

Lab6: Experimental Semivariogram

1. Semivariogram/Covariance Cloud

เครื่องมือ Semivariogram/Covariance Cloud: สามารถนำมาใช้ในการแสดงค่า semivariogram และ covariance ที่คำนวณจากข้อมูลในชุดข้อมูลทั้งหมด โดยจะแสดงค่าเหล่านี้สำหรับคู่ของตำแหน่งที่ต่างกันภายในชุดข้อมูล และทำการแสดงผลในรูปแบบของกราฟ ซึ่งค่าเหล่านี้จะแสดงตามระยะทางระหว่างตำแหน่งที่กำหนด

การทำงานของเครื่องมือ Semivariogram/Covariance Cloud:

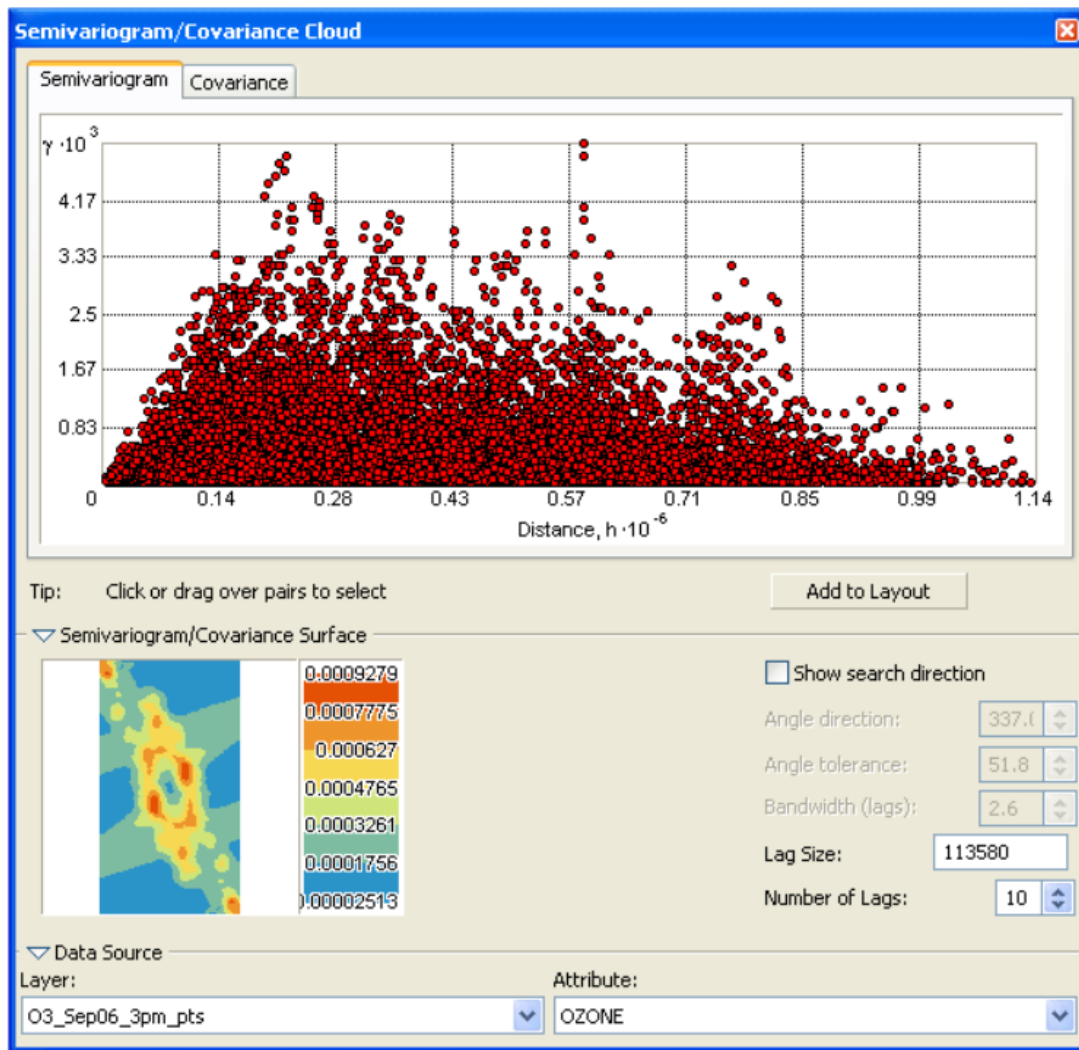
1. **Semivariogram** คือ การวัดความแตกต่างระหว่างค่าที่สองตำแหน่งในชุดข้อมูล ซึ่งจะช่วยในการศึกษาความสัมพันธ์ระหว่างตำแหน่งต่างๆ ในพื้นที่ โดยพิจารณาจากระยะทางระหว่างจุดตัวอย่าง (distance) และค่าความแตกต่างของข้อมูลที่มีระยะห่างต่างกัน
2. **Covariance** คือ การวัดความสัมพันธ์ระหว่างสองตำแหน่งในชุดข้อมูล ซึ่งสามารถแสดงให้เห็นถึงความสัมพันธ์เชิงเส้นหรือการเปลี่ยนแปลงของค่าข้อมูลตามตำแหน่งในพื้นที่

วิธีการแสดงผล:

- เครื่องมือจะแสดงค่าของ semivariogram และ covariance โดยการจับคู่ตำแหน่งข้อมูลที่มีระยะห่างต่างๆ กัน
- กราฟที่ได้จะแสดงค่าต่างๆ ตามระยะทางระหว่างตำแหน่งต่างๆ โดยค่าความสัมพันธ์ของข้อมูลจะถูกแสดงในแกนแนวดิ่ง (vertical axis) ส่วนระยะทางระหว่างตำแหน่งจะถูกแสดงในแกนแนวนอน (horizontal axis)
- ค่าของ semivariogram และ covariance จะช่วยให้เข้าใจถึงการกระจายตัวของข้อมูล และลักษณะของความสัมพันธ์ระหว่างข้อมูลในตำแหน่งต่างๆ

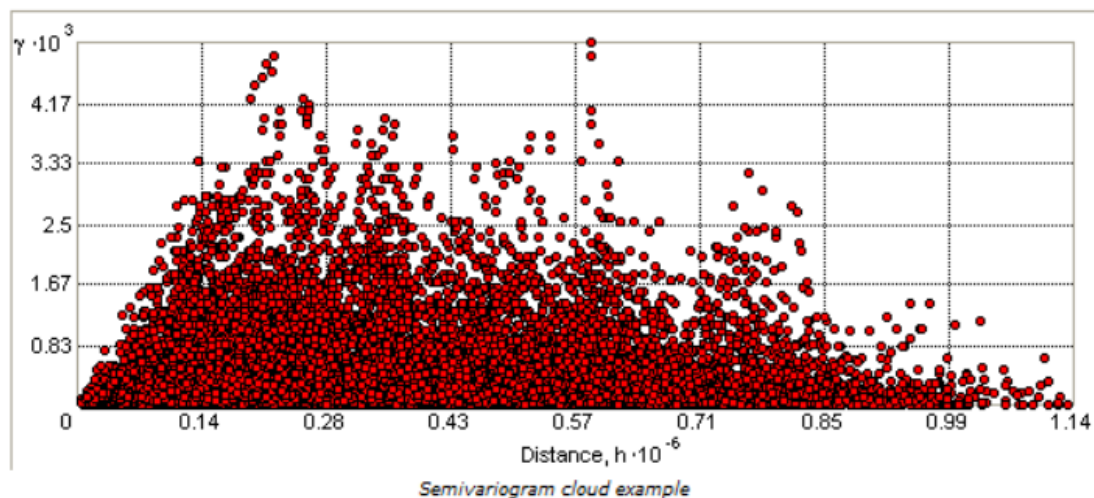
เครื่องมือ **Semivariogram/Covariance Cloud** เป็นเครื่องมือที่สำคัญในกระบวนการ **kriging** และการวิเคราะห์เชิงพื้นที่ เพราะสามารถใช้ในการตรวจสอบโครงสร้างเชิงพื้นที่ของข้อมูล รวมถึงการประมาณค่าความสัมพันธ์ระหว่างตำแหน่งข้อมูลในพื้นที่

เครื่องมือ **Semivariogram/Covariance Cloud** แสดงค่า **semivariogram** และ **covariance** แบบเชิงประจักษ์สำหรับคู่ของตำแหน่งทั้งหมดในชุดข้อมูล และทำการแสดงผลค่าเหล่านี้เป็นฟังก์ชันของระยะทางที่แยกตำแหน่งทั้งสองออกจากกัน ตามตัวอย่างที่แสดงด้านล่างนี้:



Semivariogram covariance cloud illustration

เครื่องมือ **Semivariogram/Covariance Cloud** สามารถใช้ในการตรวจสอบลักษณะเฉพาะของ **spatial autocorrelation** ภายในชุดข้อมูลและค้นหา **outliers** ภายในพื้นที่ท้องถิ่น โดย **semivariogram cloud** จะมีลักษณะประมาณดังนี้:



ในภาพประกอบด้านบน แต่ละจุดสีแดงจะแสดงค่า **semivariogram** แบบเชิงประจักษ์ (ค่าแตกต่างยกกำลังสองระหว่างค่าของข้อมูลสองจุดที่ประกอบเป็นคู่) ที่แสดงผลตามระยะทางที่แยกทั้งสองจุดออกจากกัน คุณสามารถใช้เครื่องมือ **brush** เพื่อเลือกจุดและดูคู่ข้อมูลที่เชื่อมโยงใน **ArcMap** ได้

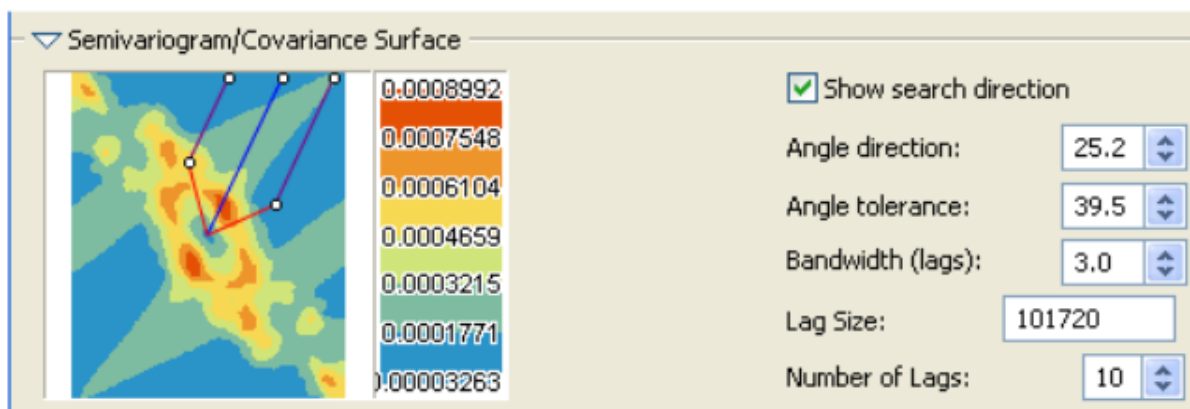
การทำงานของ Semivariogram/Covariance Cloud:

- **Semivariogram/Covariance Cloud** จะมีจุดใน "Cloud" แทนคู่ของจุดข้อมูลในชุดข้อมูล ซึ่งจำนวนจุดใน "Cloud" จะเพิ่มขึ้นอย่างรวดเร็วเมื่อจำนวนจุดในชุดข้อมูลเพิ่มขึ้น
- หากชุดข้อมูลมี **n** จุด จำนวนจุดใน **semivariogram/covariance cloud** จะเป็น $n*(n-1)/2$ จุด ดังนั้นจึงไม่แนะนำให้ใช้ชุดข้อมูลที่มีจุดมากกว่าหลายพันจุด
- หากชุดข้อมูลของคุณมีจุดข้อมูลมากกว่าหลายพันจุด คุณควรใช้เครื่องมือ **Subset Features** เพื่อสุ่มตัวอย่างข้อมูลบางส่วนและใช้ชุดข้อมูลย่อยนั้นในการสร้าง **semivariogram/covariance cloud**

การสร้าง Semivariogram Surface:

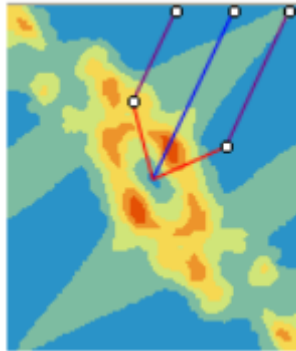
- ในตัวอย่างที่แสดงด้านล่างจะเป็น **Semivariogram Surface** ที่มีความสามารถในการควบคุมทิศทางการค้นหา
- ค่าใน **semivariogram cloud** จะถูกจัดกลุ่มในถัง (bins) ตามทิศทางและระยะทางระหว่างคู่ของตำแหน่ง
- ค่าเหล่านี้จะถูกเฉลี่ยและเรียงเรียงเพื่อสร้าง **semivariogram surface**
- ในแผนภูมิด้านล่าง, **legend** จะแสดงค่าในช่วงระหว่างการเปลี่ยนแปลงของสี
- คุณสามารถกำหนดขนาดของ **lag** ในเครื่องมือนี้ เพื่อควบคุมขนาดของถัง (bins) และจำนวนของถังจะถูกกำหนดโดยจำนวน **lags** ที่คุณระบุ
- ขอบเขตของ **semivariogram surface** จะถูกควบคุมโดยขนาดของ **lag** และจำนวน **lags**

เครื่องมือนี้ช่วยให้การศึกษา **spatial autocorrelation** ในชุดข้อมูลทำได้ง่ายขึ้น และสามารถใช้เพื่อหาค่าผิดปกติในพื้นที่เฉพาะ (local outliers) ได้



Semivariogram Surface with Search Direction

คุณสามารถดูข้อมูลย่อยของค่าใน **semivariogram cloud** โดยการเลือกช่องทำเครื่องหมาย **Show search direction** และคลิกที่ **direction controller** เพื่อปรับขนาดหรือเปลี่ยนทิศทางของมันได้



คุณสามารถเลือกชุดข้อมูลและแอตทริบิวต์โดยใช้ตัวควบคุมดังต่อไปนี้:

▼ Data Source

Layer: O3_Sep06_3pm Attribute: OZONE

ในกรณีนี้, **OZONE** คือ Field ใน Attribute ที่สนใจ (ความเข้มข้นของโอโซน) คุณยังสามารถคลิกที่ปุ่มลูกศรข้างๆ **Data Source** เพื่อซ่อนส่วนนี้ของเครื่องมือชั่วคราวได้

Binning empirical semivariograms

เมื่อสร้าง **empirical semivariogram** (หรือ semivariogram ที่ได้จากข้อมูลจริง) จากชุดของตำแหน่ง (locations) และ **semivariogram cloud** จะพบว่า การวาดกราฟแต่ละคู่ตำแหน่งนั้นมีจำนวนจุดมากจนทำให้กราฟอัดแน่นและไม่สามารถตีความได้มากนัก ดังนั้น เพื่อลดจำนวนจุดใน empirical semivariogram จะมีการจัดกลุ่มคู่ตำแหน่งตามระยะห่างระหว่างกัน การจัดกลุ่มนี้เรียกว่า binning การทำ binning ประกอบไปด้วย 2 ขั้นตอนหลัก คือ:

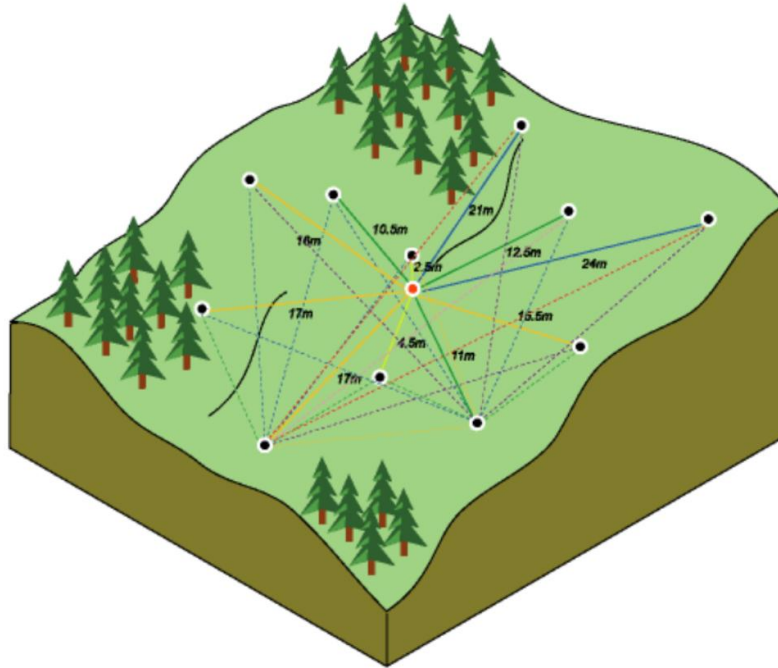
Stage one

ในขั้นตอนแรกของการทำ binning สำหรับ empirical semivariogram จะมีการทำงานในสองขั้นตอนหลัก คือ:

1. **การจับคู่ตำแหน่ง (Forming Pairs of Points):**
 - ในขั้นตอนนี้จะมีการจับคู่ตำแหน่งของจุดที่อยู่ในชุดข้อมูลทั้งหมด โดยการจับคู่จะพิจารณาจาก **ระยะทาง** และ **ทิศทาง** ที่อยู่ระหว่างจุดแต่ละคู่
 - ตัวอย่างเช่น ถ้าคุณมี **12** ตำแหน่ง ในพื้นที่ และเลือกจุดสีแดงหนึ่งจุดเป็นจุดเริ่มต้น, ทุกๆ จุดที่เหลือจะถูกจับคู่กับจุดสีแดงนั้น
2. **การจัดกลุ่มตามระยะห่างและทิศทาง (Grouping Pairs by Distance and Direction):**
 - เมื่อจับคู่จุดต่างๆ ได้แล้ว, จุดแต่ละคู่จะถูกจัดกลุ่มตามระยะห่างระหว่างจุดในคู่และทิศทางของการจับคู่

- กลุ่มที่มีระยะห่างใกล้เคียงกันจะถูกจัดให้อยู่ใน **bin** เดียวกัน
- สีที่เหมือนกันระหว่างการเชื่อมโยงของคู่ (link) แสดงถึงระยะห่างที่คล้ายกันใน **bin** นั้นๆ

การจับคู่และการจัดกลุ่มนี้ช่วยให้สามารถลดจำนวนข้อมูลที่ต้องนำเสนอใน **semivariogram cloud** ซึ่งจะทำให้การตีความข้อมูลสะดวกและชัดเจนยิ่งขึ้น



กระบวนการนี้จะดำเนินต่อไปสำหรับคู่ข้อมูลทั้งหมด โดยจะเห็นว่าในกระบวนการจับคู่จำนวนคู่ข้อมูลจะเพิ่มขึ้นอย่างรวดเร็วเมื่อมีการเพิ่มตำแหน่งใหม่เข้าไป นี่คือเหตุผลที่สำหรับแต่ละ **bin** จะมีการแสดงค่าเฉลี่ยของระยะห่างและ **semivariance** (ค่าความแปรปรวนครึ่งหนึ่ง) สำหรับคู่ข้อมูลทั้งหมดใน **bin** นั้นๆ โดยจะแสดงเป็นจุดเดี่ยวบน **semivariogram**

Stage two

ในขั้นตอนที่สองของกระบวนการ **binning** การจัดกลุ่มคู่ข้อมูลจะทำตาม ระยะห่าง และทิศทางร่วม:

1. การจัดกลุ่มตามระยะห่างและทิศทางร่วม:
 - จินตนาการถึงกราฟที่ทุกจุดมี จุดเริ่มต้นร่วม ซึ่งลักษณะนี้ทำให้ **empirical semivariogram** เป็นแบบสมมาตร (symmetric)
 - ในการทำ binning นี้ การจับคู่จะไม่เพียงแค่อภิปรายแค่ระยะห่างเท่านั้น แต่ยังรวมถึง ทิศทาง ของคู่ข้อมูลด้วย
2. การคำนวณและเฉลี่ยค่า:
 - สำหรับแต่ละ **bin**, ค่าความแตกต่างยกกำลังสอง (squared difference) จะถูกคำนวณจาก ค่าของตำแหน่งคู่ข้อมูลที่เชื่อมโยง

- หลังจากนั้นจะทำการเฉลี่ยค่าความแตกต่างยกกำลังสองทั้งหมดใน **bin** นั้นๆ และคูณด้วย 0.5 เพื่อให้ได้ค่า **empirical semivariogram** สำหรับแต่ละ **bin**

3. การแสดงผลและการควบคุม:

- ใน **Geostatistical Analyst**, คุณสามารถควบคุมขนาดของ **lag** (การแบ่งช่วงระยะทาง) และจำนวนของ **lags** (จำนวนช่วงระยะทางที่ใช้ในการแบ่ง)
- ค่าของ **empirical semivariogram** ที่คำนวณได้ในแต่ละ **bin** จะถูกแสดงเป็นจุดที่มีการใช้ สี แสดงค่าเหล่านี้ และจะเรียกว่า **semivariogram surface**

กระบวนการนี้ช่วยให้คุณสังเกตลักษณะของความสัมพันธ์ระหว่างตำแหน่งต่างๆ ในพื้นที่ และเข้าใจถึงการกระจายของข้อมูลในเชิงพื้นที่ได้ดีขึ้น

Creating empirical semivariograms

การสร้าง **empirical semivariogram** จะเริ่มต้นด้วยการคำนวณ ความแตกต่างยกกำลังสอง (**squared difference**) ระหว่างค่าของตำแหน่งที่แตกต่างกันทุกคู่:

1. การคำนวณความแตกต่างยกกำลังสอง:

- สำหรับทุกคู่ของตำแหน่งในชุดข้อมูล (สมมติว่าแต่ละคู่ประกอบด้วยตำแหน่งสองจุด), คำนวณ ความแตกต่างยกกำลังสอง ของค่าที่วัดได้ในตำแหน่งเหล่านั้น

2. การสร้าง Semivariogram Cloud:

- เมื่อค่าความแตกต่างยกกำลังสองถูกคำนวณแล้ว, จะมีการพล็อตกราฟโดยที่:
 - แกน **y** แสดงค่าความแตกต่างยกกำลังสอง (half squared difference)
 - แกน **x** แสดงระยะห่างระหว่างตำแหน่ง (distance separating the locations)
- กราฟนี้เรียกว่า **semivariogram cloud**

3. การสังเกต Spatial Autocorrelation:

- จุดมุ่งหมายหลักของการศึกษา **variography** คือการ **สำรวจและวัดการพึ่งพิงเชิงพื้นที่ (spatial dependence)** หรือที่เรียกว่า **spatial autocorrelation** ซึ่งแสดงถึงความสัมพันธ์ระหว่างค่าของตำแหน่งที่อยู่ใกล้กัน
 - **Spatial autocorrelation** หมายความว่า สิ่งที่อยู่ใกล้กันจะมีลักษณะที่คล้ายคลึงกันมากกว่าสิ่งที่อยู่ไกลกัน
 - ดังนั้น, คู่ของตำแหน่งที่ใกล้กัน (อยู่ทางซ้ายสุดของแกน x ใน semivariogram cloud) จะมีค่าที่คล้ายคลึงกัน (และจึงมีค่าความแตกต่างยกกำลังสองต่ำบนแกน y)
 - คู่ตำแหน่งที่อยู่ห่างออกไป (ทางขวาของแกน x ใน semivariogram cloud) ค่าความแตกต่างยกกำลังสองจะสูงขึ้น (แสดงการแตกต่างกันมากขึ้น)

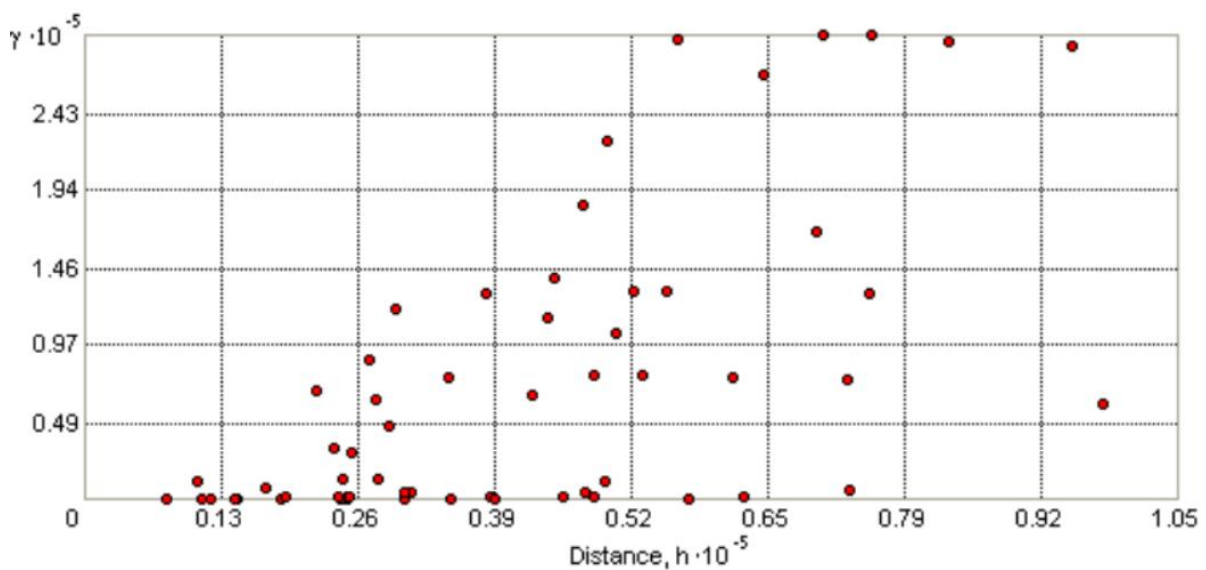
4. ข้อจำกัดทางคอมพิวเตอร์:

- เนื่องจากข้อจำกัดในด้านเวลาในการคำนวณและหน่วยความจำ หากชุดข้อมูลมีมากกว่า **5000** ข้อมูล (observations), **Geostatistical Analyst** จะสุ่มเลือก **5000** ข้อมูล จากชุดข้อมูลทั้งหมดสำหรับการวิเคราะห์เชิงโครงสร้าง

(structural analysis) และการปรับพอดี้ **semivariogram model** (โมเดลสำหรับการหาค่าความแตกต่าง)

- จะช่วยลดจำนวนข้อมูลที่จะถูกนำมาพล็อตใน semivariogram
- การสุ่มเลือกข้อมูลนี้ไม่ส่งผลกระทบต่อ **semivariogram model** ที่ได้ เนื่องจากข้อมูลทั้งหมดจะถูกใช้ในการสร้างค่าพยากรณ์
- อย่างไรก็ตาม, หากชุดข้อมูลมีค่าที่ใหญ่กว่าปกติ (outliers), ค่าดังกล่าวอาจจะไม่ได้ถูกเลือกในกลุ่มข้อมูลสุ่มที่ใช้สร้าง **empirical semivariogram**, ซึ่งอาจทำให้ **semivariogram model** ที่ได้แตกต่างจากการที่ใช้ข้อมูลทั้งหมด

สรุป: การสร้าง **empirical semivariogram** ช่วยให้เราสามารถสำรวจความสัมพันธ์เชิงพื้นที่และการพึ่งพิงระหว่างจุดข้อมูล โดยการพล็อต **semivariogram cloud** ที่แสดงความแตกต่างยกกำลังสองของข้อมูลตามระยะห่างระหว่างตำแหน่งต่างๆ



Choosing a lag size

การเลือก **lag size** (ขนาดของช่วงการแบ่งระยะ) มีผลสำคัญต่อ **empirical semivariogram** และการวิเคราะห์ **spatial autocorrelation**:

1. ผลกระทบจากการเลือกขนาดของ **lag**:
 - หาก **lag size** ใหญ่เกินไป: อาจทำให้ **short-range autocorrelation** หรือความสัมพันธ์ระยะสั้นถูกกลบไป
 - หาก **lag size** เล็กเกินไป: อาจทำให้เกิด **bins** ที่ว่างเปล่า (ไม่มีข้อมูล) หรือ **sample sizes** ที่น้อยเกินไปในบาง bin ซึ่งอาจทำให้ค่าเฉลี่ยใน bin นั้นไม่สามารถแสดงข้อมูลที่เป็นตัวแทนได้
2. การเลือกขนาดของ **lag** เมื่อข้อมูลมีการสุ่มไม่เป็นระเบียบ:
 - หากข้อมูลที่คุณใช้มาจากการสุ่มแบบไม่เป็นระเบียบหรือไม่ตรงกับตารางการสำรวจ (sampling grid), การเลือกขนาดของ **lag** จะไม่ง่ายเหมือนกับกรณีที่ข้อมูลมาจากการสำรวจที่มีการจัดตาราง
3. การปรับขนาดของ **lag**:

- หาก **range** ของ **semivariogram model** ที่ได้จากการพอดี (fitting) มีค่าค่อนข้างเล็กเมื่อเทียบกับการกระจายของ **empirical semivariogram**, คุณสามารถลดขนาดของ lag
- หาก **range** ของ **semivariogram model** ที่ได้จากการพอดี (fitting) มีค่าค่อนข้างใหญ่เมื่อเทียบกับการกระจายของ **empirical semivariogram**, คุณสามารถเพิ่มขนาดของ lag

สรุป: การเลือกขนาดของ **lag** ที่เหมาะสมมีผลโดยตรงต่อการสร้าง **empirical semivariogram** และการวิเคราะห์ **spatial autocorrelation** โดยการเลือกขนาดของ lag ที่เหมาะสมจะช่วยให้คุณสังเกตเห็นความสัมพันธ์ระหว่างจุดข้อมูลในระยะต่างๆ ได้ดีขึ้น