mean Temperature of Driest Quarter (Bio9) คืออุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด

10. Mean Temperature of Warmest Quarter (Bio10) คืออุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด

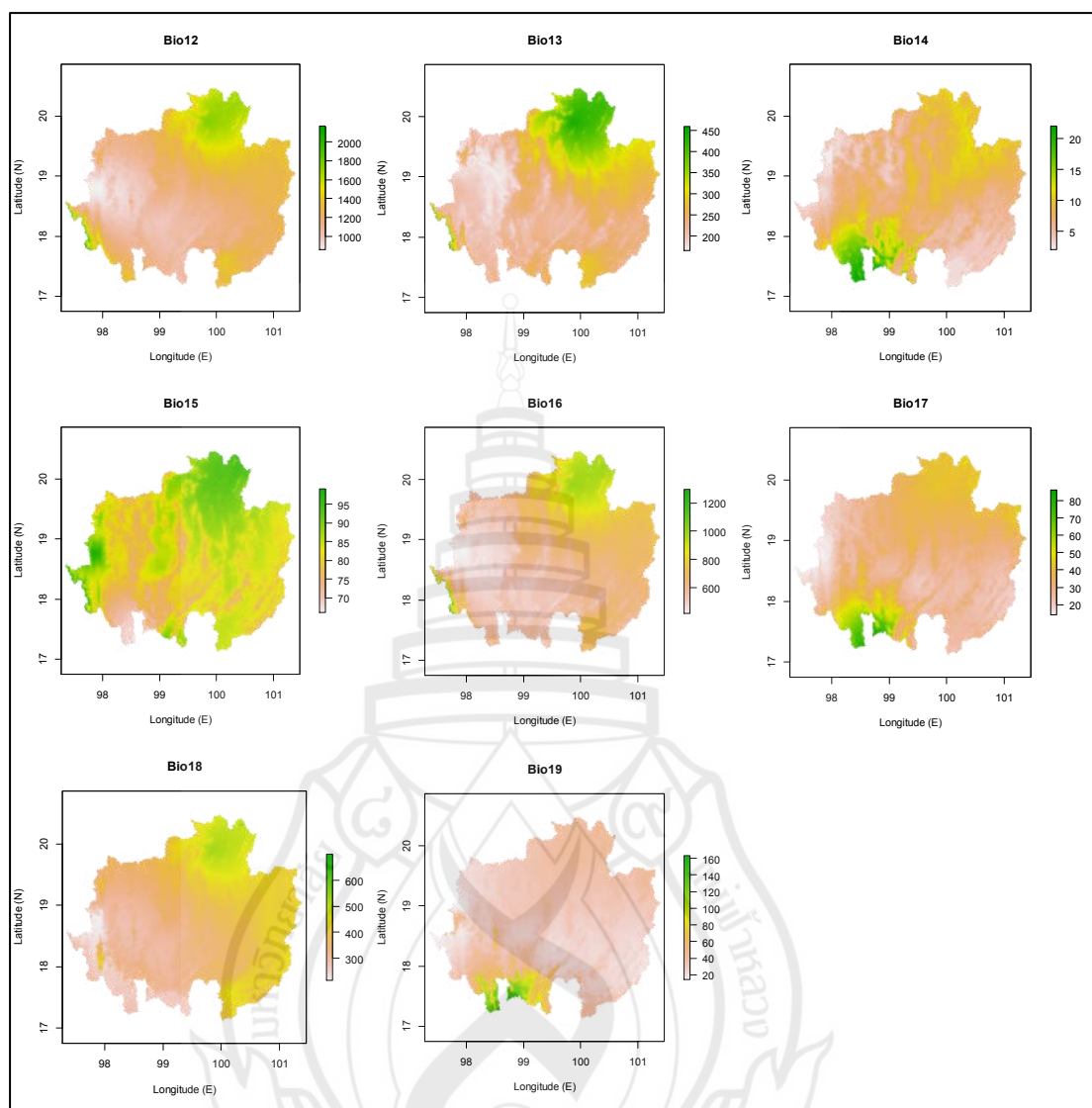
11. Mean Temperature of Coldest Quarter (Bio11) คืออุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด



ภาพที่ 3.6 ตัวแปรที่เกี่ยวข้องกับอุณหภูมิ (Bio1-Bio11) ในเขตพื้นที่ภาคเหนือตอนบน

ตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน (Bio12-Bio19) ซึ่งมีหน่วยเป็นมิลลิเมตร (mm) ต่อปี ต่อไตรมาส หรือต่อเดือน ดังแสดงในภาพที่ 3.7 ซึ่งมีรายละเอียดดังนี้

1. Annual Precipitation (Bio12) คือปริมาณน้ำฝนรายปี
2. Precipitation of Wettest Month (Bio13) คือปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด
3. Precipitation of Driest Month (Bio14) คือปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด
4. Precipitation Seasonality (Bio15) คือช่วงฤดูกาลของปริมาณน้ำฝน เป็นดัชนีที่บ่งชี้ถึงความแปรปรวนของปริมาณน้ำฝน
5. Precipitation of Wettest Quarter (Bio16) คือปริมาณน้ำฝนในช่วงไตรมาสที่ฝนตกชุกที่สุด
6. Precipitation of Driest Quarter (Bio17) คือปริมาณน้ำฝนในช่วงไตรมาสที่แห้งแล้งที่สุด
7. Precipitation of Warmest Quarter (Bio18) คือปริมาณน้ำฝนในช่วงไตรมาสที่ร้อนที่สุด
8. Precipitation of Coldest Quarter (Bio19) คือปริมาณน้ำฝนในช่วงไตรมาสที่หนาวที่สุด



ภาพที่ 3.7 ตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน (Bio12-Bio19) ในเขตพื้นที่ภาคเหนือตอนบน

3.2.2 ข้อมูลที่ได้จากการสร้างปัจจัยเพิ่มเติมจากข้อมูลเดิมที่มีอยู่

เป็นการจัดเตรียมข้อมูลปัจจัยต่างๆ เพื่อนำไปสร้างแบบจำลอง โดยเตรียมข้อมูลปัจจัยให้อยู่ในรูปแบบของไฟล์ราสเตอร์ (Raster Files) นามสกุล .tif โดยจัดเตรียมข้อมูล ดังนี้

3.2.2.1 ความลาดชัน (Slope)

จัดเตรียมโดยใช้ข้อมูลแผนที่ลักษณะความสูงของภูมิประเทศแบบดิจิทัล (Digital Elevation Model: DEM) สร้างแผนที่ความลาดชัน ซึ่งมีหน่วยเป็นองศา (Degree) จากนั้นทำการจัดแบ่งประเภทของความลาดชัน ซึ่งจัดแบ่งประเภทของความลาดชันได้ 6 ประเภท (Tang,2000) ดังนี้

ประเภทที่ 1 มีความลาดชัน 0-15 องศา เรียกว่า Gentle Slopes

ประเภทที่ 2 มีความลาดชัน 15-30 องศา เรียกว่า Moderate Slopes

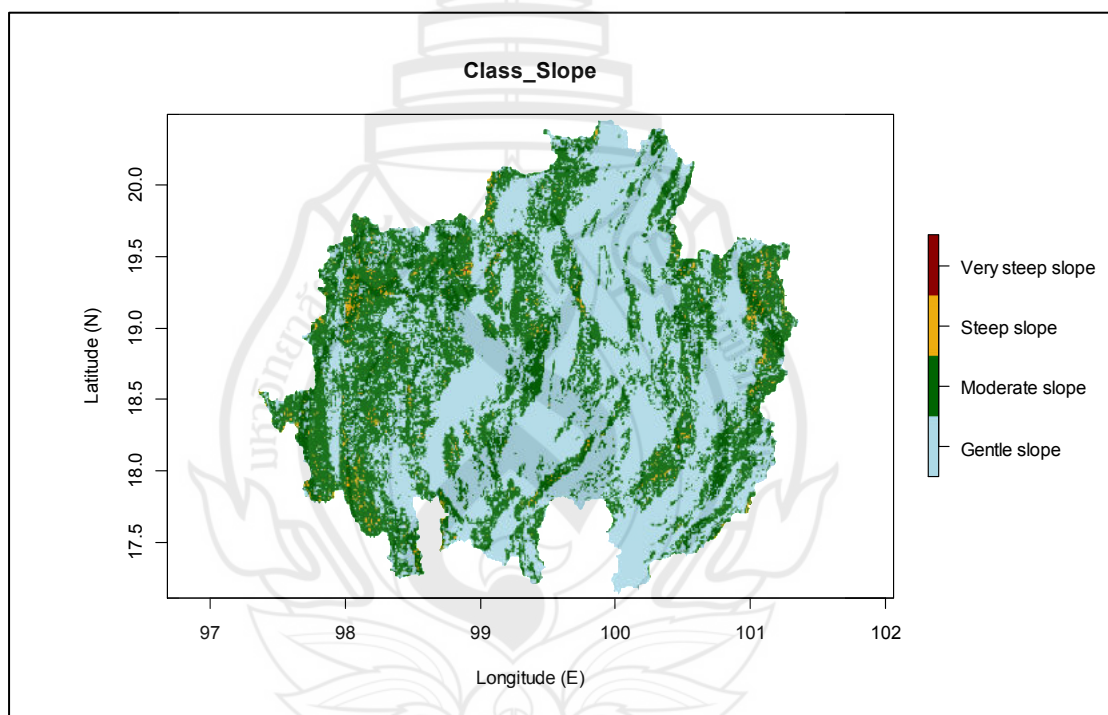
ประเภทที่ 3 มีความลาดชัน 30-45 องศา เรียกว่า Steep Slopes

ประเภทที่ 4 มีความลาดชัน 45-60 องศา เรียกว่า Very Steep Slopes

ประเภทที่ 5 มีความลาดชัน 60-75 องศา เรียกว่า Precipitous Slopes

ประเภทที่ 6 มีความลาดชัน 75-90 องศาหรือมากกว่า 90 องศา เรียกว่า Cliffed Slopes

สำหรับในเขตพื้นที่ภาคเหนือตอนบนของประเทศไทยสามารถแบ่งออกเป็น 4 ประเภทคือ Gentle Slopes, Moderate Slopes, Steep Slopes และ Very Steep Slopes ดังแสดงในภาพที่ 3.8

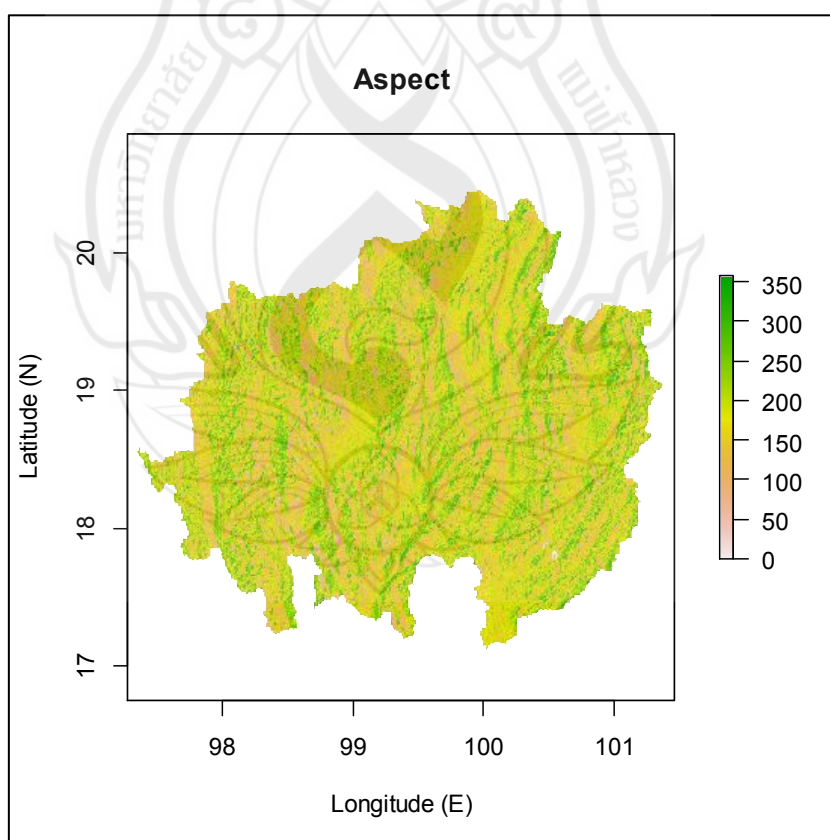


ภาพที่ 3.8 ประเภทของความลาดชันในเขตพื้นที่ภาคเหนือตอนบน

3.2.2.2 ทิศด้านลาด (Aspect)

ทิศด้านลาดเป็นการบ่งบอกทิศทางของพื้นที่ที่มีความลาดชันนั้นๆ ว่าหันไปทางทิศใด โดยใช้ข้อมูลแผนที่ลักษณะความสูงของภูมิประเทศแบบดิจิทัล (Digital Elevation Model: DEM) ในการจัดเตรียมข้อมูลทิศด้านลาด ซึ่งมีหน่วยเป็นองศา ดังแสดงในภาพที่ 3.9 โดยมีรายละเอียดดังนี้

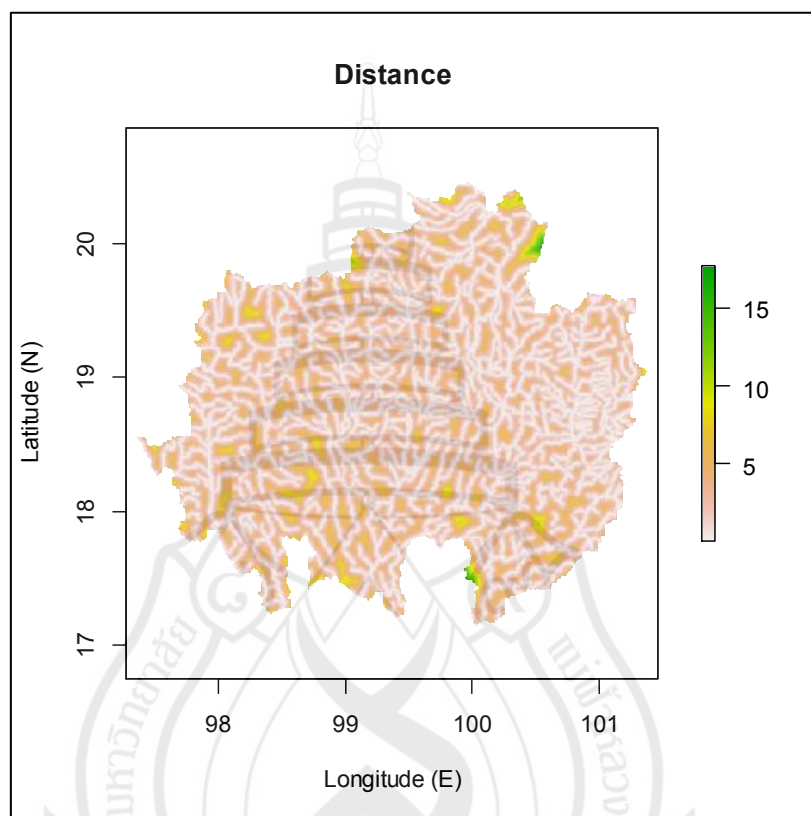
ทิศเหนือ	มีค่าอยู่ระหว่าง 0-22.5 องศา และ 337.5-360 องศา
ทิศตะวันออกเฉียงเหนือ	มีค่าอยู่ระหว่าง 22.5-67.5 องศา
ทิศตะวันออก	มีค่าอยู่ระหว่าง 67.5-112.5 องศา
ทิศตะวันออกเฉียงใต้	มีค่าอยู่ระหว่าง 112.5-157.5 องศา
ทิศใต้	มีค่าอยู่ระหว่าง 157.5-202.5 องศา
ทิศตะวันตกเฉียงใต้	มีค่าอยู่ระหว่าง 202.5-247.5 องศา
ทิศตะวันตก	มีค่าอยู่ระหว่าง 247.5-292.5 องศา
ทิศตะวันตกเฉียงเหนือ	มีค่าอยู่ระหว่าง 292.5-337.5 องศา



ภาพที่ 3.9 ทิศด้านลาดในเขตพื้นที่ภาคเหนือตอนบน

3.2.2.3 ระยะห่างจากแม่น้ำของพื้นที่ (Distance)

ใช้ข้อมูลแม่น้ำ (River) ในการคำนวณเพื่อหาระยะห่างจากแม่น้ำสายหลักและแม่น้ำสาขาในพื้นที่ ดังแสดงในภาพที่ 3.10 มีหน่วยเป็นกิโลเมตร



ภาพที่ 3.10 ระยะห่างจากแม่น้ำสายหลักและแม่น้ำสาขาในเขตพื้นที่ภาคเหนือตอนบน

3.2.2.4 การแบ่งกลุ่มของข้อมูลโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน

ทำการจัดกลุ่มข้อมูลโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความชัน ดังแสดงในภาพที่ 3.11 โดยมีรายละเอียดในแต่ละกลุ่มดังนี้

กลุ่มที่ 1 (Class1) เป็นชนิดหิน S,M และชันความชัน Gentle Slopes

กลุ่มที่ 2 (Class2) เป็นชนิดหิน S,M และชันความชัน Moderate Slopes

กลุ่มที่ 3 (Class3) เป็นชนิดหิน S,M และชันความชัน Steep Slopes

กลุ่มที่ 4 (Class4) เป็นชนิดหิน S,M และชันความชัน Very Steep Slopes

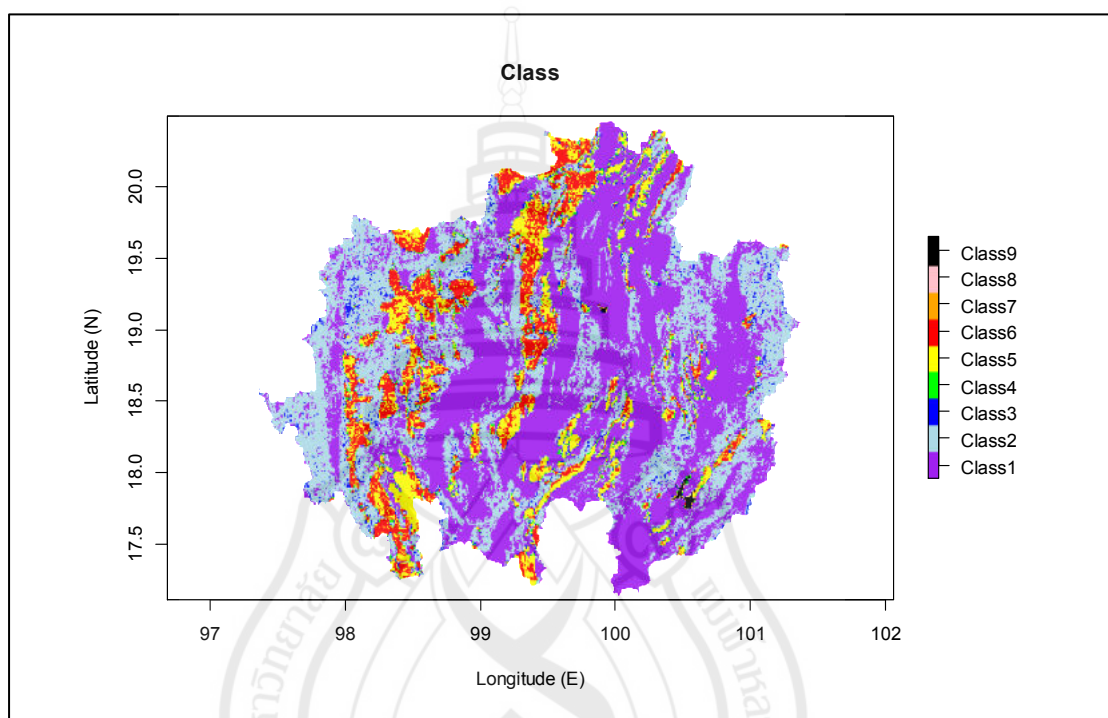
กลุ่มที่ 5 (Class5) เป็นชนิดหิน I และชันความชัน Gentle Slopes

กลุ่มที่ 6 (Class6) เป็นชนิดหิน I และชันความชัน Moderate Slopes

กลุ่มที่ 7 (Class7) เป็นชนิดหิน I และชันความชัน Steep Slopes

กลุ่มที่ 8 (Class8) เป็นชนิดหิน I และชันความชัน Very Steep Slopes

กลุ่มที่ 9 (Class9) เป็นพื้นน้ำ W และชันความชันทุกชั้น



ภาพที่ 3.11 การจัดกลุ่มของโครงสร้างธรณีวิทยาของชนิดหินและประเภทของความลาดชันในเขตพื้นที่ภาคเหนือตอนบน

เนื่องจากตัวแปร Class1-Class9 เป็นข้อมูลเชิงคุณภาพ จึงได้ทำการสร้างตัวแปรหุ่น (Dummy Variables) ดังแสดงในภาพที่ 3.12 โดยกำหนดค่าตัวแปรดังนี้

ตัวแปร Class1 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน S,M และชันความชัน Gentle Slopes
มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชันความชันอื่น ๆ

ตัวแปร Class2 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน S,M และชันความชัน Moderate Slopes
มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชันความชันอื่น ๆ

ตัวแปร Class3 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน S,M และชันความชัน Steep Slopes
มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชันความชันอื่น ๆ

ตัวแปร Class4 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน S,M และชั้นความชัน Very Steep Slopes

มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชั้นความชันอื่น ๆ

ตัวแปร Class5 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน I และชั้นความชัน Gentle Slopes

มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชั้นความชันอื่น ๆ

ตัวแปร Class6 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน I และชั้นความชัน Moderate Slopes

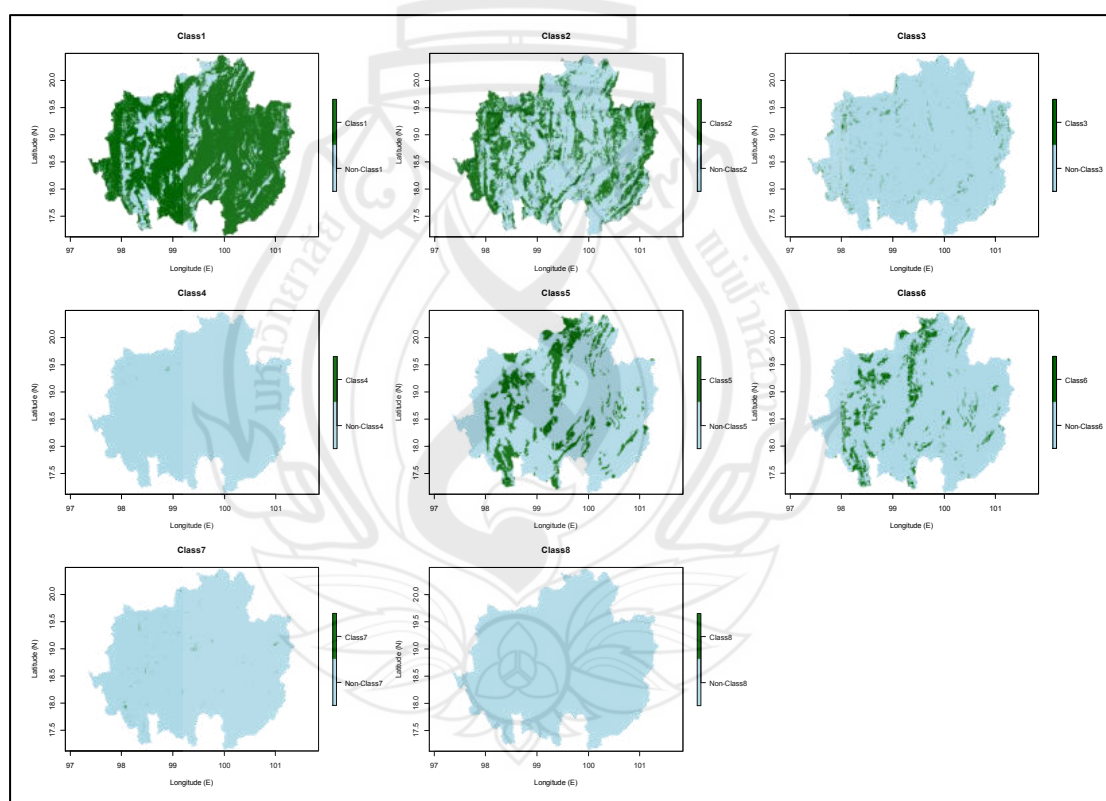
มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชั้นความชันอื่น ๆ

ตัวแปร Class7 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน I และชั้นความชัน Steep Slopes

มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชั้นความชันอื่น ๆ

ตัวแปร Class8 มีค่าเท่ากับ 1 เมื่อเป็นชั้นหิน I และชั้นความชัน Very Steep Slopes

มีค่าเท่ากับ 0 เมื่อเป็นชั้นหินและชั้นความชันอื่น ๆ



ภาพที่ 3.12 ข้อมูลของตัวแปรหุ่น Class1-Class8 ในเขตพื้นที่ภาคเหนือตอนบน

ตัวแปรทุกตัวที่ใช้ในการสร้างแบบจำลองจะอยู่ในรูปแบบของไฟล์ราสเตอร์ (Raster files) นามสกุล .tif โดยมีขนาดของกริด 1 กิโลเมตร $\times$ 1 กิโลเมตร ซึ่งมีตัวแปรดังต่อไปนี้

ตัวแปรทางด้านภูมิศาสตร์ ได้แก่ ตัวแปรความสูงจากระดับน้ำทะเล (DEM) , ระยะห่างจากแม่น้ำ (Distance), ทิศด้านลาด (Aspect) และตัวแปรการแบ่งกลุ่มของข้อมูลโครงสร้างทางธรณีวิทยาของหินและความลาดชัน (Class1-Class8)

ตัวแปรทางด้านสภาพภูมิอากาศ ได้แก่ ความชื้นสัมพัทธ์ (Humidity), ตัวแปรทางด้าน Bioclimatic (Bio1-Bio19)

3.3 การสร้างชุดข้อมูลสำหรับการสร้างแบบจำลอง

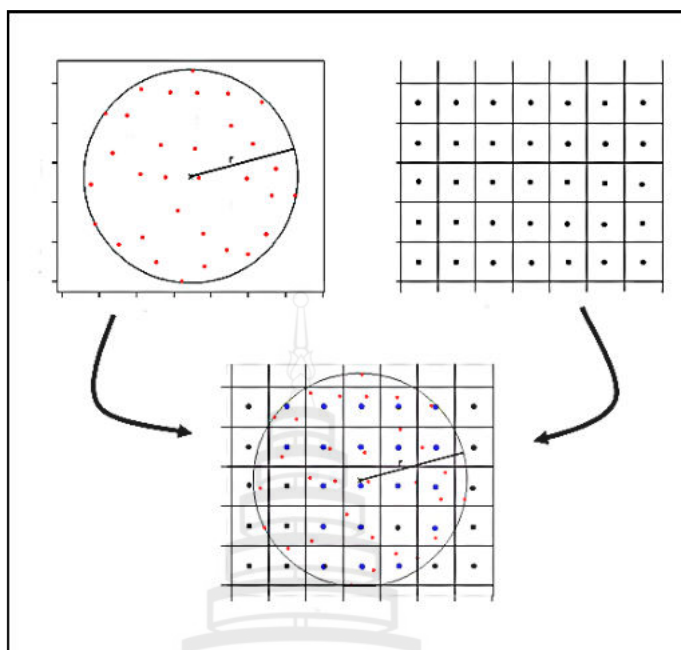
เป็นการเตรียมชุดข้อมูลเพื่อใช้ในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต ได้แก่ วิธีการสุ่มจุด Pseudo-Absence และการสร้างชุดข้อมูลเพื่อใช้สำหรับการสร้างแบบจำลองจากจุดพิกัดที่มีการค้นพบต้นชาป่า (Presence Points) และจุด Pseudo-Absence

3.3.1 การสุ่มจุด Pseudo-absence points

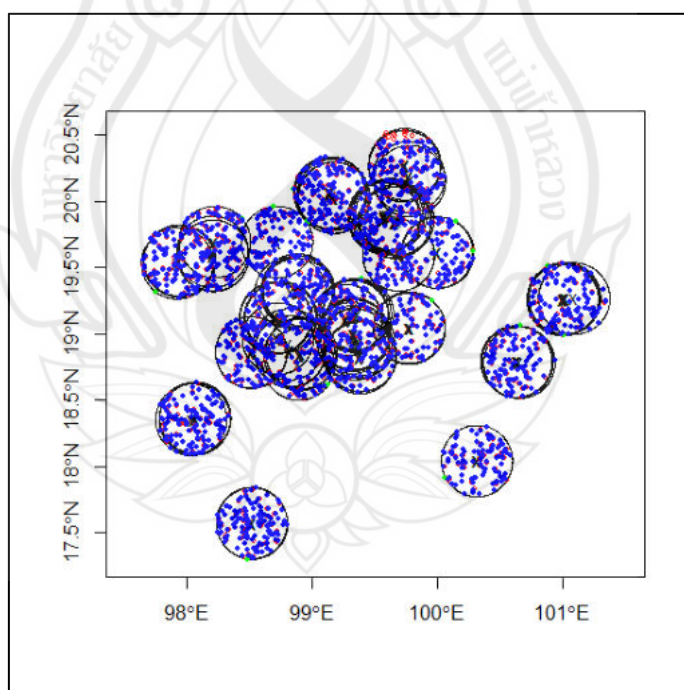
ในการศึกษาครั้งนี้ได้สร้างชุดข้อมูล Pseudo-Absence Points โดยใช้เทคนิคในการสร้าง 3 วิธีด้วยกัน ดังนี้

3.3.1.1 การเลือกจุด Pseudo-Absence โดยการสุ่มในพื้นที่ที่กำลังศึกษา

ทำการสุ่มจุด Pseudo-absence ภายในพื้นที่รอบๆ จุดที่มีการค้นพบสิ่งมีชีวิต (Presence Points) ในรัศมีที่แตกต่างกัน (VanDerWal, Shoo, Graham & Williams, 2009) ดังแสดงในภาพที่ 3.13 และ 3.14 โดยภาพที่ 3.13 จะแสดงหลักการการสุ่มจุด Pseudo-Absence โดยการซ้อนทับกันของแผนที่ 2 แผนที่ (Hijmans & Elith, 2011) โดยแผนที่แรกจะทำการสุ่มจุดภายในพื้นที่วงกลม (จุดสีแดง) ส่วนแผนที่ที่สองจะทำการสร้างกริดสี่เหลี่ยมจัตุรัส (Grid Cells) โดยจุดสีแดงเป็นจุดที่อยู่ตรงกลางของกริด เมื่อนำมาซ้อนทับกันแล้ว จะเลือกใช้กริดที่มีจุดสีแดงและไม่ได้อยู่นอกขอบเขตวงกลม ซึ่งกริดที่ถูกเลือกนั้น จุดกึ่งกลางกลางของกริด (จุดสีน้ำเงิน) จะถูกนำมาใช้เป็นจุด Pseudo-Absence และภาพที่ 3.14 แสดงตัวอย่างการสุ่มจุด Pseudo-Absence ที่รัศมี 30 กิโลเมตรและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร โดยจุดสีน้ำเงินเป็นจุด Pseudo-Absence



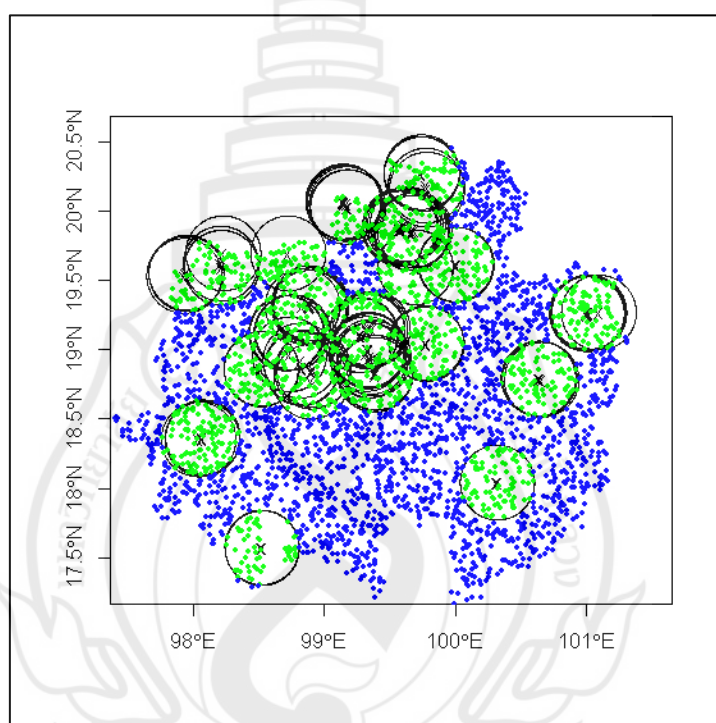
ภาพที่ 3.13 หลักการการสุ่มจุด Pseudo-Absence



ภาพที่ 3.14 ตัวอย่างการสุ่มจุด Pseudo-Absence โดยการสุ่มในพื้นที่ที่กำลังศึกษาที่รัศมี 30 กิโลเมตรและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร

3.3.1.2 การเลือกจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence

เป็นการเลือกจุด Pseudo-Absence ที่คล้ายกับแบบที่หนึ่งแต่ใช้แนวคิดที่ว่าพื้นที่โดยรอบจุด Presence เป็นพื้นที่ที่มีความเหมาะสมต่อการกระจายของสิ่งมีชีวิต ดังนั้นจึงทำการสุ่มเลือกจุด Pseudo-Absence ภายนอกวงกลมในรัศมีที่แตกต่างกัน ดังแสดงในภาพที่ 3.15 ซึ่งแสดงตัวอย่างการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ที่รัศมี 30 กิโลเมตร และมีกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร โดยจุดสีน้ำเงินที่อยู่ภายนอกวงกลมจะเป็นจุด Pseudo-Absence

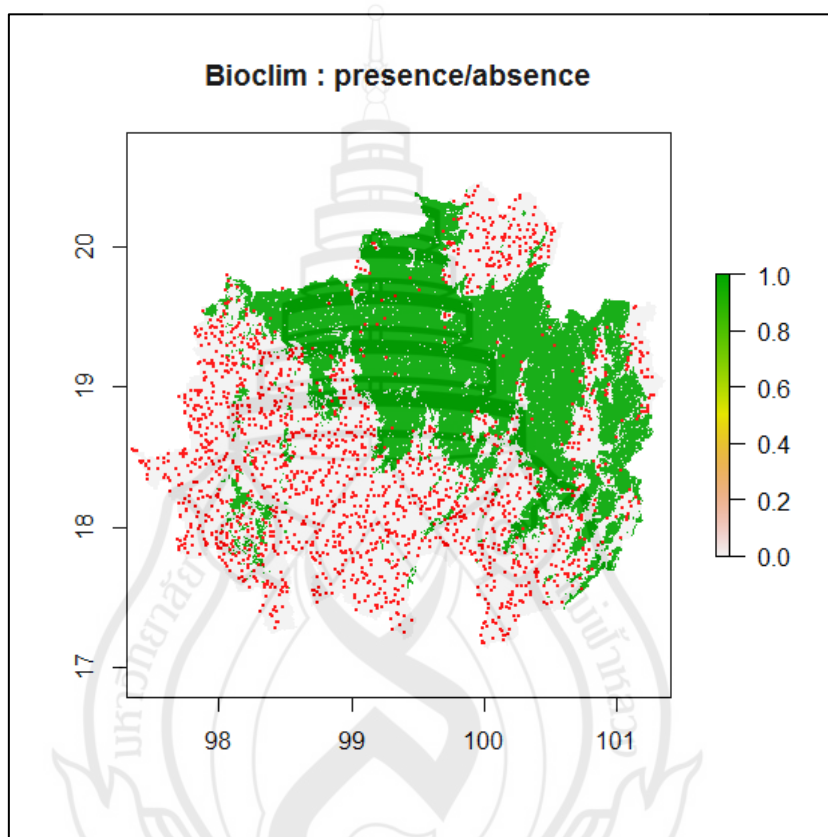


ภาพที่ 3.15 ตัวอย่างการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ที่รัศมี 30 กิโลเมตรและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร

3.3.1.3 การเลือกจุด Pseudo-Absence แบบ 2 ขั้นตอน

เป็นการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Barbet-Massiet et al., 2012; Lobo & Tognelli, 2011; Wisz & Guisan, 2009) โดยขั้นตอนแรกเป็นการสร้างพื้นที่ที่เหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิตโดยเลือกใช้เทคนิคโปรไฟล์ในการสร้างพื้นที่ที่เหมาะสม จากนั้นในขั้นตอนที่สองเป็นการสุ่มเลือกจุด Pseudo-Absence ภายนอกพื้นที่ที่เหมาะสม ซึ่งในการศึกษารั้งนี้ได้เลือกใช้วิธี BioClim ในการในการสร้างพื้นที่ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าใน

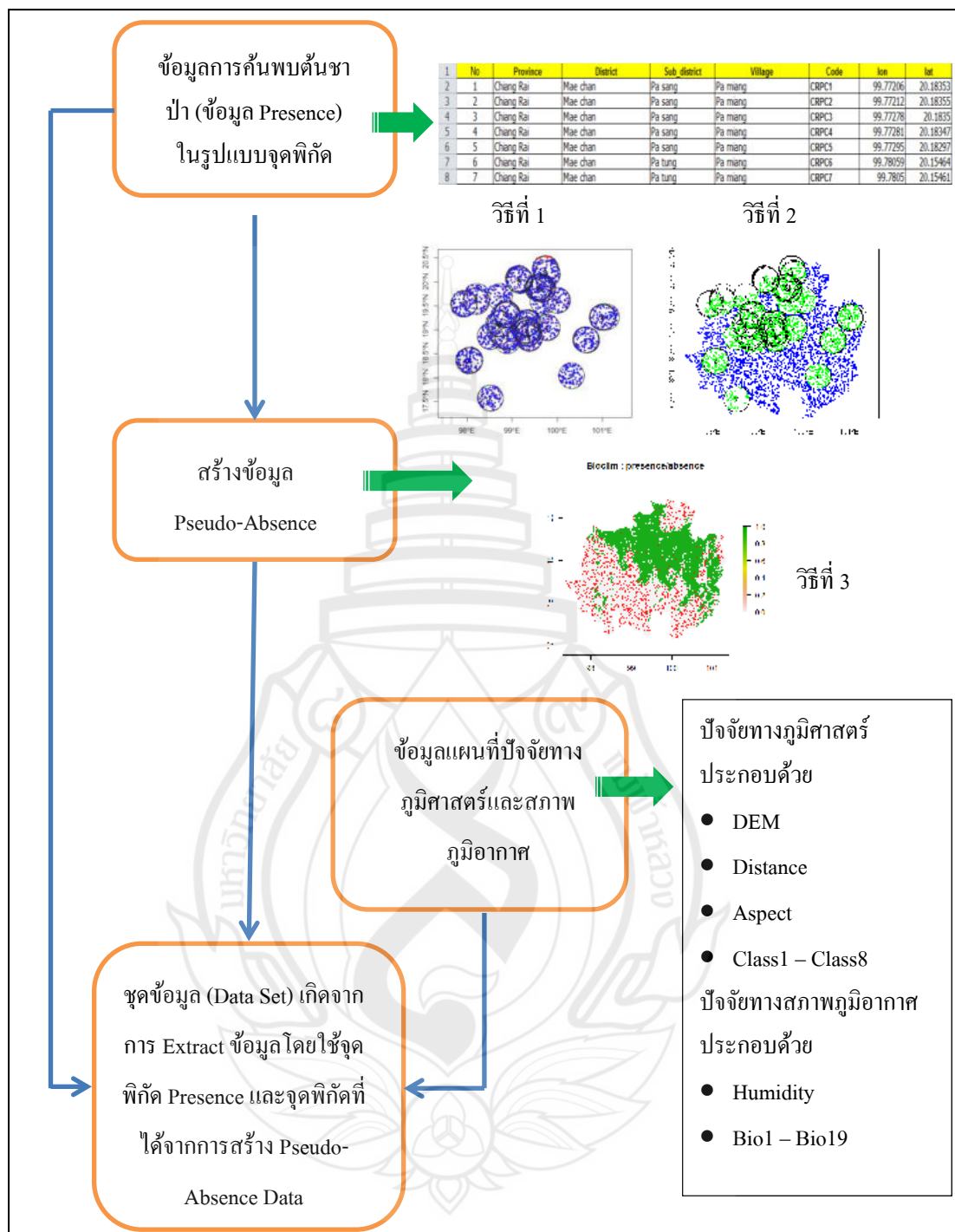
ขั้นตอนแรก โดยการสุ่มใช้หลักการเดียวกันกับวิธีที่ 1 ดังแสดงในภาพที่ 3.16 ซึ่งแสดงตัวอย่างการสุ่มจุด Pseudo-absence แบบสองขั้นตอนที่กริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร โดยพื้นที่สีเขียวซึ่งสร้างโดยวิธีการ BioClim เป็นพื้นที่ที่คาดว่ามีความเหมาะสมต่อการเจริญเติบโตของสิ่งมีชีวิต จุดสีแดงที่อยู่ภายนอกพื้นที่สีเขียวจะถูกจัดให้เป็นจุด Pseudo-Absence



ภาพที่ 3.16 ตัวอย่างการสุ่มจุด Pseudo-Absence แบบสองขั้นตอนและกริดขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร

3.3.2 การสร้างชุดข้อมูล (Data set)

จากภาพที่ 3.17 แสดงขั้นตอนการสร้างชุดข้อมูล (Data Set) โดยเริ่มจากข้อมูล Presence ซึ่งอยู่ในรูปแบบของจุดพิกัดทางภูมิศาสตร์ จะถูกนำไปใช้ในการสร้างชุดข้อมูล Pseudo-Absence โดยวิธีการทั้ง 3 วิธี จากนั้นก็นำชุดข้อมูลจุดพิกัดของ Presence และ Pseudo-Absence ไปสกัดข้อมูลแผนที่ทางภูมิศาสตร์ (DEM, Distance, Aspect, Class1-Class8) และสภาพภูมิอากาศ (Humidity, Bio1-Bio19) ก็จะได้ชุดข้อมูล (Data Set) ที่จะถูกนำไปใช้ในขั้นตอนต่อไป



ภาพที่ 3.17 ขั้นตอนการสร้างชุดข้อมูล (Data Set)

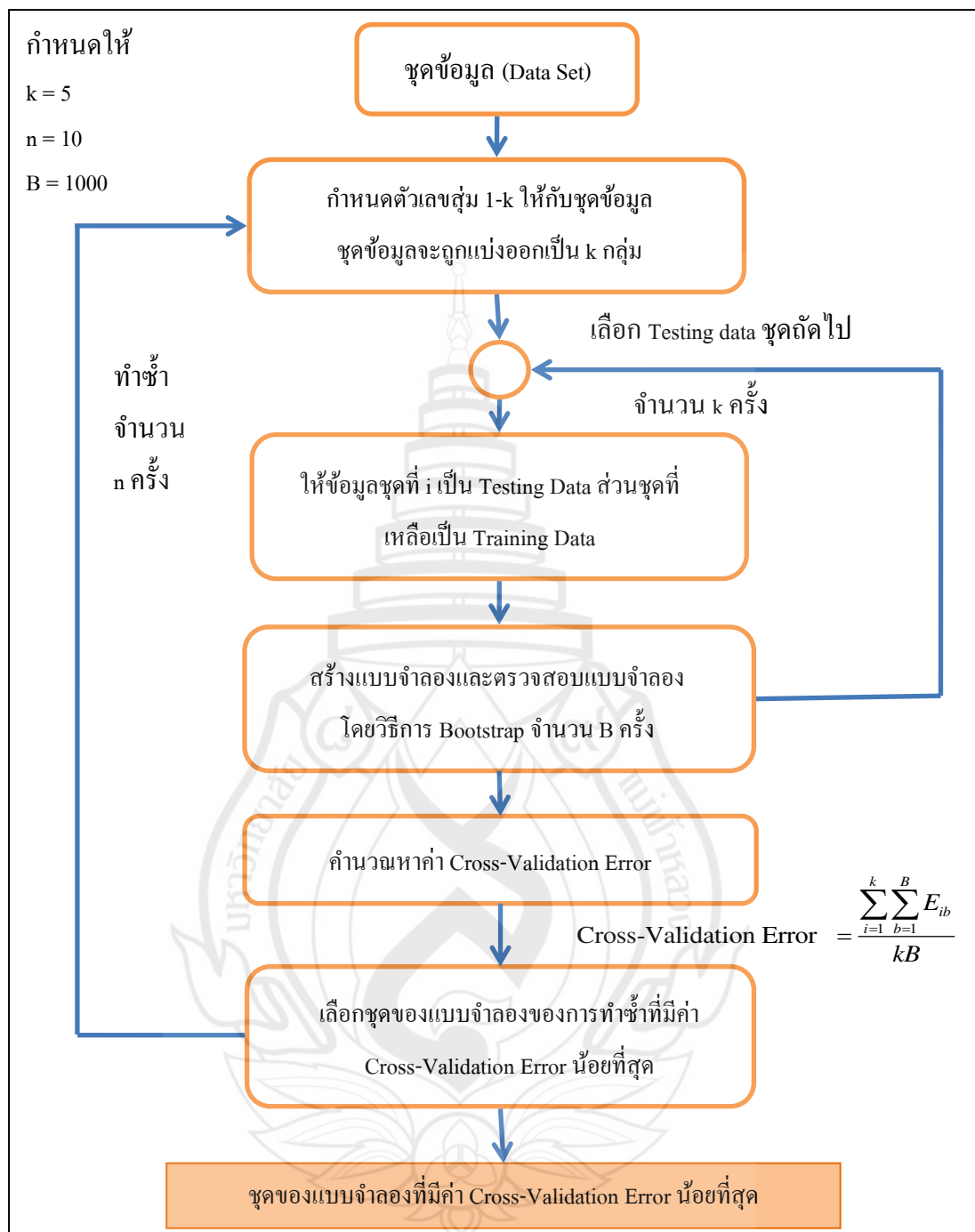
3.4 กระบวนการในการสร้างแบบจำลอง การตรวจสอบ และการพยากรณ์

3.4.1 กระบวนการในการสร้างแบบจำลองและการตรวจสอบ

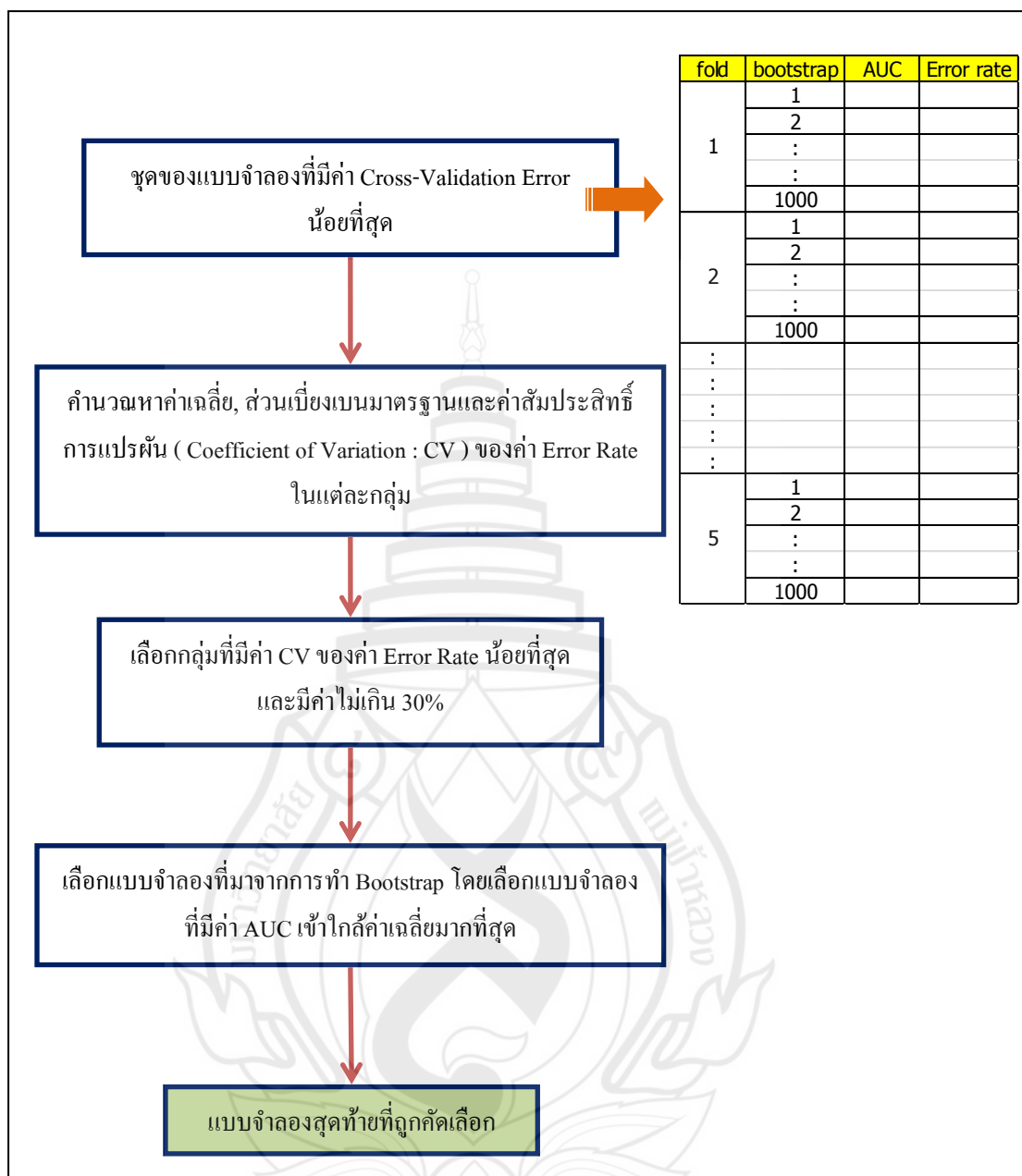
จากภาพที่ 3.18 แสดงขั้นตอนการสร้างแบบจำลองโดยนำวิธี Bootstrap และวิธี k-Fold Cross-Validation มาประยุกต์ใช้ โดยวิธี Bootstrap เป็นเทคนิคการสุ่มซ้ำ (Resampling) โดยการสุ่มตัวอย่างแบบใส่กลับคืน (Resampling With Replacement) เป็นการเพิ่มขนาดของกลุ่มตัวอย่างที่จะนำมาใช้ในการสร้างแบบจำลอง ส่วนวิธี k-Fold Cross-Validation เป็นวิธีที่ใช้ในการตรวจสอบและเปรียบเทียบแบบจำลองเพื่อนำไปสู่แบบจำลองที่ดีที่สุด

เมื่อได้ชุดข้อมูล (Data Set) แล้ว ข้อมูลจะถูกกำหนดตัวเลขอย่างสุ่ม 1-5 ทำให้ข้อมูลถูกแบ่งออกเป็น 5 กลุ่มอย่างละเท่า ๆ กัน หรือใกล้เคียงกัน กำหนดให้ข้อมูลกลุ่มที่ 1 เป็นชุดข้อมูลในส่วนของ Testing Data เอาไว้สำหรับในการตรวจสอบความถูกต้องในการพยากรณ์ของแบบจำลอง ส่วนข้อมูลในกลุ่มที่ 2, 3, 4 และ 5 จะอยู่ในส่วนของ Training Data เอาไว้สำหรับในการสร้างแบบจำลอง Training Data จะถูกนำมาทำการสุ่มซ้ำโดยวิธีการ Bootstrap จำนวน 1000 ครั้ง ซึ่งจะได้ชุดข้อมูล Training Data ใหม่จำนวน 1000 ชุด ชุดข้อมูล Training Data ใหม่ในแต่ละชุดจะถูกนำมาสร้างแบบจำลองและถูกตรวจสอบความถูกต้องในการพยากรณ์โดยชุดข้อมูล Testing Data โดยคำนวณค่า Area Under Receiver Operating Characteristic Curve (AUC) ซึ่งเป็นค่าที่แสดงความถูกต้องในการจำแนก (Discriminatory Accuracy) ลักษณะที่สนใจและค่าอัตราส่วนในการทำนายผิด (Error Rate) จากนั้นย้อนกลับไปกำหนด Testing Data ชุดใหม่ โดยให้ข้อมูลกลุ่มที่ 2 เป็นชุดข้อมูลในส่วนของ Testing Data ส่วนข้อมูลในกลุ่มที่ 1, 3, 4 และ 5 จะอยู่ในส่วนของ Training Data แทน ทำตามขั้นตอนทุกกระบวนการซ้ำอีกครั้ง จนกำหนดชุด Testing Data ครบทั้ง 5 กลุ่ม จะได้แบบจำลองพร้อมค่า AUC และ Error Rate จำนวน 5000 ชุด จากนั้นคำนวณค่า Cross-Validation Error ซึ่งเป็นค่าเฉลี่ยของ Error Rate ทั้งหมด 5000 ค่า

จากนั้นย้อนกลับไปทำซ้ำโดยการกำหนดตัวเลขอย่างสุ่ม 1-5 ขึ้นมาใหม่ และทำตามขั้นตอนเดิมทุกประการ ทำซ้ำอีก 10 ครั้งจะได้ค่า Cross-Validation Error มาทั้งหมด 10 ค่า ในการเลือกชุดแบบจำลองจากการทำซ้ำ 10 ครั้งจะเลือกชุดแบบจำลองที่มีค่า Cross-Validation Error น้อยที่สุดเอามาเป็นตัวแทนเพื่อนำไปหาแบบจำลองในขั้นตอนต่อไป



ภาพที่ 3.18 ขั้นตอนการสร้างแบบจำลองโดยนำวิธี Bootstrap และ วิธี k-Fold Cross-Validation มาประยุกต์ใช้



ภาพที่ 3.19 ขั้นตอนการคัดเลือกแบบจำลองจากชุดของแบบจำลองที่มีค่า Cross-Validation Error น้อยที่สุด

จากภาพที่ 3.19 แสดงขั้นตอนการคัดเลือกแบบจำลอง เมื่อได้ชุดแบบจำลองที่มีค่า Cross-Validation Error น้อยที่สุดแล้ว ชุดแบบจำลองในกลุ่มที่ 1 คือชุดแบบจำลองที่เกิดจากการใช้กลุ่มข้อมูลกลุ่มที่ 1 เป็นกลุ่ม Testing Data กลุ่มข้อมูลกลุ่มที่ 2, 3, 4 และ 5 เป็น Training Data ชุดแบบจำลองในกลุ่มที่ 2 คือชุดแบบจำลองที่เกิดจากการใช้กลุ่มข้อมูลกลุ่มที่ 2 เป็นกลุ่ม Testing Data กลุ่มข้อมูลกลุ่มที่ 1, 3, 4 และ 5 เป็น Training Data ชุดแบบจำลองในกลุ่มที่ 3 คือชุดแบบจำลองที่เกิดจากการใช้กลุ่มข้อมูลกลุ่มที่ 3 เป็นกลุ่ม Testing Data กลุ่มข้อมูลกลุ่มที่ 1, 2, 4 และ 5 เป็น Training Data ชุดแบบจำลองในกลุ่มที่ 4 คือชุดแบบจำลองที่เกิดจากการใช้กลุ่มข้อมูลกลุ่มที่ 4 เป็นกลุ่ม Testing Data กลุ่มข้อมูลกลุ่มที่ 1, 2, 3 และ 5 เป็น Training Data และสุดท้ายชุดแบบจำลองในกลุ่มที่ 5 คือชุดแบบจำลองที่เกิดจากการใช้กลุ่มข้อมูลกลุ่มที่ 5 เป็นกลุ่ม Testing Data กลุ่มข้อมูลกลุ่มที่ 1, 2, 3 และ 4 เป็น Training Data ในแต่ละกลุ่มแบบจำลองจะมีแบบจำลองพร้อมทั้งค่า AUC และ Error Rate 1000 ชุด คำนวณหาค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน และค่าสัมประสิทธิ์การแปรผัน (Coefficient of Variation : CV) ของค่า Error Rate ในแต่ละกลุ่มทำการเลือกกลุ่มที่มีค่าสัมประสิทธิ์การแปรผันที่น้อยที่สุดและไม่ควรมีค่าเกิน 30%

เมื่อได้กลุ่มแบบจำลองที่มีค่าสัมประสิทธิ์การแปรผันที่น้อยที่สุดแล้ว ทำการคำนวณหาค่าเฉลี่ยของ AUC ทั้ง 1000 ค่า จากนั้นหาแบบจำลองใน 1000 ชุดที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด แบบจำลองมีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุดจะเป็นแบบจำลองสุดท้ายที่ถูกคัดเลือกแล้ว ดังนั้น ชุดข้อมูล Pseudo-Absence จำนวน 1 ชุด จะได้แบบจำลองที่ถูกคัดเลือกแล้ว 1 แบบจำลอง

ในการสร้างชุดข้อมูล Pseudo-Absence โดยวิธีที่ 1 และ 2 จะกำหนดวงกลมที่มีรัศมีตั้งแต่ 10, 15, 20, 25, 30, 35, 40, 45 และ 50 กิโลเมตร โดยในแต่ละรัศมีจะสร้างชุดข้อมูล Pseudo-Absence จำนวน 20 ชุด ส่วนวิธีที่ 3 จะสร้างแผนที่พื้นที่ที่เหมาะสมโดยวิธี Bioclim จำนวน 5 แผนที่ โดยในแต่ละแผนที่จะสร้างชุดข้อมูล Pseudo-Absence จำนวน 20 ชุดเช่นเดียวกัน

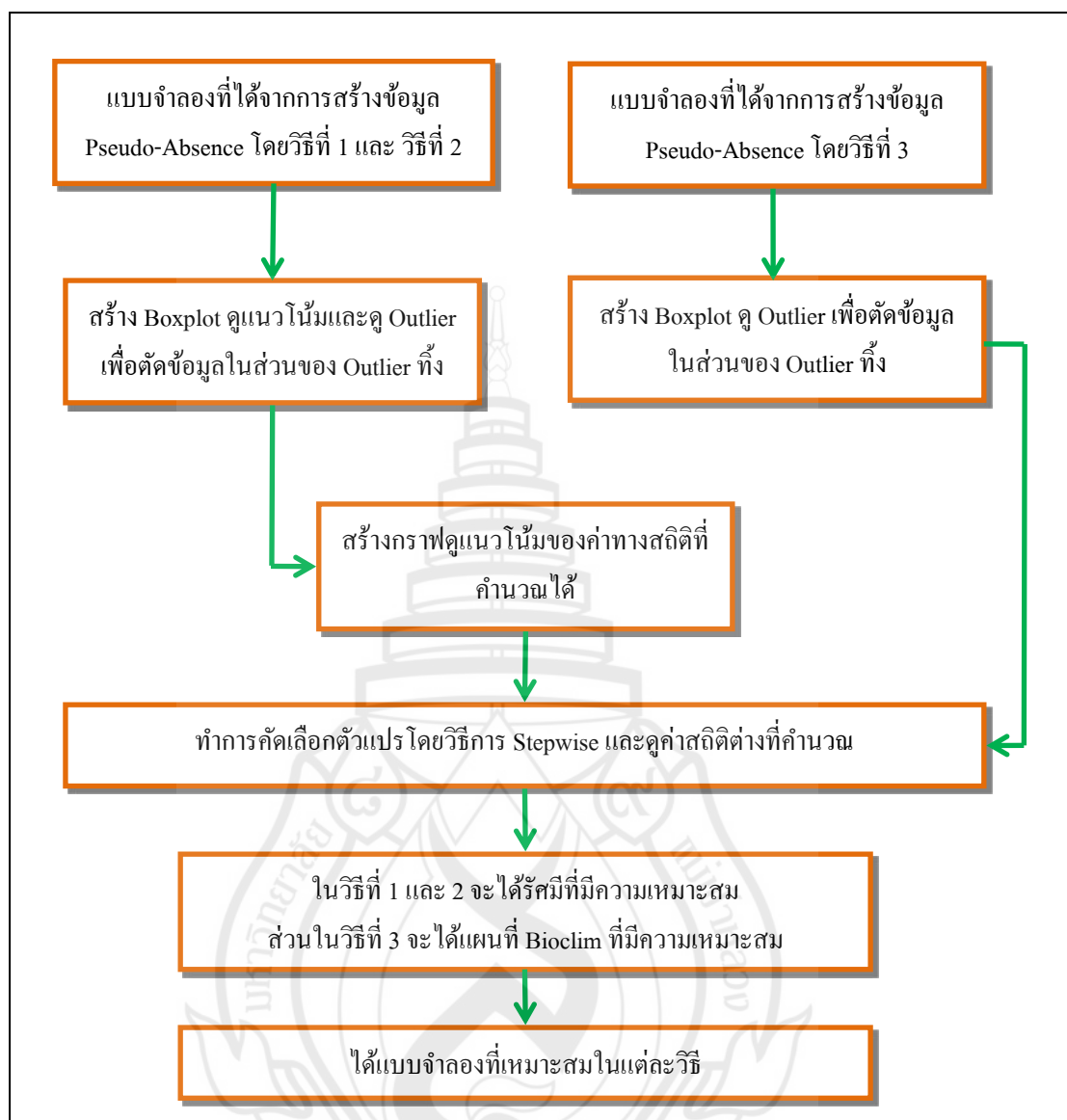
3.4.2 วิธีการหาแบบจำลองที่เหมาะสมและการพยากรณ์แบบจำลอง

จากภาพที่ 3.20 แสดงวิธีการหาแบบจำลองที่เหมาะสมในแต่ละวิธีในการสร้างชุดข้อมูล Pseudo-Absence โดยเริ่มจากนำแบบจำลองที่ถูกคัดเลือกได้โดยวิธีที่ 1 และวิธีที่ 2 นำมาสร้าง Boxplot เพื่อดูแนวโน้มและดูค่าผิดปกติ (Outlier Value) โดยเปรียบเทียบในแต่ละรัศมี รัศมีใดที่มีค่า Outlier กลุ่มตัวอย่างนั้นก็จะถูกตัดทิ้ง จากนั้นหาช่วงรัศมีที่เหมาะสมจากเส้นกราฟแนวโน้มที่ได้ เมื่อได้ช่วงของรัศมีที่เหมาะสมแล้ว จะทำการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้วนำค่าสถิติต่าง ๆ ที่คำนวณได้มาเปรียบเทียบเพื่อหารัศมีที่เหมาะสมในแต่ละวิธีต่อไป ส่วนแบบจำลองที่ถูกคัดเลือกได้โดยวิธีที่ 3 จะถูกนำมาสร้าง Boxplot เพื่อดูค่าผิดปกติ (Outlier Value) เช่นเดียวกัน จากนั้นทำการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้วนำค่าสถิติต่าง ๆ ที่คำนวณได้มา

เปรียบเทียบเพื่อหาแผนที่ Bioclim ที่มีความเหมาะสมต่อไป จากนั้นในแต่ละวิธีจะคำนวณหา ค่าเฉลี่ยของ AUC เพื่อหาแบบจำลองที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด โดยแบบจำลองที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุดจะเป็นแบบจำลองที่มีความเหมาะสมในแต่ละวิธีการสร้างชุดข้อมูล Pseudo-Absence

ในขั้นตอนแรกของการสร้างแบบจำลองจะใช้แบบจำลองการกระจายของสิ่งมีชีวิตโดย วิธีการ Generalized Linear Models (GLMs) ก่อน เมื่อได้แบบจำลองที่มีความเหมาะสมในแต่ละ วิธีการสร้างชุดข้อมูล Pseudo-Absence แล้วจะนำชุด Training Data และ Testing Data ของ แบบจำลองที่ได้มาทำการสร้างแบบจำลองใหม่โดยใช้วิธีการ Generalized Additive Models (GAMs) และวิธีการ Maximum Entropy (MAXENT) ต่อไป

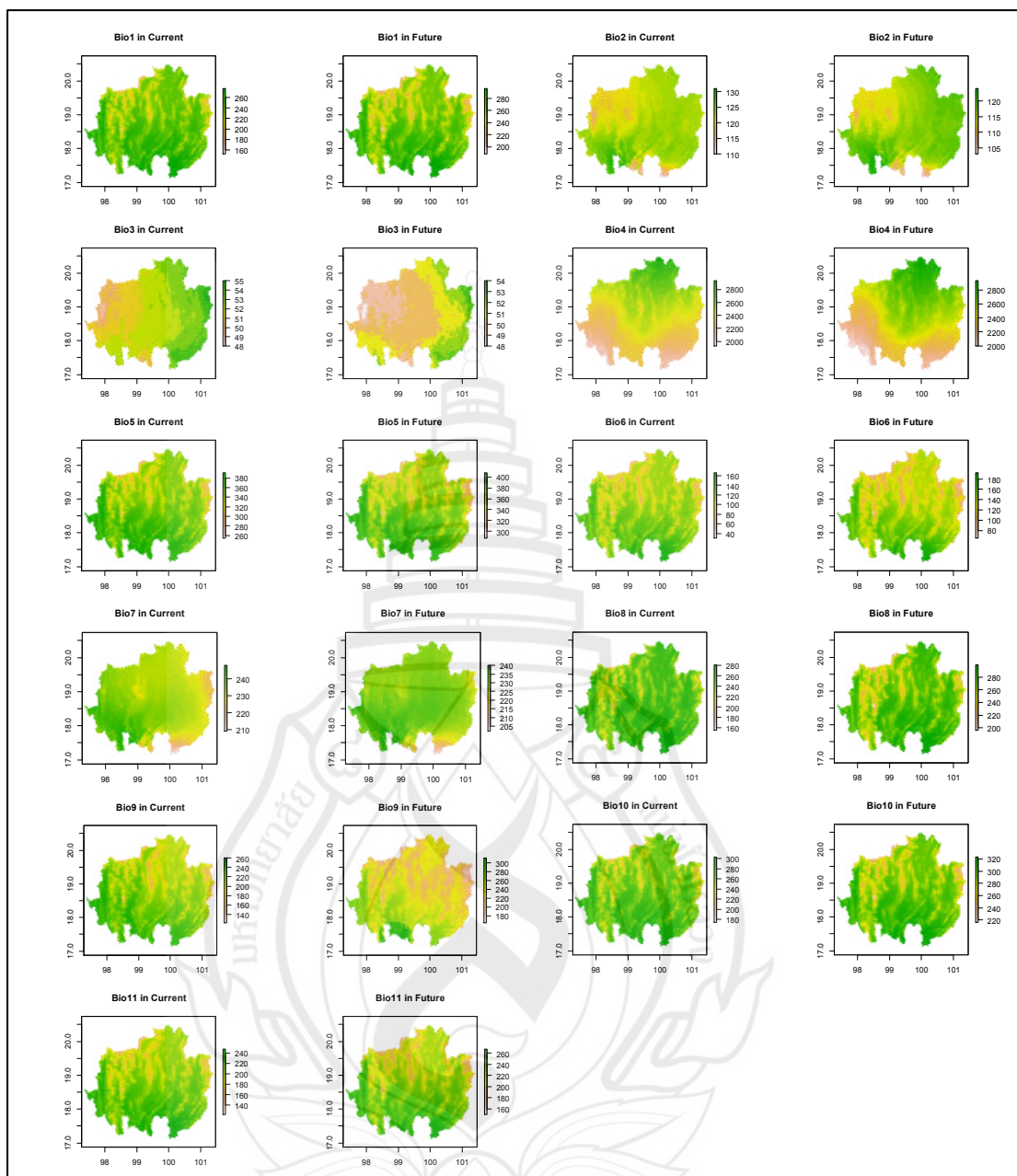
ในส่วนของการพยากรณ์ เมื่อได้แบบจำลองที่มีความเหมาะสมแล้วจะนำมาสร้างแผนที่ พยากรณ์เพื่อหาพื้นที่การกระจายที่มีความเหมาะสมต่อไป เนื่องจากการศึกษาการกระจายของ ต้นชาป่า ดังนั้นในขั้นตอนสุดท้ายของการพยากรณ์จะพิจารณาเฉพาะพื้นที่ที่เป็นพื้นที่ป่าเท่านั้น จากนั้นเปรียบเทียบแผนที่ในแต่ละวิธีว่ามีความแตกต่างกันหรือไม่โดยใช้วิธี Cochran's Q Test ในการทดสอบ



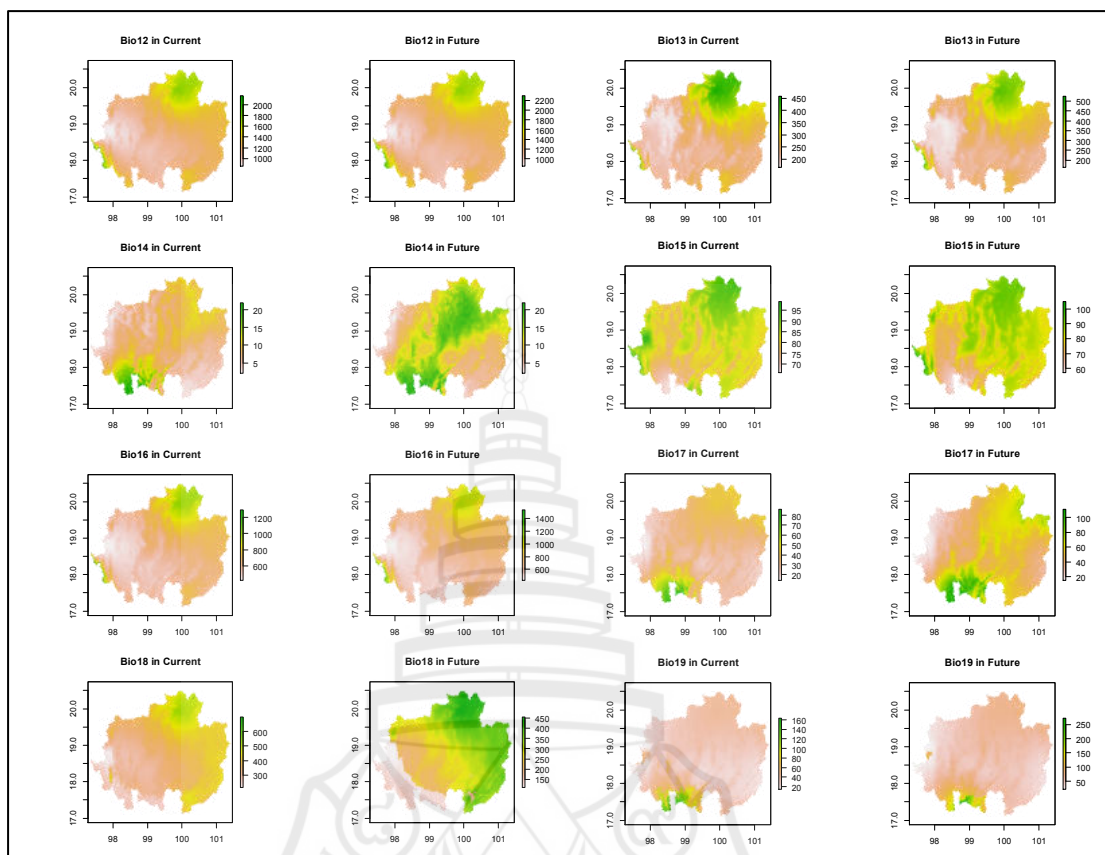
ภาพที่ 3.20 วิธีการหาแบบจำลองที่เหมาะสมในแต่ละวิธีของการสร้างชุดข้อมูล Pseudo-Absence

3.5 การเปรียบเทียบผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศ (Climate Change)

ในขั้นตอนนี้จะทำการเปรียบเทียบผลกระทบต่อการเปลี่ยนแปลงสภาพภูมิอากาศ โดยการเปลี่ยนตัวแปรสภาพภูมิอากาศในปัจจุบัน (ข้อมูลอุณหภูมิตั้งแต่ปี ค.ศ. 1950-2000) เป็นสภาพภูมิอากาศในปี ค.ศ. 2050 เพื่อศึกษาผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศของแบบจำลองที่ได้ในแต่ละวิธีต่อไป โดยตัวแปรสภาพภูมิอากาศในปี ค.ศ. 2050 (ค่าเฉลี่ยระหว่างปี ค.ศ. 2041-2060) เป็นข้อมูลออนไลน์จาก http://www.worldclim.org/cmip5_30s#com สร้างโดยแบบจำลอง ACCESS1-0 Scenario rcp45 ซึ่งเป็นการคาดการณ์สภาพภูมิอากาศจากแบบจำลองสภาพภูมิอากาศโลก (Global Climate Models: GCMs) โดยใช้ภาพฉายการปล่อยก๊าซเรือนกระจกแบบใหม่ (New emission scenarios) หรือ Representative Concentration Pathways (RCP) ในรายงานการประเมินฉบับที่ 5 (IPCC's Fifth Assessment Report: AR5) โดยคณะกรรมการระหว่างรัฐบาลว่าด้วยการเปลี่ยนแปลงภูมิอากาศ (Intergovernmental Panel on Climate Change : IPCC) Work group 1 ซึ่งถือเอาความเข้มข้นของก๊าซเรือนกระจกเป็นจุดเริ่มต้นแล้วประเมินว่าที่ความเข้มข้นของก๊าซเรือนกระจกในระดับต่างๆ จะกระทบต่อการเปลี่ยนแปลงภูมิอากาศและกระบวนการที่เกี่ยวข้องอย่างไรบ้างในการศึกษาครั้งนี้เลือกซีเนริโอ (Scenario) RCP4.5 หมายถึงสภาพโลกอนาคตที่ต้องการให้พลังงานถูกกักอยู่ในบรรยากาศคงที่หรือไม่เพิ่มไปกว่า 4.5 วัตต์ต่อตารางเมตร ดังแสดงในภาพที่ 3.21 และภาพที่ 3.22



ภาพที่ 3.21 ตัวแปรภูมิอากาศที่เกี่ยวข้องกับอุณหภูมิ (Bio1-Bio11) ในปัจจุบันเปรียบเทียบกับปี ค.ศ. 2050



ภาพที่ 3.22 ตัวแปรภูมิอากาศที่เกี่ยวข้องกับปริมาณน้ำฝน (Bio12-Bio19) ในปัจจุบันเปรียบเทียบกับปีค.ศ. 2050

3.6 การวิเคราะห์ข้อมูลโดยวิธีทางสถิติ

3.6.1 Cochran's Q test

การทดสอบความแตกต่างของแผนที่ยากรณ์ว่าในแต่ละวิธีของการสร้างแบบจำลองการกระจายของต้นชาป่ามีแผนที่พยากรณ์แตกต่างกันหรือไม่ และมีวิธีไหนบ้างที่มีแผนที่พยากรณ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยเลือกใช้วิธี Cochran's Q test

Cochran's Q test ใช้สำหรับทดสอบกลุ่มตัวอย่างหลายกลุ่มที่สัมพันธ์กันโดยการจับคู่และให้ได้รับการจัดกระทำที่แตกต่างกัน (Treatment) จากนั้นเปรียบเทียบผลที่เกิดขึ้นว่ามีความแตกต่างกันหรือไม่ หรือใช้สำหรับทดสอบกลุ่มตัวอย่างเพียงกลุ่มเดียวแต่ได้รับการจัดกระทำ (Treatment) ที่มีความแตกต่างกัน แล้วเปรียบเทียบว่าผลแตกต่างกันหรือไม่ โดยข้อมูลที่ได้จากการวัดอยู่ในมาตรานามบัญญัติที่จำแนกข้อมูลออกเป็นกลุ่มเป็นพวกได้ 2 พวก เช่น ใช่-ไม่ใช่ ชอบ-ไม่ชอบ สอบได้-สอบตก ทำถูก-ทำผิด และกำหนดให้ค่าของข้อมูลเป็น 1 และ 0 ซึ่งการศึกษาครั้งนี้ให้ 1 แทนพื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่าและ 0 แทนพื้นที่ที่คาดว่าจะไม่มีการกระจายของต้นชาป่า โดยมีสูตรในการคำนวณค่าสถิติ Q ดังนี้

$$Q = \frac{(k-1) \left[k \sum_{j=1}^k G_j^2 - \left(\sum_{j=1}^k G_j \right)^2 \right]}{k \sum_{i=1}^N L_i^2 - \left(\sum_{i=1}^N L_i \right)^2}$$

โดย k คือ จำนวนการจัดกระทำที่แตกต่างกัน (Treatment) ในที่นี้คือจำนวนแผนที่พยากรณ์ที่ต้องการนำมาเปรียบเทียบกัน

G_j คือ จำนวนกลุ่มตัวอย่างที่อยู่ในกลุ่ม j ที่ตอบรับหรือประสบความสำเร็จ ในที่นี้คือจำนวนพิกเซลที่อยู่ในกลุ่ม j โดยเป็นพื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า หรือมีค่าเท่ากับ 1

L_i คือ จำนวนกลุ่มตัวอย่างที่อยู่ในแถวที่ i ที่ตอบรับหรือประสบความสำเร็จ ในที่นี้คือจำนวนพิกเซลที่อยู่ในแถว i ที่เป็นพื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า หรือมีค่าเท่ากับ 1

ตารางที่ 3.1 ตัวอย่างข้อมูลที่นำมาทดสอบโดยวิธี Cochran's Q Test

จุดที่ถูกสุ่มในแผนที่ พยากรณ์	การจัดกระทำ (Treatment)			ผลรวม
	แผนที่พยากรณ์ที่ 1		แผนที่พยากรณ์ที่ k	
1				L_1
2				L_2
3	ข้อมูลอยู่ในมาตราแบบบัญญัติที่จำแนกข้อมูลออกเป็น กลุ่มเป็นพวกได้ 2 กลุ่ม มีค่าเพียงสองค่าคือ 0 และ 1			L_3
:				:
:				:
N				L_N
ผลรวม	G_1		G_k	

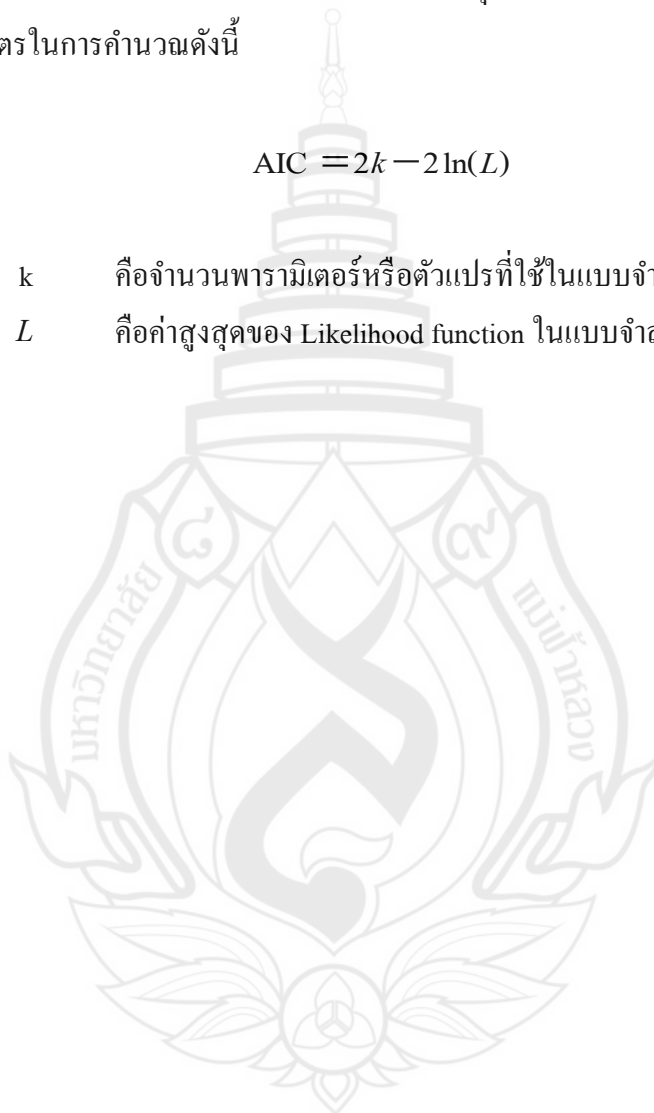
จากตารางที่ 3.1 แสดงตัวอย่างข้อมูลที่นำมาทดสอบโดยวิธี Cochran's Q Test โดยเริ่มจากการสุ่มจุดเพื่อที่จะสกัด (Extract) ข้อมูลจากแผนที่พยากรณ์ที่นำมาเปรียบเทียบกับจุดที่ถูกสุ่มนั้นตกอยู่ในแต่ละแผนที่พยากรณ์เป็นพื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (มีค่าเป็น 1) หรือเป็นพื้นที่ที่คาดว่าจะไม่มีการกระจายของต้นชาป่า (มีค่าเป็น 0) จากนั้นหาผลรวมของค่า L_i ในแต่ละแถวและผลรวมของค่า G_j ในแต่ละแผนที่พยากรณ์ จากนั้นนำไปคำนวณหาค่าสถิติ Q ตามสูตรข้างต้น โดยการแจกแจงแบบ Q จะเหมือนกับการแจกแจงไคสแควร์ที่มีค่าองศาความเป็นอิสระ (Degree of freedom; df) เท่ากับ $k-1$ ถ้าค่า Q ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าไคสแควร์ที่เปิดจากตารางสถิติที่ df เท่ากับ $k-1$ ณ ระดับนัยสำคัญที่กำหนดจะปฏิเสธ H_0 นั่นคือ สอดคล้องการตอบรับในแต่ละกลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (ศูนย์พัฒนาทรัพยากรการศึกษา มหาวิทยาลัยมหาสารคาม, ม.ป.ป.) หลังจากที่มีการทดสอบความแตกต่างของแผนที่พยากรณ์ที่มากกว่า 2 แผนที่โดยใช้วิธี Cochran's Q test แล้ว หากพบว่ามีความแตกต่างกันก็จะทำการทดสอบเปรียบเทียบพหุคูณ (Multiple Comparisons) โดยในการศึกษาครั้งนี้เลือกใช้วิธี Wilcoxon Sign Test ซึ่งเป็นการนำขนาดของความแตกต่างในข้อมูลแต่ละคู่มาพิจารณาพร้อมกับเครื่องหมาย

3.6.2 เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion)

เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) สร้างจากการประมาณความแปรปรวนของข้อสนเทศคูลล์แบ็ค-ไลท์เบอร์ (Kullback-Leibler Information) ระหว่างตัวแบบจริงกับตัวแบบที่เหมาะสม ที่มีคุณสมบัติไม่เอนเอียง การคัดเลือกตัวแบบโดยใช้เกณฑ์ AIC จะคัดเลือกตัวแบบที่ให้ค่า AIC ต่ำสุดเป็นตัวแบบที่ใช้ในการอธิบายตัวแปรตามได้ โดยมีสูตรในการคำนวณดังนี้

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

โดย k คือจำนวนพารามิเตอร์หรือตัวแปรที่ใช้ในแบบจำลอง
 L คือค่าสูงสุดของ Likelihood function ในแบบจำลอง



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในบทนี้จะแสดงผลการวิเคราะห์ข้อมูลทั่วไป ผลของการสร้างแบบจำลองการกระจายของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบนของประเทศไทยโดยเปรียบเทียบการสร้างแบบจำลองการกระจายในแต่ละวิธี ซึ่งได้แก่แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี Generalized Linear Models (GLMs), วิธี Generalized Additive Models (GAMs) และวิธี Maximum Entropy (MAXENT) รวมถึงผลการศึกษาผลกระทบที่เกิดจากการเปลี่ยนแปลงสภาพภูมิอากาศที่มีต่อการกระจายของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน

4.1 การวิเคราะห์ข้อมูลทั่วไป

การวิเคราะห์ข้อมูลทั่วไป ในที่นี้เป็นการวิเคราะห์ปัจจัยทางสภาพทางภูมิศาสตร์และสภาพทางภูมิอากาศในพื้นที่ที่มีการค้นพบต้นชาป่าจำนวน 185 จุด ในเขตพื้นที่ภาคเหนือตอนบนของประเทศไทย

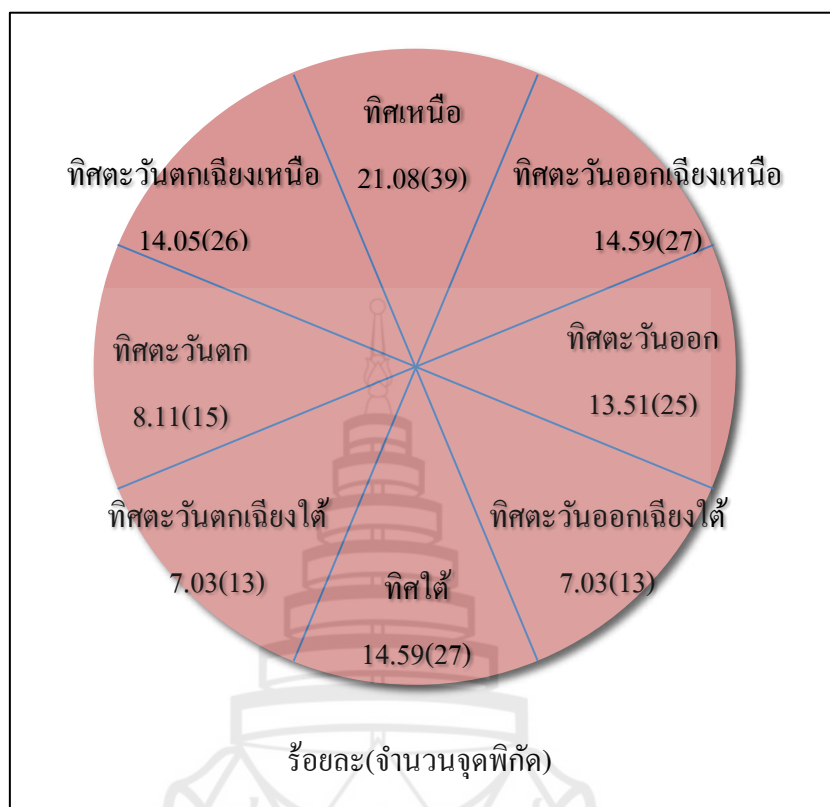
4.1.1 การวิเคราะห์ปัจจัยทางสภาพภูมิศาสตร์

ปัจจัยทางภูมิศาสตร์ที่นำมาศึกษาได้แก่ ความสูงจากระดับน้ำทะเล (DEM), ระยะห่างจากแม่น้ำ (Distance), ทิศด้านลาด (Aspect), โครงสร้างทางธรณีวิทยาของชนิดของหิน (Geology), ความลาดชัน (Slope) และขอบเขตของพื้นที่ป่าไม้ (Forest)

ตารางที่ 4.1 ค่าสถิติของความสูงจากระดับน้ำทะเล ความลาดชัน และระยะห่างจากแม่น้ำของพื้นที่ที่มีการค้นพบต้นชาป่าจำนวน 185 จุด

ปัจจัย	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
ความสูงจากระดับน้ำทะเล (เมตร)	963.26	406.93	275.00	2243.00
ความลาดชัน (องศา)	10.28	6.28	1.00	26.64
ระยะห่างจากแม่น้ำ (กิโลเมตร)	2.39	1.63	0.03	6.13

จากตารางที่ 4.1 แสดงค่าสถิติของความสูงจากระดับน้ำทะเล ความลาดชันและระยะห่างจากแม่น้ำของพื้นที่ที่มีการค้นพบต้นชาป่าจำนวน 185 จุดพบว่า พื้นที่ที่มีการค้นพบต้นชาป่าส่วนใหญ่อยู่บนพื้นที่สูง โดยมีระดับความสูงเฉลี่ยประมาณ 960 เมตรจากระดับน้ำทะเล ซึ่งพื้นที่สูงส่วนใหญ่ในเขตภาคเหนือตอนบนของประเทศไทยเป็นภูเขาสูงในเขตป่าธรรมชาติ มีความลาดชันโดยเฉลี่ยประมาณ 10 องศา ซึ่งถูกจัดให้อยู่ในกลุ่ม Gentle slope นั่นคือ พื้นที่ที่มีความลาดเอียงเล็กน้อย มีความลาดชันไม่เกิน 15 องศา โดยพื้นที่ส่วนใหญ่อยู่ห่างจากแม่น้ำโดยเฉลี่ยประมาณ 2 กิโลเมตร



ภาพที่ 4.1 จำนวนจุดพิกัดและร้อยละของการค้นพบต้นชาป่าจำนวน 185 จุด จำแนกตามทิศด้านลาด

จากภาพที่ 4.1 แสดงจำนวนจุดพิกัดและร้อยละของการค้นพบต้นชาป่าจำแนกตามทิศด้านลาด ซึ่งทิศด้านลาดจะเป็นการบอกทิศทางของพื้นที่ที่มีความลาดชันนั้นว่าหันไปทางทิศใด จะเห็นได้ว่าโดยส่วนใหญ่แล้ว ต้นชาป่าจะถูกค้นพบในพื้นที่ที่มีความลาดชันหันหน้าก่อนไปทางทิศตะวันออกและทิศเหนือ แต่ก็เป็นที่น่าสังเกตที่มีการค้นพบต้นชาป่าในพื้นที่ที่มีความลาดชันหันหน้าไปทางทิศใต้มากเป็นอันดับสองซึ่งเท่ากับพื้นที่ที่หันหน้าไปทางทิศตะวันออกเฉียงเหนือ ซึ่งอาจเนื่องมาจากปรากฏการณ์ตะวันอ้อมใต้หรือตะวันอ้อมข้าว (Winter Solstice) อธิบายได้ว่าโดยปกติแกนหมุนของโลกเอียงประมาณ 23 องศาจากระนาบที่โลกหมุนรอบดวงอาทิตย์ ถ้าประเทศอยู่ตรงบริเวณเส้นศูนย์สูตร ตะวันก็จะอ้อมเหนือ 6 เดือน และอ้อมใต้ 6 เดือน แต่เนื่องจากประเทศไทยเป็นประเทศเหนือเส้นศูนย์สูตรก็จะเกิดความไม่เท่ากันของการอ้อมเหนืออ้อมใต้ ยิ่งอยู่เหนือเส้นศูนย์สูตรมากเท่าไร ตะวันก็จะอ้อมใต้มากขึ้น โดยเฉพาะในฤดูหนาวดวงอาทิตย์จะปรากฏขึ้นจากขอบฟ้าทางทิศตะวันออกเฉียงใต้ โคจรอ้อมลงสู่ขอบฟ้าทางทิศตะวันตกเฉียงใต้ ซึ่งเป็นช่วงเวลากลางวันสั้นกว่าเวลากลางคืน ดังนั้นจึงเป็นไปได้ว่าต้นชาป่ามักเจริญเติบโตได้ดีในพื้นที่ที่มีความลาดชันหันหน้าไปทางแสงจากดวงอาทิตย์

ตารางที่ 4.2 จำนวนและร้อยละของจุดพิคัดของการค้นพบต้นชาป่าจำนวน 185 จุด จำแนกตาม
โครงสร้างทางธรณีวิทยาของชนิดหิน ประเภทของความลาดชันและการจัดกลุ่มของ
โครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน

ปัจจัยทางภูมิศาสตร์		จำนวนจุด พิคัด	ร้อยละ
โครงสร้างทางธรณีวิทยา	หินตะกอนและหินแปร	114	61.62
ของชนิดหิน	หินอัคนี	71	38.38
ประเภทของความลาดชัน	Gentle Slopes (0 - 15 องศา)	143	77.30
	Moderate Slopes (15 – 30 องศา)	42	22.70
การจัดกลุ่มของโครงสร้าง	Class1	84	45.41
ทางธรณีวิทยาของชนิดหินและ	Class2	30	16.22
ประเภทของความลาดชัน	Class5	59	31.89
	Class6	12	6.49
รวม		185	100.00

จากตารางที่ 4.2 แสดงจำนวนและร้อยละของจุดพิคัดของการค้นพบต้นชาป่าจำนวน 185 จุด จำแนกตามโครงสร้างทางธรณีวิทยาของชนิดหินซึ่งโผล่ที่ผิวดินหรือถูกดินปกคลุมอยู่ไม่เกิน 1 หรือ 1.5 เมตร ประเภทของความลาดชันและการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน พบว่าพื้นที่ที่มีการค้นพบต้นชาป่าส่วนใหญ่เป็นบริเวณที่เป็นหินตะกอนและหินแปรประมาณร้อยละ 62 ส่วนที่เหลือพบในบริเวณพื้นที่ที่เป็นหินอัคนี โดยพื้นที่ส่วนใหญ่เป็นบริเวณ Gentle Slopes กล่าวคือมีความลาดชันระหว่าง 0-15 องศา ส่วนที่เหลือเป็นบริเวณ Moderate Slopes กล่าวคือมีความลาดชันระหว่าง 15-30 องศา เมื่อแยกตามการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1-Class8) พบว่ามีเพียง Class1, Class2, Class5 และ Class6 เท่านั้นที่มีการค้นพบต้นชาป่า โดยส่วนใหญ่แล้วพบต้นชาป่าบริเวณ Class1 กล่าวคือเป็นพื้นที่ที่มีความลาดชัน 0-15 องศาและเป็นพื้นที่ที่ปรากฏหินตะกอนและหินแปร รองลงมาเป็นพื้นที่ที่มีความลาดชัน 0-15 องศาและเป็นพื้นที่ที่ปรากฏหินอัคนี

4.1.2 การวิเคราะห์ปัจจัยทางสภาพภูมิอากาศ

ปัจจัยทางภูมิอากาศที่นำมาศึกษาได้แก่ ความชื้นสัมพัทธ์ (Humidity), ตัวแปรด้าน Bioclimatic (Bio1-Bio19) โดยตัวแปร Bio1-Bio11 เป็นตัวแปรที่เกี่ยวข้องกับอุณหภูมิ ส่วน Bio12-Bio19 เป็นตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน

ตารางที่ 4.3 ค่าสถิติของความชื้นสัมพัทธ์ (Humidity) ในพื้นที่ที่มีการค้นพบต้นชาป่า

ความชื้นสัมพัทธ์ (Humidity)	
ค่าเฉลี่ย	74.94
ส่วนเบี่ยงเบนมาตรฐาน	2.98
ค่าต่ำสุด	70.05
ค่าสูงสุด	83.43
จำนวน	185.00

จากตารางที่ 4.3 แสดงค่าสถิติของความชื้นสัมพัทธ์ (Humidity) พบว่าในบริเวณที่มีการค้นพบต้นชาป่ามีความชื้นสัมพัทธ์โดยเฉลี่ยประมาณร้อยละ 75 ซึ่งความชื้นสัมพัทธ์ในภาคเหนือของประเทศไทยมีปริมาณสูงอยู่แล้ว เนื่องจากพื้นที่เป็นที่สูงและมีป่าไม้จำนวนมาก โดยความชื้นสัมพัทธ์ผันแปรไปตามช่วงเวลาของวัน กล่าวคือ ความชื้นสัมพัทธ์จะสูงสุดในช่วงเช้ามืด และต่ำสุดในช่วงบ่าย

ตารางที่ 4.4 ค่าสถิติของตัวแปรด้าน Bioclimatic (Bio1-Bio11) ที่เป็นตัวแปรที่เกี่ยวข้องกับ
อุณหภูมิ (หน่วย: °C × 10) จำนวน 185 จุด

ตัวแปร	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
Annual Mean Temperature (Bio1)	223.88	19.13	159	255
Mean Diurnal Range (Bio2)	121.92	1.54	119	126
Isothermality (Bio3)	52.16	1.15	50	55
Temperature Seasonality (Bio4)	2519.45	151.35	2090	2854
Max Temperature of Warmest Month (Bio5)	327.06	20.31	257	356
Min Temperature of Coldest Month (Bio6)	95.20	19.03	31	124
Temperature Annual Range (Bio7)	231.92	3.31	224	241
Mean Temperature of Wettest Quarter (Bio8)	237.69	19.34	176	267
Mean Temperature of Driest Quarter (Bio9)	191.02	21.11	122	224
Mean Temperature of Warmest Quarter (Bio10)	249.77	19.41	184	280
Mean Temperature of Coldest Quarter (Bio11)	185.93	18.76	122	217

จากตารางที่ 4.4 แสดงค่าสถิติของตัวแปร Bio1-Bio11 ซึ่งเป็นตัวแปรที่เกี่ยวข้องกับอุณหภูมิ โดยผลการวิเคราะห์ข้อมูลมีรายละเอียดดังต่อไปนี้

ในบริเวณที่มีการค้นพบต้นชาป่ามีอุณหภูมิเฉลี่ยรายปี (Bio1) ประมาณ 22 องศาเซลเซียส โดยพบว่าอุณหภูมิเฉลี่ยต่ำสุดประมาณ 16 องศาเซลเซียสและอุณหภูมิเฉลี่ยสูงสุดประมาณ 26 องศาเซลเซียส เมื่อพิจารณาถึงความผันผวนของอุณหภูมิรายวัน (ความแตกต่างระหว่างอุณหภูมิกลางวันและกลางคืน) พบว่า ความแตกต่างของอุณหภูมิรายวัน (Bio2) โดยเฉลี่ยประมาณ 12 องศาเซลเซียส โดยมีค่าสูงสุดและค่าต่ำสุดไม่ห่างจากค่าเฉลี่ยมากนัก เมื่อพิจารณาค่า Isothermality (Bio3) ซึ่งเป็นดัชนีที่ใช้อธิบายการผันผวนของความแตกต่างของอุณหภูมิรายวันเปรียบเทียบกับความแตกต่างของอุณหภูมิรายปี ซึ่งมีค่าเฉลี่ยโดยประมาณร้อยละ 52

เมื่อพิจารณาค่า Temperature Seasonality (Bio4) หรือช่วงฤดูกาลของอุณหภูมิ ซึ่งเป็นดัชนีที่ใช้สำหรับวัดการเปลี่ยนแปลงของอุณหภูมิในช่วงปีที่กำหนดหรือเฉลี่ยรายปี โดยขึ้นอยู่กับส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของอุณหภูมิเฉลี่ยรายเดือน ถ้าหาค่า Temperature Seasonality มีค่าสูงแสดงว่าอุณหภูมิมีความแปรปรวนมากขึ้น ในการสำรวจบริเวณที่มีการค้นพบต้นชาป่าพบว่ามีค่า Temperature Seasonality โดยเฉลี่ย 2519.45 หรือประมาณ 2.5 องศาเซลเซียส แสดงว่าความแปรปรวนของอุณหภูมิมิมีไม่มากนัก

ในช่วงเดือนที่ร้อนที่สุดมีอุณหภูมิเฉลี่ยสูงสุด (Bio5) ประมาณ 33 องศาเซลเซียส ในขณะที่ช่วงเดือนที่หนาวที่สุดมีอุณหภูมิเฉลี่ยต่ำสุด (Bio6) ประมาณ 10 องศาเซลเซียส นั่นคือความแตกต่างของอุณหภูมิรายปี (Bio7) ซึ่งคำนวณได้จากผลต่างระหว่างอุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5) และอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6) โดยเฉลี่ยประมาณ 23 องศาเซลเซียส ซึ่งมีความแตกต่างกันค่อนข้างมากเนื่องจากทางภาคเหนือของประเทศไทยมีลักษณะภูมิประเทศส่วนใหญ่เป็นพื้นที่สูง ในช่วงฤดูหนาวจะได้รับอิทธิพลจากลมมรสุมฤดูหนาวหรือลมมรสุมตะวันออกเฉียงเหนือ ในขณะที่ฤดูร้อนอุณหภูมิจะสูง เนื่องจากอยู่ไกลจากทะเลจึงทำให้ได้รับอิทธิพลจากทะเลน้อย และเป็นระยะที่ดวงอาทิตย์ส่องแสงตั้งฉากกับภาคเหนือ

ในช่วงไตรมาสที่ฝนตกชุกที่สุดมีอุณหภูมิโดยเฉลี่ย (Bio8) ประมาณ 24 องศาเซลเซียส ในขณะที่ในช่วงไตรมาสที่แห้งแล้งที่สุดจะมีอุณหภูมิเฉลี่ย (Bio9) โดยประมาณ 19 องศาเซลเซียส

ในช่วงไตรมาสที่ร้อนที่สุดมีอุณหภูมิเฉลี่ย (Bio10) ประมาณ 25 องศาเซลเซียส ในขณะที่ช่วงไตรมาสที่หนาวที่สุดจะมีอุณหภูมิเฉลี่ย (Bio11) โดยประมาณ 19

ตารางที่ 4.5 ค่าสถิติของตัวแปรด้าน Bioclimatic (Bio12-Bio19) ที่เป็นตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน (หน่วย : มิลลิเมตร) จำนวน 185 จุด

ตัวแปร	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่าต่ำสุด	ค่าสูงสุด
Annual Precipitation (Bio12)	1273.56	182.76	1023	1640
Precipitation of Wettest Month (Bio13)	260.90	68.94	179	399
Precipitation of Driest Month (Bio14)	8.11	2.12	4	19
Precipitation Seasonality (Bio15)	81.27	4.95	69	91
Precipitation of Wettest Quarter (Bio16)	663.22	134.07	476	930
Precipitation of Driest Quarter (Bio17)	36.18	6.35	21	65
Precipitation of Warmest Quarter (Bio18)	381.91	59.98	232	490
Precipitation of Coldest Quarter (Bio19)	37.16	9.80	22	117

จากตารางที่ 4.5 แสดงค่าสถิติของตัวแปร Bio12-Bio19 ซึ่งเป็นตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝน โดยผลการวิเคราะห์ข้อมูลมีรายละเอียดดังต่อไปนี้

ในบริเวณที่มีการค้นพบต้นชาป่ามีปริมาณน้ำฝนรายปี (Bio12) โดยเฉลี่ยประมาณ 1274 มิลลิเมตรต่อปี โดยพบว่ามีปริมาณน้ำฝนรายปีต่ำสุดอยู่ที่ 1023 มิลลิเมตรต่อปี สูงสุดอยู่ที่ 1640 มิลลิเมตรต่อปี ในเดือนที่ฝนตกชุกที่สุดมีปริมาณน้ำฝนเฉลี่ย (Bio13) ประมาณ 261 มิลลิเมตรต่อเดือน ในขณะที่เดือนที่แห้งแล้งที่สุดมีปริมาณน้ำฝนเฉลี่ย (Bio14) ประมาณ 8 มิลลิเมตรต่อเดือน เมื่อพิจารณาค่า Precipitation Seasonality (Bio15) ซึ่งเป็นดัชนีชี้วัดความแปรปรวนของปริมาณน้ำฝนโดยขึ้นอยู่กับค่าสัมประสิทธิ์ความผันแปร (Coefficient of Variation) ของอุณหภูมิเฉลี่ยราย

เดือน โดยมีค่าเฉลี่ยประมาณร้อยละ 81.27 ซึ่งถือว่ามีความผันแปรของปริมาณน้ำฝนค่อนข้างสูง กล่าวคือพื้นที่มีช่วงที่ฝนตกชุกที่สุดและมีช่วงที่แล้งที่สุดภายในปีเดียวกัน

ในช่วงไตรมาสที่ฝนตกชุกที่สุดมีปริมาณน้ำฝนเฉลี่ย (Bio16) ประมาณ 663 มิลลิเมตรต่อไตรมาสหรือ 221 มิลลิเมตรต่อเดือน ในขณะที่ช่วงไตรมาสที่แห้งแล้งที่สุดกลับพบว่ามีปริมาณน้ำฝนเฉลี่ย (Bio17) เพียง 36 มิลลิเมตรต่อไตรมาสหรือ 12 มิลลิเมตรต่อเดือน ส่วนในช่วงไตรมาสที่ร้อนที่สุดมีปริมาณน้ำฝนเฉลี่ย (Bio18) ประมาณ 382 มิลลิเมตรต่อไตรมาสหรือ 127 มิลลิเมตรต่อเดือน ในขณะที่ช่วงไตรมาสที่หนาวที่สุดมีปริมาณน้ำฝนเฉลี่ย (Bio19) เพียง 17 มิลลิเมตรต่อไตรมาสหรือประมาณ 6 มิลลิเมตรต่อเดือน ซึ่งในช่วงฤดูหนาวทางภาคเหนือของประเทศไทยจะได้รับอิทธิพลจากลมมรสุมตะวันออกเฉียงเหนือ อากาศจะหนาวเย็นและแห้งแล้ง ดังนั้นในช่วงไตรมาสที่หนาวที่สุดจึงมีปริมาณน้ำฝนน้อย

4.2 การสร้างแบบจำลองการกระจายของสิ่งมีชีวิต

การสร้างแบบจำลองการกระจายของสิ่งมีชีวิต ในที่นี้คือการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยเปรียบเทียบกัน 3 วิธี คือ วิธี Generalized Linear Models (GLMs), วิธี Generalized Additive Models (GAMs) และวิธี Maximum Entropy (MAXENT)

4.2.1 แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs เป็นวิธีการสร้างแบบจำลองโดยใช้เทคนิคการจำแนกกลุ่ม (Discrimination Technique) ซึ่งเป็นเทคนิคที่ต้องการทั้งข้อมูลจุดที่มีการค้นพบสิ่งมีชีวิต (Presence Points) และข้อมูลจุดที่ไม่มีการค้นพบสิ่งมีชีวิต (Absence Points) ซึ่งโดยปกติแล้ว ข้อมูล Absence มักจะไม่มีการบันทึกหรือเก็บรวบรวมไว้ ดังนั้นจึงมีการจำลองข้อมูล Absence ขึ้นมาเป็นข้อมูลเสมือนของจุดที่ไม่มีการค้นพบสิ่งมีชีวิต โดยเรียกข้อมูลนี้ว่า “Pseudo-Absence” ซึ่งในการศึกษาครั้งนี้จะทำการสุ่มเลือกจุด Pseudo-Absence เปรียบเทียบกัน 3 วิธี ได้แก่ แบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา แบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence และแบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence แบบสองขั้นตอน โดยมีรายละเอียดดังต่อไปนี้

4.2.1.1 GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

การเลือกจุด Pseudo-Absence โดยวิธีนี้จะทำการสุ่มจุด Pseudo-Absence ภายในพื้นที่รอบ ๆ จุดที่มีการค้นพบสิ่งมีชีวิต (Presence Points) ในรัศมี 10, 15, 20, 25, 30, 35, 40, 45 และ 50 กิโลเมตรตามลำดับ โดยในแต่ละรัศมีจะสุ่มจุด Pseudo-Absence จำนวน 20 จุด ซึ่งในแต่ละชุดจะถูกนำไปสร้างแบบจำลองเพื่อหาแบบจำลองที่เหมาะสม ซึ่งวิธีคัดเลือกรัศมีที่มีความเหมาะสมสามารถศึกษาได้จากภาคผนวก ก สรุปได้ว่าที่รัศมี 45 กิโลเมตรมีความเหมาะสม เนื่องจากมีค่าเฉลี่ยของค่า Cross-Validation Error น้อยและมีความผันแปรที่เกิดจากการทำซ้ำน้อยที่สุด

ตารางที่ 4.6 ค่า Cross-Validation Error ของชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตร

รอบการทำซ้ำ (Iteration)	Cross-Validation Error
1	0.31397
2	0.31357
3	0.31442
4	0.31407
5	0.31297
6	0.31395
7	0.31311
8	0.3125
9	0.31392
10	0.31175

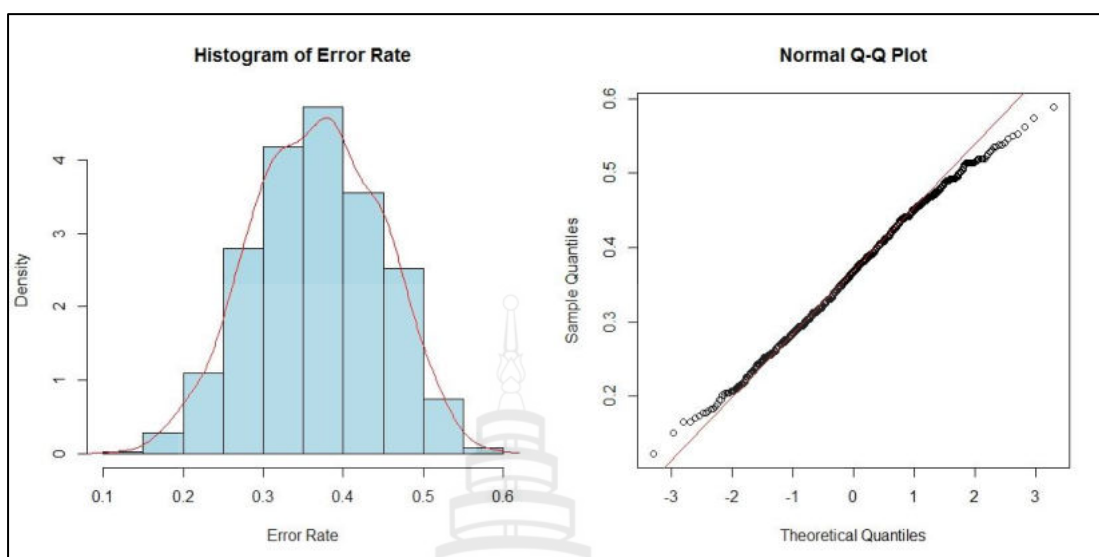
จากตารางที่ 4.6 แสดงค่า Cross-Validation Error ของชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตรในการทำซ้ำ 10 รอบ พบว่าค่า Cross-Validation Error ในแต่ละรอบมีค่าใกล้เคียงกัน ไม่มีความแตกต่างกันมากนัก ซึ่งรอบของการทำซ้ำที่จะถูกเลือกจะต้องให้ค่า Cross-Validation Error น้อยที่สุด นั่นคือรอบที่ 10 ซึ่งมีค่าเท่ากับ 0.31175 (ตัวหนา)

เมื่อได้รอบการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุดแล้ว ข้อมูลจะถูกแบ่งออกเป็นกลุ่ม Training Data และ Testing Data ซึ่งจะแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่ม โดยกลุ่มที่ 1 จะหมายถึงกลุ่มที่ใช้กลุ่มที่ 1 เป็นกลุ่ม Testing Data ในขณะที่กลุ่มอื่น ๆ เป็นกลุ่ม Training Data กลุ่มที่ 2, กลุ่มที่ 3, กลุ่มที่ 4 และกลุ่มที่ 5 ก็จะเปลี่ยนกลุ่ม Testing Data ในลักษณะเดียวกันกับกลุ่มที่ 1 โดยในแต่ละกลุ่มจะมีการทำ Bootstrap จำนวน 1000 ครั้ง

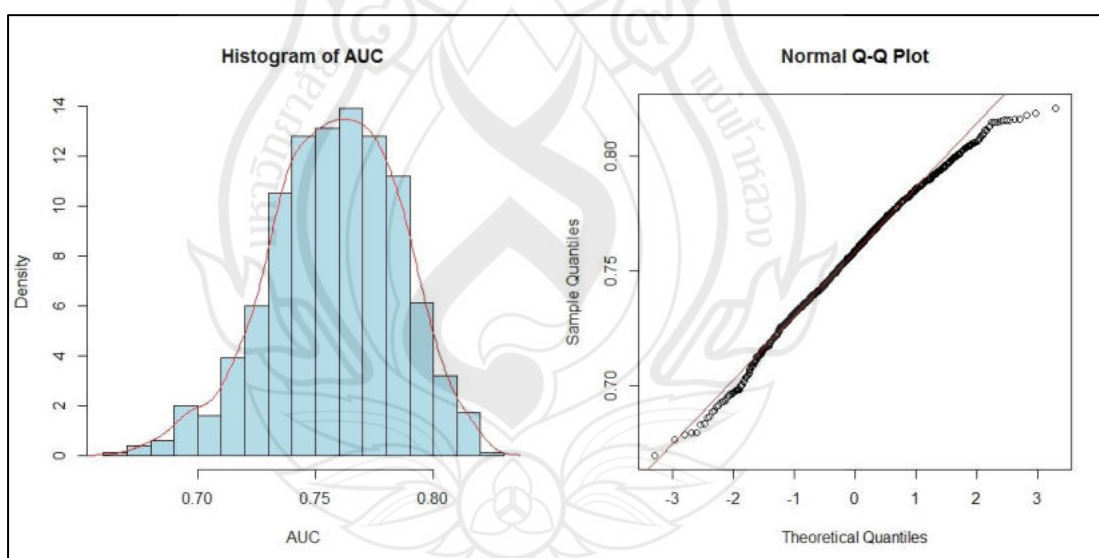
ตารางที่ 4.7 ค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 10 ชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตร

กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
1	0.25798	0.06147	0.23826	1000
2	0.32005	0.07991	0.24967	1000
3	0.36667	0.07958	0.21702	1000
4	0.27882	0.08391	0.30093	1000
5	0.33522	0.08330	0.24850	1000

จากตารางที่ 4.7 แสดงค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 10 ชุด Pseudo-Absence ที่ 17 ที่รัศมี 45 กิโลเมตรในการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ค่าเฉลี่ยของ Error Rate ในกลุ่มที่ 1 มีค่าน้อยที่สุด แต่ค่าสัมประสิทธิ์ความผันแปรในกลุ่มที่ 3 กลับมีค่าน้อยที่สุดแทน และเนื่องจากต้องการลดความผันแปรจากการทำ Bootstrap ดังนั้นจึงเลือกกลุ่มที่ 3 (ตัวหนา) ซึ่งให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุดและตรงตามเงื่อนไขคือค่าสัมประสิทธิ์ความผันแปรไม่ควรมีค่าเกินร้อยละ 30 ซึ่งกลุ่มที่ 3 ให้ค่าสัมประสิทธิ์ความผันแปรประมาณร้อยละ 22



ภาพที่ 4.2 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า Error Rate ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง



ภาพที่ 4.3 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง

ในการทำ Bootstrap จำนวน 1000 ครั้ง มีความจำเป็นที่ต้องมีการตรวจสอบการแจกแจงของค่าที่ต้องการศึกษาว่ามีการแจกแจงแบบปกติหรือไม่โดยการพล็อตฮิสโตแกรมและ Normal Q-Q Plot ดังแสดงในภาพที่ 4.2 พบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมซึ่งเป็นข้อมูลจริงวิ่งไปตามเส้นตรงสีแดงซึ่งเป็นค่าคาดหวัง (Expected Value) ดังนั้นจึงอนุมานได้ว่าค่า Error Rate ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติ นอกเหนือจากค่า Error Rate ที่สนใจที่จะศึกษาแล้ว อีกค่าที่สนใจคือค่า AUC ดังแสดงในภาพที่ 4.3 ซึ่งแสดงแผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมที่วิ่งไปตามเส้นสีแดงเช่นเดียวกันกับค่า Error Rate จึงอนุมานได้ว่าค่า AUC ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติเช่นเดียวกัน

ตารางที่ 4.8 ค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3

กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
3	0.75794	0.02710	0.03575	1000
แบบจำลอง	:	591		
AUC	:	0.75790		
Error Rate	:	0.31532		

จากตารางที่ 4.8 แสดงค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3 ซึ่งเป็นกลุ่มที่ถูกคัดเลือกแล้วเนื่องจากให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุด พบว่า ในกลุ่มที่ 3 มีค่าเฉลี่ยของ AUC อยู่ที่ 0.75794 และแบบจำลองที่ถูกคัดเลือกจาก 1000 แบบจำลองคือแบบจำลองที่ 591 เนื่องจากมีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุดคือ 0.75790 โดยมีค่า Error Rate เท่ากับ 0.31532

เมื่อได้แบบจำลองที่ 591 เป็นตัวแทนที่เหมาะสมแล้ว ขั้นตอนต่อไปคือการคัดเลือกตัวแปรด้วยวิธีการของ Stepwise โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) มาคัดเลือกตัวแปร

ตารางที่ 4.9 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษาโดยมี
การคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว

Factors	Estimate	Standard Error	Z Test
Intercept	-86.043	445.587	-0.193
DEM	0.005	0.002	2.431 *
Distance	-0.093	0.064	-1.451
Humidity	-0.127	0.055	-2.300 *
Bio3	0.890	0.222	4.000 ***
Bio4	0.007	0.002	3.631 ***
Bio8	0.135	0.060	2.256 *
Bio10	-0.080	0.051	-1.565
Bio12	0.021	0.005	4.002 ***
Bio14	-0.135	0.060	-2.247 *
Bio16	-0.039	0.009	-4.423 ***
Class1	14.651	445.198	0.033
Class2	14.514	445.198	0.033
Class5	14.266	445.198	0.032
Class6	13.562	445.198	0.030
Null Deviance = 892.55, Residual Deviance = 699.71			
Pseudo-R ² = 21.61, AIC = 729.71			

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001

* หมายถึงมีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 4.9 แสดงแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษาโดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้วพบว่า มีตัวแปรเข้าไปในแบบจำลองคือ ความสูงจากระดับน้ำทะเล (DEM), ระยะห่างจากแม่น้ำ (Distance), ความชื้นสัมพัทธ์ (Humidity), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6)

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 892.55 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 699.71 ซึ่งสามารถคำนวณหาค่า Pseudo- R^2 ได้เท่ากับ 21.61 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 21.61 ได้สมการพยากรณ์ดังนี้

$$\begin{aligned} \text{Logit}(\mu) = & -86.043 + 0.005\text{DEM} - 0.093\text{Distance} - 0.127\text{Humidity} + 0.890\text{Bio3} \\ & + 0.007\text{Bio4} + 0.135\text{Bio8} - 0.080\text{Bio10} + 0.021\text{Bio12} - 0.135\text{Bio14} \\ & - 0.039\text{Bio16} + 14.651\text{Class1} + 14.514\text{Class2} + 14.266\text{Class5} \\ & + 13.562\text{Class6} \end{aligned}$$

จากสมการพบว่าตัวแปร DEM, Bio3, Bio4, Bio8, Bio12, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวกแสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Distance, Humidity, Bio10, Bio14 และ Bio16 มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio3, Bio4 และ Bio12 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร DEM และ Bio8 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.05 สามารถอธิบายได้ดังต่อไปนี้

ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างช่วงความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลเชิงบวกหมายความว่าถ้าหากค่า Isothermality (Bio3) มีค่ามากหรือมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็จะ

เพิ่มขึ้น เช่นเดียวกันกับตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4) ซึ่งคำนวณมาจากค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) คูณด้วย 100 และปริมาณน้ำฝนรายปี (Bio12) นั่นคือพื้นที่ใดที่มีความผันแปรของอุณหภูมิมากหรือมีการเปลี่ยนแปลงของอุณหภูมิที่ค่อนข้างแตกต่าง ซึ่งหมายถึงพื้นที่ที่มีช่วงของฤดูกาลโอกาสของการค้นพบต้นชาป่าในพื้นที่ที่เพิ่มขึ้นหรือพื้นที่ใดที่มีปริมาณน้ำฝนรายปีที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่เพิ่มขึ้นด้วยเช่นกัน นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

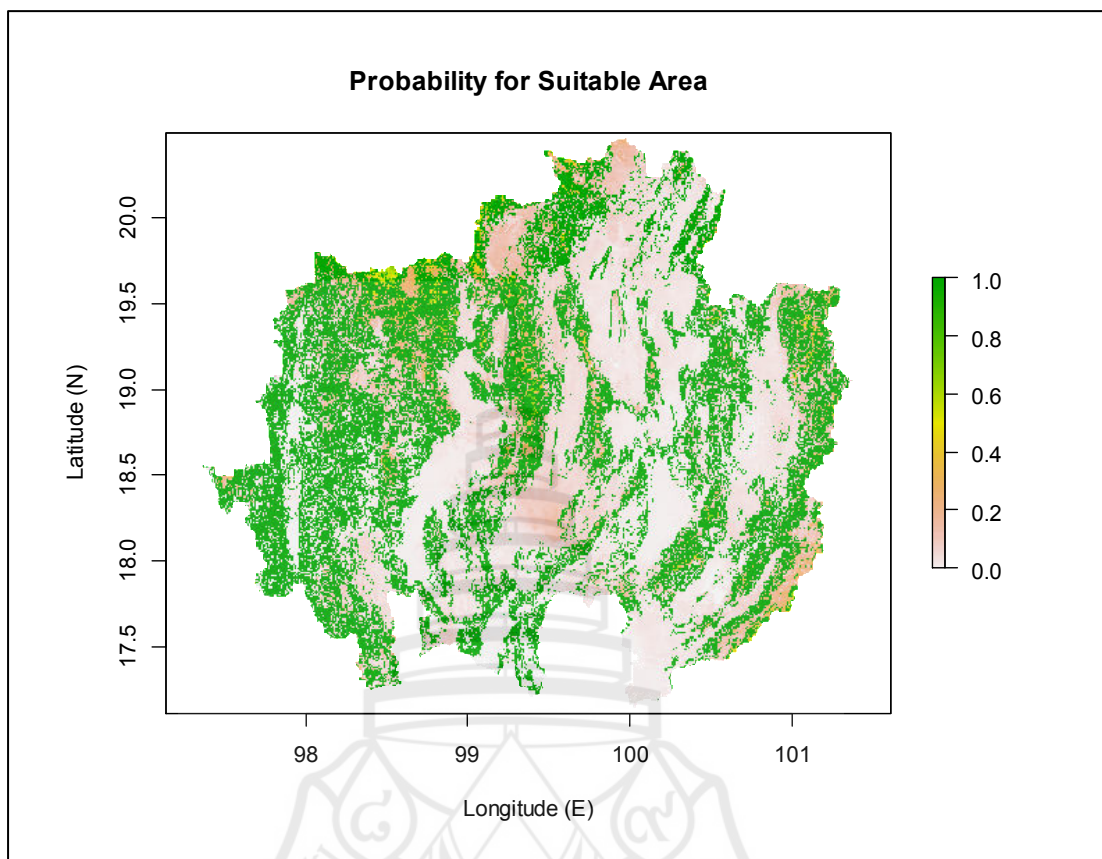
ตัวแปรความสูงจากระดับน้ำทะเล (DEM) มีผลเชิงบวกหมายความว่า ยิ่งพื้นที่อยู่สูงจากระดับน้ำทะเลเท่าไรโอกาสที่จะค้นพบต้นชาป่ายังมีมากขึ้น นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ส่วนตัวแปรชั้นของข้อมูลธรณีวิทยาและความลาดชัน (Class1, Class2, Class5 และ Class6) ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติ แต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ โดยมีผลเชิงบวก นั่นคือโครงสร้างทางธรณีวิทยาของชนิดหินไม่มีผลต่อแบบจำลองแต่พื้นที่จะต้องมีความลาดชันไม่เกิน 30 องศา โอกาสที่จะค้นพบต้นชาป่ามากยิ่งขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Bio16 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Humidity และ Bio14 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.05 สามารถอธิบายได้ดังต่อไปนี้

ตัวแปรปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) มีผลเชิงลบ หมายความว่าพื้นที่ใดที่มีปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุดมากเกินไป โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลงได้ เช่นเดียวกับตัวแปรความชื้นสัมพัทธ์ (Humidity) และปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) มีผลเชิงลบ กล่าวคือพื้นที่ใดมีความชื้นสัมพัทธ์สูง ๆ หรือพื้นที่ใดในช่วงเดือนที่แห้งแล้งที่สุดมีปริมาณน้ำฝนมากโอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลง

ส่วนตัวแปรระยะห่างจากแม่น้ำ (Distance) และตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติ แต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ โดยตัวแปรระยะห่างจากแม่น้ำ (Distance) และอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio10) มีผลเชิงลบ กล่าวคือพื้นที่ใดมีความชื้นสัมพัทธ์สูง ๆ หรือมีอุณหภูมิเฉลี่ยสูง ๆ ในไตรมาสที่ร้อนที่สุด โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลงเช่นกัน



ภาพที่ 4.4 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

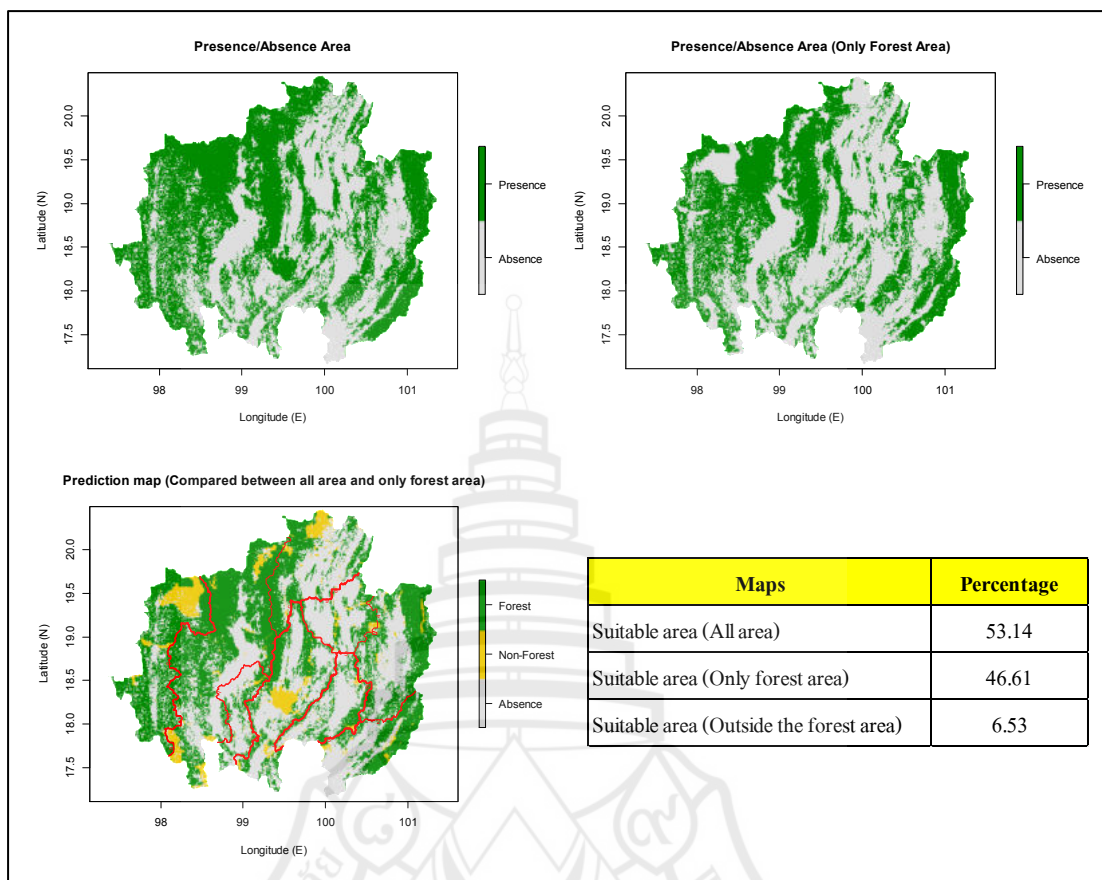
จากภาพที่ 4.4 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษาโดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 4.10 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	295
AUC	0.76236
max TPR+TNR	-2.36536
Threshold	0.08585

จากตารางที่ 4.10 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษาโดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้วพบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absence ใช้ทั้งหมด 295 จุด มีค่า AUC เท่ากับ 0.76236 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 76.24 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.08585 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.5 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.5 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษาโดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ซึ่งแผนที่จะถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนจะแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 53 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนจะแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 47 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า แต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีมีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 7 ของพื้นที่

ทั้งหมด ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอน เชียงรายและลำปาง ดังแสดงในภาพซ้ายมือล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัดในเขตพื้นที่ภาคเหนือตอนบน เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่อยู่ทางทิศตะวันตกก่อนไปทางทิศเหนือของพื้นที่ ซึ่งอยู่แถว ๆ ป่าในจังหวัดเชียงใหม่และแม่ฮ่องสอนเป็นส่วนใหญ่ มีบางส่วนอยู่ระหว่างป่ารอยต่อของจังหวัดเชียงใหม่และเชียงราย นอกจากนี้ยังพบว่าพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าบางส่วนกระจายอยู่ค่อนข้างห่างจากทางทิศตะวันออกของพื้นที่แถวป่าชายแดนในเขตจังหวัดน่าน

4.2.1.2 GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence

การสุ่มจุด Pseudo-Absence โดยวิธีนี้จะทำการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบ ๆ จุดที่มีการค้นพบสิ่งมีชีวิต (Presence Points) ในรัศมี 10, 15, 20, 25, 30, 35, 40, 45 และ 50 กิโลเมตรตามลำดับ โดยในแต่ละรัศมีจะสุ่มจุด Pseudo-Absence จำนวน 20 จุด ซึ่งในแต่ละชุดจะถูกนำไปสร้างแบบจำลองเพื่อหาแบบจำลองที่เหมาะสม โดยวิธีการในการพิจารณาหารัศมีที่มีความเหมาะสม การหาขอบของการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุด การหากลุ่ม Testing Data ที่ดีที่สุดและการหาแบบจำลองที่ดีที่สุดใช้วิธีการพิจารณาเช่นเดียวกับแบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา พบว่ารัศมีที่มีความเหมาะสมต่อการสร้างแบบจำลองคือ 25 กิโลเมตร โดยมีชุด Pseudo-Absence ที่ 3 ที่ให้ค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด มีรอบการทำซ้ำที่ 2 ที่ให้ค่า Cross-Validation Error น้อยที่สุด และมีกลุ่ม Testing Data กลุ่มที่ 5 ที่ให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate น้อยที่สุด โดยมีแบบจำลองที่ 937 ถูกเลือกเป็นตัวแทนที่เหมาะสม (ดูรายละเอียดได้ที่ภาคผนวก ข) เมื่อได้แบบจำลองที่เป็นตัวแทนที่เหมาะสมแล้ว ขั้นตอนต่อไปคือการคัดเลือกตัวแปรด้วยวิธีการของ Stepwise โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) มาคัดเลือกตัวแปร

ตารางที่ 4.11 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว

Factors	Estimate	Standard Error	Z Test
Intercept	-72.965	668.418	-0.109
Aspect	-0.003	0.001	-2.400 *
Humidity	-0.192	0.079	-2.444 *
Bio2	-0.766	0.245	-3.131 **
Bio3	1.509	0.389	3.882 ***
Bio4	0.015	0.007	2.191 *
Bio5	-0.798	0.405	-1.971 *
Bio6	0.642	0.404	1.587
Bio7	1.065	0.392	2.717 **
Bio8	0.147	0.090	1.630
Bio10	-0.510	0.252	-2.029 *
Bio11	0.490	0.283	1.733 .
Bio12	0.009	0.004	2.483 *
Bio13	-0.049	0.012	-4.293 ***
Bio14	-0.259	0.112	-2.301 *
Bio17	0.417	0.185	2.248 *
Bio19	-0.247	0.173	-1.433
Class1	15.429	667.762	0.023
Class2	14.652	667.762	0.022

ตารางที่ 4.11 (ต่อ)

Factors	Estimate	Standard Error	Z Test
Class5	15.313	667.762	0.023
Class6	14.971	667.762	0.022

Null Deviance = 915.74, Residual Deviance = 548.73

Pseudo-R² = 40.08, AIC = 590.73

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 4.11 แสดงแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว พบว่ามีตัวแปรเข้าไปในแบบจำลองคือ ทิศด้านลาด (Aspect), ความชื้นสัมพัทธ์ (Humidity), ช่วงความแตกต่างของอุณหภูมิรายวัน (Bio2), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), ช่วงความแตกต่างของอุณหภูมิรายปี (Bio7), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), อุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17), ปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19) และกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) ซึ่งมีตัวแปร Bio3 และ Bio13 มีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Bio2 และ Bio7 มีนัยสำคัญทางสถิติที่ 0.01 ตัวแปร Aspect, Humidity, Bio4, Bio5, Bio10, Bio12, Bio14 และ Bio17 มีนัยสำคัญทางสถิติที่ 0.05 และตัวแปร Bio11 มีนัยสำคัญทางสถิติที่ 0.10

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 915.74 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 548.73 ซึ่งสามารถคำนวณหาค่า Pseudo- R^2 ได้เท่ากับ 40.08 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 40.08 ได้สมการพยากรณ์ดังนี้

$$\begin{aligned} \text{Logit}(\mu) = & -72.965 - 0.003\text{Aspect} - 0.192\text{Humidity} - 0.766\text{Bio2} + 1.509\text{Bio3} \\ & + 0.015\text{Bio4} - 0.798\text{Bio5} + 0.642\text{Bio6} + 1.065\text{Bio7} + 0.147\text{Bio8} \\ & - 0.510\text{Bio10} + 0.490\text{Bio11} + 0.009\text{Bio12} - 0.049\text{Bio13} - 0.259\text{Bio14} \\ & + 0.417\text{Bio17} - 0.247\text{Bio19} + 15.429\text{Class1} + 14.652\text{Class2} \\ & + 15.313\text{Class5} + 14.971\text{Class6} \end{aligned}$$

จากสมการพบว่าตัวแปร Bio3, Bio4, Bio6, Bio7, Bio8, Bio11, Bio12, Bio17, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Aspect, Humidity, Bio2, Bio5, Bio10, Bio13, Bio14 และ Bio19 มีค่าสัมประสิทธิ์เป็นลบ แสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio3 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.001 อธิบายได้ว่า ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างความแตกต่างของอุณหภูมิรายวัน โดยเฉลี่ย (Bio2) ต่อความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลเชิงบวกนั้นคือ ยิ่งค่า Isothermality (Bio3) มีค่ามากหรือมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

ตัวแปร Bio7 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.01 อธิบายได้ว่า ยิ่งตัวแปรช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีค่าสูงขึ้น โอกาสที่จะค้นพบต้นชาป่ามีมากขึ้น หมายความว่าพื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุด (Bio5) กับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุด (Bio6) มาก พื้นที่นั้นมีโอกาสที่จะค้นพบต้นชาป่ามากยิ่งขึ้น

ตัวแปร Bio4, Bio12 และตัวแปร Bio17 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.05 สามารถอธิบายได้ดังนี้ ตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4) คือค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) คูณด้วย 100 นั่นคือในพื้นที่ที่มีความผันแปรของอุณหภูมิมากหรือ

มีการเปลี่ยนแปลงของอุณหภูมิที่ค่อนข้างแตกต่าง ซึ่งหมายถึงพื้นที่ที่มีช่วงของฤดูกาล โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น เช่นเดียวกับตัวแปรปริมาณน้ำฝนรายปี (Bio12) อธิบายได้ว่าในพื้นที่ที่มีปริมาณน้ำฝนรายปีที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วย เช่นกันนั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า ส่วนตัวแปรปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) มีผลเชิงบวกเช่นเดียวกัน สามารถอธิบายได้ว่าพื้นที่ใดที่มีปริมาณน้ำฝนสูง ๆ ในช่วงไตรมาสที่แห้งแล้งที่สุด จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่มากขึ้น

ตัวแปร Bio11 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.10 สามารถอธิบายว่า พื้นที่ใดในช่วงไตรมาสที่หนาวที่สุด มีอุณหภูมิที่ไม่หนาวเกินไปหรือมีอุณหภูมิสูง ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีมากขึ้น

ส่วนตัวแปร Bio6, Bio8, Class1, Class2, Class5 และ Class6 ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ ซึ่งสามารถอธิบายได้ดังต่อไปนี้

ตัวแปรอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6) มีผลเชิงบวกหมายความว่าพื้นที่ใดในช่วงเดือนที่หนาวที่สุดมีอุณหภูมิไม่หนาวเกินไปหรือมีอุณหภูมิสูง โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีมากขึ้น

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8) มีผลเชิงบวก หมายความว่าพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีมากขึ้น

ส่วนกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) มีผลเชิงบวก หมายความว่าโครงสร้างทางธรณีวิทยาของชนิดหินไม่มีผลต่อแบบจำลองแต่พื้นที่ที่มีความลาดชันไม่เกิน 30 องศา โอกาสที่จะค้นพบต้นชาป่ายังมีมากขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Bio13 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.001 หมายความว่าพื้นที่ใดในช่วงเดือนที่ฝนตกชุกที่สุดมีปริมาณน้ำฝนมากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะลดลง

ตัวแปร Bio2 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.01 หมายความว่าพื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) มาก ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงด้วย

ตัวแปร Aspect, Humidity, Bio5, Bio10 และ Bio14 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.05 สามารถอธิบายได้ดังนี้ ตัวแปรทิศด้านลาด (Aspect) ซึ่งมีค่าตั้งแต่ 0 ถึง 360 องศา จะถูกแบ่งออกเป็น 8 ทิศ ทิศละ 45 องศา ทิศเหนือ, ตะวันออกเฉียงเหนือ, ตะวันออก, ตะวันออกเฉียงใต้, ใต้, ตะวันตกเฉียงใต้, ตะวันตก และตะวันตกเฉียงเหนือตามลำดับ ยังมีตัวแปร

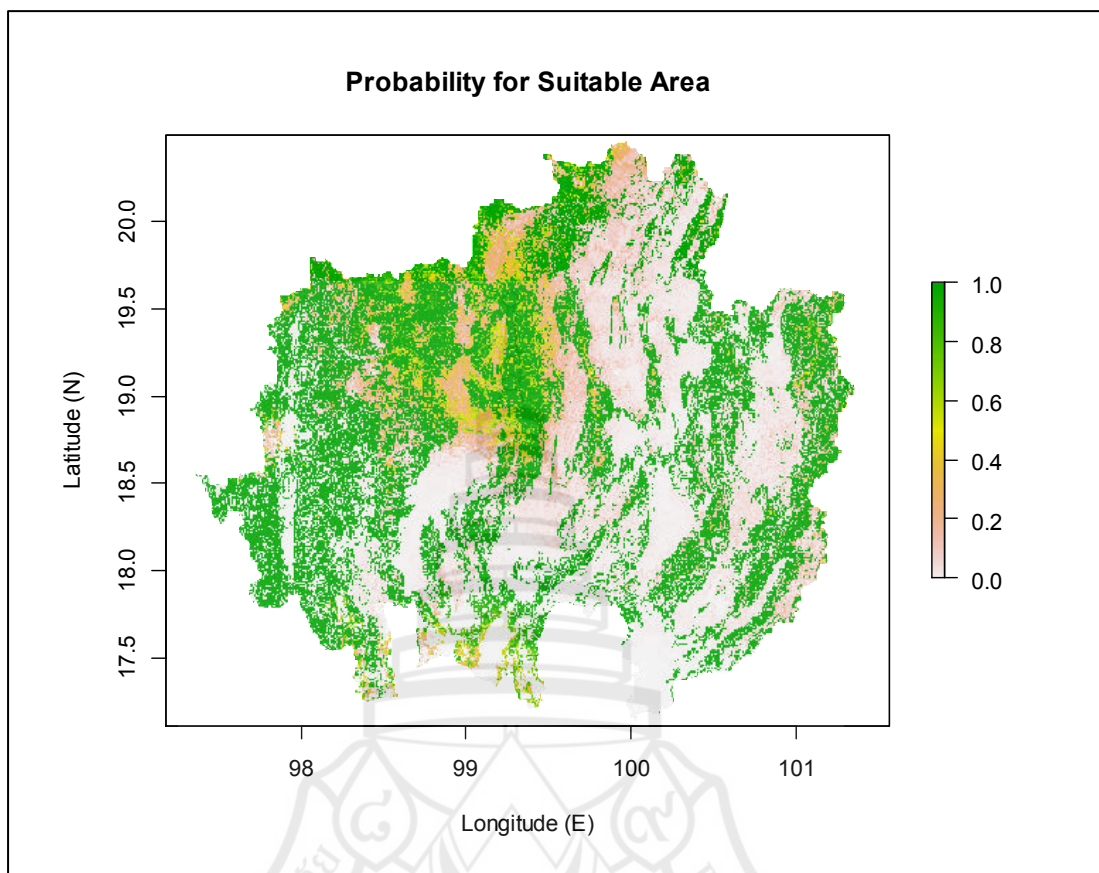
Aspect มีค่าสูงขึ้นทิศด้านลาดก็จะหันหน้าไปทางทิศตะวันตกเฉียง ๆ ดังนั้นถ้าทิศด้านลาด (Aspect) มีค่าสูงขึ้นคือพื้นที่มีการหันหน้าไปทางทิศตะวันตกมากขึ้น โอกาสที่จะค้นพบต้นชาป่าก็น้อยลง

ตัวแปรความชื้นสัมพัทธ์ (Humidity) มีผลเชิงลบหมายความว่า พื้นที่ใดมีความชื้นสัมพัทธ์ (Humidity) สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลง เช่นเดียวกับพื้นที่ใดมีอุณหภูมิเฉลี่ยในเดือนที่ร้อนที่สุด (Bio5) สูง ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) มีผลเชิงลบหมายความว่า พื้นที่ใดในช่วงไตรมาสที่ร้อนที่สุดมีอุณหภูมิเฉลี่ยสูงเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง เช่นเดียวกับตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) หากพื้นที่ใดมีปริมาณน้ำฝนมากเกินไปโอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงเช่นกัน

ส่วนตัวแปร Bio19 ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ ซึ่งมีผลเชิงลบหมายความว่า พื้นที่ใดในช่วงไตรมาสที่หนาวที่สุดแล้วมีปริมาณน้ำฝนที่มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลง





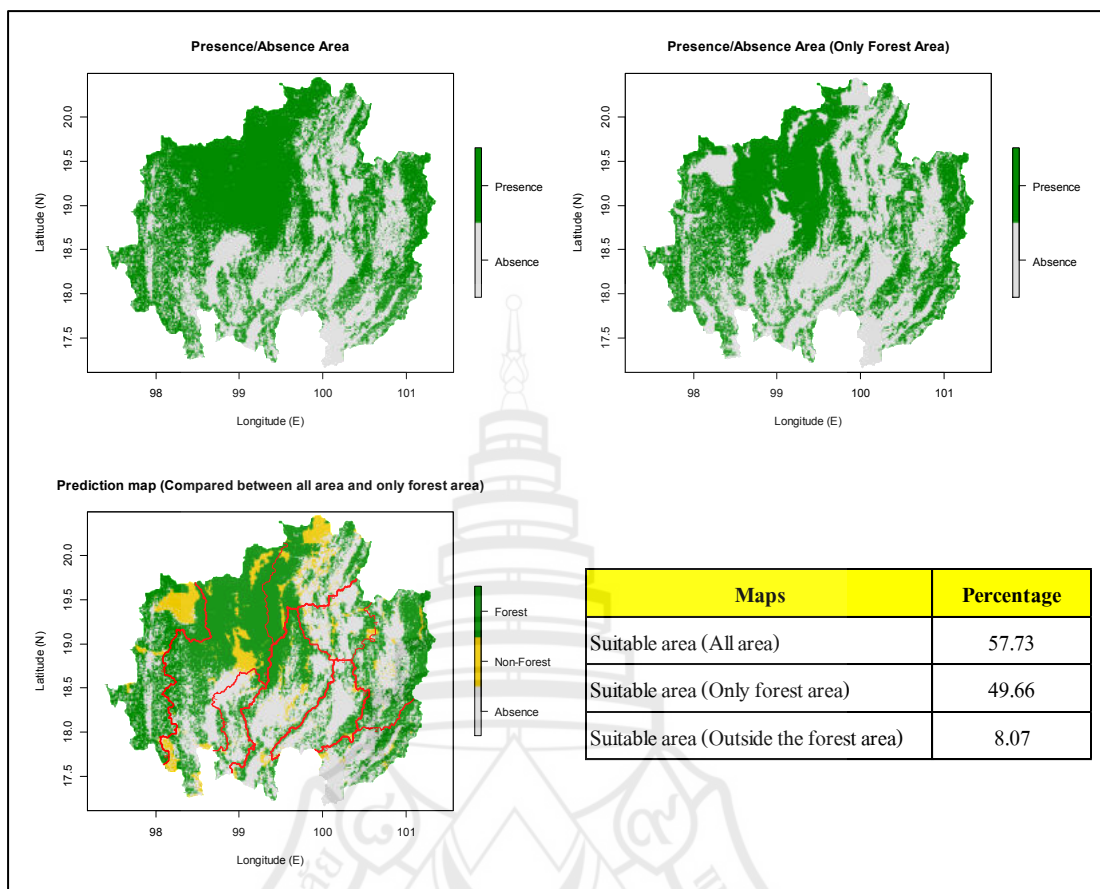
ภาพที่ 4.6 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence

จากภาพที่ 4.6 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิอากาศที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่ หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ 4.12 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการ
สุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	37
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	369
AUC	0.89284
max TPR+TNR	-2.53361
Threshold	0.07354

จากตารางที่ 4.12 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้วพบว่าจำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 37 จุด จำนวนจุด Pseudo-Absence ใช้ทั้งหมด 369 จุด มีค่า AUC เท่ากับ 0.89284 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 89.28 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.07354 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.7 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.7 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence โดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในเขตพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 58 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 50 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่า

มีอยู่ประมาณร้อยละ 8 ของพื้นที่ทั้งหมด ซึ่งส่วนใหญ่อยู่ในจังหวัด แม่ฮ่องสอน เชียงใหม่และ เชียงราย ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด ในเขตพื้นที่ภาคเหนือตอนบน เมื่อพิจารณาพื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่อยู่ทางทิศตะวันตกเฉียงเหนือของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่อย่างหนาแน่นในป่าแถว ๆ จังหวัดเชียงใหม่ตอนบนและกระจายอยู่โดยทั่วไปในพื้นที่ป่าของจังหวัด แม่ฮ่องสอน และอยู่อย่างหนาแน่นในป่ารอยต่อของจังหวัดเชียงรายและเชียงใหม่ โดยมีบางส่วนกระจายอยู่ทางป่าชายแดนทางทิศตะวันออกของพื้นที่แถว ๆ จังหวัดน่าน

4.2.1.3 GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

การสุ่มจุด Pseudo-Absence โดยวิธีนี้ในขั้นตอนแรกจะสร้างพื้นที่ที่มีความเหมาะสมต่อการกระจายของต้นชาป่าโดยวิธีการ BIOCLIM โดยสร้างเป็นแผนที่แสดงพื้นที่ที่มีความเหมาะสมจำนวน 5 แผนที่ จากนั้นในขั้นตอนที่ 2 จะทำการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่ที่มีความเหมาะสมในแต่ละแผนที่ โดยในแต่ละแผนที่จะสุ่มจุด Pseudo-Absence จำนวน 20 จุด ซึ่งในแต่ละจุดจะถูกนำไปสร้างแบบจำลองเพื่อหาแบบจำลองที่เหมาะสม โดยการหาจุด Pseudo-Absence ที่ให้ค่า AUC ที่เข้าใกล้ค่าเฉลี่ยมากที่สุดเป็นตัวแทนในแต่ละแผนที่แสดงพื้นที่ที่มีความเหมาะสม จากนั้นหาค่าของการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุด การหากลุ่ม Testing Data ที่ดีที่สุด และการหาแบบจำลองที่ดีที่สุดใช้วิธีการพิจารณาเช่นเดียวกับแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในวิธีการที่ 1 และ 2 พบว่าแผนที่แสดงพื้นที่ที่มีความเหมาะสมที่ 1 มีความเหมาะสมต่อการสร้างแบบจำลอง โดยมีจุด Pseudo-Absence ที่ 1 ที่ให้ค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด มีรอบการทำซ้ำที่ 3 ที่ให้ค่า Cross-Validation Error น้อยที่สุด และมีกลุ่ม Testing Data กลุ่มที่ 3 ที่ให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate น้อยที่สุด โดยมีแบบจำลองที่ 681 ถูกเลือกเป็นตัวแทนที่เหมาะสม (ดูรายละเอียดได้ที่ภาคผนวก ก) เมื่อได้แบบจำลองที่เป็นตัวแทนที่เหมาะสมแล้ว ขั้นตอนต่อไปคือการคัดเลือกตัวแปรด้วยวิธีการของ Stepwise โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) มาคัดเลือกตัวแปร

ตารางที่ 4.13 แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน ซึ่งมีการ
คัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว

Factors	Estimate	Standard Error	Z Test
Intercept	-176.800	844.500	-0.209
Aspect	0.003	0.001	2.526 *
Distance	-0.390	0.079	-4.950 ***
Humidity	-0.110	0.077	-1.440
Bio1	0.314	0.154	2.039 *
Bio2	0.260	0.173	1.501
Bio3	0.538	0.312	1.722 .
Bio6	-0.917	0.138	-6.671 ***
Bio8	0.532	0.076	7.035 ***
Bio12	0.042	0.009	4.891 ***
Bio13	0.064	0.021	3.093 **
Bio14	-0.627	0.150	-4.182 ***
Bio16	-0.107	0.022	-4.945 ***
Bio17	0.156	0.040	3.853 ***
Class1	17.360	844.100	0.021
Class2	17.920	844.100	0.021
Class5	17.680	844.100	0.021

ตารางที่ 4.13 (ต่อ)

Factors	Estimate	Standard Error	Z Test
Class6	17.300	844.100	0.020
Class7	-70.300	3370.000	-0.021
Null Deviance = 975.20, Residual Deviance = 430.55			
Pseudo-R ² = 55.85, AIC = 468.55			

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 4.13 แสดงแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว พบว่ามีตัวแปรเข้าไปในแบบจำลองคือ ทิศด้านลาด (Aspect), ระยะห่างจากแม่น้ำ (Distance), ความชื้นสัมพัทธ์ (Humidity), อุณหภูมิเฉลี่ยรายปี (Bio1), ช่วงความแตกต่างของอุณหภูมิรายวัน (Bio2), Isothermality (Bio3), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5, Class6 และ Class7) ซึ่งมีตัวแปร Distance, Bio6, Bio8, Bio12, Bio14, Bio16 และ Bio17 มีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Bio13 มีนัยสำคัญทางสถิติที่ 0.01 ตัวแปร Aspect และ Bio1 มีนัยสำคัญทางสถิติที่ 0.05 และตัวแปร Bio3 มีนัยสำคัญทางสถิติที่ 0.10

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L₀) เท่ากับ 975.20 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L₁) เท่ากับ 430.55 ซึ่งสามารถคำนวณหาค่า Pseudo-R² ได้เท่ากับ 55.85 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 55.85 ได้สมการพยากรณ์ดังนี้

$$\begin{aligned}\text{Logit}(\mu) = & -176.800 + 0.003\text{Aspect} - 0.390\text{Distance} - 0.110\text{Humidity} \\ & + 0.314\text{Bio1} + 0.260\text{Bio2} + 0.538\text{Bio3} - 0.917\text{Bio6} + 0.532\text{Bio8} \\ & + 0.042\text{Bio12} + 0.064\text{Bio13} - 0.627\text{Bio14} - 0.107\text{Bio16} + 0.156\text{Bio17} \\ & + 17.360\text{Class1} + 17.920\text{Class2} + 17.680\text{Class5} + 17.300\text{Class6} \\ & - 70.300\text{Class7}\end{aligned}$$

จากสมการพบว่าตัวแปร Aspect, Bio1, Bio2, Bio3, Bio8, Bio12, Bio13, Bio17, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Distance, Humidity, Bio6, Bio14, Bio16 และ Class7 มีค่าสัมประสิทธิ์เป็นลบ แสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio8, Bio12 และ Bio17 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.001 สามารถอธิบายได้ดังต่อไปนี้

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8) มีผลเชิงบวก หมายความว่าพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่มีมากขึ้น

ตัวแปรปริมาณน้ำฝนรายปี (Bio12) อธิบายได้ว่าในพื้นที่ที่มีปริมาณน้ำฝนรายปีที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วยเช่นกัน นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ตัวแปรปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) มีผลเชิงบวกสามารถอธิบายได้ว่าพื้นที่ใดที่มีปริมาณน้ำฝนสูง ๆ ในช่วงไตรมาสที่แห้งแล้งที่สุด จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่มากขึ้น

ตัวแปร Bio13 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.01 อธิบายได้ว่าพื้นที่ใดในช่วงเดือนที่ฝนตกชุกที่สุดแล้วมีปริมาณน้ำฝนที่เพียงพอ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะมากขึ้น

ตัวแปร Aspect และ Bio1 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.05 อธิบายได้ว่า ตัวแปรทิศด้านลาด (Aspect) มีค่าตั้งแต่ 0 ถึง 360 องศา ซึ่งถูกแบ่งออกเป็น 8 ทิศ ทิศละ 45 องศา ทิศเหนือ, ตะวันออกเฉียงเหนือ, ตะวันออก, ตะวันออกเฉียงใต้, ใต้, ตะวันตกเฉียงใต้, ตะวันตก และตะวันตกเฉียงเหนือตามลำดับ ยิ่งค่า Aspect มีค่ามาก ทิศด้านลาดก็จะหันหน้าไปทางทิศตะวันตกเรื่อย ๆ ดังนั้นถ้า Aspect มีค่าสูงขึ้นคือพื้นที่มีการหันหน้าไปทางทิศตะวันตกเฉียงเหนือมากขึ้น โอกาสที่จะค้นพบต้นชาป่าก็มีมากขึ้น เช่นเดียวกับ ตัวแปร Bio1 ที่มีผลต่อแบบจำลองเชิง

บวก อธิบายได้ว่า พื้นที่ใดมีอุณหภูมิเฉลี่ยรายปีเพียงพอ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่ก็มีมากขึ้นเช่นกัน

ตัวแปร Bio3 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.10 อธิบายได้ว่า ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อความแตกต่างของอุณหภูมิรายปี (Bio7) ซึ่งได้มาจากผลต่างระหว่างอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุด (Bio5) ลบกับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุด (Bio6) นั่นคือยิ่งค่า Isothermality (Bio3) มีค่ามาก หมายถึงมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

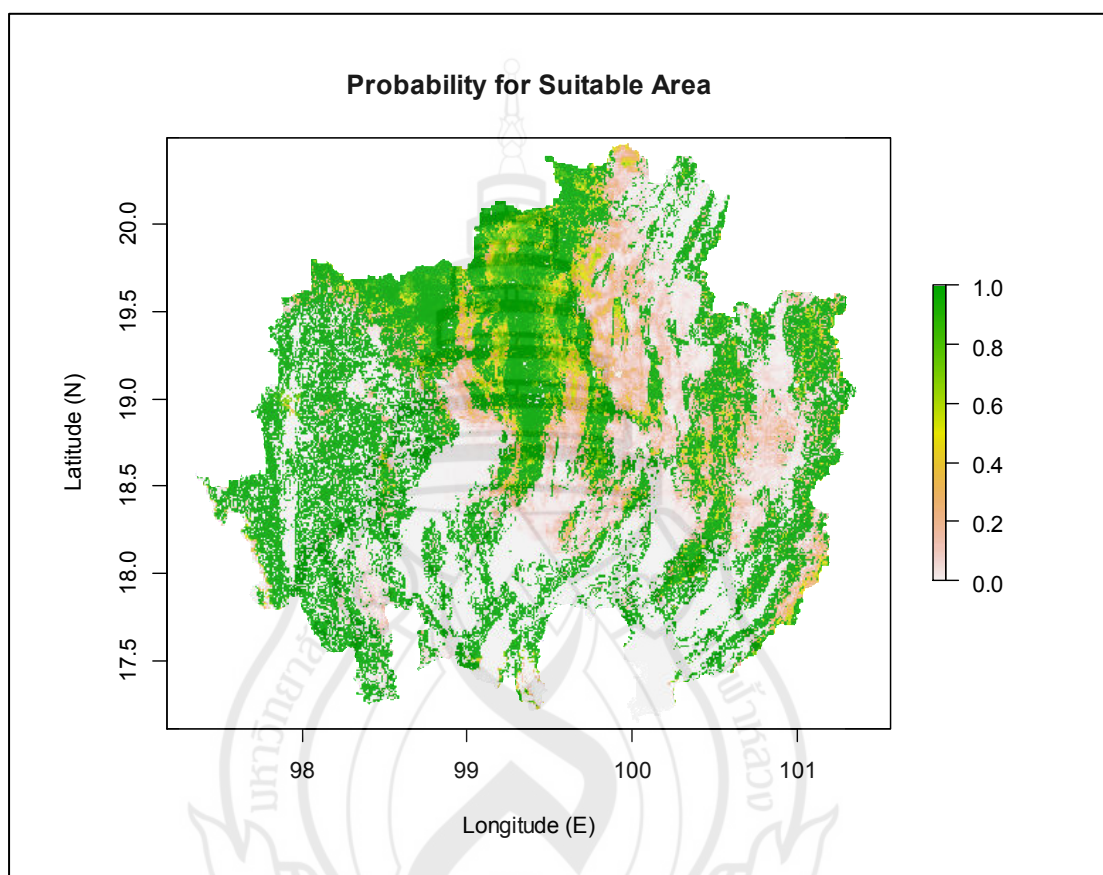
ส่วนตัวแปร Bio2, Class1, Class2, Class5 และ Class6 ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ ซึ่งสามารถอธิบายได้ว่า ตัวแปร Bio2 มีผลต่อแบบจำลองเชิงบวกหมายความว่าพื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) มาก ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะมากขึ้นด้วย เช่นเดียวกับกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) มีผลเชิงบวกหมายความว่าพื้นที่ใดที่มีความลาดชันไม่เกิน 30 องศาและมีโครงสร้างทางธรณีวิทยาของชนิดหินประเภทใดก็ได้ จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นมากขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Distance, Bio6, Bio14 และ Bio16 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.001สามารถอธิบายได้ดังต่อไปนี้

ตัวแปร Distance มีผลเชิงลบหมายความว่า พื้นที่ใดที่อยู่ห่างจากแม่น้ำมากเท่าไร โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลง เช่นเดียวกับพื้นที่ใดมีอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6) ต่ำเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลง หรือพื้นที่ใดมีปริมาณน้ำฝนในช่วงเดือนที่แห้งแล้งที่สุด (Bio14) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลงเช่นกัน ส่วนตัวแปร Bio16 ก็มีผลเชิงลบเช่นเดียวกัน หมายความว่าพื้นที่ใดมีปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลง

ส่วนตัวแปร Humidity และ Class7 ถึงแม้จะทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในแบบจำลองโดยอิทธิพลจากตัวแปรตัวอื่น ๆ ซึ่งสามารถอธิบายได้ดังต่อไปนี้ ตัวแปรความชื้นสัมพัทธ์ (Humidity) มีผลเชิงลบหมายความว่า พื้นที่ใดมีความชื้นสัมพัทธ์ (Humidity) สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่

ก็ลดลง เช่นเดียวกับพื้นที่ใดที่มีอุณหภูมิเฉลี่ยในเดือนที่ร้อนที่สุด (Bio5) สูง ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง หรือพื้นที่ใดที่อยู่ในเขตชั้นหินอัคนีและมีความลาดชันเกิน 30 องศา พื้นที่นั้นจะมีโอกาสค้นพบต้นชาป่าน้อยลงเช่นเดียวกัน



ภาพที่ 4.8 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

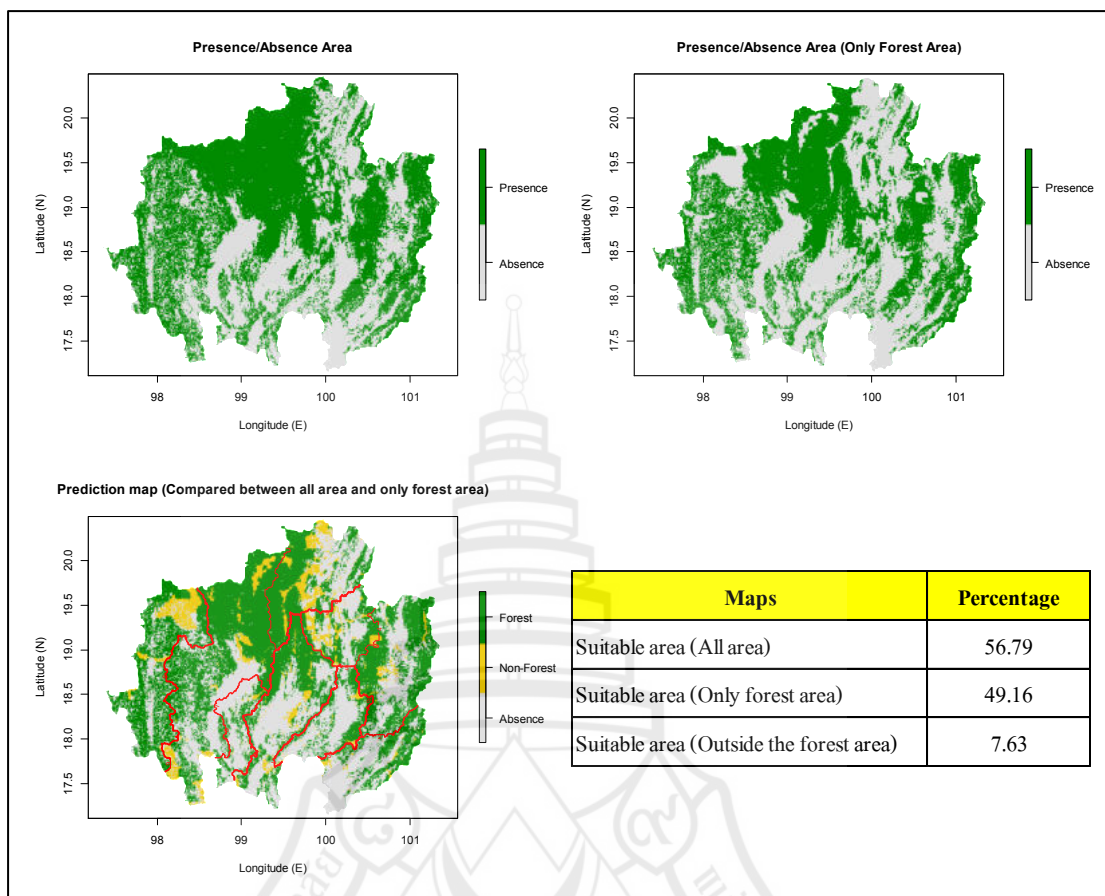
จากภาพที่ 4.8 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละ พิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึงพื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 4.14 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง GLMs ที่เกิดจากการ สุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	366
AUC	0.95262
max TPR+TNR	-1.90547
Threshold	0.12949

จากตารางที่ 4.14 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GLMs ที่เกิด จากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน ซึ่งมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absence ใช้ทั้งหมด 366 จุด มีค่า AUC เท่ากับ 0.95262 หมายความว่ามีความ ถูกต้องของการพยากรณ์เท่ากับร้อยละ 95.26 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไป คำนวณหาค่า Threshold ได้เท่ากับ 0.12949หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็น น้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็น มากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.9 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมด และเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.9 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน โดยมีการคัดเลือกตัวแปรด้วยวิธีการ Stepwise แล้ว ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน โดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในเขตพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 57 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 49 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในบริเวณพื้นที่ป่ามีอยู่ประมาณร้อยละ 8 ของพื้นที่

ทั้งหมด ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอนและเชียงราย ดังแสดงในภาพซ้ายมือล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัดในเขตพื้นที่ภาคเหนือตอนบน เมื่อพิจารณาพื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่อยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่ในป่าแถว ๆ จังหวัดเชียงใหม่ ตอนบนและป่ารอยต่อระหว่างจังหวัดเชียงใหม่-เชียงราย บางส่วนกระจายอยู่โดยทั่วไปในพื้นที่ป่าของจังหวัดแม่ฮ่องสอน นอกจากนี้ยังพบว่ามีบางส่วนกระจายอยู่ทางป่าชายแดนทางทิศตะวันออกของพื้นที่แถว ๆ จังหวัดน่านและป่ารอยต่อระหว่างจังหวัดพะเยาและจังหวัดน่าน

4.2.2 แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs เป็นวิธีการสร้างแบบจำลองโดยใช้เทคนิคการจำแนกกลุ่มเช่นเดียวกับแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs แต่ต่างกันตรงที่ฟังก์ชันการเชื่อมโยง (Link Functions) โดยในวิธีการของ GLMs ฟังก์ชันการเชื่อมโยง จะอยู่ในรูปแบบเชิงเส้น (Linear Model) ในขณะที่ฟังก์ชันการเชื่อมโยงของ GAMs จะไม่อยู่ในรูปแบบเชิงเส้น แต่จะอยู่รูปแบบของแบบจำลองเชิงบวก (Additive Model) ที่ประกอบไปด้วยฟังก์ชันปรับเรียบ (Smooth Functions) ของตัวแปรที่นำมาศึกษา โดยจะนำตัวแปรจากแบบจำลองที่ถูกคัดเลือกโดยวิธี GLMs ในแต่ละวิธีของการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธีมาปรับปรุงแบบจำลองโดยวิธี GAMs

4.2.2.1 GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

จากแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา พบว่ามีตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 คือ ความสูงจากระดับน้ำทะเล (DEM), ความชื้นสัมพัทธ์ (Humidity), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) และปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) ซึ่งในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs จะทำการสร้างฟังก์ชันการเชื่อมโยง โดยนำตัวแปรมาสร้าง Smooth Functions ก่อนนำไปสร้างแบบจำลอง จากนั้นทำการตัด Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองแล้วตรวจสอบค่า AIC อีกครั้ง หากพบว่าค่า AIC มีค่าลดลงแสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นสามารถตัดออกจากแบบจำลองได้ แต่ถ้าหากตัด Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองแล้วพบว่าค่า AIC ไม่ได้ลดลงแสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นยังมีอิทธิพลต่อแบบจำลองการกระจายของสิ่งมีชีวิต จึงไม่สามารถตัดออกจากแบบจำลองได้

ตารางที่ 4.15 ค่าสถิติทดสอบในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

Smooth Functions	Chi-Square Test	
	GAMs	Improved GAMs
s(DEM)	36.48 ***	41.09 ***
s(Humidity)	30.87 ***	28.20 ***
s(Bio3)	8.10 **	7.77 **
s(Bio4)	24.35 ***	26.91 ***
s(Bio8)	40.18 ***	42.14 ***
s(Bio12)	21.10 ***	24.63 ***
s(Bio14)	4.34 **	7.72 **
s(Bio16)	1.87	
Linear Functions	Z Test	
	GAMs	Improved GAMs
Bio16		-1.95 .
AIC	652.95	652.95
R ²	32.00	32.00

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 4.15 แสดงค่าสถิติในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธีการ GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 จากแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษามาสร้างแบบจำลอง GAMs พบว่า ค่า AIC มีค่า

เท่ากับ 652.9531 โดยมีค่าลดลงจากวิธีการ GLMs ส่วนค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination: R^2) มีค่าเท่ากับ 32.0 โดยมีค่าเพิ่มขึ้นจากวิธีการ GLMs แต่จะสังเกตได้ว่า Smooth Functions ของตัวแปร Bio16 ไม่มีผลนัยสำคัญทางสถิติ จึงทำการปรับปรุงแบบจำลองใน 2 แนวทาง โดยแนวทางแรก จะทำการตัด Smooth Functions ของตัวแปรที่ไม่มีผลนัยสำคัญทางสถิติ ออกจากแบบจำลอง จากนั้นตรวจสอบค่า AIC ส่วนในแนวทางที่สองจะเปลี่ยน Smooth Functions ของตัวแปรที่ไม่มีผลนัยสำคัญทางสถิติเป็นฟังก์ชันเชิงเส้น (Linear function) ของตัวแปรนั้น จากนั้นตรวจสอบค่า AIC

เมื่อเปลี่ยน Smooth Functions ของตัวแปร Bio16 เป็นฟังก์ชันเชิงเส้น พบว่าจะมีค่า AIC ต่ำสุด แสดงว่าตัวแปร Bio16 มีผลต่อแบบจำลองในลักษณะเชิงเส้น เมื่อพิจารณาค่า R^2 พบว่ามีค่าไม่แตกต่างกัน ดังนั้นจึงเลือกแบบจำลอง GAMs ที่มีการปรับปรุงแล้วไปวิเคราะห์ในขั้นตอนต่อไป ซึ่งจะพิจารณาค่าองศาความเป็นอิสระ (Degree of Freedom) ของ Smooth Functions ถ้าหากค่าองศาความเป็นอิสระของ Smooth Function มีค่าเท่ากับ 1 แสดงว่าปัจจัยนั้นจะส่งผลกระทบต่อแบบจำลองในลักษณะฟังก์ชันเชิงเส้น (Linear Function) ดังนั้นจึงต้องเปลี่ยน Smooth Functions นั้นให้อยู่ในรูปแบบของฟังก์ชันเชิงเส้น

ตารางที่ 4.16 แบบจำลอง GAMs และค่าองศาความเป็นอิสระของ Smooth Function ในแบบจำลองที่มีการปรับปรุงทั้ง 2 ครั้ง

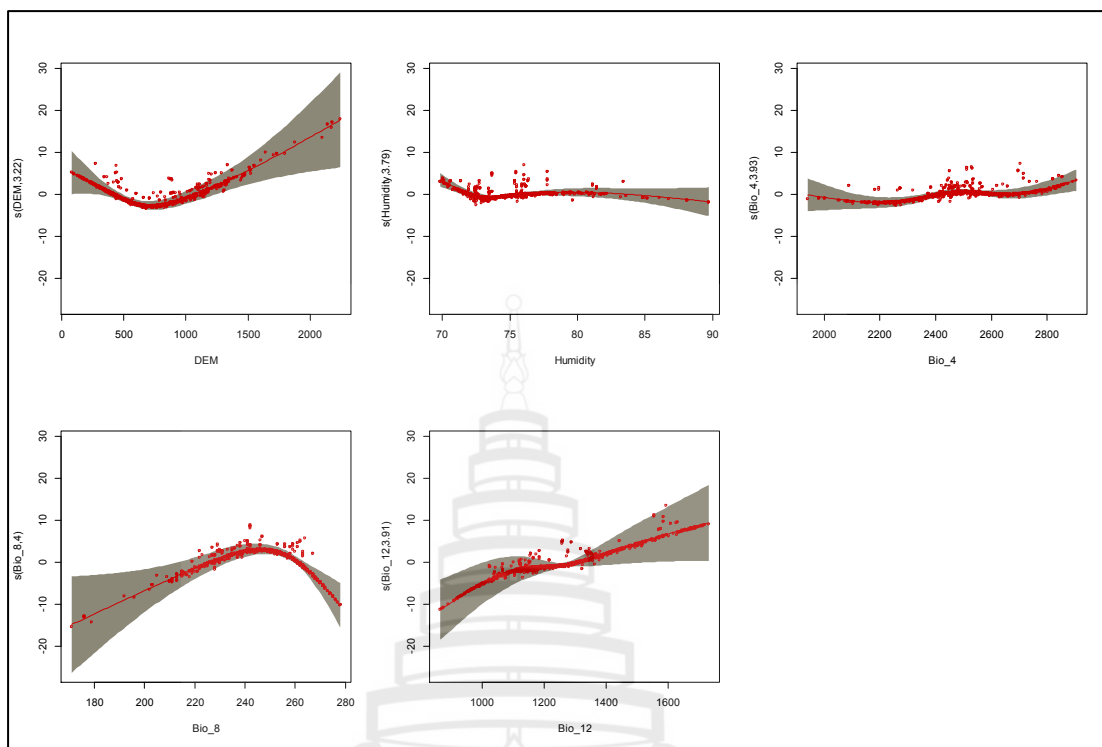
Smooth Functions	Improved GAMs (1)			Improved GAMs (2)		
	Degree of Freedom	Chi-Square Test		Degree of Freedom	Chi-Square Test	
s(DEM)	3.22	41.09 ***		3.22	36.86 ***	
s(Humidity)	3.79	28.20 ***		3.79	27.95 ***	
s(Bio3)	1.00	7.77 **				
s(Bio4)	3.93	26.91 ***		3.93	26.26 ***	
s(Bio8)	4.00	42.14 ***		4.00	35.69 ***	

ตารางที่ 4.16 (ต่อ)

Smooth Functions	Improved GAMs (1)		Improved GAMs (2)	
	Degree of Freedom	Chi-Square Test	Degree of Freedom	Chi-Square Test
s(Bio12)	3.91	24.63 ***	3.91	24.61 ***
s(Bio14)	1.00	7.72 **		
Linear Functions	Estimate	Z Test	Estimate	Z Test
(Intercept)	15.03	1.60	-17.69	-1.53
Bio3			0.66	2.80 **
Bio14			-0.23	-3.05 **
Bio16	-0.03	-1.95 .	-0.03	-1.95 .
R^2	32.00		32.00	

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 4.16 แสดงแบบจำลอง GAMs และค่าองค์การความเป็นอิสระของ Smooth Function ในแบบจำลองที่มีการปรับปรุงทั้ง 2 ครั้งพบว่า แบบจำลอง GAMs ที่มีการปรับปรุงครั้งแรกนั้น มี Smooth Functions ของตัวแปร Bio3 และ Bio14 ที่มีค่าองค์การความเป็นอิสระเท่ากับ 1 แสดงว่าตัวแปร Bio3 และ Bio4 ส่งผลกระทบบต่อแบบจำลองในลักษณะฟังก์ชันเชิงเส้น ดังนั้นจึงทำการปรับปรุงครั้งที่สอง โดยเปลี่ยน Smooth Functions ของตัวแปร Bio3 และ Bio14 ให้อยู่ในรูปแบบของฟังก์ชันเชิงเส้นแทน ซึ่งพบว่าตัวแปรทั้ง 2 ตัวต่างมีนัยสำคัญทางสถิติที่ 0.01 โดยที่ค่า R^2 ยังคงมีค่าเท่าเดิม



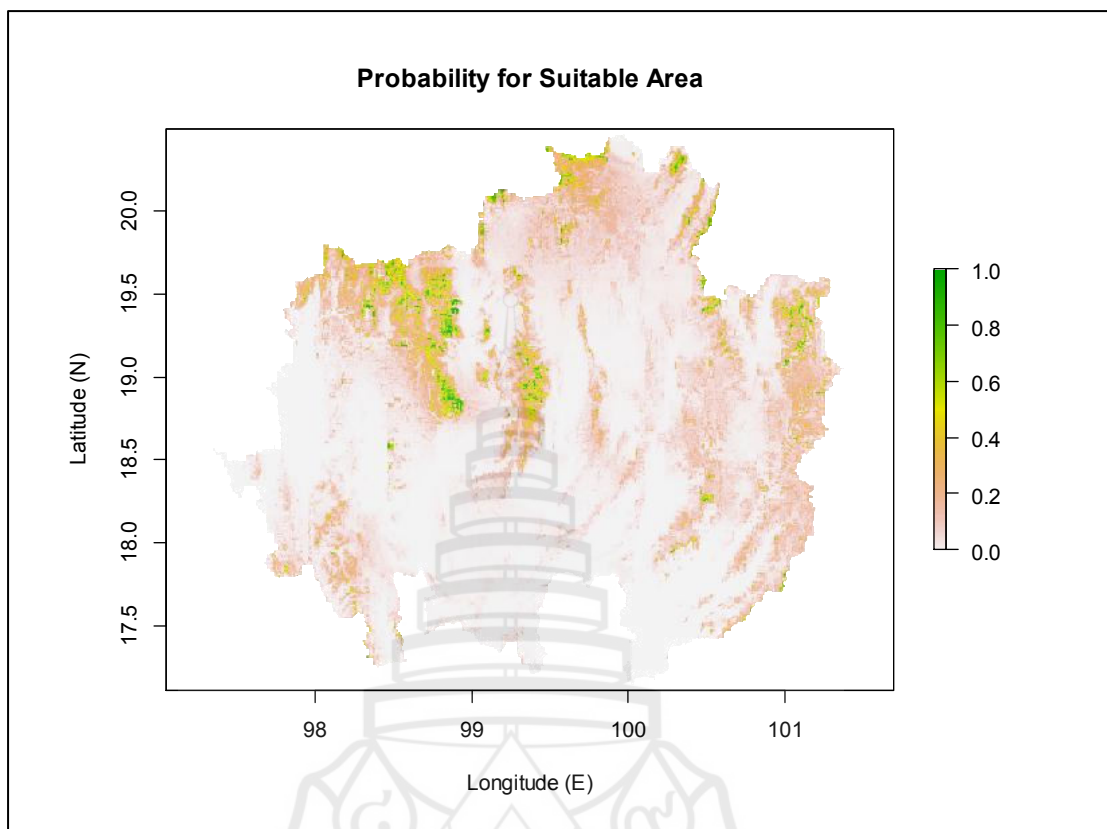
ภาพที่ 4.10 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

จากภาพที่ 4.10 แสดงเส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร DEM, Humidity, Bio4, Bio8 และ Bio12 ในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว โดยเลือกใช้วิธี Cubic Regression Splines ในการสร้าง Smooth Functions ซึ่งกำหนดให้จุด Knot เท่ากับ 5 โดยแกน x คือค่าของตัวแปรและแกน y คือค่าของการพยากรณ์ที่เกิดจาก Smooth Functions ซึ่งค่าแกน y จะอยู่ในรูปแบบของ Additive Component (บทที่ 2 หน้า 33) ซึ่งอยู่ในรูปแบบเชิงเส้น ดังนั้นค่ากลาง (Center Value) จะมีค่าเท่ากับ 0 โดยแถบที่แรเงาเป็นแถบช่วงความเชื่อมั่นที่ร้อยละ 95 ของเส้นกราฟพยากรณ์ (เส้นสีแดง) โดยจุดสีแดงคือค่าความคลาดเคลื่อนบางส่วน (Partial Residual) ที่เกิดจากการประมาณ ซึ่งค่าความคลาดเคลื่อนส่วนใหญ่จะกระจายอยู่ภายในแถบช่วงความเชื่อมั่น จากเส้นกราฟแสดงว่า ตัวแปร DEM, Humidity, Bio4, Bio8 และ Bio12 ส่งผลกระทบท่อบแบบจำลองในลักษณะที่ไม่เป็นเชิงเส้น (Non-Linear) ดังนั้นการสร้าง Smooth Functions ของตัวแปร DEM, Humidity, Bio4, Bio8 และ Bio12 เพื่อใช้ในแบบจำลอง GAMs จึงมีความเหมาะสม

ตารางที่ 4.17 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

Variables	Coefficients	Variables	Coefficients
s(DEM)	-3.348	s(Bio8)	2.165
	-3.497		8.498
	-4.123		13.860
	17.536		-7.940
s(Humidity)	-2.504	s(Bio12)	5.507
	0.429		4.419
	-2.414		7.870
	-1.810		10.641
s(Bio4)	-1.201	Bio3	0.664
	0.805	Bio14	-0.229
	0.170	Bio16	-0.028
	3.377		
intercept = -17.693			

จากตารางที่ 4.17 แสดงค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว ซึ่งตัวแปร Bio3, Bio14 และ Bio16 จะอยู่ในรูปแบบของเชิงเส้น ในขณะที่ตัวแปร DEM, Humidity, Bio4, Bio8 และ Bio12 จะอยู่ในรูปแบบ Smooth Functions โดยวิธีการ Cubic Regression Splines ซึ่งกำหนดจุด Knot เท่ากับ 5



ภาพที่ 4.11 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

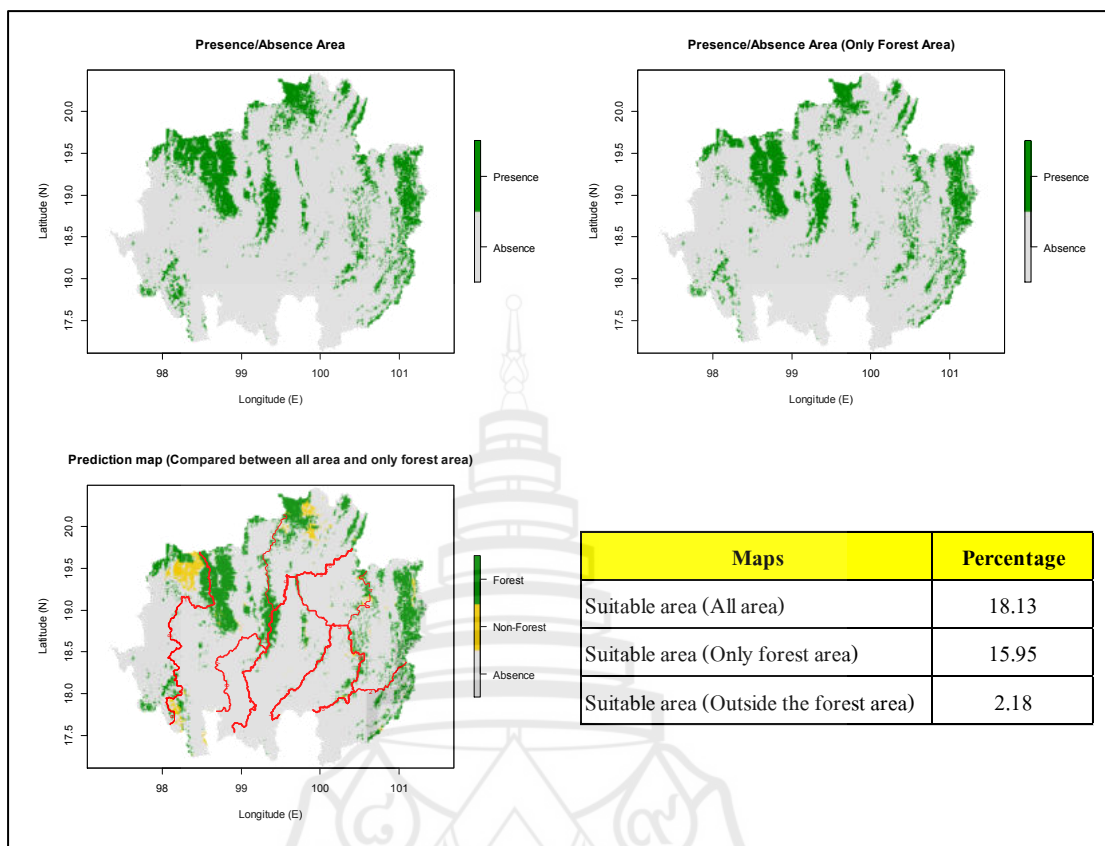
จากภาพที่ 4.11 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 4.18 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	295
AUC	0.83283
max TPR+TNR	-2.04047
Threshold	0.11502

จากตารางที่ 4.18 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absence ใช้ทั้งหมด 295 จุด มีค่า AUC เท่ากับ 0.83283 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 83.28 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.11502 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.12 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้วเปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.12 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 18 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 15 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 2 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอน ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าใน

เขตพื้นที่ป่าพบว่า ส่วนใหญ่จะกระจุกอยู่ทางทิศตะวันตกเฉียงเหนือของพื้นที่ ซึ่งอยู่แถว ๆ ป่าในเขตจังหวัดเชียงใหม่รวมถึงบริเวณป่าที่เป็นพื้นที่รอยต่อระหว่างจังหวัดเชียงใหม่และแม่ฮ่องสอน โดยมีบางส่วนกระจายอยู่ในบริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดเชียงใหม่ ลำปางและเชียงราย และมีบางส่วนกระจายอยู่ทางทิศเหนือของพื้นที่ ซึ่งเป็นบริเวณป่าในเขตจังหวัดเชียงรายตอนบน นอกจากนี้ยังพบว่ามีการกระจายอยู่ทางป่าชายแดนด้านทิศตะวันออกของพื้นที่ในเขตจังหวัดน่าน

4.2.2.2 GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence

จากแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence พบว่ามีตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 คือ ทิศด้านลาด (Aspect), ความชื้นสัมพัทธ์ (Humidity), ช่วงความแตกต่างของอุณหภูมิรายวัน (Bio2), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), ช่วงความแตกต่างของอุณหภูมิรายปี (Bio7), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) และปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) ซึ่งในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs จะทำการสร้างฟังก์ชันการเชื่อมโยง โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 มาสร้าง Smooth Functions ก่อนนำไปสร้างแบบจำลอง จากนั้นตรวจสอบค่า AIC โดยทำการตัด Smooth functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองเพื่อตรวจสอบค่า AIC อีกครั้ง หากพบว่าค่า AIC มีค่าลดลง แสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นสามารถตัดออกจากแบบจำลองได้ แต่ถ้าหากตัด Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองแล้วพบว่าค่า AIC ไม่ได้ลดลง แสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นยังมีอิทธิพลต่อแบบจำลองการกระจายของสิ่งมีชีวิต จึงไม่สามารถตัดออกจากแบบจำลองได้

ตารางที่ 4.19 ค่าสถิติทดสอบ Chi-Square และค่าองศาความเป็นอิสระของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence

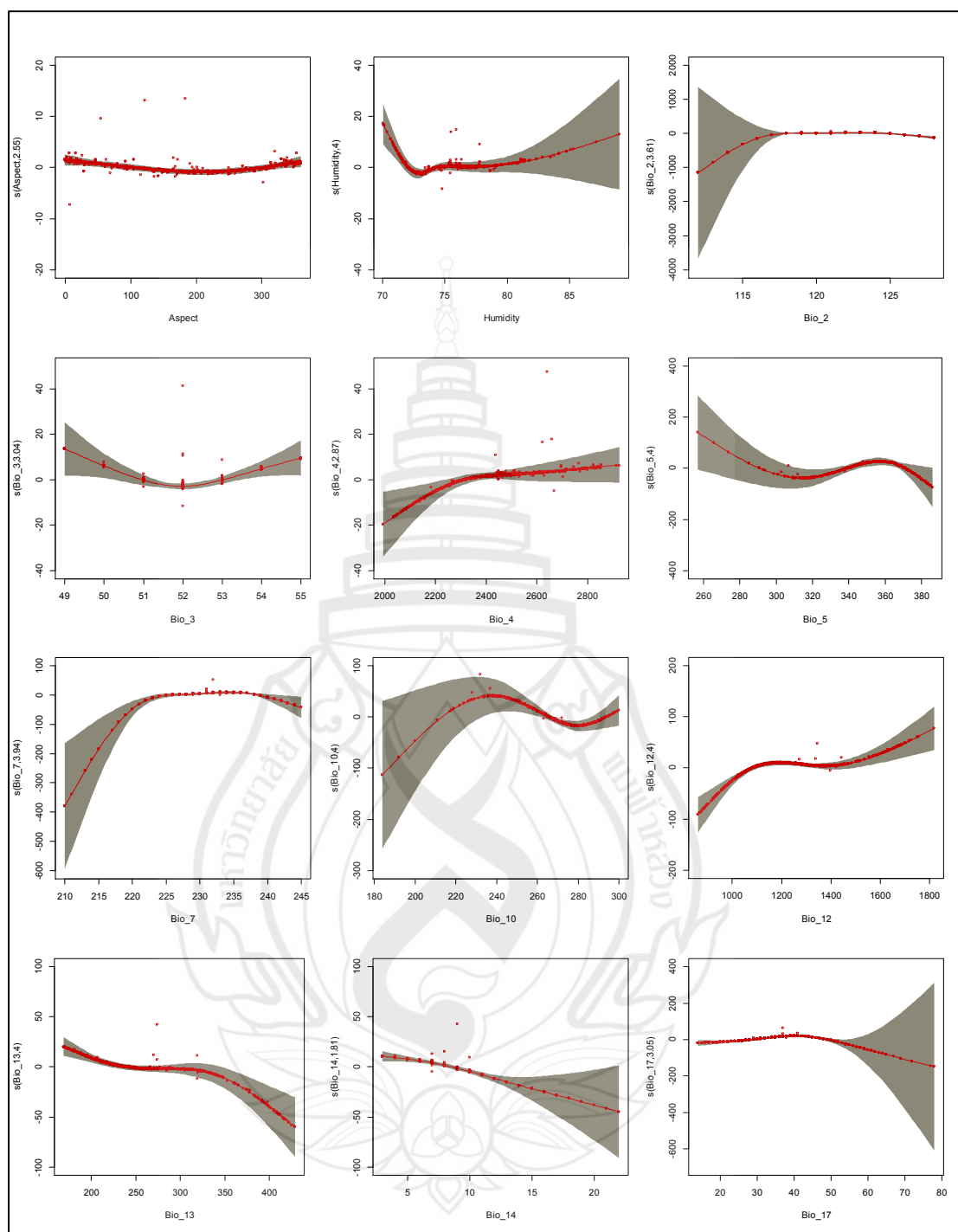
Smooth Functions	Degree of Freedom	Chi-Square Test
s(Aspect)	2.55	14.33 **
s(Humidity)	4.00	20.73 ***
s(Bio2)	3.61	28.64 ***
s(Bio3)	3.04	23.39 ***
s(Bio4)	2.87	8.17 *
s(Bio5)	4.00	26.07 ***
s(Bio7)	2.94	21.44 ***
s(Bio10)	4.00	22.36 ***
s(Bio12)	4.00	36.10 ***
s(Bio13)	4.00	25.74 ***
s(Bio14)	1.81	36.38 ***
s(Bio17)	3.05	55.02 ***
AIC	285.10	
R ²	78.00	

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 4.19 แสดงค่าสถิติทดสอบ Chi-Square และค่าองศาความเป็นอิสระ (Degree of Freedom) ของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธีการ GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 ของแบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence มาสร้างแบบจำลอง GAMs พบว่าค่า AIC มีค่าเท่ากับ

285.095 โดยมีค่าลดลงจากวิธีการ GLMs มีค่า R^2 มีค่าเท่ากับร้อยละ 78.0 โดยมีค่าเพิ่มขึ้นจากวิธีการ GLMs และจะสังเกตได้ว่าทุกตัวแปรต่างก็มีนัยสำคัญทางสถิติ และเมื่อพิจารณาค่าองศาความเป็นอิสระพบว่าไม่มีตัวแปรไหนเลยที่มีค่าองศาความเป็นอิสระเท่ากับหนึ่ง แสดงว่าไม่มีตัวแปรใดที่ส่งผลกระทบต่อแบบจำลองในลักษณะเชิงเส้น ดังนั้นจึงไม่ต้องมีการปรับปรุงแบบจำลอง

จากภาพที่ 4.13 แสดงเส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Aspect, Humidity, Bio2, Bio3, Bio4, Bio5, Bio7, Bio10, Bio12, Bio13, Bio14 และ Bio17 ในแบบจำลอง GAMs โดยเลือกใช้วิธี Cubic Regression Splines ในการสร้าง Smooth Functions ซึ่งกำหนดให้จุด Knot เท่ากับ 5 โดยแกน x คือค่าของตัวแปรและแกน y คือค่าของการพยากรณ์ที่เกิดจาก Smooth Functions ซึ่งค่าแกน y จะอยู่ในรูปแบบของ Additive Component (บทที่ 2 หน้า 33) ซึ่งอยู่ในรูปแบบเชิงเส้น ดังนั้นค่ากลาง (Center Value) จะมีค่าเท่ากับ 0 โดยแถบที่แรเงาเป็นแถบช่วงความเชื่อมั่นที่ ร้อยละ 95 ของเส้นกราฟพยากรณ์ (เส้นสีแดง) ของแต่ละตัวแปร โดยจุดสีแดงคือค่าความคลาดเคลื่อนบางส่วน (Partial Residual) ที่เกิดจากการประมาณ ซึ่งค่าความคลาดเคลื่อนส่วนใหญ่จะกระจายอยู่ภายในแถบช่วงความเชื่อมั่น จากเส้นกราฟแสดงว่า ตัวแปร Aspect, Humidity, Bio2, Bio3, Bio4, Bio5, Bio7, Bio10, Bio12, Bio13, Bio14 และ Bio17 ส่งผลกระทบต่อแบบจำลองในลักษณะที่ไม่เป็นเชิงเส้น (Non-Linear) ดังนั้นการสร้าง Smooth Functions ของตัวแปรแต่ละตัวเพื่อใช้ในแบบจำลอง GAMs จึงมีความเหมาะสม

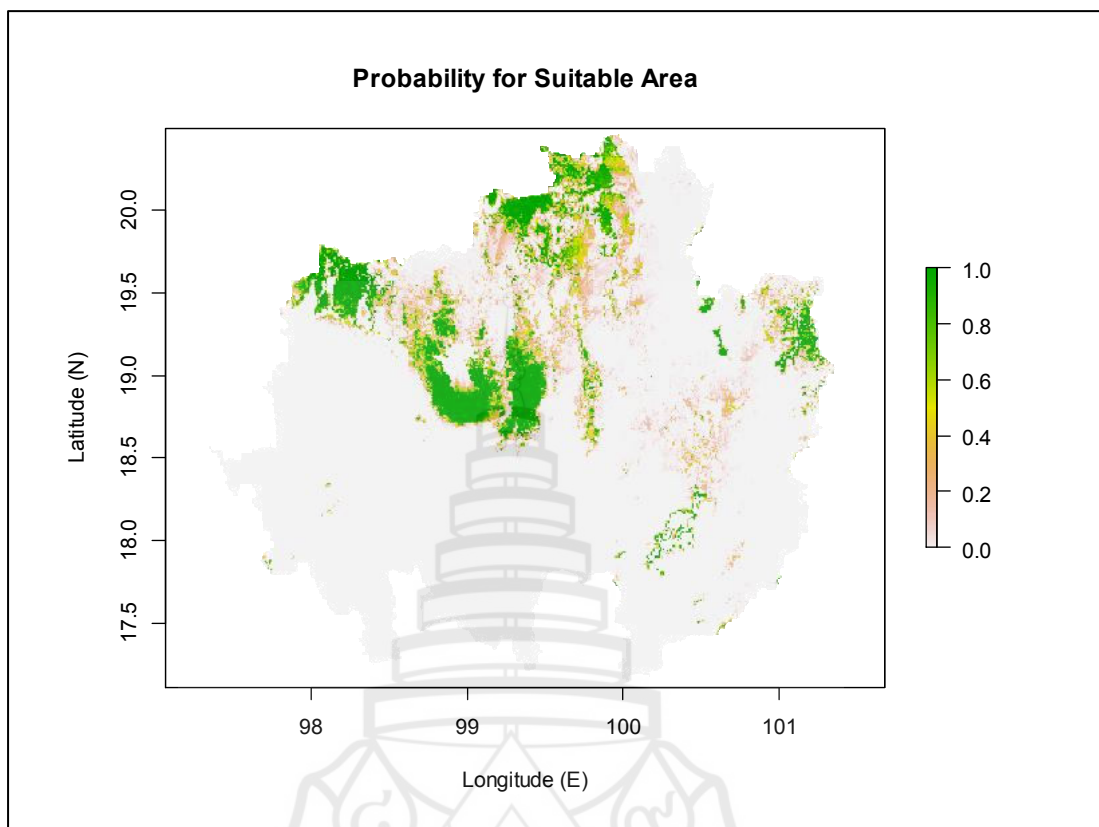


ภาพที่ 4.13 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs

ตารางที่ 4.20 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs

Variables	Coefficients	Variables	Coefficients	Variables	Coefficients
s(Aspect)	-0.604	s(Bio4)	13.243	s(Bio12)	55.996
	-1.397		13.923		41.473
	-1.141		7.978		52.933
	0.919				83.024
s(Humidity)	-12.795	s(Bio5)	-67.309	s(Bio13)	-12.751
	-1.053		-76.630		-8.419
	-11.891		-84.871		-9.528
	12.627		-97.128		-61.305
s(Bio2)	-145.857	s(Bio7)	-41.497	s(Bio14)	-5.423
	506.583		237.229		-15.741
	1040.171		298.752		-29.369
	-106.866		-18.968		-44.891
s(Bio3)	-1.384	s(Bio10)	60.082	s(Bio17)	12.745
	-13.857		67.756		26.188
	-3.884		77.206		-34.170
	9.374		31.403		-149.787
intercept = -50.175					

จากตารางที่ 4.20 แสดงค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่ตัวแปร Aspect, Humidity, Bio2, Bio3, Bio4, Bio5, Bio7, Bio10, Bio12, Bio13, Bio14 และ Bio17 จะอยู่ในรูปแบบ Smooth Functions โดยวิธีการ Cubic Regression Splines โดยกำหนดจุด Knot เท่ากับ



ภาพที่ 4.14 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs

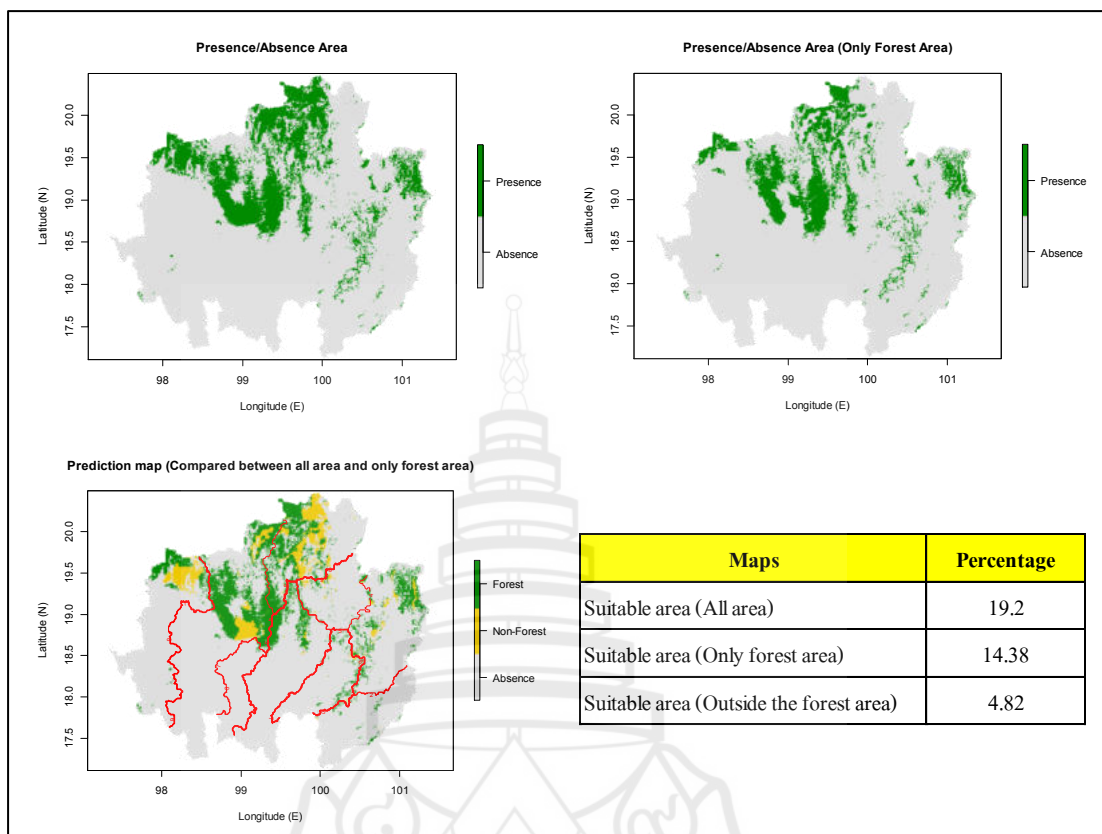
จากภาพที่ 4.14 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง GAMs ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 4.21 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองแบบจำลอง GAMs

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	37
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	369
AUC	0.94675
max TPR+TNR	-3.04164
Threshold	0.33669

จากตารางที่ 4.21 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GAMs พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 37 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด 369 จุด มีค่า AUC เท่ากับ 0.94675 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 94.68 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.33669 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.15 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.15 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงแล้ว ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 19 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 14 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 5 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอน เชียงใหม่และเชียงราย ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่จะกระจุกอยู่ทางทิศตะวันตกเฉียงเหนือของ

พื้นที่ ซึ่งอยู่แถว ๆ ป่าในเขตจังหวัดเชียงใหม่ตอนบนและบริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดเชียงใหม่และลำปาง บริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดเชียงใหม่และเชียงราย โดยมีบางส่วนกระจายอยู่บริเวณป่าในเขตจังหวัดเชียงรายตอนบน นอกจากนี้ยังพบว่าการกระจายเพียงเล็กน้อยอยู่ทางป่าด้านทิศตะวันออกของพื้นที่ในเขตจังหวัดน่าน

4.2.2.3 GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

จากแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน พบว่ามีตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 คือ ทิศด้านลาด (Aspect), ระยะห่างจากแม่น้ำ (Distance), อุณหภูมิเฉลี่ยรายปี (Bio1), ช่วงความแตกต่างของอุณหภูมิรายวัน (Bio2), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) ซึ่งในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs จะทำการสร้างฟังก์ชันการเชื่อมโยง โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 มาสร้าง Smooth Functions ก่อนนำไปสร้างแบบจำลอง จากนั้นตรวจสอบค่า AIC โดยทำการตัด Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองเพื่อตรวจสอบค่า AIC อีกครั้ง หากพบว่าค่า AIC มีค่าลดลงแสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นสามารถตัดออกจากแบบจำลองได้ แต่ถ้าหากตัด Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองแล้วพบว่าค่า AIC ไม่ได้ลดลง แสดงว่า Smooth Functions ที่ไม่มีผลนัยสำคัญทางสถิตินั้นยังมีอิทธิพลต่อแบบจำลองการกระจายของสิ่งมีชีวิต จึงไม่สามารถตัดออกจากแบบจำลองได้

ตารางที่ 4.22 ค่าสถิติทดสอบของแบบจำลองในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี
GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

Smooth Functions	Chi-Square Test			
	GAMs		Improved GAMs	
s(Aspect)	25.02	***	24.00	***
s(Distance)	22.26	***	23.29	***
s(Bio1)	33.92	***	33.24	***
s(Bio6)	52.42	***	41.69	***
s(Bio8)	1.26			
s(Bio12)	2.18			
s(Bio13)	4.90	*	6.85	***
s(Bio14)	29.78	***	32.77	***
s(Bio16)	13.83	**	26.47	***
s(Bio17)	15.19	**	16.87	***
Linear Functions	Z Test			
	GAMs		Improved GAMs	
Bio8			-1.55	
AIC	377.02		372.88	
R ²	67.60		67.50	

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 4.22 แสดงค่าสถิติในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธีการ GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 จากแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษามาสร้างแบบจำลอง GAMs พบว่า ค่า AIC มีค่าเท่ากับ 377.0156 โดยมีค่าลดลงจากวิธีการ GLMs ค่า R^2 มีค่าเท่ากับ 67.6 โดยมีค่าเพิ่มขึ้นจากวิธีการ GLMs แต่จะสังเกตได้ว่า Smooth Functions ของ Bio8 และ Bio12 ไม่มีผลนัยสำคัญทางสถิติ จึงทำการปรับปรุงแบบจำลองใน 2 แนวทาง โดยแนวทางแรก จะทำการตัด Smooth Functions ของตัวแปรที่ไม่มีผลนัยสำคัญทางสถิติออกจากแบบจำลองทีละตัว จากนั้นตรวจสอบค่า AIC ส่วนในแนวทางที่สอง จะเปลี่ยน Smooth Functions ของตัวแปรที่ไม่มีผลนัยสำคัญทางสถิติเป็นฟังก์ชันเชิงเส้น (Linear Function) ของตัวแปรนั้นทีละตัว จากนั้นตรวจสอบค่า AIC จากแบบจำลอง GAMs ที่ได้พบว่า เมื่อตัด Smooth Functions ของ Bio12 ออกจากแบบจำลองและเปลี่ยน Smooth Functions ของ Bio8 เป็นฟังก์ชันเชิงเส้น จะมีค่า AIC ต่ำสุด แสดงว่าตัวแปร Bio12 ไม่มีผลต่อแบบจำลองและตัวแปร Bio8 มีผลต่อแบบจำลองในลักษณะเชิงเส้น เมื่อพิจารณาค่า R^2 พบว่ามีค่าลดลงเล็กน้อยไม่แตกต่างจากแบบจำลองแรกมากนัก ดังนั้นจึงเลือกแบบจำลอง GAMs ที่มีการปรับปรุงแล้วไปวิเคราะห์ในขั้นตอนต่อไป

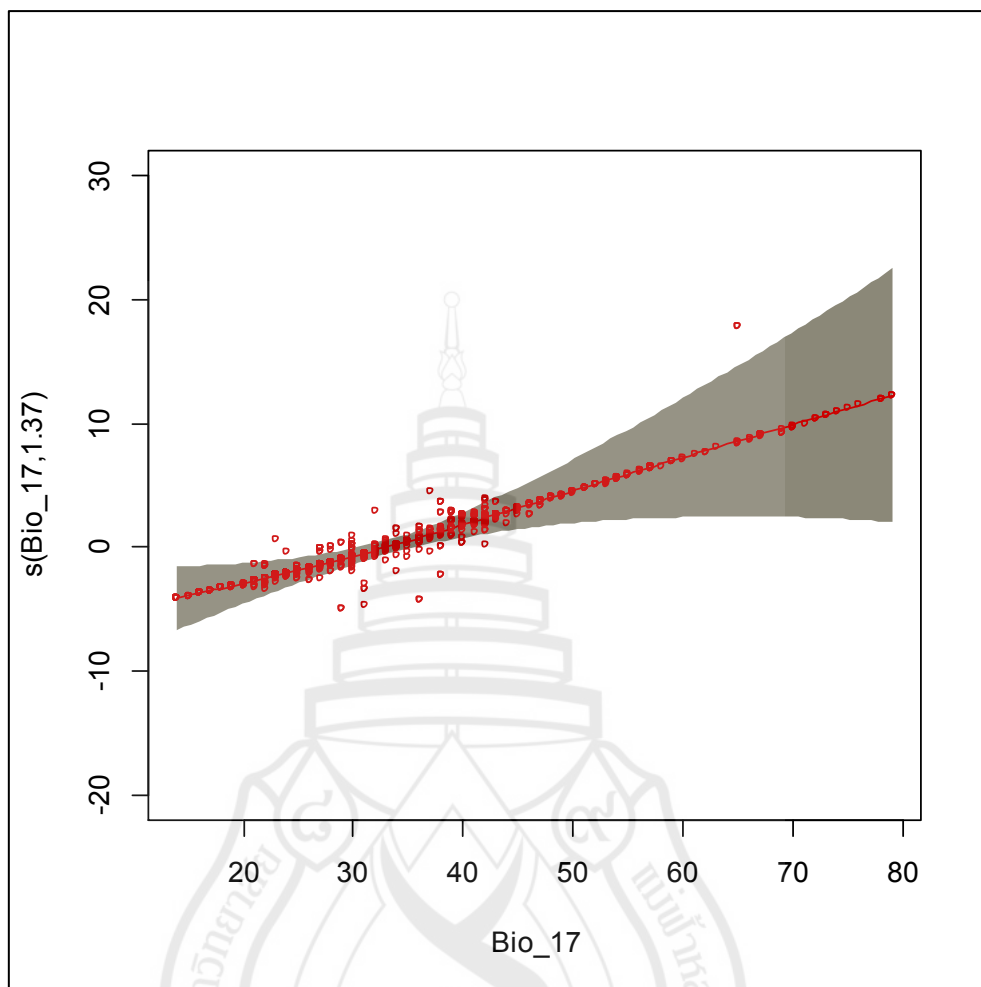
ในขั้นตอนถัดไป จะพิจารณาค่าองศาความเป็นอิสระ (Degree of Freedom) ของ Smooth Functions ถ้าหากค่าองศาความเป็นอิสระของ Smooth Function มีค่าเท่ากับ 1 แสดงว่าปัจจัยนั้นจะส่งผลกระทบต่อแบบจำลองในลักษณะฟังก์ชันเชิงเส้น ดังนั้นจึงต้องเปลี่ยน Smooth Functions ให้อยู่ในรูปแบบของฟังก์ชันเชิงเส้น (Linear Functions) ดังแสดงในตารางที่ 4.23

ตารางที่ 4.23 ค่าสถิติทดสอบ Chi-Square และค่าองศาความเป็นอิสระของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน

Smooth Functions	Degree of Freedom	Chi-Square Test
s(Aspect)	3.54	24.00 ***
s(Distance)	3.23	23.29 ***
s(Bio1)	3.80	33.24 ***
s(Bio6)	3.84	41.69 ***
s(Bio13)	2.85	6.85
s(Bio14)	3.58	32.77 ***
s(Bio16)	3.88	26.47 ***
s(Bio17)	1.37	16.87 ***
AIC	372.88	
R²	67.50	

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

จากตารางที่ 4.23 แสดงค่าสถิติทดสอบ Chi-Square และค่าองศาความเป็นอิสระ (Degree of Freedom) ของ Smooth Functions ในแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธีการ GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence โดยนำตัวแปรที่มีนัยสำคัญทางสถิติ 0.05, 0.01 และ 0.001 ของแบบจำลอง GLMs ที่เกิดจากการเลือกจุด Pseudo-Absence แบบ 2 ขั้นตอน มาสร้างแบบจำลอง GAMs พบว่าค่า AIC มีค่าเท่ากับ 372.88 โดยมีค่าลดลงจากวิธีการ GLMs มีค่า R² มีค่าเท่ากับร้อยละ 67.5 โดยมีค่าเพิ่มขึ้นจากวิธีการ GLMs และจะสังเกตได้ว่าทุกตัวแปรต่างก็มีนัยสำคัญทางสถิติ ยกเว้น Smooth Function ของ Bio13 แต่ก็ไม่สามารถตัดออกจากแบบจำลองได้ เพราะจะทำให้ค่า AIC เพิ่มขึ้น และเมื่อพิจารณาค่าองศาความเป็นอิสระพบว่า ไม่มีตัวแปรไหนเลยที่มีค่า องศาความเป็นอิสระเท่ากับหนึ่ง มีเพียง Smooth Function ของ Bio17 ที่มีค่าองศาความเป็นอิสระใกล้เคียงกับ 1 ดังนั้นในขั้นตอนต่อไปจะพิจารณาว่าเส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Bio17 เป็นลักษณะเชิงเส้นตรงหรือไม่



ภาพที่ 4.16 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Bio17

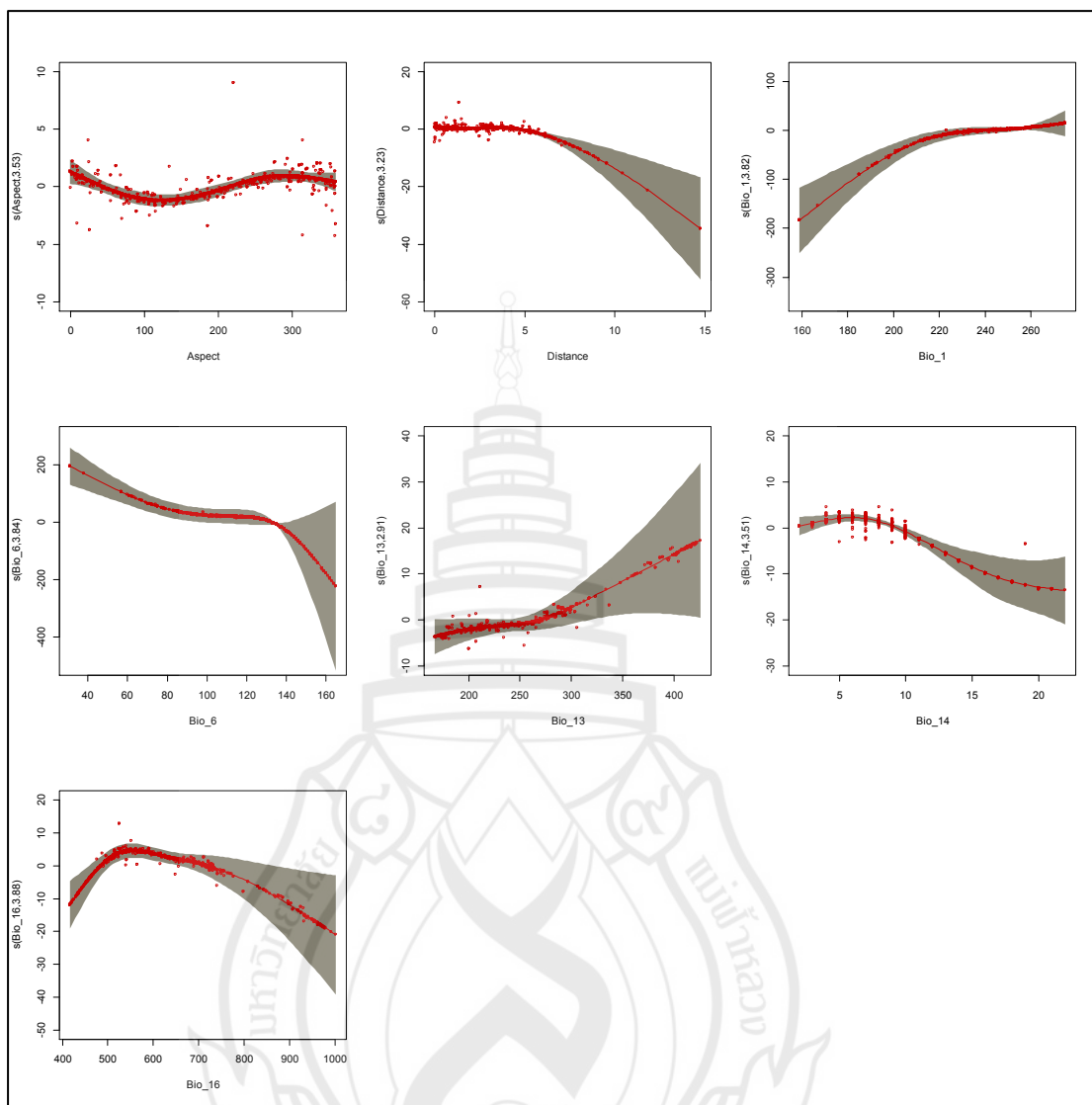
จากภาพที่ 4.16 แสดงเส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Bio17 ในแบบจำลอง GAMs โดยเลือกใช้วิธี Cubic Regression Splines ในการสร้าง Smooth Functions โดยมีค่าองศาความเป็นอิสระเท่ากับ 1.370 โดยแถบที่แรเงาเป็นแถบช่วงความเชื่อมั่นที่ร้อยละ 95 ของเส้นกราฟพยากรณ์ (เส้นสีแดง) โดยจุดสีแดงคือค่าความคลาดเคลื่อนบางส่วน (Partial Residual) จากเส้นกราฟแสดงว่า ตัวแปร Bio17 มีความสัมพันธ์ในรูปแบบเชิงเส้น ดังนั้นจึงเปลี่ยน Smooth Functions ของตัวแปร Bio17 เป็นฟังก์ชันเชิงเส้น (Linear Functions) แทน

ตารางที่ 4.24 ค่าสถิติทดสอบของแบบจำลองใน GAMs หลังจากเปลี่ยน Smooth Function ของตัวแปร Bio17 เป็นฟังก์ชันเชิงเส้นแล้ว

Smooth Functions		Chi-Square Test	
s(Aspect)		24.051	***
s(Distance)		23.293	***
s(Bio1)		32.408	***
s(Bio6)		41.219	***
s(Bio13)		7.057	
s(Bio14)		36.891	***
s(Bio16)		31.210	***
Linear Functions		Z Test	
Bio8		-1.51	
Bio17		3.693	***
AIC		372.93	
R ²		67.4	

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

จากตารางที่ 4.24 แสดงค่าสถิติทดสอบของแบบจำลองใน GAMs หลังจากเปลี่ยน Smooth Function ของตัวแปร Bio17 เป็นฟังก์ชันเชิงเส้นแทน พบว่าตัวแปร Bio17 มีผลต่อแบบจำลองในลักษณะเชิงเส้นอย่างมีนัยสำคัญทางสถิติที่ 0.001 โดยมีค่า AIC เพิ่มขึ้นเล็กน้อยและค่า R² ลดลงเพียงเล็กน้อย



ภาพที่ 4.17 เส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

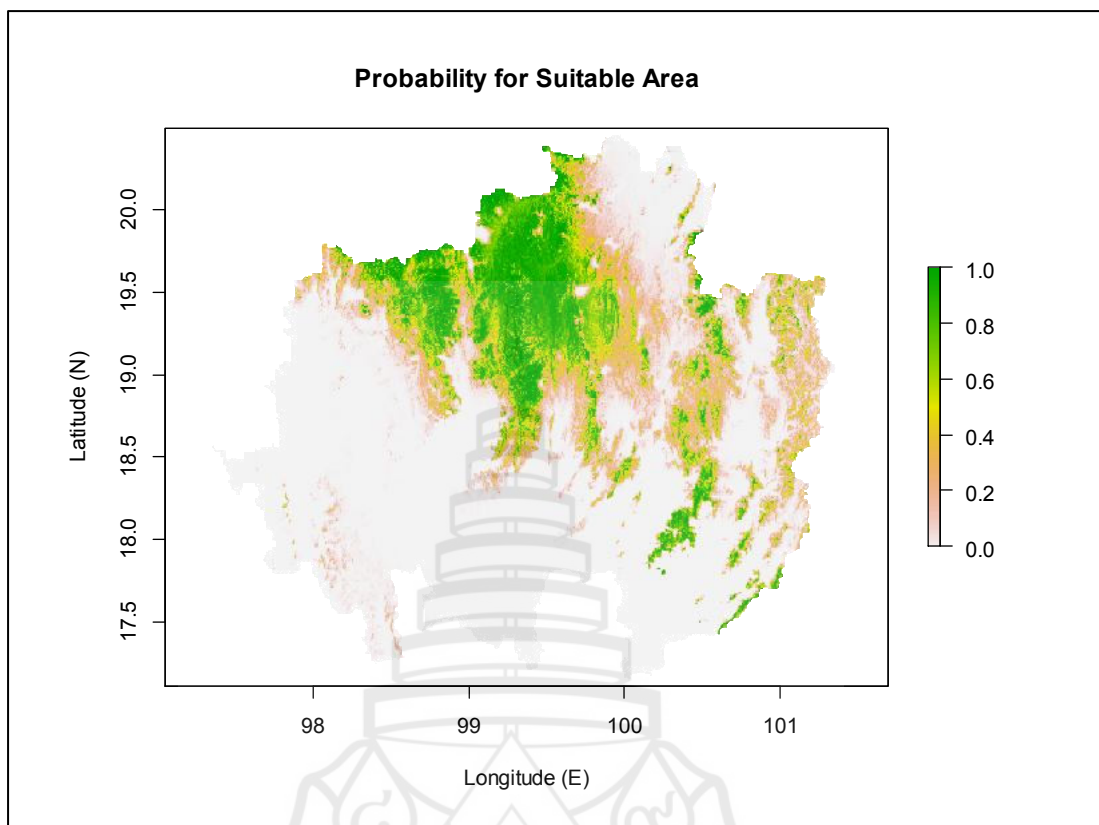
จากภาพที่ 4.17 แสดงเส้นกราฟพยากรณ์ที่เกิดจาก Smooth Functions ของตัวแปร Aspect, Distance, Bio1, Bio6, Bio13, Bio14 และ Bio16 ในแบบจำลอง GAMs โดยเลือกใช้วิธี Cubic Regression Splines ในการสร้าง Smooth Functions ซึ่งกำหนดให้จุด Knot เท่ากับ 5 โดยแกน x คือค่าของตัวแปรและแกน y คือค่าของการพยากรณ์ที่เกิดจาก Smooth Functions ซึ่งค่าแกน y จะอยู่ในรูปแบบของ Additive Component (บทที่ 2 หน้า 33) ซึ่งอยู่ในรูปแบบเชิงเส้น ดังนั้นค่ากลาง (Center Value) จะมีค่าเท่ากับ 0 โดยแถบที่แรเงาเป็นแถบช่วงความเชื่อมั่นที่ร้อยละ 95 (95% Confidence Interval) ของเส้นกราฟพยากรณ์ (เส้นสีแดง) โดยจุดสีแดงคือค่าความคลาดเคลื่อน

บางส่วน (Partial Residual) ที่เกิดจากการประมาณ ซึ่งค่าความคลาดเคลื่อนส่วนใหญ่จะกระจายอยู่ภายในแถบช่วงความเชื่อมั่น จากเส้นกราฟแสดงว่า ตัวแปร Aspect, Distance, Bio1, Bio6, Bio13, Bio14 และ Bio16 มีความสัมพันธ์ในรูปแบบที่ไม่เป็นเชิงเส้น (Non-Linear) ดังนั้นการสร้าง Smooth Functions ของตัวแปรแต่ละตัวเพื่อใช้ในแบบจำลอง GAMs จึงมีความเหมาะสม

ตารางที่ 4.25 ค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงค่าสุดท้ายแล้ว

Variables	Coefficients	Variables	Coefficients	Variables	Coefficients
s(Aspect)	-1.690	s(Bio6)	0.843	s(Bio16)	11.525
	-1.132		-106.533		7.581
	0.220		-162.749		4.785
	0.174		-228.069		-19.817
s(Distance)	0.153	s(Bio13)	1.029	Bio8	-0.149
	0.326		0.743	Bio17	0.238
	0.044		5.091		
	-34.535		17.612		
s(Bio1)	7.842	s(Bio14)	1.731		
	53.541		-3.982		
	170.227		-10.880		
	50.873		-13.622		
intercept = 7.129					

จากตารางที่ 4.25 แสดงค่า Coefficients ของแต่ละตัวแปรในแบบจำลอง GAMs ที่มีการปรับปรุงค่าสุดท้ายแล้ว โดยตัวแปร Bio8 และ Bio17 อยู่ในรูปแบบฟังก์ชันเชิงเส้น ส่วนตัวแปร Aspect, Distance, Bio1, Bio6, Bio13, Bio14 และ Bio16 จะอยู่ในรูปแบบ Smooth Functions โดยวิธีการ Cubic Regression Splines โดยกำหนดจุด Knot เท่ากับ 5



ภาพที่ 4.18 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

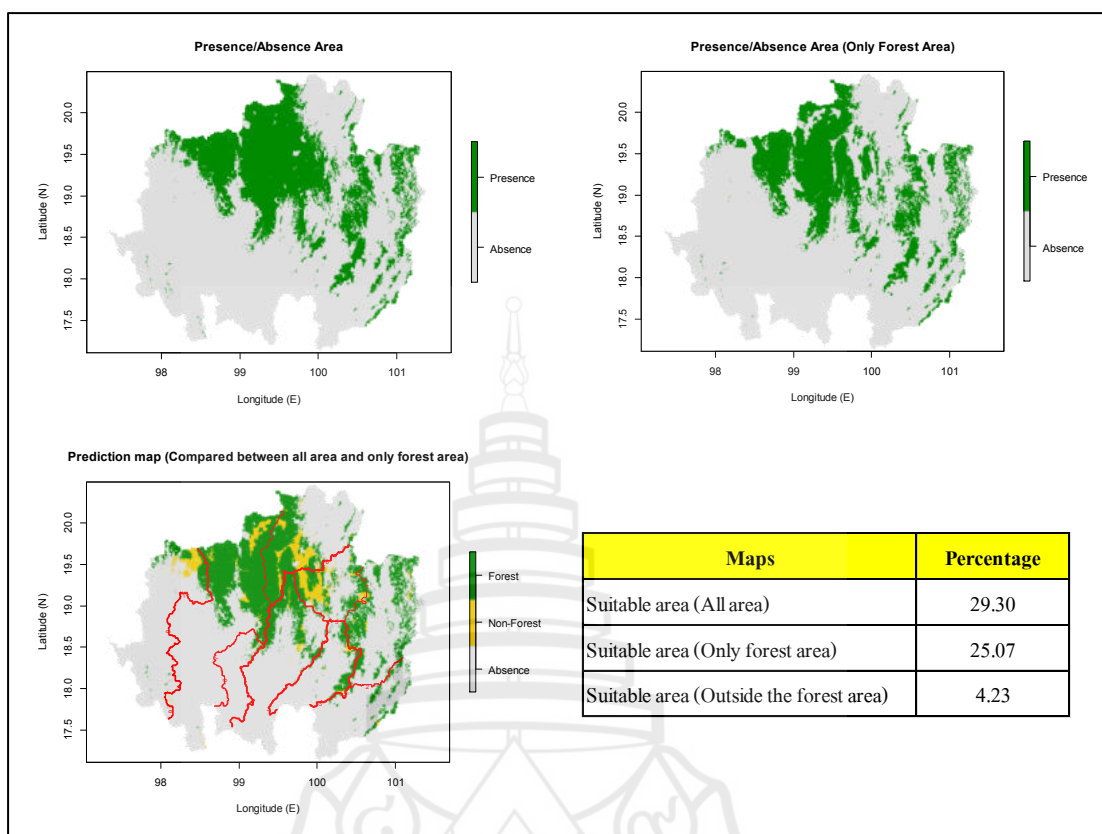
จากภาพที่ 4.18 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 4.26 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	366
AUC	0.97670
max TPR+TNR	-1.40610
Threshold	0.19685

จากตารางที่ 4.26 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด 366 จุด มีค่า AUC เท่ากับ 0.97670 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 97.67 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.19685 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.19 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้วเปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

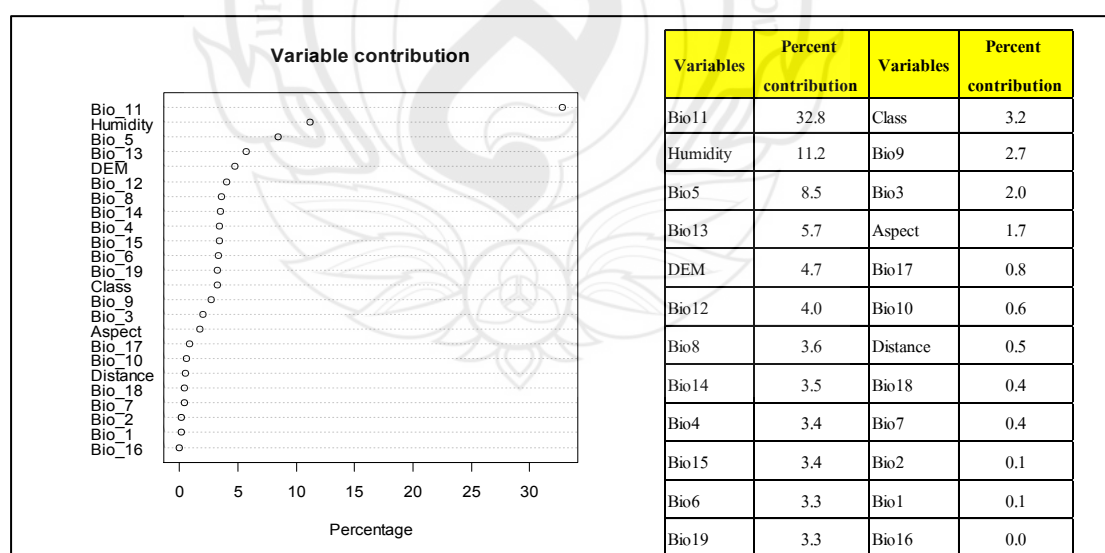
จากภาพที่ 4.19 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs ที่มีการปรับปรุงล่าสุดแล้ว ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 29 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 25 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 4 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในเขตจังหวัดแม่ฮ่องสอน เชียงรายและพะเยา ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมี

การค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่กระจายอยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของพื้นที่ ซึ่งอยู่แถว ๆ ป่าในเขตจังหวัดเชียงใหม่ตอนบน บริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดเชียงใหม่และแม่ฮ่องสอน บริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดเชียงใหม่ เชียงรายและลำปาง นอกจากนี้ยังพบว่าการกระจายบริเวณพื้นที่ป่ารอยต่อระหว่างจังหวัดพะเยาและจังหวัดน่าน รวมถึงบริเวณป่าชายแดนทิศตะวันออกของพื้นที่ในเขตจังหวัดน่านด้วย

4.2.3 แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี MAXENT

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี MAXENT เป็นวิธีการสร้างแบบจำลองที่อยู่บนพื้นฐานของ Machine-Learning โดยใช้ข้อมูลค้นพบ (Presence Data) ร่วมกับข้อมูลลักษณะสิ่งแวดล้อมในพื้นที่ (Background Data) ในการสร้างแบบจำลอง ซึ่งแตกต่างจากวิธี GLMs และ GAMs

ในการสร้างแบบจำลอง MAXENT ในครั้งนี้ ใช้ข้อมูล Presence Points 93 จุดสำหรับสร้าง Training data และทำการสุ่ม 10,021 จุดสำหรับ Background Points และ Presence Points ในการกำหนดการกระจายแบบ MAXENT โดยมีปัจจัยสภาพแวดล้อมประกอบไปด้วยตัวแปรเชิงปริมาณคือตัวแปร Aspect, DEM, Distance, Humidity, Bio1, Bio2, Bio3, Bio4, Bio5, Bio6, Bio7, Bio8, Bio9, Bio10, Bio11, Bio12, Bio13, Bio14, Bio15, Bio16, Bio17, Bio18, Bio19 และตัวแปรเชิงกลุ่มคือตัวแปร Class



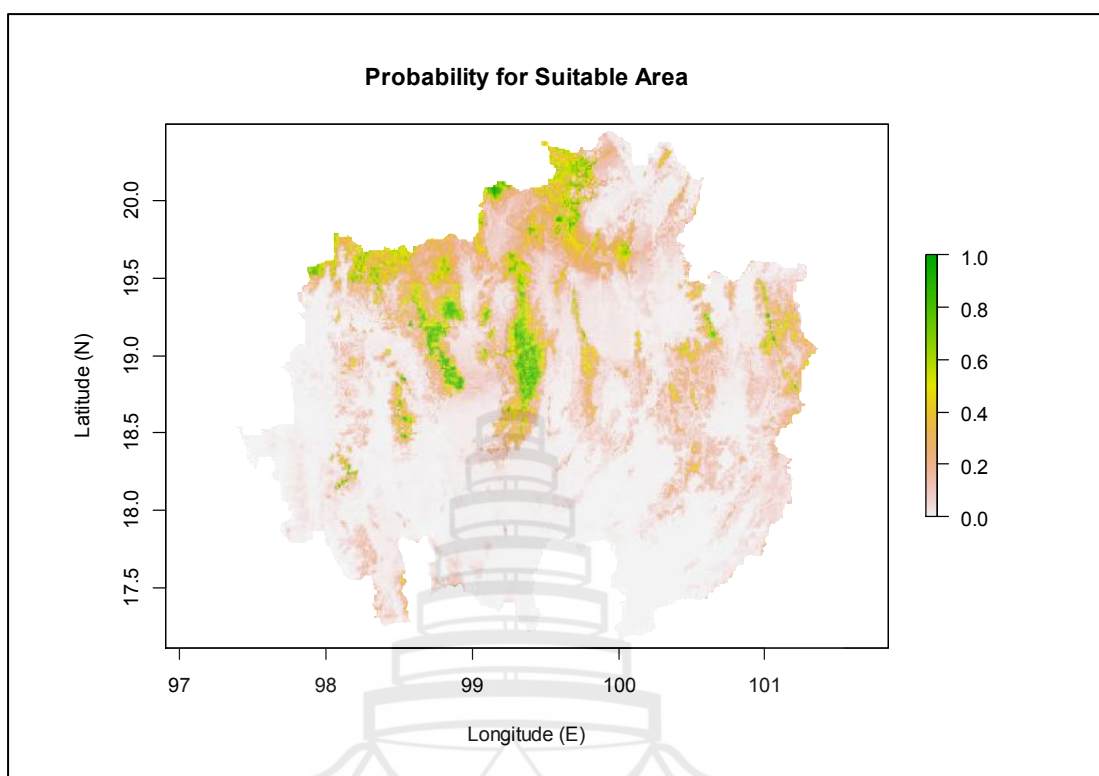
ภาพที่ 4.20 ค่า Variable Contribution ของทุกตัวแปร

จากภาพที่ 4.20 แสดง Variable Contribution ของทุกตัวแปรพบว่าตัวแปร Bio11 มีค่า Variable Contribution มากที่สุดถึงร้อยละ 32.8 แสดงว่าอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11) เป็นปัจจัยหลักที่สำคัญต่อการกระจายของชาป่า ถัดมาได้แก่ ความชื้นสัมพัทธ์ (Humidity), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13) ตามลำดับ ซึ่งทั้ง 3 ตัวแปรนี้มีค่า Variable Contribution ไม่ต่ำกว่าร้อยละ 5 ส่วนตัวแปรที่เหลือต่างก็มีค่า Variable Contribution ต่ำกว่าร้อยละ 5

ตารางที่ 4.27 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง MAXENT

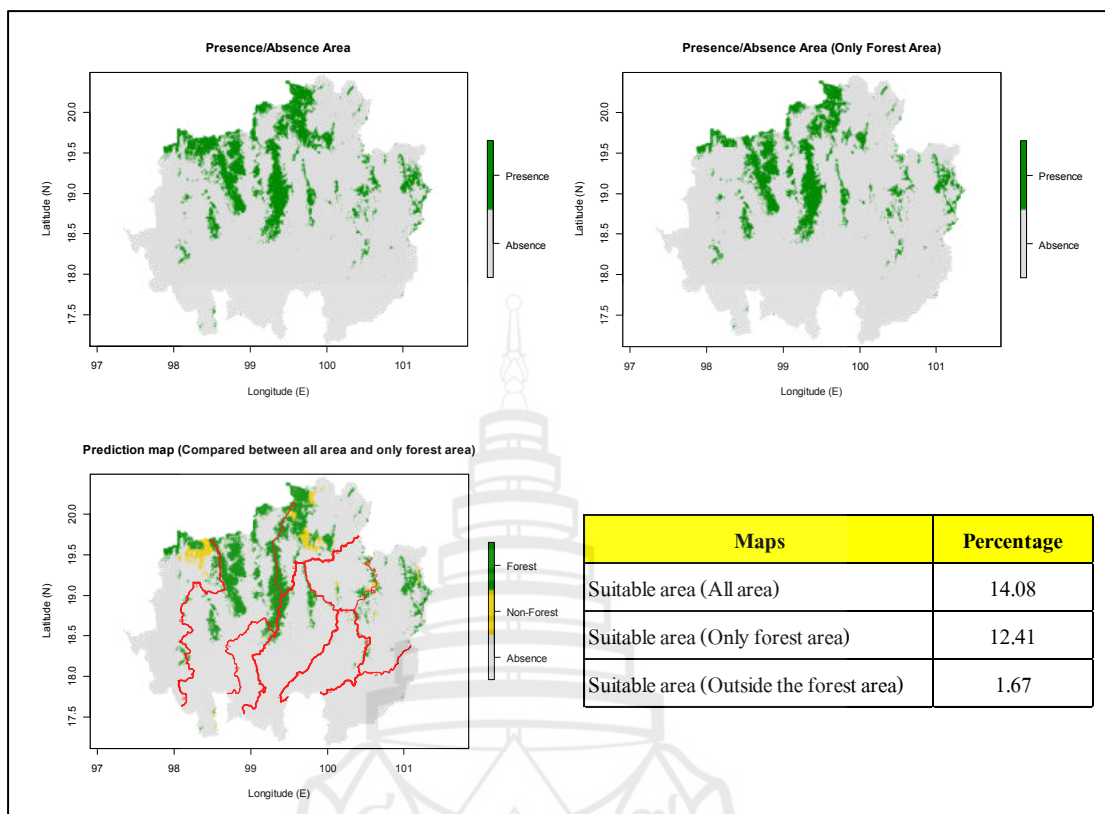
การตรวจสอบความถูกต้อง (Model evaluation)	Value
จำนวนจุด Training Samples	93
จำนวนจุด Background	10,021
AUC	0.95980
Threshold by max TPR+TNR	0.24300

จากตารางที่ 4.27 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง MAXENT พบว่าจำนวนใช้จุด Training samples ทั้งหมด 93 จุด จำนวนจุด Background points ใช้ทั้งหมด 10,021 จุด มีค่า AUC เท่ากับ 0.9598 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 95.98 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งใช้เป็นค่า Threshold เท่ากับ 0.243 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.21 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง MAXENT

จากภาพที่ 4.21 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง MAXENT ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่ หลังจากนั้นใช้ค่า Threshold สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence



ภาพที่ 4.22 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.22 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 14 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นเขตพื้นที่ป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 12 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 2 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอนและเชียงราย ดังแสดงในภาพล่างซ้าย (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า ส่วนใหญ่กระจุกอยู่ทางทิศตะวันตกเฉียงเหนือของพื้นที่ ซึ่งอยู่แถว ๆ ป่าในเขตจังหวัด

เชียงราย เชียงใหม่ และมีบางส่วนกระจายอยู่บริเวณป่าทางด้านทิศตะวันออกของพื้นที่ ซึ่งอยู่ในเขต จังหวัดน่าน

เนื่องจากตัวแปรที่นำมาศึกษามีจำนวนมาก เพื่อเป็นการลดตัวแปรที่มีความสัมพันธ์ลง จึงทำการรวมกลุ่มตัวแปรที่คาดว่าจะมีความสัมพันธ์กันให้อยู่ในกลุ่มเดียวกันโดยพิจารณาความสัมพันธ์ของแต่ละตัวแปร จากนั้นสร้างปัจจัยใหม่ขึ้นมา โดยใช้เทคนิควิธีการวิเคราะห์องค์ประกอบ (Factors Analysis) ซึ่งปัจจัยใหม่ที่ได้จะถูกนำไปสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี MAXENT ในขั้นตอนต่อไป

การวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรหลายตัว (Multivariate Statistical Technique) เพื่อตรวจสอบความสัมพันธ์ภายในระหว่างตัวแปรต่าง ๆ ซึ่งได้ทำการศึกษาความสัมพันธ์ของตัวแปร DEM, Aspect, Distance, Humidity, Bio1, Bio2, Bio3, Bio4, Bio5, Bio6, Bio7, Bio8, Bio9, Bio10, Bio11, Bio12, Bio13, Bio14, Bio15, Bio16, Bio17, Bio18 และ Bio19 ในเขตพื้นที่ภาคเหนือตอนบนทั้งหมด ซึ่งในเบื้องต้นจะทำการตรวจสอบความสัมพันธ์ของแต่ละตัวแปรโดยใช้ Correlation Matrix ในการตรวจสอบว่าตัวแปรที่นำมาใช้มีความสัมพันธ์กันหรือไม่ จากนั้นจะใช้เทคนิควิธี Parallel Analysis ในการกำหนดจำนวนปัจจัยที่เหมาะสม

ตารางที่ 4.28 Correlation Matrix ของแต่ละตัวแปร

	Aspect	DEM	Humidity	Distance	Bio1	Bio2	Bio3
Aspect	1.000	0.008	0.057	-0.066	-0.021	0.020	0.067
DEM	0.008	1.000	0.059	0.134	-0.895	0.070	-0.164
Humidity	0.057	0.059	1.000	-0.047	-0.261	0.243	0.582
Distance	-0.066	0.134	-0.047	1.000	-0.093	0.066	0.000
Bio1	-0.021	-0.895	-0.261	-0.093	1.000	-0.018	-0.063
Bio2	0.020	0.070	0.243	0.066	-0.018	1.000	0.424
Bio3	0.067	-0.164	0.582	0.000	-0.063	0.424	1.000
Bio4	-0.028	-0.043	0.195	-0.044	-0.266	-0.171	0.088
Bio5	-0.028	-0.754	-0.362	-0.053	0.961	0.047	-0.226
Bio6	-0.017	-0.846	-0.281	-0.069	0.986	-0.076	-0.087
Bio7	-0.048	0.233	-0.357	0.053	0.046	0.472	-0.558
Bio8	-0.019	-0.973	-0.120	-0.121	0.938	-0.060	0.087
Bio9	-0.033	-0.611	-0.401	-0.039	0.871	-0.039	-0.344
Bio10	-0.026	-0.893	-0.281	-0.091	0.998	-0.030	-0.096
Bio11	-0.015	-0.808	-0.297	-0.070	0.979	0.022	-0.093
Bio12	0.017	-0.207	0.507	-0.004	-0.085	0.103	0.531
Bio13	0.005	-0.289	0.442	-0.020	-0.009	0.016	0.456
Bio14	-0.010	0.313	0.045	0.082	-0.225	0.635	0.139
Bio15	-0.001	-0.442	0.324	-0.076	0.211	-0.372	0.098
Bio16	0.014	-0.276	0.502	-0.026	-0.052	0.035	0.524
Bio17	-0.013	0.313	0.030	0.118	-0.275	0.495	0.193
Bio18	0.022	-0.250	0.534	-0.043	-0.115	0.015	0.667
Bio19	-0.014	0.210	-0.153	0.123	-0.063	0.326	0.024

ตารางที่ 4.28 (ต่อ)

	Bio4	Bio5	Bio6	Bio7	Bio8	Bio9	Bio10
Aspect	-0.028	-0.028	-0.017	-0.048	-0.019	-0.033	-0.026
DEM	-0.043	-0.754	-0.846	0.233	-0.973	-0.611	-0.893
Humidity	0.195	-0.362	-0.281	-0.357	-0.120	-0.401	-0.281
Distance	-0.044	-0.053	-0.069	0.053	-0.121	-0.039	-0.091
Bio1	-0.266	0.961	0.986	0.046	0.938	0.871	0.998
Bio2	-0.171	0.047	-0.076	0.472	-0.060	-0.039	-0.030
Bio3	0.088	-0.226	-0.087	-0.558	0.087	-0.344	-0.096
Bio4	1.000	-0.399	-0.352	-0.240	0.030	-0.536	-0.224
Bio5	-0.399	1.000	0.967	0.274	0.821	0.945	0.962
Bio6	-0.352	0.967	1.000	0.020	0.892	0.899	0.982
Bio7	-0.240	0.274	0.020	1.000	-0.145	0.312	0.069
Bio8	0.030	0.821	0.892	-0.145	1.000	0.682	0.941
Bio9	-0.536	0.945	0.899	0.312	0.682	1.000	0.868
Bio10	-0.224	0.962	0.982	0.069	0.941	0.868	1.000
Bio11	-0.451	0.978	0.987	0.114	0.855	0.923	0.970
Bio12	0.553	-0.231	-0.125	-0.438	0.160	-0.344	-0.080
Bio13	0.706	-0.167	-0.058	-0.436	0.254	-0.330	0.007
Bio14	0.021	-0.115	-0.233	0.430	-0.289	-0.172	-0.213
Bio15	0.610	0.061	0.173	-0.417	0.433	-0.060	0.231
Bio16	0.658	-0.226	-0.105	-0.492	0.225	-0.366	-0.043
Bio17	0.136	-0.196	-0.269	0.245	-0.292	-0.226	-0.261
Bio18	0.610	-0.332	-0.173	-0.654	0.173	-0.464	-0.120
Bio19	-0.115	0.046	-0.021	0.259	-0.173	0.039	-0.054

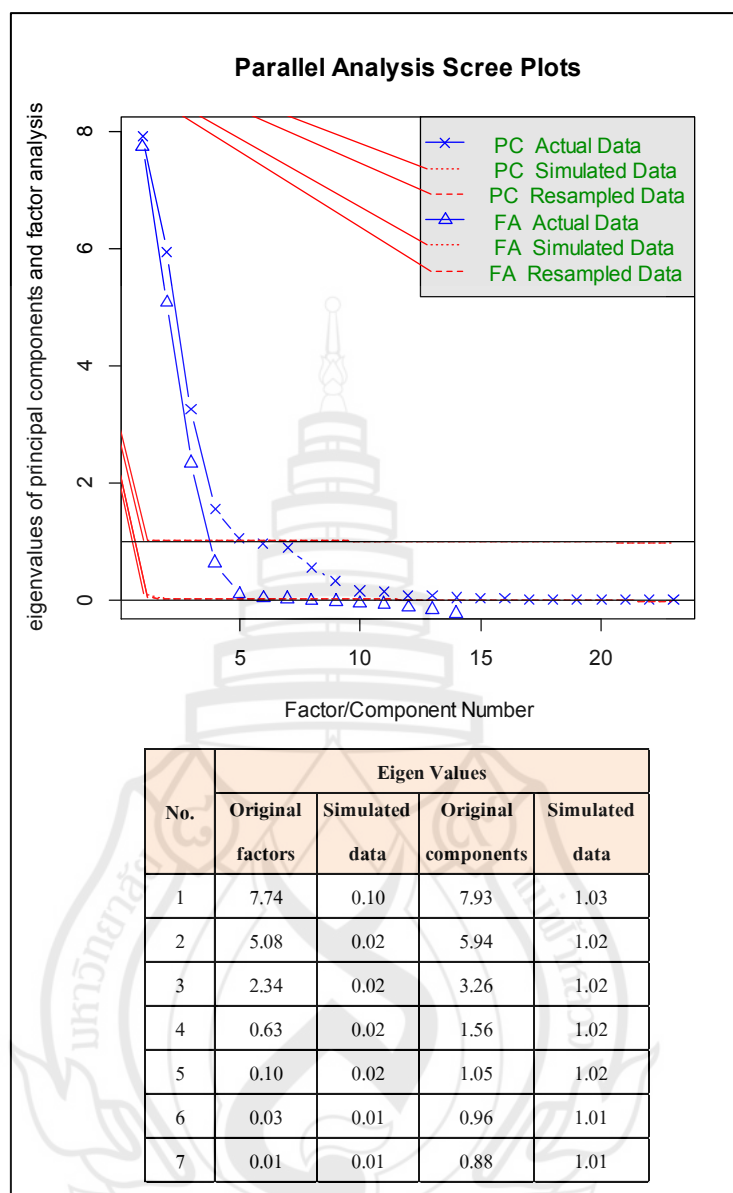
ตารางที่ 4.28 (ต่อ)

	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16
Aspect	-0.015	0.017	0.005	-0.010	-0.001	0.014
DEM	-0.808	-0.207	-0.289	0.313	-0.442	-0.276
Humidity	-0.297	0.507	0.442	0.045	0.324	0.502
Distance	-0.070	-0.004	-0.020	0.082	-0.076	-0.026
Bio1	0.979	-0.085	-0.009	-0.225	0.211	-0.052
Bio2	0.022	0.103	-0.016	0.635	-0.372	0.035
Bio3	-0.093	0.531	0.456	0.139	0.098	0.524
Bio4	-0.451	0.553	0.706	0.021	0.610	0.658
Bio5	0.978	-0.231	-0.167	-0.115	0.061	-0.226
Bio6	0.987	-0.125	-0.058	-0.233	0.173	-0.105
Bio7	0.114	-0.438	-0.436	0.430	-0.417	-0.492
Bio8	0.855	0.160	0.254	-0.289	0.433	0.225
Bio9	0.923	-0.344	-0.330	-0.172	-0.060	-0.366
Bio10	0.970	-0.080	0.007	-0.213	0.231	-0.043
Bio11	1.000	-0.202	-0.157	-0.194	0.065	-0.194
Bio12	-0.202	1.000	0.925	0.110	0.613	0.979
Bio13	-0.157	0.925	1.000	0.101	0.763	0.966
Bio14	-0.194	0.110	0.101	1.000	-0.391	0.039
Bio15	0.065	-0.613	0.763	-0.391	1.000	0.713
Bio16	-0.194	0.979	0.966	0.039	0.713	1.000
Bio17	-0.263	0.293	0.266	0.878	-0.302	0.213
Bio18	-0.250	0.811	0.811	-0.136	0.583	0.857
Bio19	-0.011	0.051	0.041	0.751	-0.369	-0.034

ตารางที่ 4.28 (ต่อ)

	Bio17	Bio18	Bio19
Aspect	-0.013	0.022	-0.014
DEM	0.313	-0.250	0.210
Humidity	0.030	0.534	-0.153
Distance	0.118	-0.043	0.123
Bio1	-0.275	-0.115	-0.063
Bio2	0.495	0.015	0.326
Bio3	0.193	0.667	0.024
Bio4	0.136	0.610	-0.115
Bio5	-0.196	-0.332	0.046
Bio6	-0.269	-0.173	-0.021
Bio7	0.245	-0.654	0.259
Bio8	-0.292	0.173	-0.173
Bio9	-0.226	-0.464	0.039
Bio10	-0.261	-0.120	-0.054
Bio11	-0.263	-0.250	-0.011
Bio12	0.293	0.811	0.051
Bio13	0.266	0.811	0.041
Bio14	0.878	-0.136	0.751
Bio15	-0.302	0.583	-0.369
Bio16	0.213	0.857	-0.034
Bio17	1.000	0.038	0.850
Bio18	0.038	1.000	-0.200
Bio19	0.850	-0.200	1.000

จากตารางที่ 4.28 แสดง Correlation Matrix ของแต่ละตัวแปร ซึ่งใช้เพื่อบ่งบอกขนาดความสัมพันธ์ของสองตัวแปรใด ๆ โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) เป็นเกณฑ์ในการวัดความสัมพันธ์ ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเข้าใกล้ 1 หรือ -1 จะแสดงว่าตัวแปรทั้งสองตัวนั้นมีความสัมพันธ์กันในระดับสูง แต่ถ้าหากค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเข้าใกล้ 0 จะแสดงว่าตัวแปรทั้งสองตัวนั้นมีความสัมพันธ์กันในระดับน้อยหรือไม่มีเลย ซึ่งเครื่องหมาย +, - หน้าตัวเลขจะแสดงทิศทางของความสัมพันธ์ ถ้าค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นบวก แสดงว่าตัวแปรทั้งสองตัวมีความสัมพันธ์ไปในทิศทางเดียวกัน แต่ถ้าหากค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเป็นลบ แสดงว่าตัวแปรทั้งสองตัวมีความสัมพันธ์ในทิศทางที่ตรงข้ามกัน โดยขนาดของค่าสัมประสิทธิ์สหสัมพันธ์ที่มีขนาด 0.7-1 จะแสดงว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันในระดับสูงถึงสูงมาก (Hinkle, William & Stephen, 1998) พบว่า Aspect มีค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 0 แสดงว่า Aspect มีความสัมพันธ์กับตัวแปรอื่น ๆ ในระดับน้อยหรือไม่มีเลย ในขณะที่ตัวแปร DEM มีความสัมพันธ์ในระดับสูงถึงสูงมากกับตัวแปร Bio1, Bio5, Bio6, Bio8, Bio10 และ Bio11 ซึ่งเป็นตัวแปรเหล่านี้เป็นตัวแปรที่เกี่ยวข้องกับอุณหภูมิทั้งสิ้น แสดงว่าความสูงจากระดับน้ำทะเลมีความสัมพันธ์กับอุณหภูมิ ส่วนตัวแปร Bio12 ซึ่งเป็นปริมาณน้ำฝนรายปีมีความสัมพันธ์ในระดับสูงถึงสูงมากกับตัวแปร Bio13, Bio16 และ Bio18 ซึ่งทั้ง 3 ตัวแปรนี้ต่างก็เป็นตัวแปรที่เกี่ยวข้องกับปริมาณน้ำฝนทั้งสิ้น แสดงว่าตัวแปรในกลุ่มของปริมาณน้ำฝนต่างก็มีความสัมพันธ์กันเองค่อนข้างสูงเช่นกัน ดังนั้นจึงจำเป็นที่จะต้องมีการจัดกลุ่มปัจจัยที่มีความสัมพันธ์กันให้อยู่ในกลุ่มเดียวกัน โดยใช้เทคนิควิธี Parallel Analysis ในการกำหนดจำนวนกลุ่มปัจจัยที่เหมาะสม



ภาพที่ 4.23 ผลการวิเคราะห์ของ Parallel Analysis โดย Scree Plots

จากภาพที่ 4.23 แสดงผลการวิเคราะห์ของ Parallel Analysis โดยใช้ Scree Plots ในการแสดงผล ซึ่งแสดงค่า Eigen Values ของแต่ละ Factors โดยเรียงลำดับจากมากไปหาน้อย พบว่าจำนวน Component ที่แนะนำคือ 5 ตัว เนื่องจากค่า Eigen Values มีค่ามากกว่า 1 ในขณะที่จำนวน Factors ที่แนะนำคือ 7 ตัว เนื่องจากค่า Eigen Values มีค่ามากกว่า 1 สรุปได้ว่าจำนวน Factors ที่จะสร้างขึ้นใหม่โดยใช้เทคนิควิธี Factor Analysis คือ 7 ตัว

ตารางที่ 4.29 ค่า Loading Factors ของแต่ละตัวแปรในการสร้างปัจจัย 7 ตัว

Variables	WLS1	WLS2	WLS3	WLS4	WLS5	WLS6	WLS7
Aspect	-0.015	-0.002	-0.010	-0.003	0.106	0.000	0.008
DEM	-0.881	-0.293	0.193	0.032	-0.015	0.175	-0.165
Distance	-0.080	-0.018	0.101	0.016	-0.066	-0.014	-0.132
Humidity	-0.231	0.452	-0.084	0.168	0.601	0.056	-0.012
Bio1	0.994	-0.018	-0.080	0.002	-0.046	-0.009	0.044
Bio2	0.024	-0.004	0.398	0.840	0.251	-0.144	-0.070
Bio3	-0.017	0.432	0.113	0.153	0.662	-0.506	-0.069
Bio4	-0.295	0.686	-0.021	-0.056	-0.283	-0.047	0.582
Bio5	0.961	-0.179	0.011	0.077	-0.126	0.132	-0.033
Bio6	0.988	-0.073	-0.049	-0.090	-0.015	0.031	-0.038
Bio7	0.039	-0.431	0.233	0.638	-0.432	0.401	0.009
Bio8	0.924	0.255	-0.161	-0.012	-0.046	-0.011	0.179
Bio9	0.870	-0.316	-0.026	0.007	-0.149	0.196	-0.178
Bio10	0.990	-0.005	-0.068	0.001	-0.090	0.011	0.076
Bio11	0.981	-0.166	-0.047	0.013	0.001	0.018	-0.076
Bio12	-0.047	0.964	0.119	0.028	0.101	-0.052	-0.158
Bio13	0.018	0.968	0.137	-0.051	0.070	0.025	0.129
Bio14	-0.161	-0.008	0.851	0.364	0.027	0.042	0.051
Bio15	0.191	0.763	-0.342	-0.220	0.030	0.330	0.250
Bio16	-0.026	0.988	0.048	-0.012	0.097	-0.064	0.000
Bio17	-0.198	0.169	0.929	0.162	-0.067	-0.055	-0.059
Bio18	-0.107	0.830	-0.130	-0.040	0.241	-0.339	0.087
Bio19	0.011	-0.066	0.916	-0.034	-0.068	0.020	-0.116

จากตารางที่ 4.29 แสดง Loading factors ของแต่ละตัวแปรในการสร้าง Factors 7 ด้วยวิธีการหมุนแกนปัจจัย (Factor rotation) ด้วยวิธี Varimax ซึ่งค่า Loading factors จะใช้ในการพิจารณาว่ามีตัวแปรใดบ้างที่ควรอยู่ในกลุ่ม Factors เดียวกัน โดยพิจารณาค่า Loading factors ของแต่ละตัวแปร ถ้าค่า Loading factors ของตัวแปรใดมีค่ามาก (เข้าสู่ +1 หรือ -1) ก็จะถูกจัดให้อยู่ในกลุ่ม Factors นั้น (ส่วนที่แรงเงา) พบว่าสามารถจัดให้ตัวแปรอยู่ในกลุ่มเดียวกันได้ดังรายละเอียดต่อไปนี้

กลุ่มปัจจัยที่ 1 จะประกอบด้วยตัวแปรความสูงจากระดับน้ำทะเล (DEM), อุณหภูมิเฉลี่ยรายปี (Bio1), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) และตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11) เนื่องจากมีค่า Loading factors มาก (เข้าสู่ +1 หรือ -1) และเมื่อเรียงลำดับค่า Loading Factors แล้วพบว่าตัวแปร Bio1 มีค่า Loading factors มากที่สุด (ไม่คำนึงถึงเครื่องหมาย) รองลงมาได้แก่ Bio10, Bio6, Bio11, Bio5, Bio8, DEM และ Bio9 ตามลำดับ ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยของอุณหภูมิที่มีความสัมพันธ์กับระดับความสูงของน้ำทะเล ใช้สัญลักษณ์เป็น Factor1

กลุ่มปัจจัยที่ 2 จะประกอบด้วยตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ช่วงฤดูกาลของปริมาณน้ำฝน (Bio15), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และตัวแปรปริมาณน้ำฝนในไตรมาสที่ร้อนที่สุด (Bio18) เนื่องจากมีค่า Loading factors มาก (เข้าสู่ +1 หรือ -1) และถึงแม้ว่าค่า Loading Factors ของ Bio4 จะมีค่าเพียง 0.686 แต่เมื่อเปรียบเทียบกับแต่ละกลุ่ม Factors แล้ว พบว่า ในกลุ่ม Factor ที่ 2 ของ Bio4 จะมีค่า Loading factors มากที่สุด และเมื่อเรียงลำดับค่า Loading Factors แล้วพบว่าตัวแปร Bio16 มีค่า Loading factors มากที่สุด (ไม่คำนึงถึงเครื่องหมาย) รองลงมาได้แก่ Bio13, Bio12, Bio18, Bio15 และ Bio4 ตามลำดับ ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยของปริมาณน้ำฝนที่มีความสัมพันธ์กับฤดูกาล (Season) ใช้สัญลักษณ์เป็น Factor2

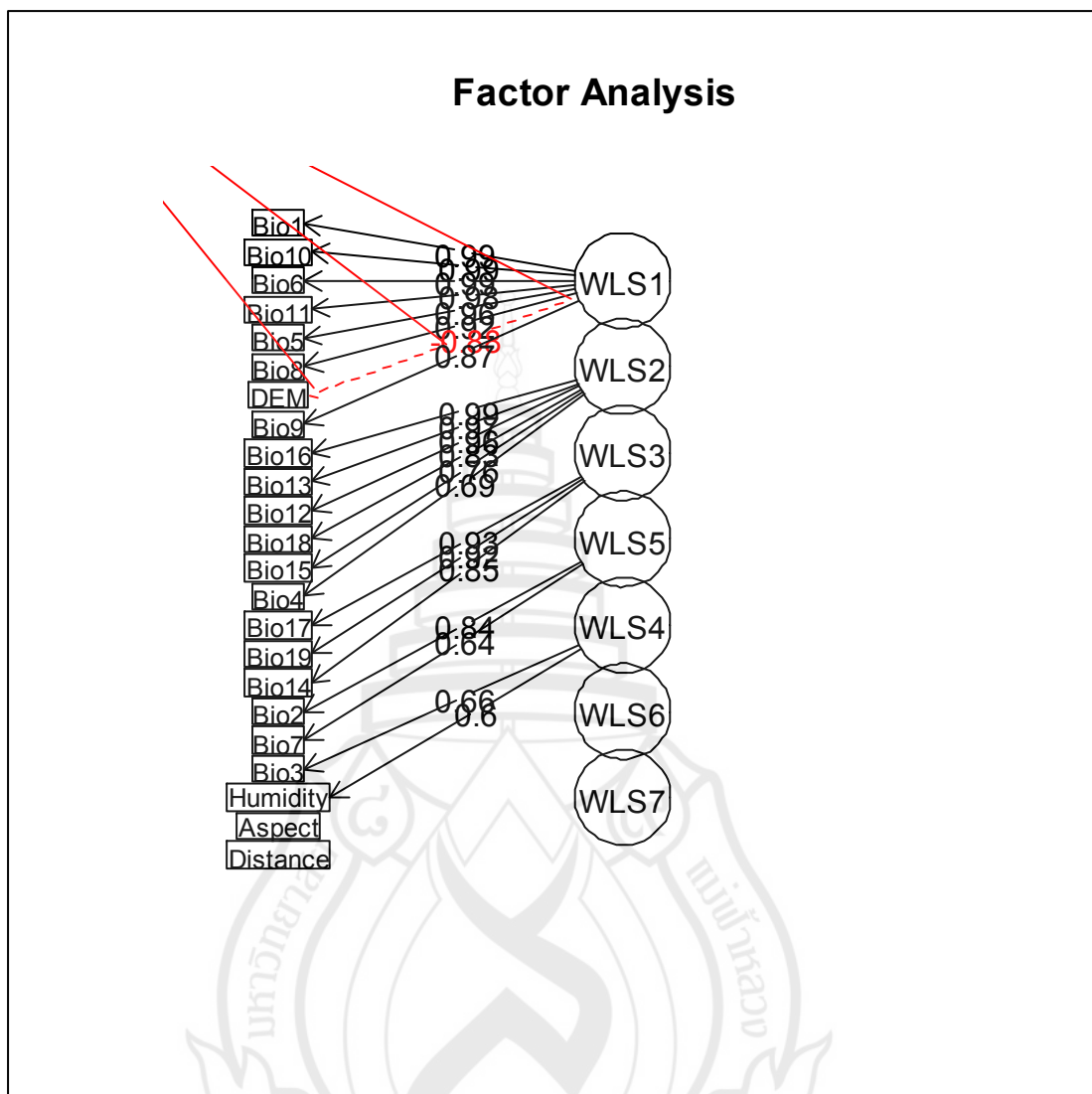
กลุ่มปัจจัยที่ 3 จะประกอบด้วยตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และตัวแปรปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19) เนื่องจากมีค่า Loading factors มาก (เข้าสู่ +1 หรือ -1) และเมื่อเรียงลำดับค่า Loading Factors แล้วพบว่าตัวแปร Bio17 มีค่า Loading factors มากที่สุด (ไม่คำนึงถึงเครื่องหมาย) รองลงมาได้แก่ Bio19 และ Bio14 ตามลำดับ ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยของปริมาณน้ำฝนในช่วงที่แห้งแล้ง ใช้สัญลักษณ์เป็น Factor3

กลุ่มปัจจัยที่ 4 จะประกอบด้วยตัวแปรช่วงอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) และตัวแปรช่วงของอุณหภูมิรายปี (Bio7) เนื่องจากมีค่า Loading factors มาก และเมื่อเรียงลำดับค่า Loading Factors แล้วพบว่าตัวแปร Bio2 มีค่า Loading factors มากที่สุด (ไม่คำนึงถึงเครื่องหมาย) รองลงมาได้แก่ Bio7 ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยความผันแปรของอุณหภูมิรายวันและรายปี ใช้สัญลักษณ์เป็น Factor4

กลุ่มปัจจัยที่ 5 จะประกอบด้วยตัวแปรความชื้นสัมพัทธ์ (Humidity) และตัวแปร Isothermality (Bio3) เนื่องจากมีค่า Loading factors มาก และเมื่อเรียงลำดับค่า Loading Factors แล้วพบว่าตัวแปร Bio3 มีค่า Loading factors มากที่สุด (ไม่คำนึงถึงเครื่องหมาย) รองลงมาได้แก่ Humidity ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยอุณหภูมิที่มีความสัมพันธ์กับความชื้นสัมพัทธ์ ใช้สัญลักษณ์เป็น Factor5

กลุ่มปัจจัยที่ 6 มีเพียงตัวแปรทิศด้านลาด (Aspect) ตัวแปรเดียว ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยทิศด้านลาด ใช้สัญลักษณ์เป็น Factor6

กลุ่มปัจจัยที่ 7 มีเพียงตัวแปรระยะห่างจากแม่น้ำ (Distance) ตัวแปรเดียว ดังแสดงในภาพที่ 4.24 โดยตั้งชื่อกลุ่มปัจจัยนี้ว่า กลุ่มปัจจัยระยะห่างจากแม่น้ำ ใช้สัญลักษณ์เป็น Factor7



ภาพที่ 4.24 ผลการจัดกลุ่ม Factors

จากภาพที่ 4.24 แสดงผลการจัดกลุ่ม Factors โดย WLS1-WLS7 จะแสดงถึงกลุ่มปัจจัย Factor1-Factor7 ซึ่ง WLS ย่อมาจาก “Weight Least Square” ซึ่งเป็นหนึ่งในวิธีการสกัดองค์ประกอบนั่นคือ การค้นหาจำนวนองค์ประกอบที่มีความสามารถเพียงพอในการอธิบายความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้ โดยเส้นสีแดงจะแสดงค่า Loading Factors ที่เป็นค่าลบ

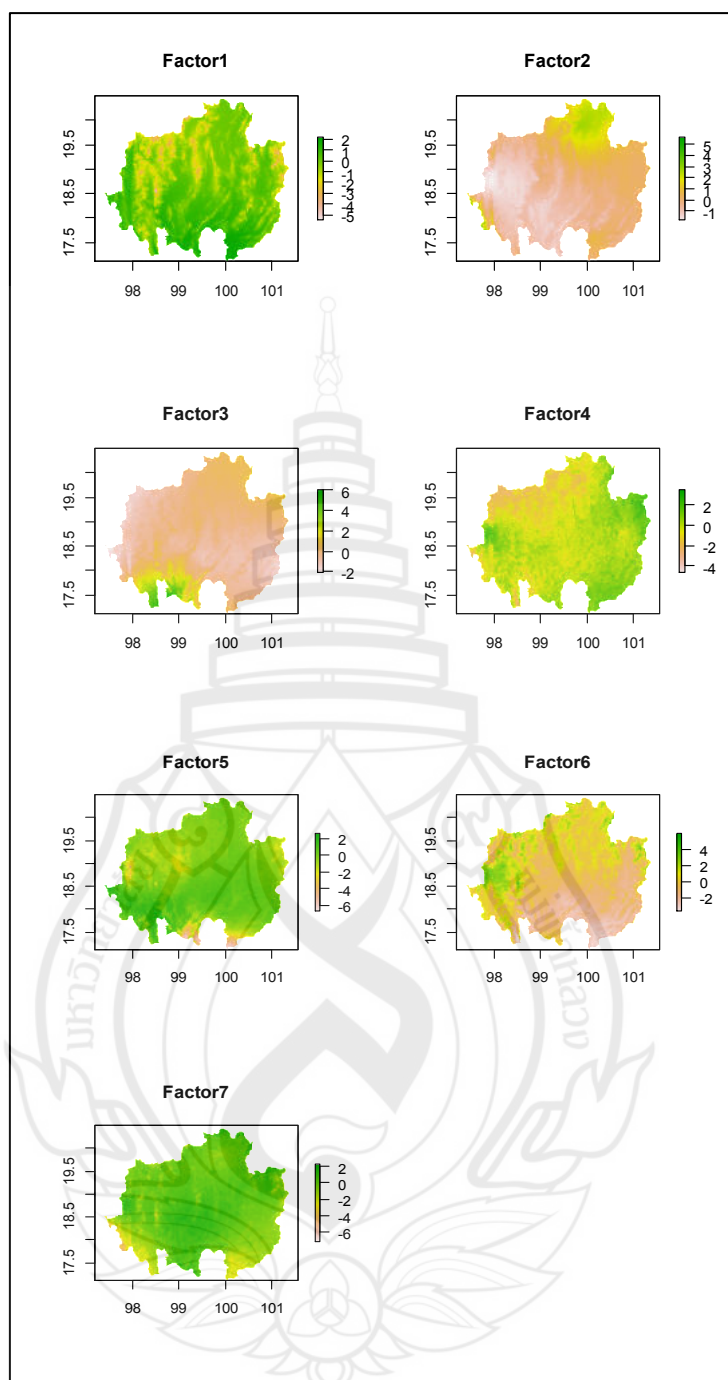
ตารางที่ 4.30 ค่า Weight Score ของแต่ละตัวแปรในการสร้างปัจจัย 7 ตัว

Variables	WLS1	WLS2	WLS3	WLS5	WLS4	WLS6	WLS7
Aspect	0.002	-0.008	0.004	-0.002	0.028	0.017	0.037
DEM	-0.011	0.040	0.020	-0.012	-0.209	-0.037	-0.118
Humidity	0.021	-0.010	0.031	0.060	0.165	0.139	0.154
Distance	-0.001	0.008	-0.009	0.011	-0.079	-0.059	-0.046
Bio1	0.383	-0.289	0.468	-0.177	3.123	1.580	2.300
Bio2	0.012	-0.041	0.076	0.234	0.880	0.000	0.269
Bio3	0.004	0.003	-0.067	0.371	0.168	-0.076	0.099
Bio4	-0.029	0.146	0.103	-0.055	-0.624	-0.578	0.162
Bio5	-0.017	0.431	-0.751	1.872	-1.605	1.060	-0.983
Bio6	0.168	-0.119	0.645	-2.008	1.958	0.196	0.275
Bio7	-0.019	0.070	-0.169	0.643	-0.684	0.126	-0.147
Bio8	0.050	0.080	-0.564	0.580	-0.470	-0.532	0.096
Bio9	-0.017	0.071	-0.174	-0.024	0.205	0.346	-0.159
Bio10	0.182	-0.359	0.533	-0.512	-1.251	-0.939	1.691
Bio11	0.262	0.297	-0.041	0.327	-2.578	-2.133	-3.453
Bio12	0.046	0.085	0.249	-0.318	0.470	0.253	-0.800
Bio13	-0.008	0.252	0.162	-0.015	-0.343	-0.178	0.397
Bio14	0.005	-0.077	0.156	-0.011	0.470	0.382	0.305
Bio15	0.042	-0.099	0.107	-0.293	1.441	1.555	0.554
Bio16	-0.041	0.807	-0.600	0.713	-1.595	-0.568	-0.991
Bio17	0.092	-0.119	0.809	-0.367	0.141	0.063	0.114
Bio18	-0.005	0.036	-0.152	0.125	-0.087	-0.206	-0.171
Bio19	-0.006	0.033	0.142	-0.087	-0.069	0.008	0.014

จากตาราง 4.30 แสดงค่า Weight Score ของแต่ละตัวแปรในการสร้างปัจจัย 7 ตัว ซึ่ง Weight Score ที่ได้จะถูกนำไปคำนวณหา Factor score ซึ่งอยู่ในรูปแบบของผลรวมเชิงเส้น โดยค่าของตัวแปรแต่ละตัวจะต้องถูกนำมา Standardized ก่อนที่นำมาคำนวณหา Factor score ยกตัวอย่างเช่น การหาค่า Factor score ของปัจจัยที่ 1

$$\begin{aligned} \text{Factor1} = & 0.002\text{Aspect} - 0.011\text{DEM} + 0.021\text{Humidity} - 0.001\text{Distance} + 0.383\text{Bio1} \\ & + 0.012\text{Bio2} + 0.004\text{Bio3} - 0.029\text{Bio4} - 0.017\text{Bio5} + 0.168\text{Bio6} - 0.019\text{Bio7} \\ & + 0.050\text{Bio8} - 0.017\text{Bio9} + 0.182\text{Bio10} + 0.262\text{Bio11} + 0.046\text{Bio12} \\ & - 0.008\text{Bio13} + 0.005\text{Bio14} + 0.042\text{Bio15} - 0.041\text{Bio16} + 0.092\text{Bio17} \\ & - 0.005\text{Bio18} - 0.006\text{Bio19} \end{aligned}$$

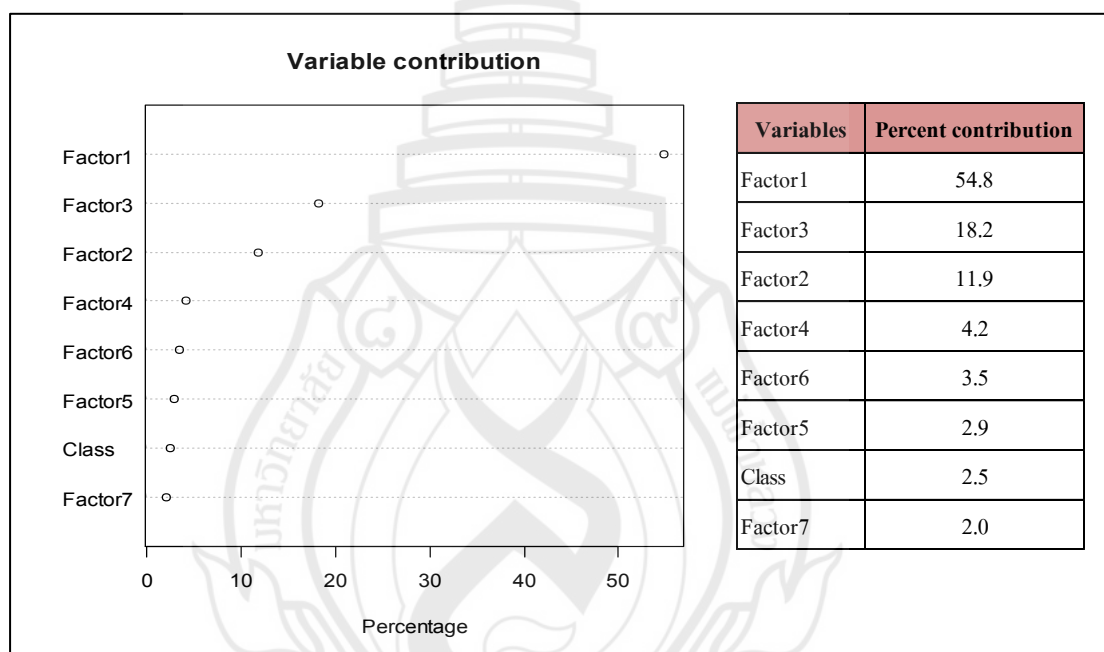
คำนวณหา Factor score จนครบทั้ง 7 ตัว ก็จะได้ factors ใหม่ 7 ตัวที่มาจากตัวแปรทั้งหมด 23 ตัว ดังแสดงในภาพที่ 4.15 ซึ่งแสดง Factors ทั้ง 7 ตัวในเขตพื้นที่ภาคเหนือตอนบนทั้งหมด ซึ่ง Factors ทั้ง 7 ตัวนี้จะถูกนำไปสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี MAXENT ในขั้นตอนต่อไป



ภาพที่ 4.25 Factors ทั้ง 7 ตัวในเขตพื้นที่ภาคเหนือตอนบน

เมื่อได้กลุ่มปัจจัยที่สร้างขึ้นใหม่แล้ว ขั้นตอนต่อไปทำการสร้างแบบจำลอง MAXENT โดยใช้ข้อมูล Presence points ร่วมกับข้อมูลลักษณะสิ่งแวดล้อมในพื้นที่ (Background data) ซึ่งก็คือ Factors ทั้ง 7 ตัวที่ทำการสร้างขึ้นใหม่โดยเทคนิค Factor analysis

ในการสร้างแบบจำลอง MAXENT ในครั้งนี้ ใช้ข้อมูลการค้นพบ (Presence data) 93 จุด สำหรับสร้าง Training data และทำการสุ่ม 10,083 จุดสำหรับ Background points และ Presence points ในการกำหนดการกระจายแบบ MAXENT โดยมีปัจจัยสภาพแวดล้อมประกอบไปด้วย Factors ทั้ง 7 ตัว ได้แก่ Factor1, Factor2, Factor3, Factor4, Factor5, Factor6 และ Factor7 และตัวแปร Class โดยกำหนดให้เป็นตัวแปรเชิงกลุ่ม



ภาพที่ 4.26 ค่า Variable Contribution ของตัวแปร Factors และตัวแปร Class

จากภาพที่ 4.26 แสดงค่า Variable Contribution ของตัวแปร Factors และตัวแปร Class พบว่ากลุ่มปัจจัยของอุณหภูมิที่มีความสัมพันธ์กับระดับความสูงของน้ำทะเล (Factor1) มีค่า Variable Contribution มากที่สุดถึงร้อยละ 54.8 แสดงว่ากลุ่มปัจจัยของอุณหภูมิที่มีความสัมพันธ์กับระดับความสูงของน้ำทะเลเป็นกลุ่มปัจจัยหลักที่สำคัญต่อการกระจายของสิ่งมีชีวิต ซึ่งประกอบไปด้วยตัวแปรความสูงจากระดับน้ำทะเล (DEM), อุณหภูมิเฉลี่ยรายปี (Bio1), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) และตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11)

ถัดมาได้แก่ กลุ่มปัจจัยของปริมาณน้ำฝนในช่วงที่แห้งแล้ง (Factors3) มี Variable Contribution ร้อยละ 18.2 ซึ่งประกอบไปด้วยตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด(Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และตัวแปรปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19) แสดงว่าปริมาณน้ำฝนในช่วงฤดูแล้งมีความสำคัญต่อการกระจายของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน

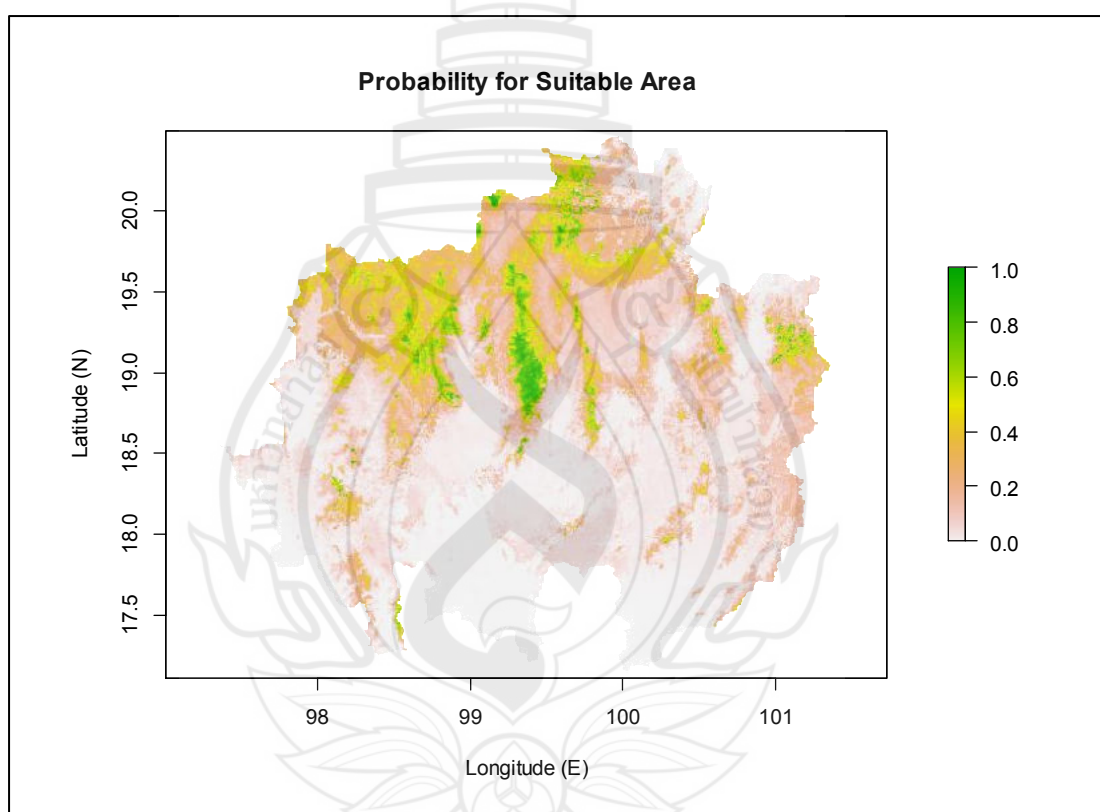
ถัดมาได้แก่ กลุ่มปัจจัยของปริมาณน้ำฝนที่มีความสัมพันธ์กับฤดูกาล (Factors2) มี Variable Contribution ร้อยละ 11.9 ซึ่งประกอบไปด้วยตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ช่วงฤดูกาลของปริมาณน้ำฝน (Bio15), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และตัวแปรปริมาณน้ำฝนในไตรมาสที่ร้อนที่สุด (Bio18) แสดงว่าปริมาณน้ำฝนที่ตกต้องตามฤดูกาลมีความสำคัญต่อการกระจายของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน

ถัดมาได้แก่ Factor4, Factor5, Factor6, Class, และ Factor7 ตามลำดับ ซึ่งมีค่า Variable Contribution ไม่ต่างกันมากนัก โดยมีค่าอยู่ระหว่างร้อยละ 2 ถึง 5

ตารางที่ 4.31 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบ

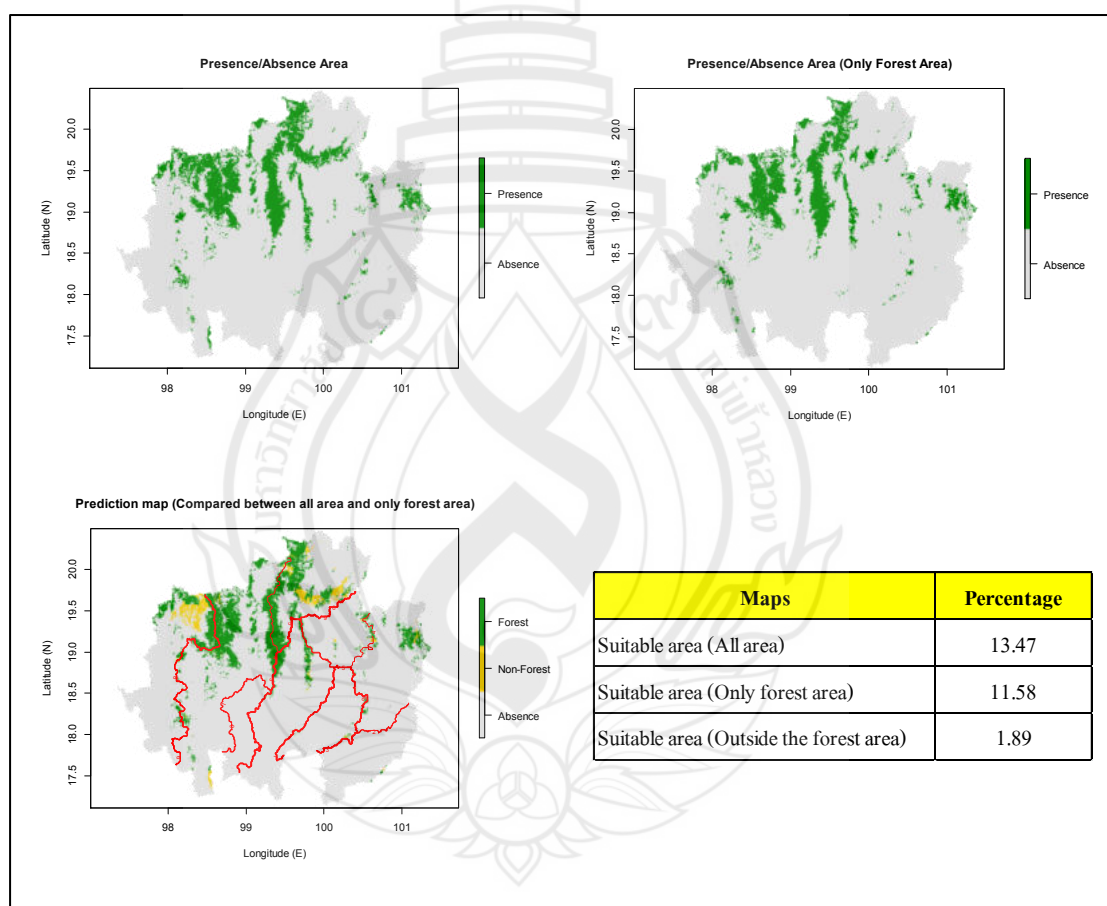
การตรวจสอบความถูกต้อง (Model Evaluation)	Value
จำนวนจุด Training Samples	93
จำนวนจุด Background	10,083
AUC	0.93510
Threshold by max TPR+TNR	0.34600

จากตารางที่ 4.31 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลอง MAXENT หลังการวิเคราะห์ห้วงค์ประกอบพบว่า จำนวนใช้จุด Training Samples ทั้งหมด 38 จุด จำนวนจุด Background Points ใช้ทั้งหมด 10,038 จุด มีค่า AUC เท่ากับ 0.9351 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 93.51 ค่า $\max \text{TPR} + \text{TNR}$ คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งใช้เป็นค่า Threshold เท่ากับ 0.3460 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 4.27 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลอง MAXENT หลังการวิเคราะห์ห้วงค์ประกอบ

จากภาพที่ 4.27 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าของแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบ ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่ หลังจากนั้นใช้ค่า Threshold สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence



ภาพที่ 4.28 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบเปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 4.28 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง MAXENT ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบนโดยไม่สนใจว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 13 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 12 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 2 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในเขตจังหวัดแม่ฮ่องสอนและเชียงราย ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) ส่วนเส้นสีแดงจะแสดงขอบเขตของแต่ละจังหวัด พบว่าพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าส่วนใหญ่จะกระจุกอยู่ในพื้นที่ป่าทางทิศตะวันตกเฉียงเหนือของภาคเหนือในเขตจังหวัดเชียงใหม่ ช่วงป่ารอยต่อระหว่างจังหวัดเชียงใหม่และแม่ฮ่องสอน รวมถึงป่ารอยต่อระหว่างจังหวัดเชียงใหม่และเชียงราย นอกจากนี้ยังพบว่าการกระจายของพื้นที่ที่คาดว่าจะค้นพบต้นชาป่าบางส่วนทางทิศตะวันออกของภาคเหนือ ซึ่งอยู่ในเขตจังหวัดน่าน

4.3 การเปรียบเทียบความแตกต่างของแบบจำลองการกระจายของสิ่งมีชีวิต

การเปรียบเทียบความแตกต่างของแบบจำลองการกระจายของสิ่งมีชีวิต ในที่นี้คือการนำแผนที่การพยากรณ์ที่ได้จากแบบจำลองการกระจายของสิ่งมีชีวิตทั้ง 3 วิธี ได้แก่วิธี Generalized Linear Models (GLMs), วิธี Generalized Additive Models (GAMs) และวิธี Maximum Entropy (MAXENT) นำมาเปรียบเทียบกัน เพื่อต้องการศึกษาว่า แบบจำลองการกระจายของสิ่งมีชีวิตในแต่ละวิธีมีแผนที่พยากรณ์แตกต่างกันอย่างไร

4.3.1 การเปรียบเทียบความแตกต่างของแบบจำลอง GLMs ในแต่ละวิธีการสุ่มจุด

Pseudo-Absence

การเปรียบเทียบความแตกต่างของแบบจำลอง GLMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence เป็นการเปรียบเทียบแบบจำลอง GLMs ที่เกิดมาจากการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี ได้แก่ วิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1), วิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3)

ตารางที่ 4.32 ค่า Coefficient Value ของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี

Variables	Coefficient Value		
	Method1	Method2	Method3
Intercept	-86.043	-72.965	-176.800
Aspect		-0.003 *	0.003 *
DEM	0.005 *		
Distance	-0.093		-0.390 ***
Humidity	-0.127 *	-0.192 *	-0.110
Bio1			0.314 *
Bio2		-0.766 **	0.260
Bio3	0.890 ***	1.509 ***	0.538 .
Bio4	0.007 ***	0.015 *	
Bio5		-0.798 *	
Bio6		0.642	-0.917 ***
Bio7		1.065 **	
Bio8	0.135 *	0.147	0.532 ***
Bio10	-0.080	-0.510 *	
Bio11		0.490 .	
Bio12	0.021 ***	0.009 *	0.042 ***
Bio13		-0.049 ***	0.064 **
Bio14	-0.135 *	-0.259 *	-0.627 ***
Bio16	-0.039 ***		-0.107 ***
Bio17		0.417 *	0.156 ***
Bio19		-0.247	

ตารางที่ 4.32 (ต่อ)

Variables	Coefficient Value		
	Method1	Method2	Method3
Class1	14.651	15.429	17.360
Class2	14.514	14.652	17.920
Class5	14.266	15.313	17.680
Class6	13.562	14.971	17.300
Class7			-70.300

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 4.32 แสดงค่า Coefficient Value ของแบบจำลอง GLMs โดยเปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี พบว่าตัวแปรที่เข้าไปในแบบจำลองทั้ง 3 วิธี (ตัวหนา) ได้แก่ตัวแปรความชื้นสัมพัทธ์ (Humidity), Isothermality (Bio3), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) และกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5, Class6) แสดงว่าตัวแปรดังกล่าวมีอิทธิพลต่อการกระจายของต้นชาป่า ถึงแม้ว่าในการสร้างแบบจำลองจะใช้วิธีการสุ่มจุด Pseudo-Absence Points ที่แตกต่างกันก็ตาม

โดยทั้ง 3 วิธีมีตัวแปร Bio3, Bio8, Bio12, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวกเหมือนกัน หมายความว่าตัวแปรดังกล่าวจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ สามารถอธิบายได้ดังนี้

ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างช่วงความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลเชิงบวก นั่นคือยิ่งค่า Isothermality (Bio3) มีค่ามากหรือมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8) มีผลเชิงบวก หมายความว่าพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่ที่มีมากขึ้น เช่นเดียวกันกับตัวแปรปริมาณน้ำฝนรายปี (Bio12) ที่มีผลเชิงบวกเช่นเดียวกัน นั่นคือในพื้นที่ที่มีปริมาณน้ำฝนรายปีมากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่เพิ่มขึ้นด้วยเช่นกัน ส่วนกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) มีผลเชิงบวก หมายความว่าโครงสร้างทางธรณีวิทยาของชนิดหินไม่มีผลต่อแบบจำลองแต่พื้นที่ที่มีความลาดชันไม่เกิน 30 องศา โอกาสที่จะค้นพบต้นชาป่ายังมีมากขึ้น

ในขณะที่ตัวแปร Humidity และ Bio14 มีค่าสัมประสิทธิ์เป็นลบเหมือนกัน ซึ่งหมายความว่าตัวแปรดังกล่าวจะไปลดโอกาสของการค้นพบต้นชาป่าในพื้นที่ นั่นคือพื้นที่ใดที่มีความชื้นสัมพัทธ์ (Humidity) สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลง เช่นเดียวกับตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) หากพื้นที่ใดมีปริมาณน้ำฝนมากเกินไปโอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงเช่นกัน

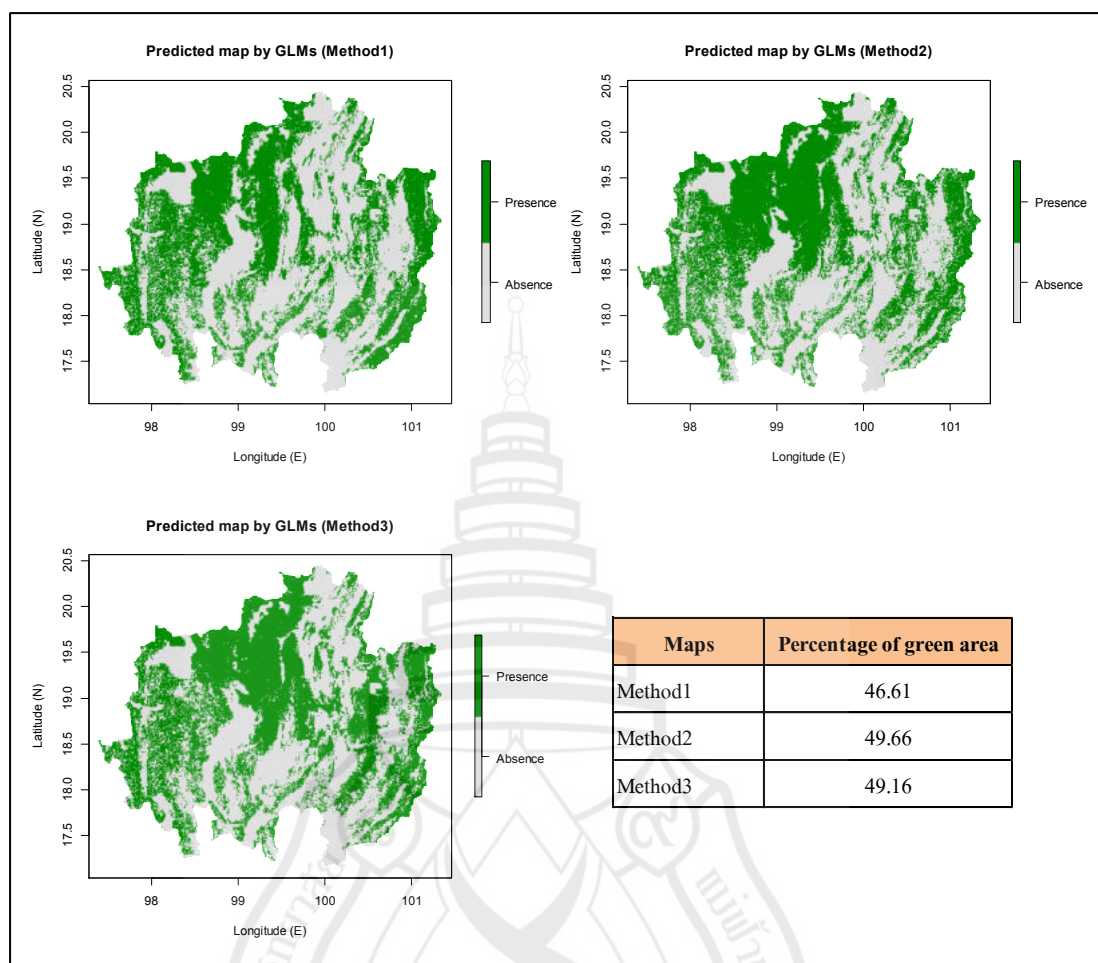
ตารางที่ 4.33 ผลการตรวจสอบแบบจำลอง GLMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธีโดยแยกตามประเภทของตัวแปร

Method of Selecting Pseudo-Absence	Geography Factors		Climate Factors		All Factors	
	R ²	AUC	R ²	AUC	R ²	AUC
Method 1	13.77	0.70214	19.50	0.76574	21.61	0.76236
Method 2	16.37	0.79426	38.27	0.89446	40.08	0.89284
Method 3	17.96	0.78099	48.39	0.93148	55.85	0.95262

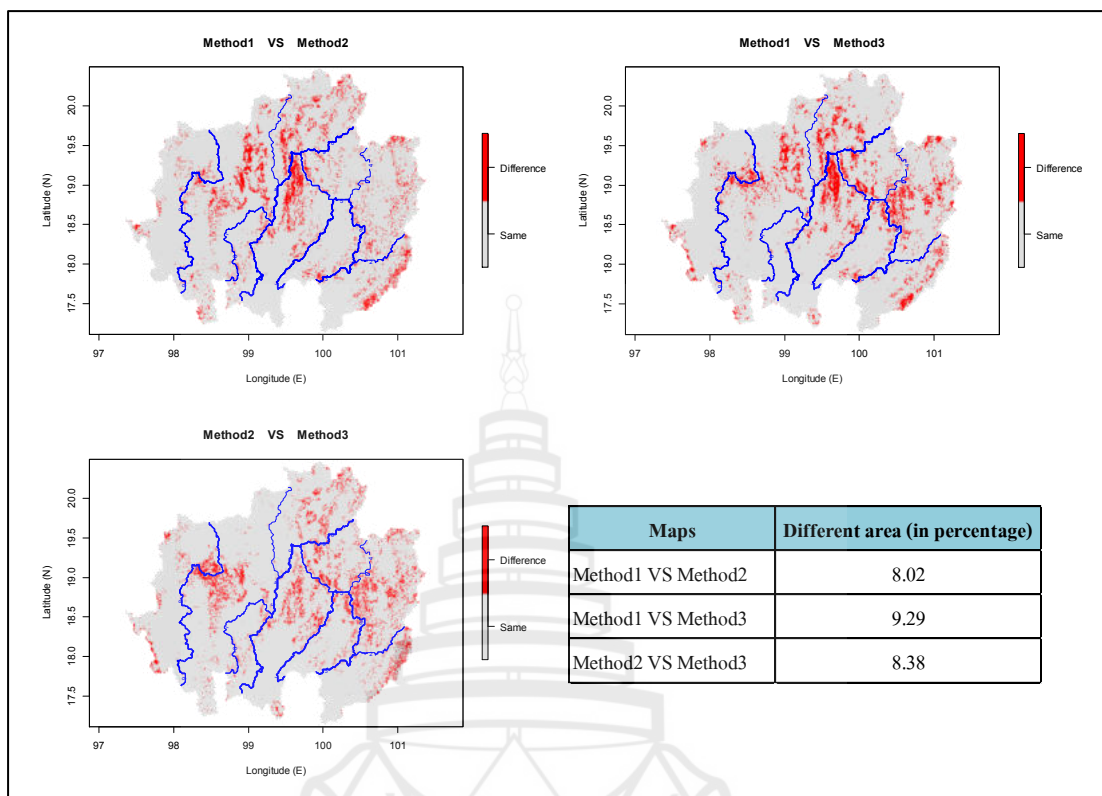
จากตารางที่ 4.33 แสดงผลการตรวจสอบแบบจำลอง GLMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธีโดยแยกตามประเภทของตัวแปร (ดูรายละเอียดแบบจำลองโดยนำเข้าตัวแปรทางด้านภูมิศาสตร์ (Geography Factors) และตัวแปรทางด้านภูมิอากาศ (Climate Factors) ในภาคผนวก ง) พบว่าในทุกวิธีการคัดเลือกจุด Pseudo-Absence นั้น การสร้างแบบจำลองโดยนำเข้าตัวแปรทุกตัวแปรและคัดเลือกตัวแปรโดยวิธี Stepwise ทำให้ความสามารถในการพยากรณ์ของแบบจำลองเพิ่มขึ้น นั่นคือ ค่า R² มีค่ามากกว่าการพิจารณาเพียงตัวแปรทางด้านภูมิศาสตร์ (Geography Factors) หรือพิจารณาเพียงตัวแปรทางด้านภูมิอากาศ (Climate Factors) และเมื่อเปรียบเทียบกันระหว่างวิธีการคัดเลือกจุด Pseudo-Absence ในแต่ละวิธีพบว่า แบบจำลองที่เกิด

จากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีค่า R^2 มากที่สุด หมายความว่าแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีความสามารถในการพยากรณ์ของแบบจำลองมากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ และเมื่อพิจารณาค่า AUC ซึ่งเป็นค่าแสดงความถูกต้องของการพยากรณ์พบว่า ในทุกวิธีของการคัดเลือกจุด Pseudo-Absence นั้นการพิจารณาเฉพาะตัวแปรทางด้านภูมิศาสตร์จะให้ค่าแสดงความถูกต้องของการพยากรณ์น้อยที่สุด ในขณะที่การพิจารณาเฉพาะตัวแปรทางด้านภูมิอากาศและการพิจารณาของทุกตัวแปรจะให้ค่าแสดงความถูกต้องของการพยากรณ์ใกล้เคียงกันหรือต่างกันเพียงเล็กน้อยเท่านั้น

จากภาพที่ 4.29 แสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธีพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า (พื้นที่สีเขียว) ไม่ค่อยมีความแตกต่างกันมากนัก ซึ่งกระจายอยู่ทั่วไปตามพื้นที่ป่าในเขตพื้นที่ภาคเหนือตอนบน โดยส่วนใหญ่จะอยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของพื้นที่ภาคเหนือตอนบน ซึ่งได้แก่บริเวณจังหวัดเชียงใหม่และแม่ฮ่องสอนตอนบน ซึ่งในวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่ามากที่สุด ซึ่งเท่ากับร้อยละ 49.66 รองลงมาคือ วิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) เท่ากับร้อยละ 49.16 และสุดท้ายวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) มีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าน้อยที่สุด ซึ่งเท่ากับร้อยละ 46.61 จะสังเกตได้ว่า วิธีการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าที่ใกล้เคียงกันมาก



ภาพที่ 4.29 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี



ภาพที่ 4.30 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี

จากภาพที่ 4.30 แสดงพื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี พบว่าพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันในแต่ละคู่มิกระจายอยู่ทั่วไปในเขตพื้นที่ภาคเหนือตอนบนเพียงเล็กน้อย และเมื่อดูอัตราส่วนของพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันพบว่า ในแต่ละคู่มีความแตกต่างกันไม่ถึงร้อยละ 10 ของพื้นที่พยากรณ์ นั่นคือระหว่างวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีพื้นที่พยากรณ์ที่แตกต่างกันเพียงร้อยละ 8.02 ระหว่างวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีพื้นที่พยากรณ์ที่แตกต่างกันเพียงร้อยละ 9.29 สุดท้ายระหว่างวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีพื้นที่พยากรณ์ที่แตกต่างกันเพียงร้อยละ 8.38

แต่ทั้งนี้ทั้งนั้นอัตราส่วนของพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันเป็นเพียงการดูโดยภาพรวมเท่านั้น ไม่สามารถที่จะอธิบายได้ว่าพื้นที่ ณ จุดเดียวกันของแผนที่พยากรณ์ทั้งสองแผนที่มีความเหมือนหรือแตกต่างกันหรือไม่ จึงทำการทดสอบความแตกต่างโดยใช้วิธี Cochran's Q Test ซึ่งเริ่มต้นจากการสุ่มจุดในแผนที่พยากรณ์ และทำการตรวจสอบค่าของแผนที่พยากรณ์ ณ จุดที่ถูกสุ่มนั้น ว่ามีความเหมือนหรือแตกต่างกันหรือไม่

ตารางที่ 4.34 ผลของการทดสอบโดยวิธี Cochran's Q Test

การสุ่มจุด	Cochran's		การทดสอบแบบคู่	
	Q test		คู่ของแผนที่พยากรณ์	Wilcoxon Sign Test (P-Value)
ครั้งที่ 1	65.73 ***		Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	0.960
ครั้งที่ 2	64.23 ***		Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	0.448
ครั้งที่ 3	82.51 ***		Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	0.557
ครั้งที่ 4	77.48 ***		Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	0.664

ตารางที่ 4.34 (ต่อ)

การสุ่มจุด	Cochran's		การทดสอบแบบคู่	
	Q test		คู่ของแผนที่พยากรณ์	Wilcoxon Sign Test (P-Value)
ครั้งที่ 5	62.30	***	Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	0.526

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

จากตารางที่ 4.34 แสดงผลของการทดสอบโดยวิธี Cochran's Q Test โดยการสุ่มจุดจำนวน 5,000 จุด จากจำนวนพิกเซลในแผนที่พยากรณ์ทั้งหมด 94,536 พิกเซลหรือประมาณร้อยละ 5 ซึ่งทำการสุ่มจุดเพื่อทดสอบความแตกต่างของแผนที่พยากรณ์ที่เกิดจากการพยากรณ์โดยแบบจำลอง GLMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี ได้แก่ วิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1), วิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) พบว่าในการสุ่มจุดเพื่อทำการทดสอบความแตกต่างทั้ง 5 ครั้งให้ผลที่เหมือนกันกล่าวคือ แผนที่พยากรณ์ที่เกิดจากการพยากรณ์โดยแบบจำลอง GLMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธีมีแผนที่พยากรณ์ที่แตกต่างกัน และเมื่อทำการทดสอบทีละคู่โดยวิธี Wilcoxon sign test พบว่าคู่ที่มีความแตกต่างกันได้แก่แผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method1 แตกต่างจากแผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method2 และแผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method1 แตกต่างจากแผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method3 โดยที่แผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method2 ไม่มีความแตกต่างจากแผนที่พยากรณ์ GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ใน Method3

สรุปได้ว่า ในการทดสอบความแตกต่างของแผนที่พยากรณ์นั้น ถึงแม้ว่าจะมองในภาพรวมจะดูเหมือนไม่มีความแตกต่างกัน แต่ถ้าหากมองในส่วนที่ย่อยลงไปในพื้นที่ในระดับพิกเซลขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร กลับพบว่ามีความแตกต่างกันของการพยากรณ์โดยแบบจำลอง GLMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธีอย่างมีนัยสำคัญทางสถิติ

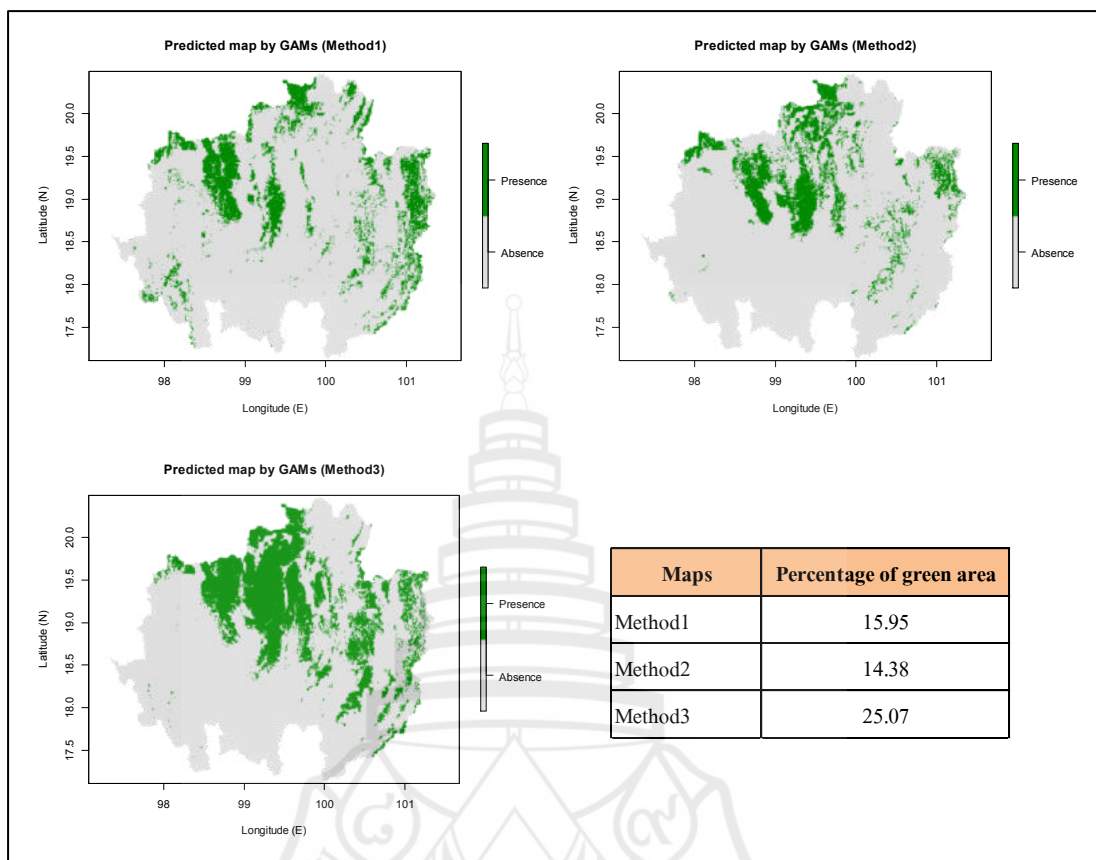
4.3.2 การเปรียบเทียบความแตกต่างของแบบจำลอง GAMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence

การเปรียบเทียบความแตกต่างของแบบจำลอง GAMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence เป็นการเปรียบเทียบแบบจำลอง GAMs ที่เกิดมาจากการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี ได้แก่ วิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1), วิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3)

ตารางที่ 4.35 ผลการตรวจสอบแบบจำลอง GAMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธี

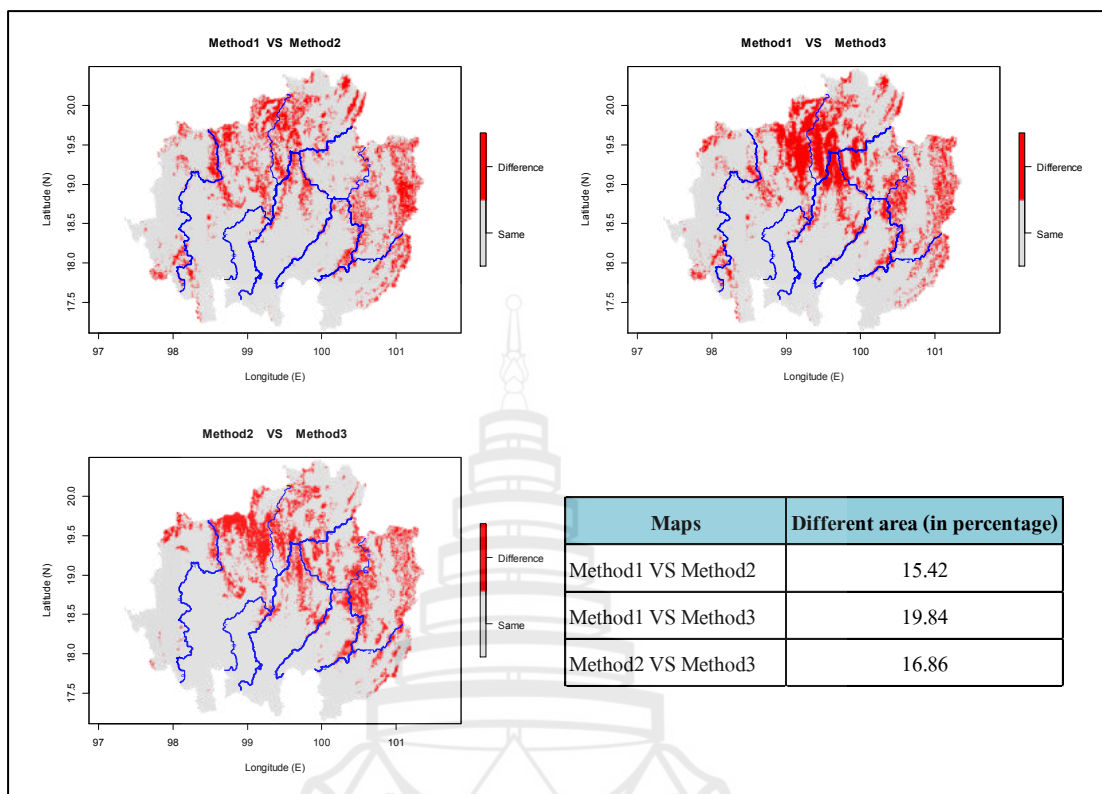
	Method of Selecting Pseudo-Absence		
	Method 1	Method 2	Method 3
R^2	32.00	78.00	67.40
AUC	0.83283	0.94675	0.97670

จากตารางที่ 4.35 แสดงผลการตรวจสอบแบบจำลอง GAMs ที่ได้จากการสุ่มจุด Pseudo-Absence ในแต่ละวิธี พบว่า ความสามารถในการพยากรณ์ของแบบจำลอง (R^2) ของแบบจำลองที่เกิดจากการสุ่ม Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีความสามารถในการพยากรณ์ของแบบจำลองมากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ ในขณะที่ค่า AUC ซึ่งเป็นค่าแสดงความถูกต้องของการพยากรณ์ที่คำนวณได้ในแต่ละวิธีการสุ่ม Pseudo-Absence พบว่า ค่า AUC ของแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีความถูกต้องของการพยากรณ์มากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ



ภาพที่ 4.31 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GLMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี

จากภาพที่ 4.31 แสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลอง GAMs เปรียบเทียบระหว่างวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธีพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า (พื้นที่สีเขียว) มีความแตกต่างกันอย่างเห็นได้ชัดเจน โดยเฉพาะแบบจำลองที่มาจากวิธีการสุ่ม Pseudo-Absence แบบ 2 ขั้นตอน (Method3) ซึ่งมีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่ามากที่สุด ซึ่งเท่ากับร้อยละ 25.07 ในขณะที่แบบจำลองที่มาจากวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าไม่แตกต่างกันมากนัก ซึ่งมีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าอยู่ระหว่างร้อยละ 14-16 ของพื้นที่ทั้งหมดเท่านั้น โดยพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่ากระจายอยู่ทั่วไปตามพื้นที่ป่าในเขตพื้นที่ภาคเหนือตอนบน โดยส่วนใหญ่จะอยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของพื้นที่ภาคเหนือตอนบน



ภาพที่ 4.32 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี

จากภาพที่ 4.32 แสดงพื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบทีละคู่ของวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี พบว่าพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันในแต่ละคู่มีกระจายอยู่ทั่วไปในเขตพื้นที่ภาคเหนือตอนบนลงมาทางด้านทิศตะวันออกของพื้นที่ และเมื่อดูอัตราส่วนของพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันพบว่า ในแต่ละคู่มีความแตกต่างกันไม่เกินร้อยละ 20 ของพื้นที่ทั้งหมด โดยพื้นที่ที่มีการพยากรณ์ระหว่างวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และวิธีการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3) มีพื้นที่พยากรณ์ที่แตกต่างกันมากที่สุดถึงร้อยละ 19.83 ซึ่งอัตราส่วนของพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันเป็นเพียงการดูโดยภาพรวมเท่านั้น ไม่สามารถที่จะอธิบายได้ว่าพื้นที่ ณ จุดเดียวกันของแผนที่พยากรณ์ทั้งสองแผนที่ มีความเหมือนหรือแตกต่างกันหรือไม่ จึงทำการทดสอบความแตกต่างโดยใช้วิธี Cochran's Q Test ซึ่งเริ่มต้นจากการสุ่มจุดในแผนที่พยากรณ์ และทำการตรวจสอบค่าของแผนที่พยากรณ์ ณ จุดที่ถูกสุ่มนั้น ว่ามีความเหมือนหรือแตกต่างกันหรือไม่

ตารางที่ 4.36 ผลของการทดสอบโดยวิธี Cochran's Q Test

การสุ่มจุด	Cochran's Q test		การทดสอบแบบคู่	
			คู่ของแผนที่พยากรณ์	Wilcoxon sign test (P-value)
ครั้งที่ 1	412.73	***	Method 1 และ Method 2	0.001 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	< 0.000 ***
ครั้งที่ 2	365.29	***	Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	< 0.000 ***
ครั้งที่ 3	368.76	***	Method 1 และ Method 2	< 0.000 ***
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	< 0.000 ***
ครั้งที่ 4	382.11	***	Method 1 และ Method 2	0.005 **
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	< 0.000 ***
ครั้งที่ 5	339.61	***	Method 1 และ Method 2	0.005 **
			Method 1 และ Method 3	< 0.000 ***
			Method 2 และ Method 3	< 0.000 ***

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

** หมายถึง มีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 4.36 แสดงผลของการทดสอบโดยวิธี Cochran's Q Test โดยการสุ่มจุดจำนวน 5,000 จุด จากจำนวนพิกเซลในแผนที่พยากรณ์ทั้งหมด 94,536 พิกเซลหรือประมาณร้อยละ 5 ซึ่งทำการสุ่มจุดเพื่อทดสอบความแตกต่างของแผนที่พยากรณ์ที่เกิดจากการพยากรณ์โดยแบบจำลอง GAMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธี ได้แก่ วิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1), วิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด

Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) พบว่า ในการสุ่มจุดเพื่อทำการทดสอบความแตกต่างทั้ง 5 ครั้งให้ผลที่เหมือนกันกล่าวคือ แผนที่พยากรณ์ที่เกิดจากการพยากรณ์โดยแบบจำลอง GAMs ในแต่ละวิธีการสุ่มจุด Pseudo-Absence ทั้ง 3 วิธีมีแผนที่พยากรณ์ที่แตกต่างกัน และเมื่อทำการทดสอบทีละคู่โดยวิธี Wilcoxon sign test พบว่าทุกคู่มีความแตกต่างกัน สรุปได้ว่าการทดสอบความแตกต่างของแผนที่พยากรณ์นั้น ถึงแม้ว่าจะมองในภาพรวมของแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) จะดูเหมือนไม่มีความแตกต่างกัน แต่ถ้าหากมองในส่วนที่ขยับลงไปในพื้นที่ในระดับพิกเซลขนาด 1 กิโลเมตร $\times$ 1 กิโลเมตร กลับพบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

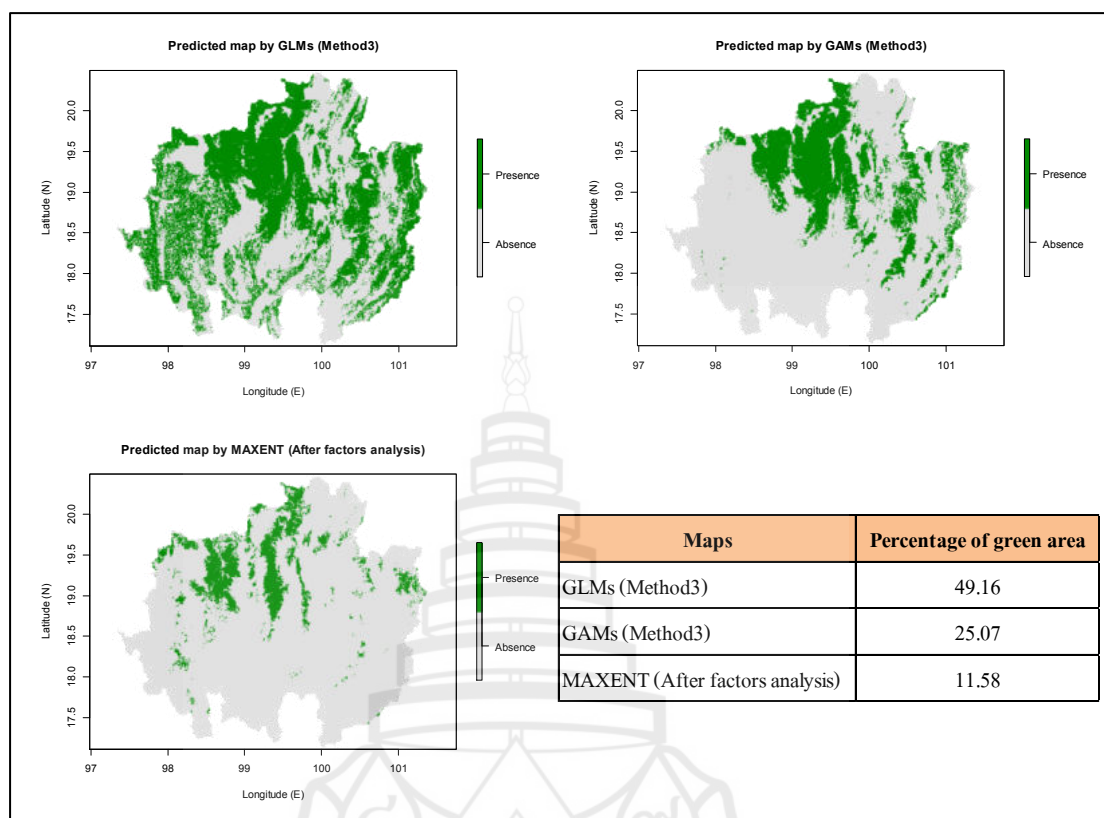
4.3.4 การเปรียบเทียบความแตกต่างของแบบจำลอง GLMs, GAMs และ MAXENT

การเปรียบเทียบความแตกต่างของแบบจำลอง GLMs , GAMs และ MAXENT เป็นการเปรียบเทียบแบบจำลอง GLMs ที่เกิดมาจากการสุ่มจุด Pseudo-Absence โดยวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method2), แบบจำลอง GAMs ที่เกิดมาจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) ซึ่งทั้ง 2 แบบจำลองต่างก็ให้ค่า AUC สูงสุดและแบบจำลอง MAXENT หลังการทำการวิเคราะห์ห่อหุ้มประกอบ

ตารางที่ 4.37 ผลการตรวจสอบแบบจำลอง GLMs, GAMs และ MAXENT

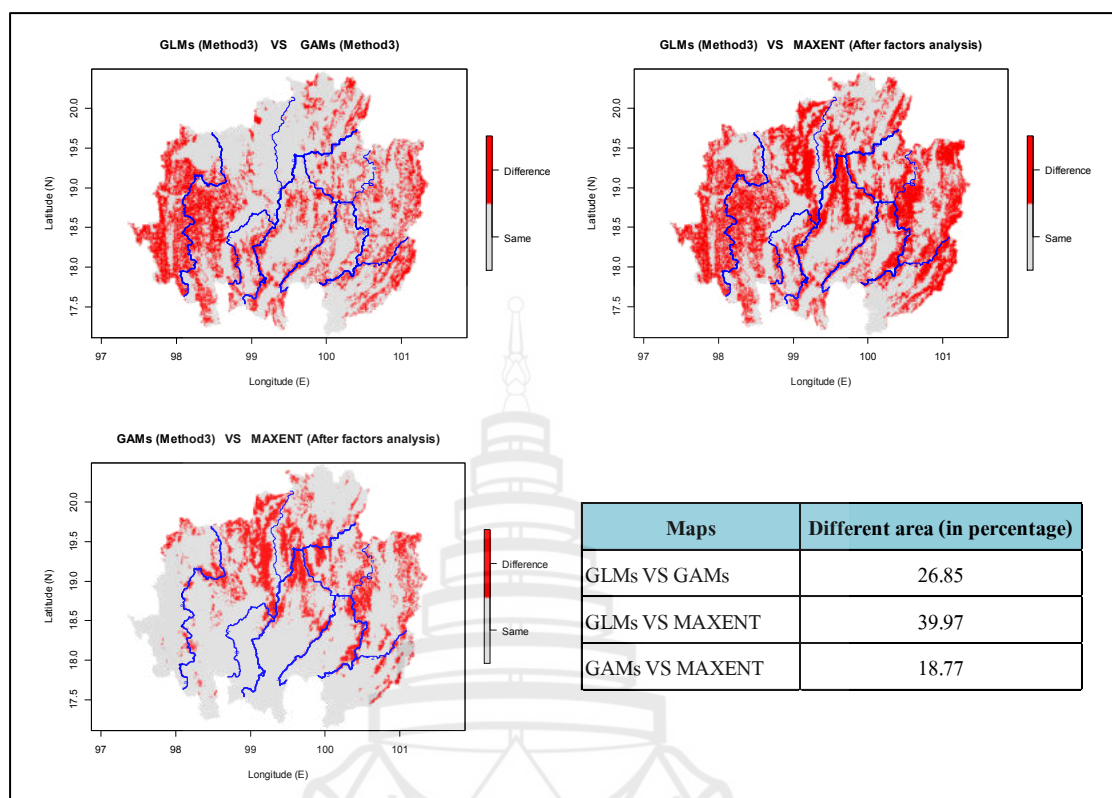
Model	R ²	AUC
GLMs (Method3)	55.85	0.95262
GAMs (Method3)	78.00	0.97670
MAXENT (After Factors Analysis)		0.93510

จากตารางที่ 4.37 แสดงผลการตรวจสอบแบบจำลอง GLMs, GAMs และ MAXENT พบว่าการสร้างแบบจำลอง GAMs ทำให้ความสามารถในการพยากรณ์ของแบบจำลองเพิ่มขึ้น นั่นคือ ค่า R² มีค่ามากกว่าการสร้างแบบจำลอง GLMs และเมื่อพิจารณาค่า AUC ซึ่งเป็นค่าแสดงความถูกต้องของการพยากรณ์พบว่าการสร้างแบบจำลอง GAMs จะให้ค่าแสดงความถูกต้องของการพยากรณ์มากที่สุด รองลงมาคือการสร้างแบบจำลอง MAXENT ก่อนการวิเคราะห์ห่อหุ้มประกอบ, การสร้างแบบจำลอง GLMs (Method3) และการสร้างแบบจำลอง MAXENT หลังการวิเคราะห์ห่อหุ้มประกอบตามลำดับ



ภาพที่ 4.33 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าเปรียบเทียบกับระหว่างแบบจำลอง GLMs, GAMs และ MAXENT

จากภาพที่ 4.33 แสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าเปรียบเทียบกับระหว่างแบบจำลอง GLMs (Method3), GAMs (Method3) และ MAXENT หลังการวิเคราะห์องค์ประกอบพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าพบว่าแบบจำลอง GLMs มีอัตราส่วนของพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่ามากที่สุด รองลงมาคือแบบจำลอง GAMs, MAXENT หลังการวิเคราะห์องค์ประกอบ ซึ่งทั้ง 3 แบบจำลอง ส่วนใหญ่จะกระจายอยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของพื้นที่



ภาพที่ 4.34 พื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบระหว่างแบบจำลอง GLMs, GAMs และ MAXENT

จากภาพที่ 4.34 แสดงพื้นที่ที่มีการพยากรณ์ที่แตกต่างกัน โดยเปรียบเทียบระหว่างแบบจำลอง GLMs (Method3), GAMs (Method3) และ MAXENT (After factor analysis) พบว่าแต่ละคู่ของแบบจำลองต่างก็มีพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันอย่างชัดเจน โดยคู่ของแบบจำลอง GAMs (Method3) และ MAXENT (After factor analysis) มีพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันมากที่สุดเกือบร้อยละ 40 ของพื้นที่ทั้งหมด รองลงมาได้แก่คู่ของแบบจำลอง GLMs (Method3) และ GAMs (Method3) ที่มีพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันประมาณร้อยละ 27 ของพื้นที่ สุดท้ายคือคู่ของแบบจำลอง GAMs (Method3) และ MAXENT (After factor analysis) มีพื้นที่ที่มีการพยากรณ์ที่แตกต่างกันน้อยที่สุด ประมาณร้อยละ 19 ของพื้นที่ทั้งหมด และเพื่อตรวจสอบความถูกต้องว่าแผนที่พยากรณ์มีความแตกต่างกันจริง จึงทำการทดสอบความแตกต่างโดยใช้วิธี Cochran's Q Test ซึ่งเริ่มต้นจากการสุ่มจุดในแผนที่พยากรณ์ และทำการตรวจสอบค่าของแผนที่พยากรณ์ ณ จุดที่ถูกสุ่มนั้น ว่ามีความเหมือนหรือแตกต่างกันหรือไม่

ตารางที่ 4.38 ผลของการทดสอบโดยวิธี Cochran's Q Test

การสุ่มจุด	Cochran's Q test		การทดสอบแบบคู่	
			คู่ของแผนที่พยากรณ์	Wilcoxon Sign Test (P-Value)
ครั้งที่ 1	2504.49	***	GLMs และ GAMs	< 0.001 ***
			GLMs และ MAXENT	< 0.000 ***
			GAMs และ MAXENT	< 0.000 ***
ครั้งที่ 2	2553.64	***	GLMs และ GAMs	< 0.000 ***
			GLMs และ MAXENT	< 0.000 ***
			GAMs และ MAXENT	< 0.000 ***
ครั้งที่ 3	2604.29	***	GLMs และ GAMs	< 0.000 ***
			GLMs และ MAXENT	< 0.000 ***
			GAMs และ MAXENT	< 0.000 ***
ครั้งที่ 4	2587.75	***	GLMs และ GAMs	< 0.001 ***
			GLMs และ MAXENT	< 0.000 ***
			GAMs และ MAXENT	< 0.000 ***
ครั้งที่ 5	2633.52	***	GLMs และ GAMs	0.005 ***
			GLMs และ MAXENT	< 0.000 ***
			GAMs และ MAXENT	< 0.000 ***

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

จากตารางที่ 4.38 แสดงผลของการทดสอบโดยวิธี Cochran's Q Test โดยการสุ่มจุดจำนวน 5,000 จุด จากจำนวนพิกเซลในแผนที่พยากรณ์ทั้งหมด 94,536 พิกเซลหรือประมาณร้อยละ 5 ซึ่งทำการสุ่มจุดเพื่อทดสอบความแตกต่างของแผนที่พยากรณ์ที่เกิดจากการพยากรณ์โดยแบบจำลอง GLMs, GAMs และ MAXENT พบว่า ในการสุ่มจุดเพื่อทำการทดสอบความแตกต่างทั้ง 5 ครั้ง ให้ผลที่เหมือนกันกล่าวคือ แผนที่พยากรณ์ที่เกิดจากการพยากรณ์ในแต่ละแบบจำลองให้ผลที่

แตกต่างกัน และเมื่อทำการทดสอบทีละคู่โดยวิธี Wilcoxon Sign Test พบว่าทุกคู่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ 0.001

4.4 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นชาป่า

ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นชาป่าเป็นการศึกษาผลกระทบของการเปลี่ยนแปลงของสภาพภูมิอากาศที่มีต่อการกระจายของต้นชาป่า โดยเปรียบเทียบระหว่างสภาพภูมิอากาศในปัจจุบัน (เก็บรวบรวมข้อมูลระหว่างปี ค.ศ.1950 – 2000) และปีในอนาคต โดยใช้ตัวแปร Bioclimatic Variables ทั้งหมด 19 ตัวแปรในปี ค.ศ. 2050 (ค่าเฉลี่ยระหว่างปี ค.ศ. 2041 – 2060) จาก http://www.worldclim.org/cmip5_30s#com

ตารางที่ 4.39 ค่าสถิติของปัจจัยอุณหภูมิในปัจจุบันเปรียบเทียบกับอนาคต

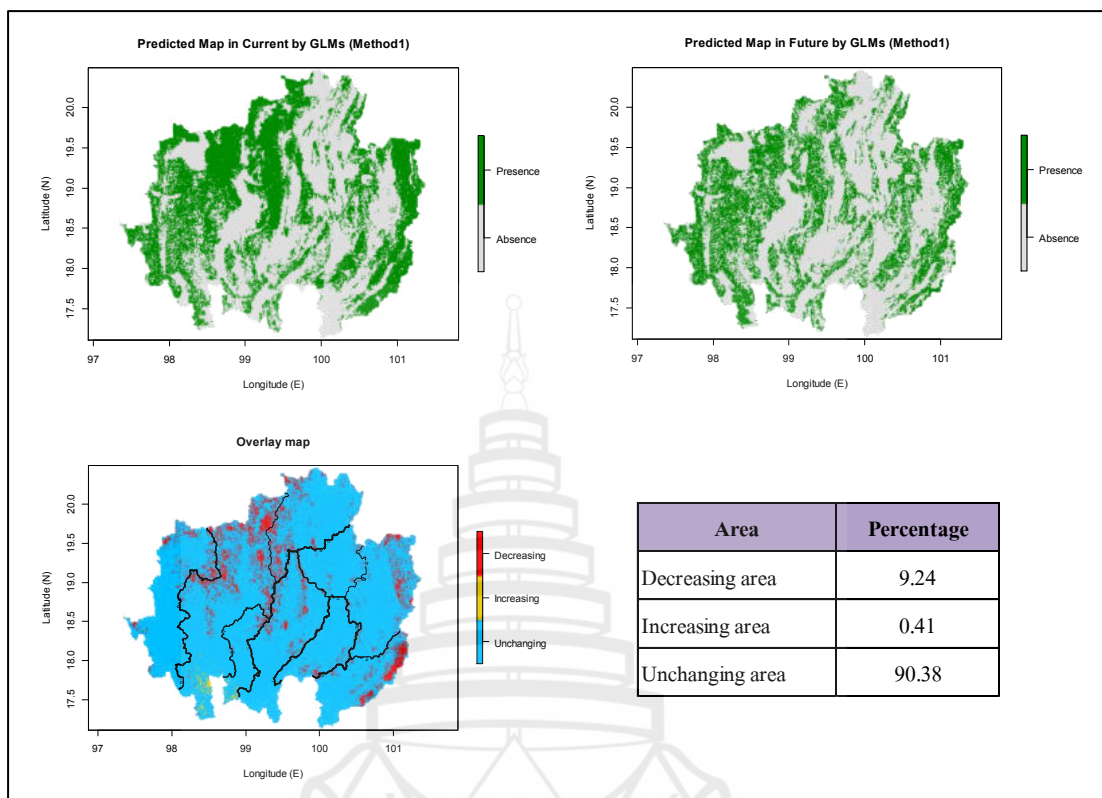
ตัวแปร	ปัจจัยอุณหภูมิในปัจจุบัน		ปัจจัยอุณหภูมิในอนาคต	
	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
Bio1	245.10	16.36	264.37	15.99
Bio2	122.44	2.25	116.73	2.52
Bio3	52.09	1.10	50.50	1.06
Bio4	2436.15	180.34	2501.95	198.94
Bio5	351.03	18.75	369.98	18.06
Bio6	117.93	18.02	141.12	18.07
Bio7	233.38	4.78	228.87	3.96
Bio8	255.52	15.57	272.21	15.01
Bio9	214.09	18.57	238.33	20.76
Bio10	270.87	16.39	293.09	15.83
Bio11	208.83	17.76	227.85	17.93

จากตารางที่ 4.39 แสดงค่าสถิติของปัจจัยทางอุณหภูมิในปัจจุบันเปรียบเทียบกับอนาคต พบว่าตัวแปรอุณหภูมิในอนาคตที่มีค่าเฉลี่ยลดลงเมื่อเทียบกับในปัจจุบัน (ตัวหนา) คือตัวแปรช่วงอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2), Isothermality (Bio3) และช่วงของอุณหภูมิรายปี (Bio7) ในขณะที่ตัวแปรอื่น ๆ จะมีค่าเฉลี่ยเพิ่มขึ้นเมื่อเปรียบเทียบกับในปัจจุบัน หมายความว่าช่วงอุณหภูมิระหว่างกลางวันและกลางคืนมีค่าลดลงและความแตกต่างของอุณหภูมิระหว่างฤดูหนาวและฤดูร้อนลดน้อยลง เมื่อพิจารณาตัวแปรอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6) พบว่าอุณหภูมิเพิ่มขึ้น 2.32 องศาเซลเซียส เช่นเดียวกับตัวแปรอุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5) พบว่าอุณหภูมิเพิ่มขึ้น 1.90 องศาเซลเซียส ซึ่งอุณหภูมิต่ำสุดในเดือนที่หนาวที่สุดเพิ่มขึ้นมากกว่าอุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด ส่งผลให้ความแตกต่างของอุณหภูมิระหว่างฤดูหนาวและฤดูร้อนลดน้อยลง

ตารางที่ 4.40 ค่าสถิติของปัจจัยปริมาณน้ำฝนในปัจจุบันเปรียบเทียบกับอนาคต

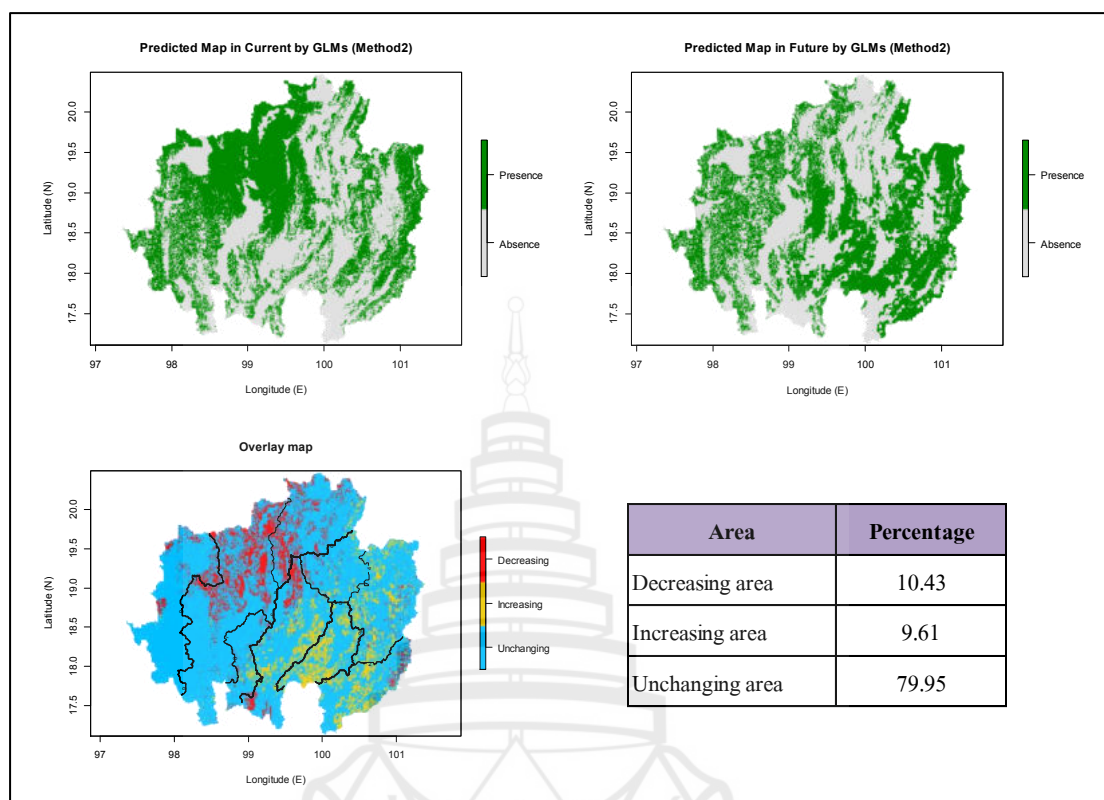
ตัวแปร	ปัจจัยปริมาณน้ำฝนในปัจจุบัน		ปัจจัยปริมาณน้ำฝนในอนาคต	
	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
Bio12	1232.46	188.76	1266.39	200.78
Bio13	256.40	61.05	269.87	64.27
Bio14	7.99	3.16	11.92	4.35
Bio15	82.65	4.94	81.27	7.18
Bio16	641.93	133.28	690.4	150.61
Bio17	33.09	9.70	51.73	16.24
Bio18	367.02	67.22	285.10	88.21
Bio19	36.62	17.24	60.62	24.47

จากตารางที่ 4.40 แสดงค่าสถิติของปัจจัยปริมาณน้ำฝนในปัจจุบันเปรียบเทียบกับอนาคต พบว่าตัวแปรปริมาณน้ำฝนในอนาคตที่มีค่าเฉลี่ยลดลงเมื่อเทียบกับในปัจจุบัน (ตัวหนา) คือตัวแปรช่วงฤดูกาลของปริมาณน้ำฝน (Bio15) และตัวแปรปริมาณน้ำฝนในไตรมาสที่ร้อนที่สุด (Bio18) ในขณะที่ตัวแปรอื่น ๆ จะมีค่าเฉลี่ยเพิ่มขึ้นเมื่อเปรียบเทียบกับในปัจจุบัน หมายความว่าฝนที่ตกต้องตามฤดูกาลมีปริมาณลดลง นั่นคือความแปรปรวนของปริมาณน้ำฝนลดลง ในขณะที่ช่วงฤดูร้อนก็จะมีปริมาณน้ำฝนลดลงด้วย แสดงว่าในช่วงเวลาดังกล่าวมีความแห้งแล้งมากขึ้น



ภาพที่ 4.35 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1)

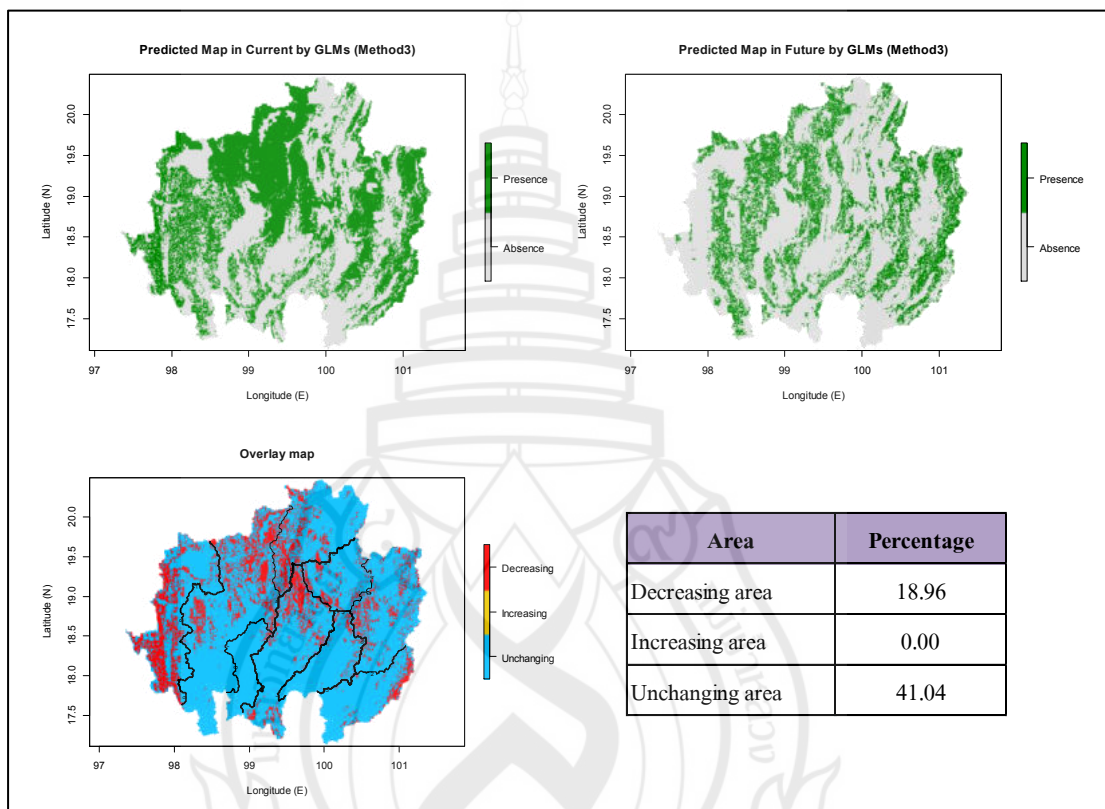
จากภาพที่ 4.35 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) พบว่าพื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลงในปี ค.ศ. 2050 และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่าพื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 90.38 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 9.24 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 0.41 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าจะอยู่ทางทิศตะวันตกเฉียงใต้ของพื้นที่ภาคเหนือตอนบนซึ่งอยู่แถว ๆ ตอนล่างของจังหวัดเชียงใหม่และจังหวัดลำพูน ส่วนพื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่โดยทั่วไปในเขตพื้นที่ภาคเหนือตอนบน



ภาพที่ 4.36 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2)

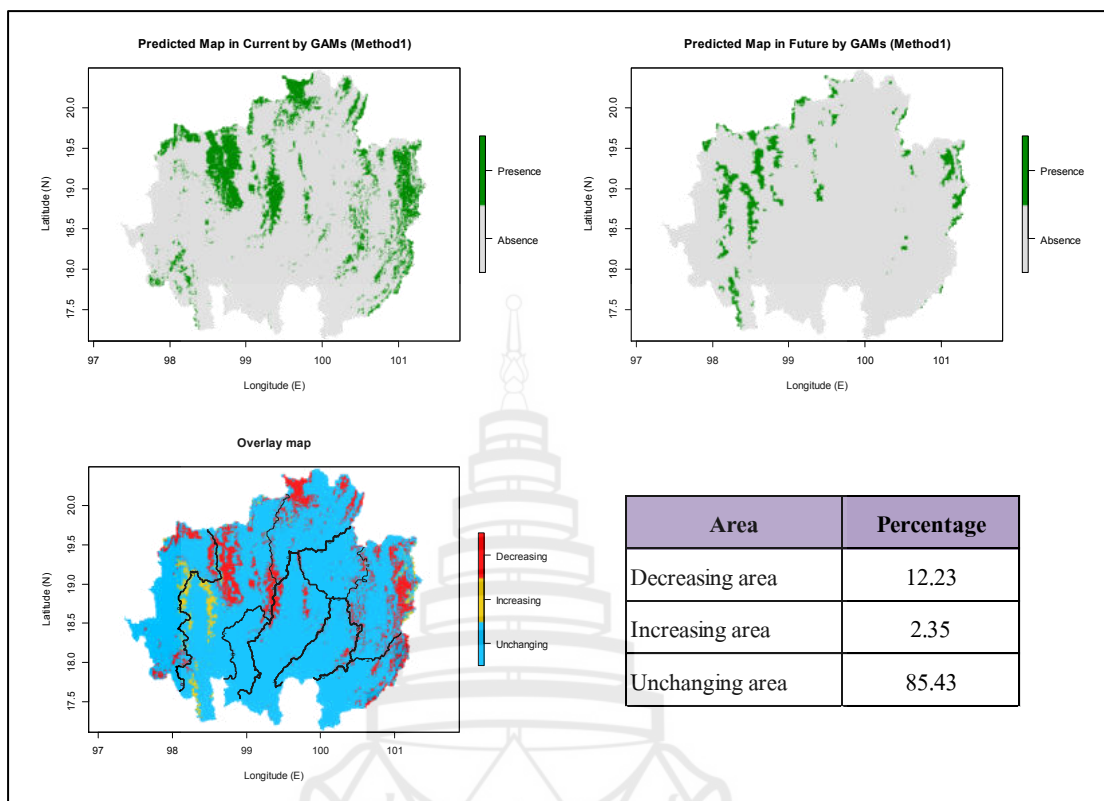
จากภาพที่ 4.36 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) พบว่าในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลงทางทิศตะวันตกเฉียงเหนือของพื้นที่ภาคเหนือตอนบนและมีแนวโน้มเพิ่มขึ้นทางทิศตะวันออกและทิศตะวันออกเฉียงใต้ของพื้นที่ภาคเหนือตอนบน และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่าพื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 79.95 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของการกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 10.43 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของการกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 9.61 ซึ่งจะสังเกตได้ว่าพื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าส่วนใหญ่จะกระจายอยู่ทางทิศตะวันออก ทิศตะวันออกเฉียงใต้และทิศใต้ของพื้นที่ภาคเหนือตอนบนแทน ซึ่งเป็นพื้นที่ของจังหวัดน่าน, แพร่, ลำปางและอุตรดิตถ์ โดยมีบางส่วนกระจายอยู่ในจังหวัดพะเยาและลำพูน ในขณะที่พื้นที่ที่

คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่าส่วนใหญ่จะกระจายอยู่ทางทิศเหนือและทิศตะวันออกเฉียงเหนือ ซึ่งแต่เดิมเป็นพื้นที่ที่คาดว่าจะมีต้นชากระจายอยู่อย่างหนาแน่น ซึ่งส่วนใหญ่เป็นพื้นที่ของจังหวัดเชียงรายและเชียงใหม่ ส่วนจังหวัดแม่ฮ่องสอนพื้นที่ส่วนใหญ่ไม่ค่อยมีการเปลี่ยนแปลง



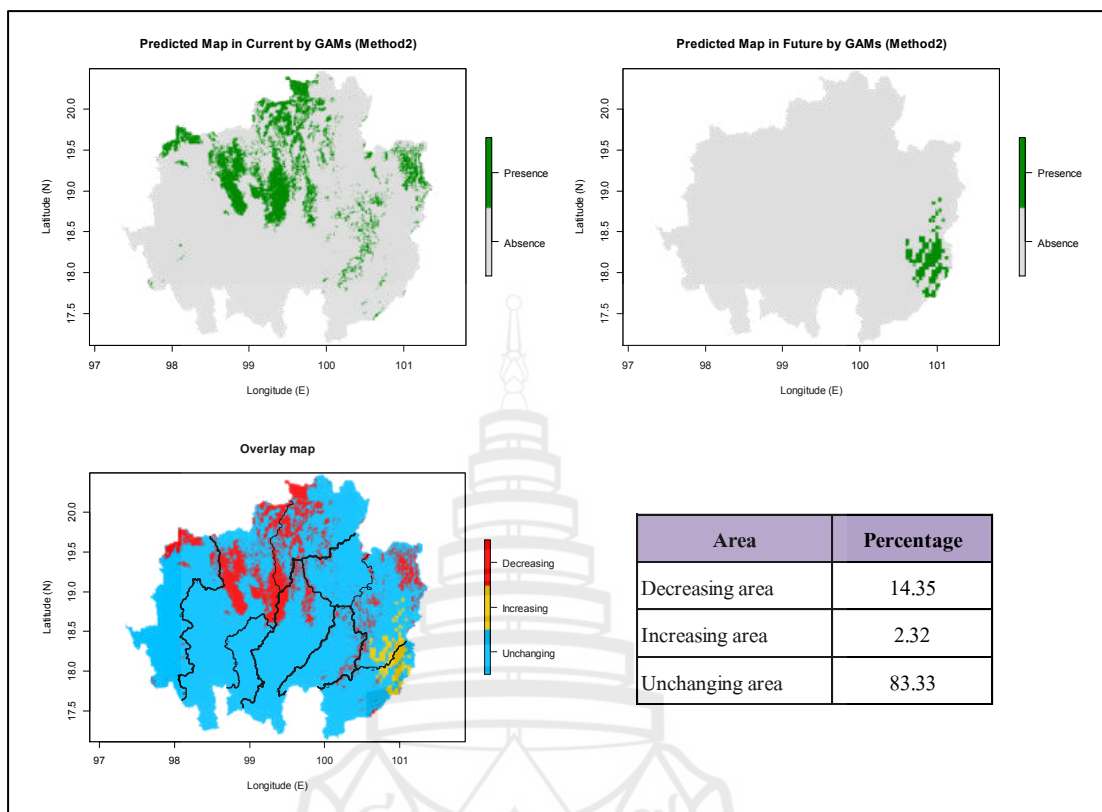
ภาพที่ 4.37 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3)

จากภาพที่ 4.37 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3) พบว่าในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลงทั่วทั้งพื้นที่ของภาคเหนือตอนบน และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่าพื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 81.04 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 18.96 ซึ่งแทบจะไม่มีพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าในพื้นที่เลย



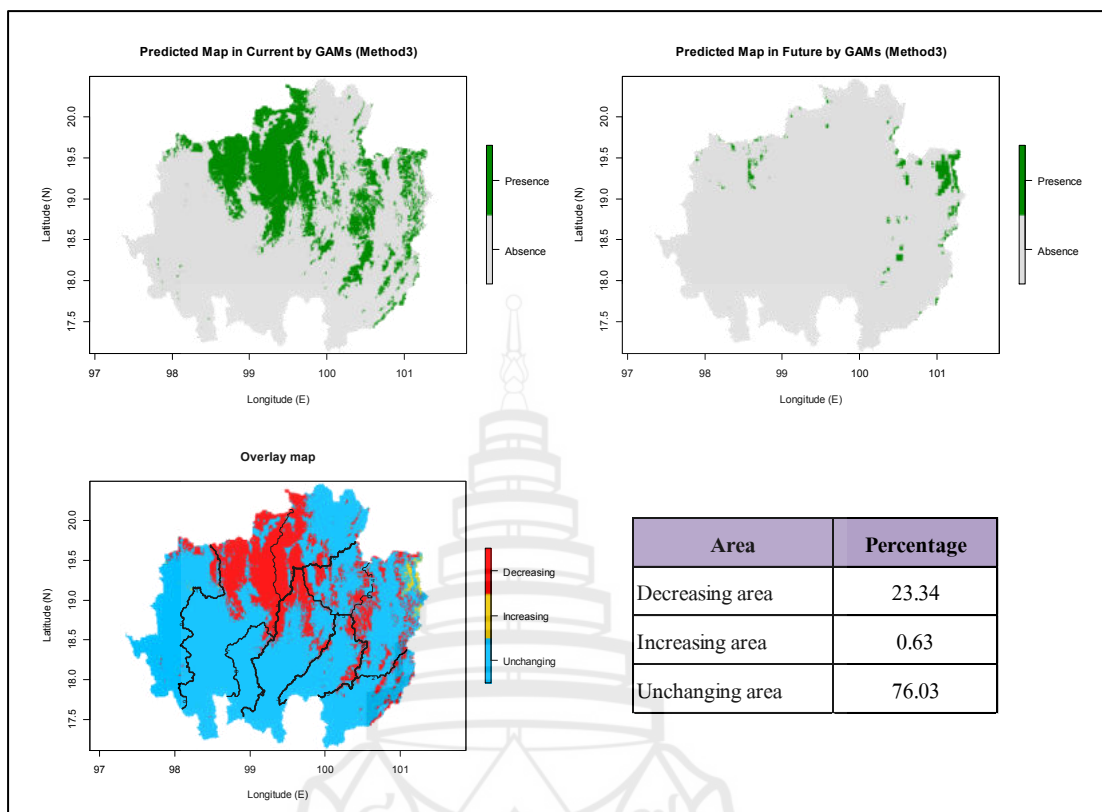
ภาพที่ 4.38 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1)

จากภาพที่ 4.38 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) พบว่าในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลง และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่าพื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 85.41 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 12.23 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 2.35 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าส่วนใหญ่จะกระจายอยู่ทางทิศตะวันตกของพื้นที่ซึ่งเป็นพื้นที่ของจังหวัดแม่ฮ่องสอนและเชียงใหม่ ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่ทั่วไปในพื้นที่



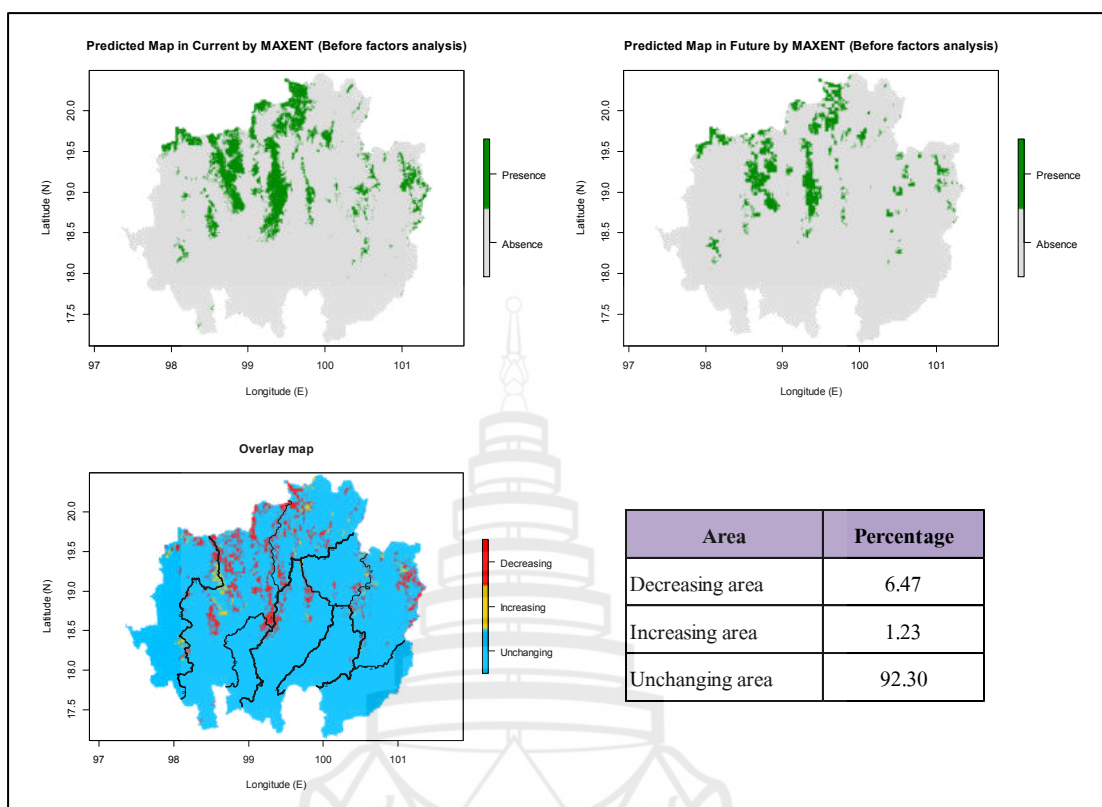
ภาพที่ 4.39 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2)

จากภาพที่ 4.39 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence (Method2) พบว่า ในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลง และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่า พื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 83.33 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 14.35 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 2.32 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่ทางทิศตะวันออกเฉียงใต้ของพื้นที่ ซึ่งเป็นพื้นที่ของจังหวัดน่านตอนล่างและอุดรดิตถ์ ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่ทั่วไปในพื้นที่



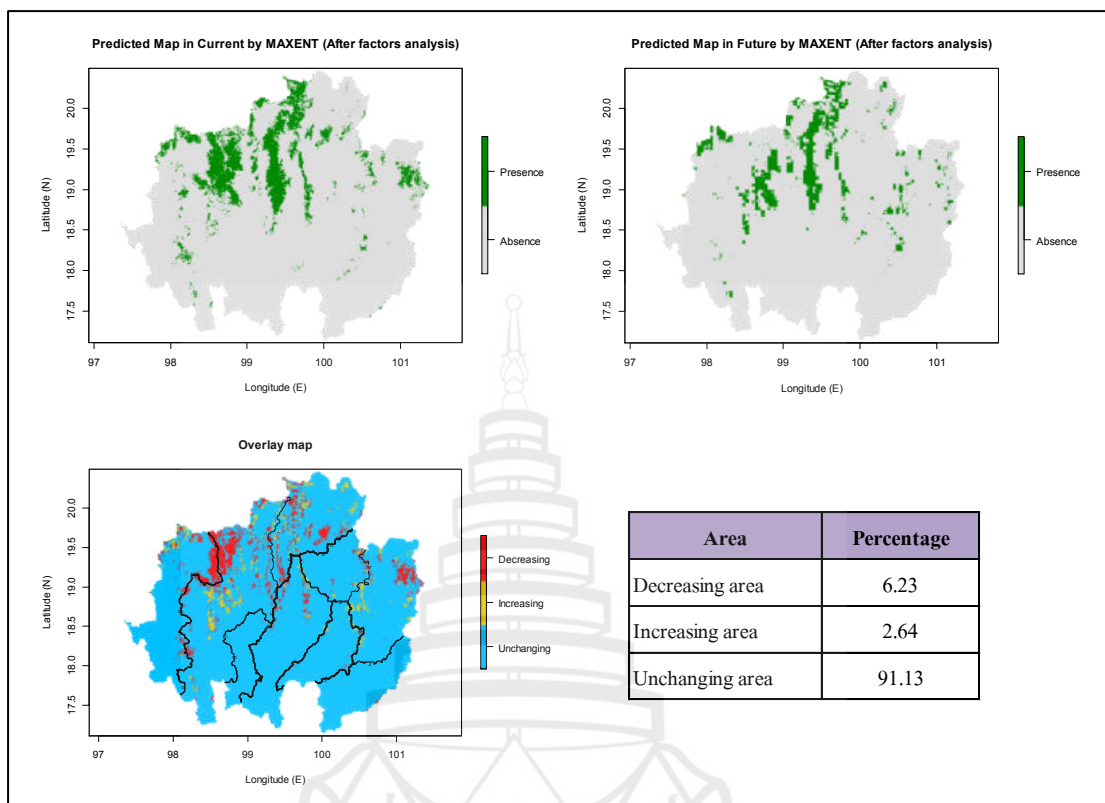
ภาพที่ 4.40 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3)

จากภาพที่ 4.40 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3) พบว่า ในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลงอย่างชัดเจนและเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่า พื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 76.03 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 23.34 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 0.63 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่ทางทิศตะวันออกของพื้นที่ ซึ่งเป็นพื้นที่ของจังหวัดน่าน ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่าจะกระจายอยู่ทั่วไปในพื้นที่



ภาพที่ 4.41 กระบวนการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง MAXENT ก่อนการวิเคราะห์องค์ประกอบ

จากภาพที่ 4.41 แสดงผลกระบวนการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง MAXENT พบว่าในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลง และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่า พื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 92.30 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 6.47 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 1.23 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าส่วนใหญ่จะกระจายอยู่ทั่วไปของพื้นที่ภาคเหนือตอนบน เช่นเดียวกับพื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่า



ภาพที่ 4.42 ผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง MAXENT หลังการวิเคราะห์องค์ประกอบ

จากภาพที่ 4.42 แสดงผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของชาป่าโดยแบบจำลอง MAXENT พบว่าในปี ค.ศ. 2050 พื้นที่ที่คาดว่าจะมีการกระจายของต้นชาป่า (พื้นที่สีเขียว) มีแนวโน้มลดลง และเมื่อนำแผนที่พยากรณ์ ณ เวลาปัจจุบันมาซ้อนทับกับแผนที่พยากรณ์ของปี ค.ศ. 2050 พบว่า พื้นที่ที่ไม่มีการเปลี่ยนแปลงมีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 91.13 ในขณะที่พื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 6.23 และพื้นที่ที่คาดว่าจะมีเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่ามีอัตราส่วนของพื้นที่คิดเป็นร้อยละ 2.64 ซึ่งจะสังเกตได้ว่า พื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของพื้นที่การกระจายของต้นชาป่าส่วนใหญ่จะกระจายอยู่ทั่วไปของพื้นที่ภาคเหนือตอนบน เช่นเดียวกับพื้นที่ที่คาดว่าจะมีลดลงของพื้นที่การกระจายของต้นชาป่า

บทที่ 5

สรุปผลการศึกษาและข้อเสนอแนะ

5.1 สรุปผลการศึกษาและอภิปรายผลการศึกษา

สรุปผลการศึกษาจากการสร้างแบบจำลองการกระจายของต้นชาป่าและสรุปผลการศึกษาปัจจัยทางสภาพทางภูมิศาสตร์และสภาพทางภูมิอากาศในพื้นที่ที่มีการค้นพบต้นชาป่าจำนวน 185 จุด ในเขตพื้นที่ภาคเหนือตอนบนของประเทศไทย รวมทั้งสรุปผลจากการศึกษาผลกระทบของแบบจำลองต่อการเปลี่ยนแปลงสภาพภูมิอากาศในปี ค.ศ. 2050 โดยเปรียบเทียบกับปัจจุบัน

5.1.1 การสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี GLMs

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs สร้างโดยใช้ข้อมูล Presence points และ Pseudo-Absence ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence เปรียบเทียบกัน 3 วิธี ซึ่งสามารถสรุปผลได้ดังนี้

5.1.1.1 การศึกษาปัจจัยที่ส่งผลกระทบต่อกระจายของต้นชาป่าจากแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี

จากการสร้างแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี ได้แก่วิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1), วิธีการสุ่มจุด Pseudo-Absence points ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence points แบบ 2 ขั้นตอน (Method3) พบว่าแบบจำลอง GLMs มีตัวแปรที่เข้าไปในตัวแบบที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธีมีความแตกต่างกัน โดยมีตัวแปรที่เข้าไปในตัวแบบทั้ง 3 วิธีได้แก่ตัวแปรความชื้นสัมพัทธ์ (Humidity), Isothermality (Bio3), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) และกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5, Class6) โดยทั้ง 3 วิธีมีตัวแปร Bio3, Bio8, Bio12, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวก ในขณะที่ตัวแปร Humidity และ Bio14 มีค่าสัมประสิทธิ์เป็นลบ ซึ่งมีเพียง 2 ตัวแปรคือตัวแปร Bio12 และ Bio14 ที่มีนัยสำคัญทางสถิติที่ 0.05 ลงไปทั้ง 3 วิธี แสดงว่าตัวแปร

Bio12 และ Bio14 มีความสำคัญอย่างยิ่งต่อตัวแบบ แต่ทั้งนี้ทั้งนั้นตัวแปรอื่น ๆ ก็มีความสำคัญต่อตัวแบบในแต่ละวิธีเช่นเดียวกัน

เมื่อพิจารณาตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8) ในพื้นที่ที่มีการค้นพบต้นชาป่า 185 จุดในเขตพื้นที่ภาคเหนือตอนบนพบว่า มีอุณหภูมิในไตรมาสที่ฝนตกชุกที่สุดโดยเฉลี่ย 23.77 องศาเซลเซียส (ตารางที่ 4.6) โดยมีผลเชิงบวกต่อแบบจำลอง ซึ่งหมายความว่าพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่ที่มีมากขึ้น เช่นเดียวกันกับตัวแปรปริมาณน้ำฝนรายปี (Bio12) ที่มีผลเชิงบวกเช่นเดียวกัน นั่นคือในพื้นที่ที่มีปริมาณน้ำฝนรายปีมากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่เพิ่มขึ้นด้วยเช่นกัน โดยในพื้นที่ที่มีการค้นพบต้นชาป่า 185 จุดในเขตพื้นที่ภาคเหนือตอนบนมีปริมาณน้ำฝนรายปีโดยเฉลี่ย 1273.56 มิลลิเมตรต่อปี (ตารางที่ 4.7)

ในขณะที่ตัวแปร ความชื้นสัมพัทธ์ (Humidity) และปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) มีค่าสัมประสิทธิ์เป็นลบ ซึ่งหมายความว่าตัวแปรดังกล่าวจะไปลดโอกาสของการค้นพบต้นชาป่าในพื้นที่ นั่นคือพื้นที่ใดมีความชื้นสัมพัทธ์ (Humidity) สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลง เช่นเดียวกับตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) หากพื้นที่ใดมีปริมาณน้ำฝนมากเกินไปโอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงเช่นกัน โดยในพื้นที่ที่มีการค้นพบต้นชาป่า 185 จุดในเขตพื้นที่ภาคเหนือตอนบนพบว่า มีความชื้นสัมพัทธ์โดยเฉลี่ย 74.94 เปอร์เซ็นต์ (ตารางที่ 4.5) และมีปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุดโดยเฉลี่ย 8.11 มิลลิเมตรต่อเดือน (ตารางที่ 4.7)

เมื่อศึกษาลักษณะภูมิอากาศโดยทั่วไปของประเทศไทยพบว่า ถ้าแบ่งตามระบบการจำแนกประเภทภูมิอากาศของเคิปปิน (Koppen's Classification) แล้ว ประเทศไทยจะจัดอยู่ในภูมิอากาศเขตร้อน (Tropical Climate) กล่าวคือมีฝนตกมากและมีอุณหภูมิสูงตลอดปี ซึ่งแบ่งเป็นเขตภูมิอากาศย่อยได้อีก 3 เขต ได้แก่ ภูมิอากาศแบบร้อนชื้นสลับแล้งหรือแบบทุ่งหญ้าสะวันนา (Tropical Wet-Dry Climate or Tropical Savannah Climate : Aw), ภูมิอากาศแบบมรสุมเขตร้อน (Tropical monsoon Climate : Am) และภูมิอากาศแบบป่าฝนเขตร้อน (Tropical Rainforest Climate : Af) โดยภาคเหนือของไทยมีลักษณะภูมิอากาศแบบร้อนชื้นสลับแล้งหรือแบบทุ่งหญ้าสะวันนา ซึ่งในเขตนี้จะมีฤดูฝนและฤดูแล้งสลับกันและจะแตกต่างกันอย่างเด่นชัด คือในฤดูมรสุมตะวันตกเฉียงใต้จะมีฝนตกตลอดฤดู แต่ในฤดูมรสุมตะวันออกเฉียงเหนือหรือฤดูหนาวอากาศจะแห้งแล้ง (โสภณ ชมชาญ, ม.ป.ป.) โดยพื้นที่ที่มีความเหมาะสมกับต้นชาป่าในช่วงที่ฝนตกชุก (ฤดูฝน) อุณหภูมิในบริเวณนั้นจะต้องมีความเหมาะสมกล่าวคือ ต้องมีอุณหภูมิสูงไม่ต่ำหรือหนาวเย็นจนเกินไป และมีปริมาณน้ำฝนมากพอในแต่ละปี ในขณะที่ช่วงเดือนที่แล้งที่สุดต้องไม่มีปริมาณ

น้ำฝนมากเกินไป ซึ่งสอดคล้องกับลักษณะภูมิอากาศของภาคเหนือของประเทศไทยแบบร้อนชื้น สลับแล้ง

เมื่อพิจารณาตัวแปรความชื้นสัมพัทธ์ (Humidity) พบว่าพื้นที่ใดมีความชื้นสัมพัทธ์ สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ที่ลดลง ซึ่งปกติแล้วประเทศไทยจะตั้งอยู่ในเขตร้อน ใกล้เส้นศูนย์สูตร จึงมีอากาศร้อนชื้นปกคลุมเกือบตลอดปี เว้นแต่บริเวณที่อยู่ลึกเข้าไปในแผ่นดิน ตั้งแต่ภาคกลางขึ้นไป ความชื้นสัมพัทธ์จะลดลงชัดเจนในช่วงฤดูหนาวเนื่องจากได้รับอิทธิพลจาก ลมมรสุมตะวันตกเฉียงเหนือ ทำให้ความชื้นสัมพัทธ์ต่ำและความชื้นสัมพัทธ์จะลดลงในฤดูร้อน เนื่องจากเป็นช่วงดวงอาทิตย์จะตั้งฉากกับภาคเหนือพอดี โดยเฉพาะฤดูร้อนจะเป็นช่วงที่ความชื้น สัมพัทธ์ลดลงต่ำสุดในรอบปี (ชลเทพ ทาตรี, 2552) แสดงให้เห็นว่าในช่วงฤดูหนาวและฤดูร้อน พื้นที่ตั้งแต่ภาคกลางขึ้นไปจนถึงทางเหนือสุดของประเทศไทยมีแนวโน้มความชื้นสัมพัทธ์ลดลง หมายความว่ายิ่งพื้นที่ที่อยู่เหนือขึ้นไปหรือลึกเข้าไปในแผ่นดินความชื้นสัมพัทธ์ก็จะลดลง โอกาส ของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วย แต่เนื่องจากพื้นที่ทางภาคเหนือของประเทศไทยส่วนใหญ่เป็นพื้นที่ที่ปกคลุมด้วยป่าเกือบร้อยละ 70 ของพื้นที่และเป็นพื้นที่ภูเขาสูง ซึ่งทำให้ความชื้น สัมพัทธ์ในพื้นที่สูงด้วย โดยเฉพาะบริเวณพื้นที่ป่าดิบเขาซึ่งเป็นป่าที่อยู่สูงจากระดับน้ำทะเลตั้งแต่ 1,000 เมตรขึ้นไปทำให้อากาศหนาวเย็นและมีความชื้นสูงตลอดปีเนื่องจากมีเมฆเข้ามาปกคลุมอยู่ เสมอ ดังนั้นโอกาสในการค้นพบต้นชาป่าในบริเวณป่าดิบเขาก็น่าจะลดลง

เมื่อพิจารณาตัวแปร Isothermality (Bio3) พบว่ามีผลต่อแบบจำลองเชิงบวก โดย Isothermality ถูกคำนวณมาจากสัดส่วนระหว่างความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อความแตกต่างของอุณหภูมิรายปี (Bio7) โดยความแตกต่างของอุณหภูมิรายปีได้มาจากผลต่าง ระหว่างอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุดลบกับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุดหรือ ความแตกต่างระหว่างอุณหภูมิในฤดูร้อนกับฤดูหนาว ซึ่งโดยปกติแล้วในเขตพื้นที่ภาคเหนือของ ประเทศไทยจะมีความแตกต่างระหว่างอุณหภูมิในฤดูร้อนกับฤดูหนาวสูงอยู่แล้ว เนื่องจากอยู่ใกล้ ทะเล จากตารางที่ 4.7 ในบริเวณที่มีการค้นพบต้นชาป่าพบว่า มีความแตกต่างของอุณหภูมิในฤดู ร้อนกับฤดูหนาวโดยเฉลี่ย 23.19 องศาเซลเซียสและมีส่วนเบี่ยงเบนมาตรฐานที่ต่ำ แสดงว่าความ แตกต่างของอุณหภูมิในฤดูร้อนกับฤดูหนาวในแต่ละบริเวณที่มีการค้นพบต้นชาป่าไม่ค่อยมีความ แตกต่างกันหรือค่อนข้างคงที่ เช่นเดียวกับความแตกต่างของอุณหภูมิรายวัน

จากตาราง 4.7 ในบริเวณที่มีการค้นพบต้นชาป่าพบว่า มีความแตกต่างของอุณหภูมิ รายวันโดยเฉลี่ย 12.19 องศาเซลเซียสและมีส่วนเบี่ยงเบนมาตรฐานที่ต่ำ แสดงว่าความแตกต่างของ อุณหภูมิรายวันในแต่ละบริเวณที่มีการค้นพบต้นชาป่าไม่ค่อยมีความแตกต่างกัน ซึ่งปัจจัยที่ส่งผล ต่อช่วงความแตกต่างของอุณหภูมิรายวันคือ ความชื้นจากน้ำ (สำนักประชาสัมพันธ์เขต 2 กรม

ประชาสัมพันธุ์, 2554) หากความชื้นจากน้ำมีมากหรืออยู่ใกล้แหล่งน้ำ จะส่งผลให้ช่วงความแตกต่างของอุณหภูมิรายวันลดลง เมื่อช่วงความแตกต่างของอุณหภูมิรายวันลดลง ก็ส่งผลให้ค่า Isothermality ลดลงด้วย โอกาสในการค้นพบต้นชาป่าในพื้นที่ที่จะลดลงด้วย ซึ่งสอดคล้องกับค่าความชื้นสัมพัทธ์ ถ้าความชื้นจากน้ำมีมากก็จะทำให้ความชื้นสัมพัทธ์มีค่าสูงด้วย ส่งผลให้โอกาสในการค้นพบต้นชาป่าในพื้นที่ที่จะลดลงเช่นกัน สรุปได้ว่าพื้นที่ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าต้องเป็นพื้นที่ที่มีความชื้นสัมพัทธ์ไม่สูงเกินไป มีช่วงพิสัยของอุณหภูมิรายวันและช่วงพิสัยอุณหภูมิในฤดูร้อนกับฤดูหนาวที่ค่อนข้างแตกต่างกันอย่างชัดเจน

เมื่อพิจารณาตัวแปรการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) พบว่ามีผลเชิงบวก หมายความว่าโครงสร้างทางธรณีวิทยาของชนิดหินในพื้นที่ไม่มีผลต่อโอกาสในการค้นพบต้นชาป่าในพื้นที่ แต่ความลาดชันกลับมีผลต่อโอกาสในการค้นพบต้นชาป่า โดยต้องมีความลาดชันไม่เกิน 30 องศา ซึ่งสอดคล้องกับผลการศึกษาของบราเบอร์ ซาโต และ ลี (Braber, Sato & Lee, 2011) ที่อธิบายความลาดชันที่มีผลกระทบต่อต้นชาว่า ต้องมีความชันไม่เกิน 30 องศา และต้องมีการจัดการเป็นพิเศษเพื่อป้องกันการชะพังทลายของหน้าดินโดยเฉพาะอย่างยิ่งต้นชาที่ยังเพิ่งเจริญเติบโตและรากยังไม่แข็งแรง

5.1.1.2 การศึกษาเปรียบเทียบแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี

เมื่อเปรียบเทียบแผนที่พยากรณ์จากแบบจำลอง GLMs ที่ได้ในแต่ละวิธีของการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธีพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า ไม่ค่อยมีความแตกต่างกันมากนัก ซึ่งกระจายอยู่ทั่วไปตามพื้นที่ป่าในเขตพื้นที่ภาคเหนือตอนบน โดยส่วนใหญ่จะอยู่อย่างหนาแน่นทางทิศตะวันตกเฉียงเหนือของพื้นที่ภาคเหนือตอนบน ซึ่งได้แก่บริเวณจังหวัดเชียงใหม่และแม่ฮ่องสอนตอนบน

ตารางที่ 5.1 ค่าเฉลี่ยของตัวแปรที่มีผลต่อแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธีโดยจำแนกตามจังหวัดที่มีการค้นพบต้นชาป่า

ตัวแปร	จังหวัด					
	เชียงราย	เชียงใหม่	แม่ฮ่องสอน	ลำปาง	น่าน	แพร่
Humidity (%)	75.603	72.448	74.203	73.069	80.398	75.366
Bio3 (%)	52.623	51.532	50.333	52.000	53.862	53.000
Bio8 ($^{\circ}\text{C} \times 0.01$)	243.000	230.968	232.381	238.071	241.621	259.000
Bio12 (mm per year)	1492.868	1147.823	1109.048	1190.143	1325.345	1155.833
Bio14 (mm per month)	8.755	8.145	5.476	10.000	8.483	5.167
จำนวนจุดที่ค้นพบ	53	62	21	14	29	6

จากตารางที่ 5.1 แสดงค่าเฉลี่ยของตัวแปรที่มีผลต่อแบบจำลอง GLMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี โดยจำแนกตามจังหวัดที่มีการค้นพบต้นชาป่าพบว่า จากจำนวนจุดที่มีการค้นพบต้นชาป่าจำนวน 185 จุด ส่วนใหญ่มีการค้นพบในจังหวัดเชียงใหม่ จำนวน 62 จุด รองลงมาจังหวัดเชียงราย, น่าน, แม่ฮ่องสอน, ลำปางและแพร่ ตามลำดับ

จากตัวแปรที่มีผลต่อแบบจำลองเชิงบวกได้แก่ Bio3, Bio8 และ Bio12 พบว่าจังหวัดที่มีค่าเฉลี่ย Bio3 โดยเรียงจากมากไปน้อยคือจังหวัดน่าน, แพร่, เชียงราย, ลำปาง, เชียงใหม่ และแม่ฮ่องสอน ตามลำดับ จังหวัดที่มีค่าเฉลี่ย Bio8 โดยเรียงจากมากไปน้อยคือจังหวัดแพร่, เชียงราย, น่าน, ลำปาง, แม่ฮ่องสอน และเชียงใหม่ตามลำดับ จังหวัดที่มีค่าเฉลี่ย Bio12 โดยเรียงจากมากไปน้อยคือจังหวัดเชียงราย, น่าน, แพร่, ลำปาง, เชียงใหม่ และแม่ฮ่องสอน ตามลำดับ จะสังเกตได้ว่าเชียงใหม่และแม่ฮ่องสอนจะมีค่าเฉลี่ยอยู่ในลำดับท้าย ๆ เมื่อเรียงค่าเฉลี่ยจากมากไปหาน้อย ในขณะที่แผนที่พยากรณ์จากแบบจำลองส่วนใหญ่มีพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าหนาแน่นในจังหวัดเชียงใหม่และแม่ฮ่องสอน ซึ่งขัดแย้งกันกับตัวแปร Bio3, Bio8 และ Bio12 ซึ่งเป็นตัวแปรที่มีผลต่อแบบจำลองเชิงบวก หมายความว่ายิ่งตัวแปรเหล่านี้มีค่ามาก โอกาสในการค้นพบต้นชาป่าในพื้นที่ก็จะมีโอกาสสูงตามไปด้วย แต่ทั้งนี้ทั้งนั้นต้องพิจารณาตัวแปรอื่น ๆ ที่มีผลต่อแบบจำลอง

ตัวแปรที่มีผลต่อแบบจำลองเชิงลบได้แก่ Humidity และ Bio14 พบว่าจังหวัดที่มีค่าเฉลี่ย Humidity โดยเรียงจากน้อยไปมากคือจังหวัดเชียงใหม่, ลำปาง, แม่ฮ่องสอน, แพร่, เชียงราย และน่านตามลำดับ จังหวัดที่มีค่าเฉลี่ย Bio14 โดยเรียงจากน้อยไปมากคือจังหวัดแพร่, แม่ฮ่องสอน, เชียงใหม่, น่าน, เชียงราย และลำปางตามลำดับ จะสังเกตได้ว่าเชียงใหม่และแม่ฮ่องสอนจะมีค่าเฉลี่ยอยู่ในลำดับต้น ๆ เมื่อเรียงค่าเฉลี่ยจากน้อยไปหามาก ซึ่งสอดคล้องกับแผนที่พยากรณ์จากแบบจำลองที่ส่วนใหญ่มีพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าหนาแน่นในจังหวัดเชียงใหม่และแม่ฮ่องสอน โดยตัวแปร Humidity และ Bio14 เป็นตัวแปรที่มีผลต่อแบบจำลองเชิงลบหมายความว่า ยิ่งตัวแปรเหล่านี้มีค่าน้อย โอกาสในการค้นพบต้นชาป่าในพื้นที่ที่จะมีโอกาสสูง

ถึงแม้ว่าค่าเฉลี่ยของ Bio3, Bio8 และ Bio12 ของจังหวัดเชียงใหม่และแม่ฮ่องสอนมีค่าน้อยซึ่งทำให้โอกาสในการค้นพบต้นชาป่าในพื้นที่ลดลงแต่ค่าเฉลี่ยของ Humidity และ Bio14 กลับทำให้โอกาสในการค้นพบต้นชาป่าในพื้นที่เพิ่มขึ้น แต่อย่างไรก็ตามตัวแปรเหล่านี้เป็นเพียงตัวแปรทางด้านภูมิอากาศ ซึ่งเป็นตัวแปรเพียงส่วนหนึ่งที่มีผลต่อแบบจำลอง GLMs ในแต่ละวิธีของการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี โดยยังไม่ได้พิจารณาตัวแปรทางด้านภูมิศาสตร์ที่มีผลต่อแบบจำลอง GLMs ในแต่ละวิธีของการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี อันได้แก่ตัวแปรของการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6)

ตารางที่ 5.2 แสดงร้อยละของจุดพิกัดของการค้นพบต้นชาป่าตามการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน โดยจำแนกตามจังหวัดที่มีการค้นพบต้นชาป่า

ตัวแปร	จังหวัด					
	เชียงราย	เชียงใหม่	แม่ฮ่องสอน	ลำปาง	น่าน	แพร่
Class1	39.62	32.26	71.43	64.29	62.07	16.67
Class2	11.32	8.06	14.29		37.93	83.33
Class5	39.62	48.39	14.29	35.71		
Class6	9.43	11.29				
รวม	100.00	100.00	100.00	100.00	100.00	100.00

จากตารางที่ 5.2 แสดงร้อยละของจุดพิคซ์ของการค้นพบต้นชาป่าตามการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน โดยจำแนกตามจังหวัดที่มีการค้นพบต้นชาป่าพบว่า ในจังหวัดเชียงรายและเชียงใหม่มีการค้นพบต้นชาป่าในพื้นที่ที่มีความลาดชันไม่เกิน 30 องศา โดยพบว่าส่วนใหญ่อยู่ในพื้นที่ที่มีความลาดชันไม่เกิน 15 องศา (Class1 และ Class5) ในขณะที่จังหวัดแม่ฮ่องสอนส่วนใหญ่มีการค้นพบต้นชาป่าในพื้นที่ที่มีโครงสร้างทางธรณีวิทยาของชนิดหินเป็นหินตะกอนและหินแปร โดยมีความลาดชันไม่เกิน 30 องศา (Class1 และ Class2) และบางส่วนมีการค้นพบต้นชาป่าในพื้นที่ที่มีโครงสร้างทางธรณีวิทยาของชนิดหินเป็นหินอัคนีโดยมีความลาดชันไม่เกิน 15 องศา (Class3) ส่วนจังหวัดลำปางมีการค้นพบต้นชาป่าในพื้นที่ที่มีความลาดชันไม่เกิน 15 องศา (Class1 และ Class5) และสังเกตได้ว่าในจังหวัดน่านและแพร่ทั้งหมดมีการค้นพบต้นชาป่าในพื้นที่ที่มีโครงสร้างทางธรณีวิทยาของชนิดหินเป็นหินตะกอนและหินแปร โดยมีความลาดชันไม่เกิน 30 องศา (Class1 และ Class2) โดยในจังหวัดน่านส่วนใหญ่พบในพื้นที่ที่มีความลาดระหว่าง 0-15 องศา ในขณะที่จังหวัดแพร่โดยส่วนใหญ่พบในพื้นที่ที่มีความลาดระหว่าง 15-30 องศา

จะสังเกตได้ว่าในจังหวัดเชียงราย, เชียงใหม่ และแม่ฮ่องสอน ปัจจัยทางด้านภูมิศาสตร์มีผลต่อการกระจายของต้นชาป่า ซึ่งได้แก่ปัจจัยโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน ส่งผลให้พื้นที่ในจังหวัดดังกล่าว มีโอกาสที่จะค้นพบต้นชาป่าได้มากกว่าจังหวัดอื่น

เมื่อนำแผนที่พยากรณ์จากแบบจำลอง GLMs ที่ได้ในแต่ละวิธีของการสุ่มเลือกจุด Pseudo-Absence มาเปรียบเทียบ โดยใช้วิธี Cochran's Q Test พบว่า เมื่อมองเปรียบเทียบด้วยตาเปล่าแล้ว โดยภาพรวมแผนที่พยากรณ์ไม่มีความแตกต่างกันมากนัก แต่ถ้าหากมองในส่วนที่ย่อยลงไปในพื้นที่ในระดับพิกเซลและทำการทดสอบความแตกต่าง กลับพบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) จะมีความแตกต่างจากวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence points แบบ 2 ขั้นตอน (Method3) ในขณะที่วิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence points แบบ 2 ขั้นตอน (Method3) ไม่มีความแตกต่างกัน

จากการพิจารณาความสามารถในการพยากรณ์ นั่นคือ ค่า R^2 พบว่าแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีค่า R^2 มากที่สุด หมายความว่าแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีความสามารถในการพยากรณ์ของตัวแบบมากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1)

ตามลำดับ ซึ่งสอดคล้องกับการศึกษาของ Wisz & Guisan, (2009) ที่ได้ศึกษาผลกระทบของวิธีการคัดเลือกจุด Pseudo-Absence ต่อแบบจำลองการกระจายของสิ่งมีชีวิต โดยกล่าวไว้ว่า แบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอนมีความสามารถในการพยากรณ์ของตัวแบบได้ดีกว่าการเลือกจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

จากการพิจารณาค่าความถูกต้องในการพยากรณ์ นั่นคือ ค่า AUC พบว่าในทุกวิธีการของการคัดเลือกจุด Pseudo-Absence นั้น การพิจารณาเฉพาะตัวแปรทางด้านภูมิศาสตร์จะให้ค่าแสดงความถูกต้องของการพยากรณ์น้อยที่สุด ในขณะที่การพิจารณาเฉพาะตัวแปรทางด้านภูมิอากาศและการพิจารณาของทุกตัวแปรจะให้ค่าแสดงความถูกต้องของการพยากรณ์ใกล้เคียงกันหรือต่างกันเพียงเล็กน้อยเท่านั้น และเมื่อเปรียบเทียบระหว่างวิธีการคัดเลือกจุด Pseudo-Absence พบว่าแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3) มีความความถูกต้องในการพยากรณ์มากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ

5.1.2 การสร้างแบบจำลองการกระจายของต้นขาป่าโดยวิธี GAMs

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs เป็นวิธีการสร้างแบบจำลองโดยใช้เทคนิคการจำแนกกลุ่มเช่นเดียวกับแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs แต่ต่างกันตรงที่ฟังก์ชันการเชื่อมโยง (Link Functions) โดยในวิธีการของ GLMs ฟังก์ชันการเชื่อมโยง จะอยู่ในรูปแบบเชิงเส้น (Linear Model) ในขณะที่ฟังก์ชันการเชื่อมโยงของ GAMs จะไม่อยู่ในรูปแบบเชิงเส้น แต่จะอยู่รูปแบบของ Additive Model ที่ประกอบไปด้วย Smooth Functions ของตัวแปรที่นำมาศึกษา และเนื่องจากแบบจำลอง GAMs มักจะเกิดปัญหาการเข้ากันเกินไป (Overfitting Problem) ดังนั้นการจำกัดจำนวนตัวแปรในการสร้างแบบจำลองจึงเป็นวิธีการหนึ่งที่ช่วยลดปัญหาดังกล่าว เพราะฉะนั้นในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GAMs ที่เกิดจากการสุ่มเลือกจุด Pseudo-Absence ทั้ง 3 วิธี จะใช้เฉพาะตัวแปรที่มีนัยสำคัญทางสถิติที่ 0.05 ลงไปมาสร้างแบบจำลอง

จากการพิจารณาความสามารถในการพยากรณ์ นั่นคือค่า R^2 พบว่าแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีค่า R^2 มากที่สุด หมายความว่าแบบจำลองมีความสามารถในการพยากรณ์ของตัวแบบมากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ ในขณะที่การพิจารณาค่าความถูกต้องในการพยากรณ์ นั่นคือค่า AUC กลับพบว่าแบบจำลองที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มี

ความความถูกต้องในการพยากรณ์มากที่สุด รองลงมาเป็นวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) ตามลำดับ แสดงว่าแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอนมีความความถูกต้องในการพยากรณ์มากที่สุดถึงแม้ว่าความสามารถในการพยากรณ์จะเป็นรองแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence ซึ่งทั้งนี้อาจเกิดจากผลกระทบของจำนวนตัวแปรที่นำไปสร้างแบบจำลอง เนื่องจากแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence ใช้จำนวนตัวแปรในการสร้างแบบจำลองมากกว่าแบบจำลอง GAMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน ดังนั้นจึงมีโอกาสสูงที่จะทำให้เกิดปัญหาการเข้ากันเกินไป (Overfitting Problem) นั่นคือส่งผลให้ค่า R^2 สูงเกินกว่าความเป็นจริง

5.1.3 การสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี MAXENT

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี MAXENT เป็นวิธีการสร้างแบบจำลองที่อยู่บนพื้นฐานของ Machine-Learning โดยใช้ข้อมูลค้นพบ (Presence Data) ร่วมกับข้อมูลลักษณะสิ่งแวดล้อมในพื้นที่ (Background Data) ในการสร้างแบบจำลอง ซึ่งแตกต่างจากวิธี GLMs และ GAMs ที่ใช้ข้อมูลทั้งจุด Presence และจุด Absence หรือ Pseudo-Absence มาใช้ในการสร้างแบบจำลอง

5.1.3.1 แบบจำลองโดยวิธี MAXENT ก่อนการวิเคราะห์องค์ประกอบ (Before FA)

จากการสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี MAXENT ก่อนการวิเคราะห์องค์ประกอบพบว่า มีตัวแปรที่มีค่า Variable Contribution เกินร้อยละ 5 อยู่ 4 ตัวแปรด้วยกันคือ ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11) ซึ่งมีค่า Variable Contribution มากที่สุดถึงประมาณร้อยละ 33 แสดงว่าอุณหภูมิเฉลี่ยในช่วงฤดูหนาวเป็นปัจจัยที่มีความสำคัญอย่างยิ่งต่อการกระจายของต้นชาป่า รองลงมาได้แก่ตัวแปรความชื้นสัมพัทธ์ (Humidity), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5) และปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13) โดยมีค่า Variable Contribution ประมาณร้อยละ 11, 9 และ 6 ตามลำดับ จะสังเกตได้ว่าอุณหภูมิในฤดูหนาว อุณหภูมิในฤดูร้อน ปริมาณน้ำฝนในฤดูฝนและความชื้นสัมพัทธ์ต่างก็เป็นปัจจัยหลักที่สำคัญ หมายความว่าในพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าควรมีช่วงระยะเวลาของฤดูหนาวนั้นมีอุณหภูมิที่ต่ำ ในขณะที่ช่วงฤดูร้อนอุณหภูมิที่สูงที่สุดในช่วงฤดูร้อนก็ไม่ควรยาวนานเกินไป สังเกตได้จากค่า Variable Contribution ของตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) ที่มีค่าเพียงร้อยละ 0.6 เท่านั้น เมื่อพิจารณาปัจจัยที่เกี่ยวข้องกับปริมาณน้ำฝนพบว่าพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าควรมีปริมาณน้ำฝนที่เพียงพอแต่ก็ไม่ควรมีปริมาณน้ำฝนมากเกินไปเป็นระยะเวลานาน ๆ สังเกตได้

จากค่า Variable Contribution ของตัวแปรปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) ที่มีค่า Variable Contribution น้อยที่สุด ส่วนตัวแปรความชื้นสัมพัทธ์เป็นอีกปัจจัยที่มีความสำคัญเนื่องจากพื้นที่ที่มีการค้นพบต้นชาป่าส่วนใหญ่พบในพื้นที่ป่า ซึ่งพื้นที่ป่าเป็นพื้นที่ที่มีความชื้นที่สูงเนื่องจากถูกปกคลุมไปด้วยป่าไม้ แต่ความชื้นสัมพัทธ์ที่สูงมาก ๆ ส่วนใหญ่เป็นพื้นที่ในป่าดิบเขาที่มีความสูงจากระดับน้ำทะเลตั้งแต่ 1,000 เมตรขึ้นไปซึ่งจะมีอุณหภูมิต่ำเกือบตลอดทั้งปีซึ่งไม่สอดคล้องกับพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า เนื่องจากพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่านั้นควรมีช่วงระยะเวลาของฤดูกาลหรือมีความผันแปรของอุณหภูมิ สังเกตได้จากค่า Variable Contribution ของตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4) ที่มีค่า Variable Contribution ประมาณร้อยละ 3.4 ซึ่งอยู่ในลำดับที่ 9 จากทั้งหมด 24 ตัวแปร

5.1.3.2 การวิเคราะห์องค์ประกอบ (Factors Analysis; FA)

เนื่องจากตัวแปรที่นำมาศึกษามีจำนวนมาก เพื่อเป็นการลดตัวแปรที่มีความสัมพันธ์ลงจึงทำการรวมกลุ่มตัวแปรที่ค่าความสัมพันธ์กันให้อยู่ในกลุ่มเดียวกันและสร้างปัจจัยใหม่ขึ้นมา โดยใช้เทคนิควิธีการวิเคราะห์องค์ประกอบ (Factors Analysis)

จากการวิเคราะห์องค์ประกอบ (Factors Analysis) พบว่า สามารถจัดให้ตัวแปรอยู่ในกลุ่มเดียวกัน ได้ 7 กลุ่มปัจจัย ดังนี้

กลุ่มปัจจัยที่ 1 กลุ่มปัจจัยของอุณหภูมิที่มีความสัมพันธ์กับความสูงจากระดับน้ำทะเล จะประกอบด้วยตัวแปรความสูงจากระดับน้ำทะเล (DEM), อุณหภูมิเฉลี่ยรายปี (Bio1), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) และตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11)

กลุ่มปัจจัยที่ 2 กลุ่มปัจจัยของปริมาณน้ำฝนที่ขึ้นอยู่กับฤดูกาล จะประกอบด้วยตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ช่วงฤดูกาลของปริมาณน้ำฝน (Bio15), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และตัวแปรปริมาณน้ำฝนในไตรมาสที่ร้อนที่สุด (Bio18)

กลุ่มปัจจัยที่ 3 กลุ่มปัจจัยของปริมาณน้ำฝนในช่วงที่แห้งแล้ง จะประกอบด้วยตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และตัวแปรปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19)

กลุ่มปัจจัยที่ 4 กลุ่มปัจจัยของความผันแปรของอุณหภูมิรายวันและรายปี จะประกอบด้วยตัวแปรช่วงอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) และตัวแปรช่วงของอุณหภูมิรายปี (Bio7)

กลุ่มปัจจัยที่ 5 กลุ่มปัจจัยอุณหภูมิที่ขึ้นอยู่กับความชื้นสัมพัทธ์ จะประกอบด้วยตัวแปรความชื้นสัมพัทธ์ (Humidity) และตัวแปร Isothermality (Bio3)

กลุ่มปัจจัยที่ 6 ปัจจัยทิศด้าน (Aspect)

กลุ่มปัจจัยที่ 7 ปัจจัยระยะห่างจากแม่น้ำ (Distance)

5.1.3.3 แบบจำลองโดยวิธี MAXENT หลังการวิเคราะห์องค์ประกอบ (After FA)

จากการสร้างแบบจำลองการกระจายของต้นชาป่าโดยวิธี MAXENT หลังการวิเคราะห์องค์ประกอบพบว่า มีกลุ่มปัจจัยที่มีค่า Variable Contribution เกินร้อยละ 5 อยู่ 3 กลุ่มปัจจัยด้วยกัน คือ กลุ่มปัจจัยที่ 1 เป็นกลุ่มปัจจัยของอุณหภูมิที่มีความสัมพันธ์กับความสูงจากระดับน้ำทะเล โดยมีค่า Variable Contribution มากที่สุดซึ่งประกอบไปด้วยตัวแปรความสูงจากระดับน้ำทะเล (DEM), อุณหภูมิเฉลี่ยรายปี (Bio1), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), อุณหภูมิต่ำสุดในเดือนที่หนาวที่สุด (Bio6), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10) และตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11)

รองลงมาได้แก่ กลุ่มปัจจัยที่ 3 เป็นกลุ่มปัจจัยของปริมาณน้ำฝนในช่วงที่แห้งแล้ง ซึ่งประกอบไปด้วยตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และตัวแปรปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19) แสดงว่าปริมาณน้ำฝนในช่วงฤดูแล้งมีความสำคัญต่อการกระจายของต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน

สุดท้ายได้แก่ กลุ่มปัจจัยที่ 2 เป็นกลุ่มปัจจัยของปริมาณน้ำฝนที่ขึ้นอยู่กับฤดูกาล ซึ่งประกอบไปด้วยตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ช่วงฤดูกาลของปริมาณน้ำฝน (Bio15), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และตัวแปรปริมาณน้ำฝนในไตรมาสที่ร้อนที่สุด (Bio18)

5.1.4 การเปรียบเทียบแบบจำลอง GLMs, GAMs และ MAXENT

แบบจำลองการกระจายของสิ่งมีชีวิตโดยวิธี GLMs, GAMs และ MAXENT เป็นวิธีการสร้างแบบจำลองโดยเทคนิคจำแนกกลุ่ม ซึ่งวิธี GLMs และ GAMs เป็นเทคนิคที่ใช้การสร้างแบบจำลองทางสถิติ (Statistical Models) ในขณะที่วิธี MAXENT เป็นวิธีการที่อยู่บนพื้นฐานของ Machine-Learning โดยใช้ข้อมูลค้นพบ (Presence Data) ร่วมกับข้อมูลลักษณะสิ่งแวดล้อมในพื้นที่ (Background Data)

จากการเปรียบเทียบแผนที่พยากรณ์ของแบบจำลอง GLMs, GAMs และ MAXENT พบว่าแผนที่พยากรณ์ของแบบจำลอง MAXENT มีความแตกต่างจากแผนที่พยากรณ์ของแบบจำลอง GLMs และ GAMs อย่างเห็นได้ชัดเจน ในขณะที่แผนที่พยากรณ์ของแบบจำลอง GLMs และ

GAMs ไม่ค่อยมีความแตกต่างกัน ซึ่งเมื่อพิจารณาค่า AUC ซึ่งเป็นค่าแสดงความถูกต้องของการพยากรณ์พบว่า การสร้างแบบจำลอง MAXENT จะให้ค่าแสดงความถูกต้องของการพยากรณ์มากกว่าแบบจำลอง GLMs แต่น้อยกว่าแบบจำลอง GAMs ซึ่งสอดคล้องกับผลการศึกษาของ Gibsan, Barrett and Burbidge (2007) ที่เปรียบเทียบแบบจำลองที่มาจาก GLMs และ MAXENT ซึ่งแบบจำลอง MAXENT มีค่า AUC ที่ค่อนข้างสูง (มากกว่า 0.8) และมีค่า AUC สูงกว่าแบบจำลอง GLMs ในขณะที่แบบจำลอง GAMs ให้ค่า AUC มากที่สุด แต่แบบจำลอง GAMs มักจะเกิดปัญหาการเข้ากันเกินไป (Overfitting problem) ซึ่งมีสาเหตุหนึ่งจากจำนวนตัวแปรในการสร้างแบบจำลองมีเป็นจำนวนมาก ดังนั้นในการสร้างแบบจำลอง GAMs ควรมีการจำกัดตัวแปรโดยเลือกเฉพาะตัวแปรที่คาดว่าจะมีความสำคัญจริง ๆ มาสร้างแบบจำลอง

5.1.5 ผลกระทบจากการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นชาป่า

การศึกษาผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศต่อการกระจายของต้นชาป่าเป็นการศึกษาสภาพการกระจายของต้นชาป่าต่อการเปลี่ยนแปลงของสภาพภูมิอากาศที่เกิดขึ้น โดยเปรียบเทียบระหว่างสภาพภูมิอากาศในปีปัจจุบันและปี ค.ศ. 2050 ซึ่งใช้ตัวแปร Bioclimatic Variables ทั้งหมด 19 ตัวแปร ในปี ค.ศ. 2050 จากผลการศึกษาพบว่า พื้นที่ส่วนใหญ่ไม่มีการเปลี่ยนแปลงมากนัก โดยเฉลี่ยอยู่ที่ประมาณร้อยละ 80-90 ของพื้นที่ทั้งหมด ส่วนพื้นที่ที่มีการเปลี่ยนแปลงพบว่าการกระจายทั่วไปในพื้นที่ ซึ่งส่วนใหญ่มีแนวโน้มลดลง โดยพื้นที่ที่คาดว่าจะมีการลดลงของการกระจายของต้นชาป่า โดยเฉลี่ยอยู่ที่ประมาณร้อยละ 6-10 เท่านั้น มีเพียงแผนที่พยากรณ์ที่เกิดจากแบบจำลอง GAMs ที่มีพื้นที่ที่คาดว่าจะมีการลดลงของการกระจายของต้นชาป่าเฉลี่ยร้อยละ 17 โดยเฉพาะแบบจำลอง GAMs ที่ได้มาจากวิธีการสุ่มเลือกจุด Pseudo-Absence แบบ 2 ขั้นตอน ที่มีพื้นที่ที่คาดว่าจะมีการลดลงของการกระจายของต้นชาป่าถึงร้อยละ 23 ของพื้นที่ทั้งหมด ส่วนพื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของการกระจายของต้นชาป่ามีเพียงส่วนน้อยเท่านั้น มีเพียงแผนที่พยากรณ์ที่เกิดจากแบบจำลอง GLMs ที่ได้มาจากวิธีการสุ่มเลือกจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) ที่มีพื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของการกระจายของต้นชาป่ามากถึงร้อยละ 9.61 ซึ่งพื้นที่ที่คาดว่าจะมีการเพิ่มขึ้นของการกระจายของต้นชาป่ามีการย้ายการกระจายมาอยู่ทางทิศตะวันออก ทิศตะวันออกเฉียงใต้และทิศใต้ของพื้นที่ภาคเหนือตอนบน แทน ซึ่งเป็นพื้นที่ของจังหวัดน่าน, แพร่ และอุตรดิตถ์ โดยมีบางส่วนกระจายอยู่ในจังหวัดพะเยา และลำปาง จากที่ส่วนใหญ่มีการกระจายอยู่อย่างหนาแน่นอยู่ทางทิศเหนือและทิศตะวันออกเฉียงเหนือ ซึ่งส่วนใหญ่เป็นพื้นที่ของจังหวัดเชียงรายและเชียงใหม่

จากตารางที่ 5.3 แสดงร้อยละของพื้นที่พยากรณ์ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่าและการเปลี่ยนแปลงของพื้นที่พยากรณ์ระหว่างเวลาปัจจุบันและอนาคต อันสืบเนื่องมาจากการเปลี่ยนแปลงสภาพภูมิอากาศ พบว่าอัตราส่วนการเปลี่ยนแปลงของพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่าเมื่อเทียบกับปีปัจจุบันพบว่า แบบจำลอง GAMs มีอัตราส่วนการเปลี่ยนแปลงที่สูงมากเมื่อเทียบกับแบบจำลองอื่น ๆ โดยเฉพาะแบบจำลอง GAMs ที่ได้มาจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) และแบบจำลอง GAMs ที่มาจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) มีอัตราส่วนการเปลี่ยนแปลงสูงถึงร้อยละ 83-90 ในขณะที่แบบจำลอง GLMs ที่ได้มาจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) มีอัตราส่วนการเปลี่ยนแปลงน้อยมากเพียงร้อยละ 1.65 ส่วนแบบจำลอง MAXENT ทั้งก่อนการวิเคราะห์ห้วงค์ประกอบ (Before FA) และก่อนการวิเคราะห์ห้วงค์ประกอบ (After FA) มีอัตราส่วนการเปลี่ยนแปลงประมาณร้อยละ 31-42

ตารางที่ 5.3 ร้อยละของพื้นที่พยากรณ์ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่าและการเปลี่ยนแปลงของพื้นที่พยากรณ์ระหว่างเวลาปัจจุบันและอนาคต อันสืบเนื่องมาจากการเปลี่ยนแปลงสภาพภูมิอากาศ

Model	Predicted percentage suitability area		% change	Species loss	Species gain	Turnover	IUCN status
				(%)	(%)		
	Current	Future					
GLMs (Method1)	46.58	37.78	-18.9	19.84	0.88	20.54	NT
GLMs (Method2)	49.66	48.84	-1.65	21.01	19.36	33.82	NT
GLMs (Method3)	49.16	30.20	-38.56	38.56	0.00	38.56	VU

ตารางที่ 5.3 (ต่อ)

Model	Predicted			Species loss (%)	Species gain (%)	Turnover	IUCN status
	percentage		% change				
	suitability area						
	Current	Future					
GAMs (Method1)	15.95	6.07	-61.94	76.64	14.7	79.63	EN
GAMs (Method2)	14.38	2.35	-83.66	99.80	16.14	99.83	CR
GAMs (Method3)	25.07	2.35	-90.62	93.13	2.51	93.29	CR
MAXENT (Before FA)	12.41	7.17	-42.21	52.15	9.94	56.47	VU
MAXENT (After FA)	11.57	7.98	-31.06	53.86	22.8	62.43	VU

เมื่อพิจารณาบัญชีแดงของสหภาพเพื่อการอนุรักษ์ธรรมชาติ (IUCN Red List of Threatened Species) ซึ่งเป็นดัชนีที่ใช้ประเมินความเสี่ยงต่อการสูญพันธุ์ โดยใช้ฐาน IUCN Red List criteria 2001 (Trisurat, Alkemade & Arets, 2009) ซึ่งมีรายละเอียดดังนี้

1. สัญลักษณ์ EX (Extinct) แทนสิ่งมีชีวิตที่ศึกษาที่มีการสูญพันธุ์ร้อยละ 100
2. สัญลักษณ์ CR (Critically endangered) แทนสิ่งมีชีวิตที่ศึกษามีโอกาสสูญพันธุ์ร้อยละ 80-100
3. สัญลักษณ์ EN (Endangered) แทนสิ่งมีชีวิตที่ศึกษาโอกาสสูญพันธุ์ร้อยละ 50-80
4. สัญลักษณ์ VU (Vulnerable) แทนสิ่งมีชีวิตที่ศึกษาโอกาสสูญพันธุ์ร้อยละ 30-50
5. สัญลักษณ์ NT (Near threatened) แทนสิ่งมีชีวิตที่ศึกษาโอกาสสูญพันธุ์น้อยกว่าร้อยละ 30
6. สัญลักษณ์ LC (Least concerned) แทนสิ่งมีชีวิตที่ศึกษาไม่มีโอกาสสูญพันธุ์

จากผลการศึกษาพบว่าแบบจำลอง GAMs มีโอกาสสูญพันธุ์สูงมาก โดยแบบจำลอง GAMs ที่ได้มาจากวิธีการสุ่มจุด Pseudo-Absence โดยการสุ่มภายนอกพื้นที่รอบจุด Presence (Method2) และแบบจำลอง GAMs ที่ได้มาจากวิธีการสุ่มจุด Pseudo-Absence แบบสองขั้นตอน (Method3) มีโอกาสสูญพันธุ์สูงถึงร้อยละ 80-100 ในขณะที่แบบจำลอง GAMs ที่ได้มาจากวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) มีโอกาสสูญพันธุ์สูงถึงร้อยละ 50-80

ส่วนแบบจำลอง GLMs มีโอกาสสูญพันธุ์ต่ำ โดยเฉพาะแบบจำลอง GAMs ที่ได้มาจากวิธีการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (Method1) และแบบจำลอง GAMs ที่มาจากวิธีการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence (Method2) ที่มีโอกาสสูญพันธุ์น้อยกว่าร้อยละ 30 ในขณะที่แบบจำลอง GLMs ที่มาจากวิธีการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน (Method3) และแบบจำลอง MAXENT ทั้งก่อนและหลังการวิเคราะห์ห่อหุ้มประกอบ ต้นชาป่ามีโอกาสสูญพันธุ์ถึงร้อยละ 30 ถึงร้อยละ 50 ในปี ค.ศ. 2050 แสดงให้เห็นว่า ในการสร้างแบบจำลองการกระจายของต้นชาป่าเพื่อประเมินผลกระทบของการเปลี่ยนแปลงสภาพภูมิอากาศ วิธีการเลือกใช้แบบจำลองและวิธีการสุ่มเลือกจุด Pseudo-Absence เพื่อนำไปใช้ในแบบจำลอง GLMs และ GAMs มีผลต่อการประเมินความเสี่ยงต่อการสูญพันธุ์

5.2 ข้อเสนอแนะ

5.2.1 การจัดเตรียมข้อมูล

ในขั้นตอนของการจัดเตรียมข้อมูล จะเห็นได้ว่าการวิเคราะห์ห่อหุ้มประกอบ (Factors Analysis) ในแบบจำลอง MAXENT นั้นตัวแปรมีความสัมพันธ์ระหว่างกันค่อนข้างสูง ซึ่งอาจส่งผลกระทบต่อแบบจำลองในวิธีการ GLMs และ GAMs ดังนั้นก่อนการสร้างแบบจำลองโดยวิธีการ GLMs และ GAMs จึงควรมีการจัดกลุ่มตัวแปรโดยใช้การวิเคราะห์ห่อหุ้มประกอบ (Factors Analysis) ก่อนซึ่งนอกจากลดความสัมพันธ์ระหว่างตัวแปรแล้ว ยังลดจำนวนตัวแปรลงอีกด้วย

5.2.2 การเลือกวิธีการสุ่มจุด Pseudo-Absence

เนื่องจากการสุ่มเลือกจุด Pseudo-Absence เพื่อนำไปใช้ในการสร้างแบบจำลอง GLMs และ GAMs มีผลต่อแบบจำลอง ดังนั้นการสุ่มเลือกจุด Pseudo-Absence จึงเป็นขั้นตอนสำคัญที่ต้องมีการศึกษาวิธีการหรือการสุ่มเลือกจุดในวิธีต่าง ๆ เพิ่มเติม โดยอาจมีการถ่วงน้ำหนักของพื้นที่ก่อนที่ทำการสุ่มเลือกจุด Pseudo-Absence หรืออาจเลือกใช้จุด Absence จริงที่ได้จากการสำรวจ

5.2.3 การเลือกใช้แบบจำลองการกระจายของสิ่งมีชีวิต

นอกเหนือจากแบบจำลอง GLMs, GAMs และ MAXENT แล้ว ยังมีแบบจำลองที่ใช้เทคนิคต่าง ๆ ในการสร้างแบบจำลอง อาทิ Regression Trees, Artificial Neural Networks, Stochastic Gradient Boosting เป็นต้น แต่ทั้งนี้ทั้งนั้นการเลือกใช้แบบจำลองการกระจายของสิ่งมีชีวิตจะขึ้นอยู่กับข้อมูลที่มีว่ามีความเหมาะสมต่อแบบจำลองที่เลือกใช้หรือไม่

5.2.4 การปรับปรุงวิธีการในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต

ในการสร้างแบบจำลองการกระจายของสิ่งมีชีวิต เมื่อข้อมูลจุดพิกซ์ของการค้นพบสิ่งมีชีวิตมีจำนวนน้อยหรือมีไม่มากพอในการสร้างแบบจำลอง วิธีในการแก้ไขปัญหาที่ดีที่สุดคือการออกสำรวจหากกลุ่มตัวอย่างเพิ่มเติม แต่วิธีการนี้เป็นวิธีการที่ยุ่งยากและเป็นการเพิ่มภาระค่าใช้จ่ายในการออกสำรวจพื้นที่ ดังนั้นวิธีการที่เหมาะสมอีกวิธีหนึ่งคือการนำเทคนิคทางสถิติมาใช้ในการปรับปรุงแบบจำลอง ยกตัวอย่างเช่น การเพิ่มจำนวนรอบของการทำซ้ำหรือการเพิ่มจำนวนรอบของการทำ Bootstrap ก็จะช่วยให้การสร้างแบบจำลองเข้าใกล้แบบจำลองที่แท้จริงยิ่งขึ้น

5.2.5 การนำไปประยุกต์ใช้

แบบจำลองการกระจายของสิ่งมีชีวิต นอกจากการศึกษาการกระจายของสิ่งมีชีวิตโดยอาศัยข้อมูลทางภูมิศาสตร์และภูมิอากาศแล้ว การศึกษาปัจจัยอื่น ๆ ที่คาดว่าจะมีผลต่อการเปลี่ยนแปลงการกระจายของสิ่งมีชีวิต เช่น ปัจจัยที่เกิดจากมนุษย์ ปัจจัยจากมลภาวะ หรือปัจจัยการขยายของเขตเมือง ก็น่าจะมีผลต่อการกระจายของสิ่งมีชีวิตในธรรมชาติเช่นกัน

การศึกษาแบบจำลองการกระจายของสิ่งมีชีวิต นอกจากจะเป็นการศึกษาการกระจายของสิ่งมีชีวิตประเภทพืชและสัตว์ต่าง ๆ ยังครอบคลุมไปถึงการศึกษาการกระจายของสิ่งมีชีวิตขนาดเล็กหรือเชื้อโรคที่ก่อให้เกิดโรคร้ายในมนุษย์หรือสัตว์ที่สัมพันธ์กับสภาพสิ่งแวดล้อมในพื้นที่นั้น ๆ



รายการอ้างอิง

รายการอ้างอิง

- ชลเทพ ทาตรี. (2552). *ร้อยเรื่องราว...น้ำ*. สืบค้นเมื่อ 10 กุมภาพันธ์ 2557, จาก <http://irrigation3.wordpress.com/2009/02/17/>
- บัวตอง จินะหล้า. (2551). *เส้นทางเมี่ยง-ชา “ศรีป่าน-ตาเวน”*. สืบค้นเมื่อ 19 มิถุนายน 2554, จาก <http://www.jokonan.org/upload/book/1000000002.pdf>
- พรชัย ปรีชาปัญญา. (2536). ภูมิปัญญาชาวบ้านด้านนิเวศน์เกี่ยวกับความยั่งยืนของสวนชาในป่าดิบเขาในภาคเหนือของประเทศไทย, *วารสารวนศาสตร์*, 12, 18-26.
- สายลม สัมพันธ์เวชโสภา, บัณฑิต วาฤทธิ์, ประวิตร พุทธานนท์, สันต์ ละอองศรี, อุดมกัณฑ์ ขาสสุวรรณ, สมเดช ศรีชัยรัตนกุล, ชวลิต กอสัมพันธ์, พนม วิญญาของ, สิริพัชร สุธีรภัทรานนท์, ชีรพงษ์ เทพกรณ์, อนัญญา เอกพันธ์, รุ่งโรจน์ นิลทอง และ ภิญโญ คงมีลาภ. (2550). *โครงการเก็บและกำหนดพันธุ์ชาที่เหมาะสมในพื้นที่ปลูกภาคเหนือของประเทศไทย*. รายงานการวิจัย พัฒนาและวิศวกรรม ฉบับสมบูรณ์. สถาบันชา มหาวิทยาลัยแม่ฟ้าหลวง, เชียงราย.
- สำนักประชาสัมพันธ์เขต 2 กรมประชาสัมพันธ์. (2554). *น้ำกับอุณหภูมิของโลก*. สืบค้นเมื่อ 10 กุมภาพันธ์ 2555, จาก http://region2.prd.go.th/ewt_news.php?nid=18320&filename=letgogreen
- สำนักอนุรักษ์และจัดการต้นน้ำ กรมอุทยานแห่งชาติ สัตว์ป่า และพันธุ์พืช. (ม.ป.ป.). *การจัดชั้นคุณภาพลุ่มน้ำ*. สืบค้นเมื่อ 27 มีนาคม 2555, จาก <http://www.dnp.go.th/watershed/class.htm>
- สุรจิต ภูักดี. (2549). *วนเกษตร (Agroforestry)*. สืบค้นเมื่อ 27 มีนาคม 2555, จาก http://www.agri.ubu.ac.th/sura/LectureN/agroforestry_Lac.pdf

โสภณ ชมชาญ. (ม.ป.ป.). *ความแห้งแล้งในประเทศไทย*. สืบค้นเมื่อ 10 กุมภาพันธ์ 2557, จาก http://www.ddd.go.th/web_UNCCD/dryland/page1.htm

ศุภนารถ เกตุเจริญ และ อัญชลี พัดมีเทศ. (ม.ป.ป.). *ชา*. สืบค้นเมื่อ 1 กุมภาพันธ์ 2554, จาก http://www.eto.ku.ac.th/neweto/e-book/plant/tree_fruit/fruit17.pdf

ศูนย์พัฒนาทรัพยากรการศึกษามหาวิทยาลัยมหาสารคาม. (ม.ป.ป.). *สถิติออนไลน์ : การทดสอบสมมติฐานสำหรับกลุ่มตัวอย่างมากกว่าสองกลุ่มที่เกี่ยวข้องกัน*. สืบค้นเมื่อ 26 กุมภาพันธ์ 2557, จาก http://www.elearning.msu.ac.th/opencourse/0504%20854/unit5_1.htm

Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where, and how many?. *Methods in Ecology and Evolution*. Retrieved February 14, 2012, from http://biology.mcgill.ca/grad/cecile/PDFs/16_BarbetMassin2012.pdf

Beerling, D. J., Huntley, B. & Bailey, J. P. (1995). Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science*, 6, 269-282.

Braber, K. D., Sato, D. & Lee, E. (2011, February). *Farm and Forestry Production and Marketing Profile for Tea (Camellia sinensis)*. Retrieved June 4, 2011, from http://agroforestry.net/scps/Tea_specialty_crop.pdf

Brown, L. H., (1966). *A report on tea growing potential in Kenya*. Kenya: Ministry of agriculture, Nairobi.

Carne, J. H. & Balerdi C. F. (2005). *Tea Growing in the Florida Home Landscape*. Retrieved June 4, 2011, from <http://edis.ifas.ufl.edu/pdf/HS/HS30800.pdf>

- Carpenter, G., Gillison, A. N. & Winter, J. (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, 2, 667 – 680.
- Drew, C. A., Wiersma, Y. F. & Huettmann, F. (2011). *Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications*. London: Springer New York Dordrecht Heidelberg,
- Elton, C. (1927). *Animal Ecology*. Sidgwick and Jackson, London.
- Franklin, J. (2010). *Mapping species distribution: spatial inference and prediction*. Cambridge University Press, New York.
- Fu, W. J., Carroll, R. J. & Wang, S. (2005, February 2). Estimating misclassification error with small samples via bootstrap cross-validation. *Bioinformatics*, 21, 1979 - 1986
- Gibson, L., Barrett, B. & Burbidge, A. (2007, November). Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions*, 13(6), 704-713
- Grinnell, J. (1917). Field tests of theories concerning distributional control. *The American Naturalist*, 51, 115-128.
- Grinnell, J. (1924). Geography and evolution. *Ecology*, 5, 225-229.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.
- Harler, C. R. (1964). *The Culture and Marketing of Tea*. London, New York: Oxford University Press.
- Hastie, T. & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science*, 1(3), 297-318.

- Hijmans, R.J. & Elith, J. (2011, August 30). *Species distribution modeling with R*. Retrieved September 12, 2011, from <http://cran.r-project.org/web/packages/dismo>
- Hinkle, D. E, William ,W. & Stephen G. J. (1998). *Applied Statistics for the Behavior Sciences*. New York: Houghton Mifflin.
- Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, 22, 415-457.
- Irwin, M. E. (2005). *Generalized Additive Models*. Retrieved May 1, 2012, from <http://www.Markirwin.net/stat135/Lecture/Lecture34.pdf>
- Lobo, J. M. & Tognelli, M. F. (2011). Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *Journal for Nature Conservation*, 19, 1-7.
- McClure, F.A. (1966). *The bamboos, a fresh perspective*. London: Oxford University Press.
- Mateo, R. G., Croat, T. B., Felicísimo, Á. M. & Muñoz, J. (2009). Profile or group discriminative techniques? Generating reliable species distribution models using pseudo – absences and target – group absences from natural history collections. *Diversity and Distribution*, 16, 84-94.
- Ministry of Agriculture, The Republic of Uganda. (2007). *A Guide to Tea Production*. Retrieved March 31, 2012, from <http://www.arenet.or.ug/uploads/Tea%20production%20guide.pdf>
- Pagella, T., Choocharoen, C., Preechapanya, P., Moss, C. & Sinclair, F.L. (2002). *Local ecological knowledge about watershed functions in northern Thailand: A guide to using the Agroecological Knowledge Toolkit (AKT)*. Retrieved June 4, 2011, from <http://akt.bangor.ac.uk/documents/ThaiGuide.pdf>

- Pearce, J. L. & Boyce, M. S. (2006). Modelling distribution and abundance with presence – only data. *Journal of Applied Ecology*, 43, 405-412.
- Pearson, R. G. (2007). *Species' Distribution Modeling for Conservation Educators and Practitioners*. Retrieved June 23, 2011, from http://biodiversityinformatics.amnh.org/files/SpeciesDistModelingSYN_1-16-08.pdf
- Phillips, S. J. & Dudík, M. (2008). Modeling of species distribution with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.
- Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecology Modelling*, 190, 231-259.
- Preechapanya, P. (1996). *Indigenous ecological knowledge about the sustainability of tea gardens in the hill evergreen forest of northern Thailand*. Doctoral Dissertation, School of Agricultural and forest Science, University of Wales, Bangor, UK.
- Preechapanya, P. (2002). *Indigenous Highland Agroforestry Systems of Northern Thailand*. Retrieved June 4, 2011, from <http://www.mekonginfo.org/assets/midocs/0002240-farming-indigenous-highland-agroforestry-systems-of-northern-thailand.pdf>
- Ranganathan, V. (n.d.). *Tea*. Retrieved June 5, 2011, from <http://www.fertilizer.org/ifa/content/download/9002/133893/version/1/file/tea.pdf>
- Robertson, M. P., Peter, C. I., Villet, M. H. & Ripley, B. S. (2003). Comparing model for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, 164, 153-167.
- SAS. (n.d.). *Generalized Additive Models*. Retrieved May 1, 2012, from <http://support.sas.com/rnd/app/da/new/dagam.html>

- Schapire, R. E., Phillips, S. J. & Dudík, M. (n.d.). *Maximum Entropy and Species Distribution Modeling*. Retrieved May 10, 2012, from <http://www.cs.princeton.edu/~schapire/talks/maxent-icml.pdf>
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423.
- Stepenson, N. L. (1998). Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation across spatial scales. *Journal of Biogeography*, 25, 855-870.
- Stokland, J.N., Halvorsen, R. & Støa, B. (2011). Species distribution modelling – Effect of Design and sample size of pseudo – absence observations. *Ecological Modelling*, 222, 1800 – 1809.
- Tang, T. (2000). Slope Profile Analysis and Classification on Limestone Residual Hills in Guilin, China. *Middle States Geographer*, 33, 40-53.
- Tea Research Association. (n.d.). *Climate and Soil*. Retrieved June 12, 2011, from <http://www.tocklai.net/Cultivation/index.aspx>
- Trisurat, Y., Alkemade, R. & Arets, E. (2009, June). Projecting forest tree distribution and adaptation to climate change in northern Thailand. *Journal of Ecology and Natural Environment*, 1(3), 055-063
- VanDerWal, J., Shoo, L. P., Graham, C. & Williams, S. E. (2009). Selecting pseudo – absence data for presence – only distribution modeling: How far should you stray from what you know?. *Ecology Modelling*, 220, 589-594.
- Wijeratne, M. A., Anandacoomaraswamy, A., Amarathunga, M. K. S. L. D., Ratnasiri, J., Basnayake, B. R. S. B. & Kalra, N. (2007). Assessment of impact of climate change on productivity of tea (*Camellia sinensis* L.) plantations in Sri Lanka. *Journal of the National Science Foundation of Sri Lanka*, 35(2), 119-126.

Wisz, M. S. & Guisan, A. (2009, April 24). Do pseudo-absence selection strategies influence species distribution models and their predictions? An information – theoretic approach based on simulated data. *BMC Ecology*, 9(8). Retrieved February 14, 2012, from <http://www.biomedcentral.com/content/pdf/1472-6785-9-8.pdf>

Wood, S.N. (2006, February 27). *Generalized Linear Models: An Introduction with R*. United State of America: Chapman and Hall.

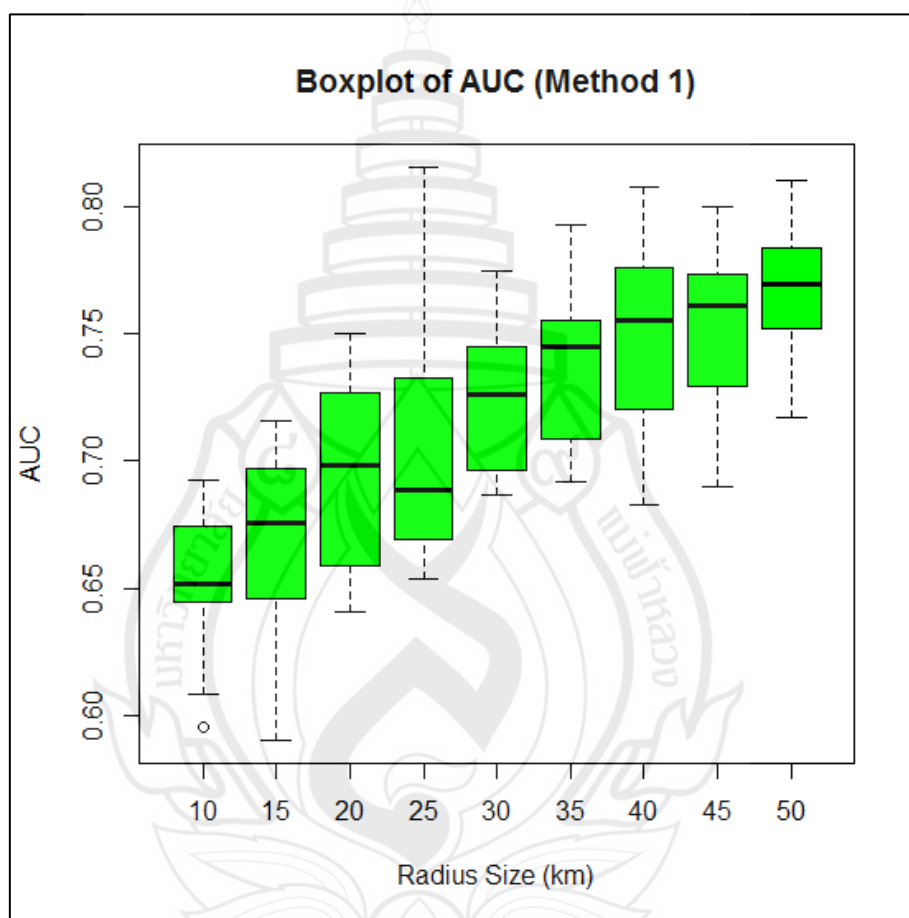




ภาคผนวก

ภาคผนวก ก

วิธีคัดเลือกรัศมีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา



ภาพที่ ก1 Boxplot ของค่า AUC ในแต่ละรัศมี

จากภาพที่ ก1 แสดง Boxplot ของค่า AUC ในแต่ละรัศมี ซึ่งเส้นสีดำตรงกลางของแต่ละกล่องจะแสดงค่ามัธยฐานของ AUC จะเห็นว่า ค่ามัธยฐานส่วนใหญ่มีแนวโน้มสูงขึ้นตามขนาดของรัศมี มีเพียงรัศมีขนาด 25 กิโลเมตรเท่านั้น ที่ค่ามัธยฐานมีค่าลดลง โดยค่ามัธยฐานของ AUC ส่วนใหญ่อยู่ระหว่าง 0.70 ถึง 0.80 ส่วนจุดวงกลมที่อยู่เหนือ Boxplot ที่รัศมี 25 กิโลเมตรจะแสดงค่าผิดปกติ (Outlier Value) ซึ่งค่าผิดปกตินี้จะไม่นำไปพิจารณา

ตารางที่ ก1 ค่าสถิติของค่า AUC จำแนกตามขนาดของรัศมี

รัศมี (กิโลเมตร)	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
10	0.65281	0.02639	0.04043	20
15	0.66775	0.03895	0.05833	20
20	0.70321	0.03814	0.05424	20
25	0.68786	0.02627	0.03819	19
30	0.73134	0.02750	0.03760	20
35	0.73702	0.02951	0.04004	20
40	0.75007	0.03548	0.04731	20
45	0.75449	0.03038	0.04026	20
50	0.76848	0.02200	0.02863	20

จากตารางที่ ก1 แสดงค่าสถิติของค่า AUC จำแนกตามขนาดของรัศมีโดยไม่นำค่าผิดปกติ (Outlier Value) ในรัศมีที่ 25 กิโลเมตรมาพิจารณาพบว่า ค่าเฉลี่ยในแต่ละรัศมีมีแนวโน้มเพิ่มขึ้นตามขนาดของรัศมี ซึ่งตั้งแต่รัศมีที่ 20 กิโลเมตรจะมีค่าเฉลี่ยของ AUC ตั้งแต่ 0.7 ขึ้นไป ยกเว้นรัศมีที่ 25 กิโลเมตรเท่านั้นที่มีค่าเฉลี่ยของ AUC น้อยกว่า 0.7 โดยค่าเฉลี่ยของ AUC ที่มีค่ามากที่สุดอยู่ที่รัศมี 50 กิโลเมตร ซึ่งมีค่าเท่ากับ 0.76848 และยังมีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุดด้วย ซึ่งมีค่าเท่ากับ 0.02863 คิดเป็นร้อยละ 2.86

ถึงแม้ว่าที่รัศมี 50 กิโลเมตรจะให้ค่า AUC สูงสุดและให้ค่าสัมประสิทธิ์ความผันแปรน้อยที่สุด ก็ไม่สามารถบ่งบอกได้ว่าที่รัศมี 50 กิโลเมตรคือรัศมีที่มีความเหมาะสมต่อการสร้างแบบจำลองการกระจายของต้นชาป่า เนื่องจากการพิจารณาเพียงแค่ค่าเฉลี่ยและค่าสัมประสิทธิ์ความผันแปรยังไม่สามารถให้คำตอบที่แน่ชัดได้ว่าที่รัศมีขนาดเท่าไรจึงจะเหมาะสมต่อการสร้างแบบจำลองการกระจายการกระจายของต้นชาป่า จึงได้ทำการพล็อตกราฟเพื่อหาช่วงของรัศมีที่มีความเหมาะสมโดยใช้การวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามมาประกอบในการคัดเลือก

ตารางที่ ก2 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า AUC

Model	AUC				
	Factors	Estimate	Std. Error	T Test	
Linear Curve	Intercept	6.31E-01	6.03E-03	104.58	***
	Radius	2.88E-03	1.84E-04	15.64	***
	$R^2 = 0.5801$				
Quadratic Curve	Intercept	6.07E-01	1.33E-02	45.763	***
	Radius	4.81E-03	9.88E-04	4.867	***
	Radius ²	-3.21E-05	1.62E-05	-1.982	*
	$R^2 = 0.5892$				
Cubic Curve	Intercept	6.11E-01	3.05E-02	20.057	***
	Radius	4.31E-03	3.76E-03	1.145	
	Radius ²	-1.34E-05	1.37E-04	-0.098	
	Radius ³	-2.07E-07	1.51E-06	-0.137	
	$R^2 = 0.5893$				
Polynomial of Degree 4 Curve	Intercept	6.20E-01	7.48E-02	8.290	
	Radius	2.69E-03	1.29E-02	0.208	
	Radius ²	8.37E-05	7.54E-04	0.111	
	Radius ³	-2.55E-06	1.79E-05	-0.142	
	Radius ⁴	1.95E-08	1.49E-07	0.131	
	$R^2 = 0.5893$				

ตารางที่ ก2 (ต่อ)

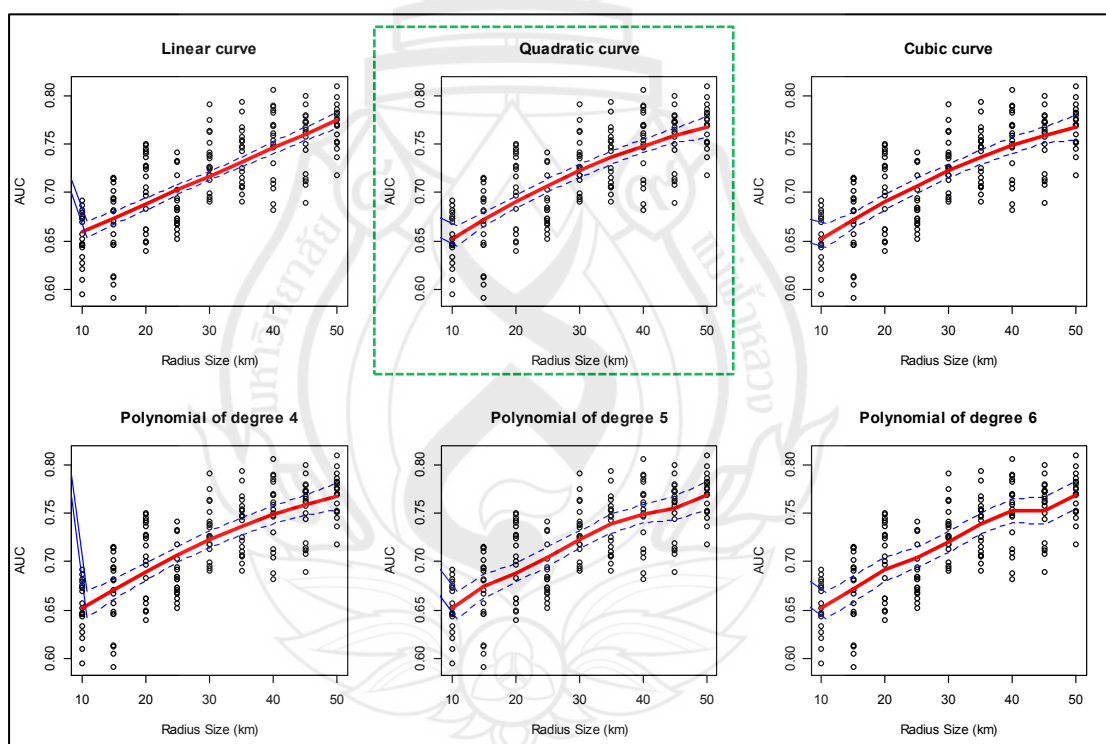
Model	AUC			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 5 Curve	Intercept	4.60E-01	2.04E-01	2.259 *
	Radius	3.91E-02	4.50E-02	0.869
	Radius ²	-2.92E-03	3.64E-03	-0.803
	Radius ³	1.11E-04	1.36E-04	0.819
	Radius ⁴	-1.99E-06	2.39E-06	-0.835
	Radius ⁵	1.34E-08	1.59E-08	0.845
$R^2 = 0.5910$				
Polynomial of Degree 6 Curve	Intercept	9.83E-01	6.27E-01	1.560
	Radius	-1.03E-01	1.68E-01	-0.612
	Radius ²	1.19E-02	1.73E-02	0.688
	Radius ³	-6.60E-04	8.90E-04	-0.742
	Radius ⁴	1.93E-05	2.44E-05	0.791
	Radius ⁵	-2.84E-07	3.39E-07	-0.837
	Radius ⁶	1.65E-09	1.88E-09	0.877
$R^2 = 0.5928$				

หมายเหตุ

*** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001

* หมายถึงมีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ ก2 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า AUC ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ตัวประมาณค่ามีความสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ให้ค่า R^2 สูงสุดคือ 0.5892 ดังนั้นจึงเลือกการถดถอยเชิงเส้นแบบพหุนามดีกรี 2 มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ก2 แผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ก2 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดิกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ 95% ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดิกรี 2 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟมีแนวโน้มเพิ่มขึ้นเรื่อยๆ ตามขนาดของรัศมี โดยช่วงรัศมีที่ 35 ถึงรัศมีที่ 50 กิโลเมตรเส้นกราฟเริ่มมีการเปลี่ยนแปลงเข้าสู่จุดสูงสุดของเส้นโค้งซึ่งเส้นกราฟก็ยังคงไม่ชัดเจน แต่เมื่อเปรียบเทียบกับเส้นกราฟในดิกรีที่สูงขึ้นพบว่าช่วงรัศมีที่ 40 ถึง 45 กิโลเมตรเส้นกราฟจะเริ่มจะคงที่ ซึ่งเห็นได้ชัดในเส้นกราฟของการวิเคราะห์การถดถอยแบบพหุนามดิกรี 6 และเนื่องจากการวิเคราะห์การถดถอยแบบพหุนามดิกรี 3 ถึงดิกรี 6 ยังมีตัวประมาณบางตัวไม่มีนัยสำคัญทางสถิติ ดังนั้นจึงไม่สามารถบอกได้ว่าช่วงรัศมี 35 ถึง 50 กิโลเมตรคือช่วงที่เหมาะสม จึงพิจารณาจากค่าอื่นเพิ่มเติมเพื่อประกอบการตัดสินใจ

ตารางที่ ก3 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Error Rate

Model	Error Rate			
	Factors	Estimate	Std. Error	T Test
Linear Curve	Intercept	4.32E-01	1.76E-02	24.47 ***
	Radius	-2.69E-03	5.40E-04	-4.988 ***
	$R^2 = 0.1232$			
Quadratic Curve	Intercept	4.14E-01	3.92E-02	10.547 ***
	Radius	-1.21E-03	2.92E-03	-0.414
	Radius ²	-2.47E-05	4.78E-05	-0.517
	$R^2 = 0.1246$			

ตารางที่ ก3 (ต่อ)

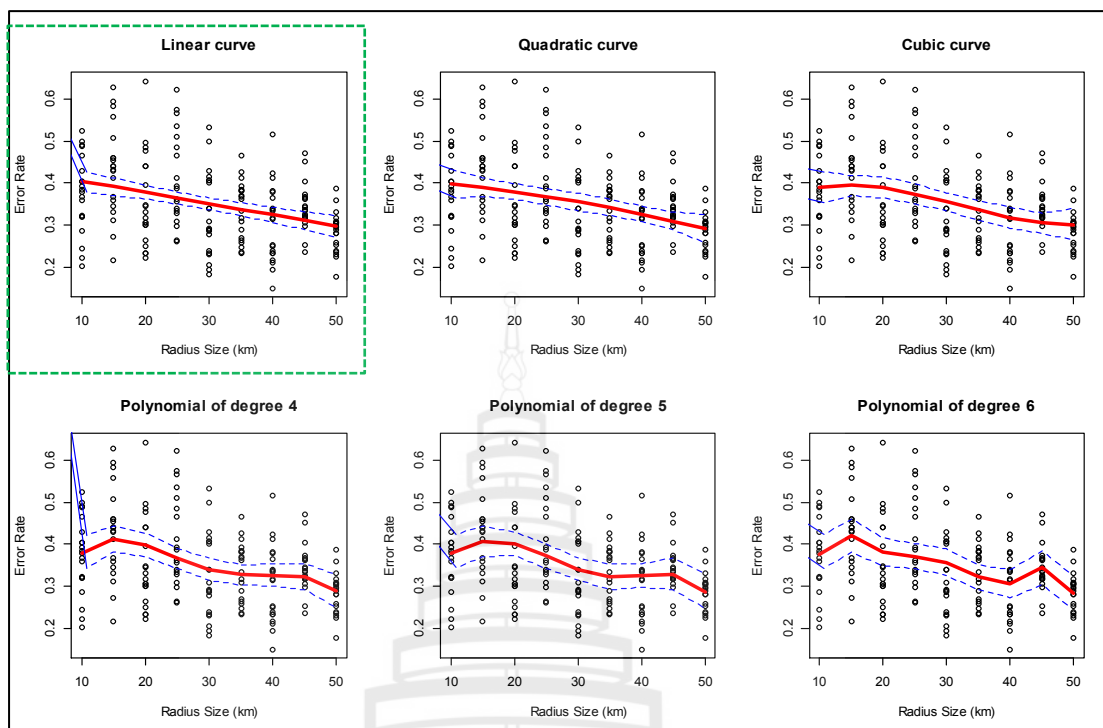
Model	Error Rate				
	Factors	Estimate	Std. Error	T Test	
Cubic Curve	Intercept	3.31E-01	8.98E-02	3.685	***
	Radius	9.76E-03	1.11E-02	0.88	
	Radius ²	-4.35E-04	4.03E-04	-1.08	
	Radius ³	4.56E-06	4.44E-06	1.025	
$R^2 = 0.1298$					
Polynomial of Degree 4 Curve	Intercept	-3.90E-02	2.18E-01	-0.179	
	Radius	7.68E-02	3.78E-02	2.034	*
	Radius ²	-4.45E-02	2.20E-03	-2.023	*
	Radius ³	1.01E-04	5.22E-05	1.937	.
	Radius ⁴	-8.06E-07	4.34E-07	-1.856	.
$R^2 = 0.1467$					
Polynomial of Degree 5 Curve	Intercept	2.94E-01	5.95E-01	0.494	
	Radius	1.07E-03	1.32E-01	0.008	
	Radius ²	1.80E-03	1.06E-02	0.169	
	Radius ³	-1.36E-04	3.97E-04	-0.342	
	Radius ⁴	3.38E-06	6.98E-06	0.485	
	Radius ⁵	-2.79E-08	4.64E-08	-0.602	
$R^2 = 0.1485$					

ตารางที่ ก3 (ต่อ)

Model	Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 6 Curve	Intercept	-2.90E+00	1.83E+00	-1.589
	Radius	8.68E-01	4.86E-01	1.784 .
	Radius ²	-8.88E-02	5.01E-02	-1.772 .
	Radius ³	4.58E-03	2.58E-03	1.775 .
	Radius ⁴	-1.27E-04	7.07E-05	-1.793 .
	Radius ⁵	1.79E-06	9.82E-07	1.819 .
	Radius ⁶	-1.01E-08	5.45E-09	-1.850 .
R ² = 0.1651				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ก3 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Error rate ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R² คือ 0.1232 ดังนั้นจึงเลือกการถดถอยเชิงเส้นแบบพหุนามดีกรี 1 มาพิจารณา นอกจากนี้อาจนำการถดถอยเชิงเส้นแบบพหุนามดีกรี 4 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 มาประกอบการพิจารณาด้วย เนื่องจากตัวประมาณเกือบทุกตัวมีนัยสำคัญทางสถิติ จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ก3 แผนภาพการกระจาย (Scatter plot) ของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม (เส้นประสีเขียวกคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ก3 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า Error Rate ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ 95% ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวนี้อาจเนื่องมาจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 1 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟแนวโน้มลดลง หมายความว่ายิ่งขนาดรัศมีใหญ่ขึ้น ค่า Error Rate ยิ่งลดลง เมื่อพิจารณาการถดถอยแบบพหุนามดีกรี 4 พบว่าค่า Error Rate ค่อนข้างคงที่ในช่วงรัศมี 35 กิโลเมตรถึง 45 กิโลเมตร แต่การถดถอยแบบพหุนามดีกรี 6 กลับพบว่าค่า Error Rate จะค่อยๆ ลดลงจนถึงรัศมีที่ 40 กิโลเมตรและจะเพิ่มสูงขึ้นที่รัศมี 45 กิโลเมตรและเมื่อพิจารณาค่า R^2 ในแต่ละดีกรีพบว่า ค่า R^2 ที่ได้จากการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามมีค่าน้อยมาก ดังนั้นจึงไม่สามารถบอกได้ว่าช่วงรัศมี 35 ถึง 50 กิโลเมตรคือช่วงที่เหมาะสม จึงพิจารณาจากค่าอื่นเพิ่มเติม

ตารางที่ ก4 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า AUC

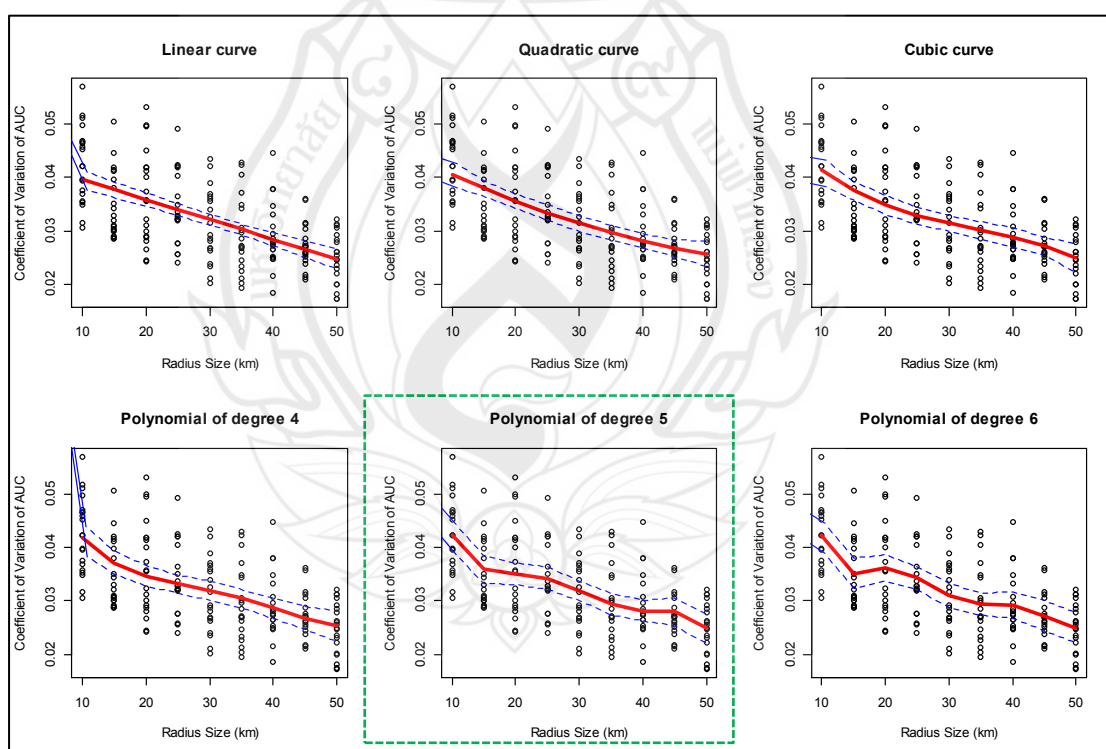
Model	Coefficient of Variation of AUC				
	Factors	Estimate	Std. Error	T Test	
Linear curve	Intercept	4.33E-02	1.24E-03	34.865	***
	Radius	-3.72E-04	3.80E-05	-9.783	***
	$R^2 = 0.3510$				
Quadratic curve	Intercept	4.64E-02	2.75E-03	16.855	***
	Radius	-6.23E-04	2.05E-04	-3.040	**
	Radius ²	4.19E-06	3.36E-06	1.247	
	$R^2 = 0.3566$				
Cubic curve	Intercept	5.31E-02	6.29E-03	8.434	***
	Radius	-1.51E-03	7.77E-04	-1.941	.
	Radius ²	3.73E-05	2.83E-05	1.322	
	Radius ³	-3.68E-07	3.12E-07	-1.182	
	$R^2 = 0.3617$				
Polynomial of degree 4 curve	Intercept	6.33E-02	1.54E-02	4.104	***
	Radius	-3.37E-03	2.67E-03	-1.260	
	Radius ²	1.49E-04	1.56E-04	0.955	
	Radius ³	-3.04E-06	3.69E-06	-0.824	
	Radius ⁴	2.23E-08	3.07E-08	0.727	
	$R^2 = 0.3637$				

ตารางที่ ก4 (ต่อ)

Model	Coefficient of Variation of AUC			
	Factors	Estimate	Std. Error	T Test
Polynomial of degree 5 curve	Intercept	1.30E-01	4.17E-02	3.102 **
	Radius	-1.84E-02	9.23E-03	-1.998 *
	Radius ²	1.39E-03	7.45E-04	1.867 .
	Radius ³	-5.02E-05	2.79E-05	-1.799 .
	Radius ⁴	8.55E-07	4.90E-07	1.747 .
	Radius ⁵	-5.55E-09	3.26E-09	-1.705 .
$R^2 = 0.3742$				
Polynomial of degree 6 curve	Intercept	2.98E-01	1.29E-01	2.319 *
	Radius	-6.42E-02	3.43E-02	-1.874 .
	Radius ²	6.18E-03	3.53E-03	1.75 .
	Radius ³	-3.00E-04	1.82E-04	-1.647
	Radius ⁴	7.73E-06	4.98E-06	1.553
	Radius ⁵	-1.02E-07	6.92E-08	-1.466
$R^2 = 0.3811$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ก4 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของ AUC ซึ่งได้จากการทำ Bootstrap1000 ครั้งในแต่ละชุดของ Pseudo-Absence Points ที่ขึ้นอยู่กับขนาดรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่าการถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 5 ตัวประมาณค่าตัวมีนัยสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 ให้ค่า R^2 สูงสุดคือ 0.3742 ดังนั้นจึงเลือก การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของ AUC ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ก4 แผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า AUC ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม (เส้นประสีเขียวนี้นคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ก4 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า Error Rate ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละรัศมีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ 95% ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 5 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟจะค่อยๆ ลดลงจนถึงรัศมีที่ 35 กิโลเมตรและจะค่อนข้างคงที่ในช่วงรัศมี 35 ถึง 45 กิโลเมตร หมายความว่าในช่วงรัศมี 35 ถึง 45 กิโลเมตรความแปรปรวนของค่า AUC ค่อนข้างคงที่และมีความแปรปรวนของค่า AUC น้อยกว่าความแปรปรวนในช่วงรัศมี 10 ถึง 30 กิโลเมตร และเพื่อลดปัญหาความผันแปรที่เกิดจากการทำ Bootstrap 1000 ครั้งจึงพิจารณาช่วงรัศมีที่มีค่าสัมประสิทธิ์ความผันแปรที่ค่อนข้างคงที่ ดังนั้นในช่วงรัศมีที่ 35 ถึง 45 กิโลเมตรจึงมีความเหมาะสมมากกว่าในช่วงรัศมีที่ 10 ถึง 30 กิโลเมตร

ตารางที่ ก5 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate

Model	Coefficient of Variation of Error Rate				
	Factors	Estimate	Std. Error	T Test	
Linear Curve	Intercept	2.74E-01	5.02E-03	54.59	***
	Radius	-1.75E-03	1.54E-04	-11.36	***
	$R^2 = 0.4218$				
Quadratic Curve	Intercept	2.64E-01	1.11E-02	23.675	***
	Radius	-8.84E-04	8.29E-04	-1.066	
	Radius ²	-1.44E-05	1.36E-05	-1.058	
$R^2 = 0.3566$					

ตารางที่ ก5 (ต่อ)

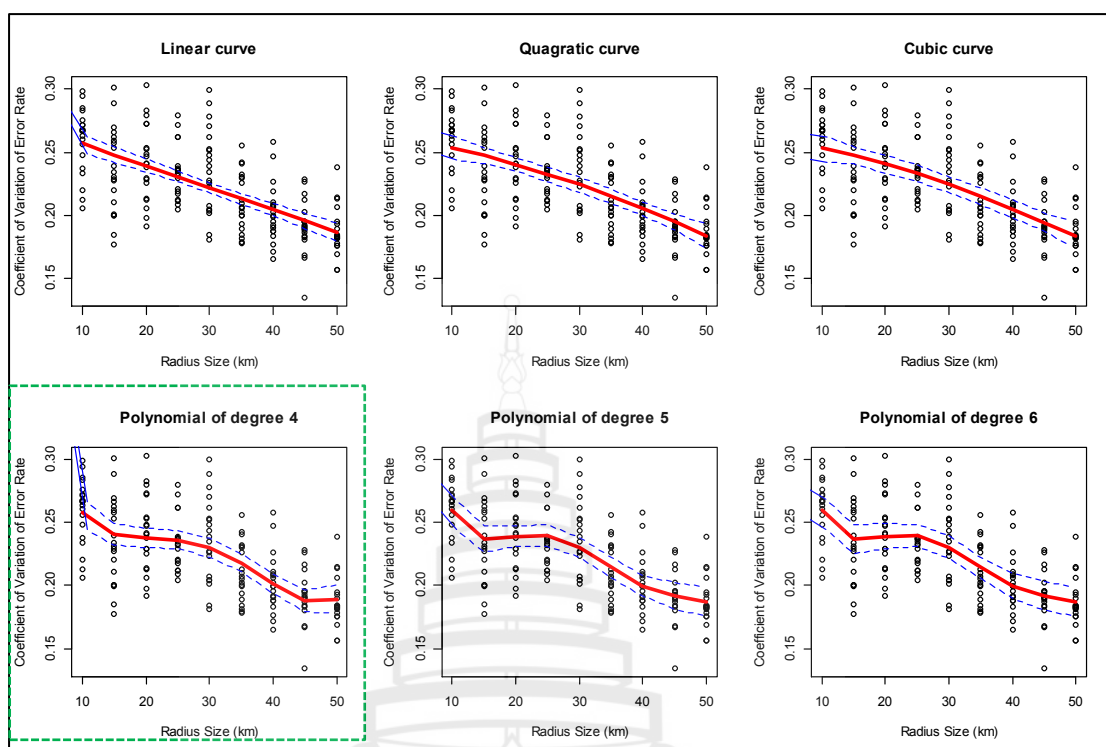
Model	Coefficient of Variation of Error Rate			
	Factors	Estimate	Std. Error	T Test
Cubic Curve	Intercept	2.59E-01	2.56E-02	10.138 ***
	Radius	-2.91E-04	3.16E-03	-0.092
	Radius ²	-3.66E-05	1.15E-04	-0.318
	Radius ³	2.46E-07	1.27E-06	0.195
		$R^2 = 0.4256$		
Polynomial of Degree 4 Curve	Intercept	4.00E-01	6.17E-02	6.474 ***
	Radius	-2.57E-02	1.07E-02	-2.408 *
	Radius ²	1.49E-03	6.22E-04	2.388 *
	Radius ³	-3.64E-05	1.48E-05	-2.463 *
	Radius ⁴	3.05E-07	1.23E-07	2.488 *
		$R^2 = 0.4453$		
Polynomial of Degree 5 Curve	Intercept	6.23E-01	1.67E-01	3.721 ***
	Radius	-7.66E-02	3.70E-02	-2.07 *
	Radius ²	5.68E-03	2.99E-03	1.901 .
	Radius ³	-1.96E-04	1.12E-04	-1.748 .
	Radius ⁴	3.12E-06	1.96E-06	1.588
	Radius ⁵	-1.87E-08	1.31E-08	-1.435
		$R^2 = 0.4518$		

ตารางที่ ก5 (ต่อ)

Model	Coefficient of Variation of Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 6 Curve	Intercept	6.27E-01	5.19E-01	1.207
	Radius	-7.76E-02	1.38E-01	-0.561
	Radius ²	5.78E-03	1.42E-02	0.406
	Radius ³	-2.01E-04	7.33E-04	-2.740
	Radius ⁴	3.26E-06	2.01E-05	0.162
	Radius ⁵	-2.08E-08	2.79E-07	-0.074
	Radius ⁶	1.13E-11	1.55E-09	0.007
$R^2 = 0.4518$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ก5 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ซึ่งได้จากการทำ Bootstrap 1000 ครั้งในแต่ละชุดของ Pseudo-Absence Points ที่ขึ้นอยู่กับขนาดรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่าการถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 4 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 ให้ค่า R^2 สูงสุดคือ 0.4453 ดังนั้นจึงเลือก การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ก5 แผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียว คือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ก5 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า Error Rate ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ 95% ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 4 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟในช่วงรัศมี 15 ถึง 25 กิโลเมตรจะค่อนข้างคงที่ และจะค่อยๆ ลดลงจนถึงรัศมีที่ 45 กิโลเมตรและจะเริ่มคงที่อีกครั้งในช่วงรัศมี 45 ถึง 50 หมายความว่าในช่วงรัศมี 45 ถึง 50 กิโลเมตรมีความแปรปรวนของค่า Error Rate น้อยกว่าความแปรปรวนในช่วงรัศมี 15 ถึง 25 กิโลเมตร ดังนั้นในช่วงรัศมี 45 ถึง 50 กิโลเมตรจึงมีความเหมาะสมมากกว่าในช่วงรัศมี 15 ถึง 25 กิโลเมตร

ตารางที่ ก6 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Cross-Validation Error

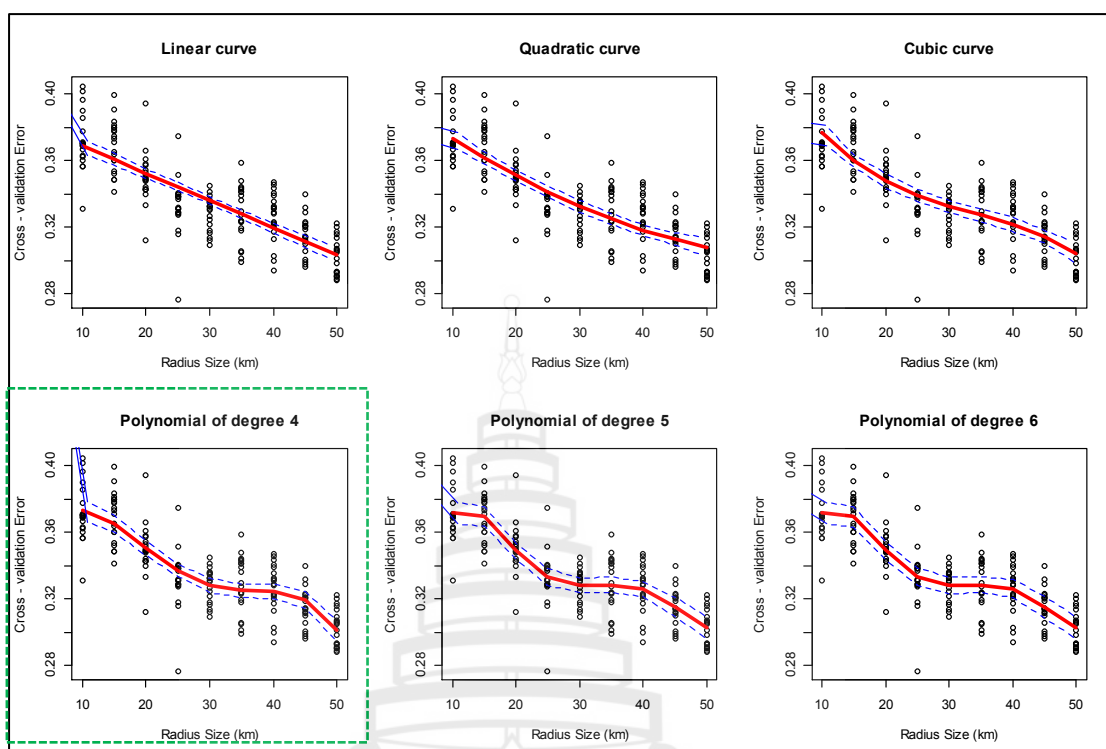
Model	Cross - validation Error				
	Factors	Estimate	Std. Error	T Test	
Linear Curve	Intercept	3.85E-01	2.98E-03	129.410	***
	Radius	-1.64E-03	9.11E-05	-18.020	***
$R^2 = 0.6471$					
Quadratic Curve	Intercept	4.00E-01	6.51E-03	61.469	***
	Radius	-2.84E-03	4.85E-04	-5.866	***
	Radius ²	2.00E-05	7.94E-06	2.523	*
$R^2 = 0.6594$					
Cubic Curve	Intercept	4.30E-01	1.47E-02	29.217	***
	Radius	-6.84E-03	1.82E-03	-3.760	***
	Radius ²	1.70E-04	6.61E-05	2.565	*
	Radius ³	-1.66E-06	7.29E-07	-2.278	*
$R^2 = 0.6693$					
Polynomial of Degree 4 Curve	Intercept	3.24E-01	3.51E-02	9.250	***
	Radius	1.24E-02	6.07E-03	2.039	*
	Radius ²	-9.81E-04	3.54E-04	-2.775	**
	Radius ³	2.60E-05	8.40E-06	3.100	**
	Radius ⁴	-2.31E-07	6.97E-08	-3.310	**
$R^2 = 0.6888$					

ตารางที่ ก6 (ต่อ)

Model	Cross-Validation Error			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 5 Curve	Intercept	9.71E-02	9.39E-02	1.034
	Radius	6.41E-02	2.08E-02	3.09 **
	Radius ²	-5.25E-03	1.68E-03	-3.132 **
	Radius ³	1.88E-04	6.27E-05	2.996 **
	Radius ⁴	-3.09E-06	1.10E-06	-2.808 **
	Radius ⁵	1.91E-08	7.32E-09	2.603 *
$R^2 = 0.7006$				
Polynomial of Degree 6 Curve	Intercept	1.45E-01	2.91E-01	0.500
	Radius	5.10E-02	7.75E-02	0.658
	Radius ²	-3.88E-03	7.98E-03	-0.486
	Radius ³	1.17E-04	4.11E-04	0.283
	Radius ⁴	-1.12E-06	1.13E-05	-0.100
	Radius ⁵	-8.40E-09	1.53E-07	-0.054
	Radius ⁶	1.53E-10	8.69E-10	0.176
$R^2 = 0.7006$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 6 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Cross-Validation Error ซึ่งได้มาจากการทำซ้ำ 10 ครั้งในแต่ละชุดของ Pseudo-Absence Points ขึ้นอยู่กับขนาดรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่าการถดถอยเชิงเส้นแบบพหุนามดีกรี 1, การถดถอยเชิงเส้นแบบพหุนามดีกรี 2, การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 4 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 ให้ค่า R^2 สูงสุดคือ 0.6888 ดังนั้นจึงเลือก การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 มาพิจารณา และอาจนำการถดถอยเชิงเส้นแบบพหุนามดีกรี 5 มาร่วมพิจารณาด้วย เนื่องจากตัวประมาณเกือบทุกตัวมีนัยสำคัญทางสถิติ จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่า Cross-Validation Error ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ 66 แผนภาพการกระจาย (Scatter Plot) ของค่า Cross-Validation Error ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ 66 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า Cross-Validation Error ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ 95% ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 4 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟจะค่อยๆ ลดลงจนถึงรัศมีที่ 30 กิโลเมตรและจะค่อนข้างคงที่ในช่วง 35 ถึง 45 กิโลเมตร หมายความว่าในช่วง 35 ถึง 45 กิโลเมตร ค่า Cross-Validation Error จะค่อนข้างคงที่มีลดลงเล็กน้อยในรัศมีที่ 45 กิโลเมตร และมีค่าน้อยกว่าค่า Cross-Validation Error ในช่วง 10 ถึง 30 กิโลเมตร

เมื่อพิจารณาทั้งเส้นกราฟที่ได้จากการวิเคราะห์การถดถอยแบบเชิงเส้นพหุนามของค่า AUC, Error Rate, สัมประสิทธิ์ความผันแปรของ AUC, สัมประสิทธิ์ความผันแปรของ Error Rate และค่า Cross-Validation Error พบว่าช่วงรัศมีที่มีความเหมาะสมที่สุด น่าจะอยู่ในช่วง 35 ถึง 50 กิโลเมตร เนื่องจากให้ค่า AUC ที่สูง ค่า Error Rate, ค่าสัมประสิทธิ์ความผันแปรของ AUC, ค่าสัมประสิทธิ์ความผันแปรของ Error Rate และค่า Cross-Validation Error ที่ต่ำและค่อนข้างคงที่ ดังนั้น จึงเลือกพิจารณาในช่วงรัศมีที่ 35 ถึง 50 กิโลเมตรเพื่อหารัศมีที่มีความเหมาะสมต่อไป

ตารางที่ ก7 ค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด

รัศมี (กิโลเมตร)	ค่าเฉลี่ยของ AUC	ชุด Pseudo - Absence	AUC
35	0.73702	19	0.74287
40	0.75007	11	0.75040
45	0.75449	17	0.75790
50	0.76848	17	0.76849

จากตารางที่ ก7 แสดงค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด โดยเลือกพิจารณาที่รัศมี 35 ถึง 50 กิโลเมตร พบว่าที่รัศมี 35 กิโลเมตร จากชุด Pseudo-Absence 20 ชุด ชุดที่ 19 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.74287 ที่รัศมี 40 กิโลเมตร ชุดที่ 11 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.75040 ที่รัศมี 45 และ 50 กิโลเมตร ชุดที่ 17 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.75790 และ 0.76849 ตามลำดับ

ชุด Pseudo-Absence ที่ถูกเลือกในแต่ละชุดจะถูกนำไปสร้างแบบจำลองโดยใช้วิธี Bootstrap และวิธี k-Fold Cross-Validation มาประยุกต์ใช้ซึ่งในขั้นตอนของการทำ Cross-Validation จะมีการทำซ้ำ 10 ครั้ง และเลือกกรอบของการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุด

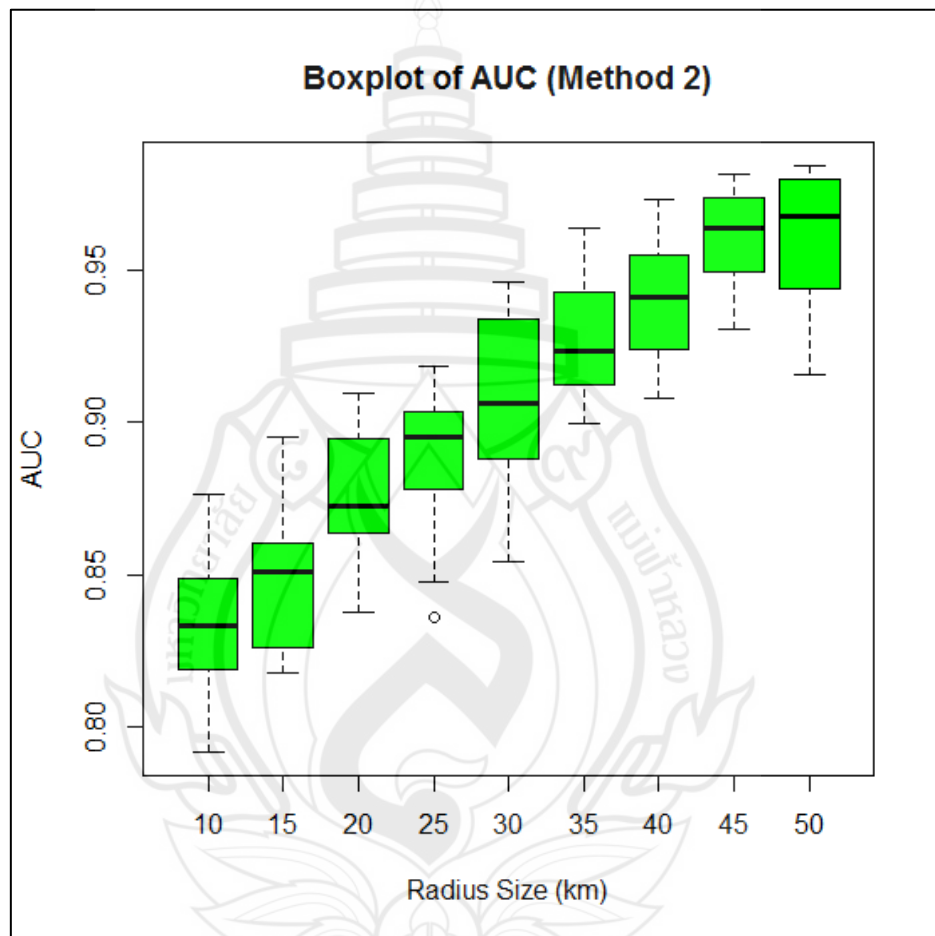
ตารางที่ ก8 ค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence ที่ถูกเลือก

รัศมี (กิโลเมตร)	ชุด Pseudo- Absence	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	สัมประสิทธิ์ ความผันแปร	จำนวน
35	19	0.32125	0.00111	0.00345	10
40	11	0.33274	0.00116	0.00349	10
45	17	0.31342	0.00083	0.00265	10
50	17	0.29157	0.00126	0.00433	10

จากตารางที่ ก8 แสดงค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence ที่ถูกเลือกในแต่ละรัศมีพบว่า ที่รัศมีที่ 50 กิโลเมตร มีค่าเฉลี่ยของ Cross-Validation Error น้อยที่สุด คือ 0.29157 ถัดมาคือรัศมีที่ 45, 35 และ 40 กิโลเมตรตามลำดับ ในขณะที่รัศมี 45 กิโลเมตรมีค่า สัมประสิทธิ์ความผันแปรน้อยที่สุดคือ 0.00265 ถัดมาคือรัศมีที่ 35, 40 และ 50 กิโลเมตรตามลำดับ ถึงแม้ว่าที่รัศมี 50 กิโลเมตรจะมีค่าเฉลี่ยน้อยที่สุดแต่ก็พบว่ามีค่าความผันแปรที่เกิดจากการทำซ้ำมากที่สุดเช่นเดียวกัน ดังนั้นเพื่อลดปัญหาความผันแปรที่เกิดจากการทำซ้ำที่รัศมี 50 กิโลเมตรจะไม่ถูก นำมาพิจารณา สรุปได้ว่าที่รัศมี 45 กิโลเมตรมีความเหมาะสม เนื่องจากมีค่าเฉลี่ยของ Cross-Validation Error น้อยและมีความผันแปรที่เกิดจากการทำซ้ำน้อยที่สุดด้วย

ภาคผนวก ข

วิธีคัดเลือกรัศมีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence



ภาพที่ ข1 Boxplot ของค่า AUC ในแต่ละรัศมี

จากภาพที่ ข1 แสดง Boxplot ของค่า AUC ในแต่ละรัศมี ซึ่งเส้นสีดำตรงกลางของแต่ละกล่องจะแสดงค่ามัธยฐานของ AUC จะเห็นว่า ค่ามัธยฐานส่วนใหญ่มีแนวโน้มสูงขึ้นตามขนาดของรัศมี โดยค่ามัธยฐานของ AUC ส่วนใหญ่อยู่ระหว่าง 0.80 ถึง 0.95 ส่วนจุดวงกลมที่อยู่ด้านล่างของ Boxplot ที่รัศมี 25 องศาจะแสดงค่าผิดปกติ (Outlier Value) ซึ่งค่าผิดปกตินี้จะไม่ถูกนำไปพิจารณา

ตารางที่ ข1 ค่าสถิติของค่า AUC จำแนกตามขนาดของรัศมี

รัศมี (กิโลเมตร)	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	สัมประสิทธิ์ความผันแปร (CV)	จำนวน
10	0.83497	0.02283	0.02734	20
15	0.85016	0.02482	0.02919	20
20	0.87943	0.02307	0.02624	20
25	0.89577	0.01437	0.01605	19
30	0.90649	0.02383	0.02628	20
35	0.92523	0.02334	0.02523	20
40	0.93578	0.01722	0.01840	20
45	0.95992	0.01409	0.01468	20
50	0.96251	0.02203	0.02289	20

จากตารางที่ ข1 แสดงค่าสถิติของค่า AUC จำแนกตามขนาดของรัศมีโดยไม่้นำค่าผิดปกติ (Outlier Value) ในรัศมีที่ 25 กิโลเมตรมาพิจารณาพบว่า ค่าเฉลี่ยในแต่ละรัศมีมีแนวโน้มเพิ่มขึ้นตามขนาดของรัศมี ซึ่งจะมีค่าเฉลี่ยของ AUC ตั้งแต่ 0.8 ขึ้นไป โดยค่าเฉลี่ยของ AUC ที่มีค่ามากที่สุดอยู่ที่รัศมี 50 กิโลเมตร ซึ่งมีค่าเท่ากับ 0.96251 แต่ที่รัศมี 45 กิโลเมตรให้ค่าสัมประสิทธิ์ความผันแปรน้อยที่สุด ซึ่งมีค่าเท่ากับ 0.01468 คิดเป็นร้อยละ 1.47

ถึงแม้ว่าที่รัศมี 50 กิโลเมตรจะให้ค่า AUC สูงสุดก็ตามแต่ก็ไม่สามารถบ่งบอกได้ว่าที่รัศมี 50 กิโลเมตรคือรัศมีที่มีความเหมาะสมต่อการสร้างแบบจำลองการกระจายของต้นชาป่าและเป็นที่น่าสังเกตว่า ค่า AUC มีค่าสูงมากถึง 0.96251 จึงมีความจำเป็นที่จะต้องพิจารณาค่าสถิติตัวอื่น ๆ เพื่อยืนยันความถูกต้อง เนื่องจากการพิจารณาเพียงค่าเฉลี่ยไม่สามารถให้คำตอบที่แน่ชัดได้ว่าที่รัศมีขนาดเท่าใดจึงจะเหมาะสมต่อการสร้างแบบจำลองการกระจายการกระจายของต้นชาป่า ดังนั้นจึงได้ทำการพล็อตกราฟของค่าต่าง ๆ ที่ได้จากแบบจำลองเพื่อหาช่วงของรัศมีที่มีความเหมาะสมโดยใช้การวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามมาประกอบในการคัดเลือก

ตารางที่ ข2 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า AUC

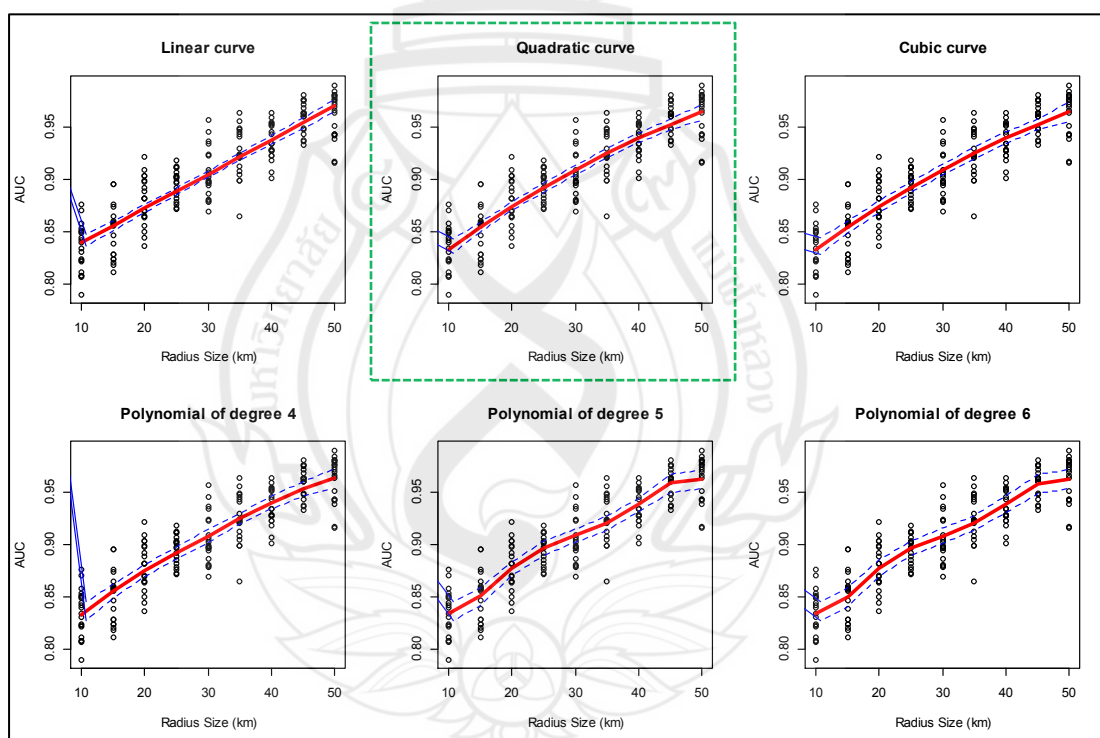
Model	AUC			
	Factors	Estimate	Std. Error	T Test
Linear Curve	Intercept	8.07E-01	4.03E-03	200.35 ***
	Radius	3.27E-03	1.23E-04	26.56 ***
	$R^2 = 0.7994$			
Quadratic Curve	Intercept	7.88E-01	8.19E-03	89.396 ***
	Radius	4.85E-03	6.56E-04	7.384 ***
	Radius ²	-2.62E-05	1.08E-05	-2.439 *
	$R^2 = 0.8060$			
Cubic Curve	Intercept	7.86E-01	2.02E-02	38.828 ***
	Radius	5.15E-03	2.50E-03	2.057 *
	Radius ²	-3.74E-05	9.09E-05	-0.412
	Radius ³	1.24E-07	1.00E-06	0.124
	$R^2 = 0.8060$			
Polynomial of Degree 4 Curve	Intercept	7.62E-01	4.96E-02	15.355 ***
	Radius	9.42E-03	8.60E-03	1.095
	Radius ²	-2.93E-04	5.01E-04	-0.585
	Radius ³	6.28E-06	1.19E-05	0.528
	Radius ⁴	-5.13E-08	9.87E-08	-0.519
	$R^2 = 0.8063$			

ตารางที่ ข2 (ต่อ)

Model	AUC			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 5 Curve	Intercept	1.02E+00	1.34E-01	7.585 ***
	Radius	-4.82E-02	2.96E-02	-1.629
	Radius ²	4.46E-03	2.39E-03	1.866 .
	Radius ³	-1.74E-04	8.95E-05	-1.945 .
	Radius ⁴	3.14E-06	1.57E-06	1.996 *
	Radius ⁵	-2.12E-08	1.05E-08	-2.033 *
$R^2 = 0.8108$				
Polynomial of Degree 6 Curve	Intercept	1.10E+00	4.15E-01	2.648 **
	Radius	-7.08E-02	1.11E-01	-0.641
	Radius ²	6.83E-03	1.14E-02	0.599
	Radius ³	-2.97E-04	5.87E-04	-0.507
	Radius ⁴	6.53E-06	1.61E-05	0.407
	Radius ⁵	-6.86E-08	2.23E-07	-0.307
	Radius ⁶	2.63E-10	1.24E-09	0.212
$R^2 = 0.8108$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ข2 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า AUC ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และ การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ให้ค่า R^2 สูงสุดคือ 0.8060 ดังนั้นจึงเลือก Quadratic Curve มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ข2 แผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวกคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ข2 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า AUC ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ร้อยละ 95 ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 2 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟมีแนวโน้มเพิ่มขึ้นเรื่อยๆ ตามขนาดของรัศมี แต่ในช่วงรัศมีที่เหมาะสมต่อการสร้างแบบจำลองการกระจายการกระจายของต้นชาป่ายังไม่ได้ เนื่องจากเส้นกราฟยังไม่ชัดเจน

ตารางที่ ข3 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Error Rate

Model	Error Rate				
	Factors	Estimate	Std. Error	T Test	
Linear Curve	Intercept	2.80E-01	6.37E-03	43.900	***
	Radius	-5.18E-03	1.95E-04	-26.590	***
	$R^2 = 0.7997$				
Quadratic Curve	Intercept	3.23E-01	1.37E-02	23.579	***
	Radius	-8.72E-03	1.02E-03	-8.557	***
	Radius ²	5.91E-05	1.67E-05	3.536	***
	$R^2 = 0.8130$				
Cubic Curve	Intercept	2.89E-01	3.13E-02	9.226	***
	Radius	-4.21E-03	3.87E-03	-1.089	
	Radius ²	-1.10E-04	1.41E-04	-0.780	
	Radius ³	1.87E-06	1.55E-06	1.208	
	$R^2 = 0.8146$				

ตารางที่ ข3 (ต่อ)

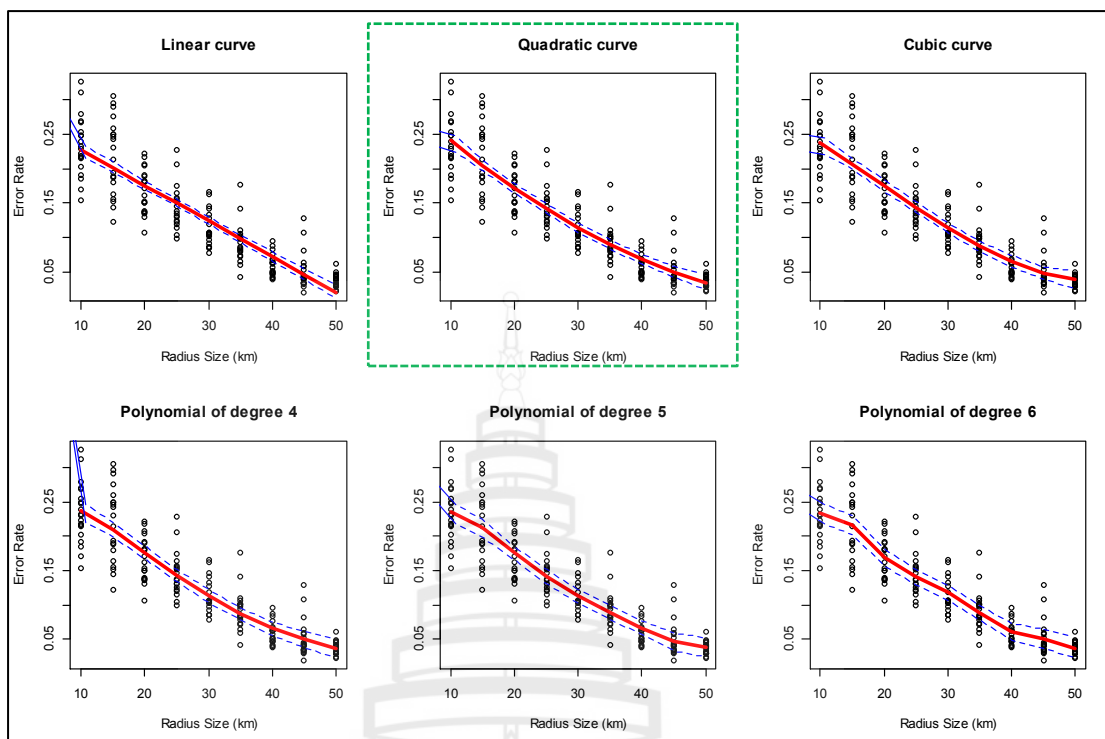
Model	Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 4 Curve	Intercept	2.53E-01	7.68E-02	3.289 **
	Radius	2.36E-03	1.33E-02	0.178
	Radius ²	-5.03E-04	7.74E-04	-0.650
	Radius ³	1.14E-05	1.84E-05	0.617
	Radius ⁴	-7.90E-08	1.53E-07	-0.517
$R^2 = 0.8148$				
Polynomial of Degree 5 Curve	Intercept	7.62E-02	2.09E-01	0.365
	Radius	4.25E-02	4.62E-02	0.920
	Radius ²	-3.82E-03	3.73E-03	-1.022
	Radius ³	1.37E-04	1.40E-04	0.980
	Radius ⁴	-2.30E-06	2.45E-06	-0.938
	Radius ⁵	1.48E-08	1.63E-08	0.907
$R^2 = 0.8157$				

ตารางที่ ข3 (ต่อ)

Model	Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of degree 6 curve	Intercept	-8.50E-01	6.44E-01	-1.320
	Radius	2.94E-01	1.72E-01	1.712 .
	Radius ²	-3.01E-02	1.77E-02	-1.702 .
	Radius ³	1.50E-03	9.10E-04	1.653
	Radius ⁴	-4.00E-05	2.49E-05	-1.605
	Radius ⁵	5.41E-07	3.46E-07	1.561
	Radius ⁶	-2.92E-09	1.92E-09	-1.520
$R^2 = 0.8182$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ข3 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Error rate ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่ การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ให้ค่า R^2 สูงสุดคือ 0.8130 ดังนั้นจึงเลือก Quadratic curve มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter plot) ของค่า Error rate ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ข3 แผนภาพการกระจาย (Scatter plot) ของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ข3 แสดงแผนภาพการกระจาย (Scatter plot) ของค่า AUC ในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ร้อยละ 95 ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 2 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟมีแนวโน้มลดลงเรื่อยๆ ตามขนาดของรัศมี ซึ่งยังหาช่วงรัศมีที่เหมาะสมต่อการสร้างแบบจำลองการกระจายการกระจายของต้นชาปายังไม่ได้ เนื่องจากเส้นกราฟยังไม่ชัดเจนเช่นเดียวกับเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ของค่า AUC

ตารางที่ ข4 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความ
ผันแปรของค่า AUC

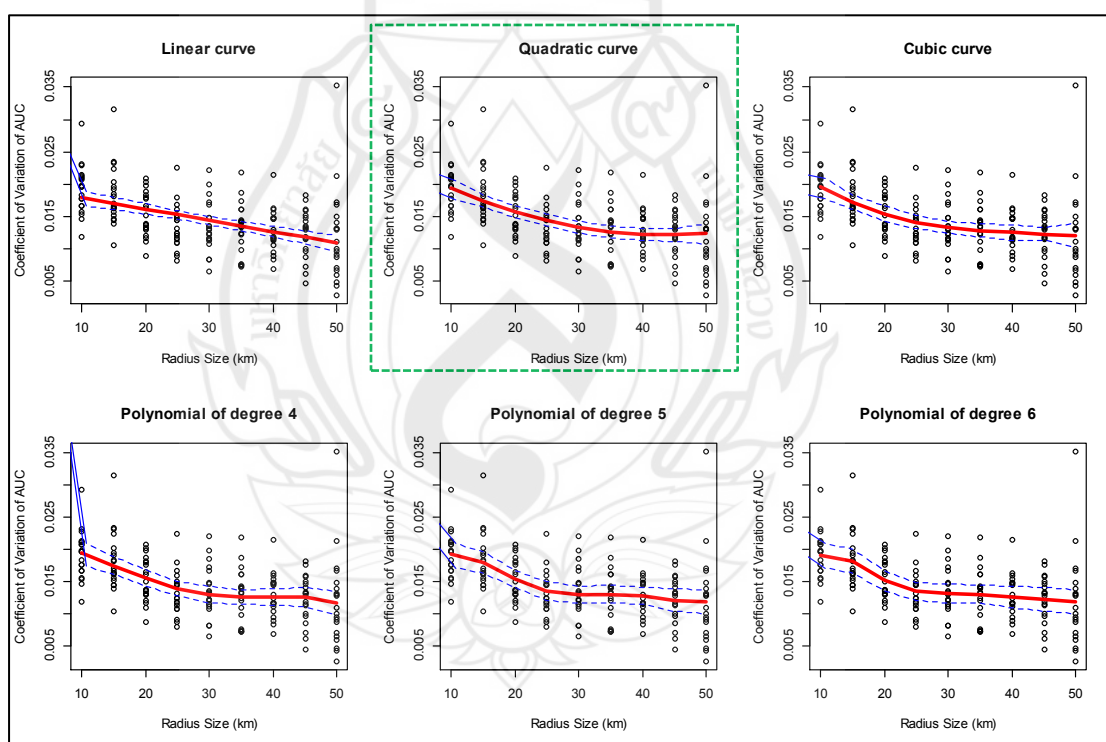
Model	Coefficient of Variance of AUC			
	Factors	Estimate	Std. Error	T Test
Linear Curve	Intercept	1.96E-02	8.30E-04	23.600 ***
	Radius	-1.75E-04	2.54E-05	-6.904 ***
		$R^2 = 0.2121$		
Quadratic Curve	Intercept	2.42E-02	1.81E-03	13.411 ***
	Radius	-5.54E-04	1.34E-04	-4.120 ***
	Radius ²	6.31E-06	2.20E-06	2.865 **
		$R^2 = 0.2473$		
Cubic Curve	Intercept	2.71E-02	4.14E-03	6.544 ***
	Radius	-9.33E-04	5.11E-04	-1.826 .
	Radius ²	2.05E-05	1.86E-05	1.104
	Radius ³	-1.58E-07	2.05E-07	-0.770
		$R^2 = 0.2498$		
Polynomial of Degree 4 Curve	Intercept	1.98E-02	1.01E-02	1.950 .
	Radius	3.92E-04	1.76E-03	0.223
	Radius ²	-5.89E-05	1.02E-04	-0.576
	Radius ³	1.75E-06	2.43E-06	0.722
	Radius ⁴	-1.59E-08	2.02E-08	-0.789
		$R^2 = 0.2525$		

ตารางที่ ๗4 (ต่อ)

Model	Coefficient of variance of AUC			
	Factors	Estimate	Std. Error	t value
Polynomial of Degree 5 Curve	Intercept	-5.66E-03	2.76E-02	-0.205
	Radius	6.18E-03	6.10E-03	1.014
	Radius ²	-5.36E-04	4.93E-04	-1.089
	Radius ³	1.99E-05	1.84E-05	1.078
	Radius ⁴	-3.36E-07	3.24E-07	-1.038
	Radius ⁵	2.13E-09	2.15E-09	0.991
$R^2 = 0.2567$				
Polynomial of Degree 6 Curve	Intercept	-4.28E-02	8.55E-02	-0.501
	Radius	1.63E-02	2.28E-02	0.714
	Radius ²	-1.59E-03	2.35E-03	-0.678
	Radius ³	7.47E-05	1.21E-04	0.619
	Radius ⁴	-1.85E-06	3.31E-06	-0.559
	Radius ⁵	2.30E-08	4.60E-08	0.506
	Radius ⁶	-1.17E-10	2.55E-10	-0.460
$R^2 = 0.2576$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 . หมายถึงมีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ข4 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า AUC ซึ่งได้จากการทำ Bootstrap 1000 ครั้งในแต่ละชุดของ Pseudo-Absence Points ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และ การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่ การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ให้ค่า R^2 สูงสุดคือ 0.2473 ดังนั้นจึงเลือก Quadratic Curve มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า AUC ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ข4 แผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า AUC ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวนี้อธิบายการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ข4 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า AUC ที่เกิดจากการทำ Bootstrap 1000 ครั้งในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ร้อยละ 95 ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 2 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟมีแนวโน้มลดลงและเริ่มคงที่ในช่วงรัศมีที่ 35 ถึง 50 กิโลเมตร หมายความว่าในช่วงรัศมี 10 ถึง 30 กิโลเมตรความแปรปรวนของค่า AUC จะค่อยๆ ลดลงและเริ่มคงที่ตั้งแต่รัศมีที่ 35 กิโลเมตรเป็นต้นไป จากเส้นกราฟนี้จะได้ช่วงรัศมีที่เหมาะสมคือช่วงรัศมีที่ 35 ถึง 50 กิโลเมตร

ตารางที่ ข5 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate

Model	Coefficient of Variance of Error Rate			
	Factors	Estimate	Std. Error	T Test
Linear Curve	Intercept	1.23E-01	6.25E-03	19.68 ***
	Radius	4.25E-03	1.91E-04	22.24 ***
		$R^2 = 0.7364$		
Quadratic Curve	Intercept	2.21E-01	1.12E-02	19.813 ***
	Radius	-3.79E-03	8.31E-04	-4.565 ***
	Radius ²	1.34E-04	1.36E-05	9.851 ***
		$R^2 = 0.8301$		
Cubic Curve	Intercept	2.57E-01	2.55E-02	10.104 ***
	Radius	-8.56E-03	3.15E-03	-2.723 **
	Radius ²	3.12E-04	1.14E-04	2.734 **
	Radius ³	-1.98E-06	1.26E-06	-1.571
		$R^2 = 0.8324$		

ตารางที่ ๗5 (ต่อ)

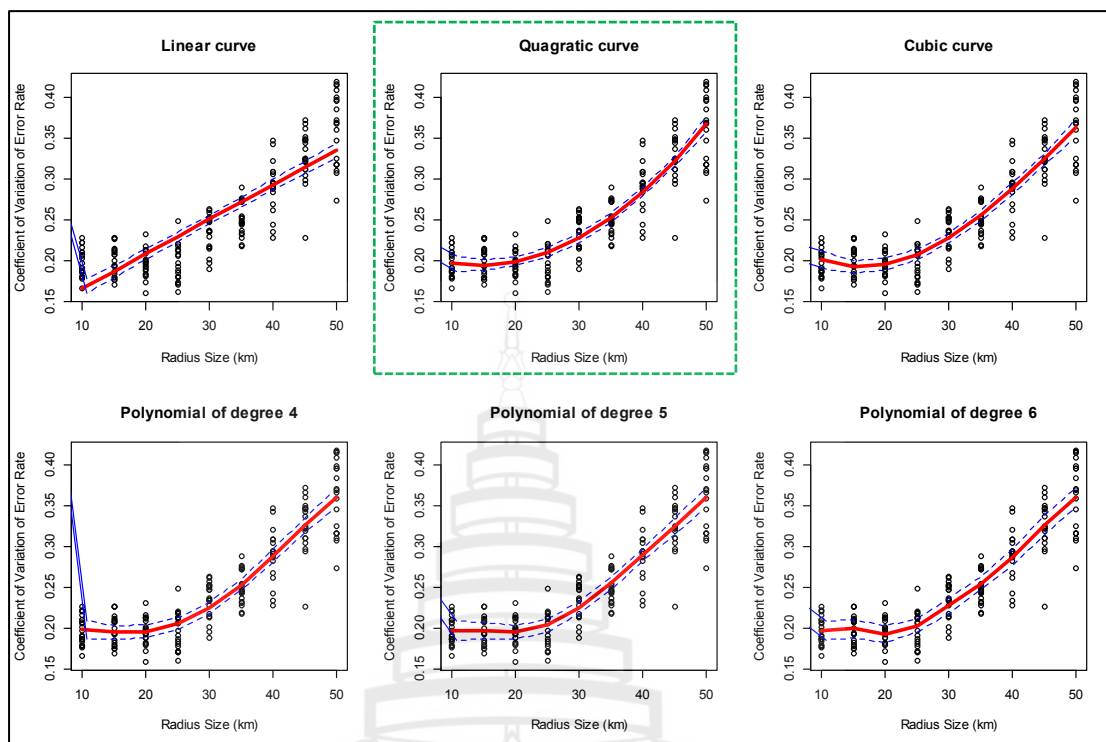
Model	Coefficient of Variance of Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 4 Curve	Intercept	1.93E-01	6.23E-02	3.104 **
	Radius	3.04E-03	1.08E-02	0.282
	Radius ²	-3.82E-04	6.28E-04	-0.609
	Radius ³	1.47E-05	1.49E-05	0.988
	Radius ⁴	-1.39E-07	1.24E-07	-1.125
$R^2 = 0.8337$				
Polynomial of Degree 5 Curve	Intercept	8.05E-02	1.70E-01	0.475
	Radius	2.87E-02	3.75E-02	0.765
	Radius ²	-2.50E-03	3.03E-03	-0.825
	Radius ³	9.50E-05	1.13E-04	0.838
	Radius ⁴	-1.56E-06	1.99E-06	-0.783
	Radius ⁵	9.46E-09	1.32E-08	0.714
$R^2 = 0.8341$				

ตารางที่ ข5 (ต่อ)

Model	Coefficient of Variance of Error Rate			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 6 Curve	Intercept	-3.88E-01	5.25E-01	-0.740
	Radius	1.56E-01	1.40E-01	1.115
	Radius ²	-1.58E-02	1.44E-02	-1.097
	Radius ³	7.87E-04	7.41E-04	1.062
	Radius ⁴	-2.06E-05	2.03E-05	-1.017
	Radius ⁵	2.76E-07	2.82E-07	0.977
	Radius ⁶	-1.48E-09	1.57E-09	-0.945
$R^2 = 0.8350$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01

จากตารางที่ ข5 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ซึ่งได้จากการทำ Bootstrap 1000 ครั้งในแต่ละชุดของ Pseudo-Absence Points ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 ให้ค่า R^2 สูงสุดคือ 0.8301 ดังนั้นจึงเลือก Quadratic Curve มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนาม



ภาพที่ ข5 แผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ข5 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่เกิดจากการทำ Bootstrap 1000 ครั้งในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ร้อยละ 95 ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 2 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟค่อนข้างคงที่ในช่วงรัศมีที่ 10 ถึง 20 กิโลเมตร และจะเริ่มมีแนวโน้มสูงขึ้นในช่วงรัศมีที่ 25 ถึง 50 กิโลเมตร หมายความว่าในช่วงรัศมีที่ 25 ถึง 50 กิโลเมตรความแปรปรวนของค่า Error Rate มีแนวโน้มสูงขึ้น หนึ่งในขั้นตอนของการคัดเลือกแบบจำลองจะกำหนดให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่เกิดจากการทำ Bootstrap 1000 ครั้งต้องมีค่าไม่เกินร้อยละ 30 ซึ่งจะเห็นได้ว่ายังกำหนดวงรัศมีที่กว้างขึ้นค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ก็จะเพิ่มมากขึ้นตามไปด้วยและมีแนวโน้มที่จะเกินร้อยละ 30 ดังนั้นเพื่อลดปัญหาความผันแปรที่เกิดขึ้น ในช่วง

รัศมีที่ 40 ถึง 50 กิโลเมตรจะไม่ถูกนำมาพิจารณา สรุปได้ว่าความแปรปรวนของค่า Error Rate จะมีค่าน้อยที่สุดและคงที่ในช่วงรัศมี 10 ถึง 20 กิโลเมตร ดังนั้นช่วงรัศมีที่เหมาะสมที่ได้คือช่วงรัศมี 10 ถึง 20 กิโลเมตร

ตารางที่ ข6 ค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Cross-Validation Error

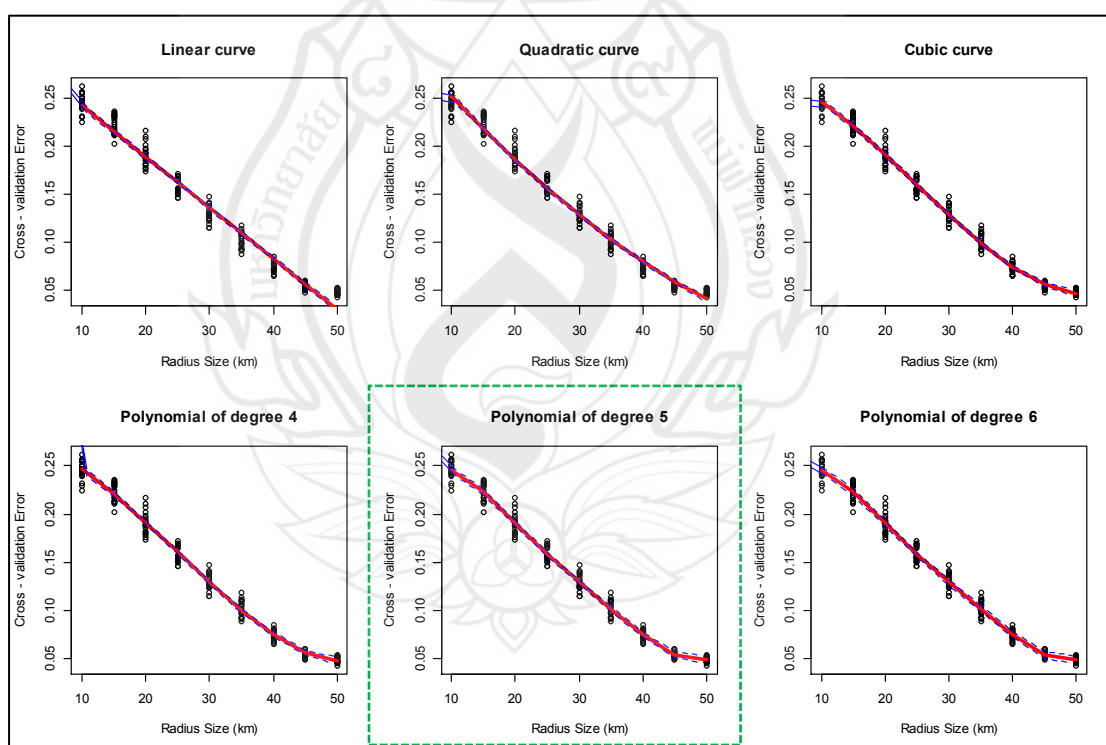
Model	Cross-Validation Error				
	Factors	Estimate	Std. Error	T Test	
Linear Curve	Intercept	2.95E-01	2.17E-03	135.720	***
	Radius	-5.29E-03	6.64E-05	-79.570	***
$R^2 = 0.9728$					
Quadratic Curve	Intercept	3.28E-01	3.91E-03	83.981	***
	Radius	-8.04E-03	2.91E-04	-27.620	***
	Radius ²	4.59E-05	4.77E-06	9.629	***
$R^2 = 0.9822$					
Cubic Curve	Intercept	2.76E-01	7.83E-03	35.267	***
	Radius	-1.11E-03	9.67E-04	-1.150	
	Radius ²	-2.13E-04	3.52E-05	-6.065	***
	Radius ³	2.88E-06	3.88E-07	7.424	***
$R^2 = 0.9865$					
Polynomial of Degree 4 Curve	Intercept	2.88E-01	1.92E-02	14.985	***
	Radius	-3.21E-03	3.32E-03	-0.964	
	Radius ²	-8.79E-05	1.94E-04	-0.454	
	Radius ³	-1.37E-07	4.60E-06	-0.030	
	Radius ⁴	-2.51E-08	3.82E-08	0.658	
$R^2 = 0.9865$					

ตารางที่ ข6 (ต่อ)

Model	Cross-Validation Error			
	Factors	Estimate	Std. Error	T Test
Polynomial of Degree 5 Curve	Intercept	1.70E-01	5.15E-02	3.293 **
	Radius	2.37E-02	1.14E-02	2.078 *
	Radius ²	-2.30E-03	9.19E-04	-2.505 *
	Radius ³	8.39E-05	3.44E-05	2.438 *
	Radius ⁴	-1.46E-06	6.04E-07	-2.417 *
	Radius ⁵	9.89E-09	4.02E-09	2.463 *
$R^2 = 0.9869$				
Polynomial of Degree 6 Curve	Intercept	1.51E-01	1.60E-01	0.943
	Radius	2.89E-02	4.25E-02	0.679
	Radius ²	-2.85E-03	4.38E-03	-0.65
	Radius ³	1.12E-04	2.26E-04	0.497
	Radius ⁴	-2.24E-06	7.70E-01	-0.363
	Radius ⁵	2.08E-08	8.59E-08	0.242
	Radius ⁶	-6.06E-11	4.77E-10	-0.127
$R^2 = 0.9869$				

หมายเหตุ *** หมายถึงมีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึงมีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึงมีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ ข6 แสดงค่าสถิติของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามของค่า Cross-Validation Error ซึ่งได้จากการทำซ้ำ 10 ครั้งในแต่ละชุดของ Pseudo-Absence Points ที่ขึ้นอยู่กับขนาดของรัศมี โดยเลือกพิจารณา 6 แบบจำลองคือ การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 (Linear Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 (Quadratic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 3 (Cubic Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 4 (Polynomial of Degree 4 Curve), การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 (Polynomial of Degree 5 Curve) และการถดถอยเชิงเส้นแบบพหุนามดีกรี 6 (Polynomial of Degree 6 Curve) พบว่า การถดถอยเชิงเส้นแบบพหุนามดีกรี 1 , การถดถอยเชิงเส้นแบบพหุนามดีกรี 2 และการถดถอยเชิงเส้นแบบพหุนามดีกรี 5 ตัวประมาณทุกตัวมีนัยสำคัญทางสถิติ แต่ การถดถอยเชิงเส้นแบบพหุนามดีกรี 5 ให้ค่า R^2 สูงสุดคือ 0.9869 ดังนั้นจึงเลือก Polynomial of Degree 5 Curve มาพิจารณา จากนั้นจึงเลือก Polynomial of Degree 5 Curve มาพิจารณา จากนั้นพล็อตแผนภาพการกระจาย (Scatter plot) ของค่า Cross-Validation Error ในแต่ละรัศมีและเส้นกราฟที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามดีกรี 5



ภาพที่ ข6 แผนภาพการกระจาย (Scatter Plot) ของค่า Cross-Validation Error ในแต่ละรัศมีและเส้นกราฟจากการวิเคราะห์การถดถอยแบบพหุนาม (เส้นประสีเขียวคือการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณา)

จากภาพที่ ข6 แสดงแผนภาพการกระจาย (Scatter Plot) ของค่า Cross-Validation Error ที่เกิดจากการทำซ้ำ 10 ครั้งในแต่ละรัศมี โดยมีเส้นกราฟสีแดงที่ได้จากการประมาณค่าของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุนามในแต่ละดีกรีและเส้นประสีน้ำเงินคือขอบเขตของช่วงความเชื่อมั่นที่ร้อยละ 95 ซึ่งการวิเคราะห์การถดถอยแบบพหุนามที่ถูกเลือกมาพิจารณาจะถูกล้อมด้วยเส้นประสีเขียวเนื่องจากตัวประมาณทุกตัวมีนัยสำคัญทางสถิติและให้ค่า R^2 สูงสุด พบว่าการวิเคราะห์การถดถอยแบบพหุนามดีกรี 5 ถูกเลือกมาพิจารณา ซึ่งจะเห็นว่าเส้นกราฟมีแนวโน้มลดลงตามขนาดของรัศมี และจุดเปลี่ยนของโค้งจะมีอยู่ 2 ช่วงคือช่วง 10 ถึง 20 กิโลเมตรและช่วงรัศมี 40 ถึง 50 กิโลเมตร แต่ช่วงรัศมีที่ 40 ถึง 50 กิโลเมตรจะไม่ถูกนำมาพิจารณาเนื่องจากมีค่าสัมประสิทธิ์ความผันแปรของค่า Error rate มากกว่าร้อยละ 30 ดังนั้นจากภาพนี้ช่วงรัศมีที่เหมาะสมน่าจะอยู่ในช่วง 10 ถึง 20 กิโลเมตร

เมื่อพิจารณาทั้งเส้นกราฟที่ได้จากการวิเคราะห์การถดถอยแบบเชิงเส้นพหุนามของค่า AUC, Error rate, สัมประสิทธิ์ความผันแปรของ AUC, สัมประสิทธิ์ความผันแปรของ Error Rate และค่า Cross-Validation Error พบว่าช่วงรัศมีที่มีความเหมาะสมที่สุด น่าจะอยู่ในช่วง 10 ถึง 20 กิโลเมตร ถึงแม้ว่าค่า AUC จะมีค่าน้อยกว่าช่วงรัศมีอื่นหรือค่า Error Rate จะมีค่ามากกว่าช่วงรัศมีอื่นก็ตาม แต่ค่าสัมประสิทธิ์ความผันแปรของ Error Rate กลับมีค่าน้อยและค่อนข้างคงที่ ซึ่งสอดคล้องกับกราฟของค่าสัมประสิทธิ์ความผันแปรของ Error Rate ที่เกิดจากการเลือกจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา (ภาพที่ 4.6) แต่เนื่องจากหลายๆ กราฟที่นำมาวิเคราะห์ยังไม่มีผลชัดเจนแน่นอน ดังนั้นจึงเลือกพิจารณาในช่วงรัศมีที่ 10 ถึง 35 กิโลเมตรโดยไม่นำช่วงรัศมีที่ 40 ถึง 50 กิโลเมตรจะไม่ถูกนำมาพิจารณาเนื่องจากมีค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate มากกว่าร้อยละ 30 จากนั้นหารัศมีที่มีความเหมาะสมต่อไป

ตารางที่ ข7 ค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence Points ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด

รัศมี (กิโลเมตร)	ค่าเฉลี่ยของ AUC	ชุด Pseudo-Absence Points	AUC
10	0.83497	20	0.83424
15	0.85016	17	0.84763
20	0.87943	15	0.88200
25	0.89577	3	0.89673
30	0.90649	15	0.90765
35	0.92523	14	0.91813

จากตารางที่ ข7 แสดงค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence Points ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด โดยเลือกพิจารณาที่รัศมี 10 ถึง 35 กิโลเมตร จากชุด Pseudo-Absence Points 20 ชุด พบว่าที่รัศมี 10 กิโลเมตร ชุดที่ 20 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.83424 ที่รัศมี 15 กิโลเมตร ชุดที่ 17 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.84763 ที่รัศมี 20 กิโลเมตร ชุดที่ 15 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.88200 ที่รัศมี 25 กิโลเมตร ชุดที่ 3 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.89673 ที่รัศมี 30 กิโลเมตร ชุดที่ 15 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.90765 ที่รัศมี 35 กิโลเมตร และชุดที่ 14 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.91813 ชุด Pseudo-Absence Points ที่ถูกเลือกในแต่ละชุดจะถูกนำไปสร้างแบบจำลองโดยใช้วิธี Bootstrap และวิธี k-Fold Cross-Validation มาประยุกต์ใช้ซึ่งในขั้นตอนของการทำ Cross-Validation จะมีการทำซ้ำ 10 ครั้ง และจะเลือกรอบของการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุด

ตารางที่ ข8 ค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ถูกเลือก

รัศมี (กิโลเมตร)	ชุด Pseudo- Absence Points	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	สัมประสิทธิ์ ความผันแปร	จำนวน
10	20	0.23038	0.0008	0.00345	10
15	17	0.21825	0.0009	0.00413	10
20	15	0.19326	0.00062	0.00323	10
25	3	0.15557	0.00047	0.00303	10
30	15	0.1242	0.00058	0.00468	10
35	14	0.09621	0.00047	0.00487	10

จากตารางที่ ข8 แสดงค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ถูกเลือกในแต่ละรัศมีพบว่า ที่รัศมีที่ 35 กิโลเมตร มีค่าเฉลี่ยของ Cross-Validation Error น้อยที่สุดคือ 0.09621 ถัดมาคือรัศมีที่ 30, 25, 20, 15 และ 10 กิโลเมตรตามลำดับ เมื่อพิจารณาค่าสัมประสิทธิ์ความผันแปรพบว่า ที่รัศมี 25 กิโลเมตร มีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุดคือ 0.00303 ถัดมาคือรัศมีที่ 20, 10, 15, 30 และ 35 กิโลเมตรตามลำดับ ถึงแม้ว่าที่รัศมี 35 กิโลเมตรจะมีค่าเฉลี่ยน้อยที่สุดแต่ก็พบว่ามีค่าสัมประสิทธิ์ความผันแปรที่เกิดจากการทำซ้ำมากที่สุดเช่นเดียวกัน ดังนั้นเพื่อลดปัญหาความผันแปรที่เกิดจากการทำซ้ำ ที่รัศมี 35 กิโลเมตรจะไม่ถูกนำมาพิจารณาแต่จะเลือกพิจารณาค่าสัมประสิทธิ์ความผันแปรที่น้อยที่สุดแทน เนื่องจากค่าเฉลี่ยไม่ค่อยแตกต่างกันมากนัก ซึ่งที่รัศมี 25 กิโลเมตรจะมีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุด ดังนั้นรัศมี 25 กิโลเมตรจึงมีความเหมาะสม

ตารางที่ ข9 ค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ 3 ที่รัศมี 25 กิโลเมตร

รอบการทำซ้ำ	Cross-Validation Error
1	0.15547
2	0.15490
3	0.15621
4	0.15556
5	0.15529
6	0.15641
7	0.15558
8	0.15578
9	0.15546
10	0.15504

จากตารางที่ ข9 แสดงค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ 3 ที่รัศมี 25 กิโลเมตรในการทำซ้ำ 10 รอบ พบว่าค่า Cross-Validation Error ในแต่ละรอบมีค่าไม่แตกต่างกันมากนัก ซึ่งรอบทำซ้ำที่จะถูกเลือกจะต้องให้ค่า Cross-Validation Error น้อยที่สุด นั่นคือรอบที่ 2 ซึ่งมีค่าเท่ากับ 0.15490

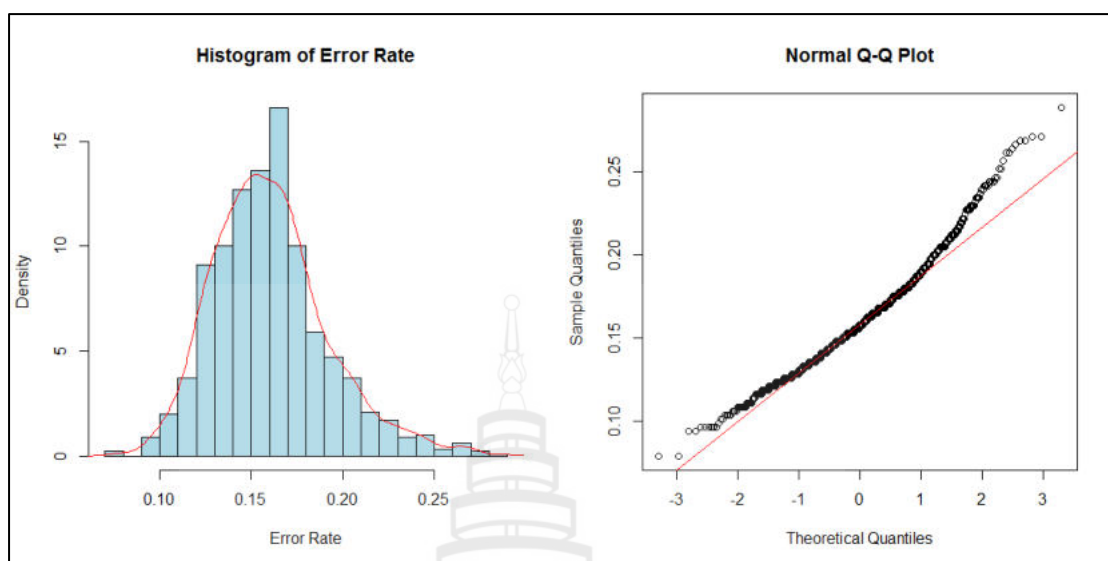
เมื่อได้รอบการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุดแล้ว ข้อมูลจะถูกแบ่งออกเป็นกลุ่ม Training Data และ Testing Data ซึ่งจะแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่ม โดยกลุ่มที่ 1 จะหมายถึงกลุ่มที่ใช้กลุ่มที่ 1 เป็นกลุ่ม Testing Data ในขณะที่กลุ่มอื่นๆ เป็นกลุ่ม Training Data กลุ่มที่ 2, กลุ่มที่ 3, กลุ่มที่ 4 และกลุ่มที่ 5 ก็จะเปลี่ยนกลุ่ม Testing Data ในลักษณะเดียวกันกับกลุ่มที่ 1 โดยในแต่ละกลุ่มจะมีการทำ Bootstrap จำนวน 1000 ครั้ง

ตารางที่ ข10 ค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 2 ชุด Pseudo-Absence Points ที่ 3 ที่รัศมี 25 กิโลเมตร

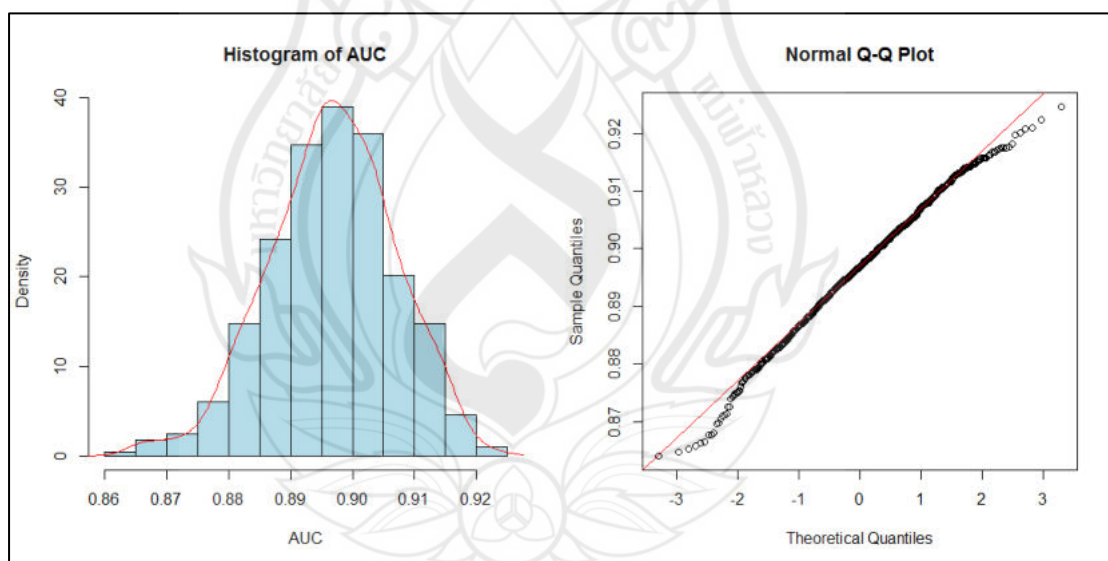
กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
1	0.14935	0.03794	0.25400	1000
2	0.19146	0.04532	0.23669	1000
3	0.13293	0.02768	0.20822	1000
4	0.14038	0.03263	0.23244	1000
5	0.16036	0.03158	0.19694	1000

จากตารางที่ ข10 แสดงค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 2 ชุด Pseudo-Absence Points ที่ 3 ที่รัศมี 25 กิโลเมตรในการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ค่าเฉลี่ยของ Error Rate ในกลุ่มที่ 3 มีค่าน้อยที่สุด แต่ค่าสัมประสิทธิ์ความผันแปรในกลุ่มที่ 5 กลับมีค่าน้อยที่สุดแทน เนื่องจากการลดความผันแปรจากการทำ Bootstrap จำนวน 1000 ครั้ง ดังนั้นจึงเลือกกลุ่มที่ 5 ซึ่งให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุดและตรงตามเงื่อนไขคือค่าสัมประสิทธิ์ความผันแปรไม่ควรมีค่าเกินร้อยละ 30 ซึ่งกลุ่มที่ 5 ให้ค่าสัมประสิทธิ์ความผันแปรที่ร้อยละ 19.69

ในการทำ Bootstrap จำนวน 1000 ครั้ง มีความจำเป็นที่ต้องมีการตรวจสอบการแจกแจงของค่าที่ต้องการศึกษาว่ามีการแจกแจงแบบปกติหรือไม่โดยการพล็อตฮิสโตแกรมและ Normal Q-Q Plot ของค่า Error Rate ดังภาพที่ ข7 พบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมที่วิ่งไปตามเส้นสีแดง จึงอนุมานได้ว่าค่า Error Rate ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติ



ภาพที่ ข7 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า Error Rate ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง



ภาพที่ ข8 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง

นอกเหนือจากค่า Error Rate ที่สนใจที่จะศึกษาแล้ว อีกค่าที่สนใจคือค่า AUC ดังภาพที่ ข8 ซึ่งแสดงแผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมที่วิ่งไปตามเส้นสีแดงเช่นเดียวกันกับค่า Error rate จึงอนุมานได้ว่าค่า AUC ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติ

ตารางที่ ข11 ค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3

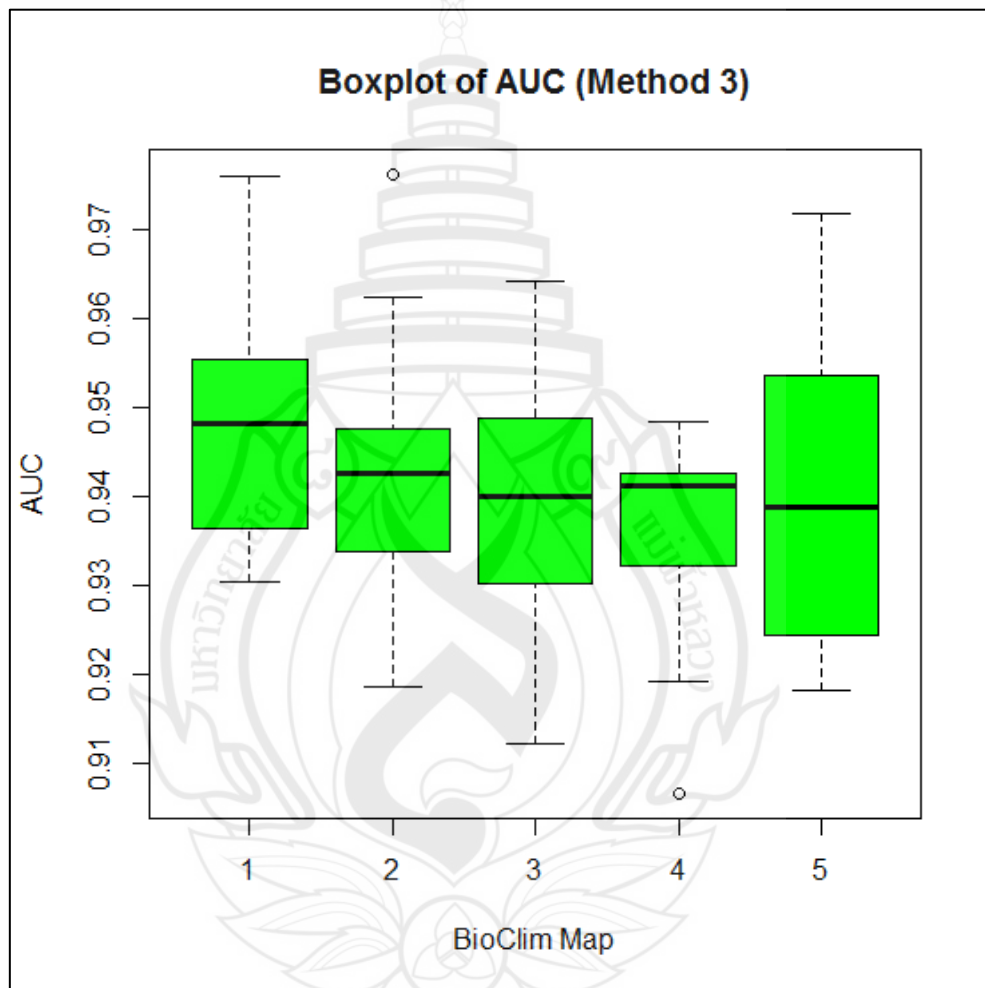
กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ ความผันแปร	จำนวน
5	0.89670	0.01019	0.01136	1000
แบบจำลอง		937		
AUC		0.89673		
Error Rate		0.17488		

จากตารางที่ ข11 แสดงค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 5 ซึ่งเป็นกลุ่มที่ถูกคัดเลือกแล้วเนื่องจากให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุดพบว่าในกลุ่มที่ 5 มีค่าเฉลี่ยของ AUC อยู่ที่ 0.89670 และแบบจำลองที่ถูกคัดเลือกจาก 1000 แบบจำลองคือแบบจำลองที่ 937 เนื่องจากมีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุดคือ 0.89673 โดยมีค่า Error Rate เท่ากับ 0.17488

เมื่อได้แบบจำลองที่ 145 เป็นตัวแทนที่เหมาะสมแล้ว ขั้นตอนต่อไปคือการคัดเลือกตัวแปรด้วยวิธีการของ Stepwise โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) มาคัดเลือกตัวแปร

ภาคผนวก ค

วิธีคัดเลือกวิธีที่มีความเหมาะสมของแบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence แบบ 2 ขั้นตอน



ภาพที่ ค1 Boxplot ของค่า AUC ในแต่ละ BIOCLIM MAP

จากภาพที่ ค1 แสดง Boxplot ของค่า AUC ในแต่ละ BIOCLIM MAP ซึ่งเส้นสีดำตรงกลางของแต่ละกล่องจะแสดงค่ามัธยฐานของ AUC จะเห็นว่า ค่ามัธยฐานส่วนใหญ่มีค่าใกล้เคียงกัน โดยค่ามัธยฐานของ AUC ส่วนใหญ่อยู่ระหว่าง 0.93 ถึง 0.95 ส่วนจุดวงกลมที่อยู่เหนือ Boxplot ที่ BIOCLIM MAP ที่ 2 จะแสดงค่าผิดปกติ (Outlier Value) ซึ่งค่าผิดปกตินี้จะไม่ถูกนำไปพิจารณา

ตารางที่ ค1 ค่าสถิติของค่า AUC จำแนกตาม BIOCLIM MAP

BIOCLIM	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ ความผันแปร	จำนวน
1	0.94840	0.01291	0.01361	20
2	0.94027	0.01131	0.01203	19
3	0.93958	0.01256	0.01336	20
4	0.93754	0.00887	0.00946	20
5	0.93978	0.01748	0.01860	20

จากตารางที่ ค1 แสดงค่าสถิติของค่า AUC จำแนกตามแผนที่ BIOCLIM โดยไม่นำค่าผิดปกติ (Outlier Value) ในแผนที่ BIOCLIM ที่ 2 มาพิจารณาพบว่า ค่าเฉลี่ยในแต่ละแผนที่ BIOCLIM มีค่าเฉลี่ยของ AUC ตั้งแต่ 0.93 ถึง 0.95 โดยแผนที่ BIOCLIM ที่ 1 มีค่า AUC สูงสุด และมีค่าสัมประสิทธิ์ความผันแปรใกล้เคียงกับแผนที่ BIOCLIM ที่ 3 ในขณะที่แผนที่ BIOCLIM ที่ 4 มีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุดและมีค่าเฉลี่ยของ AUC น้อยที่สุดเช่นเดียวกัน

ตารางที่ ค2 ค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence Points ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด

BIOCLIM	ค่าเฉลี่ยของ AUC	ชุด Pseudo-Absence Points	AUC
1	0.94840	1	0.94838
2	0.94027	13	0.93994
3	0.93958	16	0.94025
4	0.93754	3	0.93768
5	0.93978	17	0.93974

จากตารางที่ ค2 แสดงค่าเฉลี่ยของ AUC และชุดของ Pseudo-Absence Points ที่มีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุด พบว่าแผนที่ BIOCLIM ที่ 1 ชุดที่ 1 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.94838 แผนที่ BIOCLIM ที่ 2 ชุดที่ 13 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.93994 แผนที่ BIOCLIM ที่ 3 ชุดที่ 16 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.94025 แผนที่ BIOCLIM ที่ 4 ชุดที่ 3 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.93768 และแผนที่ BIOCLIM ที่ 5 ชุดที่ 17 ให้ค่า AUC ใกล้ค่าเฉลี่ยมากที่สุดคือ 0.93974 ชุด Pseudo-Absence Points ที่ถูกเลือกในแต่ละชุดจะถูกนำไปสร้างแบบจำลองโดยใช้วิธี Bootstrap และวิธี k-Fold Cross-Validation มาประยุกต์ใช้ซึ่งในขั้นตอนของการทำ Cross-Validation จะมีการทำซ้ำ 10 ครั้ง และจะเลือกรอบของการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุด

ตารางที่ ค3 ค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ถูกเลือก

BIOCLIM	ชุด Pseudo-Absence Points	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
1	1	0.12201	0.00018	0.00151	10
2	13	0.12625	0.00039	0.00307	10
3	16	0.12291	0.00068	0.00551	10
4	3	0.12405	0.00038	0.00308	10
5	17	0.11539	0.00038	0.00326	10

จากตารางที่ ค3 แสดงค่าสถิติของค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ถูกเลือกในแต่ละแผนที่ BIOCLIM พบว่า แผนที่ BIOCLIM ที่ 5 มี Cross-Validation Error โดยเฉลี่ยน้อยที่สุดคือ 0.11539 ถัดมาคือ แผนที่ BIOCLIM ที่ 1, 3, 4 และ 2 ตามลำดับ แต่เมื่อพิจารณาค่าสัมประสิทธิ์ความผันแปรพบว่า แผนที่ BIOCLIM ที่ 1 มีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุดคือ 0.00151 ถัดมาคือแผนที่ BIOCLIM ที่ 2, 4, 5 และ 3 ตามลำดับ แต่เพื่อลดปัญหาความผันแปรที่เกิดจากการทำซ้ำ จะเลือกพิจารณาค่าสัมประสิทธิ์ความผันแปรที่น้อยที่สุดแทน เนื่องจากค่าเฉลี่ยไม่ค่อยแตกต่างกันมากนัก ซึ่งแผนที่ BIOCLIM ที่ 1 มีค่าสัมประสิทธิ์ความผันแปรน้อยที่สุด ดังนั้นแผนที่ BIOCLIM ที่ 1 จึงมีความเหมาะสม

ตารางที่ ค4 ค่า Cross-validation error ของชุด Pseudo-Absence Points ที่ 1 แผนที่ BIOCLIM ที่ 1

รอบการทำซ้ำ	Cross-Validation Error
1	0.12205
2	0.12192
3	0.12179
4	0.1224
5	0.12204
6	0.12213
7	0.12216
8	0.1218
9	0.12188
10	0.12195

จากตารางที่ ค4 แสดงค่า Cross-Validation Error ของชุด Pseudo-Absence Points ที่ 1 ที่แผนที่ BIOCLIM ที่ 1 ในการทำซ้ำ 10 รอบ พบว่าค่า Cross-Validation Error ในแต่ละรอบมีค่าไม่แตกต่างกันมากนัก ซึ่งรอบทำซ้ำที่จะถูกเลือกจะต้องให้ค่า Cross-Validation Error น้อยที่สุด นั่นคือรอบที่ 3 ซึ่งมีค่าเท่ากับ 0.12179 (ส่วนที่แรเงา)

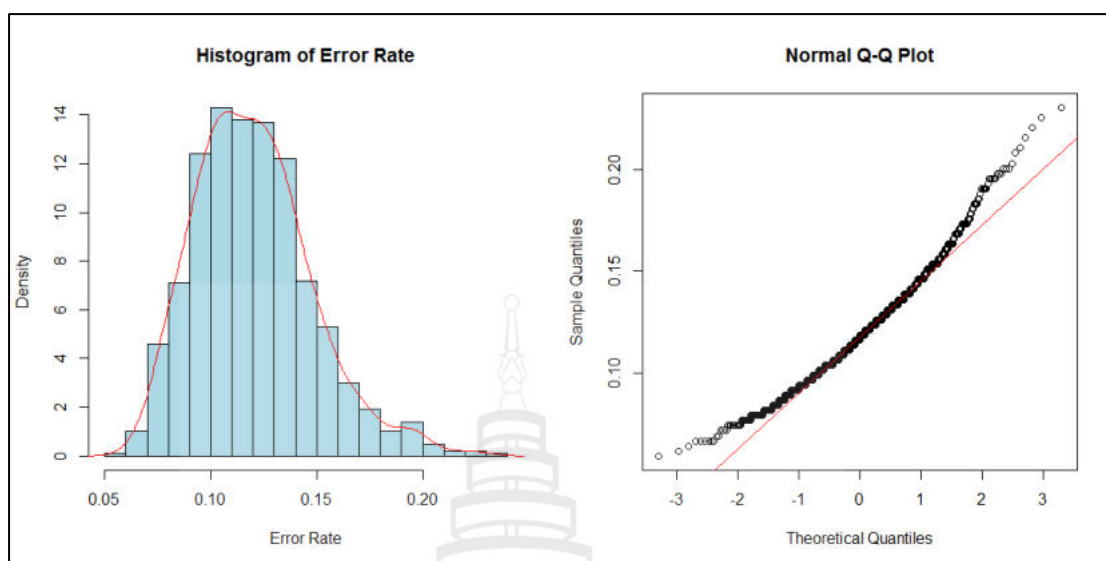
เมื่อได้รอบการทำซ้ำที่ให้ค่า Cross-Validation Error น้อยที่สุดแล้ว ข้อมูลจะถูกแบ่งออกเป็นกลุ่ม Training Data และ Testing Data ซึ่งจะแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่ม โดยกลุ่มที่ 1 จะหมายถึงกลุ่มที่ใช้กลุ่มที่ 1 เป็นกลุ่ม Testing data ในขณะที่กลุ่มอื่นๆ เป็นกลุ่ม Training Data กลุ่มที่ 2, กลุ่มที่ 3, กลุ่มที่ 4 และกลุ่มที่ 5 ก็จะเปลี่ยนกลุ่ม Testing Data ในลักษณะเดียวกันกับกลุ่มที่ 1 โดยในแต่ละกลุ่มจะมีการทำ Bootstrap จำนวน 1000 ครั้ง

ตารางที่ ค5 ค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 3 ชุด Pseudo-Absence Points ที่ 1 แผนที่ BIOCLIM ที่ 1

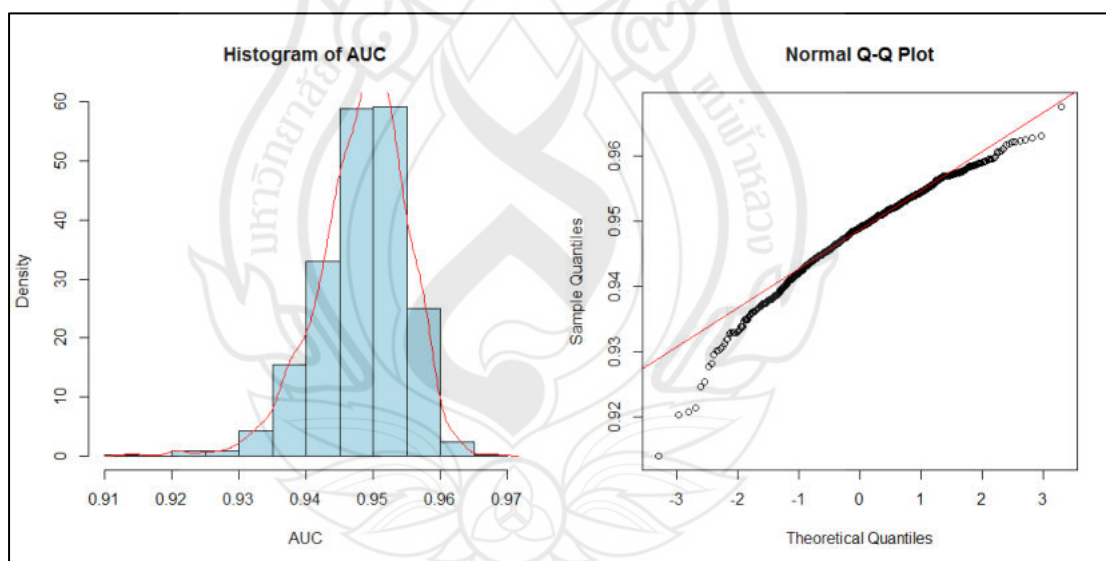
กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	สัมประสิทธิ์ความผันแปร	จำนวน
1	0.10300	0.03088	0.29979	1000
2	0.15365	0.05319	0.34618	1000
3	0.12011	0.02777	0.23122	1000
4	0.12182	0.03183	0.26127	1000
5	0.11039	0.03127	0.28325	1000

จากตารางที่ ค5 แสดงค่าสถิติของค่า Error Rate ในแต่ละกลุ่มของรอบการทำซ้ำที่ 3 ชุด Pseudo-Absence Points ที่ 1 แผนที่ BIOCLIM ที่ 1 ในการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ค่าเฉลี่ยของ Error Rate ในกลุ่มที่ 2 มีค่าน้อยที่สุด แต่ค่าสัมประสิทธิ์ความผันแปรในกลุ่มที่ 3 กลับมีค่าน้อยที่สุดแทน เนื่องจากการลดความผันแปรจากการทำ Bootstrap จำนวน 1000 ครั้ง ดังนั้นจึงเลือกกลุ่มที่ 3 ซึ่งให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุดและตรงตามเงื่อนไขคือค่าสัมประสิทธิ์ความผันแปรไม่ควรมีค่าเกินร้อยละ 30 ซึ่งกลุ่มที่ 5 ให้ค่าสัมประสิทธิ์ความผันแปรที่ร้อยละ 23.12 (ส่วนที่แรเงา)

ในการทำ Bootstrap จำนวน 1000 ครั้ง มีความจำเป็นที่ต้องมีการตรวจสอบการแจกแจงของค่าที่ต้องการศึกษาว่ามีการแจกแจงแบบปกติหรือไม่โดยการพล็อตฮิสโตแกรมและ Normal Q-Q Plot ของค่า Error Rate ดังภาพที่ ค2 พบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมที่วิ่งไปตามเส้นสีแดง จึงอนุมานได้ว่าค่า Error Rate ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติ



ภาพที่ ค2 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า Error Rate ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง



ภาพที่ ค3 แผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้ง

นอกเหนือจากค่า Error Rate ที่สนใจที่จะศึกษาแล้ว อีกค่าที่สนใจคือค่า AUC ดังภาพที่ ค3 ซึ่งแสดงแผนภาพฮิสโตแกรม (Histogram) และ Normal Q-Q Plot ของค่า AUC ที่เกิดจากการทำ Bootstrap จำนวน 1000 ครั้งพบว่า ฮิสโตแกรมมีลักษณะคล้ายโค้งระฆังคว่ำ และ Normal Q-Q Plot มีจุดวงกลมที่วิ่งไปตามเส้นสีแดงเช่นเดียวกันกับค่า Error rate จึงอนุมานได้ว่าค่า AUC ที่คำนวณได้จากการทำ Bootstrap จำนวน 1000 ครั้ง มีการแจกแจงแบบปกติ

ตารางที่ ค6 ค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3

กลุ่ม	ค่าเฉลี่ย	ส่วนเบี่ยงเบน มาตรฐาน	สัมประสิทธิ์ความ ผันแปร	จำนวน
3	0.94836	0.006565	0.006922	1000
แบบจำลอง	681			
AUC	0.94838			
Error Rate	0.08168			

จากตารางที่ ค6 แสดงค่าสถิติของค่า AUC และแบบจำลองที่ถูกคัดเลือกในกลุ่มที่ 3 ซึ่งเป็นกลุ่มที่ถูกคัดเลือกแล้วเนื่องจากให้ค่าสัมประสิทธิ์ความผันแปรของค่า Error Rate ที่น้อยที่สุด พบว่าในกลุ่มที่ 3 มีค่าเฉลี่ยของ AUC อยู่ที่ 0.94836 และแบบจำลองที่ถูกคัดเลือกจาก 1000 แบบจำลองคือแบบจำลองที่ 681 เนื่องจากมีค่า AUC เข้าใกล้ค่าเฉลี่ยมากที่สุดคือ 0.94838 โดยมีค่า Error Rate เท่ากับ 0.08168

เมื่อได้แบบจำลองที่ 145 เป็นตัวแทนที่เหมาะสมแล้ว ขั้นตอนต่อไปคือการคัดเลือกตัวแปรด้วยวิธีการของ Stepwise โดยใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion) หรือเกณฑ์เอไอซี (AIC) มาคัดเลือกตัวแปร

ภาคผนวก ง

แบบจำลอง GLMs โดยแยกวิเคราะห์ระหว่างตัวแปรทางด้านภูมิศาสตร์ และตัวแปรทางด้านภูมิอากาศ

1. แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ในพื้นที่ที่กำลังศึกษา

ตารางที่ ง1 แบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test
Intercept	-18.430	458.800	-0.040
DEM	0.003	0.000	9.717 ***
Distance	-0.093	0.061	-1.515
Class1	14.400	458.800	0.031
Class2	14.420	458.800	0.031
Class5	14.360	458.800	0.031
Class6	13.560	458.800	0.030
Null deviance = 892.55 , Residual deviance = 769.69			
Pseudo-R ² = 13.77, AIC = 783.69			

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

จากตารางที่ ง1 แสดงแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือความสูงจากระดับน้ำทะเล (DEM), ระยะห่างจากแม่น้ำ (Distance) และชั้นของข้อมูลชั้นหินและความลาดชัน (Class1, Class2, Class5 และ Class6)

โดยมีค่า Null deviance หรือค่า Log likelihood สำหรับ Constant-only Model (L_0) เท่ากับ 892.55 และค่า Residual deviance หรือค่า Log likelihood สำหรับ Full Model (L_1) เท่ากับ 769.69 ซึ่งสามารถคำนวณหาค่า Pseudo- R^2 ได้เท่ากับ 13.77 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 13.77 ได้สมการพยากรณ์ดังนี้

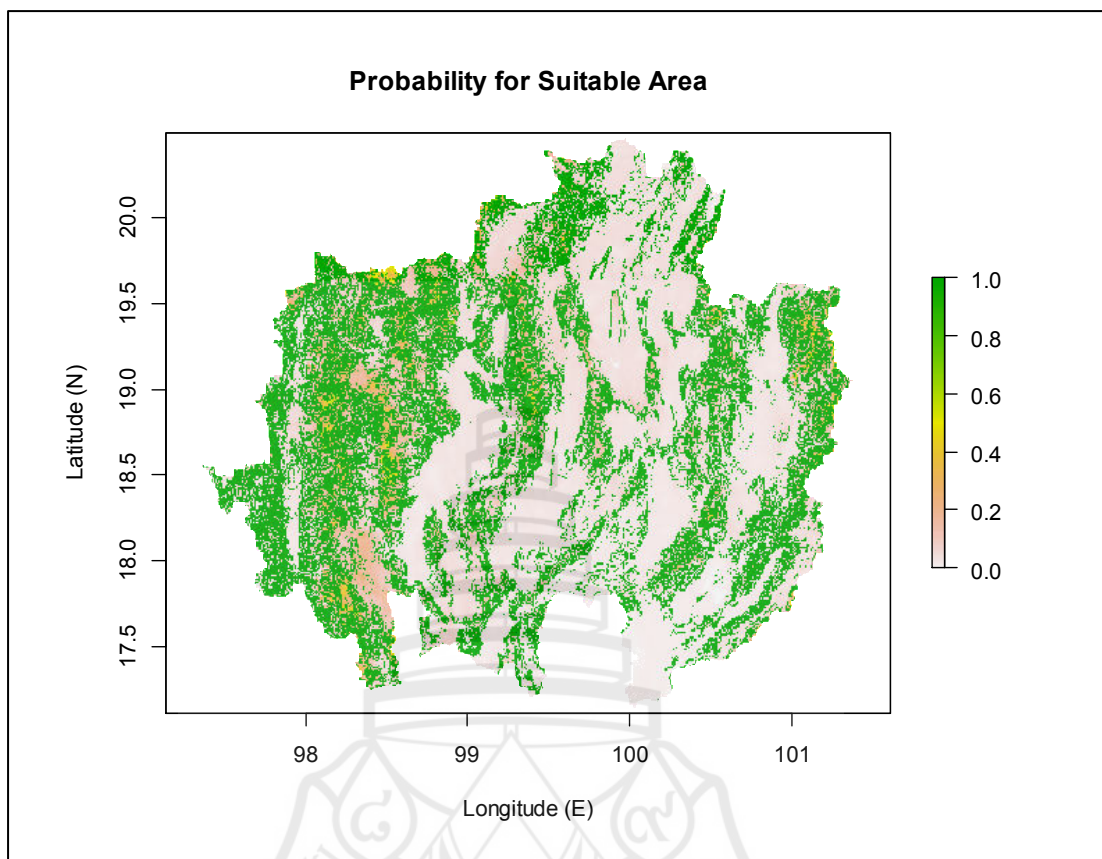
$$\text{Logit}(\mu) = -18.430 + 0.003\text{DEM} - 0.093\text{Distance} + 14.400\text{Class1} \\ + 14.420\text{Class2} + 14.360\text{Class5} + 13.560\text{Class6}$$

จากสมการพบว่าตัวแปร DEM, Class1, Class2, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวกซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Distance มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกพบว่าตัวแปรความสูงจากระดับน้ำทะเล (DEM) มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 โดยมีผลเชิงบวก หมายความว่าความสูงจากระดับน้ำทะเลถือเป็นปัจจัยทางภูมิศาสตร์ที่มีความสำคัญมาก ยิ่งพื้นที่อยู่สูงเท่าไรโอกาสที่จะค้นพบต้นชาป่ายิ่งมีมากขึ้น นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ส่วนตัวแปรอื่นๆ เมื่อทดสอบตัวแปรแต่ละตัวแล้วพบว่าไม่มีนัยสำคัญทางสถิติ แต่ก็ไม่ได้หมายความว่าตัวแปรเหล่านี้ไม่มีผลต่อแบบจำลอง โดยตัวแปรเหล่านี้กลับถูกคัดเลือกโดยวิธีการ Stepwise ซึ่งอาจเนื่องมาจากอิทธิพลจากตัวแปรตัวอื่นๆ โดยตัวแปรการจัดกลุ่มของโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5 และ Class6) มีผลเชิงบวก นั่นคือโครงสร้างทางธรณีวิทยาของชนิดหินไม่มีผลต่อแบบจำลองแต่พื้นที่จะต้องมีความลาดชันไม่เกิน 30 องศา โอกาสที่จะค้นพบต้นชาป่ามากยิ่งขึ้น

ในขณะที่ตัวแปรระยะห่างจากแม่น้ำ (Distance) มีผลเชิงลบ หมายความว่ายิ่งพื้นที่อยู่ห่างจากแม่น้ำเท่าไรโอกาสที่จะค้นพบต้นชาป่ายิ่งน้อยลง



ภาพที่ ง1 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

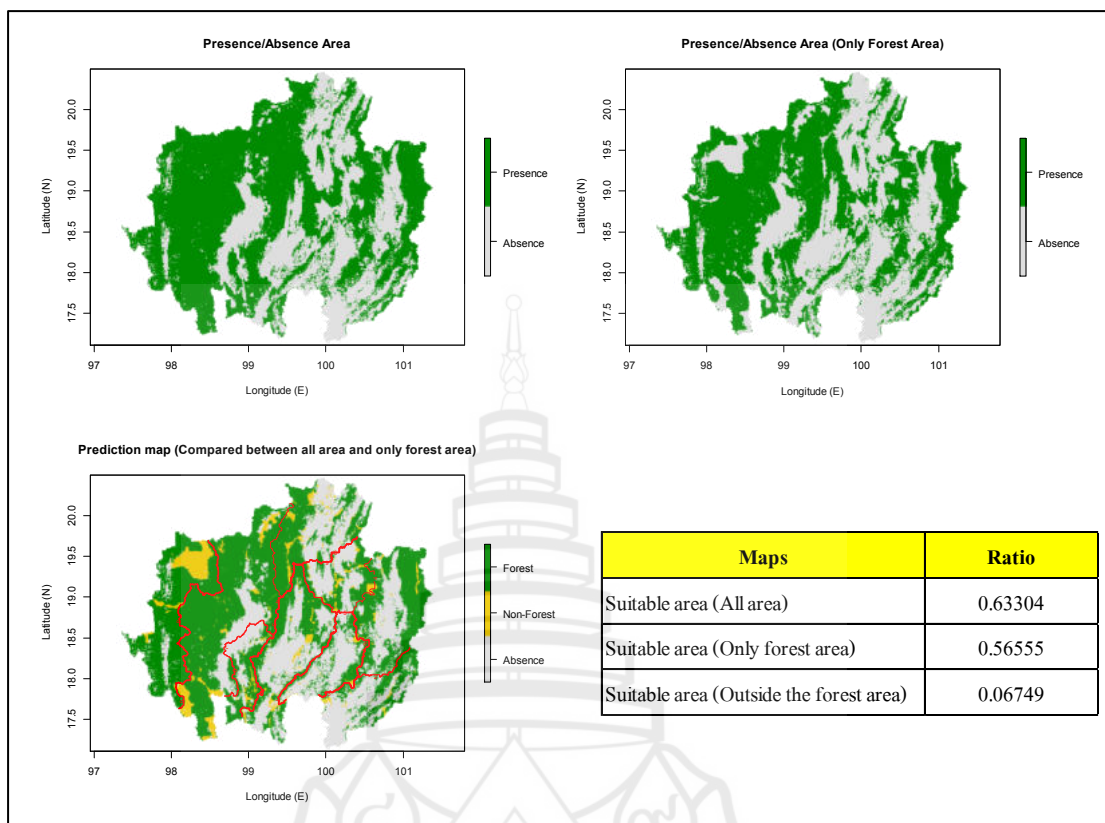
จากภาพที่ ง1 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ ง2 แสดงค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	295
AUC	0.70214
max TPR+TNR	-2.93869
Threshold	0.05027

จากตารางที่ ง2 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absence ใช้ทั้งหมด 295 จุด มีค่า AUC เท่ากับ 0.70214 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 70.21 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.05027 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ ๖2 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ ๖2 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 63 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 57 ของพื้นที่ทั้งหมดโดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิศาสตร์จะมีความเหมาะสมก็ตาม ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตป่ามีอยู่ประมาณร้อยละ 2 ของพื้นที่ทั้งหมดส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอน ดังแสดงในภาพซ้ายมือล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดง

แสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าที่อยู่ในเขตพื้นที่ป่าพบว่า พื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่แถวๆ จังหวัดเชียงใหม่ เชียงรายและแม่ฮ่องสอนเป็นส่วนใหญ่ โดยพื้นที่ส่วนใหญ่เป็นพื้นที่ป่า

ตารางที่ 33 แบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test	
Intercept	-45.503	14.657	-3.105	**
Humidity	-0.133	0.054	-2.462	*
Bio3	0.841	0.216	3.890	***
Bio4	0.007	0.001	4.022	***
Bio8	0.074	0.052	1.404	
Bio10	-0.100	0.050	-2.019	*
Bio12	0.022	0.005	4.160	***
Bio14	-0.108	0.058	-1.872	.
Bio16	-0.041	0.009	-4.757	***
Null Deviance = 892.55, Residual Deviance = 718.50				
Pseudo-R ² = 19.50, AIC = 736.50				

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 3 แสดงแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือ ความชื้นสัมพัทธ์ (Humidity), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) และปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16)

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 892.55 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 718.50 ซึ่งสามารถคำนวณค่า Pseudo- R^2 ได้เท่ากับ 19.50 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 19.50 ได้สมการพยากรณ์ดังนี้

$$\text{Logit}(\mu) = -45.503 - 0.133\text{Humidity} + 0.841\text{Bio3} + 0.007\text{Bio4} + 0.074\text{Bio8} \\ - 0.100\text{Bio10} + 0.022\text{Bio12} - 0.108\text{Bio14} - 0.041\text{Bio16}$$

จากสมการพบว่าตัวแปร Bio3, Bio4, Bio8 และ Bio12 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Humidity, Bio10, Bio14 และ Bio16 มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio3, Bio4 และ Bio12 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 สามารถอธิบายได้ว่า ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างช่วงความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลเชิงบวก หมายความว่าถ้าหากค่า Isothermality (Bio3) มีค่ามากหรือมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็จะเพิ่มขึ้น เช่นเดียวกันกับตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4) ซึ่งคำนวณมาจากค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) คูณด้วย 100 และปริมาณน้ำฝนรายปี (Bio12) นั่นคือพื้นที่ใดที่มีความผันแปรของอุณหภูมิมากหรือมีการเปลี่ยนแปลงของอุณหภูมิที่ค่อนข้างแตกต่าง ซึ่งหมายถึงพื้นที่ที่มีช่วงของฤดูกาลโอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นหรือพื้นที่ใดที่มีปริมาณน้ำฝนรายปีที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วยเช่นกัน นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ส่วนตัวแปรอุณหภูมิเฉลี่ยในไตรมาสถัดสุด (Bio8) ซึ่งถึงแม้ว่าจะทำการทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในแบบโดยอิทธิพลจากตัวแปรตัวอื่น ๆ โดยมีผลเชิงบวก หมายความว่าพื้นที่ใดในไตรมาสถัดสุดที่มีอุณหภูมิโดยเฉลี่ยสูงๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นเช่นกัน

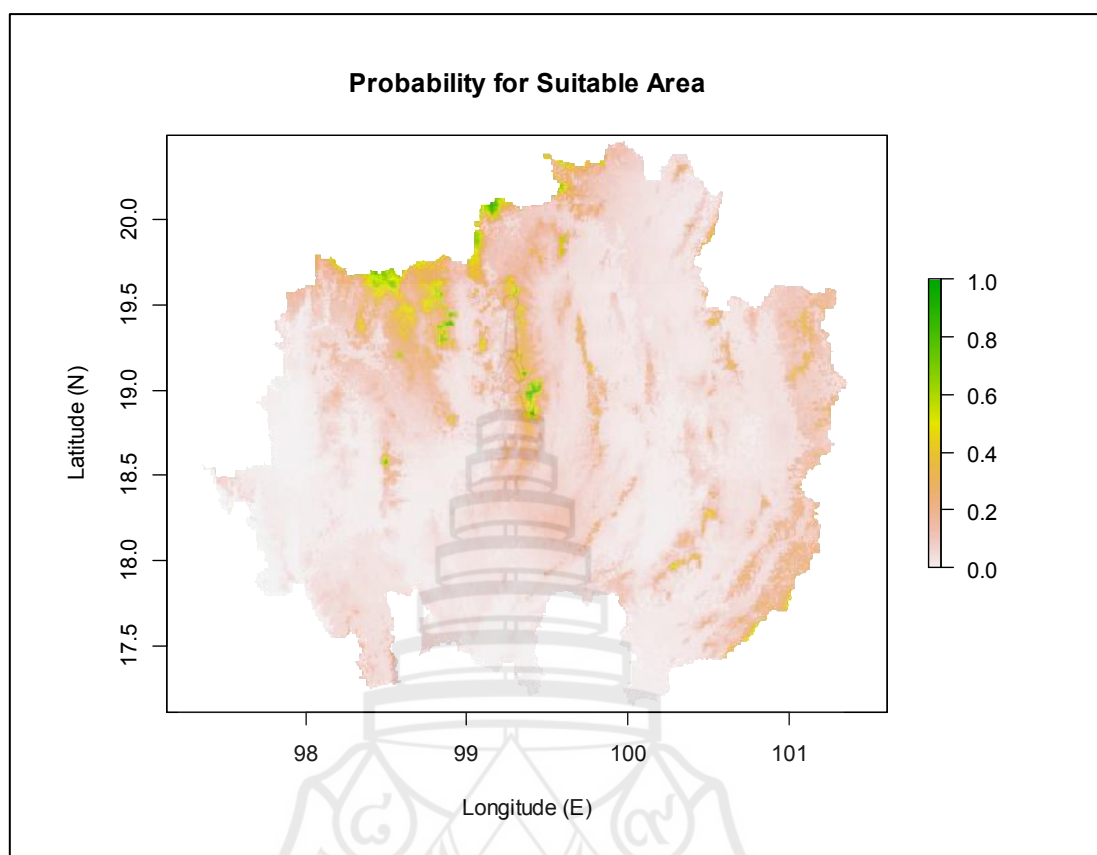
เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Bio16 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Humidity และ Bio10 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.05 และตัวแปร Bio14 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.10 สามารถอธิบายได้ดังต่อไปนี้

ตัวแปรปริมาณน้ำฝนในไตรมาสถัดสุด (Bio16) มีผลเชิงลบ หมายความว่าพื้นที่ใดที่มีปริมาณน้ำฝนในไตรมาสถัดสุดมากเกินไป โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลงได้

ตัวแปรความชื้นสัมพัทธ์ (Humidity) มีผลเชิงลบ หมายความว่าพื้นที่ใดมีความชื้นสัมพัทธ์สูงๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลง

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสถัดสุด (Bio10) ซึ่งมีผลเชิงลบ หมายความว่าพื้นที่ใดมีอุณหภูมิเฉลี่ยสูงๆ ในไตรมาสถัดสุด โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลง

ตัวแปรปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14) มีผลเชิงลบ หมายความว่าพื้นที่ใดในช่วงเดือนที่แห้งแล้งที่สุดมีปริมาณน้ำฝนมากโอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลง



ภาพที่ ๓3 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

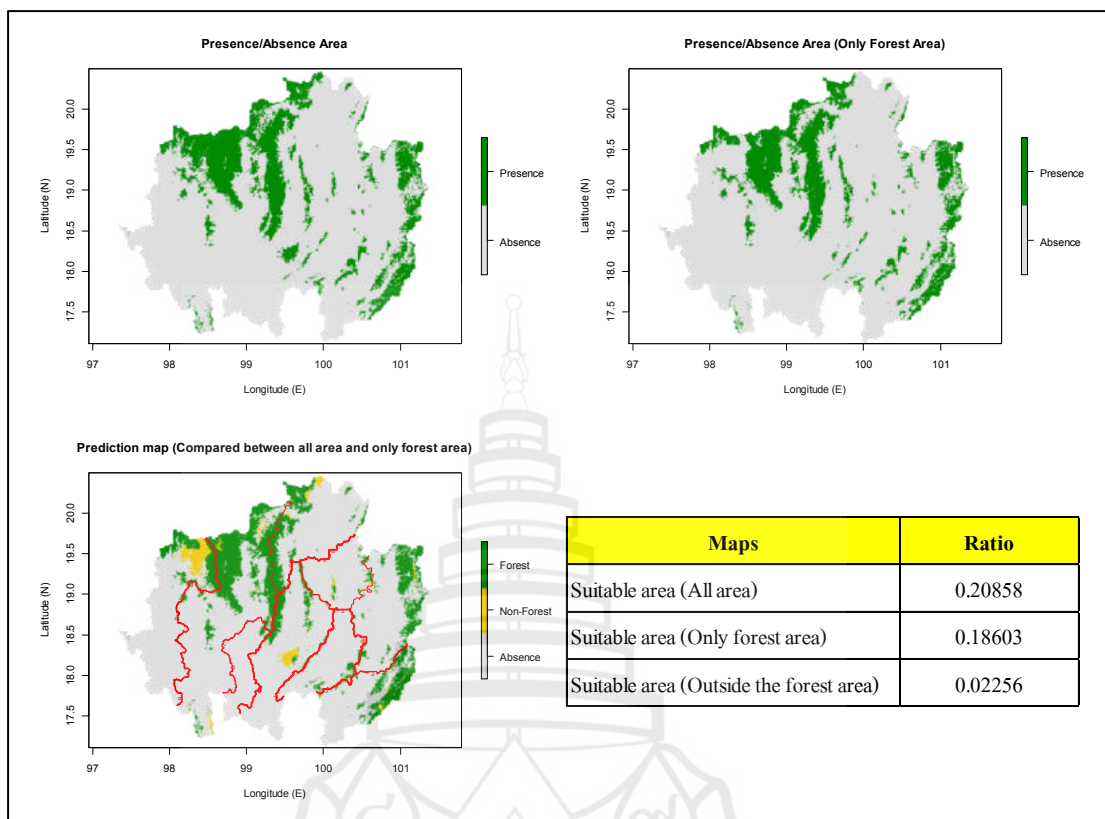
จากภาพที่ ๓3 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ ๔4 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้าน
ภูมิอากาศ

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	295
AUC	0.76574
max TPR+TNR	-2.05882
Threshold	0.11316

จากตารางที่ ๔4 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้าน
ภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการ
ตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด
295 จุด มีค่า AUC เท่ากับ 0.76574 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ
76.57 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative
Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.11316
หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่า
เท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่า
เท่ากับ 1



ภาพที่ ๔4 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ ๔4 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 21 ของพื้นที่ทั้งหมด ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 19 ของพื้นที่ทั้งหมด โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีมีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิศาสตร์จะมีความเหมาะสมก็ตาม โดยพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตป่ามีพื้นที่ประมาณร้อยละ 2 ของพื้นที่ทั้งหมด ส่วนใหญ่อยู่ในเขตจังหวัดแม่ฮ่องสอน ดังแสดงในภาพซ้ายมือล่าง (พื้นที่สีเหลือง)

โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในเขตพื้นที่ป่าพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกเฉียงเหนือของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่แถวๆ จังหวัดเชียงใหม่รวมถึงป่ารอยต่อระหว่างเชียงใหม่และแม่ฮ่องสอนตอนบน และป่ารอยต่อระหว่างเชียงใหม่และเชียงราย

2. แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-Absence ภายนอกพื้นที่รอบจุด Presence

ตารางที่ ๖5 แบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปร โดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test
Intercept	-4.946	0.392	-12.609 ***
DEM	0.003	0.000	9.967 ***
Aspect	-0.001	0.001	-1.465
Class1	0.681	0.277	2.455 *
Class5	1.150	0.300	3.829 ***
Class6	1.347	0.438	3.073 **
Null Deviance = 915.74, Residual Deviance = 765.79			
Pseudo-R ² = 16.37, AIC = 777.79			

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05

จากตารางที่ 5 แสดงแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือความสูงจากระดับน้ำทะเล (DEM), ทิศด้านลาด (Aspect) และกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class5 และ Class6) ซึ่งมีตัวแปร DEM และ Class5 มีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Class6 มีนัยสำคัญทางสถิติที่ 0.01 และตัวแปร Class1 มีนัยสำคัญทางสถิติที่ 0.05

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 915.74 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 765.79 ซึ่งสามารถคำนวณค่า Pseudo- R^2 ได้เท่ากับ 16.37 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 16.37 ได้สมการพยากรณ์ดังนี้

$$\text{Logit}(\mu) = -4.946 + 0.003\text{DEM} - 0.001\text{Aspect} + 0.681\text{Class1} \\ + 1.150\text{Class5} + 1.347\text{Class6}$$

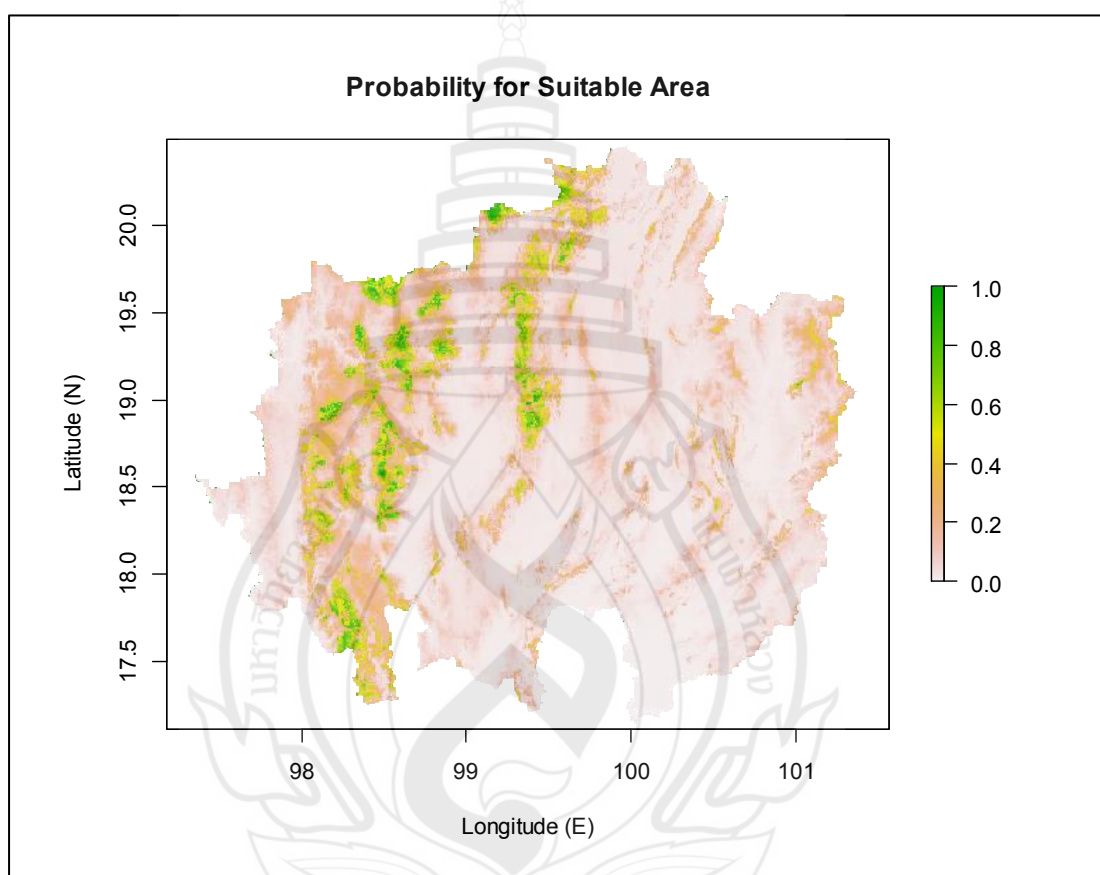
จากสมการพบว่าตัวแปร DEM, Class1, Class5 และ Class6 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Aspect มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร DEM และ Class5 มีผลเชิงบวกต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.001 หมายความว่าความสูงจากระดับน้ำทะเล (DEM) ถือเป็นปัจจัยทางภูมิศาสตร์ที่มีความสำคัญ ยิ่งพื้นที่ที่อยู่สูงเท่าไรโอกาสที่จะค้นพบต้นชาป่ายิ่งมีมากขึ้น นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่าหรือพื้นที่ที่ปรากฏหินอัคนีและมีความชัน 0 ถึง 15 องศาโอกาสที่จะค้นพบต้นชาป่ายิ่งมีมากขึ้นเช่นกัน

ตัวแปร Class6 มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.01 หมายความว่าพื้นที่ที่ปรากฏหินอัคนีและมีความชัน 15 ถึง 30 องศาโอกาสที่จะค้นพบต้นชาป่ามีมากขึ้น เช่นเดียวกับตัวแปร Class1 ซึ่งมีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติที่ 0.05 นั่นคือ พื้นที่ที่ปรากฏหินตะกอนและหินแปรโดยมีความชัน 0 ถึง 15 องศาโอกาสที่จะค้นพบต้นชาป่าก็มีมากขึ้นเช่นกัน

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Aspect มีผลเชิงลบต่อแบบจำลอง แต่เมื่อทดสอบตัวแปรแล้วพบว่าไม่มีนัยสำคัญทางสถิติ ก็ไม่ได้หมายความว่าตัวแปร Aspect ไม่มีผลต่อแบบจำลอง แต่ตัวแปร Aspect อาจถูกคัดเลือกโดยวิธีการ Stepwise โดยมาจากอิทธิพลจากตัวแปรตัวอื่นๆ ซึ่งสามารถอธิบายได้ว่า ตัวแปรทิศด้านลาด (Aspect) มีผลเชิงลบ นั่น

คือตัวแปรทิศด้านลาด (Aspect) มีค่าตั้งแต่ 0 ถึง 360 องศา ซึ่งถูกแบ่งออกเป็น 8 ทิศ ทิศละ 45 องศา ทิศเหนือ, ตะวันออกเฉียงเหนือ, ตะวันออก, ตะวันออกเฉียงใต้, ใต้, ตะวันตกเฉียงใต้, ตะวันตก และตะวันตกเฉียงเหนือตามลำดับ ยิ่งมีค่ามาก Aspect ทิศด้านลาดก็จะหันหน้าไปทางทิศตะวันตกเรื่อยๆ ดังนั้นถ้าทิศด้านลาด (Aspect) มีค่าสูงขึ้นคือพื้นที่ที่มีการหันหน้าไปทางทิศตะวันตกมากขึ้น โอกาสที่จะค้นพบต้นชาป่าก็มีน้อยลง



ภาพที่ ๓๕ แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

จากภาพที่ ๓๕ แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทาง

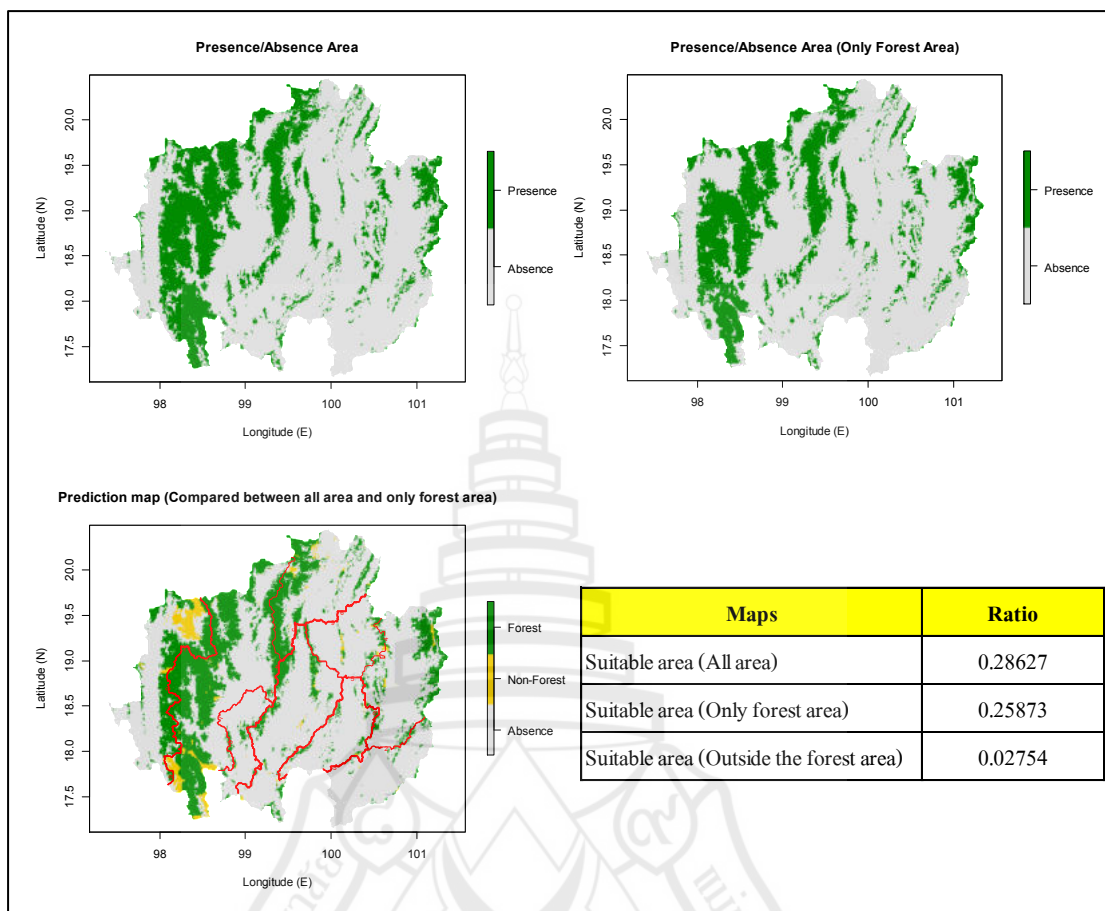
ภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ ๖ ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	37
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	369
AUC	0.79426
max TPR+TNR	-2.18505
Threshold	0.10110

จากตารางที่ ๖ แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 37 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด 369 จุด มีค่า AUC เท่ากับ 0.79426 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 79.43 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.10110 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 36 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 36 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 29 ของพื้นที่ ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 26 ของพื้นที่ โดยพิจารณาจากฟิสิกส์ไดพอยกรณออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ฟิสิกส์นั้น ไม่ได้อยู่ในบริเวณป่า ก็จะทำให้ฟิสิกส์นั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิศาสตร์จะมีความเหมาะสมก็ตาม ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตพื้นที่ป่ามีอยู่ประมาณร้อยละ 3 ของพื้นที่ ส่วนใหญ่

อยู่ในจังหวัดแม่ฮ่องสอน ดังแสดงในภาพซ้ายมือล่าง โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าที่อยู่ในเขตป่าพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกของพื้นที่ ซึ่งอยู่แถว ๆ ปารอยต่อระหว่างจังหวัดเชียงใหม่และแม่ฮ่องสอนเป็นส่วนใหญ่และบางส่วนอยู่ระหว่างปารอยต่อระหว่างจังหวัดเชียงใหม่และเชียงราย

ตารางที่ ง7 แบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test
Intercept	-58.335	28.748	-2.029 *
Humidity	-0.194	0.077	-2.51 *
Bio2	-0.789	0.236	-3.347 ***
Bio3	1.477	0.376	3.93 ***
Bio4	0.013	0.007	1.927 .
Bio5	-0.216	0.144	-1.501
Bio7	0.463	0.147	3.152 **
Bio8	0.144	0.085	1.69 .
Bio10	-0.389	0.246	-1.58
Bio11	0.438	0.273	1.604
Bio12	0.009	0.003	2.615 **

ตารางที่ ๗ (ต่อ)

Factors	Estimate	Standard Error	Z Test
Bio13	-0.047	0.011	-4.312 ***
Bio14	-0.221	0.109	-2.024 *
Bio17	0.384	0.179	2.149 *
Bio19	-0.222	0.166	-1.341
Null Deviance = 915.74, Residual Deviance = 565.26			
Pseudo-R ² = 38.27, AIC = 595.26			

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ๗ แสดงแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือ ความชื้นสัมพัทธ์ (Humidity), ช่วงความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2), Isothermality (Bio3), ช่วงฤดูกาลของอุณหภูมิ (Bio4), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), ช่วงความแตกต่างของอุณหภูมิรายปี (Bio7), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่ร้อนที่สุด (Bio10), อุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด (Bio11), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) และปริมาณน้ำฝนในไตรมาสที่หนาวที่สุด (Bio19) ซึ่งมีตัวแปร Bio2, Bio3 และ Bio13 มีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Bio7 และ Bio12 มีนัยสำคัญทางสถิติที่ 0.01 ตัวแปร Humidity, Bio14 และ Bio17 มีนัยสำคัญทางสถิติที่ 0.05 ตัวแปร Bio4 และ Bio8 มีนัยสำคัญทางสถิติที่ 0.10

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 915.74 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 565.26 ซึ่งสามารถคำนวณค่า Pseudo- R^2 ได้เท่ากับ 38.27 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 38.27 ได้สมการพยากรณ์ดังนี้

$$\begin{aligned} \text{Logit}(\mu) = & -58.335 - 0.194\text{Humidity} - 0.789\text{Bio2} + 1.477\text{Bio3} + 0.013\text{Bio4} \\ & -0.216\text{Bio5} + 0.463\text{Bio7} + 0.144\text{Bio8} - 0.389\text{Bio10} + 0.438\text{Bio11} \\ & +0.009\text{Bio12} - 0.047\text{Bio13} - 0.221\text{Bio14} + 0.384\text{Bio17} - 0.222\text{Bio19} \end{aligned}$$

จากสมการพบว่าตัวแปร Bio3, Bio4, Bio7, Bio8, Bio11, Bio12 และ Bio17 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวกซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Humidity, Bio5, Bio10, Bio13, Bio14 และ Bio19 มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio3 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญ 0.001 สามารถอธิบายได้ว่า ตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างช่วงความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลเชิงบวก นั่นคือยิ่งค่า Isothermality (Bio3) มีค่ามากหรือมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

ตัวแปร Bio7 และ Bio12 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญ 0.01 สามารถอธิบายได้ว่า ยิ่งตัวแปรช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีค่าสูงขึ้น โอกาสที่จะค้นพบต้นชาป่ามีมากขึ้น หมายความว่าพื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุดกับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุดมาก พื้นที่นั้นมีโอกาสที่จะค้นพบต้นชาป่ามากยิ่งขึ้น

เช่นเดียวกับตัวแปร Bio12 ซึ่งอธิบายได้ว่าในพื้นที่ที่มีปริมาณน้ำฝนรายปี (Bio12) ที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วยเช่นกัน นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ตัวแปร Bio17 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญ 0.05 สามารถอธิบายได้ว่าพื้นที่ใดที่มีปริมาณน้ำฝนสูงๆ ในช่วงไตรมาสที่แห้งแล้งที่สุด จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่มากขึ้น

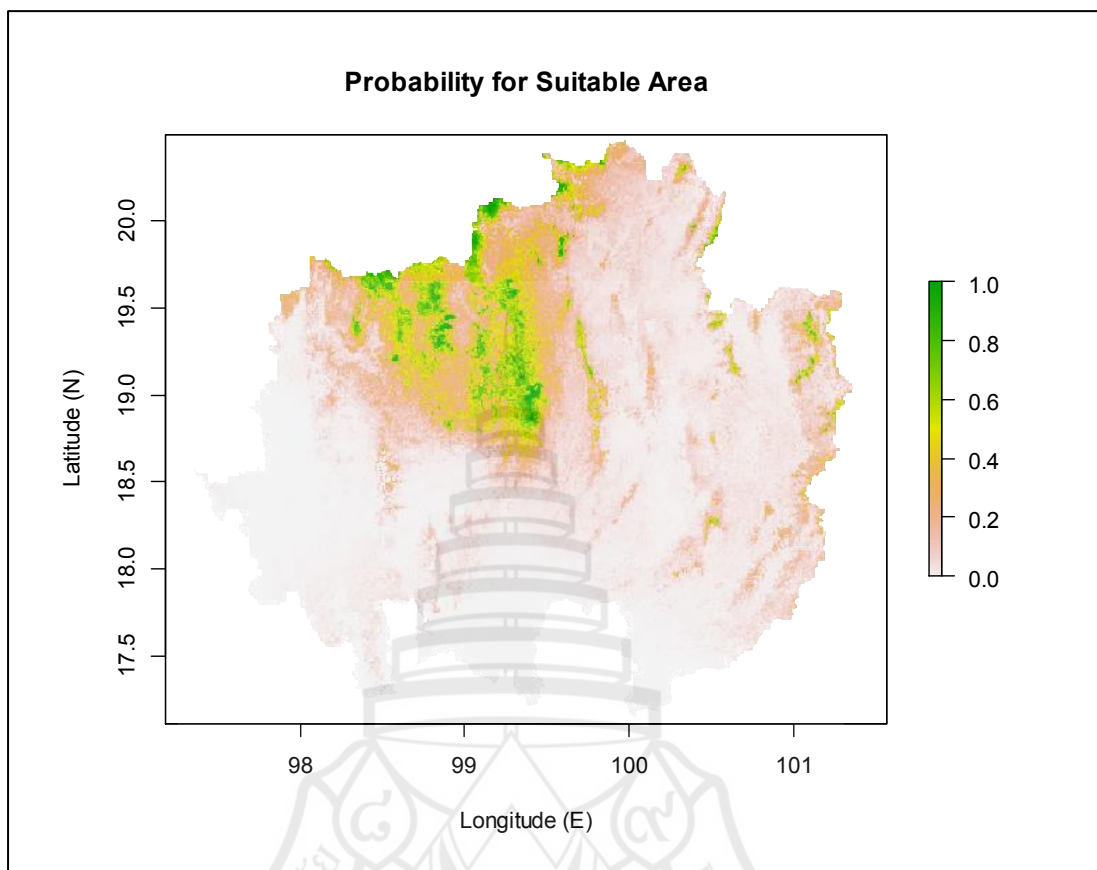
ตัวแปร Bio4 และ Bio8 ผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญ 0.10 ซึ่งสามารถอธิบายได้ดังนี้ ตัวแปรช่วงฤดูกาลของอุณหภูมิ (Bio4) คือค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) คูณด้วย 100 นั่นคือในพื้นที่ที่มีความผันแปรของอุณหภูมิมากหรือมีการเปลี่ยนแปลงของอุณหภูมิที่ค่อนข้างแตกต่าง ซึ่งหมายถึงพื้นที่ที่มีช่วงของฤดูกาลโอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น เช่นเดียวกับพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่มีมากขึ้น

ส่วนตัวแปร Bio11 หรือตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่หนาวที่สุด ซึ่งถึงแม้ว่าจะทำการทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในตัวแบบโดยอิทธิพลจากตัวแปรตัวอื่นๆ โดยมีผลเชิงบวก หมายความว่าพื้นที่ใดในช่วงไตรมาสที่หนาวที่สุด มีอุณหภูมิที่ไม่หนาวเกินไปหรือมีอุณหภูมิสูงๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีมากขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio2 และ Bio13 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญ 0.001ซึ่งสามารถอธิบายได้ดังนี้ พื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิระหว่างวันโดยเฉลี่ย (Bio2) มาก ๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงด้วย เช่นเดียวกับพื้นที่ใดมีปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง

ตัวแปร Humidity และ Bio14 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญ 0.01ซึ่งสามารถอธิบายได้ดังนี้ พื้นที่ใดมีความชื้นสัมพัทธ์ (Humidity) สูง ๆ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็ลดลง เช่นเดียวกับพื้นที่ใดมีปริมาณน้ำฝนในเดือนที่ฝนแห้งแล้งที่สุด (Bio14) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง

ส่วนตัวแปร Bio5, Bio10 และ Bio19 ถึงแม้ว่าจะทำการทดสอบตัวแปรแล้วพบว่าไม่มีผลต่อแบบจำลองอย่างมีนัยสำคัญทางสถิติแต่ก็อาจถูกคัดเลือกเข้ามาในตัวแบบโดยอิทธิพลจากตัวแปรตัวอื่นๆ โดยมีผลเชิงลบ ซึ่งสามารถอธิบายได้ดังนี้ ตัวแปร Bio5 มีผลเชิงลบหมายความว่าพื้นที่ใดในช่วงเดือนที่ร้อนที่สุดแล้วมีอุณหภูมิสูงๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีน้อยลง เช่นเดียวกับตัวแปร Bio10 ที่มีผลเชิงลบ หมายความว่าพื้นที่ใดในช่วงไตรมาสที่ร้อนที่สุดแล้วมีอุณหภูมิสูงๆ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็มีน้อยลง ส่วนตัวแปร Bio19 ก็มีผลเชิงลบ เช่นเดียวกัน หมายความว่าพื้นที่ใดในช่วงไตรมาสที่หนาวที่สุดแล้วมีปริมาณน้ำฝนที่มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงเช่นเดียวกัน



ภาพที่ ง7 แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

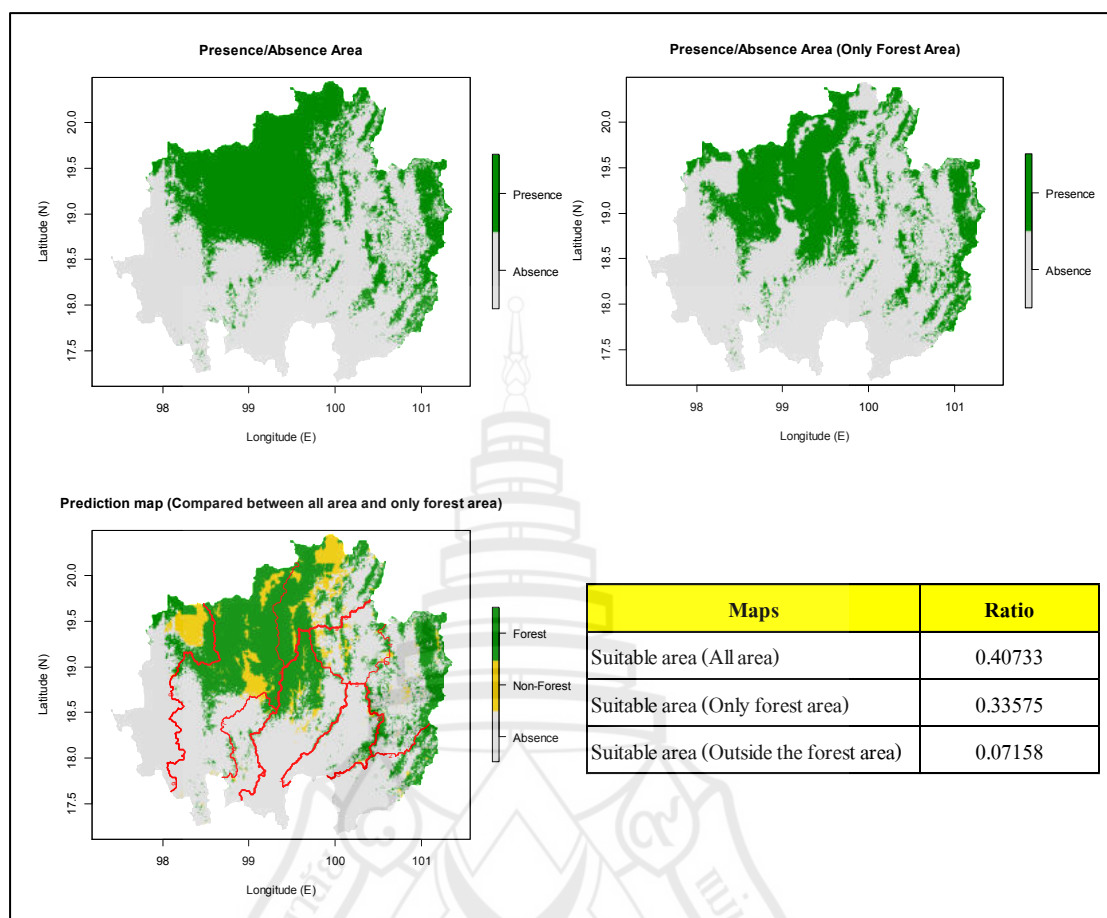
จากภาพที่ ง7 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่ว่าพื้นที่นั้นมีลักษณะทางภูมิอากาศที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ ๖๘ ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้าน
ภูมิอากาศ

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	37
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	369
AUC	0.89446
max TPR+TNR	-2.98575
Threshold	0.04807

จากตารางที่ ๖๘ แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้าน
ภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการ
ตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 37 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด
369 จุด มีค่า AUC เท่ากับ 0.89446 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ
89.45 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative
Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.04807
หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่า
เท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่า
เท่ากับ 1



ภาพที่ 38 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 38 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมีบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 41 ของพื้นที่ ส่วนภาพด้านขวามีบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 34 ของพื้นที่ โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิอากาศจะมีความเหมาะสมก็ตาม โดยพื้นที่

ที่คาดว่ามีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในพื้นที่ป่ามีอยู่ประมาณร้อยละ 7 ของพื้นที่ ส่วนใหญ่อยู่ในจังหวัด แม่ฮ่องสอน เชียงใหม่และเชียงราย ดังแสดงในภาพล่างซ้าย (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าเฉพาะในเขตพื้นที่ป่าพบว่า พื้นที่ที่คาดว่ามีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกเฉียงเหนือของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่แถว ๆ จังหวัดเชียงใหม่ เชียงรายและลำปาง โดยกระจายอย่างหนาแน่นบริเวณป่ารอยต่อของทั้ง 3 จังหวัด และมีบางส่วนอยู่ทางตะวันออกของพื้นที่ แถวป่าชายแดนในจังหวัดน่าน

3. แบบจำลอง GLMs ที่เกิดจากการสุ่มจุด Pseudo-absence แบบ 2 ขั้นตอน

ตารางที่ ๑๙ แบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปร โดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test
Intercept	-20.090	607.600	-0.033
DEM	0.003	0.000	10.879 ***
Aspect	0.002	0.001	1.748 .
Distance	-0.075	0.053	-1.409
Class1	15.550	607.600	0.026
Class2	15.640	607.600	0.026
Class5	16.010	607.600	0.026
Class6	15.780	607.600	0.026
Class7	-63.000	2426.000	-0.026
Null Deviance = 975.20 , Residual Deviance = 800.05			
Pseudo-R ² = 17.96, AIC = 818.05			

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001

. หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ 9 แสดงแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือความสูงจากระดับน้ำทะเล (DEM), ทิศด้านลาด (Aspect), ระยะห่างจากแม่น้ำ (Distance) และกลุ่มโครงสร้างทางธรณีวิทยาของชนิดหินและประเภทของความลาดชัน (Class1, Class2, Class5, Class6 และ Class7) ซึ่งมีตัวแปร DEM มีนัยสำคัญทางสถิติที่ 0.001 และตัวแปร Aspect มีนัยสำคัญทางสถิติที่ 0.10

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 975.20 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 800.05 ซึ่งสามารถคำนวณหาค่า Pseudo- R^2 ได้เท่ากับ 17.96 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 17.96 ได้สมการพยากรณ์ดังนี้

$$\text{Logit}(\mu) = -20.090 + 0.003\text{DEM} + 0.002\text{Aspect} - 0.075\text{Distance} + 15.550\text{Class1} \\ + 15.640\text{Class2} + 16.010\text{Class5} + 15.780\text{Class6} - 63.000\text{Class7}$$

จากสมการพบว่าตัวแปร DEM, Aspect, Class1, Class2, Class5 และ Class6 มีค่าเป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวก ซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Distance และ Class7 มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบ ซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

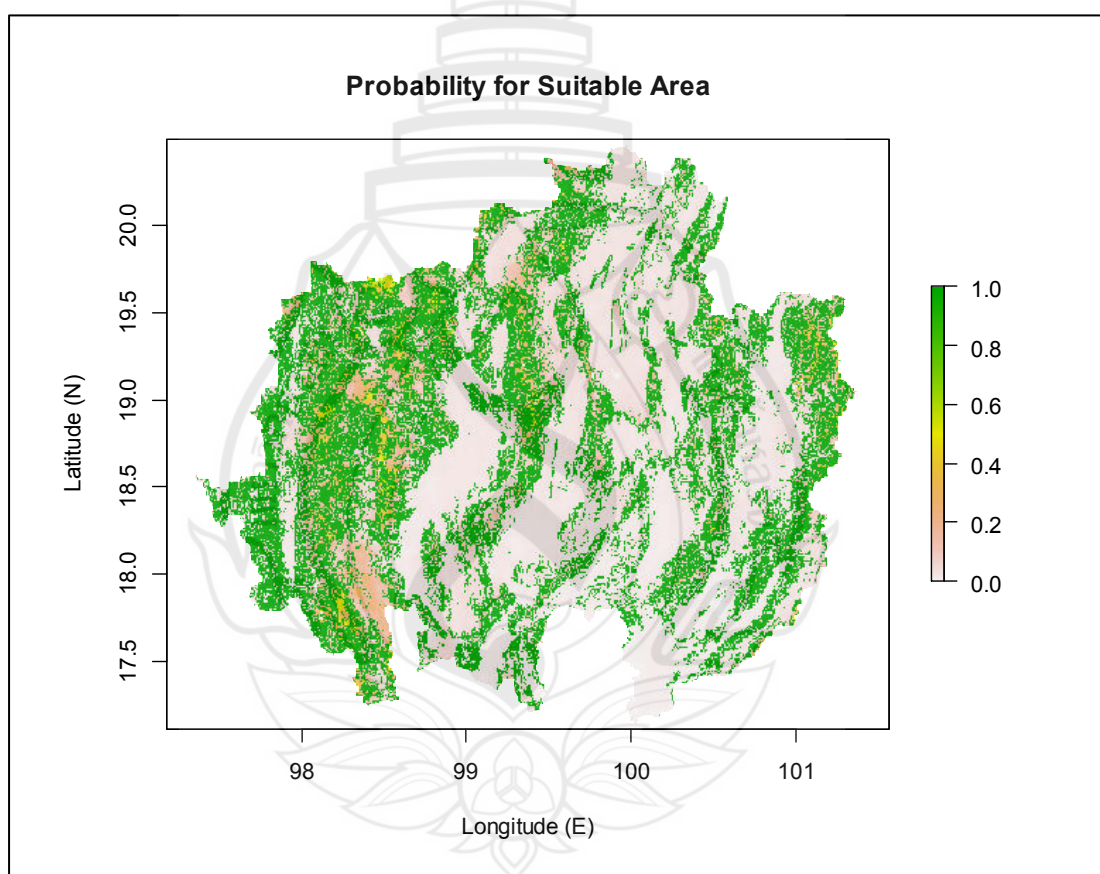
เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร DEM มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.001 อธิบายได้ว่าความสูงจากระดับน้ำทะเลถือเป็นปัจจัยทางภูมิศาสตร์ที่มีความสำคัญ ยิ่งพื้นที่อยู่สูงเท่าไรโอกาสที่จะค้นพบต้นชาป่ายิ่งมีมากขึ้น นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ตัวแปร Aspect มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.10 อธิบายได้ว่าตัวแปรทิศด้านลาด (Aspect) มีค่าตั้งแต่ 0 ถึง 360 องศา ซึ่งถูกแบ่งออกเป็น 8 ทิศ ทิศละ 45 องศา ทิศเหนือ, ตะวันออกเฉียงเหนือ, ตะวันออก, ตะวันออกเฉียงใต้, ใต้, ตะวันตกเฉียงใต้, ตะวันตก และตะวันตกเฉียงเหนือตามลำดับ ยิ่งค่า Aspect มีค่ามาก ทิศด้านลาดก็จะหันหน้าไปทางทิศตะวันตกเรื่อยๆ ดังนั้นถ้า Aspect มีค่าสูงขึ้นคือพื้นที่ที่มีการหันหน้าไปทางทิศตะวันตกเฉียงเหนือมากขึ้น โอกาสที่จะค้นพบต้นชาป่าก็มีมากขึ้น

ส่วนตัวแปร Class1, Class2, Class5 และ Class6 ถึงแม้ว่าทดสอบตัวแปรแล้วจะไม่มีนัยสำคัญทางสถิติ ก็ไม่ได้หมายความว่าตัวแปรเหล่านี้ไม่มีผลต่อแบบจำลอง แต่ตัวแปรเหล่านี้กลับถูกคัดเลือกโดยวิธีการ Stepwise ซึ่งอาจเนื่องมาจากอิทธิพลจากตัวแปรตัวอื่นๆ อธิบายได้ว่าพื้นที่มี

ความลาดชันไม่เกิน 30 องศาและมีโครงสร้างทางธรณีวิทยาของชนิดหินประเภทใดก็ได้ จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่มากยิ่งขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Distance และ Class7 มีผลต่อแบบจำลองเชิงลบ แต่เมื่อทดสอบตัวแปรแล้วพบว่าไม่มีนัยสำคัญทางสถิติ ก็ไม่ได้หมายความว่าตัวแปรเหล่านี้ไม่มีผลต่อแบบจำลอง แต่ตัวแปรเหล่านี้กลับถูกคัดเลือกโดยวิธีการ Stepwise ซึ่งอาจเนื่องมาจากอิทธิพลจากตัวแปรตัวอื่นๆ อธิบายได้ว่าพื้นที่ใดที่อยู่ห่างจากแม่น้ำมากเท่าไร โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลงหรือพื้นที่ใดที่ปรากฏหินอัคนีและความลาดชันเกิน 30 องศา พื้นที่นั้นจะมีโอกาสค้นพบต้นชาป่าน้อยลงเช่นเดียวกัน



ภาพที่ ๑๑ แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

จากภาพที่ 9 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่พื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

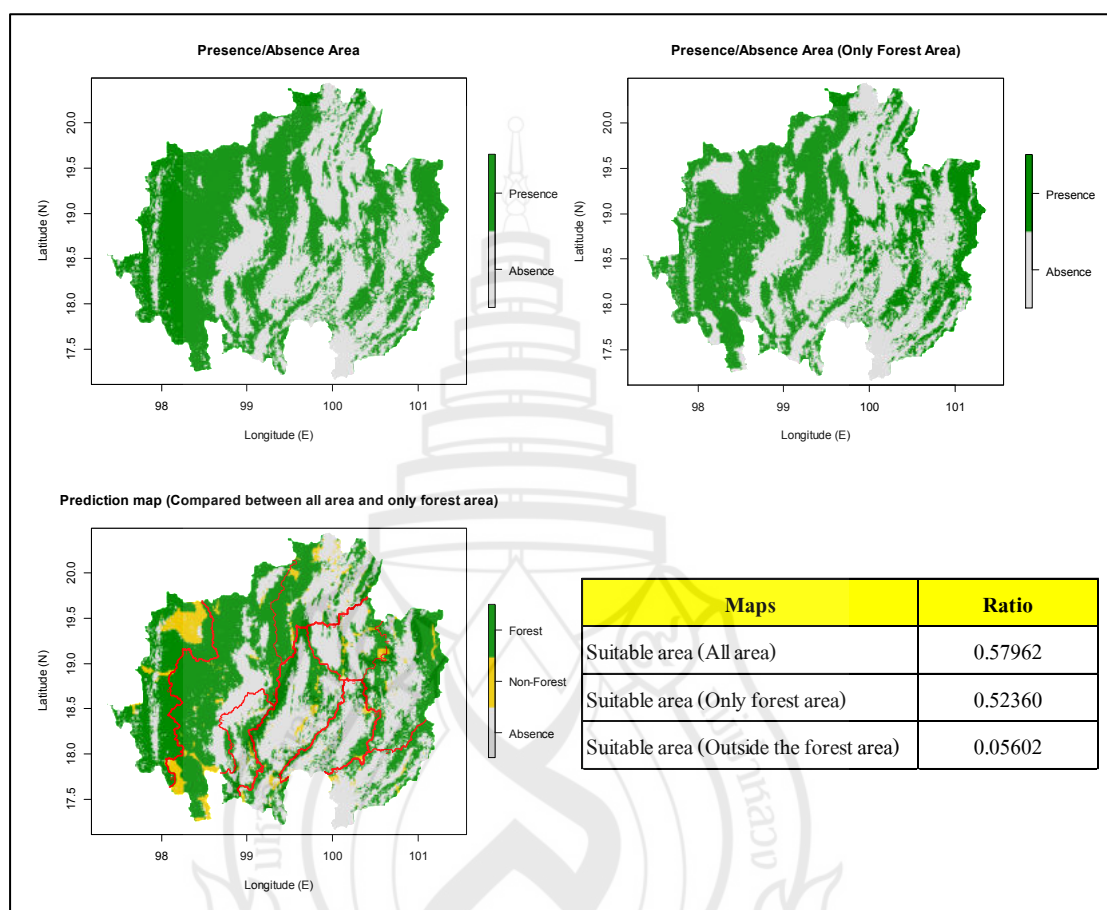
หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึงพื้นที่ Absence ค่า 1 หมายถึงพื้นที่ Presence

ตารางที่ 10 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้านภูมิศาสตร์

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	366
AUC	0.78099
max TPR+TNR	-2.82072
Threshold	0.05621

จากตารางที่ 10 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด 366 จุด มีค่า AUC เท่ากับ 0.78099 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 78.10 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหา Threshold ได้เท่ากับ 0.05621 หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่า

เท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่มีความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 10 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 10 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 58 ของพื้นที่ ส่วนภาพด้านขวามือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณป่าเท่านั้น ซึ่งพื้นที่นี้มี

อยู่ประมาณร้อยละ 52 ของพื้นที่โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิอากาศจะมีความเหมาะสมก็ตาม ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตป่ามีอยู่ประมาณร้อยละ 6 ของพื้นที่ ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอนและเชียงใหม่ตอนล่าง ดังแสดงในภาพซ้ายล่าง (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าเฉพาะในเขตพื้นที่ป่าพบว่าพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่แถวๆ จังหวัดแม่ฮ่องสอนและเชียงใหม่เป็นส่วนใหญ่ มีบางส่วนอยู่ในแนวป่ารอยต่อระหว่างจังหวัดเชียงใหม่ เชียงรายและลำปาง และบางส่วนกระจายอยู่ทางตะวันออกของพื้นที่ในเขตจังหวัดน่าน

ตารางที่ 11 แบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

Factors	Estimate	Standard Error	Z Test
Intercept	-137.900	28.140	-4.902 ***
Bio1	0.322	0.144	2.241 *
Bio3	0.609	0.224	2.717 **
Bio5	-0.687	0.118	-5.833 ***
Bio7	0.783	0.120	6.53 ***
Bio8	0.374	0.065	5.75 ***
Bio9	-0.056	0.031	-1.823 .
Bio12	0.030	0.008	3.747 ***
Bio13	0.035	0.019	1.848 .
Bio14	-0.698	0.141	-4.946 ***

ตารางที่ ง11 (ต่อ)

Factors	Estimate	Standard Error	Z Test
Bio16	-0.075	0.019	-3.872 ***
Bio17	0.205	0.041	5.023 ***

Null deviance = 975.20 Residual deviance = 503.31

Pseudo-R² = 48.39, AIC = 527.31

หมายเหตุ *** หมายถึง มีนัยสำคัญทางสถิติที่ 0.001
 ** หมายถึง มีนัยสำคัญทางสถิติที่ 0.01
 * หมายถึง มีนัยสำคัญทางสถิติที่ 0.05
 . หมายถึง มีนัยสำคัญทางสถิติที่ 0.10

จากตารางที่ ง11 แสดงแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise พบว่ามีตัวแปรเข้าไปในตัวแบบคือ อุณหภูมิเฉลี่ยรายปี (Bio1), Isothermality (Bio3), อุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5), ช่วงความแตกต่างของอุณหภูมิรายปี (Bio7), อุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8), อุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9), ปริมาณน้ำฝนรายปี (Bio12), ปริมาณน้ำฝนในเดือนที่ฝนตกชุกที่สุด (Bio13), ปริมาณน้ำฝนในเดือนที่แห้งแล้งที่สุด (Bio14), ปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) และน้ำฝนฝนไตรมาสที่แห้งแล้งที่สุด (Bio17) ซึ่งมีตัวแปร Bio5, Bio7, Bio8, Bio12, Bio14, Bio16 และ Bio17 มีนัยสำคัญทางสถิติที่ 0.001 ตัวแปร Bio3 มีนัยสำคัญทางสถิติที่ 0.01 ตัวแปร Bio1 มีนัยสำคัญทางสถิติที่ 0.05 และตัวแปร Bio9 และ Bio13 มีนัยสำคัญทางสถิติที่ 0.10

โดยมีค่า Null Deviance หรือค่า Log Likelihood สำหรับ Constant-Only Model (L_0) เท่ากับ 975.20 และค่า Residual Deviance หรือค่า Log Likelihood สำหรับ Full Model (L_1) เท่ากับ 503.31 ซึ่งสามารถคำนวณหาค่า Pseudo-R² ได้เท่ากับ 48.39 หมายความว่าสามารถพยากรณ์ได้ร้อยละ 48.39 ได้สมการพยากรณ์ดังนี้

$$\begin{aligned}\text{Logit}(\mu) = & -137.900 + 0.322\text{Bio1} + 0.609\text{Bio3} - 0.687\text{Bio5} + 0.783\text{Bio7} \\ & + 0.374\text{Bio8} - 0.056\text{Bio9} + 0.030\text{Bio12} + 0.035\text{Bio13} - 0.698\text{Bio14} \\ & - 0.075\text{Bio16} + 0.205\text{Bio17}\end{aligned}$$

จากสมการพบว่าตัวแปร Bio1, Bio3, Bio7, Bio8, Bio12, Bio13 และ Bio17 มีค่าสัมประสิทธิ์เป็นบวก แสดงว่าตัวแปรดังกล่าวมีผลเชิงบวกซึ่งจะไปช่วยเพิ่มโอกาสของการค้นพบต้นชาป่าในพื้นที่ ในขณะที่ตัวแปร Bio5, Bio9, Bio14 และ Bio16 มีค่าสัมประสิทธิ์เป็นลบแสดงว่าตัวแปรดังกล่าวมีผลเชิงลบซึ่งจะไปลดโอกาสการค้นพบต้นชาป่าในพื้นที่

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นบวกแล้วพบว่าตัวแปร Bio7, Bio8, Bio12 และ Bio17 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.001 อธิบายได้ดังต่อไปนี้

ตัวแปรช่วงความแตกต่างของอุณหภูมิรายปี (Bio7) มีผลต่อแบบจำลองเชิงบวก อธิบายได้ว่า ยิ่งตัวแปรช่วงความแตกต่างของอุณหภูมิรายปีมีค่าสูงขึ้น โอกาสที่จะค้นพบต้นชาป่ามีมากขึ้น หมายความว่าพื้นที่ใดที่มีช่วงความแตกต่างของอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุดกับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุดมาก พื้นที่นั้นมีโอกาสที่จะค้นพบต้นชาป่ามากยิ่งขึ้น

ตัวแปรอุณหภูมิเฉลี่ยในไตรมาสที่ฝนตกชุกที่สุด (Bio8) มีผลเชิงบวก หมายความว่าพื้นที่ใดที่มีอุณหภูมิสูงในช่วงไตรมาสที่ฝนตกชุกที่สุด โอกาสที่จะค้นพบต้นชาป่าในพื้นที่มีมากขึ้น

ตัวแปรปริมาณน้ำฝนรายปี (Bio12) อธิบายได้ว่าในพื้นที่ที่มีปริมาณน้ำฝนรายปีที่มากพอ โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้นด้วยเช่นกัน นั่นคือพื้นที่ที่มีความเหมาะสมต่อการเจริญเติบโตของต้นชาป่า

ตัวแปรปริมาณน้ำฝนในไตรมาสที่แห้งแล้งที่สุด (Bio17) มีผลเชิงบวกสามารถอธิบายได้ว่าพื้นที่ใดที่มีปริมาณน้ำฝนสูงๆ ในช่วงไตรมาสที่แห้งแล้งที่สุด จะมีโอกาสที่จะค้นพบต้นชาป่าในพื้นที่มากขึ้น

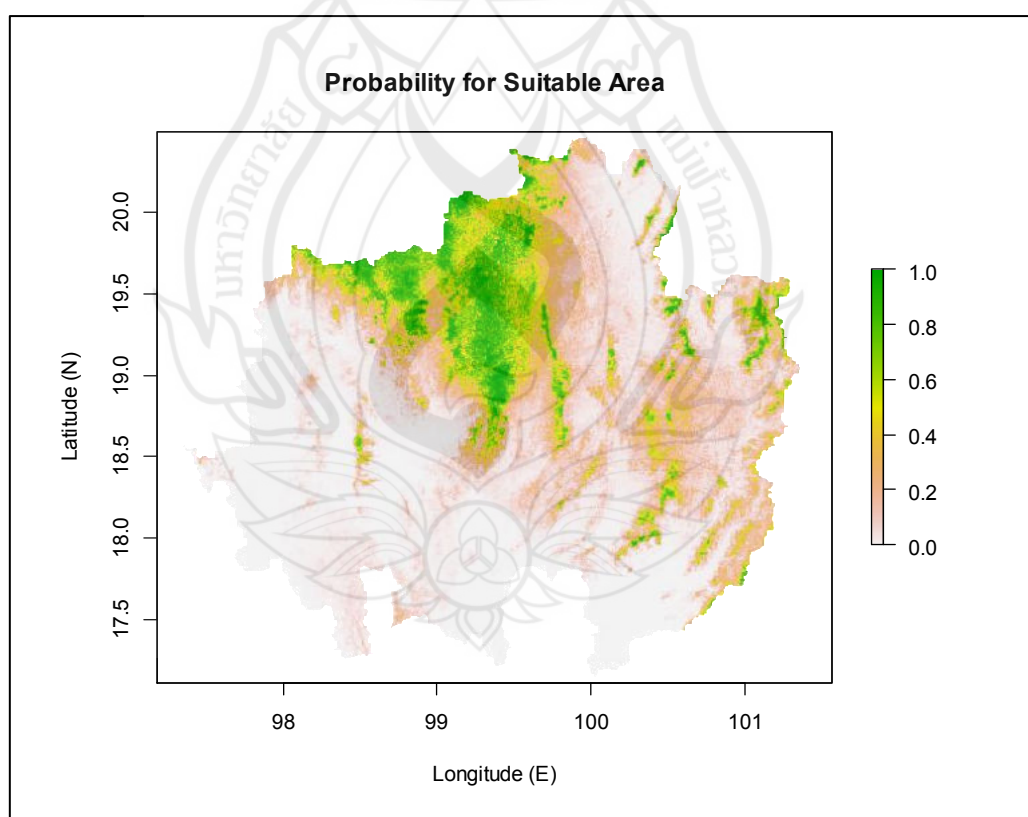
ตัวแปร Bio3 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.01 อธิบายได้ว่าตัวแปร Isothermality (Bio3) ซึ่งเกิดจากสัดส่วนระหว่างความแตกต่างของอุณหภูมิรายวันโดยเฉลี่ย (Bio2) ต่อความแตกต่างของอุณหภูมิรายปี (Bio7) ซึ่งได้มาจากผลต่างระหว่างอุณหภูมิสูงสุดในช่วงเดือนที่ร้อนที่สุด (Bio5) ลบกับอุณหภูมิต่ำสุดในช่วงเดือนที่หนาวที่สุด (Bio6) นั่นคือยิ่งค่า Isothermality (Bio3) มีค่ามาก หมายถึงมีความผันผวนของอุณหภูมิรายวันเมื่อเทียบกับช่วงความแตกต่างของอุณหภูมิรายปีมากขึ้น โอกาสของการค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

ตัวแปร Bio1 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.05 อธิบายได้ว่าพื้นที่ใดมีอุณหภูมิเฉลี่ยรายปีเพียงพอ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่ก็เพิ่มขึ้น

ตัวแปร Bio13 มีผลต่อแบบจำลองเชิงบวกที่ระดับนัยสำคัญทางสถิติที่ 0.10 อธิบายได้ว่าพื้นที่ใดในช่วงเดือนที่ฝนตกชุกที่สุดแล้วมีปริมาณน้ำฝนที่เพียงพอ โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะมากขึ้น

เมื่อพิจารณาตัวแปรที่มีค่าสัมประสิทธิ์เป็นลบแล้วพบว่าตัวแปร Bio5, Bio14 และ Bio16 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.001 อธิบายได้ว่าพื้นที่ใดมีอุณหภูมิสูงสุดในเดือนที่ร้อนที่สุด (Bio5) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง เช่นเดียวกับพื้นที่ใดมีปริมาณน้ำฝนในช่วงเดือนที่แห้งแล้งที่สุด (Bio14) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลงเช่นกัน หรือพื้นที่ใดมีปริมาณน้ำฝนในไตรมาสที่ฝนตกชุกที่สุด (Bio16) มากเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นก็จะน้อยลง

ส่วนตัวแปร Bio9 มีผลต่อแบบจำลองเชิงลบที่ระดับนัยสำคัญทางสถิติที่ 0.10 อธิบายได้ว่าพื้นที่ใดมีอุณหภูมิเฉลี่ยในไตรมาสที่แห้งแล้งที่สุด (Bio9) สูงเกินไป โอกาสที่จะค้นพบต้นชาป่าในพื้นที่นั้นจะน้อยลง



ภาพที่ ๑๑ แผนที่ความน่าจะเป็นในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise

จากภาพที่ 11 แสดงแผนที่ความน่าจะเป็นจาก 0 ถึง 1 ในการค้นพบต้นชาป่าแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ถ้าความน่าจะเป็นเท่ากับ 0 (พื้นที่สีเทา) หมายความว่าพื้นที่นั้นไม่มีโอกาสที่จะค้นพบต้นชาป่าเลย แต่ถ้าความน่าจะเป็นเท่ากับ 1 (พื้นที่สีเขียว) หมายความว่าพื้นที่นั้นมีศักยภาพเพียงพอที่จะมีการค้นพบต้นชาป่าในพื้นที่ แต่ก็ไม่ได้หมายความว่าเมื่อลงสำรวจจริงจะพบต้นชาป่าในพื้นที่จริง เพียงแต่พื้นที่นั้นมีลักษณะทางภูมิศาสตร์ที่เหมาะสมต่อการเจริญเติบโตของต้นชาป่าและมีโอกาสค่อนข้างสูงที่จะค้นพบต้นชาป่าในพื้นที่

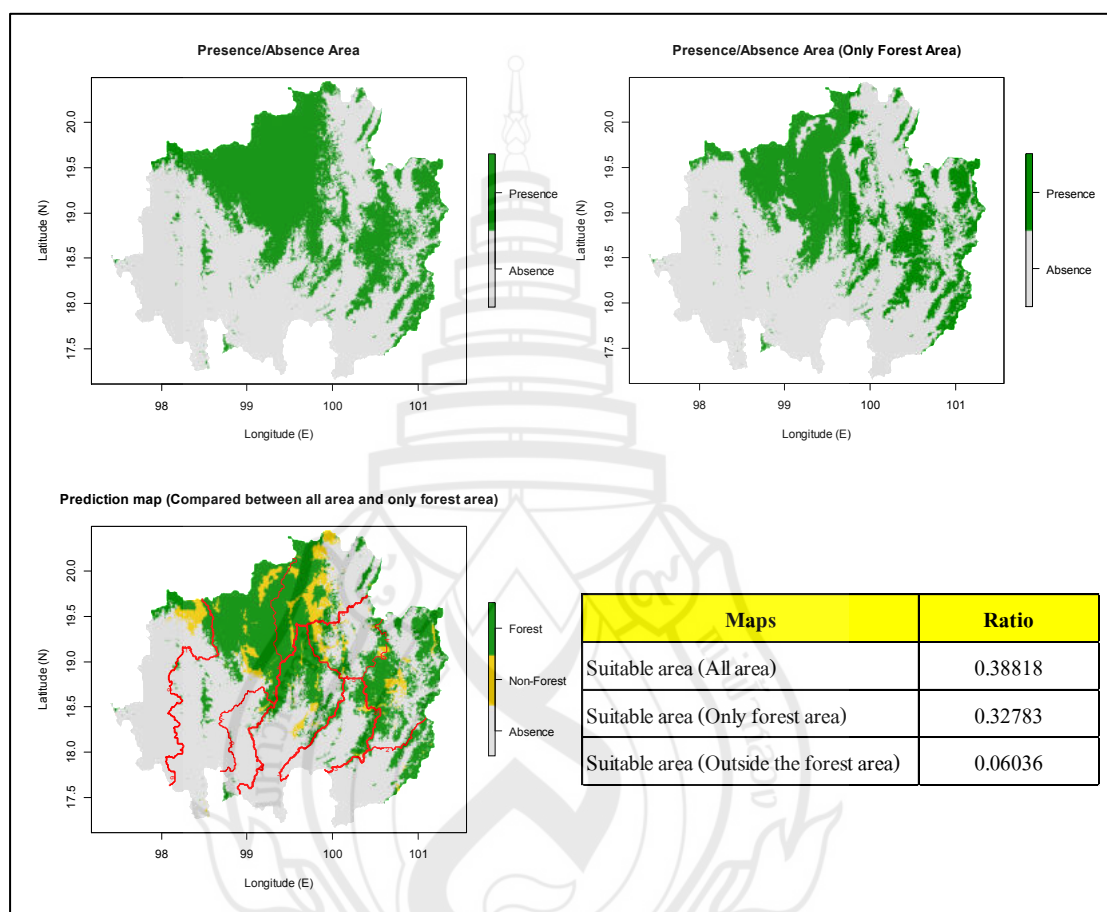
หลังจากนั้นคำนวณค่า Threshold ซึ่งใช้สำหรับแปลงค่าความน่าจะเป็นในแต่ละพิกเซลของแผนที่ให้เป็น Binary Score หรือค่า 0 และ 1 โดยค่า 0 หมายถึง พื้นที่ Absence ค่า 1 หมายถึง พื้นที่ Presence

ตารางที่ 12 ค่าที่คำนวณได้จากการตรวจสอบความถูกต้องของแบบจำลองของตัวแปรทางด้านภูมิอากาศ

การตรวจสอบความถูกต้องของแบบจำลอง (Model Evaluation)	Value
จำนวนจุด Presences ที่ใช้ตรวจสอบ	38
จำนวนจุด Pseudo-Absences ที่ใช้ตรวจสอบ	366
AUC	0.93148
max TPR+TNR	-2.19648
Threshold	0.10007

จากตารางที่ 12 แสดงค่าที่คำนวณได้จากการประเมินผลแบบจำลองของตัวแปรทางด้านภูมิศาสตร์ที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise แล้ว พบว่า จำนวนจุด Presence ที่ใช้ในการตรวจสอบความถูกต้องของแบบจำลองใช้ทั้งหมด 38 จุด จำนวนจุด Pseudo-Absences ใช้ทั้งหมด 366 จุด มีค่า AUC เท่ากับ 0.93148 หมายความว่ามีความถูกต้องของการพยากรณ์เท่ากับร้อยละ 93.15 ค่า max TPR + TNR คือผลรวมสูงสุดของค่า True Positive Rate (TPR) และ True Negative Rate (TNR) ที่เกิดจากแบบจำลอง ซึ่งค่านี้สามารถนำไปคำนวณหาค่า Threshold ได้เท่ากับ 0.10007

หมายความว่าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นน้อยกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 0 แต่ถ้าพิกเซลใดในแผนที่ที่มีค่าความน่าจะเป็นมากกว่าค่า Threshold พิกเซลนั้นจะมีค่าเท่ากับ 1



ภาพที่ 12 พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise เปรียบเทียบระหว่างพื้นที่ภาคเหนือตอนบนทั้งหมดและเฉพาะพื้นที่ป่าในภาคเหนือตอนบน

จากภาพที่ 12 รูปภาพแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าของแบบจำลองของตัวแปรทางด้านภูมิอากาศที่มีการคัดเลือกตัวแปรโดยวิธีการ Stepwise ซึ่งแผนที่ถูกแปลงให้มีค่าเฉพาะ 0 และ 1 แล้ว โดยพื้นที่สีเขียวคือพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่า โดยภาพด้านซ้ายมือบนแสดงพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าในพื้นที่ภาคเหนือตอนบน ไม่ว่าพื้นที่นั้นจะอยู่หรือไม่อยู่ในพื้นที่ป่าก็ตาม ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 39 ของพื้นที่ ส่วนภาพด้านขวามือบนแสดง

พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าโดยพิจารณาเฉพาะพื้นที่ที่เป็นบริเวณป่าเท่านั้น ซึ่งพื้นที่นี้มีอยู่ประมาณร้อยละ 33 ของพื้นที่โดยพิจารณาจากพิกเซลใดที่พยากรณ์ออกมาว่าเป็นพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่พิกเซลนั้นไม่ได้อยู่ในบริเวณป่า ก็จะทำให้พิกเซลนั้นกลายเป็นพื้นที่ที่ไม่มีการค้นพบต้นชาป่าทันที ถึงแม้ว่าปัจจัยทางด้านสภาพภูมิอากาศจะมีความเหมาะสมก็ตาม ซึ่งพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าแต่ไม่ได้อยู่ในเขตป่ามีพื้นที่ประมาณร้อยละ 6 ของพื้นที่ ส่วนใหญ่อยู่ในจังหวัดแม่ฮ่องสอน เชียงใหม่และเชียงราย ดังแสดงในภาพล่างซ้าย (พื้นที่สีเหลือง) โดยมีเส้นสีแดงแสดงขอบเขตของแต่ละจังหวัด เมื่อพิจารณาพื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าเฉพาะในเขตพื้นที่ป่าพบว่า พื้นที่ที่คาดว่าจะมีการค้นพบต้นชาป่าส่วนใหญ่อยู่ทางทิศตะวันตกเฉียงเหนือของภาคเหนือตอนบนของประเทศไทย ซึ่งอยู่แถวๆ จังหวัดเชียงใหม่ เชียงรายและลำปางตอนบน โดยกระจายอย่างหนาแน่นบริเวณป่ารอยต่อของทั้ง 3 จังหวัด และมีบางส่วนกระจายอยู่อย่างหนาแน่นในจังหวัดน่านและตอนบนของจังหวัดอุดรธานี



ภาคผนวก จ

ข้อมูลการสำรวจต้นชาป่าในเขตพื้นที่ภาคเหนือตอนบน

ตารางที่ จ1 ข้อมูลจุดพิกัดที่มีการค้นพบต้นชาป่า

No	Province	District	Sub_district	Village	lon	lat
1	Chiang Rai	Mae chan	Pa sang	Pa miang	99.77206	20.18353
2	Chiang Rai	Mae chan	Pa sang	Pa miang	99.77212	20.18355
3	Chiang Rai	Mae chan	Pa sang	Pa miang	99.77278	20.18350
4	Chiang Rai	Mae chan	Pa sang	Pa miang	99.77281	20.18347
5	Chiang Rai	Mae chan	Pa sang	Pa miang	99.77295	20.18297
6	Chiang Rai	Mae chan	Pa tung	Pa miang	99.78059	20.15464
7	Chiang Rai	Mae chan	Pa tung	Pa miang	99.78050	20.15461
8	Chiang Rai	Muang	Mae korn	Pang rim korn	99.66598	19.84161
9	Chiang Rai	Muang	Mae korn	Pang rim korn	99.66030	19.84164
10	Chiang Rai	Muang	Mae korn	Pang rim korn	99.66570	19.84174
11	Chiang Rai	Muang	Mae korn	Pang rim korn	99.66739	19.84436
12	Chiang Rai	Muang	Mae korn	Pang rim korn	99.68698	19.86164
13	Chiang Rai	Muang	Mae korn	Pang rim korn	99.66986	19.86136
14	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.57517	19.87300
15	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.57725	19.87352
16	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.57889	19.87764
17	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.57953	19.88081
18	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.58048	19.88078
19	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.58128	19.88106
20	Chiang Rai	Mae suai	Va vee	Mai phatthana	99.58139	19.88108

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
21	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.32502	19.13825
22	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.32786	19.13242
23	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.32778	19.13166
24	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.32764	19.14697
25	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.34767	19.14650
26	Chiang Rai	Wiang pa pao	Mae jedee	Huai nam guen	99.36764	19.14864
27	Chiang Rai	Muang	Huai chomphu	Pang khon	99.58253	19.88262
28	Chiang Rai	Muang	Huai chomphu	Pang khon	99.57556	19.89261
29	Chiang Rai	Muang	Huai chomphu	Pang khon	99.57734	19.89447
30	Chiang Rai	Muang	Huai chomphu	Pang khon	99.57736	19.89436
31	Chiang Rai	Muang	Huai chomphu	Pang khon	99.59995	19.89797
32	Chiang Rai	Muang	Huai chomphu	Pang khon	99.61447	19.90094
33	Chiang Rai	Muang	Huai chomphu	Pang khon	99.63362	19.90519
34	Chiang Rai	Phan	San klang	Mae khao luang	99.69220	19.59450
35	Chiang Rai	Phan	San klang	Mae khao luang	99.69258	19.59334
36	Chiang Rai	Phan	San klang	Mae khao luang	99.69256	19.59409
37	Chiang Rai	Phan	San klang	Mae khao luang	99.69303	19.59388
38	Chiang Rai	Phan	San klang	Mae khao luang	99.69395	19.59361
39	Chiang Rai	Phan	San klang	Mae khao luang	99.69364	19.59381
40	Chiang Rai	Thoeng	Chiang kien	Chiang kien	99.99812	19.61445
41	Chiang Rai	Thoeng	Chiang kien	Chiang kien	99.99817	19.61556
42	Chiang Rai	Thoeng	Chiang kien	Chiang kien	100.00656	19.61445
43	Chiang Rai	Thoeng	Chiang kien	Chiang kien	100.00642	19.61455

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
44	Chiang Rai	Thoeng	Chiang kien	Chiang kien	100.00625	19.61436
45	Chiang Rai	Thoeng	Chiang kien	Chiang kien	100.00623	19.61403
46	Chiang Rai	Thoeng	Chiang kien	Chiang kien	100.00653	19.61464
47	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.74089	20.28075
48	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.74092	20.2687
49	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.74117	20.2688
50	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.74131	20.26867
51	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.7412	20.26878
52	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.74187	20.26672
53	Chiang Rai	Mae fah luang	Mae fah luang	Samakkhi	99.73928	20.26522
54	Lampang	Muang pan	Chaeson	Pa miang	99.38750	18.83261
55	Lampang	Muang pan	Chaeson	Pa miang	99.39228	18.83500
56	Lampang	Muang pan	Chaeson	Pa miang	99.39306	18.83550
57	Lampang	Muang pan	Chaeson	Pa miang	99.39392	18.83519
58	Lampang	Muang pan	Chaeson	Pa miang	99.38925	18.83486
59	Lampang	Muang pan	Chaeson	Pa miang	99.38281	18.81859
60	Lampang	Muang pan	Chaeson	Pa miang	99.38339	18.81925
61	Lampang	Ngao	Ban rong	Park borg	99.77861	19.04244
62	Lampang	Ngao	Ban rong	Park borg	99.78003	19.04184
63	Lampang	Ngao	Ban rong	Park borg	99.78017	19.04139
64	Lampang	Ngao	Ban rong	Park borg	99.78011	19.04120
65	Lampang	Ngao	Ban rong	Park borg	99.77980	19.04081
66	Lampang	Ngao	Ban rong	Park borg	99.78042	19.04256

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
67	Lampang	Ngao	Ban rong	Park borg	99.77423	19.04950
68	Nan	Pua	Sa kat	Sa kat	101.01225	19.26725
69	Nan	Pua	Sa kat	Sa kat	101.01275	19.26675
70	Nan	Pua	Sa kat	Sa kat	101.01039	19.26634
71	Nan	Pua	Sa kat	Sa kat	101.00892	19.26586
72	Nan	Pua	Sa kat	Sa kat	101.01442	19.27164
73	Nan	Pua	Sa kat	Sa kat	101.00026	19.26380
74	Nan	Pua	Sa kat	Sa kat	101.00728	19.26800
75	Nan	Pua	Sa kat	Sa kat tai	101.09136	19.26767
76	Nan	Pua	Sa kat	Sa kat tai	101.01873	19.26764
77	Nan	Pua	Sa kat	Sa kat tai	101.01831	19.26791
78	Nan	Pua	Sa kat	Sa kat tai	101.01778	19.26803
79	Nan	Pua	Sa kat	Sa kat tai	101.01714	19.26773
80	Nan	Pua	Sa kat	Sa kat tai	101.01631	19.26822
81	Nan	Pua	Sa kat	Sa kat tai	101.01675	19.26922
82	Nan	Pua	Sa kat	Sa kat klang	101.00012	19.26333
83	Nan	Pua	Sa kat	Sa kat klang	101.01620	19.26381
84	Nan	Pua	Sa kat	Sa kat klang	101.00153	19.26391
85	Nan	Pua	Sa kat	Sa kat klang	101.00237	19.25586
86	Nan	Pua	Sa kat	Sa kat klang	101.00223	19.25606
87	Nan	Pua	Sa kat	Sa kat klang	101.00226	19.25619
88	Nan	Pua	Sa kat	Sa kat klang	101.00251	19.25652
89	Nan	Maung	Rueang	Sri na pan	100.65519	18.78353

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
90	Nan	Maung	Rueang	Sri na pan	100.63347	18.79266
91	Nan	Maung	Rueang	Sri na pan	100.63334	18.79233
92	Nan	Maung	Rueang	Sri na pan	100.63292	18.79236
93	Nan	Maung	Rueang	Sri na pan	100.63084	18.79292
94	Nan	Maung	Rueang	Sri na pan	100.63539	18.79206
95	Nan	Maung	Rueang	Sri na pan	100.64084	18.79186
96	Nan	Maung	Rueang	Sri na pan	100.63934	18.78920
97	Phrae	Maung	Pa daeng	Pa klai	100.31783	18.04133
98	Phrae	Maung	Pa daeng	Pa klai	100.31795	18.04119
99	Phrae	Maung	Pa daeng	Pa klai	100.31745	18.04131
100	Phrae	Maung	Pa daeng	Pa klai	100.31742	18.04117
101	Phrae	Maung	Pa daeng	Pa klai	100.31739	18.04136
102	Phrae	Maung	Pa daeng	Pa klai	100.31758	18.04148
103	Chiang Mai	Chiang dao	Mae na	Mae mae	98.90025	19.32023
104	Chiang Mai	Chiang dao	Mae na	Mae mae	98.90014	19.32041
105	Chiang Mai	Chiang dao	Mae na	Mae mae	98.89422	19.31965
106	Chiang Mai	Chiang dao	Mae na	Mae mae	98.89030	19.31975
107	Chiang Mai	Chiang dao	Mae na	Mae mae	98.89030	19.31957
108	Chiang Mai	Chiang dao	Mae na	Mae mae	98.88454	19.30295
109	Chiang Mai	Chiang dao	Mae na	Mae mae	98.88239	19.32011
110	Chiang Mai	Maung	Su thep	Chang kieran	98.89634	18.77000
111	Chiang Mai	Maung	Su thep	Chang kieran	98.90244	18.84141
112	Chiang Mai	Maung	Su thep	Chang kieran	98.90439	18.84215

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
113	Chiang Mai	Maung	Su thep	Chang kieran	98.91756	18.85217
114	Chiang Mai	Maung	Su thep	Chang kieran	98.91801	18.85602
115	Chiang Mai	Doi saket	Thep sadet	Pang hi	99.33241	18.92282
116	Chiang Mai	Doi saket	Thep sadet	Dong	99.35216	18.92399
117	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33125	19.00139
118	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33218	18.99974
119	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33688	18.99682
120	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33344	19.00080
121	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33821	18.99465
122	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33958	18.99395
123	Chiang Mai	Doi saket	Thep sadet	kew tam	99.33900	18.98839
124	Chiang Mai	Doi saket	Thep sadet	pang bong	99.32008	18.97011
125	Chiang Mai	Doi saket	Thep sadet	Kam pang hin	99.35128	18.94890
126	Chiang Mai	Doi saket	Thep sadet	Mae torn	99.37676	18.93543
127	Chiang Mai	Doi saket	Thep sadet	Mae torn	99.37627	18.93741
128	Chiang Mai	Mae ai	Mae ai	Na muen	99.16930	20.05066
129	Chiang Mai	Mae ai	Mae ai	Na muen	99.17961	20.02671
130	Chiang Mai	Mae ai	Mae ai	Na muen	99.17411	20.03012
131	Chiang Mai	Mae ai	Mae ai	Pha hom pok	99.14227	20.06965
132	Chiang Mai	Mae ai	Mae ai	Pha hom pok	99.14363	20.06790
133	Chiang Mai	Mae ai	Mae ai	Pha hom pok	99.14569	20.06800
134	Chiang Mai	Mae ai	Mae ai	Pha hom pok	99.15289	20.06200
135	Chiang Mai	Mae ai	Mae ai	Pha hom pok	99.15546	20.05700

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
136	Chiang Mai	Mae Chaem	Mae na chon	Mae jae	98.51718	18.86273
137	Chiang Mai	Mae Chaem	Mae na chon	Mae jae	98.51677	18.86228
138	Chiang Mai	Mae Chaem	Mae na chon	Mae jae	98.51871	18.86318
139	Chiang Mai	Mae Chaem	Mae na chon	Mae jae	98.51869	18.85845
140	Chiang Mai	Mae rim			98.82500	18.87289
141	Chiang Mai	Mae rim			98.82500	18.87289
142	Chiang Mai	Mae rim			98.82574	18.87051
143	Chiang Mai	Mae rim			98.82632	18.8654
144	Chiang Mai	Mae rim			98.81568	18.92202
145	Chiang Mai	Om koi	Mon chong	Mu sur park thang	98.52184	17.57394
146	Chiang Mai	Om koi	Mon chong	Mu sur park thang	98.52245	17.57238
147	Chiang Mai	Phrao	Long khot	Mae sai	99.27592	19.10633
148	Chiang Mai	Phrao	Long khot	Mae sai	99.27496	19.10614
149	Chiang Mai	Phrao	Long khot	Mae sai	99.27214	19.10529
150	Chiang Mai	Phrao	Long khot	Mae sai	99.27114	19.10306
151	Chiang Mai	Chiang dao	Mae na	Pang hang	98.88393	19.33602
152	Chiang Mai	Chiang dao	Mae na	Pang hang	98.88450	19.33710
153	Chiang Mai	Chiang dao	Mae na	Pang hang	98.88450	19.33692
154	Chiang Mai	Chiang dao	Mae na	San pa kia	98.83266	19.31414
155	Chiang Mai	Mae taeng	Pa pae	Tha pha	98.75706	19.12638
156	Chiang Mai	Mae taeng	Pa pae	Tha pha	98.70709	19.12092
157	Chiang Mai	Mae taeng	Pa pae	Pang ma kluai	98.71230	19.11841

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
158	Chiang Mai	Mae taeng	Pa pae	Pang ma kluai	98.71419	19.11986
159	Chiang Mai	Mae taeng	Pa pae	Tha pha	98.74644	19.12345
160	Chiang Mai	Mae taeng	Pa pae	Tha pha	98.75155	19.12335
161	Chiang Mai	Mae taeng	Pa pae	Tha pha	98.75462	19.12637
162	Chiang Mai	Mae taeng	Pa pae	Pang ma kluai	98.70763	19.11688
163	Chiang Mai	Sa moeng	Sa moeng nuea	Ban teen dong	98.72983	18.96432
164	Chiang Mai	Sa moeng	Sa moeng nuea	Ban teen dong	98.73067	18.96702
165	Mae hong son	Pang mapha	Pang mapha	Mae la na	98.21456	19.59016
166	Mae hong son	Pang mapha	Pang mapha	Mae la na	98.21393	19.59291
167	Mae hong son	Pang mapha	Pang mapha	Mae la na	98.21405	19.59280
168	Mae hong son	Pang mapha	Pang mapha	Mai hung	98.20842	19.62125
169	Mae hong son	Pang mapha	Pang mapha	Mai hung	98.20805	19.62051
170	Mae hong son	Pang mapha	Pang mapha	Mai hung	98.20795	19.62071
171	Mae hong son	Pang mapha	Pang mapha	Mai hung	98.2090	19.62106
172	Mae hong son	Pang mapha	Pang mapha	Mai lun	98.22695	19.69541
173	Mae hong son	Pang mapha	Pang mapha	Mai lun	98.72268	19.69497
174	Mae hong son	Pang mapha	Pang mapha	Mai lun	98.72268	19.69497
175	Mae hong son	Pang mapha	Pang mapha	Mai lun	98.22705	19.69504
176	Mae hong son	Pang mapha	Pang mapha	Mai lun	98.22697	19.69531
177	Mae hong son	Mueang	Mok champae	Na pa paek	97.92892	19.53271
178	Mae hong son	Mueang	Mok champae	Rak thai	97.94662	19.55603
179	Mae hong son	Mueang	Mok champae	Na pa paek	97.94533	19.54788
180	Mae hong son	Mueang	Mok champae	Rak thai	97.94532	19.54791

ตารางที่ จ1 (ต่อ)

No	Province	District	Sub_district	Village	lon	lat
181	Mae hong son	Mae la noi	Huai hom	Huai hom	98.07178	18.36862
182	Mae hong son	Mae la noi	Huai hom	Huai hom	98.07150	18.36515
183	Mae hong son	Mae la noi	Huai hom	Huai hom	98.07217	18.36256
184	Mae hong son	Mae la noi	Huai hom	Huai hom	98.06680	18.35588
185	Mae hong son	Mae la noi	Huai hom	laoop	98.03944	18.35266





ประวัติผู้เขียน

ประวัติผู้เขียน

ชื่อ	นางสาวสุรัสวดี นางแล
วัน เดือน ปีเกิด	18 สิงหาคม 2525
สถานที่อยู่ปัจจุบัน	เลขที่ 77 หมู่ที่ 3 ถนนพหลโยธิน ตำบลนางแล อำเภอเมือง จังหวัดเชียงราย 57100
ประวัติการศึกษา	
พ.ศ. 2550	ปริญญาโท วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาศาสตร์เชิงคำนวณ มหาวิทยาลัยแม่ฟ้าหลวง
พ.ศ. 2547	ปริญญาตรี วิทยาศาสตร์บัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยเชียงใหม่
ประวัติการทำงาน	-