

Lab3: Explore Data II

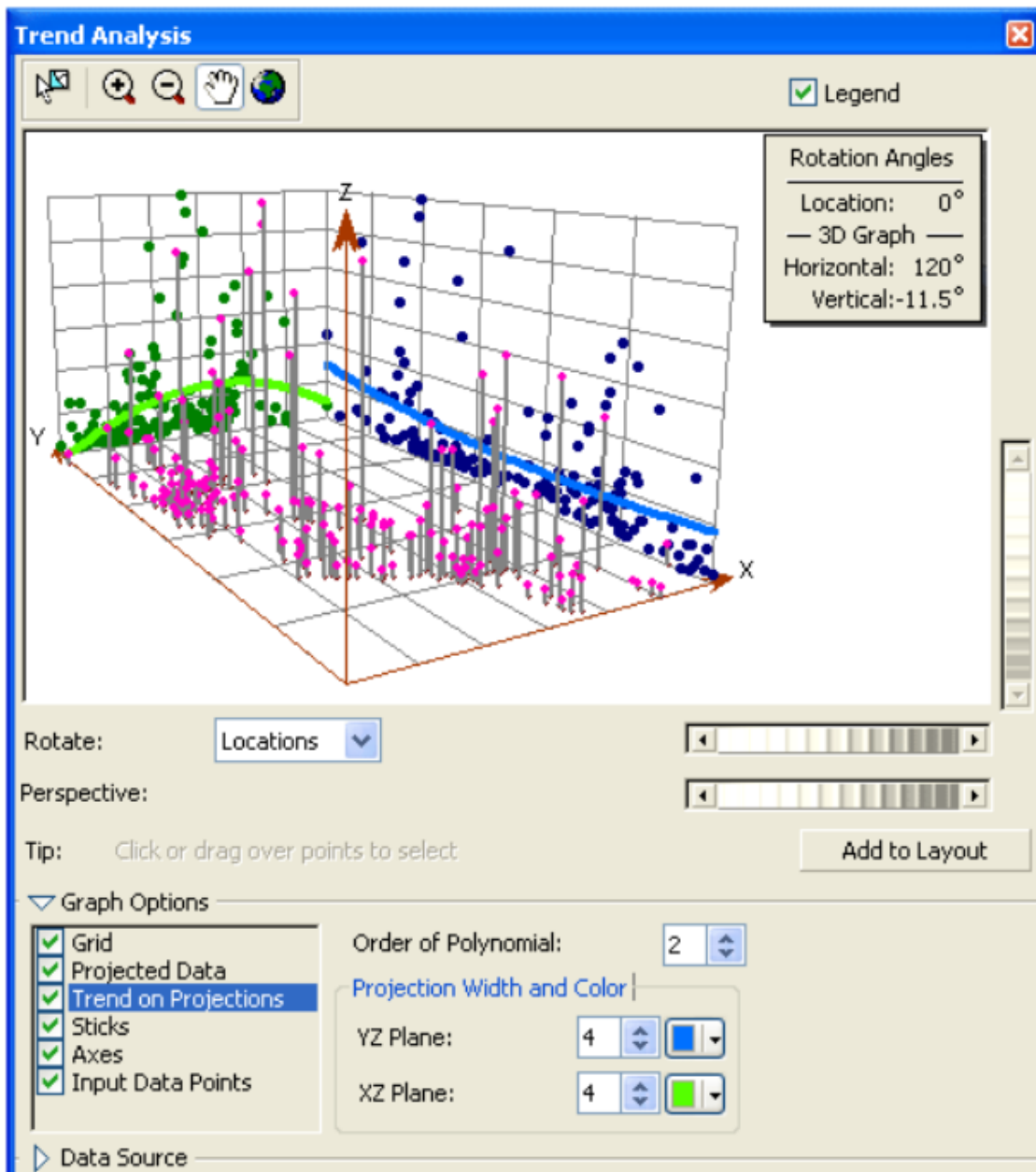
3. Trend Analysis

การวิเคราะห์แนวโน้ม (Trend Analysis): ใช้ในการศึกษาแนวโน้มของข้อมูล หรืออาจใช้เพื่อกำจัดแนวโน้มออกจากชุดข้อมูลในระหว่างการใช้ **kriging** (การประมาณค่าผิวพื้นที่โดยใช้ข้อมูลตัวอย่าง) เครื่องมือนี้ช่วยในการระบุแนวโน้มในชุดข้อมูลที่ป้อนเข้าไป และสามารถให้มุมมองที่ช่วยให้เห็นแนวโน้มและพฤติกรรมของข้อมูลได้ชัดเจนยิ่งขึ้น

ฟีเจอร์หลักของเครื่องมือ Trend Analysis:

1. **มุมมองสามมิติของข้อมูล:** ข้อมูลตัวอย่างแต่ละจุดจะถูกแสดงในลักษณะสามมิติ โดยแสดงในแผนภาพระนาบ x,y (แผนภาพสองมิติของตำแหน่งตัวอย่าง) และค่าของตัวอย่างจะถูกแสดงผ่านความสูงของเสาในมิติ z ซึ่งช่วยให้มองเห็นข้อมูลจากมุมมองที่แตกต่าง
2. **การฉายข้อมูลบนแผนภาพสองมิติ:** ข้อมูลจะถูกฉายลงบนระนาบ x,z และ y,z เพื่อแสดงเป็น scatterplots ซึ่งเหมือนการมองข้อมูลจากด้านข้าง ผ่านมุมมองสามมิติ ทำให้สามารถเห็นแนวโน้มในแต่ละทิศทางได้ง่ายขึ้น
3. **การปรับพหุนาม:** เมื่อแสดงข้อมูลบนระนาบ x,z และ y,z เครื่องมือจะทำการปรับพหุนาม (polynomials) ผ่านจุดข้อมูลใน scatterplots ที่ฉายออกไป ซึ่งจะช่วยให้สามารถระบุแนวโน้ม (trend) ของข้อมูลในทิศทางต่างๆ ได้ชัดเจน โดยปกติแล้วเครื่องมือจะเลือกพหุนามอันดับที่สอง (second-order polynomials) เพื่อแสดงแนวโน้มในข้อมูล แต่คุณสามารถเลือกพหุนามอันดับหนึ่งหรือสามเพื่อทดสอบว่าแนวโน้มไหนเหมาะสมที่สุดกับข้อมูลของคุณ
4. **การหมุนและปรับมุมมอง:** คุณสามารถหมุนข้อมูลเพื่อแยกแยะแนวโน้มที่มีทิศทางเฉพาะ และปรับมุมมองของภาพเพื่อสำรวจข้อมูลในทิศทางต่าง ๆ ได้ เครื่องมือยังมีฟีเจอร์ในการปรับขนาดและสีของจุดข้อมูลและเส้น รวมถึงการลบแผนภาพหรือจุดที่ไม่ต้องการออกไป

เครื่องมือนี้มีประโยชน์ในการทำความเข้าใจแนวโน้มในข้อมูลทั้งในลักษณะทั่วไปหรือเฉพาะเจาะจง โดยการหมุนและปรับมุมมองช่วยให้สามารถมองเห็นแนวโน้มที่อาจไม่สามารถสังเกตได้จากมุมมองธรรมดา



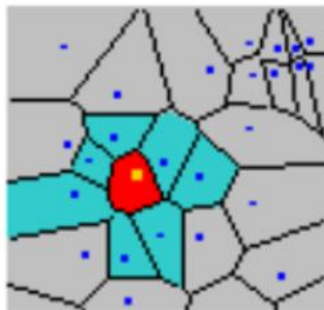
4. Voronoi Map

Voronoi Map คือแผนที่ที่สร้างขึ้นจากชุดของโพลีกอน (Polygon) ที่ล้อมรอบจุดตัวอย่าง (Sample Point) ในแต่ละจุดบนแผนที่ การใช้ Voronoi Map ช่วยในการแสดงถึงความแปรผันของข้อมูลในท้องถิ่น โดยแบ่งพื้นที่ออกเป็นหลายๆ โซนตามตำแหน่งของจุดตัวอย่างต่างๆ **Voronoi Map** ใน ArcMap คือแผนที่ที่แสดงพื้นที่ที่ถูกแบ่งโดยใช้จุดต้นทาง (Input Points) เป็นฐานในการคำนวณ พื้นที่ที่ได้เรียกว่า **Voronoi Polygons** หรือ **Thiessen Polygons** ซึ่งเป็นขอบเขตที่กำหนดว่าพื้นที่ใดอยู่ใกล้จุดใดมากที่สุด

หลักการของ Voronoi Map

- แต่ละ Polygon จะครอบคลุมพื้นที่ที่ใกล้กับจุดต้นทางของตนมากที่สุด
- เส้นขอบของแต่ละโพลีกอนคือระยะกึ่งกลาง (Perpendicular Bisector) ระหว่างจุดต้นทางสองจุด
- ใช้ในงานวิเคราะห์เชิงพื้นที่ เช่น การแบ่งพื้นที่ให้บริการ (Service Areas) หรือการศึกษาความสัมพันธ์เชิงพื้นที่

ตัวอย่างเช่น ในภาพด้านล่าง จุดตัวอย่างสีเหลืองซึ่งถูกล้อมรอบด้วย Polygon สีแดง โดยทุกตำแหน่งภายใน Polygon สีแดงจะอยู่ใกล้กับจุดตัวอย่างสีเหลืองมากกว่าจุดตัวอย่างอื่นๆ (ซึ่งแสดงด้วยจุดสีน้ำเงินเข้มขนาดเล็ก) ส่วน Polygon สีน้ำเงินทั้งหมดที่แชร์ขอบร่วมกับ Polygon สีแดง จะถือว่าจุดตัวอย่างในโพลีกอนสีน้ำเงินเหล่านี้เป็นเพื่อนบ้าน (neighbors) ของจุดตัวอย่างสีเหลือง



การใช้ Geostatistical Wizard เพื่อสร้าง Voronoi Map มีจุดเด่นในด้านการช่วยวิเคราะห์พื้นที่โดยอิงตามจุดข้อมูล (Point Data) และสามารถปรับแต่งโมเดลได้อย่างยืดหยุ่น

คุณสมบัติเด่นของการสร้าง Voronoi Map ผ่าน Geostatistical Wizard

1. สร้างพื้นที่อิทธิพล Voronoi Map ที่ระบุพื้นที่ที่ใกล้กับจุดข้อมูลต้นทางมากที่สุด โดยแต่ละพื้นที่จะแสดงขอบเขตการครอบคลุมของจุดหนึ่งจุด
2. การผสมผสานกับการประมาณค่า (Interpolation) Voronoi Polygons เป็นพื้นฐานสำหรับการประมาณค่าข้อมูลเชิงพื้นที่ เช่น การวิเคราะห์ความหนาแน่น หรือตัวแปรด้านสิ่งแวดล้อม
3. การปรับแต่งโมเดล โดยสามารถปรับแต่งพารามิเตอร์ในกระบวนการคำนวณ เช่น ระยะความสัมพันธ์ (Semivariogram) เพื่อให้โมเดลเหมาะสมกับข้อมูล

วิธีการคำนวณค่าของ Polygon ในเครื่องมือ Voronoi Map

1. **Simple:** ค่าเฉพาะของ Polygon หรือเป็นค่าที่บันทึกที่จุดตัวอย่างภายใน Polygon นั้น
2. **Mean (ค่าเฉลี่ย):** ค่าเฉพาะของ Polygon จะเป็นค่าเฉลี่ยที่คำนวณจากค่าของ Polygon และค่าของเพื่อนบ้าน
3. **Mode (ฐานนิยม):** ทุก Polygon จะถูกจัดกลุ่มเป็น 5 class ค่าเฉพาะของ Polygon จะเป็นโหมด (ค่าอันที่เกิดบ่อยที่สุด) ของ Polygon และค่าของเพื่อนบ้าน
4. **Cluster (กลุ่ม):** ทุก Polygon จะถูกจัดกลุ่มเป็น 5 class หากช่วงคลาสของ Polygon แตกต่างจากช่วงคลาสของเพื่อนบ้าน Polygon นี้จะถูกทาสีเป็นสีเทาและใส่ในคลาสที่หกเพื่อแยกความแตกต่างจากเพื่อนบ้าน
5. **Entropy (ความไม่แน่นอน):** ทุก Polygon จะถูกจัดกลุ่มเป็น 5 class ตามการจัดกลุ่มของค่าข้อมูล (smart quantiles) ค่าเฉพาะของ Polygon จะเป็น entropy ที่คำนวณจาก Polygon และเพื่อนบ้าน ซึ่งหมายถึงการวัดความไม่แน่นอนหรือความหลากหลายของข้อมูลในพื้นที่นั้น

$$\text{Entropy} = - \sum (p * \text{Log } p),$$

การคำนวณการกระจายของค่าในแต่ละคลาส

ในกรณีนี้, p หมายถึงสัดส่วนของ Polygon ที่ได้รับการกำหนดให้กับแต่ละคลาส โดยการคำนวณจะพิจารณาจากจำนวนของ Polygon ที่ได้รับการจัดกลุ่มในแต่ละคลาส เปรียบเทียบกับจำนวนทั้งหมดของ Polygon ที่มีอยู่รอบๆ

ตัวอย่างการคำนวณ:

สมมติว่าเราใช้ Polygon ที่มีเพื่อนบ้านสี่จุด (รวมเป็น 5 Polygon ซึ่งคือ Polygon หลักที่เรากำลังพิจารณาและเพื่อนบ้านอีกสี่จุด)

1. ค่าของแต่ละ Polygon จะถูกจัดกลุ่มตามช่วงคลาสต่าง ๆ
2. Polygon ที่ถูกจัดให้อยู่ในคลาสต่างๆ จะมีการนับจำนวนว่าแต่ละคลาสมี Polygon อยู่กี่จุด
3. p คือสัดส่วนของ Polygon ที่อยู่ในคลาสใดคลาสหนึ่ง ซึ่งคำนวณได้จากจำนวน Polygon ในคลาสนั้น ๆ หารด้วยจำนวนทั้งหมดของ Polygon ในกลุ่ม (ในที่นี้คือ 5 Polygon)

การคำนวณค่า p : ถ้า Polygon ทั้ง 5 (รวม Polygon หลักและเพื่อนบ้าน) ถูกแบ่งออกเป็น 3 คลาส ดังนี้: คลาส A: มี 2 Polygon, คลาส B: มี 2 Polygon, คลาส C: มี 1 Polygon

$$p \text{ สำหรับแต่ละคลาสจะคำนวณดังนี้: } p (\text{คลาส A}) = 2/5 = 0.4, p (\text{คลาส B}) = 2/5 = 0.4, p (\text{คลาส C}) = 1/5 = 0.2$$

การคำนวณนี้ช่วยให้เรารู้ว่าแต่ละคลาสมีสัดส่วนของ Polygon ที่ได้รับการกำหนดให้กับมันอย่างไรในพื้นที่นั้น ๆ

การประยุกต์ใช้ใน Voronoi Map:

การคำนวณค่า μ และการจัดกลุ่ม Polygon ตามคลาสเหล่านี้ช่วยให้เราวิเคราะห์การกระจายตัวของค่าท้องถิ่น และสามารถมองเห็นรูปแบบต่างๆ ในพื้นที่ที่มีการกระจายตัวของข้อมูลที่แตกต่างกันได้อย่างชัดเจน

ความไม่แน่นอน (Entropy) ที่ได้รับการกำหนดให้กับ Polygon จะเป็น

$$E = -[0.6 * \log_2 (0.6) + 0.2 * \log_2 (0.2) + 0.2 * \log_2 (0.2)] = 1.371$$

Minimum Entropy จะเกิดขึ้นเมื่อค่าของ Polygon ทั้งหมดอยู่ในคลาสเดียวกันทั้งหมด จากนั้น

$$E_{\min} = -[1 * \log_2 (1)] = 0$$

Entropy) ต่ำสุดจะเกิดขึ้นเมื่อค่าของ Polygon ทั้งหมดอยู่ในคลาสเดียวกัน จากนั้น

$$E_{\max} = -[0.2 * \log_2 (0.2) + 0.2 * \log_2 (0.2) + 0.2 * \log_2 (0.2) + 0.2 * \log_2 (0.2) + 0.2 * \log_2 (0.2)] = 2.322$$

การอธิบายค่าต่างๆ ที่ถูกกำหนดให้กับ Polygon ใน Voronoi Map:

1. **Median (ค่ามัธยฐาน):** ค่าของ Polygon คือค่ามัธยฐานที่คำนวณจากการแจกแจงความถี่ของ Polygon และเพื่อนบ้าน โดยค่ามัธยฐานเป็นค่ากลางที่แสดงถึงการจัดเรียงค่าจากน้อยไปหามากและเลือกค่ากลางที่สุด
2. **Standard Deviation (ส่วนเบี่ยงเบนมาตรฐาน):** ค่าของ Polygon คือส่วนเบี่ยงเบนมาตรฐานที่คำนวณจาก Polygon และเพื่อนบ้าน โดยส่วนเบี่ยงเบนมาตรฐานแสดงถึงการกระจายของข้อมูลจากค่าเฉลี่ย ค่าที่สูงหมายถึงข้อมูลมีการกระจายกว้าง, ส่วนค่าต่ำหมายถึงข้อมูลมีการกระจายแคบ
3. **Interquartile Range:** ค่าของ Polygon คือช่วงระหว่างควอไทล์ที่คำนวณจากการแจกแจงความถี่ของ Polygon และเพื่อนบ้าน โดยช่วงนี้ได้แก่ค่าควอไทล์ที่สาม (Q3) ลบด้วยค่าควอไทล์แรก (Q1) ซึ่งใช้ในการวัดการกระจายของข้อมูลในส่วนกลางของชุดข้อมูล (ระหว่าง 25% ถึง 75% ของข้อมูล)

การใช้สถิติ Voronoi สามารถนำไปใช้เพื่อวัตถุประสงค์ที่หลากหลาย และสามารถแบ่งออกเป็นกลุ่มฟังก์ชันหลัก ๆ ดังนี้:

- การวิเคราะห์การกระจายของค่าท้องถิ่น (Local)
- การหาค่าความแปรปรวนในพื้นที่
- การศึกษาความสัมพันธ์ระหว่าง Polygon และเพื่อนบ้าน

General functional categories

Functional category	Voronoi statistics
Local Smoothing	Mean, Mode, Median
Local Variation	Standard deviation, Interquartile range, Entropy
Local Outliers	Cluster
Local Influence	Simple