บทที่ 2

การตรวจเอกสาร

2.1 การสำรวจและจำแนกดิน

ประเทศสหรัฐอเมริกามีการสำรวจดินครั้งแรกโดย The U.S. Department of Agriculture Appropriations Act สำหรับงบประมาณ ปี ค.ศ. 1896 ศึกษาความสัมพันธ์ของดิน ภูมิอากาศกับสิ่งมีชีวิต เนื้อดิน องค์ประกอบของเนื้อดินในสนาม และในห้องปฏิบัติการ มุ่งเน้นเพื่องานทางเกษตรกรรม การปศุสัตว์ และป่าไม้ แล้วพัฒนากว้างขวางครอบคลุมงานด้านวิศวกรรมกระทั่งปัจจุบัน (Soil Survey Division Staff,1993)

ประเทศไทยเริ่มมีการสำรวจและจำแนกดินโดย Dr. R. L. Pendleton ดร. สาโรช มนตระกูล และ ดร. เริ่ม บูรณฤกษ์ โดยจำแนกดินตามระบบของสหรัฐอเมริกา แต่เป็นการสำรวจที่ยังไม่ละเอียดนัก แผนที่ที่ทำขึ้นเป็นแผนที่มาตราส่วน 1:2,500,000 หน่วยของแผนที่ดินอยู่ในระดับกลุ่มดินหลักและหน่วยดินสัมพันธ์ (association) ของกลุ่มดินหลัก (great soil group) ถึงแม้มีการแบ่งถึงขั้นชุดดินบ้างแต่ไม่มาก การปรับปรุงระบบการสำรวจและจำแนกดินมีเรื่อยมาจนกระทั่งปี พ.ศ. 2504 ประเทศไทยได้รับความร่วมมือจากองค์การอาหารและเกษตรแห่งสหประชาชาติ (FAO) โดย Dr. F. R. Moormann ได้ทำการสำรวจดินเป็นรายจังหวัด ต่อมาปีในปี พ. ศ. 2506 รัฐบาลไทยได้จัดตั้งกรมพัฒนาที่ดินขึ้นและได้มีการปรับปรุงแผนที่รวมทั้งได้จัดพิมพ์แผนที่ดินมาตราส่วน 1:250,000 (Moormann and Rojanasoonthon, 1968) โดยยังคงจำแนกในระดับกลุ่มดินหลักและหน่วยดินสัมพันธ์ของกลุ่มดินหลัก ปัจจุบันจึงมีแผนที่มาตราส่วนระดับจังหวัด 1:100,000 ซึ่งเป็นหน่วยแผนที่ดินระดับชุดดิน ประเภท (phase) ของชุดดิน หน่วยดินสัมพันธ์ของชุดดิน และหน่วยดินเบ็ดเตล็ด (miscellaneous) นอกจากนี้ยังได้มีการสำรวจข้อมูลดินขั้นกึ่งละเอียดในบางบริเวณเพื่อให้ข้อมูลมีความละเอียดมากขึ้น ช่วงแรกใช้การจำแนกลักษณะกลุ่มดินหลัก (Dudal and Moormann, 1964) ที่ต่อมาได้พัฒนามาเป็นระบบอนุกรมวิธานดิน 1975 โดยพิมพ์แผนที่ดินระดับใช้งานทั่วไป มาตราส่วน 1:1,000,000 นอกจากนั้นยังได้จัดพิมพ์แผนที่ดินระดับภาค มาตราส่วน 1:500,000 (เอิบ, 2542)

กล่าวได้ว่าการสำรวจดินของประเทศไทยเกือบทั้งหมดดำเนินการโดยกองสำรวจและจำแนกดิน กรมพัฒนาที่ดิน กระทรวงเกษตรและสหกรณ์ โดยความละเอียดที่จัดทำขึ้นนั้นผันแปรไปตามวัตถุประสงค์ในการใช้งานด้านต่างๆ ได้แก่ (1) *การสำรวจแบบละเอียดมากพิเศษ* (Special

Very detailed Soil Surveys) การสำรวจลักษณะนี้จะให้ข้อสนเทศทางดินละเอียดมาก ใช้ในงานซึ่งต้องมีการประเมินลักษณะของดินเข้มข้นมาก เน้นบริเวณเล็กๆ มาตราส่วนของแผนที่พื้นฐานต้องใหญ่กว่า 1:5,000 (Dent and Changprai, 1973) หน่วยของแผนที่ดินที่ใช้ เช่น ประเภทของชนิดดิน (phase of soil type) ของชุดดิน (soil series) หรือของหน่วยดินคล้าย (soil variants) ที่กำหนดช่วงความแตกต่างของประเภทดินอย่างแคบ (2) *การสำรวจแบบละเอียดมาก* (Very Detailed Soil Surveys) ให้ข้อสนเทศของดินอย่างละเอียดมากเช่นกัน แต่มีจุดประสงค์ในการประเมินลักษณะต่างๆ ของชนิดดินนั้นอย่างเข้มข้นหรือเลือกศึกษาอย่างละเอียด เช่น การวางแผนอยู่ในช่วงมาตราส่วน 1:5,000-1:10,000 แผนที่พื้นฐานใช้มาตราส่วนใหญ่ (1:2,000-1:10,000) ซึ่งค่อนข้างมีความละเอียดสูง (3) *การสำรวจดินแบบละเอียด* (Detaied Soil Surveys) เหมาะสำหรับใช้ในการประเมินลักษณะดินเพื่อใช้ในการวางแผนฟาร์มระดับไร่นาหรือวางแผนการชลประทาน ชนิดของหน่วยแผนที่ดินที่เหมาะสม ได้แก่ ประเภทดินของชุดดิน ชุดดินคล้าย ชุดดินสัมพันธ์ และชุดดินเชิงซ้อน แผนที่พื้นฐานมาตราส่วน 1:5,000-1:30,000 (4) *การสำรวจแบบค่อนข้างละเอียด* (Detailed Soil Surveys) เหมาะกับงานการศึกษาความเป็นไปได้ในระดับท้องถิ่นและการจัดการระดับที่กว้างขวางขึ้น แผนที่พื้นฐานมีมาตราส่วนประมาณ 1:20,000-1:50,000 ชนิดของหน่วยแผนที่ดินที่เหมาะสมจะใกล้เคียงกับการสำรวจดินละเอียด บริเวณที่เล็กที่สุดที่แสดงได้ในแผนที่ระดับนี้ คือ ประมาณ 6-36 เฮกตาร์ (5) *การสำรวจแบบค่อนข้างหยาบ* (Detailed Reconnaissance Soil Surveys) การสำรวจในลักษณะนี้จะให้ข้อมูลการแจกกระจายทางภูมิศาสตร์ของดินอย่างกว้างๆ หน่วยแผนที่ดินที่ใช้คือ วงศ์ดิน (Family) ชุดดิน ชุดดินคล้าย หน่วยดินสัมพันธ์ของชุดดิน และประเภทดินของหน่วยดินต่างๆ บริเวณที่เล็กที่สุดที่แสดงได้ในแผนที่ระดับนี้คือ ประมาณ 25-100 เฮกตาร์ เหมาะสำหรับใช้ประกอบการวางแผนระดับค่อนข้างกว้าง แผนที่พื้นฐานมาตราส่วน 1:40,000-1:100,000 (6) *การสำรวจแบบหยาบ* (Reconnaisance Soil Surveys) เป็นการสำรวจดินเพื่อจุดประสงค์ในการเก็บรวบรวมข้อมูลทั่วไปของดิน ทั้งการแจกกระจายเชิงภูมิศาสตร์ ชนิด และลักษณะของดิน หน่วยของแผนที่ดินที่ใช้มักเป็น กลุ่มดินใหญ่ กลุ่มดินหลัก หน่วยดินสัมพันธ์ของหน่วยดินเหล่านี้ ชุดดิน วงศ์ดิน หรือกลุ่มดินย่อย และหน่วยดินประเภทที่ดิน (land types) แผนที่พื้นฐานที่เหมาะสมควรมีมาตราส่วน 1:75,000-1:200,000 ใช้ในการวางแผนระดับภาค หรือประเทศ (7) *การสำรวจแบบหยาบมาก* (Exploratory Soil Surveys) ใช้ในการวางแผนระดับประเทศ และพิจารณาเพื่อหาพื้นที่ศึกษาในระดับที่ละเอียดขึ้น หน่วยแผนที่ดินที่ปรากฏจะเป็นหน่วยดินในขั้นการจำแนกระดับสูงใกล้เคียงกับการสำรวจแบบหยาบ แผนที่พื้นฐานมาตราส่วน 1:100,000-1:250,000 การกำหนดขอบเขตของหน่วยแผนที่ดินระดับนี้จะใช้สภาพภูมิประเทศ สภาพภูมิอากาศ พืชพรรณ และธรณีวิทยาที่มีอยู่แล้วนั้นมาประกอบการกำหนดขอบเขต และ (8) *การสำรวจแบบกว้าง* (Synthesis or

Schematic Soil Surveys) ใช้ในงานประเมินดินอย่างกว้างๆ ชนิดของหน่วยดินที่ใช้ก็เป็นหน่วยแผนที่ดินที่มีการจำแนกระดับสูง แผนที่พื้นฐานมาตราส่วนตั้งแต่ 1:100,000 ขึ้นไป การกำหนดขอบเขตแผนที่ดินใช้ข้อมูลที่มีอยู่แล้ว เช่น แผนที่ภูมิประเทศ ในการประกอบการแบ่งขอบเขต (เอิบ, 2542 ; ยงยุทธ และคณะ, 2543; Dent and Changprai, 1973)

2.2 ชุดดินและกลุ่มชุดดินของประเทศไทย

2.2.1 ข้อมูลชุดดิน (S_SERIES)

การสำรวจจำแนกและผลิตแผนที่ในระดับชุดดินและประเภทดินของชุดดิน รวมทั้งหน่วยดินสัมพันธ์ และหน่วยดินเบ็ดเตล็ดนั้น ได้มีการตีพิมพ์ผลการสำรวจดินทั่วประเทศ เป็นรายงานการสำรวจรายจังหวัด พร้อมกับแผนที่ดินในระดับชุดดิน มาตราส่วน 1:100,000 และต่อมาได้มีการผลิตแผนที่มาตราส่วน 1:1,000,000 (Vijarnsorn and Changpakdee, 1979) หลังจากนั้นกรมพัฒนาที่ดินได้ดำเนินการจัดทำแผนที่ชุดดินมาตราส่วน 1:50,000 และ 1:250,000 ให้ครอบคลุมทุกอำเภอทั่วประเทศ (ชาลี, 2541) ทั้งยังได้มีการจัดทำระบบข้อสนเทศทรัพยากรดิน (Soil Information System) ซึ่งรวบรวมและจัดเก็บคุณสมบัติต่างๆ ของชุดดินตัวแทน (typifying pedon) ซึ่งใช้เป็นบรรทัดฐานของหน่วยตัวแทนชุดดินประมาณ 244 ชุดดิน โดยระบบนี้เป็นระบบจัดการฐานข้อมูล (Database Management System, DBMS) ซึ่งอำนวยความสะดวกในการนำเข้า จัดเก็บ เรียกใช้ และรายงานผลข้อมูลชุดดิน (ทวีศักดิ์ และขนิษฐศรี, 2534) ปัจจุบันกรมพัฒนาที่ดินได้ทำการจัดจำแนกชุดดินไว้ทั้งหมดทั่วประเทศไทยประมาณ 300 ชุดดิน นอกจากนี้ได้ตีพิมพ์เอกสารวิชาการด้านกำหนดลักษณะและวินิจฉัยความเหมาะสมของชุดดินในภาคต่างๆ ของประเทศ ซึ่งใช้ระบบการจำแนกตามระบบอนุกรมวิธานดินปี 1975 ประกอบด้วยชุดดินมาตรฐานอย่างเป็นทางการทั้งหมด 237 ชุดดิน (เฉลียว และคณะ, 2531; วุฒิชาติ และคณะ, 2533; นิพันธ์ และบุญยงค์, 2536; ปราโมทย์ และคณะ, 2536) ในปัจจุบันกรมพัฒนาที่ดินใช้ข้อมูลชุดดินตามระบบอนุกรมวิธานดินปี 1998 ประกอบด้วยชุดดินมาตรฐานอย่างเป็นทางการทั้งหมด 240 ชุดดิน (นิพันธ์, 2542; สันต์ และบุรี, 2542; สถิระ, 2542; วุฒิชาติ, 2542)

อย่างไรก็ตาม กรมพัฒนาที่ดินยังไม่ได้มีการออกแบบฐานข้อมูลอรรถาธิบาย (attribute data) เหล่านี้ให้มีความเชื่อมโยงกับฐานข้อมูลเชิงพื้นที่ (spatial data) ของชุดดินแต่อย่างใด นอกจากนี้ยังไม่มีระบบการประมาณค่าคุณสมบัติดินของหน่วยแผนที่ (mapping unit) ดินประเภทชุดดินสัมพันธ์ และชุดดินที่ไม่สามารถระบุสัดส่วนได้ชัดเจน (undifferentiated soil) แต่ได้ใช้หน่วยชุดดินหลักเพียงชุดดินเดียวเป็นตัวแทน (เมธี และคณะ, 2543) จึงทำให้ข้อมูลอรรถาธิบายยังไม่สมบูรณ์

2.2.2 ข้อมูลกลุ่มชุดดิน (DLDGROUP)

นอกจากข้อมูลดินในระดับชุดดินแล้ว กรมพัฒนาที่ดินยังได้มีการจัดทำข้อมูลดินในระดับกลุ่มชุดดิน (soil group) โดยแบ่งชุดดินออกเป็น 62 กลุ่ม และแบ่งออกเป็น 2 ส่วนใหญ่ๆ คือ กลุ่มดินที่พบบนที่ราบต่ำ (low land soils) ได้แก่กลุ่มชุดดินที่ 1-25 และ 57-59 อีกส่วนหนึ่งคือกลุ่มดินที่พบบนที่ดอน (upland soils) ได้แก่กลุ่มชุดดินที่ 26-58 และ60-62 การจำแนกดังกล่าวอาศัยลักษณะและศักยภาพการใช้ประโยชน์ที่คล้ายกันเฉพาะชั้นดินบนความลึก 0-30 เซนติเมตร ร่วมกับลักษณะอื่นๆ เช่น การระบายน้ำ สภาพภูมิอากาศ ความลึกของดิน เนื้อดิน สีดิน ความเป็นกรดเป็นด่างของดิน และสภาพพื้นที่ดิน เป็นต้น โดยมีวัตถุประสงค์เพื่อความสะดวกต่อการแลกเปลี่ยนข้อมูลและการถ่ายทอดเทคโนโลยีในการเพิ่มประสิทธิภาพการผลิตของทรัพยากรดิน และสามารถกำหนดมาตรฐาน รวมทั้งวิธีการปรับปรุงและอนุรักษ์ทรัพยากรดินให้ถูกต้องเหมาะสมตามหลักวิชาการ (กรมพัฒนาที่ดิน, 2540; 2541; พิสุทธ์, 2541) นอกจากนี้กรมพัฒนาที่ดินยังได้จัดพิมพ์เป็นรายงานแผนที่ความเหมาะสมของดินกับพืชเศรษฐกิจเบื้องต้น มาตราส่วน 1:50,000 ในระดับจังหวัด และต่อมากรมพัฒนาที่ดินได้ร่วมกับคณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่ ได้จัดทำฐานข้อมูลดินระดับกลุ่มชุดดินในระบบดิจิตอลและพัฒนาระบบเรียกใช้ข้อมูลดินและความเหมาะสมของดินโดยการนำเอาแผนที่กลุ่มชุดดิน ข้อมูลความเหมาะสมของดินสำหรับพืชชนิดต่างๆ รวมทั้งข้อมูลอรรถาธิบาย มาผนวกเข้ากับโปรแกรม ArcView ซึ่งเป็นโปรแกรมด้านสารสนเทศทางภูมิศาสตร์ ทั้งนี้เพื่อให้สะดวกต่อการเรียกใช้และแก้ไขปรับปรุง (พนมศักดิ์ และเมธี, 2539) อย่างไรก็ตามข้อมูลอรรถาธิบายของฐานข้อมูลชุดดินที่ได้จัดทำขึ้นยังไม่สมบูรณ์ เนื่องจากข้อมูลในส่วนที่เป็นดินหน่วยชุดดินสัมพันธ์ หน่วยดินเชิงซ้อน หน่วยดินคล้ายและประเภทของชุดดิน ยังคงใช้หน่วยดินตัวแทนของชุดดินหลักเป็นตัวแทน ซึ่งไม่สะท้อนความเป็นจริงเท่าที่ควร อีกทั้งข้อมูลกลุ่มชุดดินที่มีอยู่ในปัจจุบันยังคงขาดในส่วนของคุณสมบัติของกลุ่มชุดดินอื่นๆ อีกหลายประการ รวมทั้งยังขาดความเชื่อมโยงกับข้อมูลชุดดินที่มีอยู่ จึงจำเป็นที่จะต้องมีการปรับปรุงเพิ่มเติมในส่วนนี้ต่อไป

2.2.3 ความสัมพันธ์ระหว่างข้อมูลชุดดิน (S_SERIES) และกลุ่มชุดดิน (DLDGROUP)

เนื่องจากข้อมูลกลุ่มชุดดินเกิดจากการสำรวจ รวบรวม และจัดกลุ่มชุดดินต่างๆ ที่มีลักษณะและศักยภาพในการใช้ประโยชน์ที่ดินที่คล้ายคลึงกัน รวมทั้งใช้การจำแนกในระดับวงศ์ตามระบบอนุกรมวิธานดินเพื่อจัดแบ่งชุดดินเข้าไว้เป็นกลุ่มชุดดินต่างๆ (กรมพัฒนาที่ดิน, 2540; 2541) ดังนั้นคุณสมบัติต่างๆ ทั้งทางกายภาพ เคมี ของข้อมูลกลุ่มชุดดินโดยส่วนใหญ่ย่อมมีความสัมพันธ์ไปในทิศทางเดียวกับคุณสมบัติของข้อมูลชุดดิน อาจกล่าวได้ว่าชุดดินเป็นสมาชิกของกลุ่มชุดดิน จึงน่า

จะมีความเป็นไปได้ที่จะนำเอาความเกี่ยวพันธ์ทางด้านข้อมูลอรรถาธิบายนี้มาใช้ในการปรับปรุง แก้ไข และเพิ่มเติมข้อมูลกลุ่มชุดดินในส่วนที่ขาดหายไปได้เป็นอย่างดี

2.3 ระบบฐานข้อมูล (Database system)

ข้อมูล (data) เป็นค่าเชิงปริมาณหรือคุณภาพ หรือเป็นเซ็ตของค่าที่ยังไม่ได้มีความหมายจนกว่าจะสัมพันธ์กับเรื่องที่ทำการวัดค่า ชุดข้อมูลที่สามารถสื่อความหมายได้นั้นถือว่าเป็นข้อสนเทศ (information) ซึ่งจะเป็นค่าข้อมูลที่มีโครงสร้างและสื่อความหมายให้คนทั่วไปเข้าใจได้ แต่เดิมการจัดเก็บข้อมูลมักกระทำในรูปของแฟ้มเอกสารต่างๆ เรียกว่าระบบแฟ้มข้อมูล (file system) ถึงแม้ว่าระบบการเก็บแบบนี้มีความปลอดภัยและความคล่องตัวสูง แต่ในขณะเดียวกันก็ทำให้เกิดปัญหามากมาย ทั้งนี้เพราะข้อมูลมักมีจำนวนมาก ทำให้เกิดความซ้ำซ้อน (data redundancy) และเมื่อมีการซ้ำซ้อนเกิดขึ้นก็จะส่งผลให้เกิดชุดข้อมูลที่ขัดแย้งกันเอง (data inconsistency) หรือเกิดการสูญหายไปของข้อมูล (data anomaly) ระหว่างการเปลี่ยนแปลง เพิ่ม หรือลบข้อมูล (กิตติ และจำลอง, 2542 ; ระพีพรรณ, 2529) ดังนั้นจึงมีการพัฒนารูปแบบการเก็บข้อมูลแบบใหม่ที่เรียกว่า "ฐานข้อมูล" (database) ที่มีการจัดเก็บข้อมูลโดยอาศัยความสัมพันธ์กันของข้อมูล และเมื่อนำเอาข้อมูลนี้มารวมกันอย่างมีระเบียบและสัมพันธ์กันจะเกิดเป็นระบบฐานข้อมูล "Database System"

ระบบฐานข้อมูลประกอบด้วยตัวข้อมูล (data) ฮาร์ดแวร์ (hardware) ซอฟต์แวร์ (software) ซึ่งใช้เป็นตัวกลางระหว่างผู้ใช้ข้อมูลและข้อมูลในระบบ โดยอาศัยระบบที่เรียกว่า Database Management System (DBMS) เป็นระบบคอมพิวเตอร์ที่พัฒนาขึ้นเพื่อแก้ปัญหาความเป็นอิสระต่อกัน (data dependence) ในระบบแฟ้มข้อมูล ดังนั้น DBMS จะทำงานอย่างเป็นอิสระจากฮาร์ดแวร์และตัวข้อมูลภายในฐานข้อมูล รูปแบบการอ้างถึงข้อมูลจะไม่ขึ้นอยู่กับโครงสร้างทางกายภาพของข้อมูล แต่ใช้ภาษา Query Language เพื่อติดต่อกับข้อมูล ทำให้ผู้ใช้สามารถเรียกใช้ข้อมูลจากฐานข้อมูลได้ง่ายขึ้นโดยไม่จำเป็นต้องทราบรายละเอียดของข้อมูล (กิตติ และจำลอง, 2542)

2.3.1 คำอธิบายฐานข้อมูล (Data dictionary หรือ Metadata)

คำอธิบายฐานข้อมูลเป็นส่วนที่เก็บรายละเอียดต่างๆ เกี่ยวกับข้อมูลภายในฐานข้อมูล เช่น โครงสร้างของข้อมูล โครงสร้างของตาราง โครงสร้างของดรรชนีต่างๆ หรือ กฎที่ใช้ในการควบคุมความถูกต้องของข้อมูลทั้งขนาด ชนิด และคำอธิบายต่างๆ ที่เรียกว่า "integrity rule" รวมทั้งกฎที่ใช้ในการรักษาความปลอดภัยให้กับฐานข้อมูล (security rule) ซึ่งข้อมูลเหล่านี้มีความจำเป็นและสำคัญต่อโปรแกรม DBMS ในการตัดสินใจเพื่อดำเนินการอย่างใดอย่างหนึ่งกับฐานข้อมูล

2.3.2 แบบจำลองฐานข้อมูล (Database model)

แบบจำลองฐานข้อมูลเป็นแบบจำลองที่ใช้ในการนำเสนอหรืออธิบายเกี่ยวกับโครงสร้างข้อมูลในฐานข้อมูล รวมทั้งรายละเอียดต่างๆ ในฐานข้อมูล มีหลายแบบ ได้แก่ Hierarchical Database Model, Network Database Model และ Relational Database Model โครงสร้างข้อมูลที่เป็นที่นิยมและใช้งานกันอย่างแพร่หลายในปัจจุบันคือ โครงสร้างแบบ Relational Database Model แบบจำลองชนิดนี้ได้ถูกพัฒนาขึ้นมาจากพื้นฐานความคิดเกี่ยวกับแบบจำลองทางคณิตศาสตร์ที่เรียกว่า ทฤษฎีเซ็ต(set theory) และตรรกะของการทำนาย (predicate logic) โดยนักคณิตศาสตร์ชื่อ E. F. Codd ในช่วงปีค.ศ. 1968-1969 โดยแบบจำลองแบบ Relational นี้จะเป็นฐานข้อมูลที่ประกอบด้วยรีเลชัน (relation) หรือตาราง (table) ซึ่งเป็นหน่วยที่ใช้ในการจัดเก็บข้อมูลในรูปของตาราง 2 มิติ ประกอบด้วยชุดของแถว (row) และสดมภ์ (column) ข้อมูล 1 รายการจะเทียบกับแถวหรือ เรคอร์ด (record) และ สดมภ์ จะเป็นคุณลักษณะต่างๆ ซึ่งเทียบได้กับ field ของเรคอร์ด ในระบบแฟ้มข้อมูล สำหรับในโครงสร้างข้อมูลแบบนี้จะเรียกแต่ละแถวของรีเลชัน ว่า "ทูเปิล" (tuple) และแต่ละสดมภ์เรียกว่า "แอททริบิว" (attribute) ข้อมูลแต่ละรีเลชันเป็นอิสระ แต่สามารถนำมาสร้างความสัมพันธ์ร่วมกันได้ เนื่องจากฐานข้อมูลสัมพันธ์นี้อยู่ในรูปแบบของเซ็ตที่ภายในประกอบด้วยสมาชิกที่มีค่าไม่ซ้ำกันเลย จึงต้องมีแอททริบิวอันใดอันหนึ่งที่ทำให้ทูเปิลในรีเลชันนั้นมีข้อมูลไม่ซ้ำกัน ลำดับของสมาชิกทั้งทูเปิลและแอททริบิวไม่มีผลต่อการอ้างถึงหรือเข้าถึงข้อมูล ทำให้แต่ละแอททริบิวของรีเลชันมีความหมายหนึ่งความหมายเท่านั้น ทำให้สามารถใช้งานได้โดยไม่ยุ่งยากซับซ้อนและไม่จำเป็นต้องมีขั้นตอนมากมายในการที่จะทำให้ได้ผลลัพธ์ออกมา ซึ่งถือว่าเป็นการพลิกทฤษฎีที่เกี่ยวข้องกับโครงสร้างฐานข้อมูลแบบเดิมที่ใช้กันอยู่ในเวลานั้น (Litwin, 2000; กิติ และจำลอง, 2542 ; Philip, 1999a)

2.3.3 การควบคุมความถูกต้องให้กับข้อมูล (Data integrity)

โดยทั่วไปข้อมูลภายในฐานข้อมูลจะมีเป็นจำนวนมาก เพราะประกอบขึ้นจากรีเลชันที่เกี่ยวข้องสัมพันธ์กันมากมาย ดังนั้นจึงมีโอกาสที่จะเกิดความผิดพลาดและไม่ถูกต้องตามความเป็นจริง จึงต้องตั้งกฎเกณฑ์และข้อจำกัดต่างๆ เพื่อใช้ควบคุมและตรวจสอบความถูกต้องของข้อมูลภายในกฎเกณฑ์นี้เรียกว่า "Integrity Rule" แบ่งออกเป็น 2 แบบคือ (ก) Database Specific Integrity Table : เป็นเกณฑ์ที่กำหนดขึ้นเพื่อใช้ควบคุมความถูกต้องให้กับฐานข้อมูลใดฐานข้อมูลหนึ่งโดยเฉพาะ เช่น การกำหนดความกว้าง ชนิด ทศนิยมของแอททริบิว หรือการกำหนดช่วง (range) ค่าของข้อมูลในแต่ละแอททริบิว โดยเกณฑ์ประเภทนี้จะไม่คำนึงถึงว่าฐานข้อมูลนั้นมีโครงสร้างข้อมูลแบบใดและจะมีความแตกต่างกันในแต่ละฐานข้อมูล และ (ข) General Integrity Rule : เป็นเกณฑ์และข้อ

จำกัดที่ควบคุมและรักษาความสัมพันธ์ระหว่างข้อมูลภายในฐานข้อมูลให้มีความถูกต้องและครบถ้วน สามารถนำไปใช้ได้กับทุกฐานใน Relation Model

2. 3.3.1 คีย์ (Key)

คีย์เป็นทูเปิลที่มีเอกลักษณ์เฉพาะตัวหรือค่าชุดของแอททริบิวที่ทำให้แต่ละทูเปิลของรีเลชันไม่ซ้ำกัน (unique) คีย์ที่ใช้ใน Relation Model ได้แก่ (1) Candidate Key เป็นคีย์ที่ทำให้ข้อมูลในแต่ละทูเปิลของรีเลชันมีค่าไม่ซ้ำกันเลย เป็นคีย์ที่มีขนาดเล็กสุด (ขั้นย่อยที่สุด) ต้องไม่สามารถแยกเป็นคีย์ย่อยได้อีก รีเลชันหนึ่งๆ ยังสามารถมี Candidate Key ได้มากกว่า 1 ตัว Candidate Key สามารถกำหนดขึ้นมาจากแอททริบิวเดียวหรือมากกว่าได้ โดยยึดว่าต้องทำให้แต่ละทูเปิลในรีเลชันนั้นไม่ซ้ำกัน การกำหนด Candidate Key ไม่เกี่ยวกับการกำหนด Index (2) Primary Key และ Alternate Key ตัว Candidate Key ที่เหมาะสมที่สุดจะกลายเป็น Primary Key ส่วน Candidate Key ที่เหลือจะกลายเป็น Alternate Key ดังนั้น Primary Key จึงเป็นคีย์ที่สำคัญในแต่ละรีเลชัน เพราะ Primary Key จะถูกใช้เป็นคีย์ในการตรวจสอบการซ้ำกันของข้อมูล ระหว่างที่ทำการป้อนข้อมูลหรือกำหนดข้อมูลใหม่ให้กับรีเลชัน (Litwin , 2000) การที่จะเลือก Candidate ใดบ้างมาเป็น Primary Key นั้นไม่มีกฎเกณฑ์ตายตัวที่แน่นอน แต่ขึ้นอยู่กับความเหมาะสมและความสะดวกของผู้ออกแบบ การตัดสินใจเลือก Primary Key นั้น จะขึ้นอยู่กับหลักเกณฑ์พื้นฐานต่างๆ เช่น เลือกแอททริบิวน้อยที่สุดและเลือกเท่าที่จำเป็นเท่านั้น (mimimality) ความคงทนเสถียร โดยเลือกแอททริบิวที่ไม่มีการเปลี่ยนแปลงบ่อยนัก (Stability) และความไม่ยุ่งยากหรือง่าย (Simplicity) เพื่อให้ผู้ใช้มีความคุ้นเคยและเข้าใจได้ง่าย (Litwin, 2000; Mahesh,1999) Primary Key จึงเป็น Candidate Key ที่ถูกเลือกให้เป็นคีย์ที่มีความเป็นเอกลักษณ์ในทุกๆ แอททริบิวในทุกๆ ทูเปิล และต้องเป็นคีย์ที่ไม่มีค่าไม่มีข้อมูล (null) (3) Foreign Key ได้แก่แอททริบิวใดแอททริบิวหนึ่งในรีเลชันที่ใช้อ้างอิงไปยังแอททริบิวที่เป็น Candidate Key ของอีกรีเลชันหนึ่งที่มีความสัมพันธ์กับ Foreign Key อาจเกิดจากแอททริบิวมากกว่า 1 แอททริบิว เช่นเดียวกับ Primary Key (Foreign Key มักเขียนอยู่ในเครื่องหมาย {}) ค่าของ Foreign key ในรีเลชันหนึ่งๆ จะต้องปรากฏอยู่ใน Candidate Key ของรีเลชันที่สัมพันธ์กัน แต่ทุกค่าของ Candidate Key ไม่จำเป็นต้องปรากฏใน Foreign Key จำนวนแอททริบิวที่จะถูกนำมาประกอบกันเป็น Foreign Key ขึ้นอยู่กับจำนวนแอททริบิวที่ได้กำหนดมาเป็น Candidate Key หรือ Primary Key นอกจากนี้ Domain ของแอททริบิวของ Foreign Key ต้องเท่ากับ Domain ของแอททริบิวของ Candidate Key ของรีเลชันที่สัมพันธ์กันกับ Foreign Key และไม่จำเป็นต้องเป็น Candidate Key ของรีเลชัน แต่สามารถทำหน้าที่เป็น Foreign Key ได้ รีเลชันที่เป็นเจ้าของ

Foreign Key จะถูกเรียกว่า "Referenced Relation " หรือ ""Target Relation" รีเลชันใดๆ สามารถมีความสัมพันธ์ด้วย Foreign Key กับตัวมันเองได้เรียกว่า "Self -Referencing"

2. 3. 4 การดำเนินการกับข้อมูล (Data manipulation)

การกระทำที่เกิดขึ้นในฐานข้อมูลประกอบด้วยการกระทำพื้นฐานใน 2 ลักษณะคือ แบบ relational algebra เป็นการกระทำพื้นฐานที่ประกอบกันขึ้นเป็นขั้นตอน (procedure) เพื่อใช้ในการกำหนดให้ฐานข้อมูลทำงานตามที่ต้องการ บางครั้งเรียกว่า procedure query language เกิดจากทฤษฎีเซ็ตทางคณิตศาสตร์ โดยรีเลชันที่นำมากระทำต่อกันต้องมีแอททริบิวเหมือนกันหรือเท่ากันมากระทำต่อกัน กลุ่มการกระทำในลักษณะนี้ได้แก่ คำสั่ง Product, Union, Intersection และ Difference เป็นต้น อีกกลุ่มหนึ่งเป็นการกระทำพิเศษที่จำเป็นในการนำรีเลชันมากระทำต่อกันโดยแอททริบิวไม่จำเป็นที่จะต้องเท่ากัน เพียงแต่ต้องมีชื่อเหมือนกันและมีโดเมนของข้อมูลเดียวกัน

การกระทำที่เป็นกลุ่มคำสั่งที่ใช้ในการเรียกใช้ข้อมูลที่ต้องการจากรีเลชันโดยไม่สนใจวิธีการที่จะได้มาซึ่งผลลัพธ์นั้น บางครั้งเรียกการดำเนินการแบบนี้ว่า non-procedure query language

2.4 ความสัมพันธ์ (Relationships)

ความสัมพันธ์ระหว่างข้อมูลต่างๆ ที่จัดเก็บอยู่ในฐานข้อมูล จัดเป็นส่วนที่สำคัญที่มีผลต่อโครงสร้างของฐานข้อมูลในแต่ละแบบ ทั้งนี้เนื่องจากสิ่งต่างๆ ที่ระบุได้ในความเป็นจริง อาจเป็นรูปธรรม (entity) หรือนามธรรม (abstract) โดยทั่วไปจะมีข้อมูลและคุณสมบัติจำนวนมากมายซับซ้อนที่สัมพันธ์กันอยู่ ดังนั้นไม่ว่าจะเป็นฐานข้อมูลแบบใดเมื่อนำมาพิจารณาจะสามารถพิจารณาได้ครั้งละ 2 รีเลชัน เท่านั้น ความสัมพันธ์นั้นแบ่งออกเป็น 3 ประเภท คือ (1) แบบ One-to-One เป็นความสัมพันธ์ที่แต่ละรายการ หรือแต่ละแถวของรีเลชันหนึ่งจะมีความสัมพันธ์กับอีกรีเลชันหนึ่งเพียงรายการเดียวเท่านั้น เป็นความสัมพันธ์ที่เกิดขึ้นบ่อยครั้งในการทำข้อมูลและจำเป็นมากเพื่อจะจัดการกับข้อมูลโดยการแบ่งรีเลชันใหญ่ๆ ให้เล็กลง เพื่อเพิ่มประสิทธิภาพแสดงข้อมูลรวมทั้งเรื่องความปลอดภัยของข้อมูล (2) แบบ One-to-Many เป็นความสัมพันธ์ของแต่ละรายการของรีเลชันหนึ่งมีความสัมพันธ์กับอีกรีเลชันหนึ่งได้มากกว่า 1 รายการ เช่น ลูกค้าธนาคาร 1 คนสามารถมีเงินฝากในบัญชีได้หลายบัญชี หรือร้านส่งพิซซ่าสามารถมีลูกค้าได้มากมาย ความสัมพันธ์แบบนี้จะพบบ่อยใน แบบจำลองแบบรีเลชันทั่วไป (3) แบบ Many-to-Many เป็นความสัมพันธ์มากกว่า 1 รายการ กล่าวคือรายการในรีเลชันหนึ่งมีความสัมพันธ์กับอีกหลายรายการในอีกรีเลชันหนึ่ง เช่น คนไข้แต่ละคนสามารถทำประกันกับหลายบริษัทและหลายบริษัทประกันเองก็มีลูกค้าได้หลายราย (Litwin, 2000; กิติและจำลอง, 2542)

2.5 ระบบสารสนเทศทางภูมิศาสตร์และการพัฒนาฐานข้อมูลดิน

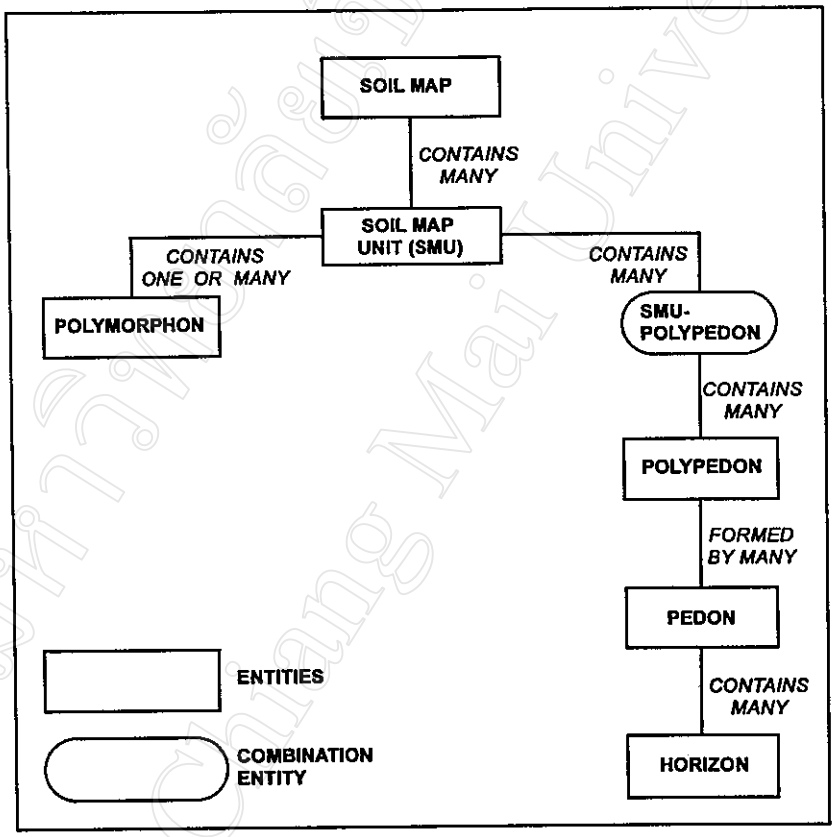
ระบบสารสนเทศทางภูมิศาสตร์ (Geographic Information System, GIS) เป็นระบบของคอมพิวเตอร์ที่ประกอบด้วย ฮาร์ดแวร์ ซอฟต์แวร์ ข้อมูลทางภูมิศาสตร์ บุคคล และวิธีการที่ทำให้เกิดการจัดเก็บ ปรับปรุง ตรวจสอบ ประมวลผล วิเคราะห์ พร้อมทั้งแสดงผลทั้งหมดในรูปของข้อมูลและสารสนเทศของสิ่งที่สามารถระบุตำแหน่งบนผิวโลก (geographically referenced information) ข้อมูลเหล่านี้อาจได้จากการสำรวจระยะไกลและการสำรวจรูปแบบอื่นๆ (Burrough, 1986; สงำ, 2533; David et al., 1991; Yue Hong Chou, 1997; Philip, 1999b)

ระบบสารสนเทศทางภูมิศาสตร์ในปัจจุบันประกอบด้วยวิธีการจัดการข้อมูลทั้งการเก็บข้อมูล การบันทึก การปรับปรุงแก้ไข และการวิเคราะห์ข้อมูลเชิงพื้นที่ และการแสดงผลในรูปกราฟิก ทำให้สามารถมองเห็นได้และง่ายต่อการเข้าใจ ระบบนี้ได้ถูกนำมาใช้ในงานหลายด้าน ทั้งทางวิศวกรรม อุตสาหกรรม การบริหาร และทางด้านการเกษตร สำหรับงานด้านข้อมูลดินนั้นได้มีการนำเอาระบบสารสนเทศทางภูมิศาสตร์เข้ามาใช้ร่วมกับระบบฐานข้อมูลเพื่อพัฒนาฐานข้อมูลดินกันอย่างกว้างขวาง หน่วยงานในหลายประเทศได้มีการพัฒนาฐานข้อมูลดินให้อยู่ในระบบสารสนเทศทางภูมิศาสตร์ที่ใช้กันแพร่หลายในปัจจุบัน และได้มีการการออกแบบข้อมูลเชิงอรรถาธิบายให้มีความเชื่อมโยงกับข้อมูลเชิงพื้นที่

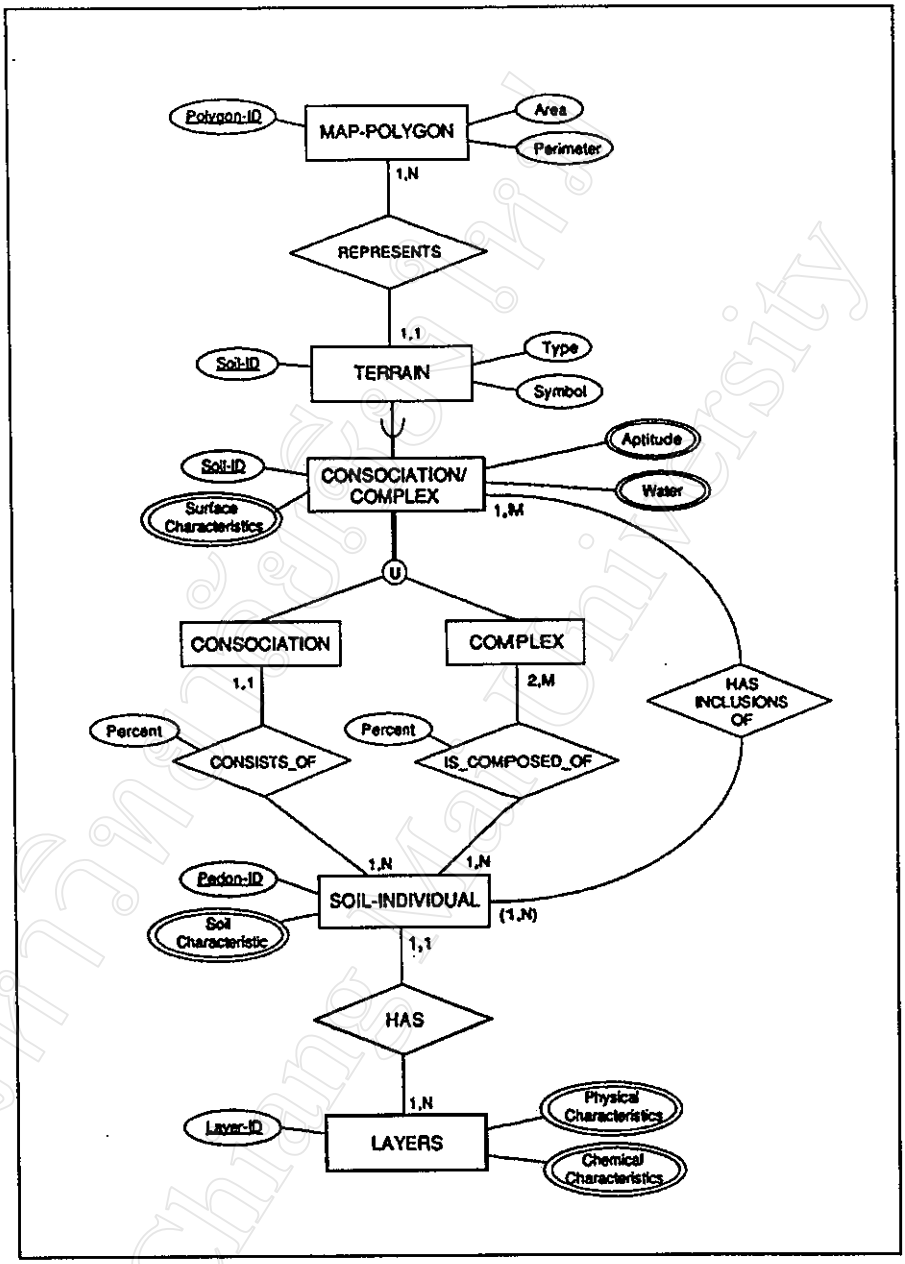
Zinck and Valenzuela (1990) ได้พัฒนาระบบข้อมูลดินและระบบองค์ประกอบข้อมูลเชิงภูมิศาสตร์ซึ่งมีการเก็บข้อมูลการแบ่งขอบเขตหน่วยแผนที่ดินไว้ในรูปข้อมูลเชิงพื้นที่ และจัดเก็บสมบัติต่างๆ ของดินไว้ในระบบฐานข้อมูลดิน การออกแบบระบบฐานข้อมูลดินนั้นในขั้นแรกเป็นการออกแบบเชิงตรรกะ (logical design) ดังแสดงในรูปที่ 1 ซึ่งจะเชื่อมโยงส่วนต่างๆ เข้าด้วยกันโดยใช้คีย์ที่เรียกว่า "Soil Map Unit" (SMU) และหมายเลขหน่วยแผนที่ดินของแต่ละพีดอน (pedon) ตามหลักการของตารางสัมพันธ์ ข้อมูลที่ได้สามารถแปลงไปใช้ร่วมกับเครื่องมือหรือโปรแกรมที่ใช้ในการวิเคราะห์ วางแผนการใช้ประโยชน์ที่ดิน หรืองานด้านอื่นๆ และสามารถเชื่อมโยงกับข้อมูลเชิงพื้นที่ในการแสดงผลการวิเคราะห์นั้นในรูปแผนที่

Fernandez and Rusinkiewiez (1993) ได้ออกแบบและพัฒนาฐานข้อมูลดินที่สามารถใช้ร่วมกับระบบสารสนเทศทางภูมิศาสตร์ โดยแบ่งขั้นตอนการออกแบบเป็น 2 กระบวนการใหญ่ๆ คือ การออกแบบเชิงตรรกะและการออกแบบเชิงกายภาพ ในฐานข้อมูลนี้ได้ใช้รูปแบบของ Extended Entity-Relationship (EER) ซึ่งเป็นรูปแบบที่สามารถประยุกต์ใช้กับข้อมูลที่มีความซับซ้อนได้ (รูปที่ 2)โดยฐานข้อมูลเชิงอรรถาธิบายนี้จะประกอบไปด้วยแฟ้มข้อมูลต่างๆ ที่เกี่ยวกับชุดดิน 5 แฟ้ม และเป็นแฟ้มข้อมูลที่เกี่ยวกับหน่วยแผนที่ดิน (soil mapping unit, SMU) อีก 6 แฟ้มข้อ

มูล โดยได้สร้างให้สัมพันธ์กับตารางประกอบชั้นข้อมูลเชิงพื้นที่ของรูปหลายเหลี่ยมของแผนที่ชุดดิน (polygon attribute table, PAT) ในโปรแกรม ARC/INFO (ESRI, 1994a) นอกจากนี้ได้ใช้หมายเลขกำกับหน่วยแผนที่ดินเป็นตัวเชื่อมโยง (primary key) ที่สำคัญ ทั้งนี้เพื่อให้ง่ายต่อการเรียกใช้แก้ไขปรับปรุง และลดความซ้ำซ้อนของข้อมูล ฐานข้อมูลชุดดินนี้ได้แยกเก็บข้อมูลออกเป็นตารางสัมพันธ์ต่างๆ เช่น ตารางชนิดภูมิประเทศ (terrain) ตารางจัดเก็บหน่วยดินสัมพันธ์ และหน่วยดินผสม ตารางแสดงสัดส่วนของหน่วยดินทั้งสองชนิดดังกล่าว ตารางแสดงคุณสมบัติของหน่วยดินตัวแทน รวมทั้งคุณสมบัติทางเคมีและฟิสิกส์ของดินในแต่ละชั้นดิน (horizon) ด้วย รูปที่ 3 แสดงแนวคิดในการเชื่อมโยงตารางต่างๆ ในฐานข้อมูลอรรถาธิบายและฐานข้อมูลเชิงพื้นที่เข้าด้วยกัน

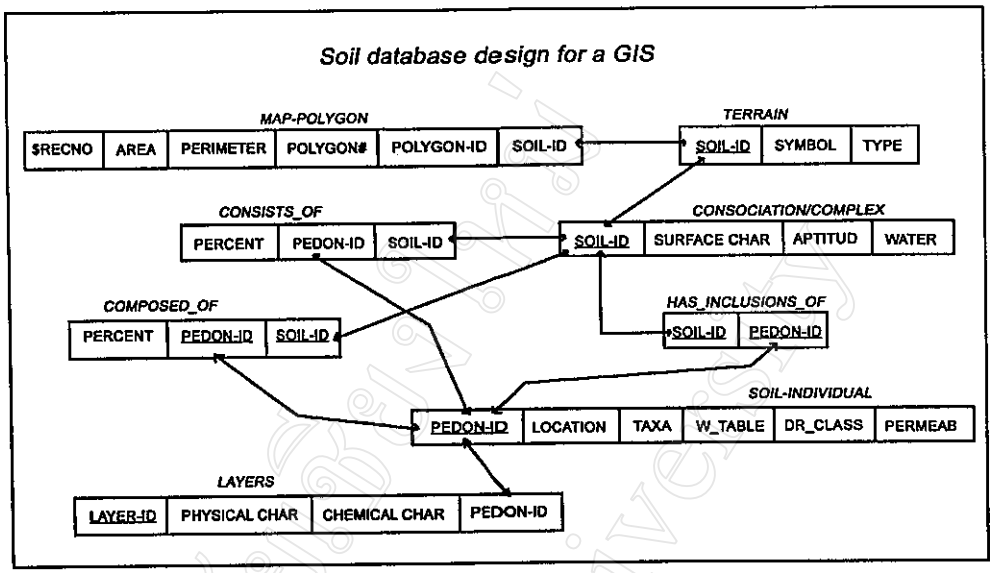


รูปที่ 1 การออกแบบฐานข้อมูลเชิงตรรกะ (Zinck and Valenzuela, 1990)



รูปที่ 2 โครงสร้างฐานข้อมูลดินในรูปแบบ Extended Entity-Relationship, ERR

(Fernandez and Rusinkiewiez, 1993)



รูปที่ 3 การเชื่อมโยงตารางข้อมูลดินและตารางฐานข้อมูลเชิงพื้นที่ตามแนวคิดของ Fernandez and Rusinkiewiez (1993)

King et al. (1994) ได้พัฒนาฐานข้อมูลดินจากแผนที่ดินของกลุ่มประเทศในยุโรป โดยสร้างจากแผนที่มาตราส่วน 1:1,000,000 และข้อมูลอื่นๆจากการสำรวจดิน ซึ่งจัดเก็บ และสามารถจัดการบนระบบสารสนเทศทางภูมิศาสตร์และสร้างแบบจำลองเชิงพื้นที่ (Spatial Organization Model, SOM) ซึ่งสามารถจัดการและวิเคราะห์ข้อมูลอรรถาธิบายของดิน ทั้งนี้เพื่อให้ฐานข้อมูลดินนี้สามารถที่จะบอกรายละเอียดของข้อมูลได้ใกล้เคียงกับความเป็นจริงตามธรรมชาติได้ดีขึ้น โดยสามารถอธิบายรายละเอียดได้ในแต่ละหน่วยแผนที่ดิน (SMUs) ที่สร้างขึ้นโดยวิธีการซ้อนทับแผนที่ขอบเขตการปกครอง แผนที่หน่วยดินสัมพันธ์ และแผนที่หน่วยดินย่อยเข้าด้วยกัน

องค์การอาหารและเกษตรของสหประชาชาติ (Food and Agriculture Organization of the United Nations, FAO) ร่วมกับศูนย์ข้อมูลดินสากล (International Soil Reference and Information Centre, ISRIC) และสถาบันทรัพยากรธรรมชาติและชีวเกษตร (Institute of Natural Resources and Agro-biology, CSIC) เพื่อพัฒนา ระบบสนับสนุนการจัดการ จัดเก็บ และเรียกใช้ข้อมูลดินพื้นฐานบนคอมพิวเตอร์ส่วนบุคคล โดยปรับปรุงมาจากข้อมูลดินเดิมของ FAO/ISRIC (SDB) ให้สามารถใช้งานได้ 3 ภาษา คือ ภาษาอังกฤษ ภาษาฝรั่งเศส และภาษาสเปน รวมทั้งได้เพิ่มส่วนของการวิเคราะห์ทางสถิติเพื่อช่วยในการปรับปรุงฐานข้อมูลดิน ระบบนี้พัฒนาขึ้นจากภาษา CLIPPER และภาษา C ซึ่งออกแบบให้สามารถตอบโต้กับผู้ใช้ได้ง่าย สามารถนำไปเชื่อมโยงกับระบบสารสนเทศทางภูมิศาสตร์เพื่อแสดงรายละเอียดของข้อมูลดินร่วมกับแผนที่ได้ และสำเนาแฟ้มข้อมูลในรูปแบบ

ที่สามารถนำไปใช้ได้กับซอฟต์แวร์อื่นๆ สามารถส่งออกข้อมูลไปยังเครื่องพิมพ์ในหลายรูปแบบ เช่น รูปกราฟ คำอธิบายโปรไฟล์ดิน (FAO, 1995)

ฐานข้อมูลดินของสำนักอนุรักษ์ทรัพยากรธรรมชาติ (Natural Resources Conservation Service, NRCS) แห่งประเทศสหรัฐอเมริกาประกอบด้วยข้อมูล 3 ระดับคือ (1) Soil Survey Geographic (SSURGO) database เป็นฐานข้อมูลที่สร้างขึ้นเพื่อใช้ประโยชน์ในระดับฟาร์ม หมู่บ้าน และเมือง พัฒนามาจากแผนที่มาตราส่วน 1:12,000-1:630,000 จัดเก็บเป็นข้อมูลประเภทเวกเตอร์ (vector) ส่วนข้อมูลเชิงอรรถาธิบายจะเชื่อมโยงกับข้อมูล Map Unit Interpretation Record (MUIR) ซึ่งจะเป็นกลุ่มของตารางสัมพันธ์ที่แสดงถึงคุณสมบัติของดินและองค์ประกอบของดิน รวมทั้งปัจจัยสภาพแวดล้อมที่ทำให้เกิดดินนั้นด้วย (2) State Soil Geographic (STATSGO) database สร้างขึ้นจากแผนที่มาตราส่วน 1:250,000 ซึ่งใช้ในการจัดการระดับที่ใหญ่กว่าแบบแรก เช่นระดับรัฐและภาค ข้อมูลเชิงอรรถาธิบายของข้อมูลดินระดับนี้จะเชื่อมโยงกับฐานข้อมูล Soil Interpretation Record (SIR) โดยมีรายละเอียดคล้ายคลึงกับข้อมูลอรรถาธิบายของงานข้อมูลระดับแรก (3) National Soil Geographic (NATSGO) database เป็นข้อมูลเชิงพื้นที่ที่สร้างขึ้นจากแผนที่มาตราส่วน 1:5,000,000 ใช้ในการจัดการระดับประเทศและภาค สำหรับข้อมูลอรรถาธิบายในระดับนี้จะเชื่อมโยงกับกลุ่มตารางสัมพันธ์ Map Unit Interpretation Record (MUIR) ทั้งนี้ข้อมูลอรรถาธิบายของทั้ง 3 ระดับ จัดเก็บอยู่ในรูปแบบของตารางสัมพันธ์ที่สามารถเชื่อมโยงกับข้อมูลเชิงพื้นที่ในระบบสารสนเทศทางภูมิศาสตร์ และผู้ใช้สามารถเรียกใช้ ค้นหาได้ถึงระดับชั้นดิน (soil horizon) อีกทั้งยังสามารถช่วยในการจำแนกดินตามระบบอนุกรมวิธานดินได้ (USDA, 1996)

สำหรับฐานข้อมูลดินเชิงพื้นที่ของประเทศแคนาดานั้น แบ่งเป็นหลายระดับตามรายละเอียดของการใช้งาน คือ (1) แผนที่นิเวศวิทยาระดับชาติ (National Ecological Framework, EcoZone, EcoRegions and EcoDistricts) สร้างจากแผนที่มาตราส่วน 1:36,000,000 ถึง1:1,000,000 เป็นข้อมูลที่เกี่ยวกับกลุ่มของดิน จัดทำขึ้นเพื่อใช้วางแผนในระดับรัฐหรือจังหวัด (2) แผนที่ดินและศักยภาพการผลิตของที่ดิน (Soil Map of Canada/Land Potential Database, LPDB) เป็นข้อมูลที่ใช้เพื่อการวางแผนระดับประเทศ ในแผนที่มาตราส่วน 1:5,000,000 สามารถแสดงพื้นที่ดินทั้งหมดของประเทศ (3) แผนที่เขตนิเวศเกษตร (Agroecological Resource Areas, ARAs) เป็นข้อมูลเชิงพื้นที่ที่แสดงสภาพภูมิอากาศ เขตภูมิอากาศ สภาพเศรษฐกิจ การเพาะปลูก ดิน และสภาพภูมิทัศน์ สร้างจากแผนที่มาตราส่วน 1:2,000,000 (4) แผนที่ดินระดับภูมิทัศน์ (Soil Landscape of Canada, SLC) จัดเก็บดินหลัก (major soil) และแสดงลักษณะภูมิประเทศ สร้างจากแผนที่มาตราส่วน 1:1,000,000 (5) แผนที่เพื่อการประเมินพื้นที่ดิน (Canada Land Inventory, CLI) สร้างจากแผนที่ต้นฉบับมาตรา

ส่วน 1:250,000 เป็นแผนที่ที่แสดงพื้นที่การเกษตรทั้งหมดในประเทศ รวมทั้งข้อมูลสมรรถนะของการทำการเกษตร ป่าไม้ พื้นที่สงวนเพื่อสัตว์ป่า พื้นที่เพื่อการนันทนาการ (6) แผนที่สำรวจดินขั้นละเอียด (Detailed Soil Survey) เป็นแผนที่ครอบคลุมพื้นที่เกษตรสำคัญๆ ของประเทศ ใช้สำหรับการวางแผนในระดับจังหวัดหรืออำเภอ มีมาตราส่วน 1:20,000-1:250,000 (7) ข้อมูลเชิงพื้นที่อื่นๆ เป็นข้อมูลเชิงพื้นที่เฉพาะแห่งและสร้างจากมาตราส่วนเล็ก ใช้สำหรับการวางแผนจัดการในระดับพื้นที่ขนาดเล็กและเฉพาะเจาะจง ในส่วนของข้อมูลเชิงอรรถาธิบายได้จัดเก็บอยู่ในตารางสัมพันธ์ที่สามารถเชื่อมโยงกับหมายเลขกำกับหน่วยแผนที่ดินภายในข้อมูลเชิงพื้นที่ดิน ซึ่งสามารถเชื่อมโยงกับตารางอื่นๆ เช่น ตารางชื่อและรายละเอียดของดิน ตารางแสดงชั้นข้อมูลดินและตารางแสดงคุณสมบัติอื่นๆ รวมทั้งสามารถเรียกใช้ เปลี่ยนแปลงแก้ไข แสดงผลทางกราฟิก และสามารถเชื่อมโยงกับระบบสารสนเทศทางภูมิศาสตร์อื่นๆ ได้เป็นอย่างดี (Schut , 2001)

ระบบฐานข้อมูลดิน SOTER (van Engelen, 1999; WRI, 2001) เป็นฐานข้อมูลดินอีกระบบหนึ่งที่มีการพัฒนาและใช้กันอย่างแพร่หลาย SOTER เป็นระบบจัดการฐานข้อมูลเชิงสัมพันธ์ของข้อมูลทรัพยากรธรรมชาติ ระบบนี้ทำให้ผู้ใช้สามารถเข้าถึงและวิเคราะห์ข้อมูลจากความสัมพันธ์ของความต้องการใช้ทรัพยากรกับผลกระทบต่อสิ่งแวดล้อมและการอนุรักษ์ โดยระบบนี้จะจัดการข้อมูลทรัพยากรธรรมชาติที่สร้างขึ้นในรูปดิจิตอลนี้อย่างเป็นระเบียบ ประกอบด้วยชั้นข้อมูลต่างๆ ดังต่อไปนี้คือ ภูมิประเทศของบริเวณที่เกิดดิน สภาพภูมิอากาศ พืชพรรณและการใช้ประโยชน์ที่ดิน ซึ่งสามารถเชื่อมโยงสู่ระบบสารสนเทศทางภูมิศาสตร์ให้สามารถแสดงข้อมูลอรรถาธิบายที่ประกอบกันขึ้นเป็นเรื่องราวไม่ว่าจะเป็นการแสดงข้อมูลแยกเป็นชั้นดิน การซ้อนทับชั้นข้อมูล หรือการแสดงผลในรูปตาราง การพัฒนาฐานข้อมูลทรัพยากรธรรมชาติได้เริ่มพัฒนามาตั้งแต่ปี ค.ศ. 1986 โดยสหพันธ์ปฐพีศาสตร์สากล (International Union of Soil Science, IUSS) ร่วมกับศูนย์ทรัพยากรที่ดินแห่งประเทศแคนนาดา และ FAO ได้เสนอระบบนี้ในการประชุมปฐพีศาสตร์โลกครั้งที่ 13 ปี ค. ศ. 1986 (World Congress Soil Science) ภายใต้โครงการ UNEP ISRIC ได้ศึกษาและพัฒนาวิธีการสร้างข้อมูลดินโลกและฐานข้อมูลผิวดินในระบบดิจิตอล (World SOil and TERrain Digital Database, SOTER) มาตราส่วนของแผนที่ที่ใช้ 1:1,000,000 และได้ถูกทดลองใช้ใน 3 พื้นที่ครอบคลุมพื้นที่ของ 5 ประเทศ (อาร์เจนตินา บราซิล อุรุกวัย สหรัฐอเมริกา และแคนาดา) มีการรายงานผลการทดสอบในการประชุมปฐพีศาสตร์โลกครั้งที่ 14 ใน 1990 และต่อมาในปี ค. ศ. 1992 FAO, UNEP, IUSS และ ISRIC ได้ให้การรับรองและได้ทำการตีพิมพ์คู่มือไปทั่วโลก ในปี ค. ศ. 1993 นอกจากนั้นยังได้มีการนำเอาวิธีการประเมินความเสี่ยงของการเกิดการชะล้างพังทลาย และระบบประเมินคุณภาพที่ดินอัตโนมัติ (Automated Land Evaluation System, ALES) เข้ามาประกอบกับระบบทำให้สามารถใช้งานได้กว้างขวางขึ้นและยังคงมีการพัฒนาต่อไป เนื่องจากยัง

คงมีความต้องการข้อมูลทรัพยากรดินเพื่อการวางแผนการใช้ประโยชน์ทรัพยากรดินในภูมิภาคต่างๆ ของโลก

ข้อมูลดินและแผนที่ดินมีความสำคัญต่อการพัฒนางานในหลายด้าน ดังจะเห็นได้จาก มีการใช้ GIS เข้ามาช่วยในการจัดเก็บแผนที่ข้อมูลดินและนำวิธีการออกแบบฐานข้อมูลที่มีความสามารถในการเชื่อมโยงเข้ากับฐานข้อมูลเชิงพื้นที่มาใช้กันอย่างแพร่หลาย สำหรับประเทศไทยถึงแม้ได้มีการสำรวจและจัดเก็บข้อมูลเป็นจำนวนมาก แต่ข้อมูลยังอยู่ในรูปที่จัดการได้ยากและไม่สะดวกต่อการเรียกใช้ นอกจากนี้การมีฐานข้อมูลดินหลายระดับเพื่อการใช้ประโยชน์ที่แตกต่างกันทำให้จำเป็นต้องมีวิธีการเชื่อมโยงฐานข้อมูลดินในหลายระดับนั้นเข้าด้วยกัน เพื่อเพิ่มประสิทธิภาพในการจัดเก็บและเรียกใช้ รวมทั้งสามารถนำไปใช้ในการสนับสนุนงานได้หลากหลายลักษณะและใกล้เคียงกับความเป็นจริงมากขึ้น

2.6 การประมาณค่าเชิงพื้นที่ (Spatial interpolation)

การประมาณค่าเชิงพื้นที่เป็นขั้นตอนการประมาณค่าคุณสมบัติของข้อมูลในตำแหน่งที่ไม่ได้มีการวัดค่าข้อมูล จากค่าข้อมูลที่มีการวัดค่าได้ในตำแหน่ง (point) ที่ทราบค่าพิกัดแน่ชัด การประมาณค่าใช้หลักการที่ว่าจุดสำรวจที่อยู่ใกล้กันบนพื้นผิวโลกย่อมมีลักษณะ และคุณสมบัติคล้ายคลึงกันมากกว่าจุดที่อยู่ไกลออกไป ตามกฎของ Tobler's law of geography (Brian, 1997) การประเมินค่าเชิงพื้นที่แบ่งได้เป็น 2 ลักษณะคือ การประมาณค่าของข้อมูลที่มีตำแหน่งเฉพาะจุด (interpolation of point) และการประมาณค่าแบบพื้นที่ (interpolation of areal data)

2.6.1 การประมาณจากค่าเฉพาะจุด

(ก) วิธีการถ่วงน้ำหนัก (Weighting)

วิธีการนี้จะมีการกำหนดค่าน้ำหนัก (weight) ให้กับจุดต่างๆ ที่มีผลกับจุดที่ต้องการประมาณค่า เช่น วิธีของ distance-weighting ซึ่งเป็นวิธีการที่ใช้ระยะทางระหว่างจุดข้อมูลข้างเคียงเป็นค่าน้ำหนัก โดยให้ความสำคัญกับจุดที่อยู่ใกล้ที่สุดว่าเป็นจุดที่มีค่าข้อมูลใกล้เคียงกับจุดที่ต้องการประมาณค่ามากที่สุด (Lam, 1983) สามารถเขียนในรูปสมการได้ดังนี้

$$f(x,y) = \frac{\sum_{i=1}^{n} w(d_i) z_i}{\sum_{i=1}^{n} w(d_i)} \quad \text{..... .....} \quad (2.1)$$

โดย f(x, y) = ค่าข้อมูล ณ จุด (x, y) ที่ต้องการประมาณค่า

w = เป็นฟังก์ชันของน้ำหนัก

z_i = เป็นค่าข้อมูล ณ จุดที่ i

d_i = เป็นระยะทางจากจุดที่ i ไปยังจุด (x, y)

ถึงแม้ว่าการถ่วงน้ำหนักเป็นวิธีการเชิงเส้นตรง (Sampson, 1978) ฟังก์ชันค่าถ่วงน้ำหนักจะเป็น $w=d^{-1}$ โดย $w(0)=\infty$ แต่ถ้าเป็น negative exponential weighting function จะประมาณค่าเฉพาะค่าเริ่มต้นที่ตำแหน่งของตัวอย่างหลายๆ จุด สำหรับการประมาณค่าด้วยวิธีการนี้มีข้อเสียหลายประการคือ มักทำให้เกิดความกำกวม (ambiguity) โดยเฉพาะใช้กับคุณสมบัติของสิ่งที่อยู่ภายใต้พื้นผิวที่ไม่ทราบค่าแน่นอน วิธีการนี้อาจง่ายต่อการเกิดผลกระทบจากความไม่สม่ำเสมอจากการกระจายตัวของข้อมูล เช่น การเกิดค่าถ่วงน้ำหนักที่เท่ากัน (equal weight) ปัญหานี้ถูกค้นพบโดย Delfiner and Delhomme (1975) และถูกแก้ไขโดยการเฉลี่ยค่าข้อมูลหรือเลือกข้อมูลเพียงจุดเดียวในการแสดงค่ากลุ่มข้อมูล (Sampson, 1978) วิธีการนี้ถูกจำกัดขอบเขตภายใต้สมการ $\min(z_i) \leq f(x,y) \leq \max(z_i)$ เช่นเดียวกันกับ $w(d_i)$ ต้องมากกว่า 0 หรืออีกนัยหนึ่งคือ การประมาณค่าแบบนี้จำเป็นที่จะต้องมีขั้นตอนการทำให้ข้อมูลสม่ำเสมอและต่อเนื่องขึ้น (Crain and Bhattacharyya, 1967)

(ข) วิธีการคริจจิ้ง (Kriging)

วิธีการประมาณค่าดังกล่าวนี้ค่อนข้างแตกต่างออกไปจากวิธีอื่น โดย Krige (1976) ได้เริ่มพัฒนาวิธีการนี้มาจากการสังเกตและหาแนวสายแร่ทองคำในประเทศแอฟริกาและพัฒนามาเป็นทฤษฎีตัวแปรเชิงพื้นที่ (regionalized variable theory) โดยอาศัยการคำนวณทางคณิตศาสตร์ เพื่อหาแนวโน้มในการปรับข้อมูลที่มีทิศทางและเส้นที่ไม่ต่อเนื่องให้มีความต่อเนื่องและสม่ำเสมอขึ้น ต่อมาเรียกวิธีการนี้ว่า Universal Kriging และใช้กันอย่างแพร่หลายในทางภูมิศาสตร์สถิติ (geostatistic) รวมทั้งการทำแผนที่เส้นชั้นความสูง และแผนที่พื้นที่ผิวต่อเนื่อง (surface mapping) วิธีการนี้จะประมาณค่าข้อมูลที่ต้องการจากค่าสังเกตที่เก็บมาจากตำแหน่งใกล้เคียง โดยอาศัยวาริโอแกรม (variogram) ซึ่งเป็นเครื่องมือทางสถิติเชิงพื้นที่ในการศึกษาลักษณะทางธรรมชาติของตัวแปรและเป็นฟังก์ชันคณิตศาสตร์ที่แสดงความแปรปรวนของตัวแปรที่อ้างอิงกับตำแหน่งได้ หรือสามารถแสดงความสัมพันธ์ระหว่างวาเรียนซ์ (variance) กับระยะทางและทิศทางในการวิเคราะห์ค่าวาเรียนซ์แล้วจึงประมาณค่าข้อมูล ณ ตำแหน่งที่ไม่เคยมีการบันทึกข้อมูลมาก่อน Trangmar et al. (1985) การประเมินค่าโดยวิธีการนี้ได้รับการพิสูจน์ทางคณิตศาสตร์ว่าเป็นการประเมินที่ไม่ลำเอียงและให้ค่าความผิดพลาดในการประมาณค่าต่ำที่สุด ซึ่งเป็นการกำหนด

คุณภาพของสิ่งที่ประมาณค่าและสามารถนำไปใช้ในการคำนวณความน่าจะเป็นและขอบเขตความเชื่อมั่นของค่าที่ต้องการประมาณต่อไป (พันธุ์ลพ, 2540)

ได้มีการนำวิธีการคริจจิ้งมาใช้ในการ ประมาณค่าข้อมูลดินตั้งแต่ปี 1980 เพื่อประโยชน์ในการปรับปรุงการใช้ประโยชน์ที่ดิน การจำแนกอนุกรมวิธานดิน รวมทั้งการศึกษาเกี่ยวกับมลภาวะของดิน แต่วิธีการนี้ยังคงมีข้อจำกัดของเทคนิคธรณีสถิติที่เกี่ยวกับความเสถียรของข้อมูลดินที่ได้จากการสำรวจ ในภาคสนามซึ่งต้องการข้อมูลตัวอย่างจำนวนมากเพื่ออธิบายความสัมพันธ์ของข้อมูลบนพื้นที่นั้น เนื่องจากข้อมูลดินเป็นข้อมูลที่มีความซับซ้อนและมีการกระจายตัวที่ไม่เป็นระเบียบแบบแผนที่แน่นอน (McBratney and Odeh, 1998)

Mueller et al. (2000) ได้ศึกษาผลจากขนาดของกริดที่ใช้ในการเก็บตัวอย่างของดิน และผลของวิธีการประมาณค่าโดยวิธีการคริจจิ้ง และการประมาณค่าแบบระยะทางผกผัน (inverse distance squared) ต่อแผนที่ความอุดมสมบูรณ์ของดินที่ได้จากคุณสมบัติของดิน 4 อย่างคือ ความเป็นกรดเป็นด่างของดิน (pH) ความเป็นกรดเป็นด่างโดยวิธี BrayII (BpH) ปริมาณฟอสฟอรัส (P) และปริมาณโพแทสเซียม(K) พบว่าขนาดกริดที่เหมาะสมในการใช้เก็บข้อมูลดินคือ ขนาด 30 เมตรกริด และเมื่อนำจุดข้อมูลตัวอย่างเหล่านี้ไปทำการประมาณค่าโดยวิธีการทั้งสอง พบว่าวิธีการคริจจิ้งไม่ได้ให้ค่าผลลัพธ์ที่ดีกว่าวิธีการประมาณค่าแบบระยะทางผกผันมากนัก อีกทั้งวิธีการก็ซับซ้อนกว่าทำให้ต้องใช้เวลาในการทำงานมากกว่า

2.6.2 การประมาณค่าเชิงพื้นที่

วิธีการนี้ใช้กับข้อมูลที่เป็นแผนที่ซึ่งแสดงการกระจายตัวของข้อมูลในลักษณะเป็นเส้นที่เกิดจากการเชื่อมโยงจุดที่มีค่าเท่ากัน (isopleth mapping) โดยการแปลงข้อมูลจากชุดข้อมูลของพื้นที่หนึ่งเรียกว่าโซนเริ่มต้น (source zones) แล้วส่งค่าไปยังอีกชุดข้อมูลของพื้นที่หนึ่งที่เรียกว่าโซนเป้าหมาย (target zones) ที่มีขอบเขตหรือพื้นที่ที่มีการซ้อนทับกันอาจไม่ตรงกันหรือไม่สอดคล้องกันก็ได้ แต่จะมีความยุ่งยากขึ้นถ้าขอบเขตของโซนเป้าหมายเป็นอิสระหรือไม่เกี่ยวข้องกับโซนเริ่มต้น เช่น การประมาณค่าประชากรที่อยู่ในแผนขอบเขตการใช้ไฟฟ้า จากแผนที่ประชากรที่แบ่งขอบเขตตามการปกครองซึ่งมีขอบเขตไม่ตรงกัน การศึกษาการประมาณค่าในลักษณะนี้ยังมีค่อนข้างน้อย แต่ก็มีบ้างโดยเฉพาะด้านที่เกี่ยวข้องกับภูมิศาสตร์ (Brian, 1997; Lam, 1983)

2.6.2.1 การประมาณค่าแบบไม่มีปริมาตร (Non-volume)

การประมาณค่าแบบนี้จะยึดเอาการประมาณค่าแบบจุดเป็นพื้นฐาน จึงมีวิธีการประมาณค่าคล้ายคลึงกับการประมาณค่าแบบจุด ดังที่กล่าวมาแล้วข้างต้น Brian (1997) ได้ยกตัวอย่างของการประมาณค่าประชากรจากข้อมูลสถิติการสำมะโนประชากรไปยังประชากรแบ่งตามเขตการศึกษา ขั้นตอนในการประมาณค่าวิธีนี้เริ่มจากการคำนวณหาค่าความหนาแน่นประชากรสำหรับแต่ละโซนเริ่มต้น โดยหารจำนวนประชากรด้วยพื้นที่ แล้วกำหนดจุดศูนย์กลางหรือจุดควบคุม (control points) สำหรับทุกเขต จากนั้นจึงใช้จุดควบคุมเหล่านั้นมาทำการประมาณค่ากริดของพื้นที่ผิวความหนาแน่นประชากร แล้วจึงเริ่มแปลงค่าจากกริดเซลที่ได้ให้เป็นจำนวนประชากร แล้วซ้อนทับกริดที่ได้จากการประมาณค่าลงบนโซนเป้าหมาย สุดท้ายคือการคำนวณค่าประชากรรวมในโซนเป้าหมาย

การใช้วิธีนี้ในการประมาณค่าต้องระมัดระวังเรื่องของการกำหนดจุดควบคุมของเขตให้ถูกต้อง เพราะสามารถทำให้เกิดข้อผิดพลาดได้ เนื่องจากรูปร่างของพื้นที่มักเป็นรูปที่ไม่สมมาตร ส่วนการประมาณค่าจากจุดที่นำมาใช้ร่วมกับวิธีการนี้ก็ยังคงมีข้อบกพร่องในแต่ละวิธีการอยู่ก่อนแล้ว นอกจากนี้ยังอาจเกิดความไม่แน่นอนขึ้นระหว่างการกำหนดค่าให้แก่กริดในบางสภาวะ โดยเฉพาะอย่างยิ่งเมื่อจุดกริดที่อยู่ตรงกันข้ามกันของจุดศูนย์กลางที่ไม่เชื่อมโยงกันมีค่าเหมือนกัน ซึ่งผิดไปจากคู่อื่น เป็นต้น (Lam, 1983)

2.6.2.2 การประมาณค่าเชิงปริมาตร (Volume)

การประมาณค่าแบบนี้จะใช้ตัวขอบเขตของข้อมูลเองมาพิจาณามากกว่าการเลือกกำหนดจุดควบคุมเอาเองอย่างวิธีการประมาณค่าจากพื้นที่โดยไม่มีปริมาตรที่กล่าวมา ดังนั้นจึงไม่มีการประมาณค่าจากจุดเข้ามาเกี่ยวข้อง การประมาณค่าจากพื้นที่แบบมีปริมาตรนี้มี 2 วิธีคือ การซ้อนทับ (overlay) และ pycnophylatic

การซ้อนทับ (Overlay)

Flowerdew and Green (1994); Ballard and Schut (1997) และ Lam (1983) อธิบายวิธีการซ้อนทับชั้นข้อมูลหนึ่งบนชั้นข้อมูลอื่นๆ ที่มีตำแหน่งทางภูมิศาสตร์ที่สามารถอ้างอิงถึงกันได้ ขั้นตอนการประมาณค่าเริ่มจากการนำโซนเป้าหมายมาซ้อนทับบนโซนเริ่มต้น จากนั้นทำการกำหนดสัดส่วนของแต่ละโซนเริ่มต้นตามพื้นที่ที่ปรากฏอยู่ในโซนเป้าหมาย แล้วจึงทำการคำนวณค่าข้อมูลรวมภายในโซนเป้าหมายจากข้อมูลโซนเริ่มต้น ตามสัดส่วนของพื้นที่โซนเริ่มต้นที่ปรากฏอยู่ในโซนเป้าหมายทั้งหมด (areal weight) กล่าวอีกนัยหนึ่ง

คือการใช้พื้นที่ของโซนเริ่มต้นที่ปรากฏอยู่ในโซนเป้าหมายเป็นค่าน้ำหนักในการคำนวณข้อมูลรวมภายในโซนเป้าหมายดังแสดงในสมการ

$$y_t = \frac{\sum_s y_{st} A_{st}}{\sum_s A_{st}} \quad \text{.....} \quad (2.2)$$

โดย Y_t = เป็นค่าข้อมูลผลลัพธ์รวมในโซนเป้าหมาย

Y_{st} = เป็นค่าข้อมูลของโซนเริ่มต้นที่ที่ปรากฏในโซนแผนที่เป้าหมาย

A_{st} = เป็นพื้นที่ของโซนเริ่มต้นที่ปรากฏในโซนเป้าหมาย

การประมาณค่าโดยวิธีการซ้อนทับนี้ต้องพิจาณาถึงชนิดของข้อมูลที่แตกต่างกัน (Lam, 1983; Ballard and Schut,1997) เพราะชนิดของข้อมูลที่แตกต่างกันนั้นจะส่งผลกระทบต่อการประมาณค่าไปยังโซนเป้าหมายกล่าวคือ (1) ถ้าข้อมูลเป็นข้อมูลชนิดที่เป็นนาม ชื่อ หรือข้อมูลที่มีความเฉพาะ (nominal data) และข้อมูลที่เป็นชนิด หรือเลขลำดับ (ordinal data) นั้นค่าของข้อมูลในโซนเป้าหมายจะแปรผันตามเปอร์เซ็นต์พื้นที่ของโซนเริ่มต้นที่เกิดขึ้นในโซนเป้าหมายนั้นๆ ดังแสดงในสมการ

$$y_t = \sum_s \left(y_{st} \times \frac{A_{st}}{A_t} \right) \quad \text{.....} \quad (2.3)$$

ส่วนข้อมูลชนิดที่เป็นค่าความสัมพันธ์ ค่าเฉลี่ย หรือค่าความหนาแน่น (interval data) นั้นค่าข้อมูลรวมที่โซนเป้าหมายจะได้จากผลรวมของค่าข้อมูลที่โซนเริ่มต้นคูณด้วยพื้นที่ในโซนย่อยที่เกิดจากซ้อนทับหารด้วยพื้นที่ทั้งหมด

$$y_t = \frac{\left(\sum_s y_s \times A_{st} \right)}{A_t} \quad \text{.....} \quad (2.4)$$

สำหรับข้อมูลที่เป็นสัดส่วน หรือจำนวนนับ (ratio) นั้นค่าข้อมูลที่โซนเป้าหมายจะประมาณได้จากสมการ

$$y_t = \frac{\left(\sum_s y_s \times A_{st} \right)}{A_t} \quad \text{.....} \quad (2.5)$$

โดย Y_t = เป็นค่าข้อมูลผลลัพธ์รวมในโซนเป้าหมาย

Y_s = เป็นค่าข้อมูลของโซนเริ่มต้น

Y_{st} = เป็นค่าข้อมูลของโซนเริ่มต้นที่ปรากฏในโซนเป้าหมาย

A_s = เป็นพื้นที่ของโซนเริ่มต้น

A_t = เป็นพื้นที่ทั้งหมดของโซนเป้าหมาย

A_{st} = เป็นพื้นที่ของโซนเริ่มต้นที่ปรากฏในโซนเป้าหมาย

Ballard and Schut (1997) กล่าวว่าการประมาณค่าโดยวิธีการซ้อนทับชั้นข้อมูลนี้มีความสมเหตุสมผลในการอ้างข้อมูลจากโซนเริ่มต้นมาสู่โซนเป้าหมาย เนื่องจากโซนทั้งสองนั้นเป็นโซนที่มีตำแหน่งอ้างอิงบนผิวโลกที่ตรงกัน และผลลัพธ์ที่ได้นั้นควรพิจารณาถึงข้อมูลตามความเป็นจริง เพื่อป้องกันข้อผิดพลาดที่อาจเกิดขึ้นได้ เช่นการประมาณค่าประชากรจากแผนแสดงจำนวนประชากร โดยมีแผนที่นิเวศวิทยาเป็นแผนที่เป้าหมาย และเนื่องจากแผนที่เป้าหมายนี้บางหน่วยแผนที่เป็นพื้นที่ที่เป็นทะเลสาป หรือพื้นที่ที่คนไม่สามารถใช้เป็นที่อยู่อาศัยได้ การประมาณค่าจากโซนเริ่มต้นจึงควรใช้วิจารณญาณและคำนึงถึงความเป็นจริงอีกทางหนึ่งด้วย

ปัญหาที่สำคัญในการประมาณค่าด้วยวิธีนี้คือ ความสม่ำเสมอ (homogeneity) ภายในแต่ละโซนเริ่มต้น กล่าวคือถ้าค่าของแต่ละโซนเริ่มต้นเหมือนกันทุกจุดทุกที่แล้ว การรวมกันเพื่อเป็นโซนเป้าหมายจะให้ผลการประมาณค่าที่เท่าเดิมทุกโซนเป้าหมาย แต่ถ้าค่าของแต่ละโซนเริ่มต้นมีการกระจายที่ไม่สม่ำเสมอการประมาณค่านั้นก็ขาดความน่าเชื่อถือได้ นอกจากนั้นความน่าเชื่อถือของการประมาณค่าแบบนี้ยังขึ้นอยู่กับธรรมชาติและระดับของความไม่เหมือนกัน (inhomogeneity) ของโซนเริ่มต้นรวมทั้งขนาดของโซนเป้าหมายที่สัมพันธ์กับโซนเริ่มต้นที่อยู่ในตำแหน่งเดียวกันด้วย

ได้มีการศึกษาปัญหาของการรวมกันอีกครั้ง (reaggregation) ของเส้นชั้นความสูงโดยใช้การประมาณค่าเชิงพื้นที่แบบจุด (point-base areal interpolation method) สรุปได้ว่ามีสภาวะหลายอย่างที่ควรพิจารณา เนื่องจากสภาวะเหล่านั้นมีผลกระทบต่อการได้มาและความน่าเชื่อถือของผลลัพธ์นั้น โดยเฉพาะอย่างยิ่งเรื่องของระดับความสม่ำเสมอของข้อมูลเชิงพื้นที่ (spatial homogeneity) (Lam, 1983)

แต่ในความเป็นจริงแล้วมีข้อมูลโซนเริ่มต้นหลายประเภทที่มีการกระจายตัวของข้อมูลที่เหมือนกัน ส่วนความไม่เหมือนกันเกิดจากข้อเท็จจริงที่ว่าแผนที่ thematic นั้นเกิดจากการสำรวจโดยการสุ่มตัวอย่าง ดังนั้นขนาดของตัวอย่างและวิธีการสุ่มตัวอย่างจึงมี

ความสำคัญต่อการกำหนดคุณภาพและความแม่นยำของแผนที่ นอกจากนั้นความรู้เกี่ยวกับการกระจายตัวเชิงพื้นที่ของข้อมูลตามธรรมชาติ ค่าของข้อมูล หรือลักษณะเฉพาะของข้อมูลที่จะเป็นตัวบ่งชี้ไปยังโซนเป้าหมายจะเป็นตัวกำหนดความถูกต้องแม่นยำของโซนเป้าหมาย

การทำ Pycnophylatic

วิธีการนี้ได้รับการพัฒนาขึ้นมาโดย Tobler (1979) มีวัตถุประสงค์เพื่อสร้างข้อมูลให้มีความสม่ำเสมอ ไม่เปลี่ยนค่าโดยฉับพลัน โดยเฉพาะบริเวณขอบของพื้นที่ และต้องการให้ค่าข้อมูลรวมภายในโซนเป้าหมายมีความถูกต้องมากขึ้น โดยจะทำการซ้อนทับ dense raster บน choropleth map แบ่งแต่ละโซนของข้อมูลรวมให้เท่ากันทุกเซลของชั้นข้อมูลราสเตอร์ที่เหลื่อมกันอยู่ จากนั้นปรับค่าข้อมูลให้สม่ำเสมอด้วยการแทนค่าแต่ละเซลด้วยค่าเฉลี่ยของเซลข้างเคียง (neighbor) ก่อนทำการรวมค่าของเซลในแต่ละโซน แล้วปรับค่าของข้อมูลภายในแต่ละโซนอย่างเป็นสัดส่วน ดังนั้นโซนเป้าหมายรวมจะเหมือนกับโซนเริ่มต้น เมื่อทำการปรับข้อมูลด้วยการแทนค่าด้วยค่าเฉลี่ยเซลข้างเคียงไปเรื่อยจนข้อมูลไม่มีการเปลี่ยนแปลง การประมาณค่าจะผิดพลาดได้หากไม่มีค่าข้อมูลบริเวณใกล้เคียงกับบริเวณที่ต้องการจะประมาณค่า เนื่องจากต้องอาศัยข้อมูลของเซลข้างเคียงมาคำนวณ

2.7 การตรวจสอบความสอดคล้อง

ในการวิเคราะห์ข้อมูลเชิงพื้นที่โดยการซ้อนทับชั้นข้อมูลที่มากกว่าหนึ่งชั้นข้อมูล จำเป็นที่จะต้องทำการเปรียบเทียบความถูกต้องและความสอดคล้องของข้อมูลในแต่ละชั้นที่นำมาวิเคราะห์ร่วมกัน ทั้งนี้เพื่อประโยชน์ต่อการอธิบายความสัมพันธ์ของข้อมูลและการกระจายตัวของข้อมูลในแต่ละชั้น การเปรียบเทียบข้อมูลเชิงพื้นที่หรือการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลเชิงพื้นที่จะแตกต่างกันไปตามชนิดของชั้นข้อมูล เช่นถ้าข้อมูลเป็นข้อมูลที่เป็นนาม ชื่อ หรือข้อมูลที่มีความเฉพาะ จะใช้ตารางความคลาดเคลื่อน (error matrix; contingency table หรือ confusion matrix) ในการแสดงความสัมพันธ์กันระหว่างข้อมูล (Bonham-Carter, 1994; Yue Hong Chou, 1997) และยังสามารถแสดงค่าความถูกต้องในส่วนของผู้ผลิต (producer accuracy หรือ observed proportions matrix) ผู้ใช้ (user accuracy หรือ expected proportion matrix) และค่าความถูกต้องโดยรวม (overall accuracy) นอกจากนั้นยังสามารถใช้วิธีวิเคราะห์สถิติ kappa statistic ซึ่งเป็นวิธีการมาตรฐานที่ใช้ในการวิเคราะห์ความสัมพันธ์ของข้อมูลเชิงพื้นที่ทั้งข้อมูลแบบราสเตอร์และเวกเตอร์ (Congalton, 1991)

Joseph (1999) ได้แนะนำการเปรียบเทียบความสอดคล้องหรือความคล้ายคลึงกันของข้อมูลเชิงพื้นที่โดยวิธีการซ้อนทับว่า ถ้าข้อมูลเป็นแบบราสเตอร์ควรใช้การนับเซล (cells) และใช้พื้นที่รวมมาใช้ในการสร้างเป็นตารางความคลาดเคลื่อน แต่ถ้าข้อมูลเป็นแบบเวกเตอร์ สามารถใช้พื้นที่ของรูปหลายเหลี่ยมที่ได้จากการซ้อนทับนั้นมาสร้างเป็นตารางความคลาดเคลื่อนก่อนนำมาวิเคราะห์เชิงสถิติพบว่าสามารถใช้ในเปรียบเทียบความสอดคล้องของข้อมูลได้เป็นอย่างดี

เนื่องจากข้อมูลดินเป็นข้อมูลที่มีความแปรปรวนสูง และมีลักษณะและคุณสมบัติที่แตกต่างกันออกไปตามสภาพภูมิประเทศ ภูมิอากาศ ต้นกำเนิดของดินเอง และปัจจัยอื่นๆ อีกมากมาย รวมทั้งอาจเกิดจากการกระทำของมนุษย์ เป็นผลให้ข้อมูลดินค่อนข้างจะมีการเปลี่ยนแปลงไปตามเวลา ดังนั้นในการพัฒนาและปรับปรุงฐานข้อมูลเชิงพื้นที่และฐานข้อมูลอรรถาธิบายข้อมูลกลุ่มชุดดินและข้อมูลชุดดินจากข้อมูลที่มีอยู่เดิมให้สมบูรณ์และเป็นประโยชน์มากขึ้นนั้น จำเป็นที่จะต้องใช้วิธีการทางภูมิศาสตร์และการจัดการฐานข้อมูลเข้ามาใช้ร่วมกัน และระบบสารสนเทศทางภูมิศาสตร์เป็นระบบหนึ่งที่สามารถนำข้อมูลเหล่านั้นมาเชื่อมโยงและสร้างความสัมพันธ์กันเพื่อให้สามารถใช้ประโยชน์ข้อมูลที่มีอยู่อย่างเป็นรูปธรรมมากขึ้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Chiang Mai University