

การจัดเตรียมข้อมูลสำหรับ
โครงการ Pixel House

คำนำ

คู่มือการจัดเตรียมข้อมูลภาพถ่ายจากดาวเทียมเพื่อใช้ในการจำแนกสิ่งปกคลุมดินของโครงการห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) ระยะที่ 1 นี้จัดทำขึ้นสำหรับใช้ในการจัดเตรียมข้อมูลภาพถ่ายจากดาวเทียมเพื่อใช้ในโครงการห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) ระยะที่ 1 โดยวัตถุประสงค์ของโครงการคือเพื่อการศึกษา วิจัย และพัฒนา นวัตกรรมใหม่ๆ ในด้านการประมวลผลภาพถ่ายดาวเทียม และพัฒนาองค์ความรู้ต่างๆ เพื่อสร้างนักวิจัย นักวิชาการ และผู้ปฏิบัติการ ที่มีความรู้ ความสามารถในการประมวลผลภาพถ่ายดาวเทียม

โดยในคู่มือจะเป็นการนำเสนอขั้นตอนการจัดเตรียมข้อมูลในส่วนข้อมูลของ Training data จากข้อมูลจากดาวเทียม LANDSAT-8 ครอบคลุมพื้นที่ทั้งประเทศไทย เพื่อนำไปใช้ในการพัฒนาอัลกอริทึมในการจำแนกสิ่งปกคลุมดินทั้ง 9 ชนิด ตามเป้าหมายของโครงการ และยังสามารถนำไปประยุกต์ใช้งานในการจัดเตรียมข้อมูลเพื่อการประมวลผลภาพถ่ายจากดาวเทียมทั้งเพื่อการจำแนก หรือเพื่อการประมวลผลอื่นๆ ได้อีกด้วย

สารบัญ

1. ความเป็นมา.....	1
1.1 โครงการห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) ระยะที่ 1.....	1
1.2 การจัดเตรียมข้อมูลดาวเทียมสำหรับการจำแนกสิ่งปกคลุมดิน.....	1
2. การจัดเตรียมข้อมูล.....	2
2.1 การคัดเลือกข้อมูลดาวเทียม.....	3
2.2 การกำหนดขนาดและตำแหน่งของ Training data.....	5
2.3 การตัดข้อมูลจากภาพถ่ายจากดาวเทียม	6
3. การจำแนกสิ่งปกคลุมดิน.....	11
3.1 การจัดทำข้อมูลและการเปรียบเทียบข้อมูลอ้างอิง.....	11
3.2 การจำแนกข้าว.....	12
3.3 การจำแนกข้าวโพด.....	13
3.4 การจำแนกอ้อย	13
3.5 การจำแนกมันสำปะหลัง.....	14
3.6 การจำแนกปาล์มน้ำมัน	15
3.7 การจำแนกยางพารา.....	15
3.8 การจำแนกป่าไม้.....	16
3.9 การจำแนกพื้นที่เมือง.....	16
3.10 การจำแนกแหล่งน้ำ.....	12

1. ความเป็นมา

1.1 โครงการห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) ระยะที่ 1

สำนักงานพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ หรือ สทอภ. เป็นหน่วยงานภาครัฐที่มีการศึกษาโดยตรงในการพัฒนาเทคโนโลยีอวกาศและภูมิสารสนเทศ โดยเฉพาะการประมวลผลภาพถ่ายจากดาวเทียม เพื่อใช้ในการผลิตชั้นข้อมูลและผลิตภัณฑ์ภูมิสารสนเทศ และนำไปประยุกต์ใช้ในการศึกษาด้านต่างๆ ต่อไป อย่างไรก็ตาม การแปลและตีความข้อมูลดาวเทียมเพื่อจัดทำชั้นข้อมูลสิ่งปกคลุมดิน นั้น ถือว่าเป็นสิ่งที่สำคัญมาก และจำเป็นต้องใช้ทักษะและความรู้ของผู้ที่มีความชำนาญการในการจำแนกสิ่งปกคลุมดินในภาพถ่ายจากดาวเทียม จึงทำให้ขั้นตอนในการผลิตชั้นข้อมูลสิ่งปกคลุมดินใช้เวลานานและทรัพยากรคนค่อนข้างมาก จึงเป็นข้อจำกัดในการเพิ่มความดีสำหรับปรับปรุงชั้นข้อมูลสิ่งปกคลุมดินให้มีความทันสมัยและรวดเร็วตามความต้องการของผู้ใช้งาน ในปัจจุบัน เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) โดยเฉพาะกระบวนการการเรียนรู้ของเครื่อง (Machine learning) มีบทบาทมากในการนำมาใช้ร่วมกันข้อมูลขนาดใหญ่ (Big Data) เพื่อเพิ่มประสิทธิภาพการทำงานของระบบปฏิบัติการแบบอัตโนมัติและให้ผลลัพธ์เทียบเท่ากับสิ่งที่ดำเนินการด้วยคน ด้วยเหตุนี้ เพื่อให้ สทอภ. สามารถสร้างนวัตกรรมใหม่ๆ ในการประมวลผลภาพถ่ายดาวเทียม ให้สอดคล้องกับการศึกษา และแนวทาบนโยบายของประเทศที่มุ่งเน้นให้ประเทศไทยเข้าสู่ยุคของ Thailand 4.0 สถาบันวิทยาการอวกาศและภูมิสารสนเทศ (GISTDA Academy) จึงมุ่งเน้นให้มีการตั้งห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) เพื่อใช้เป็นแหล่งในการคิดค้น วิจัย และพัฒนา นวัตกรรมใหม่ๆ ในด้านการประมวลผลภาพถ่ายดาวเทียม และพัฒนาองค์ความรู้ต่างๆ เพื่อสร้างนักวิจัย นักวิชาการ และผู้ประกอบการ ที่มีความรู้ ความสามารถในการประมวลผลภาพถ่ายดาวเทียมต่อไป

1.2 การเตรียมข้อมูลดาวเทียมสำหรับการจำแนกสิ่งปกคลุมดิน

สำหรับโครงการห้องปฏิบัติการประมวลผลภาพถ่ายดาวเทียมอัจฉริยะ (Pixel House) ระยะที่ 1 (โครงการ Pixel House) มีเป้าหมายที่จะทำการจำแนกสิ่งปกคลุมดินจำนวน 9 ชนิดจากภาพถ่ายดาวเทียม ด้วยเทคนิคแบบ Deep learning เพื่อลดความซับซ้อน ประหยัดเวลา และลดค่าใช้จ่ายในการใช้คน เพื่อทำการจำแนกข้อมูลครั้งต่อไป

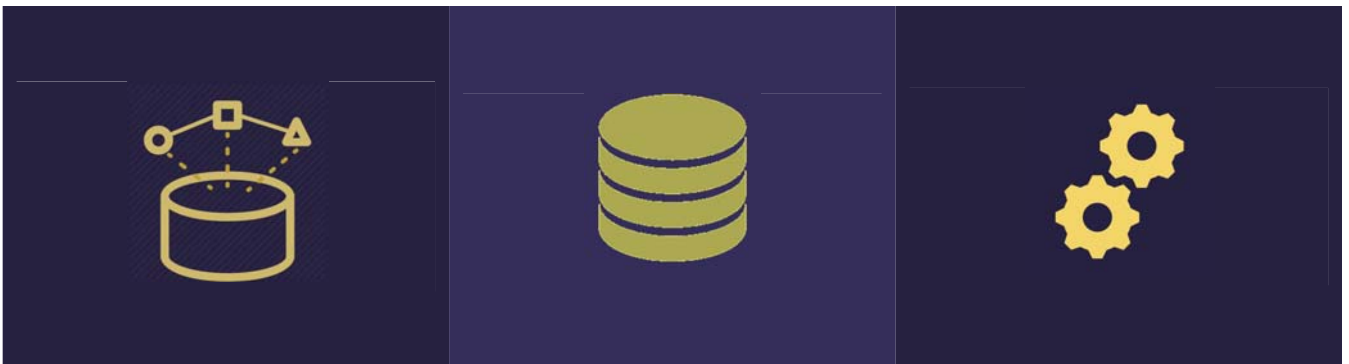
ดังนั้นเพื่อให้ผลการจำแนกมีความถูกต้องสูง หรือใกล้เคียงกับการจำแนกในรูปแบบเดิม จึงต้องมีการสอนให้ระบบการจำแนกรู้จักกับสิ่งปกคลุมดินทั้ง 9 ชนิดก่อน ด้วยการคัดเลือกข้อมูลตัวอย่าง หรือ Training data ทั้ง 9 ชนิด ให้มีความหลากหลาย และกระจายทั่วทั้งประเทศ ทำให้ต้องใช้ข้อมูล Training data ปริมาณมาก และพื้นที่ของ training data ต้องมีขนาดที่เหมาะสม จึงจะเป็นตัวแทนที่ดีของข้อมูลในแต่ละชนิด โดยก่อนที่ทำการคัดเลือก Training data นั้น จะต้องมีการคัดเลือกข้อมูลดาวเทียมที่จะนำมาใช้ในการจำแนกก่อน และยังคงจัดให้อยู่ในรูปแบบที่สะดวกและเหมาะสมต่อการนำไปจำแนกอีกด้วย

ซึ่งการจัดเตรียมข้อมูลเหล่านี้แม้จะไม่ใช่งานที่ง่าย แต่ก็ยังมีขั้นตอนที่หลายหลายพอสมควร หากขาดการวางแผน หรือการจัดการที่ดีอาจจะทำให้เสียเวลาได้ นอกจากนี้ยังสามารถนำวิธีการการจัดเตรียมข้อมูลนี้ไปประยุกต์ใช้ในการจัดเตรียมข้อมูลให้กับการประมวลผลภาพถ่ายจากดาวเทียมในแบบต่างๆ ได้อีกด้วย

2. การจัดเตรียมข้อมูล

ในการจัดทำระบบจำแนกสิ่งปกคลุมดินของโครงการ Pixel House นั้นสามารถแบ่งการดำเนินงานได้เป็น 3 ส่วนหลัก คือ

- การจัดเตรียมข้อมูล: ในส่วนนี้จะเริ่มตั้งแต่การดาวน์โหลดข้อมูล LANDSAT ไปจนถึงการจัดทำ Training data เพื่อไว้สำหรับเป็นข้อมูลสอนให้กับระบบ
- การสร้างฐานข้อมูล: เป็นส่วนของการออกแบบและการจัดการข้อมูลดาวเทียมให้อยู่ในรูปแบบที่เหมาะสมกับการนำไปจำแนกสิ่งปกคลุมดินของโครงการ Pixel House
- การจำแนกข้อมูล: เป็นส่วนของการนำความก้าวหน้าของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) โดยเฉพาะกระบวนการการเรียนรู้ของเครื่อง (Machine learning) เข้ามาช่วยบทบาทในการพัฒนาวิธีการและอัลกอริทึมในการจำแนกข้อมูลสิ่งปกคลุมดินจากภาพถ่ายดาวเทียม

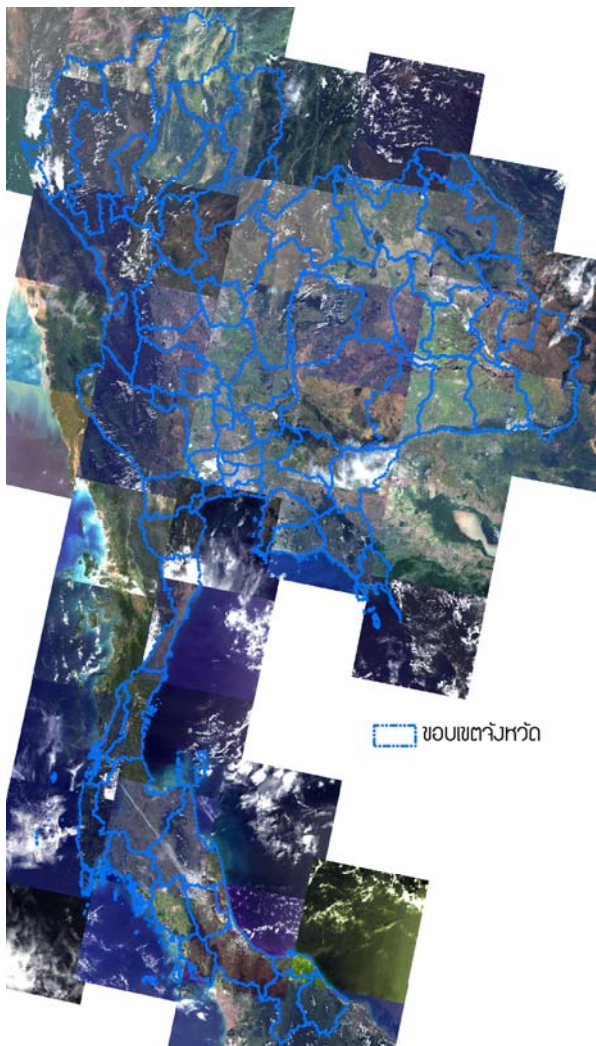


โดยในคู่มือนี้จะกล่าวถึงเฉพาะเรื่องของการจัดเตรียมข้อมูล Training data เพื่อใช้เป็นชุดข้อมูลสำหรับสอนในการพัฒนาอัลกอริทึมเพื่อการจำแนกสิ่งปกคลุมดินในโครงการ Pixel House ซึ่งสามารถแยกกระบวนการดำเนินงานได้เป็น 3 ส่วนเช่นกัน ดังนี้

- การคัดเลือกข้อมูลดาวเทียม: โครงการนี้ใช้ข้อมูลดาวเทียม LANDSAT-8 ในการจำแนกสิ่งปกคลุมดิน ซึ่งมีจำนวน 43 scene ที่ครอบคลุมพื้นที่ประเทศไทย และข้อมูลดาวเทียม LANDSAT-8 ทำการบันทึกข้อมูลตั้งแต่ปี 2556 จนถึงปัจจุบัน เมื่อคำนวณแล้วมีจำนวนข้อมูลภาพประมาณ 5,000 scene ที่ครอบคลุมประเทศไทย
- การกำหนดขนาดและตำแหน่งของ Training Data: ซึ่งได้มีการวางแผนให้ window size ของอัลกอริทึมในการจำแนกประเภทข้อมูลของสิ่งปกคลุมดินมีขนาด 500x500 pixel ดังนั้นในการจัดเตรียมข้อมูล training data ควรจะมีขนาดที่ใหญ่กว่า window size ของอัลกอริทึม (ขนาด 1,000x1,000 pixel) และควรมีการกระจายไปทั่วทั้งประเทศ เพื่อให้ได้ชุดข้อมูลตัวแทนที่ดี ซึ่งจะมีผลต่อคุณภาพของการจำแนกสิ่งปกคลุมดิน

- การตัดข้อมูลภาพถ่ายจากดาวเทียม: เมื่อกำหนดขอบเขตและตำแหน่งของ Training Data ที่กระจายอยู่ทั่วประเทศแล้วจะมีทั้งหมด 210 แห่ โดยในขั้นตอนนี้จะเป็นการตัดข้อมูลภาพถ่ายดาวเทียมให้มีขนาดและตำแหน่งตรงเท่ากับขอบเขตที่กำหนด

2.1 การคัดเลือกข้อมูลดาวเทียม



ภาพที่ 1 จำนวนและขอบของข้อมูลดาวเทียม LANDSAT-8
ที่ใช้ในการเตรียมข้อมูล

ดาวเทียม LANDSAT เป็นชื่อของชุดดาวเทียมสำรวจทรัพยากรธรรมชาติ โดยองค์การ NASA ต่อมาได้โอนกิจการให้บริษัทเอกชน EOSAT เพื่อดำเนินการเชิงพาณิชย์ โดยดาวเทียม LANDSAT-1 ส่งขึ้นสู่อวกาศเมื่อปี 2515 นับเป็นดาวเทียมสำรวจทรัพยากรดวงแรกของโลก และถูกพัฒนามาจนถึงปัจจุบันคือดาวเทียม LANDSAT-8 ซึ่งเริ่มปฏิบัติการวันที่ 30 พฤษภาคม 2556 ภายใต้การบริหารจัดการของ USGS มีวัตถุประสงค์เพื่อติดตามและสำรวจทรัพยากรบนโลก พร้อมทั้งยังเผยแพร่เพื่อการใช้ประโยชน์ของพลเรือน โดยเปิดให้คนทั่วไปสามารถเข้าไปดาวน์โหลดข้อมูลได้ตั้งแต่วันที่ 2013

ดาวเทียม LANDSAT-8 โคจรสูงเหนือพื้นโลก 705 กิโลเมตร โคจรซ้ำตำแหน่งเดิมทุกๆ 16 วัน มีความกว้างของแนวก้นกภาพ 185 กม. และมีเซ็นเซอร์บันทึกข้อมูล 11 แบนด์ ด้วยคุณสมบัติต่างๆ นั้น ทำให้มีการวิจัย พัฒนา และประยุกต์ใช้ข้อมูลจากดาวเทียม LANDSAT-8 ในการจำแนกสิ่งปกคลุมดินบนพื้นโลก โดยต้องใช้ภาพถ่ายจากดาวเทียม LANDSAT-8 จำนวน 43 scene เพื่อให้ครอบคลุมขอบเขตพื้นที่ของประเทศไทยทั้งหมด และ

ยังต้องทำการคัดเลือกภาพถ่ายจากดาวเทียมที่ปลอดเมฆเพื่อให้ผลการจำแนกสิ่งปกคลุมดินออกมามีความถูกต้องแม่นยำมากที่สุด โดยโปรแกรม Pixel House นี้ได้ทำการดาวน์โหลดข้อมูลดาวเทียม LANDSAT-8 จากปัจจุบัน (เมษายน 2561) ไปจนถึงปี 2556

ซึ่งในการทำ Training data เพื่อเป็นชุดข้อมูลสำหรับสอนระบบและอัลกอริทึมในการจำแนกนั้น ได้เลือกข้อมูลที่ปลอดเมฆ ที่อยู่ในช่วงระหว่างปี 2558 - 2560 เนื่องจากขั้นตอนนี้เป็นส่วนแรกของการดำเนินงานของ

โครงการ และได้เริ่มดำเนินงานมาตั้งแต่ช่วงเดือน ตุลาคม 2560 จึงทำให้ข้อมูลล่าสุดที่ใช้ในปี 2560 ซึ่งรายละเอียดในการจัดเตรียมข้อมูลแสดงได้ด้วยภาพที่ 1 และ ตารางที่ 1

No.	Scene	Date	Zone
1	12649	20170104	48
2	12650	20151030	48
3	12748	20171026	48
4	12749	20171026	48
5	12750	20151021	48
6	12751	20151224	48
7	12752	20161108	48
8	12755	20171010	48
9	12756	20160414	47
10	12847	20151028	48
11	12848	20151028	48
12	12849	20161115	48
13	12850	20161201	48
14	12851	20151231	47
15	12854	20171017	47
16	12855	20160320	47
17	12856	20160320	47
18	12947	20161122	47
19	12948	20151104	47
20	12949	20151104	47
21	12950	20151104	47
22	12951	20151104	47

No.	Scene	Date	Zone
23	12952	20170314	47
24	12953	20170314	47
25	12954	20170210	47
26	12955	20160311	47
27	13046	20171031	47
28	13047	20171031	47
29	13048	20151026	47
30	13049	20161113	47
31	13050	20161129	47
32	13051	20170201	47
33	13052	20160318	47
34	13053	20170201	47
35	13054	20170217	47
36	13055	20171031	47
37	13146	20160410	47
38	13147	20161120	47
39	13148	20161120	47
40	13149	20161120	47
41	13150	20161222	47
42	13246	20161127	47
43	13247	20161127	47

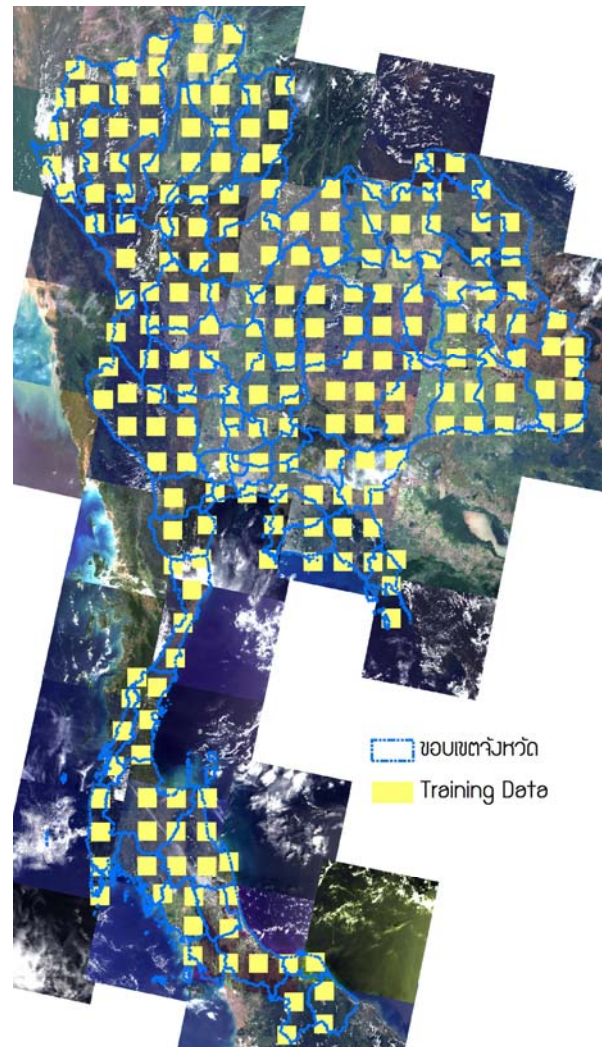
ตารางที่ 1 รายละเอียดข้อมูล LANDSAT-8 ที่ใช้ในโครงการ

2.2 การกำหนดขนาดและตำแหน่งของ Training Data

Training data คือข้อมูลตัวอย่างของสิ่งปกคลุมประเภทต่างๆ ที่ถูกจัดทำขึ้นเพื่อใช้เป็นชุดข้อมูลสำหรับสอนให้ระบบหรืออัลกอริทึมรู้จักและจดจำเพื่อนำไปใช้ในการจำแนก แยกแยะ ประเภทของสิ่งปกคลุมดินได้ โดยการกำหนดขนาดหรือขอบเขตของ Training data นั้นควรคำนึงถึงความเหมาะสมของความสามารถของระบบหรืออัลกอริทึมในการจำแนกสิ่งปกคลุมดิน นอกจากนี้ยังต้องคำนึงถึงลักษณะของภาพถ่ายจากดาวเทียมที่ใช้ด้วย โดยในโครงการ Pixel House นี้ได้ออกแบบให้ window size ของการจำแนกสิ่งปกคลุมดินจากภาพถ่ายจากดาวเทียม LANDSAT-8 มีขนาด 500x500 pixel ดังนั้นเพื่อเป็นการทดสอบประสิทธิภาพของอัลกอริทึมจึงได้กำหนดให้ขอบเขต Training data มีขนาดใหญ่กว่า window size ของอัลกอริทึมที่ใช้ จึงได้กำหนดให้มีขนาดเท่ากับ 1,000x1,000 pixel หรือมีขนาด 30 กม. x 30 กม.

โดยในแต่ละขอบเขตของ Training data นั้น ควรมีสิ่งปกคลุมดินประเภทต่างๆ ร่วมอยู่ในนั้นให้มากที่สุด ซึ่งโครงการนี้ต้องการจำแนกสิ่งปกคลุมดินทั้งหมด 9 ประเภทคือ 1.ข้าว 2.ข้าวโพด 3.

อ้อย 4.มันสำปะหลัง 5.ป่าสนน้ำมัน 6.ยางพารา 7.ป่าไม้ 8.เมือง และ 9.แหล่งน้ำ และควรกำหนดให้ Training data กระจายไปทั่วทั้งประเทศ เพื่อให้มีความหลากหลาย เหมาะสมกับการเป็นตัวแทนที่ดีของแต่ละประเภทของสิ่งปกคลุมดิน เนื่องจากสิ่งปกคลุมประเภทต่างๆ ในแต่ละพื้นที่ แต่ละภาคของประเทศมีความแตกต่างกัน เช่นการเพาะปลูก แม้ว่าจะเป็นพืชชนิดเดียวกัน แต่กิจกรรม ช่วงเวลา และลักษณะของการเพาะปลูกในแต่ละพื้นที่มีความแตกต่างกัน หรือแม้แต่ฤดูกาลก็ยังมีผลทำให้ลักษณะของสิ่งปกคลุมดินที่ปรากฏบนภาพถ่ายดาวเทียมมีความแตกต่างกัน ซึ่งในโครงการ Pixel House นี้ได้กำหนด Training data ไว้จำนวน 210 พื้นที่ กระจายไปทั่วทั้งประเทศ ดังได้แสดงในภาพที่ 2



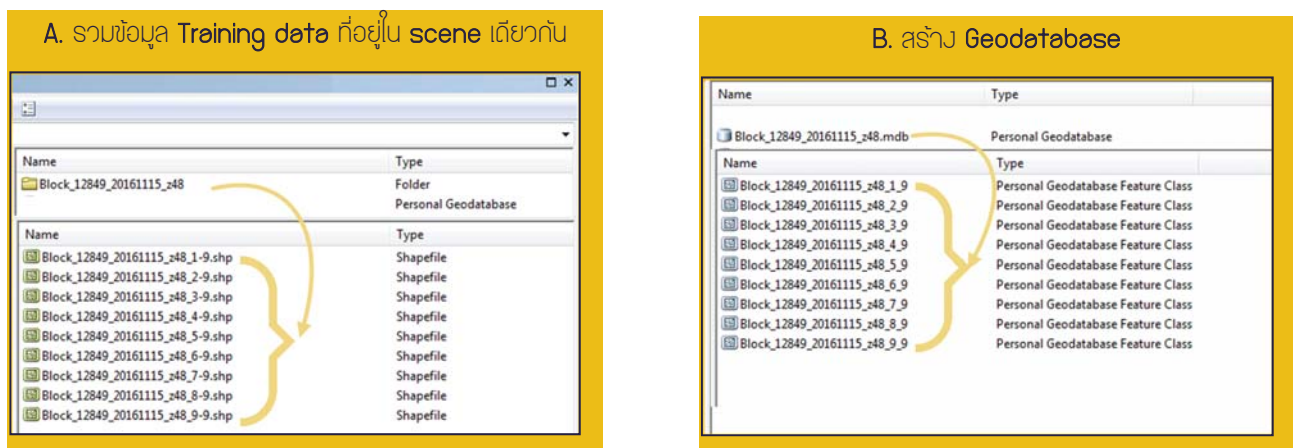
ภาพที่ 2 จำนวนและขอบของข้อมูลดาวเทียม LANDSAT-8

ที่ใช้ในการเตรียมข้อมูล

2.3 การตัดข้อมูลภาพถ่ายจากดาวเทียม

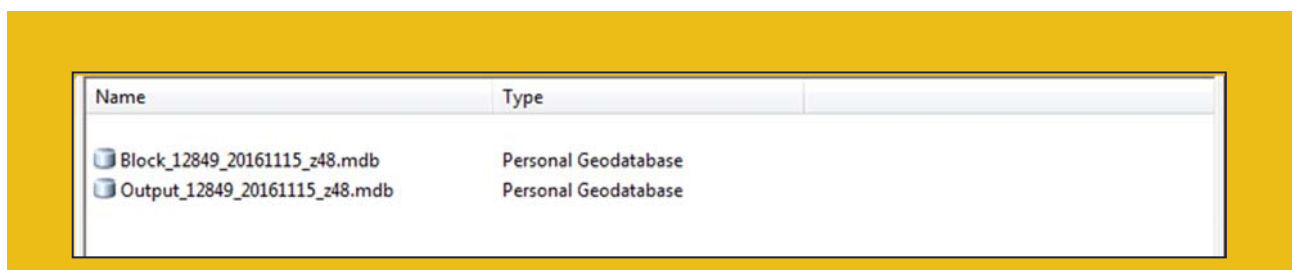
เนื่องจากข้อมูลดาวเทียม LANDSAT -8 ที่ใช้มีจำนวน 43 scene ประกอบกับมี Training data ถึง 210 พื้นที่ ซึ่งจะทำให้เสียเวลาและสิ้นเปลืองพื้นที่ในการจัดเตรียมและจัดเก็บข้อมูลเป็นอย่างมาก หากต้องการตัดแบ่งข้อมูล LANDSAT-8 ให้ครบทุกพื้นที่ของ Training data แบบ 1:1 ดังนั้นเพื่อให้การจัดเตรียมและจัดเก็บข้อมูล Training data มีประสิทธิภาพมากที่สุดจึงได้ดำเนินการตามขั้นตอนดังนี้ (โดยในส่วนของการตัดข้อมูลภาพถ่ายจากดาวเทียมนี้ใช้โปรแกรม ArcMap ในการดำเนินการ)

2.3.1 กำหนดรูปแบบของ Training data โดยเมื่อทำการคัดเลือกข้อมูลจากดาวเทียม LANDSAT-8 และกำหนดขอบตำแหน่งและเขตของ Training data ตามข้อ 2.1 และ 2.3 ได้แล้ว ก็ทำการรวมข้อมูล Training data ที่อยู่ใน scene ของภาพถ่ายจากดาวเทียมเดียวกันเอาไว้ด้วยกัน เพื่อความสะดวกในการสืบค้นข้อมูลภาพถ่ายจากดาวเทียม โดยทำการสร้าง Geodatabase ขึ้นมาเพื่อจัดเก็บขอบเขตและตำแหน่งของ Training data พร้อมทั้งกำหนดชื่อให้เชื่อมโยงกับ scene และวันที่ยับข้อมูลของดาวเทียม LANDSAT-8 ดังแสดงได้ในภาพที่ 3

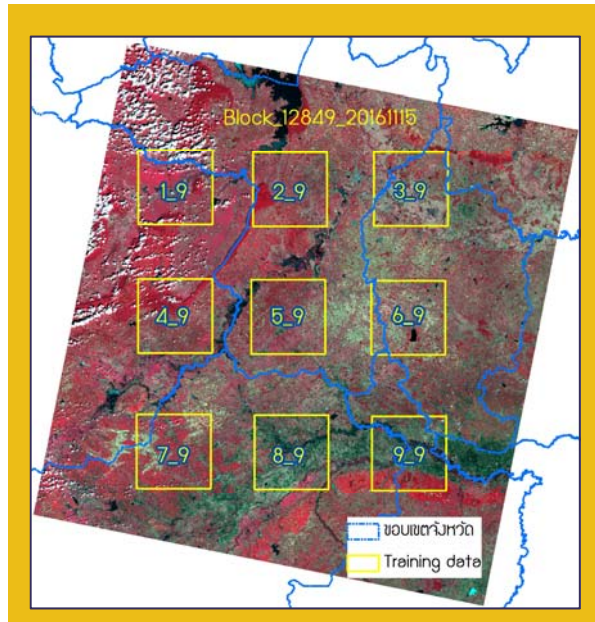


ภาพที่ 3 การกำหนดรูปแบบ และการจัดเก็บข้อมูล Training data

2.3.2 ในขั้นตอนนี้จะเป็นการสร้าง Geodatabase เพื่อรองรับ output หรือข้อมูลภาพถ่ายจากดาวเทียม ที่ผ่านการตัดตามขอบเขตของ Training data โดยกำหนดชื่อให้สอดคล้องกับ training data และภาพถ่ายจากดาวเทียม LANDSAT-8 เช่นเดียวกับการสร้าง Geodatabase ให้กับ Training data แสดงได้ดังภาพที่ 4 และภาพที่ 5



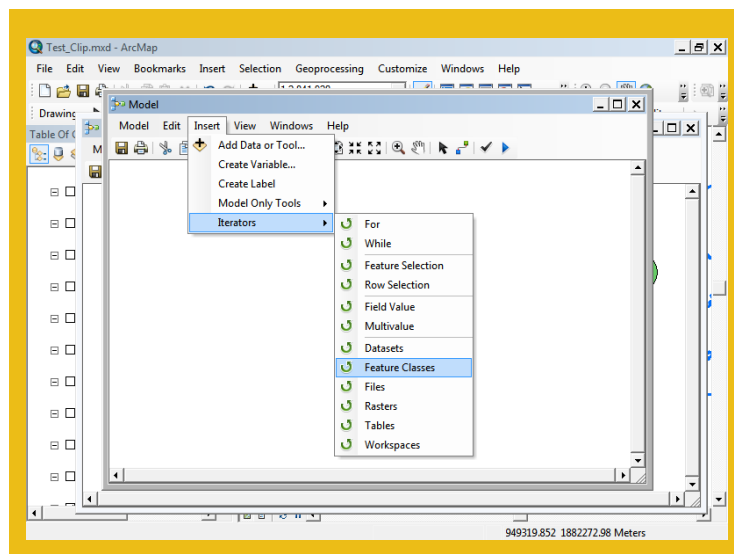
ภาพที่ 4 การสร้าง Geodatabase เพื่อรองรับ output



ภาพที่ 5 ตัวอย่างขอบเขตและตำแหน่งของ Training data บนภาพถ่ายจากดาวเทียม LANDSAT-8

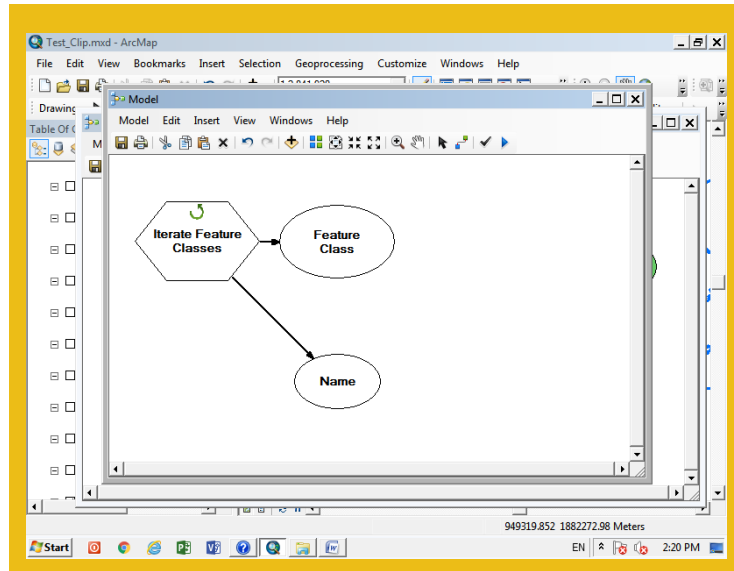
2.3.3 เมื่อทำการจัดเตรียมข้อมูล input และไฟล์ที่รองรับ output เรียบร้อยแล้ว ต่อไปคือขั้นตอนการตัดข้อมูลภาพถ่ายจากดาวเทียม LANDSAT-8 ตามขอบเขตของ Training data ที่กำหนดไว้ โดยมีทั้งหมด 210 พื้นที่ใน 43 scene ของภาพถ่ายจากดาวเทียม LANDSAT-8 ดังนั้นเพื่อเป็นการประหยัดเวลา และประหยัดพื้นที่ในการจัดเก็บข้อมูล จึงได้ทำการตัดข้อมูลในโปรแกรม ArcMap ด้วยการออกแบบให้ทำการตัดแบบกึ่งอัตโนมัติแบบราย scene ตามข้อมูลของ LANDSAT-8 ด้วยการสร้าง ModelBuilder ดังนี้

- สร้าง ModelBuilder โดยการเลือก Geoprocessing > ModelBuilder เมื่อหน้าต่าง Model x ปรากฏขึ้น ให้ทำการเลือก Iterators > Feature Classes ดังภาพที่ 6 เพื่อเป็นการกำหนดเงื่อนไขในการประมวลผลการดำเนินการจาก feature Classes ที่เป็นชุด



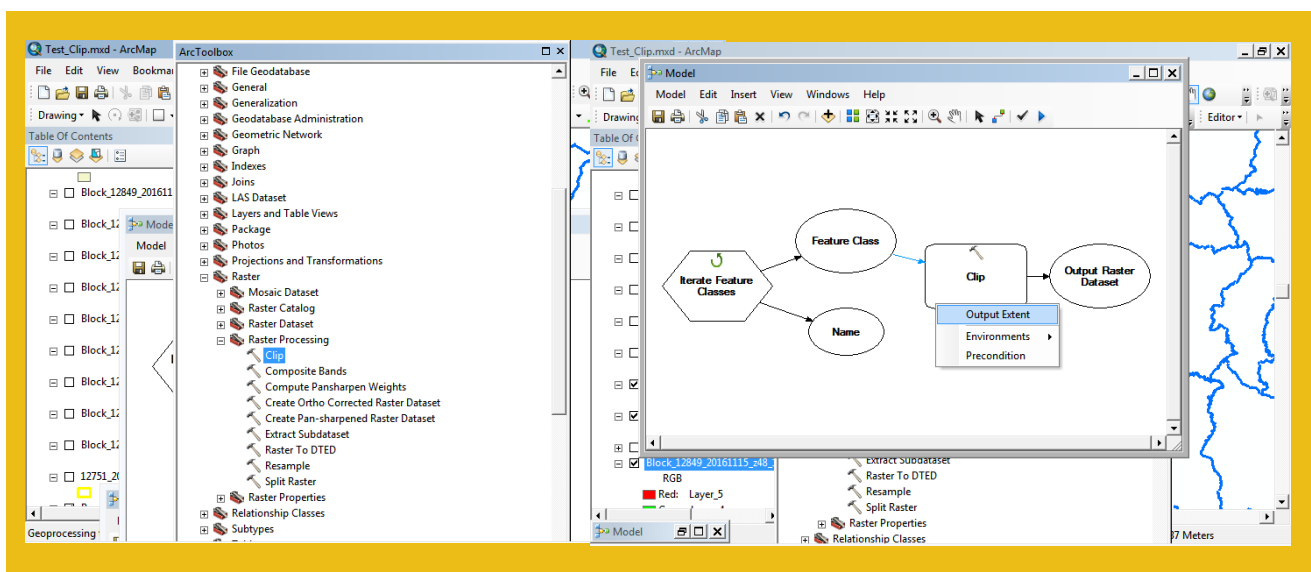
ภาพที่ 6 การสร้าง Model Builder

- ดับเบิลคลิกในกล่อง Iterate Feature Classes โดยเลือกข้อมูลขอบเขต Training data ที่ได้จัดทำไว้ในรูปแบบ Geodatabase เพื่อใช้เป็นข้อมูล Output Extent ในการดำเนินการตัดข้อมูล LANDSAT-8 ดังภาพที่ 7



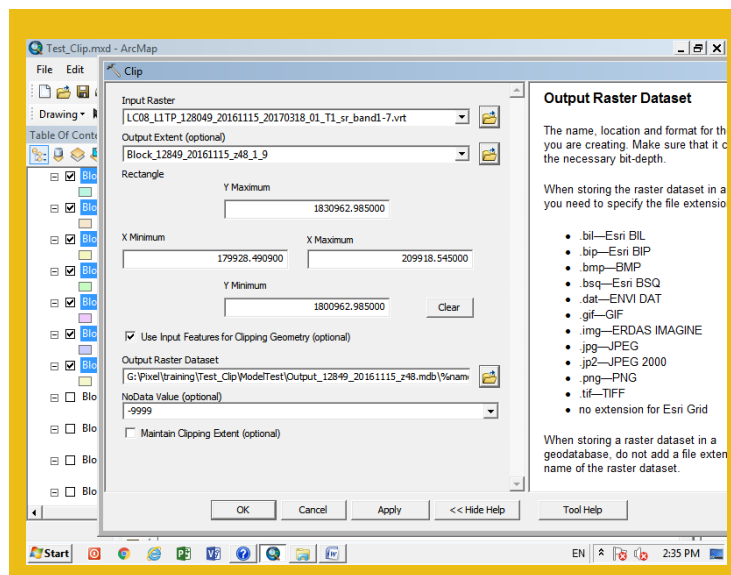
ภาพที่ 7 กำหนดขอบเขตที่จะทำการตัดข้อมูล (Output Extent)

- จากนั้นไปที่ Arc Tool Box เพื่อเลือกเครื่องมือในการตัดข้อมูลภาพถ่ายจากดาวเทียมให้มีขนาดเท่ากับขอบเขตและตำแหน่งของ Training data ที่ได้กำหนดไว้ ArcToolBox > Data Management Tool > Raster > Clip โดยคลิกลากเครื่องมือ Clip เข้ามาไว้ในกล่อง Model และทำการเชื่อมต่อข้อมูล Feature Class ของ Iterate Feature Classes เข้ากับ ฟังก์ชัน Clip ที่เพิ่งลากเข้ามาด้วยเครื่องมือ Connect แล้วเลือกเป็น Output Extent ดังภาพที่ 8



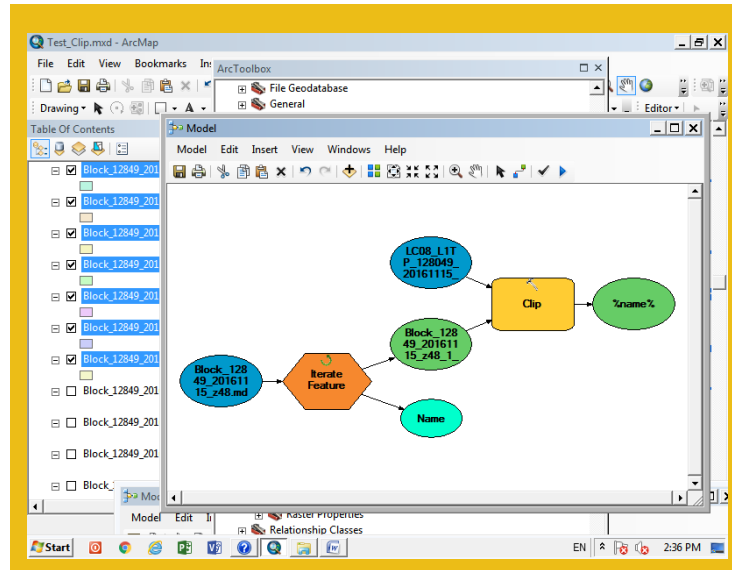
ภาพที่ 8 ฟังก์ชันในการตัดข้อมูล

- กำหนดเงื่อนไขในส่วนของการตัดข้อมูลโดย Input Raster คือ scene ของข้อมูลดาวเทียม LANDSAT-8 ที่จะทำการตัด (ทำที่ละ scene) ส่วน Output Extent คือข้อมูล Geodatabase ของขอบเขต Training data ที่ได้ทำการจัดเตรียมไว้แล้ว และ Output Raster Data Set คือ Output Geodatabase ที่ได้เตรียมไว้ในตอนต้น โดยในขั้นตอนนี้ให้ทำการเลือก Use Input Features for Clipping Geometry (optional) เพื่อให้ผลลัพธ์ที่ได้มีขนาดเท่ากับขอบเขตที่เลือกไว้พอดี นอกจากนี้ในส่วนของ Output Raster Dataset นั้น หลังจากที่ได้เลือกไฟล์ Output Geodatabase ที่ได้ทำการเตรียมไว้แล้วนั้น ในส่วนของชื่อไฟล์ให้ตั้งเป็น %name% เพื่อให้ผลลัพธ์ที่ได้มีชื่อตรงกับชื่อของขอบเขต Training data ที่ได้ตั้งไว้ เมื่อกำหนดเงื่อนไขเรียบร้อยแล้วก็ให้ทำการประมวลผลข้อมูลโดยกด Run เพื่อทำการประมวลผล และตรวจสอบผลลัพธ์ต่อไป โดยรูปที่ 9 เป็นรายละเอียดในการกำหนดเงื่อนไขต่างๆ



ภาพที่ 9 การกำหนดเงื่อนไขในการตัดข้อมูลดาวเทียม LANDSAT-8

- ภาพที่ 10 แสดงภาพรวมของการกำหนดโมเดลแบบกึ่งอัตโนมัติในการตัดข้อมูล LANDSAT-8 โดยใช้โมเดลดำเนินการกับข้อมูลที่เหลืออีก 42 scene โดยทำการเปลี่ยนเงื่อนไขตามข้อมูลที่เปลี่ยนไป ส่วนรูปที่ 11 คือผลลัพธ์ที่ได้โดยผลลัพธ์ที่ได้นั้นจะถูกตัดให้มีขนาดเท่ากับขอบเขตของ Training data และจัดเก็บให้อยู่ในรูปแบบ Geodatabase จึงเป็นไฟล์ขนาดเล็ก เหมาะสมกับนำไปใช้จำแนกสิ่งปกคลุมดินด้วยสายตา เพราะจะทำให้ไม่เสียเวลาและพื้นที่ในการโหลด ผสมแบนด์ และประมวลผล



ภาพที่ 10 ภาพรวมของโมเดลที่ใช้ในการตัดข้อมูล LANDSAT-8 ตามขนาดขอบเขตของ Training data



ภาพที่ 11 ภาพรวมของโมเดลที่ใช้ในการตัดข้อมูล LANDSAT-8 ตามขนาดขอบเขตของ Training data

3. การจำแนกข้อมูลสิ่งปกคลุมดิน

ในโครงการ **Pixel House** นี้ได้กำหนดให้มีการจำแนกสิ่งปกคลุมดินจากภาพถ่ายจากดาวเทียม **LANDSAT-8** ออกเป็น 9 ชนิด ตามความสำคัญของทรัพยากรและพืชเศรษฐกิจในประเทศ คือ 1.ข้าว 2.ข้าวโพด 3.อ้อย 4.มันสำปะหลัง 5.ปาล์มน้ำมัน 6.ยางพารา 7.ป่าไม้ 8.เมือง และ 9.แหล่งน้ำ โดยในการแปลตีความด้วยสายตาจากภาพถ่ายดาวเทียมนี้ต้องอาศัยหลักการและองค์ประกอบหลายอย่างเพื่อความถูกต้องแม่นยำของผลลัพธ์ที่ได้ แต่ในส่วนนี้ขอนำเสนอเฉพาะขั้นตอนการดำเนินการเลือก **Training data** ทั้ง 9 ประเภทที่ใช้ในโครงการนี้

3.1 การจัดทำข้อมูล และการเปรียบเทียบข้อมูลอ้างอิง

- ทำการสร้างฐานข้อมูลไฟล์จัดเก็บ Shapefile โดยตั้งชื่อแต่ละ Scene
- สร้าง Shapefile โดยตั้งชื่อเป็น ชื่อภาพดาวเทียม-ชื่อ Scene - วันที่ถ่ายภาพ-จำนวนบล็อกที่ทำ เช่น LS8-12649-20170401-LC1-9 (LC1-9 คือ บล็อกที่ 1 จากทั้งหมด 9 บล็อก)

Geometry Type: Polygon

Projected Coordinate System: WGS_1984_UTM_Zone_47N

- สร้างคุณลักษณะข้อมูล (Attribute) Addfield ชื่อ Name / Type : Text และ Code / Type : Short Integer ซึ่ง Code จะยึดตาม Attribute Landuse: LU_ID_L3 กรมพัฒนาที่ดิน โดย Code ทั้ง 9 ประเภท มีดังนี้

นาข้าว (Active paddy field) = 2101

ข้าวโพด (Corn) = 2202

มันสำปะหลัง (Cassava) = 2204

อ้อย (Sugarcane) = 2203

ยางพารา (Para rubber) = 2302

ปาล์มน้ำมัน (Oil palm) = 2303

ป่าไม้ (Forest) = 3000

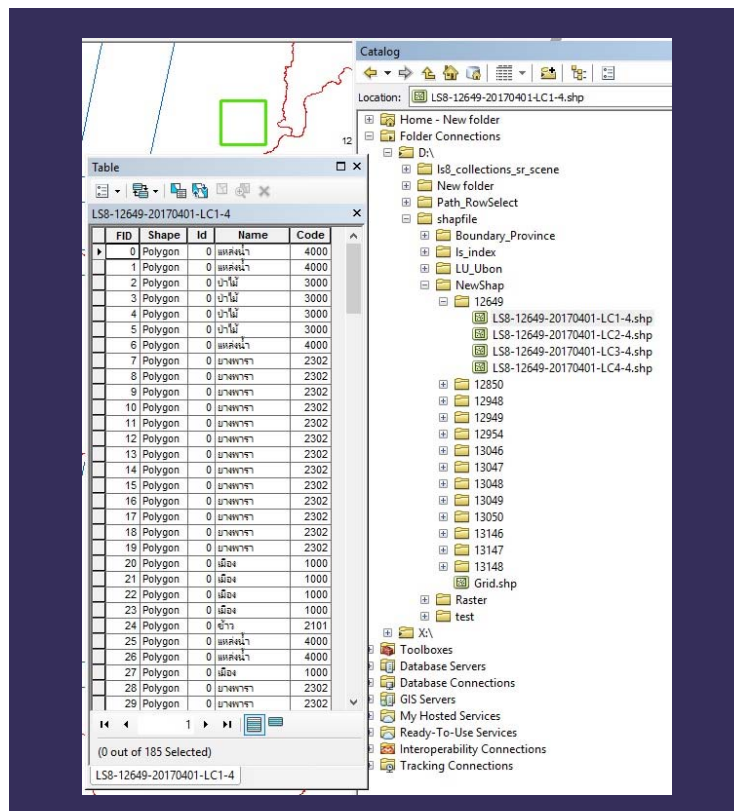
แหล่งน้ำ (water) = 4000

เมือง (Urban) = 1000

- สร้าง Symbology ไว้ก่อนโดย Add Values ชื่อคลาสทั้ง 9 ประเภท ตั้งค่าสีและเส้น
- Edit Shapefile เพื่อ Digitize ข้อมูลในแต่ละบล็อก โดยเลือกประเภทข้อมูลแต่ละอย่างที่แตกต่างกัน และมีลักษณะที่ตรงกับข้อมูล โดยพิจารณาทั้งรูปร่าง ลักษณะ สี และอื่นๆ หากลักษณะไม่ตรงหรือข้อมูลไม่เพียงพอจะไม่ทำการ Digitize ข้อมูล การ Digitize จะ Digitize ให้อยู่ในกรอบของรูปร่าง ลักษณะ สี และอื่นๆ ไม่ให้ล้นออกมาปะปนกับรูปร่าง

ลักษณะ สี ของพิกเซลอื่นๆ และพิจารณาลักษณะของการอยู่รวมกันเป็นกลุ่มที่เป็นเนื้อ
รูปร่าง ลักษณะ สี โดยเฉพาะพืชเศรษฐกิจที่มีลักษณะรูปร่างคล้ายกัน แต่การผสมสีแต่ละ
สีของพืชจะแตกต่างกันพอสมควร

- ข้อมูลอ้างอิงจะใช้ข้อมูล Landuse ของ GISTDA Portal 25k เพื่อจัดทำแนกรายละเอียด
การใช้ประโยชน์ที่ดินได้มากขึ้นของแต่ละภาค เปรียบเทียบกับ Basemap ภาพดาวเทียม
ความคมชัดสูง โดยรูปที่ 12 เป็นตัวอย่างการสร้าง Attribute ให้กับข้อมูล Training
data



ภาพที่ 12 ตัวอย่าง Attribute ของ Training data

3.2 การจำแนกข้าว

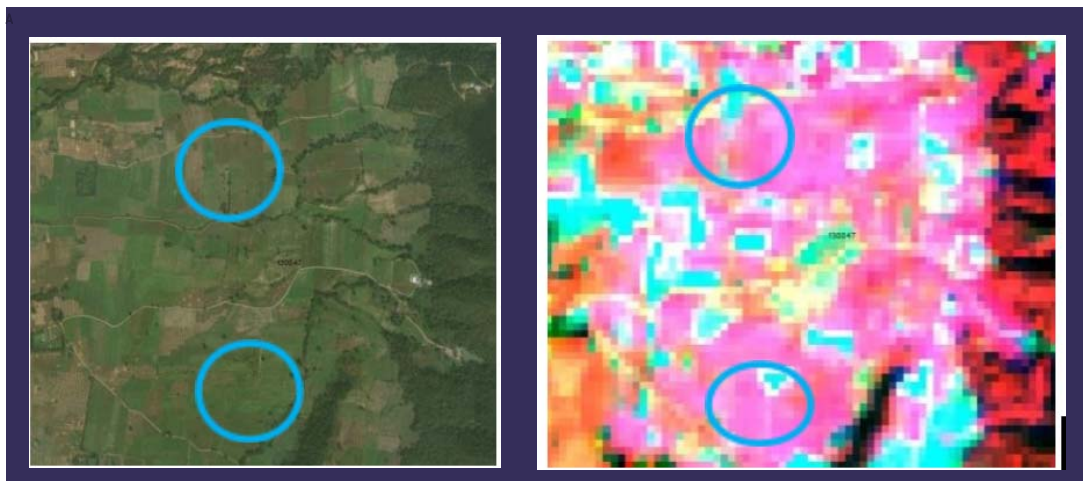
รูปแบบของนาข้าวส่วนใหญ่ภาพจะไม่ได้อยู่ในช่วงฤดูฝนที่มีเมฆมาก ถ้าหากผสมสี 541 ข้าว
จะมีสีเหลืองส้ม, ส้มอ่อน, ส้มเข้มไปจนถึงส้มแดง ขึ้นอยู่กับอายุของข้าวตั้งแต่ 6-8 สัปดาห์ ถึง 12-16
สัปดาห์ หากมีน้ำท่วมขังจะเห็นเป็นสีเขียว, เขียวอมฟ้าไปจนถึงฟ้าน้ำเงินเข้ม แต่หากเก็บเกี่ยวแล้วจะ
เป็นสีขาวที่เป็นส่วนของดิน การสังเกตจะสามารถดูได้จากคันดินที่เห็นเป็นรูปทรงสี่เหลี่ยมโดยคันดิน
มักจะทำตามแนวระดับหรือตั้งฉากกับการไหลของน้ำ และพื้นที่ส่วนมากจะอยู่ในพื้นที่ราบลุ่มทำการ
สร้างฐานข้อมูลไฟส์จัดเก็บ Shapefile โดยตั้งชื่อแต่ละ Scene โดยภาพที่ 13 แสดงลักษณะของ
พื้นที่ปลูกข้าวที่ปรากฏบนภาพถ่ายจากดาวเทียม



ภาพที่ 13 ลักษณะของพื้นที่ปลูกข้าวที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.3 การจำแนกข้าวโพด

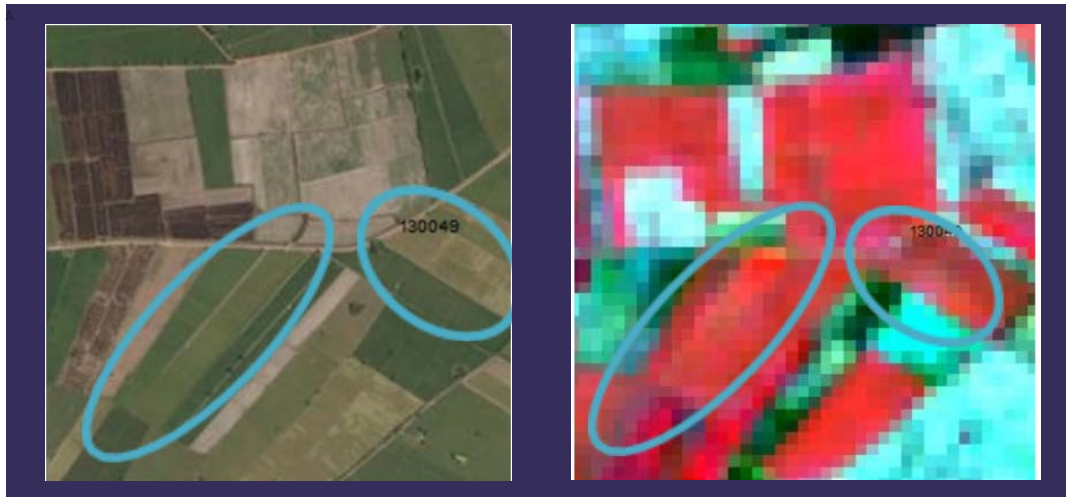
ข้าวโพดเป็นพืชไร่ ส่วนใหญ่ปลูกบริเวณเนินเขา หรือมีการระบายน้ำได้ดี รูปแบบเป็นผืนเป็นแปลง มีความหยابแบบละเอียดเรียบ มีลักษณะสีเขียวอ่อน หากผสมสี 541 จะมีลักษณะสีชมพูอ่อน, สีส้มพู ดังภาพที่ 14



ภาพที่ 14 ลักษณะของพื้นที่ปลูกข้าวโพดที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.4 การจำแนกอ้อย

อ้อยเป็นพืชไร่ ปลูกในพื้นที่การระบายน้ำดี สามารถปลูกในที่ลุ่มหรือปลูกในบริเวณนาข้าวได้ รูปแบบเป็นผืนเป็นแปลงมีความหยابแบบละเอียดเรียบ มีลักษณะสีเขียวอ่อนไปจนถึงสีเขียวเข้ม ขึ้นอยู่กับอายุของพืช หากผสมสี 541 จะมีลักษณะเป็นสีส้ม, สีส้มแดง ดังภาพที่ 15



ภาพที่ 15 ลักษณะของพื้นที่ปลูกอ้อยที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.5 การจำแนกมันสำปะหลัง

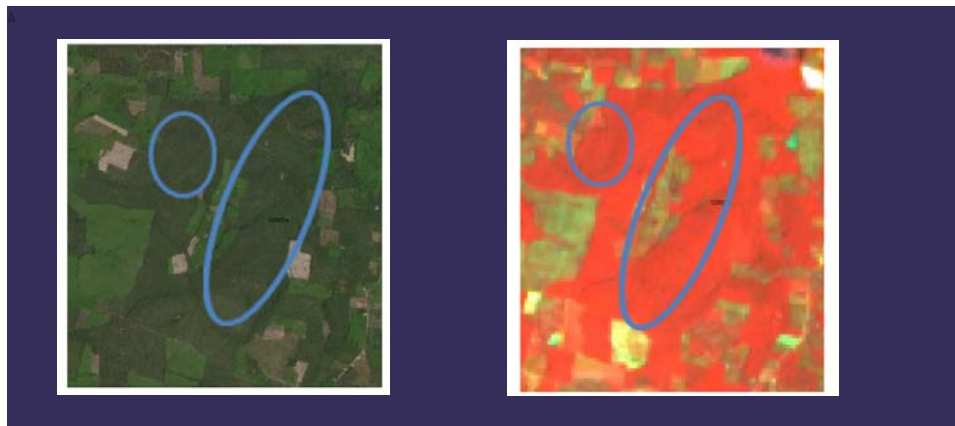
มันสำปะหลังเป็นพืชไร่ ปลูกในพื้นที่ที่สามารถปลูกได้ง่ายในพื้นที่ร้อน และร้อนชื้น มีความหยาดแบบละเอียดยเรียบ รูปแบบเป็นผืนเป็นแปลงมีเหลี่ยมด้านกว้างและยาวมากกว่าอ้อยและข้าวโพด จะมีสีเขียว, สีเขียวแก่ หากผสมสี 541 จะเป็นสีชมพู, สีชมพูเข้มอ้อยเป็นพืชไร่ ปลูกในพื้นที่การระบายน้ำดี สามารถปลูกในที่ลุ่มหรือปลูกในบริเวณนาข้าวได้ รูปแบบเป็นผืนเป็นแปลงมีความหยาดแบบละเอียดยเรียบ มีลักษณะสีเขียวอ่อนไปจนถึงสีเขียวเข้มขึ้นอยู่กัอายุของพืช หากผสมสี 541 จะมีลักษณะเป็นสีส้ม, สีส้มแดง ดังภาพที่ 16



ภาพที่ 16 ลักษณะของพื้นที่ปลูกมันสำปะหลังที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.6 การจำแนกปาล์มน้ำมัน

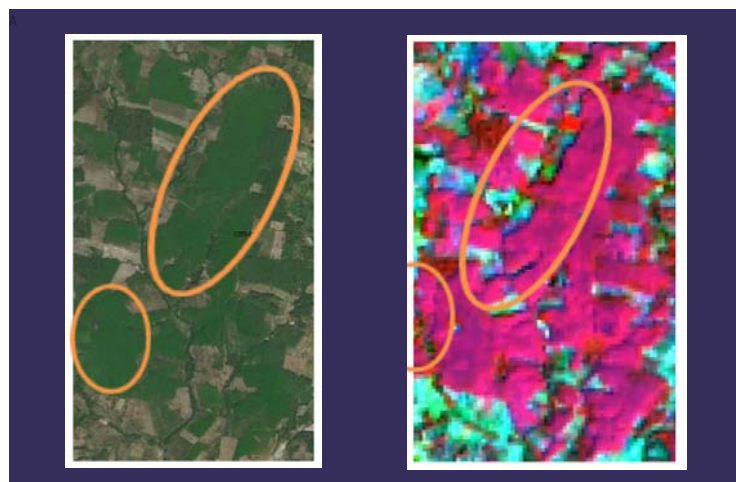
ปาล์มน้ำมัน เป็นไม้ยืนต้น สามารถพบได้ตลอดทั้งปี โดยทั่วไปจะเห็นสีคล้ำกว่าพืชไร่ และจะพบเงาซึ่งแสดงถึงความสูงของต้นไม้ที่แตกต่างจากพืชไร่ที่มีความสูงน้อยกว่า มีความหยาบแบบละเอียดเรียบ รูปแบบเป็นผืนเป็นแปลงมีเหลี่ยม ซึ่งจะมีสีเขียวแก่ หากผสมสี 541 จะเป็นสีส้มเข้ม, ส้มแดง ดังภาพที่ 17



ภาพที่ 17 ลักษณะของพื้นที่ปลูกปาล์มน้ำมันที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.7 การจำแนกยางพารา

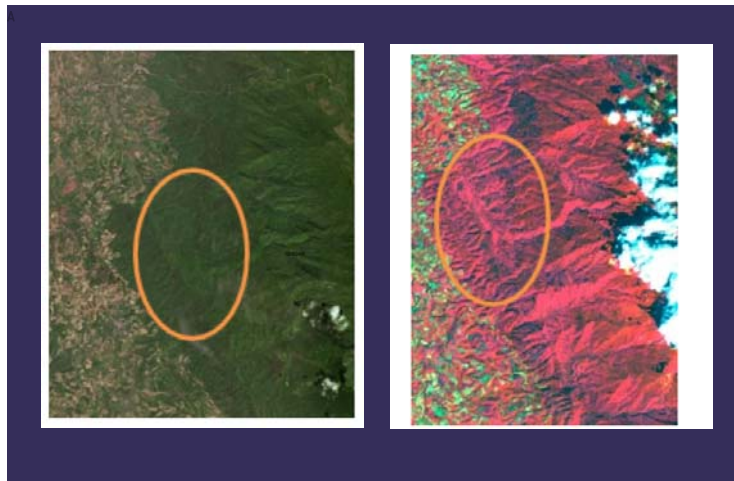
ยางพาราเป็นไม้ยืนต้น สามารถพบได้ตั้งแต่ช่วงฤดูฝนไปจนถึงปลายฤดูฝน สีคล้ำกว่าพืชไร่ และจะพบเงาซึ่งแสดงถึงความสูงของต้นไม้มีความหยาบแบบละเอียดเรียบ รูปแบบเป็นผืนเป็นแปลงมีเหลี่ยม ซึ่งจะมีสีเขียว หากผสมสี 541 จะเป็นสีชมพูเข้ม, สีชมพูม่วง แต่ยางพารามีช่วงผลัดใบในฤดูแล้ง หากผสมสี 541 จะเป็นสีเหลืองขาว, เหลืองเขียว ดังภาพที่ 18



ภาพที่ 18 ลักษณะของพื้นที่ปลูกยางพาราที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.8 การจำแนกป่าไม้

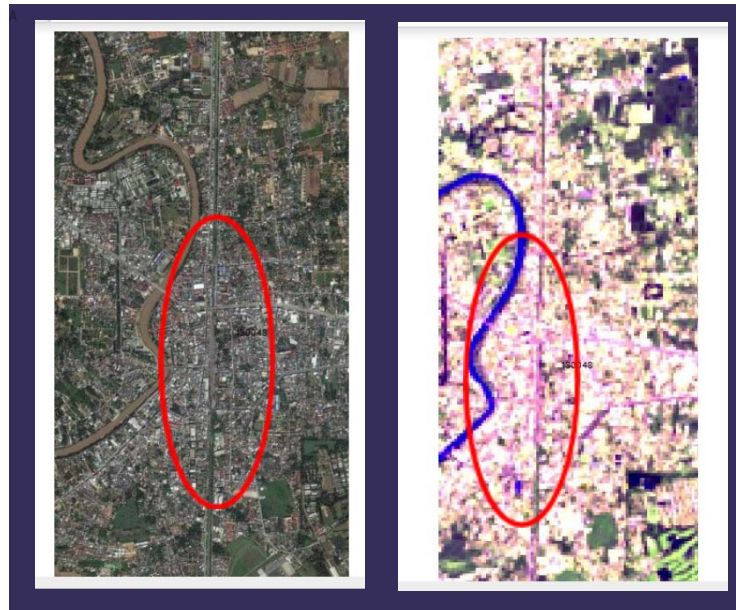
พื้นที่ป่าไม้ จะประกอบด้วยป่าชนิดต่างๆ จำ นวนมาก ซึ่งสามารถอาศัยลักษณะเฉพาะของพื้นที่ป่าช่วยในการวิเคราะห์ได้ เช่น ป่าดิบเขา จะพบเมื่อพื้นที่มีความสูงกว่าระดับน้ำทะเลปานกลาง ประมาณ 1,000 เมตร หรือ พื้นที่ป่าไม้ในภาคใต้ส่วนใหญ่จะเป็นป่าดิบชื้น เนื่องจาก ปริมาณ น้ำฝนเกินกว่า 1,600 มิลลิเมตรต่อปี ปริมาณน้ำฝนจึงเป็นปัจจัยหนึ่งในการกำหนดชนิดป่า โดยทั่วไป จะพบพื้นที่ป่าไม้บนพื้นที่สูง และในพื้นที่ชายฝั่งทะเลบริเวณที่น้ำทะเลท่วมถึงมีระดับการขึ้นลงของน้ำทะเล จะเป็นป่าชายเลน นอกจากนี้ฤดูแล้งป่าไม้ชนิดผลัดใบจะมีการผลัดใบร่วงลงสู่พื้น มีลักษณะ ความหยาบ ความสูงของพื้นผิวและกรอบด้านนอกมีลักษณะไม่สม่ำเสมอ ภาพดาวเทียมคมชัดสูงจะเห็นเป็นสีเขียวเข้มจนถึงสีเขียว หากผสมสีภาพ Landsat 8 เป็น 541 สีจะเป็นสีแดงเข้ม,แดง ดำสลับกับจนถึงสีแดง ดังภาพที่ 19



ภาพที่ 19 ลักษณะของพื้นที่ป่าไม้ที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.9 การจำแนกพื้นที่เมือง

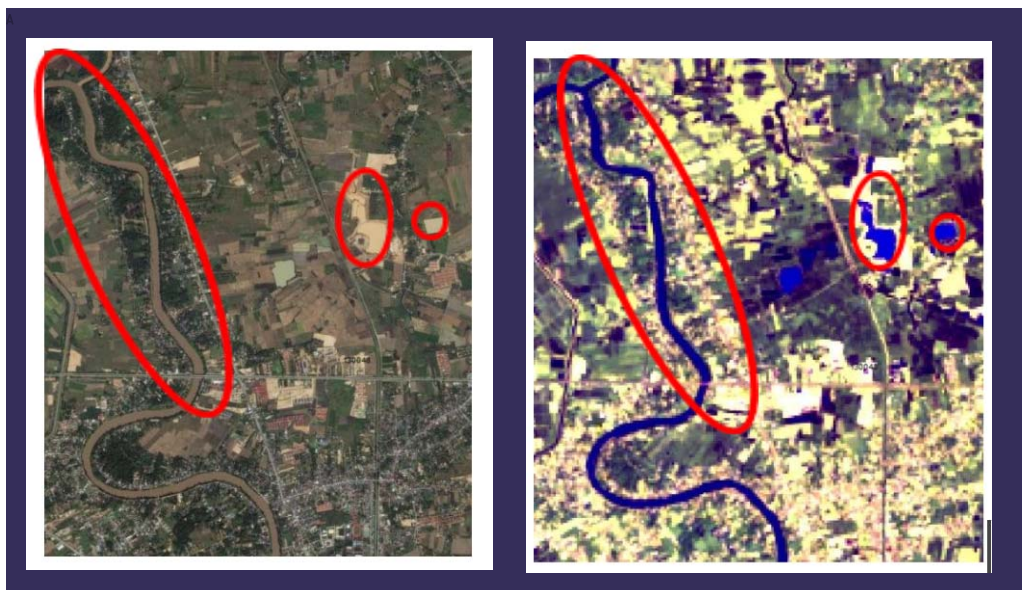
เมือง จะเห็นรูปแบบของอาคารที่หนาแน่นมีการใช้ประโยชน์ที่ดินที่หลากหลาย รวมทั้งการปลูกพืช ไม้ยืนต้นผสมรวมกันอยู่ รวมทั้งพบเส้นตรงถนนตัดผ่าน ขนาดของรูปทรงเหลี่ยมไม่เท่ากัน แสดงให้เห็นถึงการใช้ประโยชน์ที่ดินที่แตกต่างกันในย่านเมือง เช่น พื้นที่อุตสาหกรรม พื้นที่ย่านการค้า สถานที่ราชการ และที่อยู่อาศัยที่มีลักษณะเป็นบ้านเดี่ยวและบ้านจัดสรร เป็นต้น ดาวเทียม ความคมชัดสูงจะแสดงให้เห็นถึงสีผิวและสีอื่นๆที่เป็นประเภทของที่อยู่อาศัย ย่านการค้า รวมทั้งถนนที่ตัดผ่าน และสีเขียวแสดงให้เห็นถึงพืชและไม้ยืนต้นที่ปลูกสลับกัน หากผสมสีภาพ Landsat 8 เป็น 764 จะเห็นประเภทของที่อยู่อาศัย ย่านการค้าเป็นสีเขียวสลับชมพู และถนนเป็นสีชมพูอ่อนแทรกสลับกับพืช, ไม้ยืนต้นเป็นสีเขียวอ่อน, สีเขียวเข้ม ดังนั้นย่านตัวเมืองจะมีลักษณะสีพิกเซลละเอียดสลับไปมา ดังภาพที่ 20



ภาพที่ 20 ลักษณะของพื้นที่เมืองที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)

3.10 การจำแนกแหล่งน้ำ

แหล่งน้ำ ประกอบด้วยแหล่งน้ำตามธรรมชาติ และแหล่งน้ำที่สร้างขึ้น 1)พื้นที่น้ำที่เกิดขึ้นตามธรรมชาติ เช่น แม่น้ำ ลำธาร หนอง บึง 2)แหล่งน้ำที่สร้างขึ้นโดยมนุษย์ เช่น เขื่อน อ่างเก็บน้ำ บ่อน้ำในไร่นาต่างๆ เป็นต้น ลักษณะระหว่างพื้นที่น้ำทั้งสองประเภทแตกต่างกันโดยสามารถสังเกตจากแหล่งน้ำธรรมชาติ ขอบเขตพื้นที่ไม่แน่นอน ไม่มีรูปแบบ ของเส้นตรง หรือทรงสี่เหลี่ยม ส่วนแหล่งน้ำที่สร้างขึ้นจะมีรูปแบบตามภูมิประเทศ รูปแบบการสร้างที่แน่นอน มีเหลี่ยม การผสมสีจะอาศัยลักษณะของสีให้มีความแตกต่างกับเงาเงาของเมฆหรือน้ำขุ่น น้ำท่วมในนาหรือแปลงปลูกพืช ซึ่งจะผสมสีอยู่ระหว่าง 761,762 และ 763 ส่วนลักษณะการสะท้อนแสงหากน้ำมีความลึก ความใสของน้ำจะมีสีน้ำเงินเข้ม แต่ถ้าหากน้ำตื้น จะมีสีน้ำเงิน และถ้าหากมีความขุ่นของน้ำด้วยจะมีสีน้ำเงินฟ้า,สีฟ้า นอกจากนี้แหล่งน้ำที่เป็นทางน้ำ แม่น้ำ ลำธาร จะต้องมีความกว้างเกิน 60 เมตร ขึ้นไป หรือประมาณ 2 พิกเซลขึ้นไป ดังภาพที่ 21



ภาพที่ 21 ลักษณะของแหล่งน้ำที่ปรากฏบนภาพถ่ายจากดาวเทียมรายละเอียดสูง (ซ้าย) และภาพถ่ายจากดาวเทียม LANDSAT-8 (ขวา)