

Lab10: Kriging II

1. Search neighborhoods

คุณสามารถสันนิษฐานได้ว่า เมื่อระยะทางจากตำแหน่งที่ต้องการพยากรณ์เพิ่มขึ้น ค่าที่วัดได้จะมีความสัมพันธ์เชิงพื้นที่ (spatial autocorrelation) กับตำแหน่งที่ต้องการพยากรณ์น้อยลง เนื่องจากจุดเหล่านี้มีผลกระทบต่อค่าที่พยากรณ์เพียงเล็กน้อยหรือไม่มีเลย ดังนั้นสามารถตัดค่าดังกล่าวออกจากการคำนวณค่าพยากรณ์ในจุดนั้นได้โดยการกำหนดขอบเขตการค้นหา (search neighborhood)

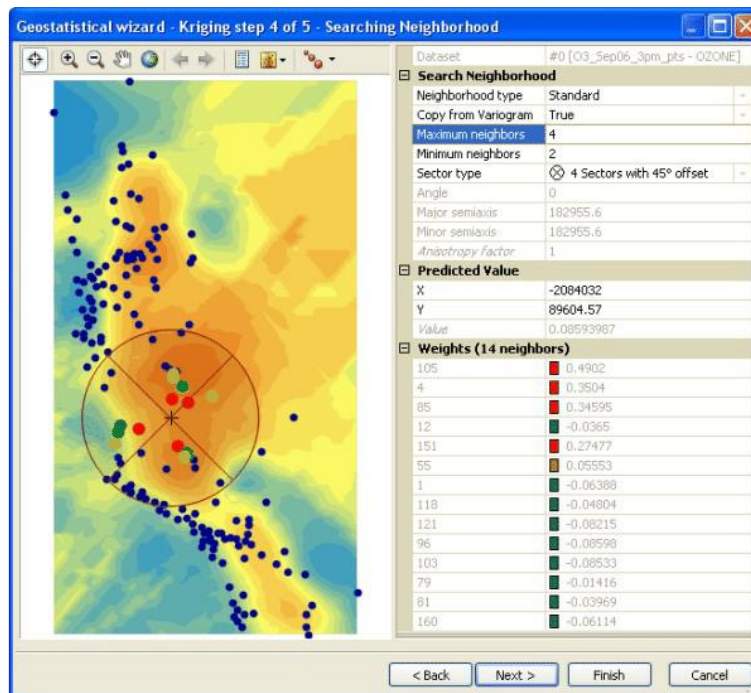
นอกจากนี้ ตำแหน่งที่อยู่ไกลออกไปอาจส่งผลกระทบในทางลบต่อค่าที่พยากรณ์ได้ หากอยู่ในพื้นที่ที่มีลักษณะแตกต่างจากตำแหน่งที่ต้องการพยากรณ์ อีกเหตุผลหนึ่งที่ต้องใช้ขอบเขตการค้นหาคือเพื่อเพิ่มความเร็วในการคำนวณ หากมีตำแหน่งข้อมูลถึง 2,000 ตำแหน่ง เมทริกซ์ที่ใช้จะมีขนาดใหญ่เกินกว่าจะคำนวณค่าพยากรณ์ได้ ดังนั้น ยิ่งขอบเขตการค้นหามีขนาดเล็ก ค่าพยากรณ์ก็จะถูกสร้างขึ้นได้เร็วขึ้น ด้วยเหตุนี้ จึงเป็นแนวปฏิบัติทั่วไปที่จะจำกัดจำนวนจุดที่ใช้ในการพยากรณ์โดยกำหนดขอบเขตการค้นหา

รูปร่างของขอบเขตการค้นหาจะกำหนดระยะทางและทิศทางของค่าที่วัดได้ที่จะถูกนำมาใช้ในการพยากรณ์ นอกจากนี้ยังสามารถกำหนดพารามิเตอร์เพิ่มเติมเพื่อจำกัดตำแหน่งที่ใช้ภายในขอบเขตการค้นหาได้ โดยสามารถเปลี่ยนขนาดและรูปร่างของขอบเขตการค้นหา หรือเปลี่ยนจำนวนจุดข้อมูลที่รวมอยู่ในการพยากรณ์

รูปร่างของขอบเขตการค้นหาจะได้รับอิทธิพลจากข้อมูลนำเข้าและพื้นที่ที่ต้องการสร้าง หากไม่มีอิทธิพลด้านทิศทางในความสัมพันธ์เชิงพื้นที่ของข้อมูล (สามารถดูรายละเอียดเพิ่มเติมในหัวข้อ "Accounting for directional influences") ควรใช้จุดข้อมูลในทุกทิศทางอย่างเท่าเทียมกัน โดยรูปร่างของขอบเขตการค้นหาจะเป็นวงกลม อย่างไรก็ตาม หากมีความสัมพันธ์เชิงพื้นที่ในทิศทางใดทิศทางหนึ่งหรือมีแนวโน้มของข้อมูล ควรกำหนดขอบเขตการค้นหาให้เป็นรูปวงรี โดยให้แกนหลักขนานกับทิศทางที่มีความสัมพันธ์เชิงพื้นที่ระยะไกลมากที่สุด (ทิศทางที่ค่าข้อมูลมีความคล้ายคลึงกันมากที่สุด)

ขอบเขตการค้นหาสามารถกำหนดได้ใน **Geostatistical Wizard** ดังตัวอย่างต่อไปนี้:

- **ประเภทของขอบเขตการค้นหา:** มาตรฐาน (Standard)
- **จำนวนเพื่อนบ้านสูงสุด:** 4
- **จำนวนเพื่อนบ้านขั้นต่ำ:** 2
- **ประเภทของเชกเตอร์ (กลยุทธ์การค้นหา):** แบ่งออกเป็น 4 เชกเตอร์ พร้อมออฟเซต 45° และรัศมี 182,955.6
- **พิกัดของจุดทดสอบ:** ($x = -2,084,032$, $y = 89,604.57$)
- **ค่าที่พยากรณ์ได้:** 0.08593987

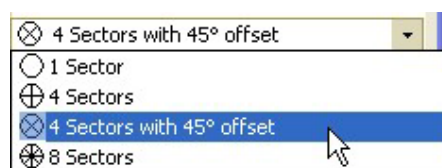


Searching neighborhood step

ส่วน **Weights** แสดงค่าน้ำหนักที่ใช้ในการประมาณค่าที่ตำแหน่งซึ่งถูกทำเครื่องหมายด้วยเครื่องหมายกากบาทบนพื้นผิวตัวอย่าง (preview surface) โดยจุดข้อมูลที่มีน้ำหนักมากที่สุดจะถูกไฮไลต์เป็นสีแดง

เมื่อกำหนดรูปร่างของขอบเขตการค้นหาแล้ว คุณสามารถจำกัดตำแหน่งภายในรูปร่างที่ควรใช้ได้ โดยสามารถกำหนดจำนวนเพื่อนบ้านสูงสุดและต่ำสุดที่รวมอยู่ในการคำนวณ และสามารถแบ่งขอบเขตการค้นหาออกเป็นเซกเตอร์ (sectors) เพื่อให้แน่ใจว่าค่าในทุกทิศทางได้รับการพิจารณา หากแบ่งขอบเขตการค้นหาออกเป็นเซกเตอร์ จำนวนเพื่อนบ้านสูงสุดและต่ำสุดที่กำหนดจะถูกนำไปใช้กับแต่ละเซกเตอร์

มีประเภทของเซกเตอร์หลายแบบที่สามารถใช้ได้ ได้แก่:



- **ประเภทของเซกเตอร์ที่สามารถใช้ได้ ได้แก่:**

- หนึ่งเซกเตอร์ (One sector)
- วงรีที่แบ่งเป็นสี่เซกเตอร์ (Ellipse with four sectors)
- วงรีที่แบ่งเป็นสี่เซกเตอร์พร้อมออฟเซต 45 องศา (Ellipse with four sectors and a 45-degree offset)
- แปดเซกเตอร์ (Eight sectors)

Kriging ใช้โครงสร้างข้อมูลที่กำหนดโดยขอบเขตการค้นหาร่วมกับแบบจำลองเซมิวาริโอแกรม (semivariogram model) เพื่อตรวจสอบค่าน้ำหนักของตำแหน่งที่วัดค่าได้ จากนั้นใช้ค่าน้ำหนักและค่าที่วัดได้เพื่อคำนวณค่าพยากรณ์ที่ตำแหน่งที่ต้องการ การคำนวณนี้จะดำเนินการ

ในทุกตำแหน่งภายในพื้นที่ศึกษาเพื่อสร้างพื้นผิวที่ต่อเนื่อง วิธีการทำอินเตอร์โพลชันอื่น ๆ ก็ใช้กระบวนการเดียวกัน แต่จะใช้เทคนิคที่แตกต่างกันในการกำหนดค่าน้ำหนัก ซึ่งไม่เกี่ยวข้องกับแบบจำลองเซมิวาริโอแกรม

จำนวนเพื่อนบ้านสูงสุดที่สามารถใช้ได้ ขึ้นอยู่กับวิธีการทำอินเตอร์โพลชัน:

- **Areal Interpolation:** 200 เพื่อนบ้าน
- **Diffusion Interpolation with Barriers:** ไม่จำกัดจำนวนเพื่อนบ้าน
- **Empirical Bayesian Kriging:** 64 เพื่อนบ้าน
- **Global Polynomial Interpolation:** ไม่ใช่ขอบเขตการค้นหา
- **Inverse Distance Weighted:** 1,000 เพื่อนบ้าน
- **Kernel Interpolation with Barriers:** ไม่จำกัดจำนวนเพื่อนบ้าน
- **Kriging (นอกเหนือจาก Empirical Bayesian Kriging):** 200 เพื่อนบ้าน
- **Local Polynomial Interpolation:** 1,000 เพื่อนบ้าน
- **Radial Basis Functions:** 64 เพื่อนบ้าน

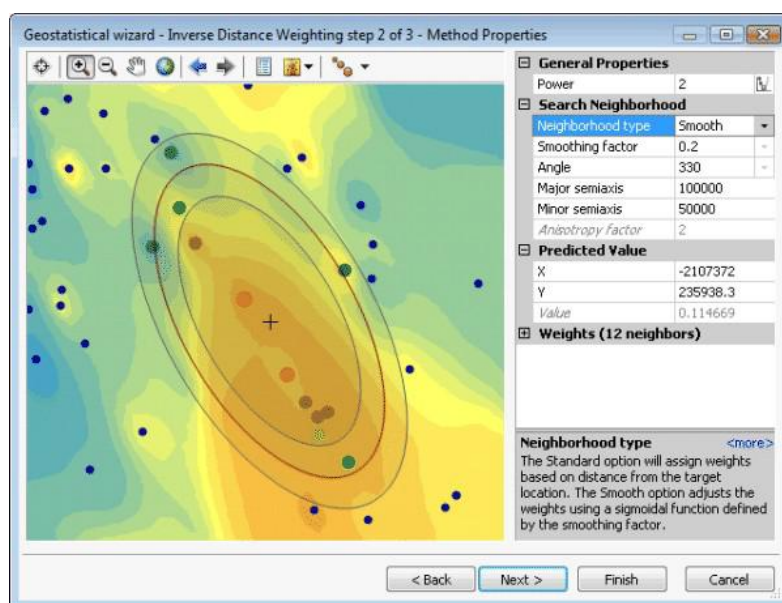
เมื่อใช้จำนวนเซกเตอร์ที่แตกต่างกัน จำนวนเพื่อนบ้านสูงสุดจะถูกหารด้วยจำนวนเซกเตอร์ ตัวอย่างเช่น

- หากใช้ **สี่เซกเตอร์** ให้หารจำนวนเพื่อนบ้านด้วย 4
- หากใช้ **แปดเซกเตอร์** ให้หารจำนวนเพื่อนบ้านด้วย 8

Smooth Interpolation Option

ขอบเขตการค้นหาแบบมาตรฐาน (Standard Search Neighborhood) ถูกกำหนดโดยพารามิเตอร์ของวงรี (Ellipse) ได้แก่:

- มุม (Angle)
- แกนกึ่งหลัก (Major Semiaxis)
- แกนกึ่งรอง (Minor Semiaxis)



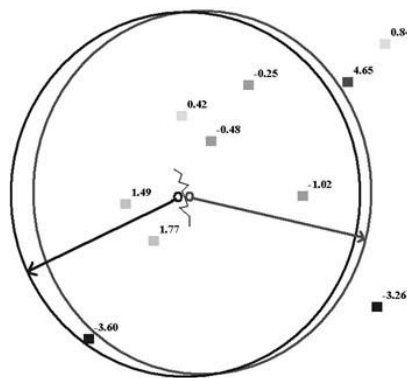
อย่างไรก็ตาม ตัวเลือก **Smooth Interpolation** จะสร้าง วงรีชั้นนอก (**Outer Ellipse**) และ วงรีชั้นใน (**Inner Ellipse**) โดยที่ระยะของวงรีทั้งสองจะขึ้นอยู่กับ ค่าปัจจัยการปรับเรียบ (**Smoothing Factor**) และความยาวของแกนกึ่งหลักและแกนกึ่งรองของวงรี

จุดข้อมูลทั้งหมดที่อยู่ภายในสามวงรีจะถูกใช้ในการทำอินเตอร์โพลชัน โดยจุดที่อยู่ นอกวงรีชั้นในแต่ยังอยู่ภายในวงรีชั้นนอก จะได้รับค่าน้ำหนักโดยใช้ฟังก์ชันซิกมอยด์ (sigmoidal function) ซึ่งมีค่าระหว่าง 0 ถึง 1

วิธีการปรับเรียบเชิงพื้นที่ (Local Smoothing Methods)

มีหลายวิธีที่ใช้สำหรับการปรับเรียบเชิงพื้นที่ โดยทั่วไป **Spline Algorithms** มักถูกใช้ในการสร้างแผนที่แบบต่อเนื่อง เนื่องจากช่วยลดความโค้งของพื้นผิวให้น้อยที่สุด ในขณะที่ **แบบจำลองทางภูมิศาสตร์เชิงสถิติ (Geostatistical Models)** มักถูกใช้สำหรับการพยากรณ์เชิงพื้นที่ที่เหมาะสมที่สุดและการทำแผนที่ในหลายสาขาวิทยาศาสตร์

อย่างไรก็ตาม **แบบจำลอง Kriging** แบบดั้งเดิม อาจสร้างพื้นผิวที่ไม่ต่อเนื่องได้เมื่อนำไปใช้กับขอบเขตการค้นหาแบบท้องถิ่น (Local Neighborhood) ตัวอย่างเช่น หากต้องการพยากรณ์ค่าที่สองจุดใกล้เคียงกัน ซึ่งสามารถแสดงเป็นศูนย์กลางของวงกลมสองวง การใช้ Kriging อาจทำให้เกิดความไม่ต่อเนื่องระหว่างจุดเหล่านี้



การพยากรณ์ค่าที่แต่ละจุดจะใช้ข้อมูลที่อยู่ภายในวงกลมแต่ละวงที่สอดคล้องกับตำแหน่งนั้น ๆ ตัวอย่างข้างต้นแสดงให้เห็นว่าความแตกต่างเพียงอย่างเดียวระหว่างขอบเขตการค้นหาแบบท้องถิ่นคือว่าตำแหน่งที่มีค่าประมาณ **-3.60** ถูกนำมาใช้หรือไม่

โดยทั่วไป **Kriging** ไม่สามารถสร้างพื้นผิวที่ต่อเนื่องได้ หากใช้ขอบเขตการค้นหาแบบท้องถิ่น ความไม่ต่อเนื่องของพื้นผิวจะเห็นได้ชัดเจนเมื่อข้อมูลมีค่าที่แตกต่างกันอย่างมีนัยสำคัญในพื้นที่ใกล้เคียงกัน

Altering the search neighborhood by changing its size and shape

ขนาดการค้นหาของ **neighborhood** กำหนดรูปร่างของ **neighborhood** และข้อจำกัดของจุดภายใน **neighborhood** ที่จะใช้ในการพยากรณ์ตำแหน่งที่ยังไม่ได้ทำการวัด

คุณสามารถตั้งค่าพารามิเตอร์ของ **neighborhood** โดยการดูตำแหน่งของจุดต่าง ๆ ในหน้าต่าง **data view** และใช้ความรู้ที่ได้รับจาก **ESDA** (Exploratory Spatial Data Analysis) และการสร้าง **semivariogram/covariance modeling**

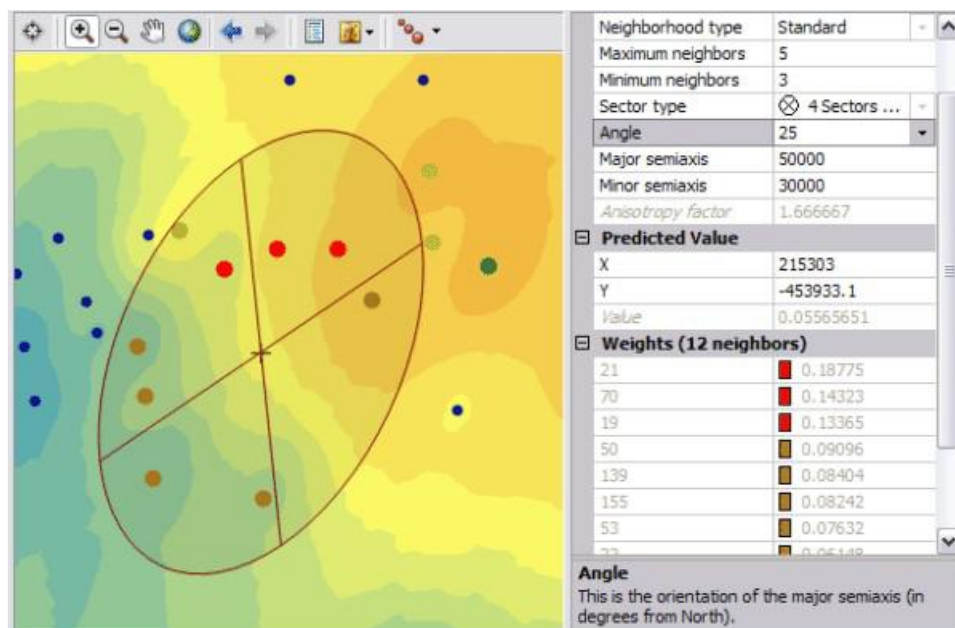
ใช้ขั้นตอนด้านล่างเป็นแนวทางในการเปลี่ยนแปลง **search neighborhood** สำหรับวิธีการ **interpolation** ใดๆ ที่มีอยู่ใน **Geostatistical Wizard** (ยกเว้น **global polynomial interpolation**)

ขั้นตอน:

1. ประเภทของ **Sector (Sector type)** จะถูกใช้ในการเปลี่ยนแปลงประเภทของ **search neighborhood** โดยเลือกจากรายการที่กำหนดไว้ล่วงหน้า
2. กำหนดค่ามุม (Angle value) ระหว่าง 0 ถึง 360 องศาหรือเลือกจากรายการรอบดาว นี่คือการวัดทิศทางของ **major semiaxis**, โดยวัดเป็นองศาจากทิศเหนือ
3. พารามิเตอร์ **Major semiaxis** และ **Minor semiaxis** จะถูกใช้ในการเปลี่ยนแปลงรูปร่างของ **ellipse**. รูปร่างที่ต้องการจะแสดงในหน้าต่างการแสดงผลหลังจากที่คุณกรอกค่าต่าง ๆ เหล่านี้

สรุป:

ขั้นตอนที่กล่าวมาช่วยให้คุณสามารถปรับการค้นหาค่าของ **neighborhood** เพื่อให้เหมาะสมกับลักษณะของข้อมูลที่คุณมี โดยการเลือกค่าที่เหมาะสมจาก **Sector type** และการปรับขนาดของ **semiaxis** เพื่อให้ได้รูปร่างที่ต้องการสำหรับการพยากรณ์



Altering the search neighborhood by changing the number of neighbors

ขนาดการค้นหาค่าของ **neighborhood** กำหนดรูปร่างของ **neighborhood** และข้อจำกัดของจุดภายใน **neighborhood** ที่จะใช้ในการพยากรณ์ตำแหน่งที่ยังไม่ได้ทำการวัด

คุณสามารถตั้งค่าพารามิเตอร์ของ **neighborhood** โดยการดูตำแหน่งของจุดต่าง ๆ ในหน้าต่าง **data view** และใช้ความรู้ที่ได้รับจาก **ESDA** (Exploratory Spatial Data Analysis) และการสร้าง **semivariogram/covariance modeling**

เคล็ดลับสำหรับการปรับ search neighborhood โดยการเปลี่ยนจำนวนเพื่อนบ้าน:

- แต่ละ **sector** จะถูกขยายออกไปหากไม่พบจำนวนจุดขั้นต่ำใน sector
- หากไม่มีจุดใดใน **searching neighborhood** สำหรับวิธีการ **interpolation** ส่วนใหญ่ จะหมายความว่าไม่สามารถพยากรณ์ที่ตำแหน่งนั้นได้
- แม้ว่าวิธีการ **interpolators** บางตัว เช่น **simple kriging** และ **disjunctive kriging** จะทำนายค่าในพื้นที่ที่ไม่มีจุดข้อมูลโดยใช้ค่าเฉลี่ยของชุดข้อมูล แต่เป็นการปฏิบัติทั่วไปในการเปลี่ยนแปลง **searching neighborhood** เพื่อให้มีจุดบางจุดใน **searching neighborhood**

ขั้นตอนในการเปลี่ยนแปลง search neighborhood สำหรับวิธีการ interpolation ที่มีอยู่ใน Geostatistical Wizard (ยกเว้น global polynomial interpolation):

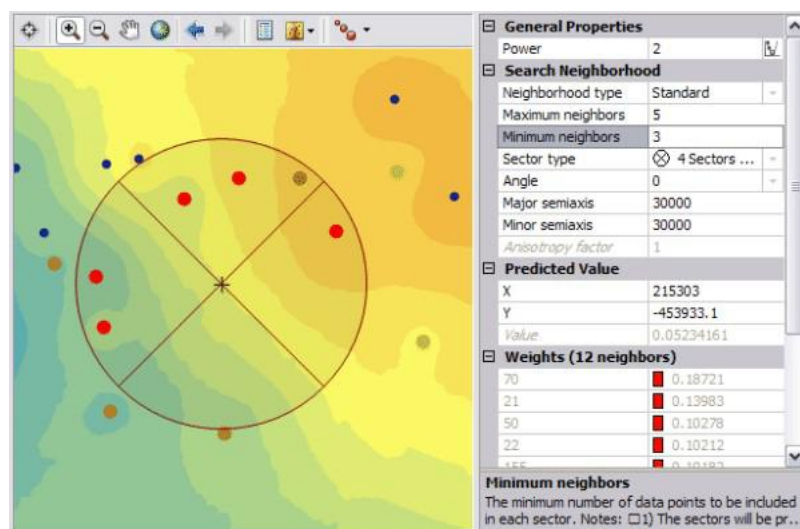
ขั้นตอนนี้ใช้เมื่อคุณได้กำหนดวิธีการ **interpolation** และข้อมูลที่ต้องการใช้แล้ว และได้ทำการดำเนินการจนถึงหน้าต่าง **Searching Neighborhood**

ขั้นตอน:

1. จำกัดจำนวนของ **neighbors** ที่จะใช้ในการพยากรณ์โดยการเปลี่ยนแปลงพารามิเตอร์ **Maximum neighbors** และ **Minimum neighbors**.
 - พารามิเตอร์เหล่านี้จะควบคุมจำนวน **neighbors** ที่รวมอยู่ในแต่ละ sector ของ **search neighborhood**
 - จำนวนและการจัดเรียงของ **sectors** สามารถเปลี่ยนแปลงได้โดยการปรับพารามิเตอร์ **Sector type**

สรุป:

ขั้นตอนนี้ช่วยให้คุณปรับขนาดของ **search neighborhood** โดยการกำหนดจำนวน **neighbors** ที่ต้องการใช้ในการพยากรณ์ เพื่อให้ได้การคำนวณที่แม่นยำและสมเหตุสมผลในการทำนายค่าของตำแหน่งที่ยังไม่ได้ทำการวัด.



เคล็ดลับ:

ผลกระทบของ **search neighborhood** สามารถประเมินได้โดยใช้เครื่องมือ **cross-validation** และ **comparison** ที่มีอยู่ใน **Geostatistical Analyst**. หากจำเป็น, คุณสามารถกำหนด **search neighborhood** ใหม่และสร้างพื้นผิวใหม่ (surface) ได้

สรุป:

การใช้เครื่องมือ **cross-validation** และ **comparison** จะช่วยให้คุณสามารถประเมินผลกระทบที่เกิดจากการตั้งค่าของ **search neighborhood** และหากพบว่าไม่เหมาะสมกับข้อมูล คุณสามารถปรับ **search neighborhood** และสร้าง **surface** ใหม่เพื่อให้การพยากรณ์แม่นยำยิ่งขึ้น

Accounting for directional influences

การมีอยู่ของส่วนประกอบทางทิศทางสามารถมีผลต่อการพยากรณ์ในพื้นที่ผลลัพธ์ของคุณได้ 2 ประเภท: แนวโน้มทั่วโลก (**global trends**) และ อิทธิพลทางทิศทางบน **semivariogram/covariance** (ที่เรียกว่า **anisotropy**)

1. **แนวโน้มทั่วโลก (Global trends):** คือกระบวนการที่มีผลกระทบต่อการวัดทั้งหมดในลักษณะที่เป็นเชิงกำหนด (deterministic). แนวโน้มทั่วโลกสามารถแสดงด้วยสมการทางคณิตศาสตร์ เช่น พหุนาม (polynomial) และสามารถลบออกจากการวิเคราะห์ข้อมูลที่ได้มา แต่จะต้องนำกลับเข้าไปก่อนที่จะทำการพยากรณ์. กระบวนการนี้เรียกว่า **detrending**
2. **Anisotropy (อิทธิพลทางทิศทาง):** คือผลที่เกิดจากการที่ข้อมูลมีความแตกต่างในลักษณะของการกระจายตามทิศทางต่าง ๆ ซึ่งสามารถนำมารวมกับการวิเคราะห์ **semivariogram/covariance** เพื่อให้การพยากรณ์ที่มีความแม่นยำมากขึ้น

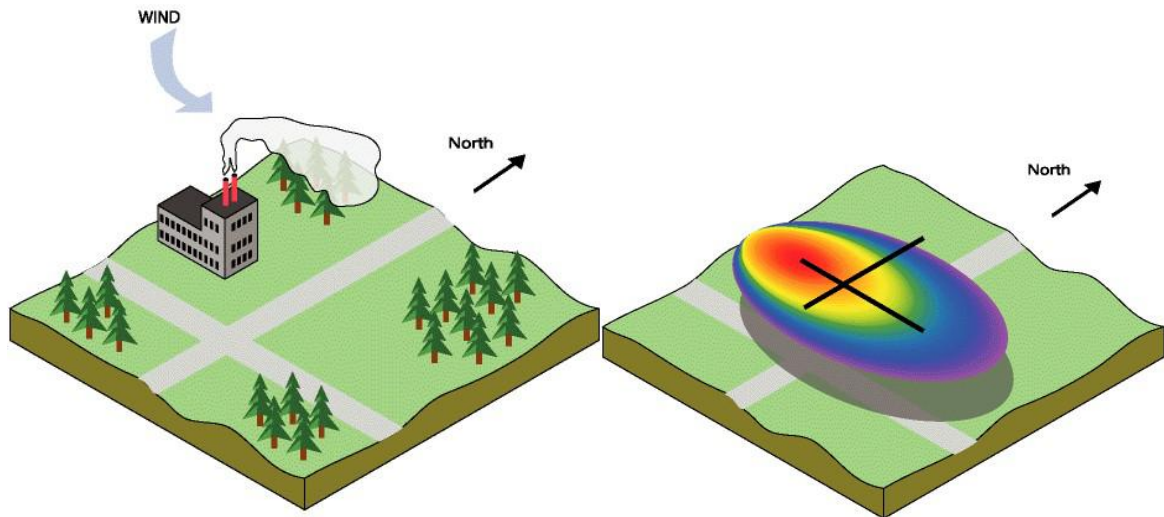
สรุป:

Detrending เป็นการลบ **global trend** ออกจากข้อมูลเพื่อให้การวิเคราะห์ไม่ถูกรบกวนจากแนวโน้มทั่วโลก และหลังจากนั้นจะสามารถนำแนวโน้มกลับมาเพื่อการพยากรณ์ที่แม่นยำ

ตัวอย่างของ global trend สามารถเห็นได้จากผลกระทบของลมที่พัดผ่านปล่องควันในโรงงาน ในภาพจะเห็นว่า ความเข้มข้นของมลพิษ ที่สูงจะถูกแสดงด้วยสีอบอุ่น (แดงและเหลือง) และความเข้มข้นที่ต่ำจะถูกแสดงด้วยสีเย็น (เขียวและน้ำเงิน). สังเกตได้ว่า ค่าของมลพิษเปลี่ยนแปลงซ้ำในทิศทาง ตะวันออก-ตะวันตก มากกว่าทิศทาง เหนือ-ใต้ เนื่องจากทิศทาง ตะวันออก-ตะวันตก สอดคล้องกับทิศทางลม ขณะที่ทิศทาง เหนือ-ใต้ ตั้งฉากกับลม

สรุป:

Global trend เช่น การกระจายตัวของมลพิษที่ได้รับผลจากทิศทางลมในตัวอย่างนี้ แสดงให้เห็นถึงการเปลี่ยนแปลงที่ซ้ำกันในทิศทางที่สอดคล้องกับลม (ตะวันออก-ตะวันตก) เมื่อเทียบกับทิศทางที่ตั้งฉากกับลม (เหนือ-ใต้)



รูปทรงของ **semivariogram/covariance curve** อาจแตกต่างกันไปตามทิศทาง (anisotropy) หลังจากที่ถูกนำออกไป หรือหากไม่มี global trend อยู่ในข้อมูล. **Anisotropy** แตกต่างจาก **global trend** ที่กล่าวถึงก่อนหน้านี้ เพราะว่า **global trend** สามารถอธิบายได้จากกระบวนการทางกายภาพ (เช่น ลมที่พัดผ่าน) และสามารถใช้สมการทางคณิตศาสตร์ มาทำการจำลองได้

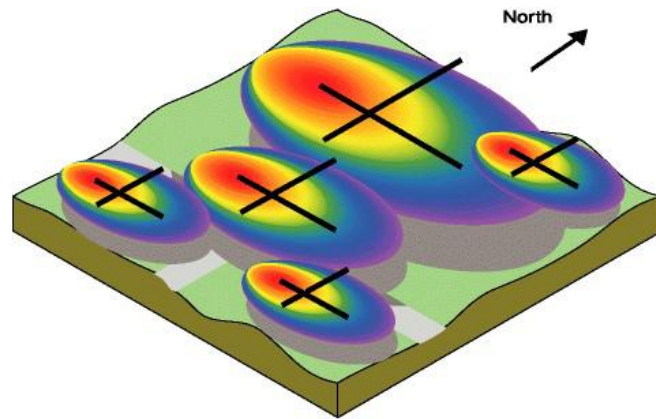
ในขณะที่ **anisotropy** หรือ ทิศทางที่มีอิทธิพล ใน **semivariogram** มักจะไม่ทราบสาเหตุที่ชัดเจน ดังนั้นจึงถือเป็น ข้อผิดพลาดสุ่ม หรือ **random error** ที่ไม่สามารถอธิบายได้. อย่างไรก็ตาม แม้ว่าจะไม่รู้สาเหตุของ anisotropy แต่ก็สามารถ คำนวณและจัดการกับอิทธิพลนี้ได้โดยการจำลอง

สรุป:

- **Anisotropy** อาจทำให้ **semivariogram/covariance curve** มีรูปทรงแตกต่างกันไปตามทิศทาง และมันเป็นอิทธิพลที่ไม่สามารถอธิบายได้จากกระบวนการทางกายภาพเหมือน **global trend**
- ถึงแม้ว่าไม่สามารถทราบสาเหตุได้ แต่ **anisotropy** ก็สามารถถูกคำนวณและนำมาพิจารณาในกระบวนการวิเคราะห์

Anisotropy โดยปกติไม่ใช่กระบวนการที่สามารถอธิบายได้ด้วยสูตรคณิตศาสตร์เดียว มันไม่มีแหล่งที่มาเดียวหรืออิทธิพลที่สามารถทำนายได้ว่าจะมีผลกระทบต่อทุกจุดที่วัดได้ Anisotropy เป็นลักษณะของกระบวนการสุ่มที่แสดงความสัมพันธ์ทางอัตโนมัติ (autocorrelation) ที่สูงกว่าในทิศทางหนึ่งเมื่อเทียบกับทิศทางอื่น ภาพด้านล่างนี้จะแสดงแนวคิดเกี่ยวกับลักษณะของกระบวนการนี้

อีกครั้งที่มีการแสดงความเข้มข้นของมลพิษที่สูงในสี่อุน (แดงและเหลือง) และความเข้มข้นที่ต่ำในสี่เยน (เขียวและน้ำเงิน) กระบวนการสุ่มนี้แสดงการบิดเบี้ยวที่มีความยาวนานน้อยกว่าในทิศทางหนึ่งเมื่อเทียบกับทิศทางอื่น การบิดเบี้ยวเหล่านี้อาจเกิดจากกระบวนการทางกายภาพที่ไม่รู้จักหรือไม่สามารถวัดได้ แต่จะถูกจำลองเป็นกระบวนการสุ่มที่มีความสัมพันธ์ทางอัตโนมัติในทิศทาง



เมื่อคุณใช้วิธีการ Kriging และเลือกที่จะพิจารณา anisotropy (ความไม่เป็นอิสระตามทิศทาง) เซมิเวรีโอแกรมเชิงประจักษ์ (empirical semivariogram) จะมีความสัมพันธ์ทางอวกาศที่แตกต่างกันสำหรับแต่ละทิศทาง ในตัวอย่างนี้ รูปร่างของเซมิเวรีโอแกรมจะเพิ่มขึ้นเร็วขึ้นในทิศทางเหนือ-ใต้ก่อนที่จะเริ่มแบนราบ (ทิศทางนี้คือทิศทางที่ความเข้มข้นเปลี่ยนแปลงเร็วขึ้น) จุดในเซมิเวรีโอแกรมเชิงประจักษ์จะได้รับการฟิตด้วยโมเดลเซมิเวรีโอแกรมที่มีรูปร่างแตกต่างกันตามทิศทางของมัน และน้ำหนักของ Kriging จะได้รับการปรับตามนั้น

Predicting values for specific locations

หากคุณต้องการพยากรณ์ค่าที่ตำแหน่งเฉพาะหรือหลายตำแหน่งโดยไม่ต้องพยากรณ์ค่าทั่วทั้งพื้นผิว คุณสามารถใช้เครื่องมือใน **ArcMap** ได้ ตัวอย่างเช่น คุณอาจต้องการตรวจสอบระดับรังสีที่บ้านหลังหนึ่ง, ระดับความสูงของหอชมวิวกำลังจะสร้าง, หรือระดับ **PCB (Polychlorinated Biphenyls)** ที่พื้นที่ว่ายน้ำ

วิธีที่ง่ายที่สุดในการดูค่าพยากรณ์

- ใช้ **MapTips** เพื่อดูค่าพยากรณ์ของตำแหน่งที่ต้องการ

หากต้องการบันทึกค่าพยากรณ์ที่ตำแหน่งเฉพาะลงในเลเยอร์ผลลัพธ์ เพื่อใช้ในการวิเคราะห์เพิ่มเติม ให้ทำตามขั้นตอนดังต่อไปนี้:

ขั้นตอนการพยากรณ์ค่าเฉพาะตำแหน่งใน ArcMap

1. คลิกขวา ที่ **Geostatistical Layer** ใน **Table of Contents** (ซึ่งสร้างขึ้นจากข้อมูลที่เกิดขึ้นได้)
2. คลิก **Validation/Prediction**
3. ในเครื่องมือ **GA Layer To Points** ; ใช้ **Point Observation Locations** เพื่อเลือกตำแหน่งที่ต้องการพยากรณ์ค่า
4. เลือก **Field to Validate On** ให้เหมาะสม
5. ระบุ **Output Feature Class** เพื่อบันทึกผลลัพธ์
6. คลิก **OK**
7. คลิกขวา ที่ **Prediction Layer** และเลือก **Open Attribute Table** เพื่อแสดงผลลัพธ์

2. Empirical Bayesian Kriging (EBK)

Introduction

Empirical Bayesian Kriging (EBK) เป็นวิธีการประมาณค่าทางธรณีสถิติที่ช่วยให้กระบวนการสร้างโมเดลคริกกิ้งที่ถูกต้องเป็นไปได้โดยอัตโนมัติ ส่วนวิธีคริกกิ้งอื่นๆ ใน Geostatistical Analyst นั้นจำเป็นต้องปรับพารามิเตอร์ด้วยตนเองเพื่อให้ได้ผลลัพธ์ที่แม่นยำ แต่ EBK สามารถคำนวณพารามิเตอร์เหล่านี้โดยอัตโนมัติผ่านกระบวนการแบ่งกลุ่มข้อมูลย่อย และการจำลองค่า

นอกจากนี้ EBK ยังแตกต่างจากวิธีคริกกิ้งอื่นๆ ตรงที่สามารถคำนึงถึงข้อผิดพลาดที่เกิดจากการประมาณค่า semivariogram ซึ่งเป็นองค์ประกอบสำคัญในการทำคริกกิ้ง วิธีคริกกิ้งทั่วไปจะคำนวณค่า semivariogram จากตำแหน่งข้อมูลที่ทราบ และใช้ค่า semivariogram เดียวกันนี้ในการพยากรณ์ค่าที่ตำแหน่งที่ไม่ทราบ ซึ่งกระบวนการนี้มีข้อสมมติว่าค่า semivariogram ที่ได้จากข้อมูลที่มีอยู่เป็นค่าจริงสำหรับทั้งพื้นที่คำนวณ อย่างไรก็ตาม การละเลยความไม่แน่นอนของการประมาณค่า semivariogram ทำให้วิธีคริกกิ้งทั่วไปประเมินค่าความคลาดเคลื่อนมาตรฐานของการพยากรณ์ต่ำกว่าความเป็นจริง

Empirical Bayesian Kriging มีให้ใช้งานใน **Geostatistical Wizard** และเป็นเครื่องมือในการประมวลผลทางภูมิศาสตร์ (geoprocessing tool)

ข้อดี

- ต้องการการสร้างโมเดลแบบโต้ตอบน้อยมาก
- ค่าความคลาดเคลื่อนมาตรฐานของการพยากรณ์มีความแม่นยำกว่าวิธีคริกกิ้งอื่นๆ
- สามารถพยากรณ์ข้อมูลที่มีความไม่แน่นอนระดับปานกลางได้อย่างแม่นยำ
- มีความแม่นยำกว่าวิธีคริกกิ้งอื่นๆ เมื่อใช้กับชุดข้อมูลขนาดเล็ก

ข้อเสีย

- เวลาในการประมวลผลจะเพิ่มขึ้นอย่างรวดเร็วเมื่อจำนวนจุดนำเข้า ขนาดของกลุ่มข้อมูลย่อย หรือค่าปัจจัยการทับซ้อนเพิ่มขึ้น นอกจากนี้ การใช้การแปลงข้อมูลจะทำให้การประมวลผลช้าลง โดยเฉพาะเมื่อเลือกใช้โมเดลกึ่งแบริโอแกรมประเภท K-Bessel หรือ K-Bessel Detrended
- การประมวลผลช้ากว่าวิธีคริกกิ้งอื่นๆ โดยเฉพาะเมื่อสร้างผลลัพธ์เป็นแรสเตอร์
- ไม่รองรับการทำโคคริกกิ้ง (Cokriging) และการแก้ไขคุณสมบัติแอนไอโซทรอปิก (anisotropic corrections)
- การแปลงแบบ Log Empirical มีความไวต่อค่าผิดปกติสูง หากใช้การแปลงนี้กับชุดข้อมูลที่มีค่าผิดปกติ อาจทำให้ค่าพยากรณ์สูงหรือต่ำกว่าค่าจุดนำเข้าหลายเท่าตัว

การประมาณค่า Semivariogram

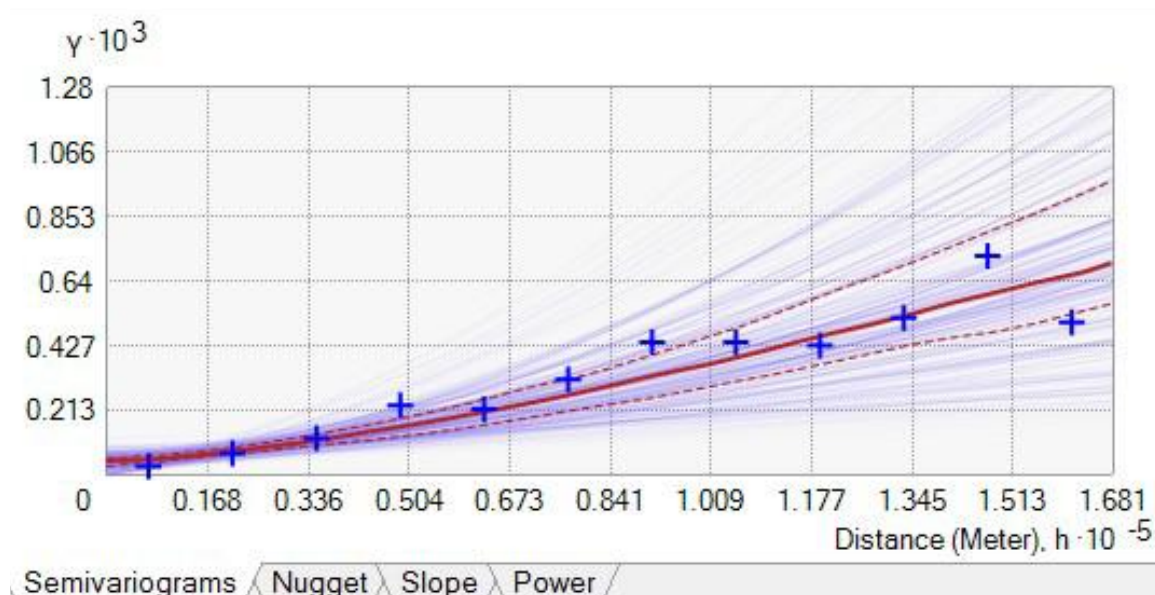
ต่างจากวิธีคริกกิงอื่นๆ (ซึ่งใช้วิธีถ่วงน้ำหนักกำลังสองน้อยที่สุด - Weighted Least Squares) พารามิเตอร์ของกึ่งแวรีโอแกรมใน **Empirical Bayesian Kriging (EBK)** ถูกประมาณค่าด้วยวิธี **Restricted Maximum Likelihood (REML)** อย่างไรก็ตาม เนื่องจาก REML มีข้อจำกัดด้านการคำนวณสำหรับชุดข้อมูลขนาดใหญ่ ข้อมูลนำเข้าจึงถูกแบ่งออกเป็นกลุ่มย่อยที่มีการทับซ้อนกัน โดยขนาดของแต่ละกลุ่มถูกกำหนดไว้ล่วงหน้า (ค่าเริ่มต้นคือ 100 จุดต่อกลุ่ม)

ในแต่ละกลุ่มย่อย กึ่งแวรีโอแกรมจะถูกประมาณค่าโดยใช้กระบวนการดังต่อไปนี้:

1. คำนวณค่ากึ่งแวรีโอแกรมจากข้อมูลภายในกลุ่มย่อย
2. ใช้กึ่งแวรีโอแกรมที่ได้เป็นโมเดลเพื่อจำลองข้อมูลใหม่แบบไม่มีเงื่อนไขที่ตำแหน่งข้อมูลนำเข้าในกลุ่มย่อย
3. คำนวณค่ากึ่งแวรีโอแกรมใหม่จากข้อมูลที่จำลองขึ้น
4. ทำซ้ำขั้นตอนที่ 2 และ 3 ตามจำนวนครั้งที่กำหนด โดยในแต่ละครั้งจะใช้กึ่งแวรีโอแกรมที่ได้จากขั้นตอนที่ 1 เป็นพื้นฐานในการจำลองข้อมูลใหม่ แล้วนำข้อมูลที่จำลองขึ้นมาใช้ในการคำนวณกึ่งแวรีโอแกรมใหม่

กระบวนการนี้สร้าง semivariogram จำนวนมากสำหรับแต่ละกลุ่มย่อย และเมื่อนำมาพล็อตรวมกัน จะได้การแจกแจงของ semivariogram ที่มีการกระจายตามความหนาแน่น (ยังมีสีน้ำเงินเข้มมาก แสดงว่ามี semivariogram จำนวนมากผ่านบริเวณนั้น)

ค่ากึ่งแวรีโอแกรมเชิงประจักษ์ (empirical semivariances) จะแสดงเป็นเครื่องหมายกากบาทสีน้ำเงิน นอกจากนี้ ค่ามัธยฐานของการแจกแจงจะแสดงด้วยเส้นสีแดงทึบ ขณะที่เปอร์เซ็นต์ไทล์ที่ 25 และ 75 จะแสดงด้วยเส้นประสีแดง ตามตัวอย่างที่แสดงด้านล่าง



จำนวน semivariogram ที่จำลองต่อกลุ่มย่อยถูกตั้งค่าเริ่มต้นไว้ที่ **100 ชุด** โดยแต่ละ semivariogram ที่ได้เป็นค่าประมาณของ semivariogram จริงในกลุ่มย่อยนั้น

สำหรับแต่ละตำแหน่งที่ต้องการพยากรณ์ ค่า **พยากรณ์** จะถูกคำนวณโดยใช้การแจกแจงของ semivariogram ใหม่ ซึ่งได้จากการสุ่มตัวอย่างโดยอิงตามค่าความน่าจะเป็น (likelihood-based sampling) ของ semivariogram แต่ละตัวจากกลุ่ม semivariogram ในบริเวณใกล้เคียง

ตัวอย่างเช่น หากตำแหน่งที่ต้องการพยากรณ์มีจุดข้อมูลเพื่อนบ้านที่อยู่ในสามกลุ่มย่อยที่แตกต่างกัน (ตามที่กำหนดโดยขอบเขตการค้นหา - searching neighborhood) ค่า **พยากรณ์** จะถูกคำนวณโดยใช้ semivariogram ที่จำลองจากแต่ละสามกลุ่มย่อยนี้ โดย semivariogram ที่นำมาใช้จะถูกเลือกตามค่าความน่าจะเป็นของแต่ละตัว

Kriging model

Empirical Bayesian Kriging (EBK) แตกต่างจากวิธีคริกกิงอื่นๆ ใน **Geostatistical Analyst** เนื่องจากใช้ **Intrinsic Random Function (IRF)** เป็นโมเดลคริกกิง

โดยปกติ วิธีคริกกิงทั่วไปจะสมมติว่าข้อมูลมีค่าเฉลี่ยโดยรวม (หรือแนวโน้มที่กำหนดไว้ล่วงหน้า) และมีค่าความแปรปรวนรอบค่าเฉลี่ยนั้น ซึ่งหมายความว่า ค่าที่เบี่ยงเบนออกไปมาก จะถูกดึงกลับเข้าหาค่าเฉลี่ยเสมอ ทำให้ค่าที่พยากรณ์ออกมาไม่เบี่ยงเบนมากเกินไป

อย่างไรก็ตาม **EBK** ไม่ได้ตั้งสมมติฐานว่าข้อมูลมีแนวโน้มไปสู่ค่าเฉลี่ยโดยรวม ซึ่งหมายความว่า ค่าเบี่ยงเบนที่มากอาจเพิ่มขึ้นต่อไปหรืออาจลดลงก็ได้ในอัตราเดียวกัน ดังนั้น **Intrinsic Random Functions** จึงสามารถแก้ไขแนวโน้มในข้อมูลได้โดยธรรมชาติ

Semivariogram model

สำหรับระยะทาง h ที่กำหนด **Empirical Bayesian Kriging (EBK)** รองรับเซมิแควเรียนกรั้มต่อไปนี้:

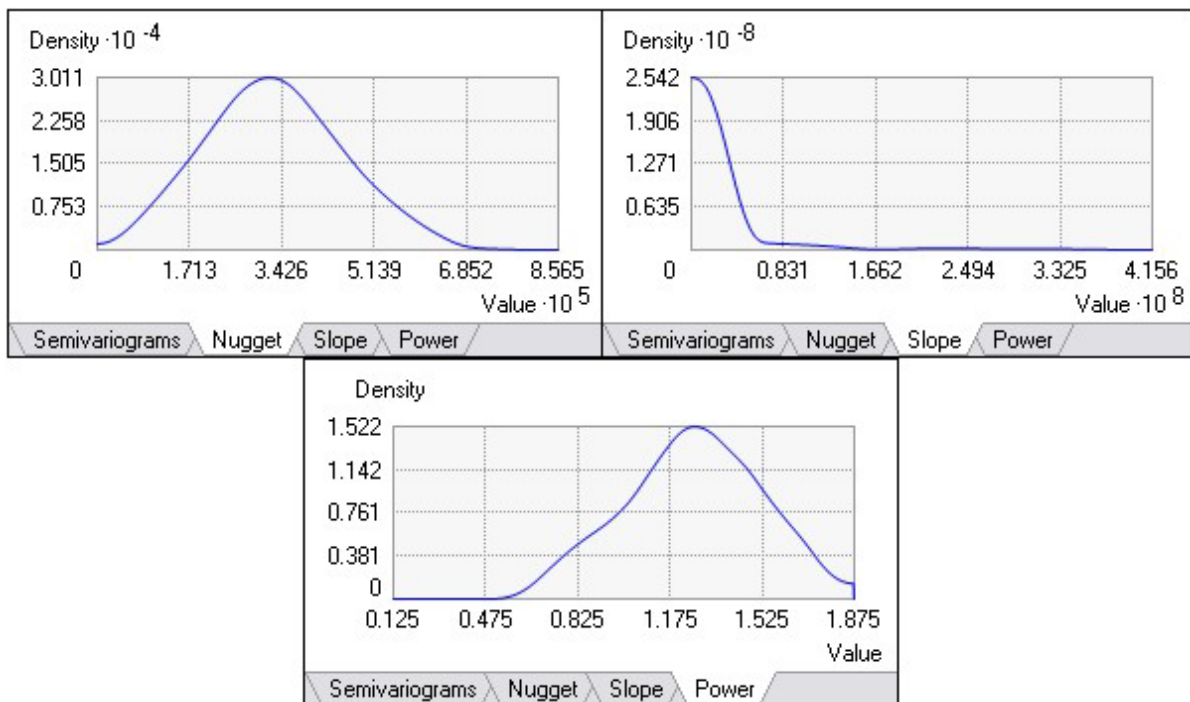
- **Power**
 $\gamma(h) = \text{Nugget} + b|h|^a$
- **Linear**
 $\gamma(h) = \text{Nugget} + b|h|$
- **Thin Plate Spline**
 $\gamma(h) = \text{Nugget} + b|h|^2 * \ln(|h|)$

ค่า **Nugget** และ **b (slope)** ต้องเป็นค่าบวกเสมอ และค่า **a (power)** ต้องอยู่ในช่วง **0.25** ถึง **1.75** ภายใต้ข้อจำกัดเหล่านี้ พารามิเตอร์จะถูกประมาณค่าด้วยวิธี **Restricted Maximum Likelihood (REML)**

โมเดลกึ่งแควรีโอแกรมใน **EBK** ไม่มีพารามิเตอร์ **Range** หรือ **Sill** เนื่องจากฟังก์ชันที่ใช้ไม่มีขอบเขตบน (upper bound)

ใน **EBK** สามารถวิเคราะห์การแจกแจงของค่าพารามิเตอร์ที่ประมาณได้ เนื่องจากการคำนวณกึ่งแควรีโอแกรมจำนวนมากในแต่ละตำแหน่ง เมื่อคลิกที่แท็บ **Nugget, Slope หรือ Power** ระบบจะแสดงการแจกแจงของพารามิเตอร์ที่เกี่ยวข้อง

ภาพด้านล่างแสดงการแจกแจงของพารามิเตอร์ semivariogram ที่ได้จาก semivariogram จำลองในตัวอย่างก่อนหน้านี้



Distributions of nugget, slope, and power

เมื่อคลิกที่ตำแหน่งอื่นบนพื้นผิวตัวอย่าง (preview surface) ระบบจะแสดงการแจกแจงของกึ่งแปรโอแกรมและพารามิเตอร์ของกึ่งแปรโอแกรมสำหรับตำแหน่งใหม่ หากการแจกแจงไม่เปลี่ยนแปลงอย่างมีนัยสำคัญทั่วทั้งขอบเขตข้อมูล แสดงว่าข้อมูลมี **ความเป็นสถานีโดยรวม (globally stationary)**

โดยปกติ การแจกแจงควรเปลี่ยนแปลงอย่าง **ราบรื่น** ตามตำแหน่งในขอบเขตข้อมูล อย่างไรก็ตาม หากพบว่าการแจกแจงมีการเปลี่ยนแปลงอย่างรวดเร็วในระยะทางสั้นๆ อาจบ่งชี้ถึงความไม่เป็นสถานีของข้อมูล หรือความผันผวนสูงของพารามิเตอร์

ในกรณีนี้ การเพิ่มค่า **Overlap Factor** สามารถช่วยทำให้การเปลี่ยนแปลงของการแจกแจง **ราบรื่นขึ้น** และลดความผันผวนระหว่างจุดที่อยู่ใกล้กัน

หมายเหตุ: ตามที่อธิบายในส่วน **Transformations** ด้านล่าง การใช้การแปลงข้อมูล จะเปลี่ยนโมเดลคริกกิงจาก **Intrinsic Random Function** เป็น **Simple Kriging Mode** เมื่อเปลี่ยนเป็น **Simple Kriging** จะมี โมเดลกึ่งแปรโอแกรมเพิ่มเติม ให้เลือกใช้งาน ซึ่งอาจช่วยปรับปรุงความแม่นยำของการพยากรณ์ในบางกรณี

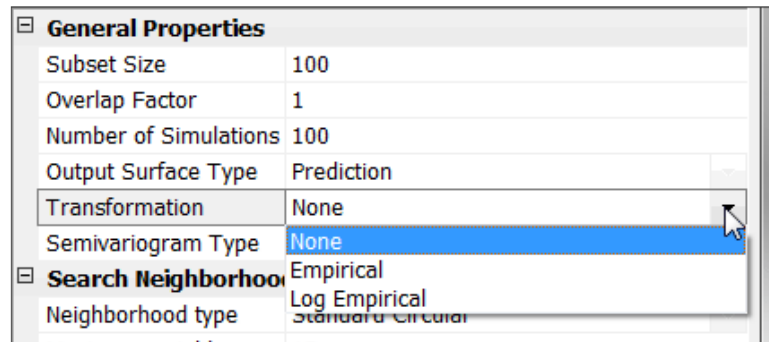
Transformations

Empirical Bayesian Kriging (EBK) มีตัวเลือก **Multiplicative Skewing Normal Score Transformation** โดยสามารถเลือกใช้หนึ่งในสองรูปแบบของการแจกแจงฐาน (**Base Distributions**) ได้แก่

1. **Empirical**
2. **Log Empirical**

Log Empirical Transformation มีข้อจำกัดที่สำคัญคือ ค่าข้อมูลทั้งหมดต้องเป็นบวก และจะ รับประกันว่าผลการพยากรณ์ทั้งหมดเป็นค่าบวก

การแปลงประเภทนี้เหมาะสำหรับข้อมูลที่ไม่สามารถมีค่าติดลบได้ เช่น ปริมาณน้ำฝน ซึ่งโดยธรรมชาติแล้วไม่สามารถมีค่าเป็นลบได้

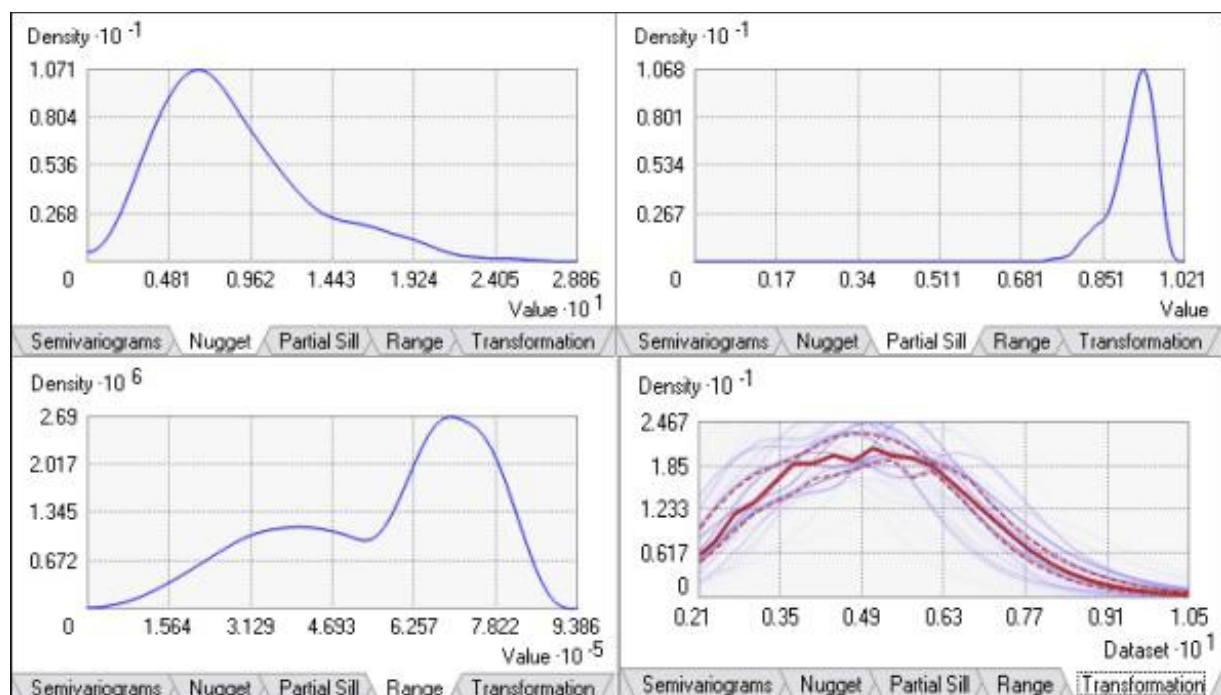


หากมีการใช้ การแปลงข้อมูล (transformation) โมเดลที่ใช้จะเปลี่ยนจาก **Intrinsic Random Function** เป็น **Simple Kriging Model**

เนื่องจากการเปลี่ยนแปลงนี้ พารามิเตอร์ของ semivariogram ก็จะเปลี่ยนไปเช่นกัน โดยแทนที่ **Nugget, Slope และ Power** ด้วยพารามิเตอร์ใหม่ ได้แก่:

- **Nugget**
- **Partial Sill**
- **Range**

พารามิเตอร์เหล่านี้ใช้กันทั่วไปใน **Simple Kriging** และช่วยกำหนดลักษณะของ semivariogram ในลักษณะที่แตกต่างจากโมเดลเดิม



หากเลือกใช้ **K-Bessel** หรือ **K-Bessel Detrended** เป็นประเภทของ **Semivariogram** จะมี กราฟเพิ่มเติม สำหรับพารามิเตอร์ **Shape** ของ K-Bessel แสดงขึ้นมา

นอกจากนี้ จะมี แท็บ **Transformation** เพิ่มเติม ที่แสดงการแจกแจงของค่าการแปลงที่คำนวณได้ (**fitted transformations**) โดยแต่ละค่าการแปลงจะสอดคล้องกับการจำลองแต่ละครั้ง

เช่นเดียวกับแท็บ semivariogram การแจกแจงของการแปลงจะแสดงเป็นสีที่ไล่ตามความหนาแน่นของข้อมูล (ยิ่งเข้มหมายถึงมีความหนาแน่นสูง) และมีเส้นเปอร์เซ็นต์ไทล์ (**quantile lines**) กำกับไว้เพื่อช่วยแสดงช่วงของค่าที่ได้

Semivariograms

ทุกวิธีการทางธรณีสถิติ (Geostatistics) สมมติฐานว่ามีการพึ่งพิงเชิงพื้นที่ (Spatial Autocorrelation) ซึ่งหมายความว่าสิ่งที่อยู่ใกล้กันมักจะมีค่าคล้ายคลึงกันมากกว่าสิ่งที่อยู่ห่างออกไป และ semivariogram จะเป็นตัวกำหนดว่าความคล้ายคลึงนี้ลดลงอย่างไรตามระยะทาง

semivariogram บางประเภท (เช่น **Exponential**) สมมติว่าความคล้ายคลึงกันลดลงอย่างรวดเร็ว ขณะที่ **Whittle Semivariogram** สมมติว่าความคล้ายคลึงกันลดลงอย่างช้าๆ ถึงแม้ว่าจะมีค่า **Nugget, Range** และ **Sill** เท่ากัน แต่ทั้งสองแบบนี้จะอธิบายการลดลงของความคล้ายคลึงกันในลักษณะที่แตกต่างกันอย่างมา

กุญแจสำคัญในการได้ผลลัพธ์ที่น่าเชื่อถือคือการเลือก semivariogram ที่ตรงกับพฤติกรรมของปรากฏการณ์ที่คุณกำลังศึกษามากที่สุด โดยรูปแบบของ semivariogram ที่สามารถใช้งานได้จะขึ้นอยู่กับการเลือก การแปลงข้อมูล (**Transformation**)

หากตั้งค่า **Transformation** เป็น "**None**"

สามารถเลือก semivariogram จากตัวเลือกต่อไปนี้:

- **Power** (ค่าเริ่มต้น)
- **Linear**
- **Thin Plate Spline**

หากตั้งค่า **Transformation** เป็น "**Empirical**" หรือ "**Log Empirical**"

สามารถเลือกกึ่งเวรีโอแกรมจากตัวเลือกต่อไปนี้:

- **Exponential** (ค่าเริ่มต้น)
- **Exponential Detrended**
- **Whittle**
- **Whittle Detrended**
- **K-Bessel**
- **K-Bessel Detrended**

เกี่ยวกับกึ่งเวรีโอแกรมแบบ **Detrended**

ทั้งสามแบบ **Detrended Semivariogram** จะมีลักษณะเหมือนกับแบบที่ไม่ใช่ Detrended

แต่มีการลบแนวโน้มระดับที่หนึ่ง (First-order Trend Removal) ออก การลบแนวโน้มนี้จะไม่มีผลกระทบต่อความเร็วในการคำนวณ ขณะเดียวกันการมีแนวโน้มในข้อมูลสามารถตรวจสอบได้โดยใช้เครื่องมือ **Trend Analysis** ใน **ESDA**

ข้อดีและข้อเสียของแต่ละแบบจำลอง Semivariograms

แต่ละ semivariogram มีข้อดีและข้อเสีย การเลือก semivariogram ต้องพิจารณาถึงเวลาในการคำนวณและความยืดหยุ่นของโมเดล (ความสามารถในการรองรับชุดข้อมูลที่หลากหลายได้อย่างแม่นยำ):

Power

- **ข้อดี:** คำนวณได้เร็วและยืดหยุ่น โดยทั่วไปเป็นตัวเลือกที่ปลอดภัยที่สามารถบาลานซ์ระหว่างประสิทธิภาพและความแม่นยำได้
- **ข้อเสีย:** ยืดหยุ่นน้อยกว่าตัวเลือกอื่นและคำนวณช้ากว่า

Linear

- **ข้อดี:** คำนวณได้เร็วมาก
- **ข้อเสีย:** เป็นโมเดลที่ยืดหยุ่นน้อยที่สุด

Thin Plate Spline

- **ข้อดี:** คำนวณได้เร็วมาก ทำงานได้ดีที่สุดเมื่อมีแนวโน้มที่แข็งแกร่ง
- **ข้อเสีย:** ยืดหยุ่นน้อยกว่า โดยเฉพาะเมื่อไม่มีแนวโน้ม

Exponential

- **ข้อดี:** มีการแปลงที่ยืดหยุ่น คำนวณได้เร็วกว่าค่า K-Bessel และ K-Bessel Detrended
- **ข้อเสีย:** รูปทรงของ semivariogram ไม่ยืดหยุ่น คำนวณช้ากว่า Power, Linear และ Thin Plate Spline

Exponential Detrended

- **ข้อดี:** มีการแปลงที่ยืดหยุ่น คำนวณได้เร็วกว่าค่า K-Bessel และ K-Bessel Detrended ลบแนวโน้มระดับที่หนึ่งออก
- **ข้อเสีย:** รูปทรงของ semivariogram ไม่ยืดหยุ่น คำนวณช้ากว่า Power, Linear และ Thin Plate Spline

Whittle

- **ข้อดี:** มีการแปลงที่ยืดหยุ่น คำนวณได้เร็วกว่าค่า K-Bessel และ K-Bessel Detrended
- **ข้อเสีย:** รูปทรงของกึ่งแวริโอแกรมไม่ยืดหยุ่น คำนวณช้ากว่า Power, Linear และ Thin Plate Spline

Whittle Detrended

- **ข้อดี:** มีการแปลงที่ยืดหยุ่น คำนวณได้เร็วกว่าค่า K-Bessel และ K-Bessel Detrended ลบแนวโน้มระดับที่หนึ่งออก
- **ข้อเสีย:** รูปทรงของ semivariogram ไม่ยืดหยุ่น คำนวณช้ากว่า Power, Linear และ Thin Plate Spline

K-Bessel

- ข้อดี: ยืดหยุ่นและแม่นยำที่สุด
- ข้อเสีย: ใช้เวลาคำนวณนานที่สุด

K-Bessel Detrended

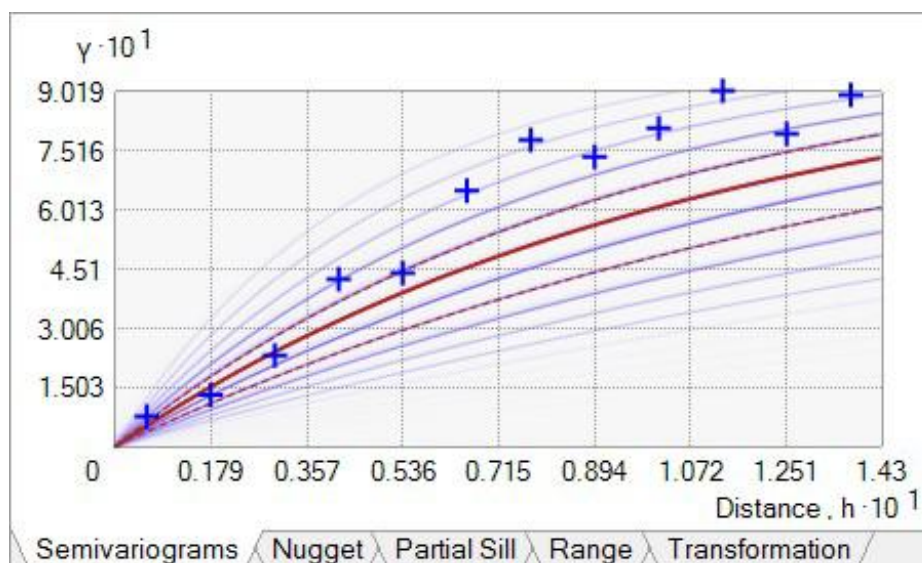
- ข้อดี: ยืดหยุ่นและแม่นยำที่สุด ลบแนวโน้มระดับที่หนึ่งออก
- ข้อเสีย: ใช้เวลาคำนวณนานที่สุด

การเลือกแบบจำลอง Semivariograms

ควรพิจารณาจากเกณฑ์ต่อไปนี้:

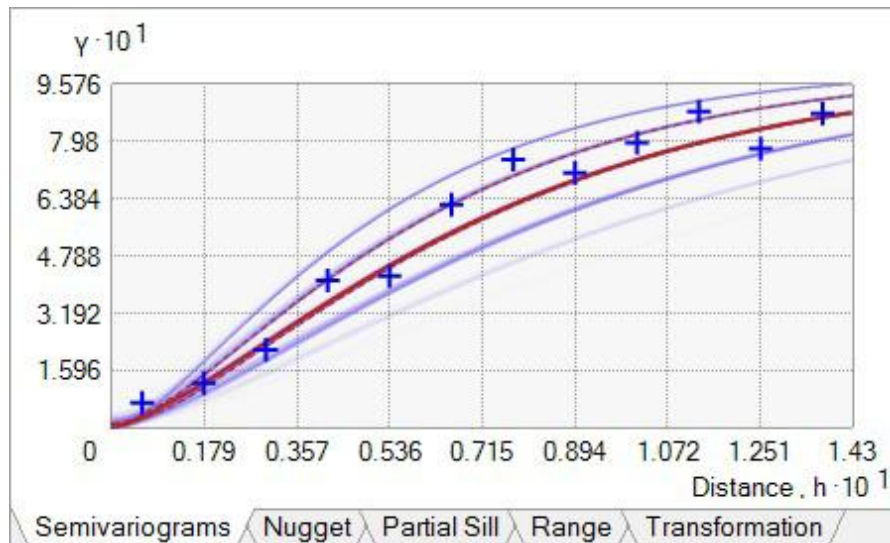
- หากคุณพร้อมที่จะรอเพื่อให้ได้ผลลัพธ์ที่แม่นยำที่สุด ควรเลือก **K-Bessel** หรือ **K-Bessel Detrended** การเลือกจะขึ้นอยู่กับการมีหรือไม่มีแนวโน้มในข้อมูล
- หากคุณต้องการผลลัพธ์อย่างรวดเร็วและยอมเสียความแม่นยำบางส่วน ควรเลือก **Linear** หรือ **Thin Plate Spline** หากไม่มีแนวโน้มหรือแนวโน้มอ่อนแอ **Linear** จะเป็นตัวเลือกที่ดีกว่า
- หากต้องการสมดุลระหว่างความแม่นยำและความเร็ว **Power** เป็นตัวเลือกที่ดี
- หากจำเป็นต้องใช้การแปลงข้อมูล แต่ไม่สามารถรอผลลัพธ์นาน ควรเลือก **Exponential** หรือ **Whittle (หรือแบบ Detrended)** โดยควรเลือกตัวที่ตรงกับ **semivariances** ที่ได้จากข้อมูลจริง ใน **Geostatistical Wizard** (อธิบายข้างล่าง) การตรวจสอบด้วย **Cross-validation** ควรนำมาพิจารณาด้วย

หากต้องการเลือกระหว่าง **Exponential, Whittle, และแบบ Detrended** ควรเลือก **Semivariograms** ที่ให้การปรับที่ดีที่สุด กับ **empirical semivariances** (จุดสีน้ำเงินในกราฟข้างล่าง) โดยปกติแล้ว **empirical semivariances** ควรจะอยู่กลางสเปกตรัมของ semivariograms เช่น ในกราฟนี้ จุดสีน้ำเงินไม่ได้อยู่กลางสเปกตรัมของ semivariograms (ส่วนใหญ่จะอยู่ด้านบนของสเปกตรัม)



Empirical semivariances do not fall in the middle of the spectrum.

แทนที่จะเลือกสเปกตรัม semivariograms ที่จุดสีน้ำเงินไม่ได้อยู่กลางสเปกตรัม ดังนั้นควรเลือก กึ่งเวรีโอแกรมที่สเปกตรัมมีจุดสีน้ำเงินอยู่กลางสเปกตรัม ซึ่งแสดงให้เห็นว่า empirical semivariations ตกอยู่ในส่วนที่เหมาะสมที่สุดภายในสเปกตรัมของกึ่งเวรีโอแกรม



Empirical semivariations fall in the middle of the spectrum.

การคำนวณระยะทางสำหรับข้อมูลในพิกัดภูมิศาสตร์

หากข้อมูลนำเข้าอยู่ในระบบพิกัดภูมิศาสตร์ ระยะทางจะถูกคำนวณโดยใช้ **ระยะทางแบบ chordal** ซึ่งหมายถึงระยะทางในแนวเส้นตรงที่เชื่อมต่อระหว่างจุดสองจุด เส้นตรงนี้จะผ่านตัวโลกแทนที่จะวิ่งไปตามผิวของโลก เพื่อให้เห็นภาพนี้ ลองนึกถึงการส่องไฟด้วยไฟฉายผ่านทรงกลมโปร่งใส ความยาวของลำแสงระหว่างจุดที่แสงเข้าทรงกลมและจุดที่แสงออกจากทรงกลม คือระยะทางแบบ chordal ระหว่างสองจุดนั้น

ข้อดีหลักของการใช้ระยะทางแบบ chordal เมื่อเทียบกับระยะทางแบบ geodesic คือ มีความต้องการในการคำนวณที่น้อยกว่า นอกจากนี้ ยังมีทฤษฎีเกี่ยวกับการประยุกต์ใช้ kriging บน spheroids อยู่ในระดับที่จำกัด

หมายเหตุ:

เนื่องจากระยะทางแบบ chordal ไม่สามารถประมาณค่าเป็นระยะทางแบบ geodesic ได้ดี สำหรับระยะทางที่เกิน 30 องศาทศนิยม รัศมีการค้นหาจึงต้องไม่เกิน 15 องศาทศนิยม (ดังนั้นเส้นผ่านศูนย์กลางจะต้องไม่เกิน 30 องศา) และตำแหน่งใด ๆ ที่ไม่มีเพื่อนบ้านภายใน 15 องศาทศนิยมจะถูกคำนวณเป็น NoData

นอกจากนี้ แบบจำลอง semivariograms บางแบบจำเป็นต้องมีการปรับระนาบราบสำหรับแต่ละชุดข้อมูลย่อยเพื่อทำการลบแนวโน้ม (trend removal) โดยระนาบนี้ไม่สามารถสร้างได้อย่างแม่นยำสำหรับชุดข้อมูลย่อยที่มีขอบเขตเกิน 30 องศาทศนิยม ดังนั้นขอบเขตของชุดข้อมูลย่อยแต่ละชุดจึงถูกจำกัดไว้ที่ 30 องศาสำหรับแบบจำลอง semivariograms ดังต่อไปนี้:

- Thin Plate Spline
- Exponential Detrended
- Whittle Detrended
- K-Bessel Detrended

พารามิเตอร์เพิ่มเติมสำหรับ Empirical Bayesian Kriging

Empirical Bayesian kriging ใช้พารามิเตอร์สามตัวที่ไม่พบในวิธีการ kriging อื่น ๆ ได้แก่:

- **Subset Size:** กำหนดจำนวนจุดในแต่ละชุดข้อมูล ขนาดของชุดข้อมูลที่ใหญ่ขึ้นจะทำให้การคำนวณของ EBK ใช้เวลานานขึ้น
- **Overlap Factor:** กำหนดระดับของการทับซ้อนระหว่างชุดข้อมูล แต่ละจุดข้อมูลสามารถตกอยู่ในหลายชุดได้ และ Overlap Factor จะระบุจำนวนชุดเฉลี่ยที่แต่ละจุดข้อมูลจะตกอยู่ เช่น หากค่า Overlap Factor เป็น 1.5 หมายความว่า จุดครึ่งหนึ่งจะถูกใช้ในชุดหนึ่ง และครึ่งหนึ่งจะถูกใช้ในสองชุด ค่าของ Overlap Factor ที่สูงขึ้นทำให้พื้นที่ที่ได้มีความเรียบเนียนขึ้น แต่จะเพิ่มเวลาในการประมวลผล
- **Number of Simulations:** กำหนดจำนวน semivariograms ที่จะถูกจำลองสำหรับแต่ละชุดข้อมูล การจำลองจำนวนมากจะทำให้การทำนายมีความแม่นยำมากขึ้น แต่เวลาในการประมวลผลจะเพิ่มขึ้น