

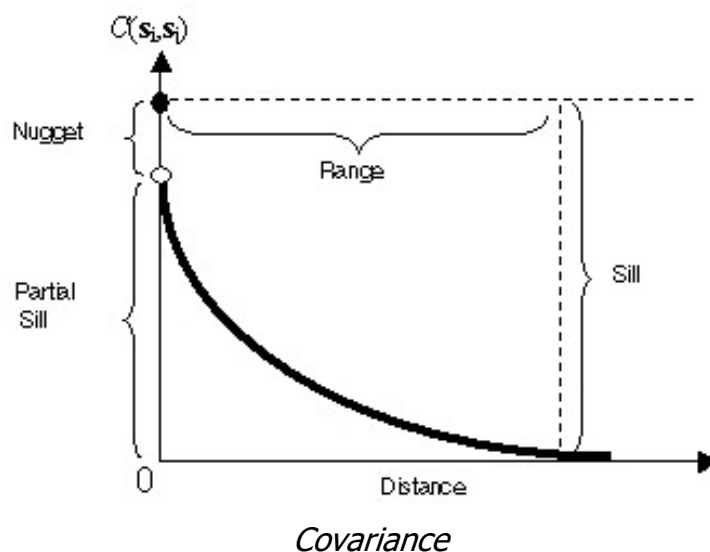
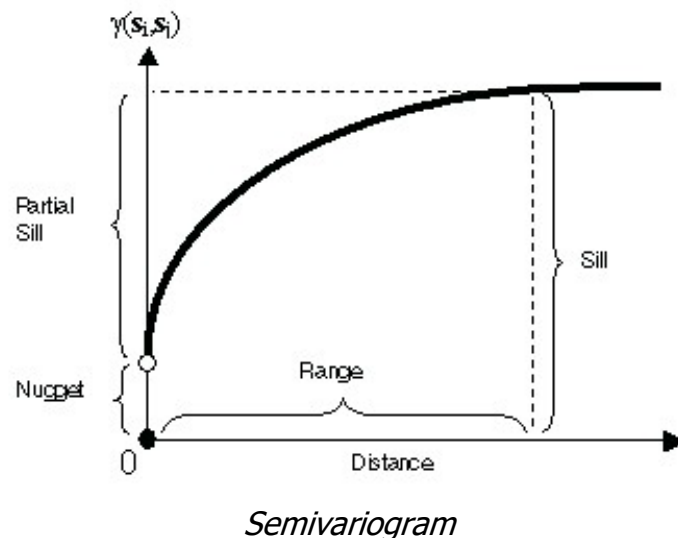
Lab7: Variogram

1. Semivariogram and Covariance Functions

เซมิวาริโอแกรมและฟังก์ชันโคเวเรียนซ์ใช้ในการหาปริมาณสมมติฐานที่ว่าข้อมูลที่อยู่ใกล้กันมักจะมีค่าคล้ายคลึงกันมากกว่าข้อมูลที่อยู่ห่างกัน เซมิวาริโอแกรมและโคเวเรียนซ์ต่างก็วัดความเชื่อมั่นของความสัมพันธ์ทางสถิติ (Statistical Correlation) โดยใช้ระยะทางเป็นตัวแปร

กระบวนการสร้างแบบจำลองเซมิวาริโอแกรมและฟังก์ชันโคเวเรียนซ์

การสร้างแบบจำลองเซมิวาริโอแกรมและฟังก์ชันโคเวเรียนซ์คือการปรับเส้นโค้งเซมิวาริโอแกรมหรือโคเวเรียนซ์ให้เข้ากับข้อมูลเชิงประจักษ์ (Empirical Data) เป้าหมายคือการหาความพอดีที่ดีที่สุด (Best Fit) และรวมเอาความรู้เกี่ยวกับปรากฏการณ์นั้นๆ ลงในแบบจำลองด้วย โดยแบบจำลองนี้จะถูกนำไปใช้ในการทำนาย (Predictions) ในภายหลังต่อไป



องค์ประกอบสำคัญของเซมิวาริโอแกรม:

- **Range (ระยะอิทธิพล):** คือระยะทางที่เซมิวาริโอแกรมไปถึงระดับ **Sill** และคงที่ ไม่เพิ่มขึ้นอีก ซึ่งแสดงถึงขอบเขตที่ความสัมพันธ์เชิงพื้นที่ (Spatial Correlation) ระหว่างจุดข้อมูลเริ่มหมดไป
- **Sill:** คือค่าความสูงที่สุดที่เส้นเซมิวาริโอแกรมไปถึงเมื่อเริ่มคงที่ (หรือไม่เพิ่มขึ้นอีก) โดยทั่วไปแล้ว ซิลล์จะประกอบด้วยสองส่วนหลัก ได้แก่:
 1. **Nugget Effect (เอฟเฟกต์นิกเก็ต):** ความไม่ต่อเนื่องที่เกิดขึ้นที่จุดกำเนิด (Origin)
 2. **Partial Sill (พาร์เชียลซิลล์):** ส่วนของความแปรปรวนที่เกิดจากโครงสร้างของข้อมูลจริง

เมื่อรวม **Nugget Effect** และ **Partial Sill** เข้าด้วยกัน จะได้ค่าของ **Sill**

- **Nugget Effect:** สามารถแบ่งย่อยได้เป็นสองส่วนคือ:
 - **Measurement Error (ข้อผิดพลาดในการวัด):** ความแปรปรวนที่เกิดจากความไม่แม่นยำในการเก็บรวบรวมข้อมูล
 - **Microscale Variation (ความแปรปรวนในระดับจุลภาค):** ความแตกต่างที่เกิดจากความแปรปรวนในรายละเอียดเล็ก ๆ ของปรากฏการณ์ที่ไม่สามารถตรวจจับได้ด้วยมาตรการที่ใช้

เอฟเฟกต์นิกเก็ตเป็นเพียงผลรวมของ **Measurement Error** และ **Microscale Variation** โดยที่องค์ประกอบใดองค์ประกอบหนึ่งอาจมีค่าเป็นศูนย์ได้ ดังนั้นเอฟเฟกต์นิกเก็ตอาจประกอบด้วยเพียงส่วนใดส่วนหนึ่งเท่านั้น

เซมิวาริโอแกรมและโคเวเรียนซ์ไม่สามารถเป็นฟังก์ชันใด ๆ ได้อย่างอิสระ เพื่อให้การพยากรณ์มีความคลาดเคลื่อนมาตรฐานแบบคริกกิง (Kriging Standard Errors) ที่ไม่เป็นค่าลบ (Nonnegative) จะต้องเลือกใช้ฟังก์ชันเฉพาะที่เหมาะสมเท่านั้น

การสร้างแบบจำลองแบบผสม (Composite Models):

- นอกจากจะเลือกแบบจำลองเดี่ยวแล้ว คุณยังสามารถสร้างแบบจำลองโดยการรวมแบบจำลองหลาย ๆ แบบเข้าด้วยกันได้ การรวมกันเช่นนี้จะยังคงให้ผลลัพธ์ที่ถูกต้องตามหลักการทางสถิติ
- ใน **Geostatistical Analyst** คุณสามารถรวมแบบจำลองได้สูงสุดถึง 4 แบบ

ข้อยกเว้นสำคัญ:

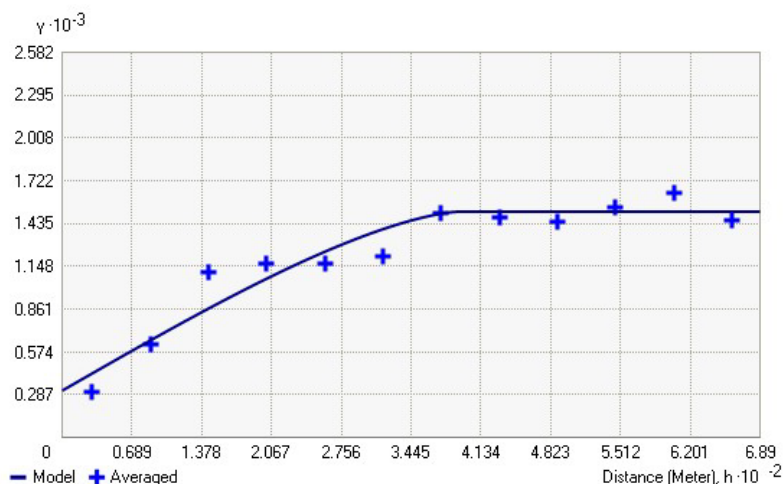
- มีบางกรณีที่เซมิวาริโอแกรมสามารถมีอยู่ได้โดยที่ไม่มีฟังก์ชันโคเวเรียนซ์ที่สอดคล้องกัน ตัวอย่างเช่น:
 - **เซมิวาริโอแกรมเชิงเส้น (Linear Semivariogram):** มีลักษณะเส้นตรง และไม่มีค่า Sill ทำให้ไม่มีฟังก์ชันโคเวเรียนซ์ที่สอดคล้องกัน
- เนื่องจากข้อจำกัดนี้ **Geostatistical Analyst** จะใช้เฉพาะแบบจำลองที่มีค่า Sill เท่านั้น

การเลือกแบบจำลองเซมิวาริโอแกรมที่ "ดีที่สุด":

- ไม่มีหลักเกณฑ์ที่ตายตัวในการเลือกแบบจำลองเซมิวาริโอแกรมที่ดีที่สุด
- คุณสามารถพิจารณาจากเซมิวาริโอแกรมเชิงประจักษ์ (Empirical Semivariogram) หรือฟังก์ชันโคเวเรียนซ์ของคุณ เพื่อเลือกแบบจำลองที่เหมาะสม
- นอกจากนี้ คุณสามารถใช้วิธีการ **Validation (การตรวจสอบความถูกต้อง)** และ **Cross-validation (การตรวจสอบข้ามชุดข้อมูล)** เพื่อช่วยในการตัดสินใจได้อีกด้วย

The range, sill, and nugget

เซมิวาริโอแกรม (Semivariogram) แสดงถึงความสัมพันธ์เชิงพื้นที่แบบอัตโนมัติ (Spatial Autocorrelation) ของจุดตัวอย่างที่ถูกวัด เมื่อมีการพล็อตค่าของแต่ละคู่ตำแหน่ง (เรียกว่า **Binning the Empirical Semivariogram**) จะมีการปรับเส้นแบบจำลองให้สอดคล้องกับข้อมูลเชิงประจักษ์ (เรียกว่า **Fitting a Model to the Empirical Semivariogram**)



คุณลักษณะสำคัญของแบบจำลองเซมิวาริโอแกรม:

1. Range (3.8)

- เมื่อคุณดูแบบจำลองของเซมิวาริโอแกรม จะสังเกตได้ว่าเส้นแบบจำลองจะเริ่มคงที่เมื่อถึงระยะทางหนึ่ง
- ระยะทางที่แบบจำลองเริ่มแบนราบ (ไม่เพิ่มขึ้นอีก) นี้เรียกว่า **Range**
- ตำแหน่งตัวอย่างที่อยู่ใกล้กันในระยะที่น้อยกว่า Range จะมีความสัมพันธ์เชิงพื้นที่แบบอัตโนมัติ (Spatial Autocorrelation)
- ในทางตรงกันข้าม ตำแหน่งที่อยู่ห่างกันเกินกว่า Range จะไม่มีความสัมพันธ์เชิงพื้นที่

2. Sill (1.5)

- ค่าที่แบบจำลองเซมิวาริโอแกรมไปถึงที่ระดับ Range (ค่าบนแกน Y) เรียกว่า **Sill**
- **Partial Sill (พาร์เชียลซิลล์)** คือค่าของ Sill ลบด้วยค่าของ **Nugget** (ค่าความไม่ต่อเนื่องที่จุดกำเนิด)

3. Nugget (0.3)

ตามทฤษฎีแล้ว ที่ระยะห่างเป็นศูนย์ ($Lag = 0$) ค่าเซมิวาริโอแกรม (Semivariogram) ควรจะเท่ากับ 0 อย่างไรก็ตาม ในความเป็นจริง เมื่อระยะห่างระหว่างจุดมีค่าน้อยมาก (ใกล้ศูนย์มาก ๆ) เซมิวาริโอแกรมมักจะแสดงค่าที่มากกว่า 0 ซึ่งเรียกปรากฏการณ์นี้ว่า **Nugget Effect** (เอฟเฟกต์นิกเก็ต)

สาเหตุของ Nugget Effect:

1. ข้อผิดพลาดในการวัด (**Measurement Errors**):
 - เกิดจากความไม่แม่นยำของอุปกรณ์ที่ใช้ในการวัดข้อมูล ซึ่งเป็นข้อผิดพลาดที่หลีกเลี่ยงไม่ได้ในกระบวนการเก็บข้อมูลจริง
2. ความแปรปรวนเชิงพื้นที่ในระดับจุลภาค (**Microscale Variation**):
 - ธรรมชาติของปรากฏการณ์บางอย่างอาจมีการเปลี่ยนแปลงในระดับเล็กมาก ๆ ที่เล็กกว่าระยะห่างระหว่างจุดตัวอย่าง (Sampling Interval)
 - การแปรปรวนเหล่านี้ไม่สามารถตรวจจับได้ด้วยระยะการเก็บตัวอย่างที่ใช้ ทำให้ปรากฏเป็นส่วนหนึ่งของ Nugget
3. ผลรวมของทั้งสองปัจจัย:
 - ในบางกรณี เอฟเฟกต์นิกเก็ตเกิดอาจเกิดจากทั้งข้อผิดพลาดในการวัดและความแปรปรวนในระดับจุลภาครวมกัน

2. Fitting a model to the empirical semivariogram

การสร้างแบบจำลอง Semivariogram/Covariance

การสร้างแบบจำลอง **Semivariogram/Covariance** เป็นขั้นตอนสำคัญที่เชื่อมโยงระหว่าง การอธิบายเชิงพื้นที่ (**Spatial Description**) และ การพยากรณ์เชิงพื้นที่ (**Spatial Prediction**) โดยการประยุกต์ใช้หลักการทาง **Geostatistics** ที่สำคัญที่สุดคือการพยากรณ์ค่าข้อมูลในตำแหน่งที่ไม่ได้เก็บตัวอย่าง (Unsampled Locations)

บทบาทของ Empirical Semivariogram และ Covariance:

- Empirical Semivariogram/Covariance: แสดงรูปแบบความสัมพันธ์เชิงพื้นที่ที่สังเกตได้จริง
- Empirical Semivariogram และ Covariance ช่วยให้ข้อมูลเกี่ยวกับความสัมพันธ์เชิงพื้นที่แบบอัตโนมัติ (Spatial Autocorrelation) ของชุดข้อมูล
- อย่างไรก็ตาม พวกมันไม่สามารถให้ข้อมูลสำหรับทุกทิศทางและทุกระยะทางได้อย่างครบถ้วน

ความจำเป็นในการสร้างแบบจำลอง (Model Fitting):

- เพื่อให้สามารถทำการพยากรณ์ด้วยวิธี **Kriging** ได้อย่างถูกต้อง และเพื่อให้ได้ **ค่าความแปรปรวนของ Kriging (Kriging Variance)** ที่เป็นบวก จำเป็นต้องปรับแบบจำลอง (Fit a Model) กับ **Empirical Semivariogram/Covariance**
- การสร้างแบบจำลองในที่นี้หมายถึงการปรับฟังก์ชันต่อเนื่อง (Continuous Function) หรือเส้นโค้ง (Curve) ให้สอดคล้องกับข้อมูลเชิงประจักษ์
- **Model Fitting:** การปรับฟังก์ชันเชิงทฤษฎีเพื่อให้สามารถนำไปใช้ในการพยากรณ์ได้อย่างแม่นยำ

มุมมองที่แตกต่างกันของค่า Empirical Semivariogram/Covariance

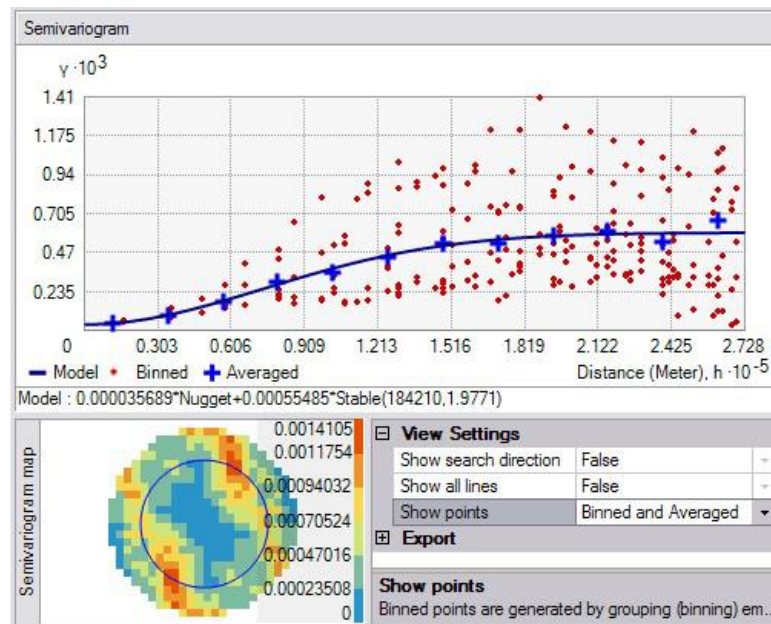
ใน **Geostatistical Wizard** มีมุมมอง (Views) ที่แตกต่างกัน 3 แบบสำหรับแสดงค่าของ **Empirical Semivariogram/Covariance** ซึ่งคุณสามารถเลือกใช้มุมมองแบบใดแบบหนึ่ง สองแบบ หรือใช้ทั้งสามแบบร่วมกันเพื่อช่วยในการปรับแบบจำลอง (Model Fitting) ให้เหมาะสมกับข้อมูลของคุณ

มุมมองทั้ง 3 แบบ ได้แก่:

1. **มุมมองเริ่มต้น (Default View):**
 - จะแสดงค่า **Empirical Semivariogram/Covariance** ที่ถูกจัดกลุ่ม (Binned) และเฉลี่ย (Averaged) แล้ว
 - ช่วยให้มองเห็นแนวโน้มโดยรวมของข้อมูลได้อย่างชัดเจน ทำให้ง่ายต่อการปรับแบบจำลอง
2. **มุมมองเสริม (Optional Views):**
 - อาจรวมถึงมุมมองที่แสดงข้อมูลในรูปแบบจุดดิบ (Raw Data Points) เพื่อให้เห็นรายละเอียดของแต่ละจุดตัวอย่าง
 - อาจมีมุมมองที่เน้นการแสดงความแตกต่างเชิงทิศทาง (Directional Variograms) เพื่อวิเคราะห์ความแปรปรวนในแต่ละทิศทางเฉพาะ
3. **มุมมองแบบผสม (Combined Views):**
 - สามารถเปิดดูหลายมุมมองพร้อมกัน เพื่อเปรียบเทียบและวิเคราะห์ได้อย่างครอบคลุม
 - วิธีนี้ช่วยเพิ่มความแม่นยำในการเลือกแบบจำลองที่สอดคล้องกับข้อมูลจริงมากที่สุด

ประโยชน์ของการใช้มุมมองที่แตกต่างกัน:

- ช่วยในการระบุรูปแบบของความสัมพันธ์เชิงพื้นที่ (Spatial Patterns) ได้ชัดเจนยิ่งขึ้น
- ช่วยในการปรับแต่งพารามิเตอร์ของแบบจำลอง เช่น **Range, Sill, Nugget** ได้อย่างมีประสิทธิภาพ
- เพิ่มความมั่นใจในการเลือกแบบจำลองที่เหมาะสมสำหรับการพยากรณ์ (Kriging Prediction)



Binned Values, Averaged Values:

- **Binned Values:**
 - แสดงเป็น จุดสีแดง (Red Dots)
 - สร้างขึ้นโดยการจัดกลุ่ม (Binning) จุดของ **Empirical Semivariogram/Covariance** เข้าด้วยกันใน เซลล์สี่เหลี่ยม (Square Cells) ที่มีความกว้างเท่ากับระยะห่างหนึ่งช่วง (One Lag Wide)
 - แสดงถึง ความแปรปรวนในระดับท้องถิ่น (Local Variation) ของค่า Semivariogram/Covariance
- **Averaged Values:**
 - แสดงเป็น เครื่องหมายกากบาทสีน้ำเงิน (Blue Crosses)
 - สร้างขึ้นโดยการจัดกลุ่ม (Binning) จุดของ **Empirical Semivariogram/Covariance** ที่อยู่ภายใน เชกเตอร์เชิงมุม (Angular Sectors)
 - แสดงถึงการเปลี่ยนแปลงของค่า Semivariogram/Covariance ที่ ราบรื่น (Smooth) มากกว่าเมื่อเทียบกับค่าที่ถูกจัดกลุ่ม

ข้อแตกต่างที่สำคัญ:

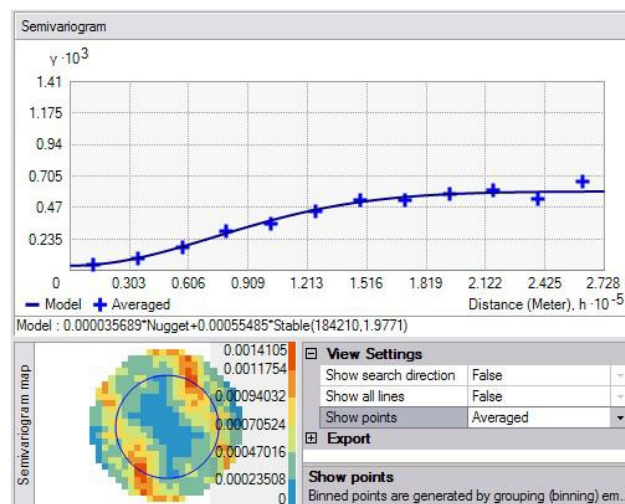
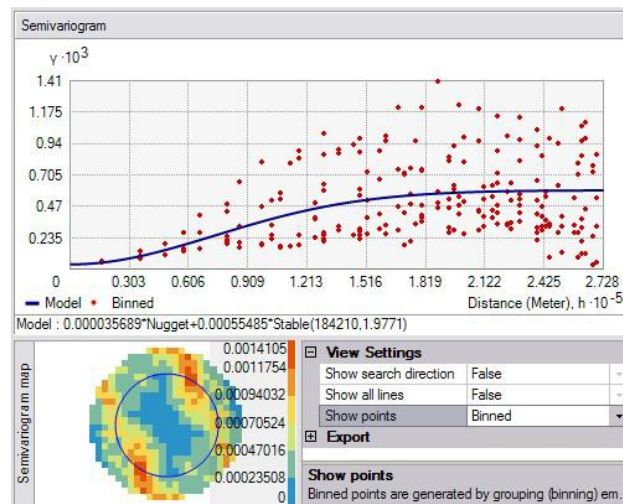
- **Binned Points (จุดสีแดง):** แสดงความแปรปรวนของข้อมูลในแต่ละจุดอย่างละเอียด เหมาะสำหรับการดูการเปลี่ยนแปลงเฉพาะจุด
- **Averaged Points (กากบาทสีน้ำเงิน):** ช่วยให้มองเห็นแนวโน้มโดยรวมได้ชัดเจนขึ้น เพราะการเปลี่ยนแปลงจะราบรื่นกว่า

Model Fitting:

- ในหลายกรณี การปรับแบบจำลองให้สอดคล้องกับ ค่าเฉลี่ย (Averaged Values) จะทำได้ง่ายกว่า เพราะข้อมูลไม่รกเกินไป (Less Cluttered) และแสดงถึงการเปลี่ยนแปลงที่ราบรื่น ทำให้เห็นแนวโน้มของ **Spatial Autocorrelation** ได้ชัดเจนขึ้น

Show Points Control:

1. **Binned and Averaged:** แสดงทั้งจุดสีแดงและกากบาทสีน้ำเงินพร้อมกัน
2. **Binned:** แสดงเฉพาะจุดสีแดง
3. **Averaged:** แสดงเฉพาะกากบาทสีน้ำเงิน

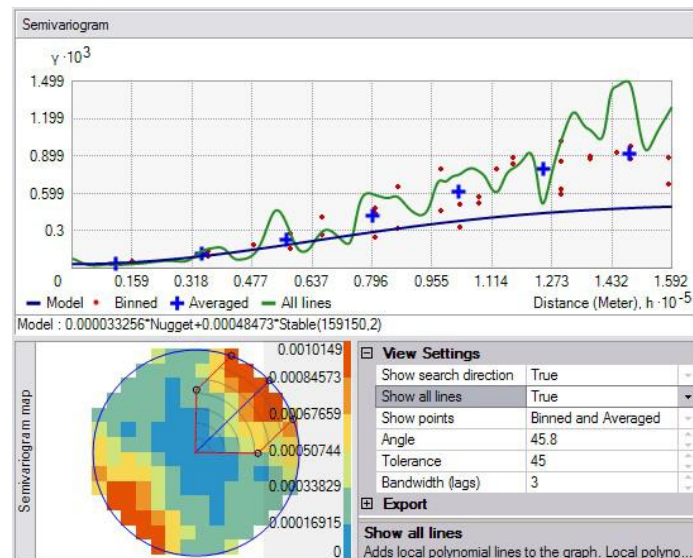


Green Lines:

- คุณสามารถเพิ่ม **เส้นสีเขียว (Green Lines)** ลงในกราฟได้โดยการเลือกตัวเลือก **"Show all lines"**
- เส้นสีเขียวเหล่านี้เป็น **พหุนามเชิงเส้น (Local Polynomials)** ที่ถูกปรับให้พอดีกับค่าของ **Empirical Semivariogram/Covariance** ที่ถูกจัดกลุ่ม (Binned Values)

Show Search Direction:

- หากตั้งค่า **"Show search direction"** เป็น **True** ระบบจะแสดงเฉพาะเส้นพหุนามเชิงเส้นที่ถูกปรับให้พอดีกับผิวของ **Empirical Semivariogram/Covariance** ในแนวแกนกลาง (Central Axis Transect) ของเครื่องมือ **"Show search direction"**
- ซึ่งช่วยในการวิเคราะห์ความสัมพันธ์เชิงพื้นที่ในทิศทางที่สนใจได้อย่างเฉพาะเจาะจง



หลักเกณฑ์ในการปรับแบบจำลอง Semivariogram/Covariance

เมื่อคุณปรับแบบจำลอง **Semivariogram/Covariance** ให้เข้ากับข้อมูลเชิงประจักษ์ (Empirical Data) ควรพิจารณาให้แบบจำลองมีลักษณะดังนี้:

1. ผ่านจุดกึ่งกลางของกลุ่มค่าที่ถูกจัดกลุ่ม (**Binned Values** - จุดสีแดง)
2. ผ่านใกล้เคียงที่สุดกับค่าที่ถูกเฉลี่ย (**Averaged Values** - กากบาทสีน้ำเงิน)
3. ผ่านใกล้เคียงที่สุดกับจุดกึ่งกลางของเส้นพหุนามเชิงเส้น (**Local Polynomials** - เส้นสีเขียว)

ข้อควรคำนึง:

- ความรู้ของคุณเกี่ยวกับปรากฏการณ์ที่วิเคราะห์อาจกำหนด **รูปแบบของแบบจำลอง (Model Shape)** รวมถึงค่าของ **Nugget, Range, Partial Sill** และ **Anisotropy (ความไม่สมมาตรของพื้นที่)**
- แม้ว่าแบบจำลองจะดูไม่พอดีกับข้อมูลเชิงประจักษ์ทั้งหมด แต่อย่าลืมว่าข้อมูลเหล่านี้เป็นเพียง **ตัวอย่าง (Sample)** ของปรากฏการณ์จริง ซึ่งอาจไม่สะท้อนทุกแง่มุมของความสัมพันธ์เชิงพื้นที่และเชิงสถิติได้อย่างครบถ้วน

ประเภทของแบบจำลอง Semivariogram/Covariance

โปรแกรม ArcGIS ชุดคำสั่ง **Geostatistical Analyst** มีฟังก์ชันสำหรับสร้างแบบจำลอง Semivariogram/ Covariance ให้เลือกหลายแบบ ได้แก่:

1. **Circular:** เหมาะสำหรับปรากฏการณ์ที่มีรูปแบบการกระจายเป็นวงกลม
2. **Spherical:** แบบจำลองยอดนิยม ใช้สำหรับข้อมูลที่แปรปรวนและมีขอบเขตชัดเจน
3. **Tetraspherical:** รุ่นที่ละเอียดกว่า Spherical ช่วยเพิ่มความยืดหยุ่นในการปรับแบบจำลอง
4. **Pentaspherical:** แบบจำลองที่ซับซ้อนมากขึ้นสำหรับการปรับเข้ากับข้อมูลที่มีรายละเอียดสูง
5. **Exponential:** ใช้สำหรับข้อมูลที่มีความสัมพันธ์ลดลงอย่างรวดเร็วเมื่อระยะทางเพิ่มขึ้น
6. **Gaussian:** เหมาะกับข้อมูลที่มีการเปลี่ยนแปลงแบบราบรื่น (Smooth)

7. **Rational Quadratic:** ให้ความยืดหยุ่นสูงในการปรับรูปแบบของความแปรปรวน
8. **Hole Effect:** ใช้กับข้อมูลที่มีความแปรปรวนแบบเป็นจังหวะ (Periodic Variations)
9. **K-Bessel:** สำหรับการวิเคราะห์ข้อมูลที่มีรูปแบบซับซ้อนทางสถิติ
10. **J-Bessel:** คล้ายกับ K-Bessel แต่เน้นในแง่ของฟังก์ชันพิเศษทางคณิตศาสตร์
11. **Stable:** ใช้สำหรับข้อมูลที่มีพฤติกรรมไม่เสถียรหรือมีการกระจายที่ไม่เป็นปกติ

การแสดงตัวอย่างของแบบจำลองที่แตกต่างกัน

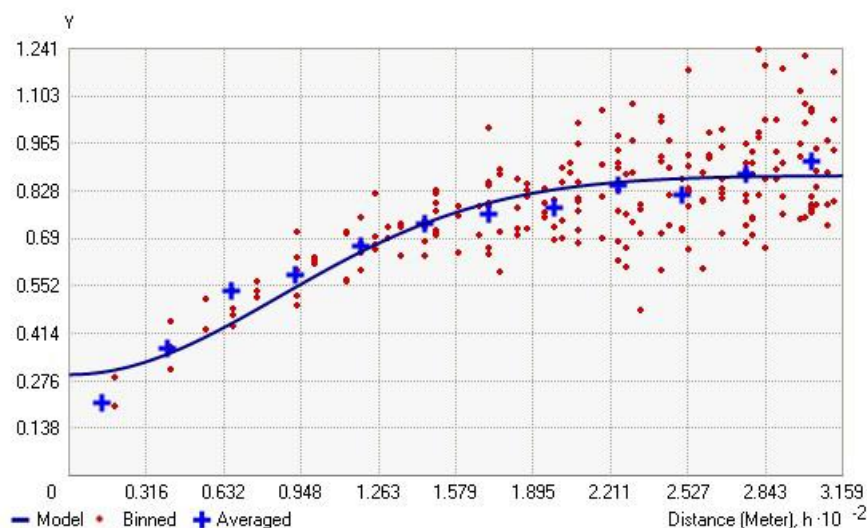
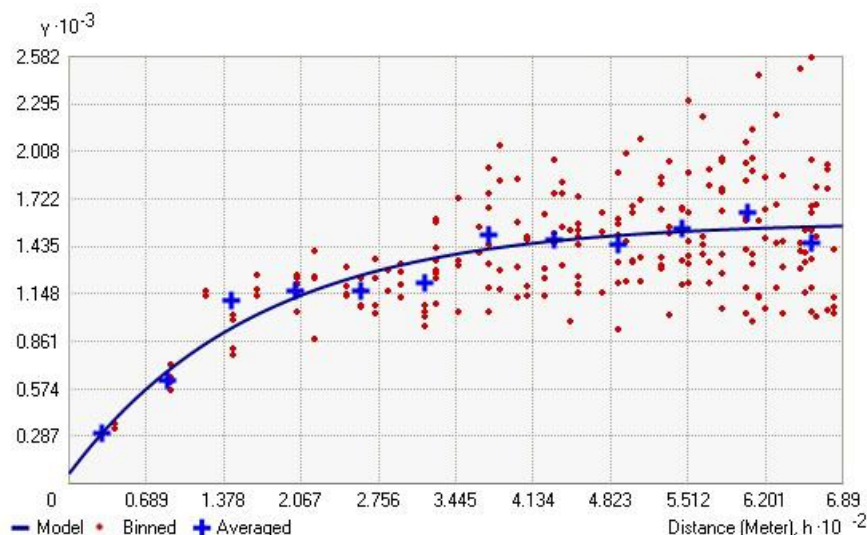
กราฟด้านล่างแสดง สองแบบจำลองที่พบบ่อย และแสดงถึงความแตกต่างของฟังก์ชัน:

1. Exponential Semivariogram

- เส้นโค้งที่ลดลงอย่างรวดเร็วเมื่อระยะทางเพิ่มขึ้น
- เหมาะสำหรับข้อมูลที่มีความสัมพันธ์ลดลงอย่างรวดเร็วตามระยะทาง

2. Gaussian Semivariogram

- เส้นโค้งที่ลดลงอย่างราบรื่นและค่อนข้างคงที่
- เหมาะสำหรับข้อมูลที่แปรปรวนอย่างต่อเนื่องหรือมีความสัมพันธ์ที่ราบรื่น



3. Search neighborhoods

การใช้ Search Neighborhood ในการพยากรณ์ค่า

เมื่อจุดต่างๆ ห่างออกจากตำแหน่งที่ต้องการพยากรณ์ (Prediction Location) มากขึ้น ค่าในการวัดจะมีความสัมพันธ์เชิงพื้นที่ (Spatial Autocorrelation) ที่น้อยลงหรืออาจจะไม่มีเลยกับตำแหน่งที่ต้องการพยากรณ์ ดังนั้น ค่าจากสถานที่ที่ห่างไกลจะไม่มีผลต่อค่าที่พยากรณ์ ดังนั้นสามารถตัดออกจากการคำนวณในจุดพยากรณ์นั้น ๆ โดยการกำหนด **search neighborhood**.

อีกเหตุผลหนึ่งในการใช้ **search neighborhood** คือเพื่อเพิ่มความเร็วในการคำนวณ หากมีจุดข้อมูลมากเกินไป (เช่น 2,000 จุด) เมทริกซ์ที่ใช้ในการคำนวณจะมีขนาดใหญ่เกินไป ซึ่งจะทำให้ไม่สามารถคำนวณค่าได้. เมื่อคุณลดขนาดของ **search neighborhood** คุณจะสามารถคำนวณค่าที่พยากรณ์ได้เร็วขึ้น.

การใช้ **search neighborhood** จะช่วยจำกัดจำนวนจุดที่ใช้ในการคำนวณค่าพยากรณ์ โดยกำหนดรูปร่างและขนาดของพื้นที่ที่ใช้ค้นหาค่าที่ใช้ในการพยากรณ์. คุณสามารถปรับขนาด, รูปร่าง หรือจำนวนของเพื่อนบ้านที่ใช้ใน **search neighborhood** ได้.

การกำหนดรูปร่างของ Search Neighborhood

- รูปร่างของ **search neighborhood** ขึ้นอยู่กับข้อมูลที่มีและลักษณะของพื้นผิวที่คุณพยายามจะสร้าง.
- ถ้าไม่มี **Directional Influence** (อิทธิพลตามทิศทาง) ในข้อมูลของคุณ, คุณจะต้องการใช้จุดข้อมูลในทุกทิศทางเท่า ๆ กัน และรูปร่างของ **search neighborhood** จะเป็นวงกลม.
- แต่ถ้ามี **Directional Autocorrelation** (ความสัมพันธ์ตามทิศทาง) หรือมีแนวโน้มในข้อมูล, คุณอาจต้องการให้รูปร่างของ **search neighborhood** เป็นรูปวงรี (Ellipse) ที่มีแกนใหญ่ขนานกับทิศทางที่มีความสัมพันธ์เชิงพื้นที่ในระยะทางไกล (Long-Range Autocorrelation).

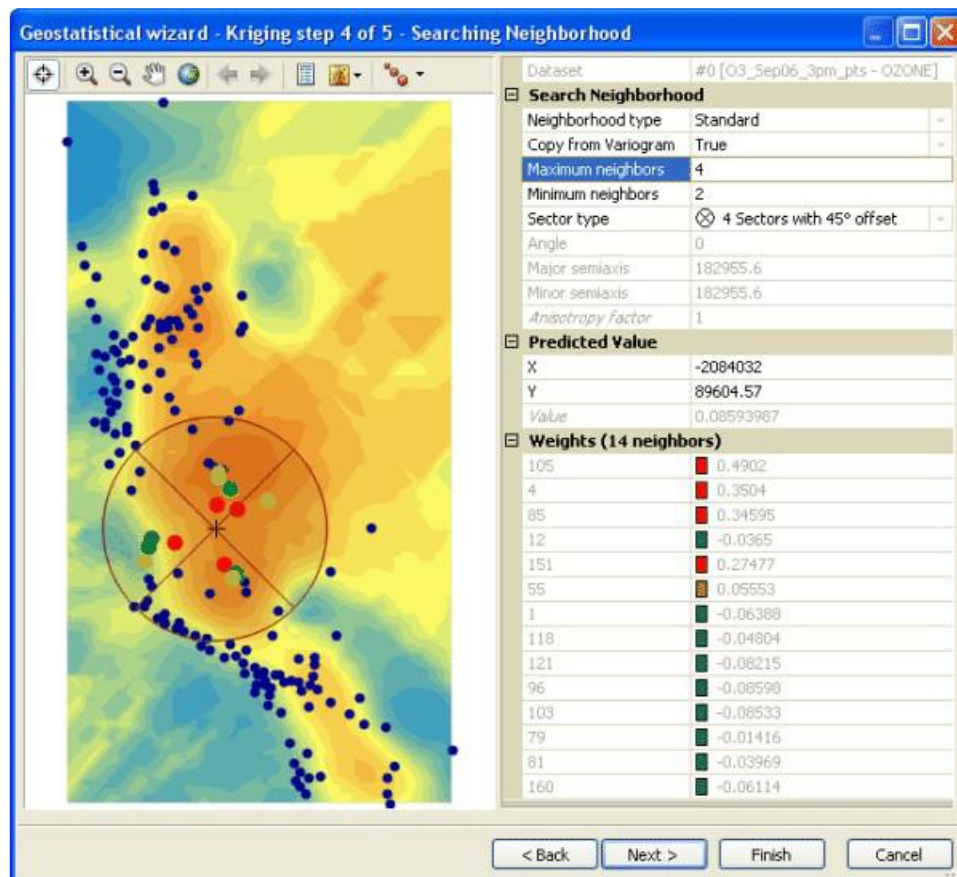
ตัวอย่างการใช้ Search Neighborhood ใน Geostatistical Wizard:

- ประเภทของ Neighborhood: Standard
- จำนวนเพื่อนบ้านสูงสุด: 4
- จำนวนเพื่อนบ้านขั้นต่ำ: 2
- ประเภท Sector (กลยุทธ์การค้นหา): Four Sectors with 45° offset
- รัศมี: 182955.6
- พิกัดของจุดทดสอบ (Test Point): ($x = -2084032$, $y = 89604.57$)
- ค่าที่พยากรณ์: 0.08593987

Weights

- ในส่วน **Weights** จะมีการแสดงน้ำหนักที่ใช้ในการประเมินค่าที่ตำแหน่งที่ระบุโดยเครื่องหมาย **Crosshair** บนพื้นผิวพริ้ว. จุดข้อมูลที่มีน้ำหนักมากที่สุดจะถูกเน้นด้วยสีแดง.

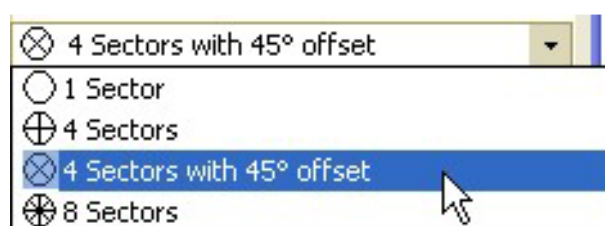
- เมื่อกำหนดรูปร่างของ **search neighborhood** แล้ว, คุณสามารถจำกัดตำแหน่งที่ควรใช้ในพื้นที่นั้น ๆ ได้. คุณสามารถกำหนด **จำนวนเพื่อนบ้านสูงสุด (Maximum Neighbors)** และ **จำนวนเพื่อนบ้านขั้นต่ำ (Minimum Neighbors)** ที่จะรวมไว้และแบ่งพื้นที่ **neighborhood** ออกเป็น **sectors** เพื่อให้มั่นใจว่าค่าจากทุกทิศทางจะถูกใช้. หากคุณแบ่งพื้นที่ออกเป็น **sectors**, จำนวนเพื่อนบ้านสูงสุดและขั้นต่ำที่กำหนดจะถูกนำไปใช้กับแต่ละ **sector**.



จะมี **ประเภทของ sector** หลายประเภทที่สามารถใช้งานได้ โดยสามารถปรับได้ตามลักษณะข้อมูลและวิธีการพยากรณ์ที่ต้องการ.

การใช้ **Weights** ช่วยให้เข้าใจถึงค่าที่ถูกคำนวณจาก **search neighborhood** โดยแสดงน้ำหนักของแต่ละจุดข้อมูลที่มีอิทธิพลต่อการพยากรณ์. คุณสามารถปรับขนาดและรูปร่างของ **neighborhood** เพื่อให้การพยากรณ์แม่นยำขึ้น โดยการกำหนด **sectors** เพื่อรวมค่าจากทุกทิศทาง.

ประเภทของ Sectors



4. Prediction errors

ใน Geostatistical Wizard ของ ArcGIS ค่า Prediction Errors ที่ใช้ในการประเมินผล โมเดลธรณีสัณติมีหลายค่าที่สำคัญ ซึ่งแต่ละค่านี้จะบ่งชี้ถึงคุณภาพของการทำนาย (Prediction) ที่ทำจากโมเดล โดยแต่ละค่าใช้ในการประเมินและการเปรียบเทียบความเหมาะสมของโมเดล ในกรณีที่การทำนายไม่เป็นที่พอใจจะช่วยให้เราปรับปรุงโมเดลได้

1. Mean Prediction Error (MPE)

- เป็นค่าเฉลี่ยของข้อผิดพลาดในการทำนาย ซึ่งคำนวณจากความแตกต่างระหว่างค่าจริงและค่าที่ทำนายออกมา โดยมีค่าเป็นบวกหรือเป็นลบขึ้นอยู่กับว่าการทำนายมีแนวโน้มที่จะสูงกว่าหรือต่ำกว่าค่าจริง
- หากค่า MPE ใกล้ 0 หมายความว่าในการทำนายโดยรวมไม่ค่อยมีอคติ (Bias-free) แต่ถ้าค่า MPE มีค่าสูงมาก แสดงว่าโมเดลมีอคติในการทำนาย
- ค่าที่ดีคือใกล้ 0 หรือไม่มีอคติอย่างชัดเจน

2. Root Mean Square Error (RMSE)

- ค่า RMSE เป็นการวัดความผิดพลาดเฉลี่ยในรูปของหน่วยของข้อมูลต้นฉบับ โดยคำนวณจากการหารากที่สองของค่าเฉลี่ยของข้อผิดพลาดที่ยกกำลังสอง
- ใช้เพื่อประเมินว่าความแตกต่างระหว่างค่าจริงและค่าทำนายมีขนาดใหญ่หรือไม่ หากค่า RMSE ต่ำแสดงว่าโมเดลทำนายได้ดี
- ค่าที่ต่ำแสดงว่าโมเดลทำนายได้แม่นยำ ส่วนค่าที่สูงอาจบ่งชี้ว่าโมเดลมีข้อผิดพลาดมาก

3. Mean Standardized Prediction Error (MSPE)

- เป็นค่าเฉลี่ยของข้อผิดพลาดที่ได้มาตรฐาน ซึ่งคำนวณจากการหารากที่สองของค่าเฉลี่ยของข้อผิดพลาดที่ได้มาตรฐาน (Standardized) และหารค่าเฉลี่ยของมัน
- ใช้ในการเปรียบเทียบผลการทำนายระหว่างหลายๆ โมเดลในกรณีที่มีหน่วยไม่เหมือนกัน โดยค่านี้จะช่วยให้เข้าใจได้ว่าโมเดลมีแนวโน้มที่จะมีความผิดพลาดที่ได้มาตรฐานเป็นอย่างไร
- ค่าที่ใกล้ 0 ถือว่าโมเดลมีการทำนายที่ไม่มีอคติในทุกช่วง

4. Root Mean Square Standardized Error (RMSSE)

- เป็นการหารากที่สองของค่าเฉลี่ยของข้อผิดพลาดที่ได้มาตรฐาน เพื่อให้สามารถเปรียบเทียบความแม่นยำได้ในรูปของมาตรฐาน
- ใช้ในการประเมินว่าโมเดลมีความแม่นยำในการทำนายหรือไม่ เมื่อเทียบกับการกระจายของข้อมูล
- ค่าที่ต่ำแสดงว่าโมเดลมีความแม่นยำในการทำนาย และค่านี้ควรใกล้ 0

5. Average Standard Error (ASE)

- ค่านี้คำนวณจากค่าเฉลี่ยของข้อผิดพลาดมาตรฐาน ซึ่งเป็นการวัดการกระจายของข้อผิดพลาดจากการทำนาย
- ใช้เพื่อประเมินว่าข้อผิดพลาดของโมเดลมีความกระจายตัวแค่ไหน
- ค่า ASE ที่ต่ำแสดงให้เห็นว่าโมเดลสามารถทำนายได้ดีและมีข้อผิดพลาดกระจายตัวน้อย

สรุป

- ค่าที่ใกล้ 0 คือค่าที่ดีที่สุดสำหรับ MPE, MSPE และ RMSSE
- ค่า RMSE และ ASE ควรต่ำเพื่อแสดงว่าโมเดลทำนายได้แม่นยำ
- การเลือกค่าที่ดีที่สุดนั้นขึ้นอยู่กับว่าคุณต้องการให้โมเดลมีความแม่นยำหรือมีอคติที่ต่ำที่สุด