

지리통계 분석 - Geostatistical Analyst

지리통계 분석(Geostatistical Analyst)은 향상된 연속면 모델링을 위한 결정론이나 지리적 방법론에서 사용된다. 실질적인 공간 데이터 분석이나 통계적으로 유효한 연속면을 만드는 과정으로 지리통계 마법사를 통해 생성한다. 지리통계 분석을 통해 만들어진 연속면은 GIS모델로 빈번히 사용되고 다른 익스텐션인 Spatial Analyst와 3D Analyst와 함께 지도 시각화 방법으로 사용된다.

지리통계 분석은 지리통계와 GIS 사이에 위치하고 있다. 지리통계 도구는 잘 조직된 GIS모델링 환경을 제외하고 언제나 사용이 가능하다. 통합은 GIS전문가에 의해서 예견되는 연속면의 통계적 오류들을 측정함으로써 연속면 모델의 질을 정량화할 수 있기 때문에 중요하다.

지리통계 분석을 이용하여 연속면을 유용하게 만드는 세가지 방법이 있다.

- 실질적인 공간데이터 분석
- 구조 분석
- 연속면 예상에 따른 결과물

지리통계 분석 도구는 이런 세가지 방법으로 사용자가 쉽게 만들 수 있고 구조적 공간 데이터 분석을 할 수 있는 유일한 도구를 제시하고 있다.

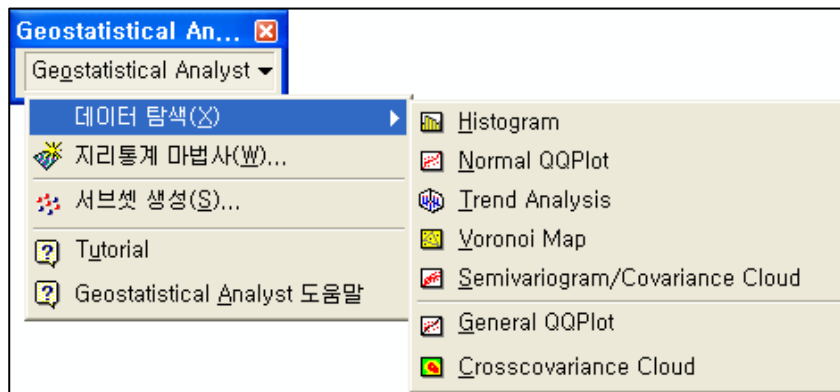


그림 1 Geostatistical Analyst의 메뉴

데이터 탐색(Explore Data) – 막대그래프(Histogram)

정규 QQPlot (Normal QQPlot)

추세 분석(Trend Analysis)

보노로이 맵(Voronoi Map)

반배리오그램/공분산 Cloud(Semivariogram/Covariance Cloud)

일반 QQPlot(General QQPlot)

교차 공분산 Cloud(Crosscovariance Cloud)

지리통계 마법사(Geostatistical Wizard)

서브 셋 생성(Create Subsets)

적절한 연속면 생성 방법론

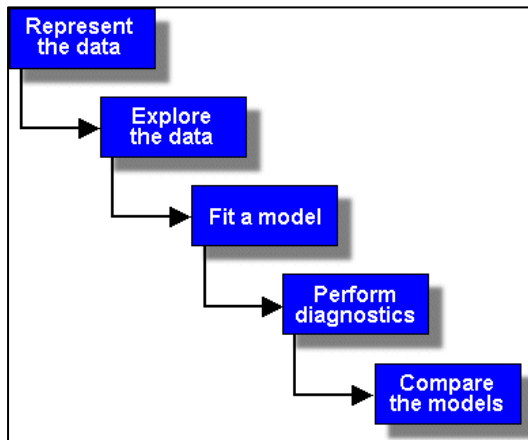


그림 2 연속면 생성 방법론의 구조

데이터입력(Represent the data)

ArcMap 데이터 뷰에 레이어를 추가하고 디스플레이

데이터탐색(Explore the data)

입력한 데이터 셋의 통계적 속성을 조사

데이터를 탐색하거나 연속면을 생성할 때 지리 통계도구 사용 가능

적합한 모델 찾기(Fit a model)

연속면을 생성하기 위한 모델 구상. 탐구를 위한 데이터는 적당한 모델을 찾는 데 도움

실행 결과 진단(Perform diagnostics)

만들어진 연속면의 평가. 모델이 알려지지 않는 값을 어떻게 예상했는지 이해하는 데 도움

모델의 비교(Compare the models)

만일 하나 이상의 연속면을 만들었다면 결과물이 샘플이 외 알지 못했던 값이 위치한 연속면을 가장 잘 나타내는 지와 결과물끼리 비교가 가능한지 확인

기능소개

막대그래프(Histogram)

데이터탐색(Explore Data) -> 막대그래프(Histogram)

- 현재 ArcMap 상에 레이어의 필드값을 통해 히스토그램 작성

공간 보간 방법으로 연속면을 산출하는데 만약에 데이터가 정규 분포되어 있다면 좋은 결과를 만들 수 있다. 만일 데이터가 벗어나도 그 데이터를 표준적인 형태로 전환할 수 있다. 그러므로 연속면을 만들기 전에 데이터들의 분포를 이해할 필요가 있다. 히스토그램 도구는 데이터 셋 안에 속성값으로 막대그래프를 그릴 수 있다.

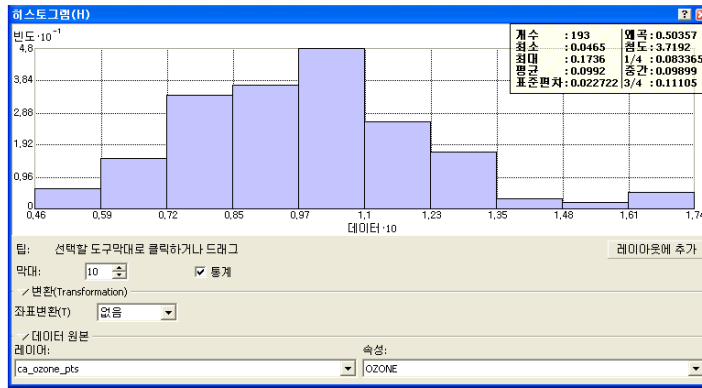


그림 3 히스토그램

정규 QQPlot(Normal QQPlot)

데이터 탐색(Explore Data) -> 정규 QQPlot(Normal QQPlot)

- 정규분포의 분위수(quantiles, 모집단 혹은 표본의 전체 도수를 몇 등분)와 이에 대응하는 자료 분포의 분위수를 좌표 평면에서 각각 수평축과 수직축의 좌표로 하여 점을 찍은 것 (표준 정규분포에 비해 실제 데이터가 분포하는 것을 비교)
- None, Log, Arcsin(역사인 함수), Box-Cox(데이터들이 등분산을 갖는 정규분포가 되도록 변환) 와 같은 변환기법 이용 가능

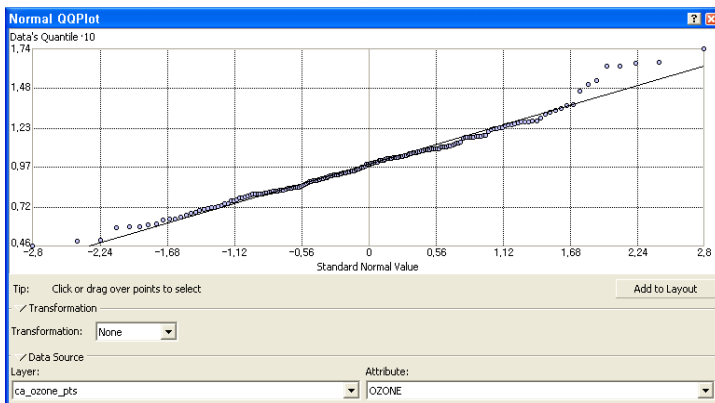
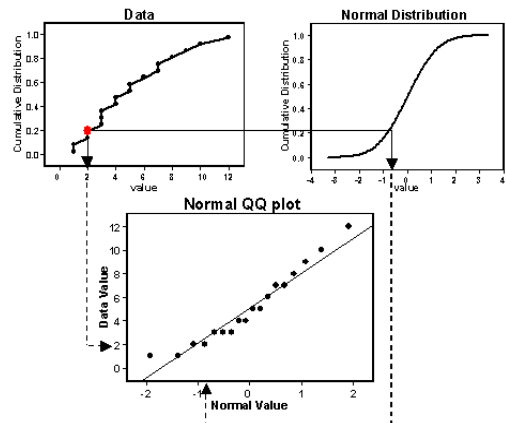


그림 4 정규 QQPlot



QQ Plot는 데이터표준 정규분포와 데이터의 분포를 비교한다.

점들이 밀접할 수록 그래프에서 직선으로 나타나고 밀접한 표본 데이터는 표준 분포를 따라간다. 일반 QQ Plot는 실제 데이터와 정규 분포를 대비해서 표시한 분위수의 그래프이다. 두 개의 분포가 일치하기 위해 QQ Plot는 직선으로 나타난다. 그러므로 표준정규분포의 분위수에 비교해 데이터들의 분위수의 오존 데이터 정규도를 확인할 수 있다. 정규 QQ Plot보다 일반 QQ Plot는 직선에 매우 밀접하다. 이 선에서의 주요 시작점은 집중된 오존의 높은 데이터로부터 발생된다. 만약 히스토그램이나 정규 QQ Plot로부터 정규분포를 나타낼 수 없을 때 정규분포와 같게 만들기 위한 변환 크리깅 보간기법을 사용하기 전에 필요하다.

일반 QQPlot(General QQPlot)

데이터탐색(Explore Data) -> 일반QQPlot(General QQPlot)

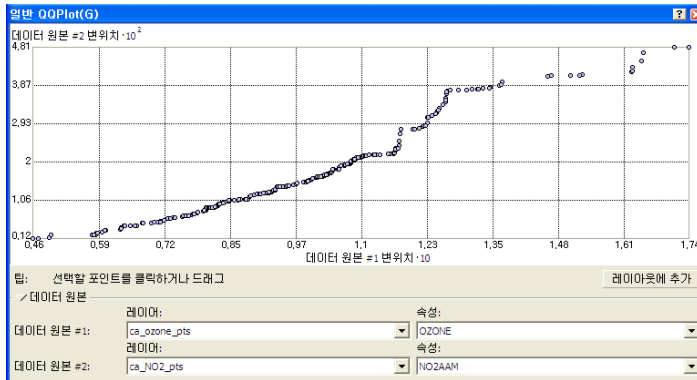
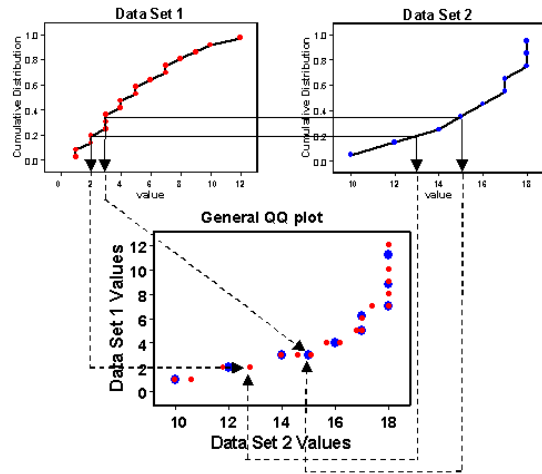


그림 5 일반 QQPlot



일반 QQPlot은 두 데이터 셋의 분포 유사성을 찾는 데 사용한다. 두 데이터 셋의 값이 하나의 그래프에서 동등한 위치에 표시된다.

추세 분석(Trend Analysis)

데이터 탐색(Explore Data) -> 추세 분석(Trend Analysis)

- 변화하는 데이터의 움직임으로부터 추출되는 추세 선을 통한 데이터 분석

만약 데이터에 추세(일정한 방향으로 나가는 경향)가 존재한다면 그것은 수학적 공식에 의해 임의적으로 구성된 연속면을 표현한다고 할 수 있다.

추세분석 도구는 입력된 데이터 셋에 대해 추세가 있는 지 여부를 확인한다.

추세 분석에서 각각의 수직 막대는 데이터 포인트들의 값과 위치를 표현한다. 포인트들은 동-서 축과 북-남 축으로 해서 수직 면으로 나오게 된다. 가장 적합한 라인은 예상된 포인트들을 통해 그려지고 특별한 경향을 보이는 것을 통해 모델 경향이 나타난다. 만약 라인이 수평이라면 이것은 기울기가 없다는 것을 나타낸다. 그러나 만약 녹색 선이 이미지보다 더 위에 나타난다면 낮은 값을 표현하는 것이고 평평히 밖으로 나갈 때까지 이동한다면 증가하는 것이다. 이 시연을 통해 데이터들이 동-서 방향의 라인이 강한 추세를 보여주고 북-서 방향의 라인은 그보다 약한 것으로 확인할 수 있다.

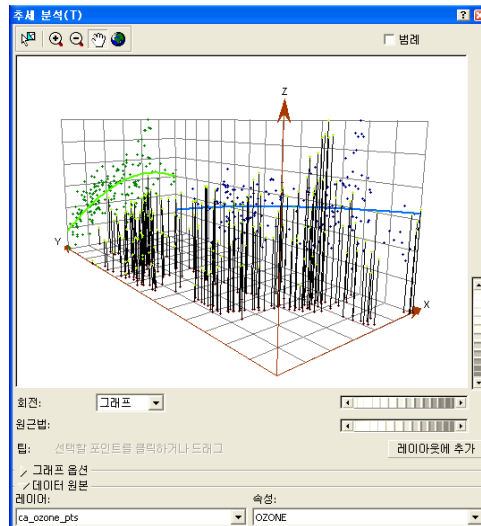


그림 6 추세 분석

반배리오그램 / 공분산¹ Cloud(Semivariogram/Covariance Cloud)

데이터 탐색(Explore Data) -> 반배리오그램/공분산(Semivariogram/Covariance Cloud)

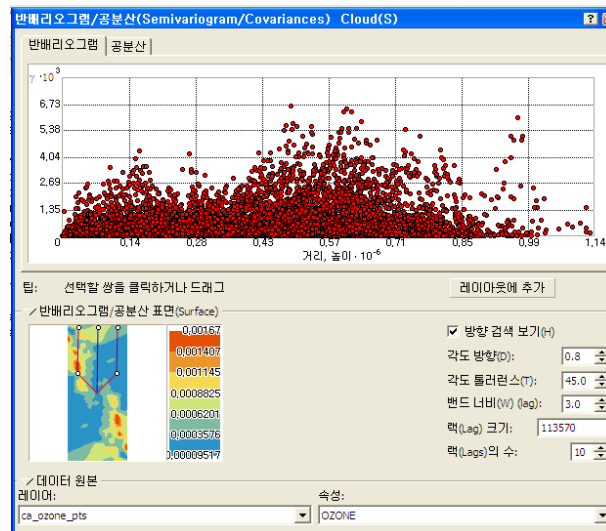


그림 7 반배리오그램/ 공분산 Cloud

- 데이터 셋 안에서 공간적 자기상관²에 의한 위치의 특징을 확인하거나 데이터 셋 밖에서 관찰 가능
- 방향 검색에 따른 반배리오그램 연속면 확인 가능

¹ 공분산 - 두 변량의 공통된 분산(통계에서 각 변량의 값이 평균값과 차이)을 의미

² 공간적 자기 상관 - 서로 다른 2가지 변량 사이의 상관 관계를 상호 상관 관계라 하고, x(t)를 하나의 임의 프로세스로 하여 시각 t1일 때의 값 x(t1)과 시각 t2일 때의 값 x(t2) 사이에 존재하는 상관. 자기 상관을 표시하는 것으로 자기 상관 함수라는 것이 있으며, x(t1)×x(t2)의 조화 평균으로 정의된다

반배리오그램 cloud의 값은 한 쌍의 위치들간의 방향과 거리를 기반으로 만들어진다. 또한 이 값으로 평균값이나 매끄러운 반배리오그램 연속면을 만들 수 있다. 반배리오그램 연속면의 범위는 락(lag)의 크기나 개수에 조종된다.

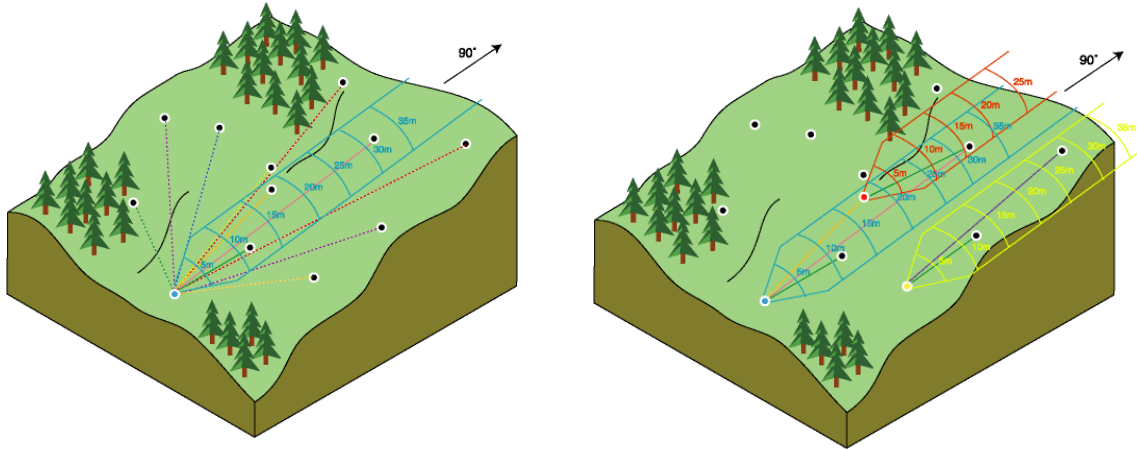


그림 8 방향 검색에 대한 예

보로노이 맵(Voronoi Map)

데이터 탐색(Explore Data) -> 보로노이 맵(Voronoi Map)

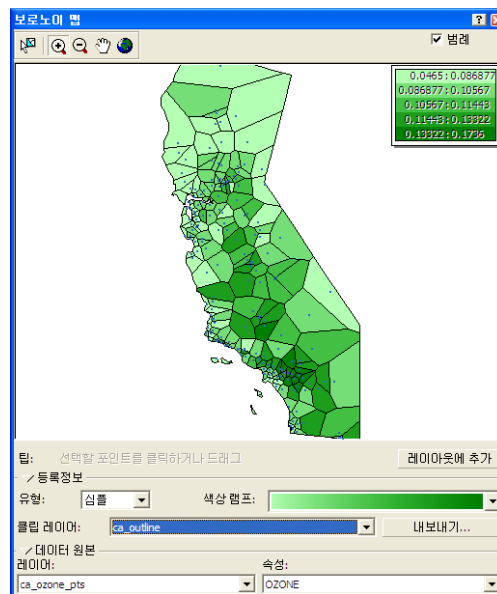


그림 9 보로노이 맵

- 보로노이 맵은 표본이 되는 포인트의 위치 주위에 연속되는 폴리곤 생성

보로노이 폴리곤은 Thiessen, Dirichlet 폴리곤이라고 불리기도 한다. 폴리곤 내의 어떠한 위치에서도 폴리곤 내부에 위치한 한 점까지의 거리가 다른 폴리곤 내에 위치한 거리보다 가깝도

록 폴리곤의 경계가 설정된다. 클립 레이어 설정으로 레이어의 경계에 맞게 만들 수도 있고 나온 결과물을 내보내기를 통해 셰이프 파일로 내보낼 수 있다.

보로노이 맵을 만들 때 다양한 통계기법(심플, 평균, 모드, 클러스터³, 엔트로피⁴, 중앙값, 표준편차, IQR⁵) 이 적용된 맵 생성도 가능하다.

교차공분산 Cloud(Crosscovariance Cloud)

데이터 탐색(Explore Data) -> 교차공분산 Cloud(Crosscovariance Cloud)

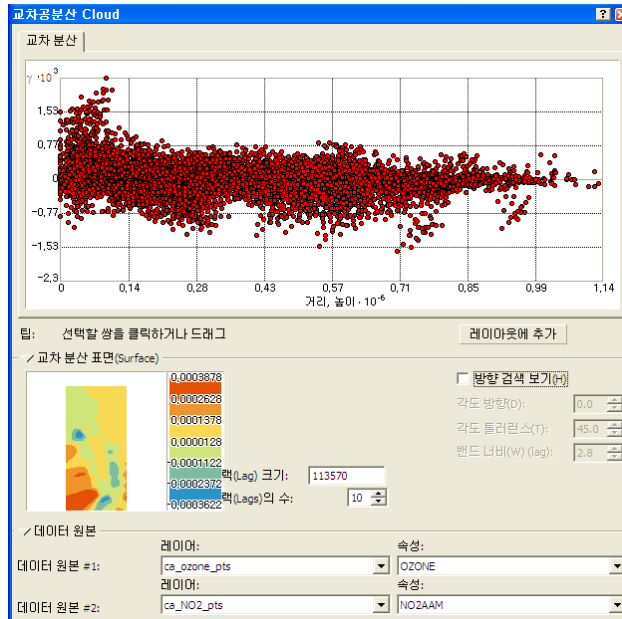


그림 10 교차공분산 cloud

- 두 데이터 셋간 공간적 자기상관의 지역적 특징을 확인
- 빨간 점은 데이터 셋의 값을 표현하는 것으로 ArcMap과 연결되어 빨간 점이 선택 시 ArcMap에도 같이 선택
- 방향 검색을 통해 원하는 방향에서 바라보는 데이터 연속면 확인이 가능

³ Cluster - 모든 셀을 다섯 개의 클래스로 분류하여 만일 이웃에 접한 셀들과 값이 다르면 셀의 색깔을 흐리게 표현하여 구분
⁴ Entropy - 모든 셀을 데이터 값의 자연스러운 그룹화(분위수) 기반으로 다섯 개의 클래스로 구분하여 이웃한 셀들과 계산($Entropy = - \sum (p_i * \text{Log } p_i)$)하여 나타냄
⁵ Interquartile Range - 사분위수범위라고 하는데, 세번째 사분위수와 첫번째 사분위수의 차이. 다시 말해 중심에서 50%에 해당하는 수의 범위를 말함.

서브셋 생성(Create Subsets)

- 서브 셋으로부터 만들어진 지리통계 레이어의 유효성 검사 실행

출력 지표의 질을 가장 엄격한 확인할 수 있는 방법은 특정 위치의 예측된 값과 필드에서 측정된 값을 비교하는 것이다. 항상 연구지역에 대한 독립적이고 유효성 있는 데이터 셋을 가지기는 어렵다. 이를 해결하기 위한 방법으로 기존의 데이터 셋을 두 개 부분으로 나누는 방법이 있다. 한 부분은 모델링을 위해 사용되고 나머지는 부분은 테스트나 결과 지표 유효성 확인을 위해 사용한다.

- 서브 셋으로부터 만들어진 지리통계 레이어의 유효성 검사 실행 방법

서브 셋 생성

1. 지리통계 분석 도구막대에서 서브 셋 생성 클릭
2. 입력 레이어 선택
3. 슬라이드 바를 이용해 두 개로 나눌 Training(교육)/Testing(테스트) 백분율과 서브 셋 이름 설정
4. 완료하면 설정 이름으로 서브 셋 두 개 생성완료 (저장 시에 GDB로 저장)

서브 셋을 이용한 유효성 검사 실행

1. TOC에 만들어진 교육/테스트 레이어 추가
2. 추가된 교육/테스트 레이어 모두 선택
3. 지리 통계 분석 도구막대에서 지리통계 마법사 실행
4. 입력 데이터에 교육용 레이어 선택
5. 알맞은 메소드 선택
6. 오른쪽 탭에 있는 유효성에서 입력 데이터 설정(테스트 레이어)
7. 데이터 셋과 유효성에 설정한 레이어의 같은 속성 선택
8. 매개변수 설정, 교차 검증, 유효성이라는 세 가지 단계를 거쳐 생성
9. 새로 만들어진 레이어 매뉴에서 유효성 추가되어 입력데이터와 결과를 저장위치 설정
10. 생성된 레이어에 Included(포함여부), Predicted(예측치), Error(오류치)라는 필드가 추가된 것을 확인

The image shows two screenshots of ArcGIS software windows. The top window, titled '속성 val', displays a table with columns: OBJECTID*, Shape*, LATITUDE, LONGITUDE, ELEVATION, OZONE, Included, Predicted, and Error. The bottom window, titled '속성 ca_ozone_pts', displays a table with columns: FID, Shape*, LATITUDE, LONGITUDE, ELEVATION, and OZONE. Both tables contain data for several points.

OBJECTID*	Shape*	LATITUDE	LONGITUDE	ELEVATION	OZONE	Included	Predicted	Error
1	포인트	39.1447	-123.2065	194	0.0465	Yes	0.067593	0.021093
2	포인트	39.403	-123.3491	420	0.04969	Yes	0.04969	0
3	포인트	37.7661	-122.3978	5	0.05	Yes	0.080742	0.030742
4	포인트	37.9508	-122.3569	23	0.05799	Yes	0.084298	0.026308
5	포인트	36.6986	-121.6354	36	0.0586	Yes	0.0586	0
6	포인트	35.3668	-120.845	18	0.05889	Yes	0.05889	0
7	포인트	37.8014	-122.2672	7	0.0599	Yes	0.081052	0.021152
8	포인트	37.012	-122.1929	0	0.0606	Yes	0.078965	0.018365

FID	Shape*	LATITUDE	LONGITUDE	ELEVATION	OZONE
0	포인트	39.1447	-123.2065	194	0.0465
1	포인트	39.403	-123.3491	420	0.04969
2	포인트	37.7661	-122.3978	5	0.05
3	포인트	37.9508	-122.3569	23	0.05799
4	포인트	36.6986	-121.6354	36	0.0586
5	포인트	35.3668	-120.845	18	0.05889
6	포인트	37.8014	-122.2672	7	0.0599

그림 11 서브 셋을 이용한 유효성 검사

지리통계 마법사(Geostatistical Wizard)

지리통계 마법사를 이용해 적절한 메소드의 적용과 데이터를 입력하면 지리통계 레이어 생성

메소드는 5가지가 존재 – IDW 내삽법(Inverse Distance Weighting)

- 글로벌 다항식 내삽(Global Polynomial Interpolation)
- 로컬 다항식 내삽(Local Polynomial Interpolation)
- RBF 내삽(Radial Basis Functions)
- 크리깅 내삽(Kriging)
- 코크리깅 내삽(Cokriging)

방법, 데이터의 개수, 유효성 데이터의 유무에 따라 단계가 달라짐. 크게 1~5단계까지 적용

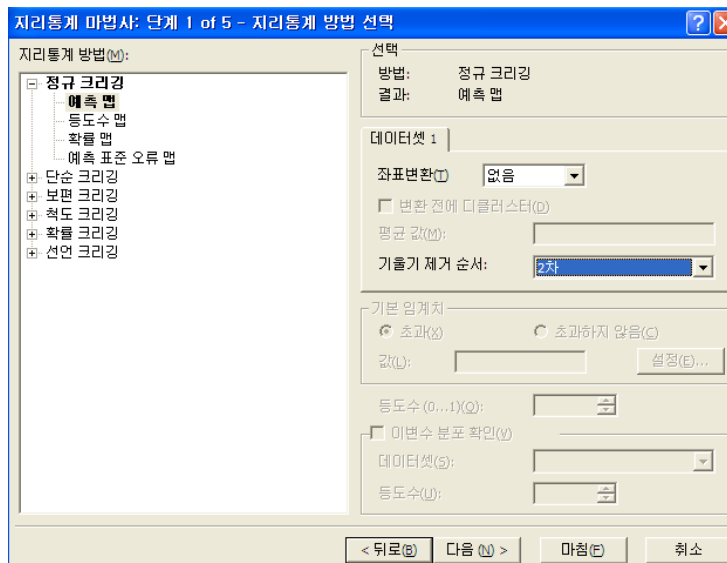
0. 입력 데이터와 방법 선택
1. 지리통계방법 선택
2. 추세제거 중
3. 반배리오그램/공분산 모델링
4. 네이버후드 검색
5. 교차 검색

<Geostatistical Analyst Tutorial Exercise 3으로 단계 설명>

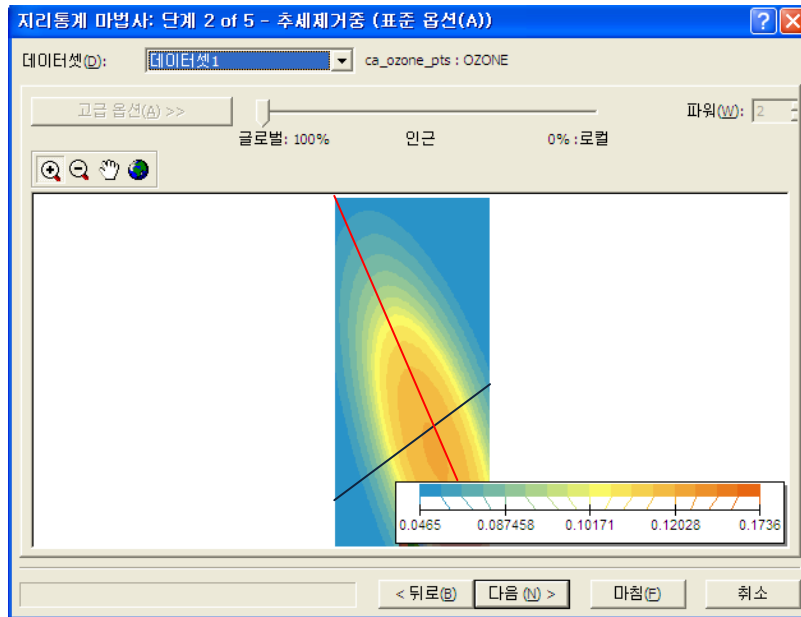
지리통계 마법사를 이용한 지리통계 레이어 생성



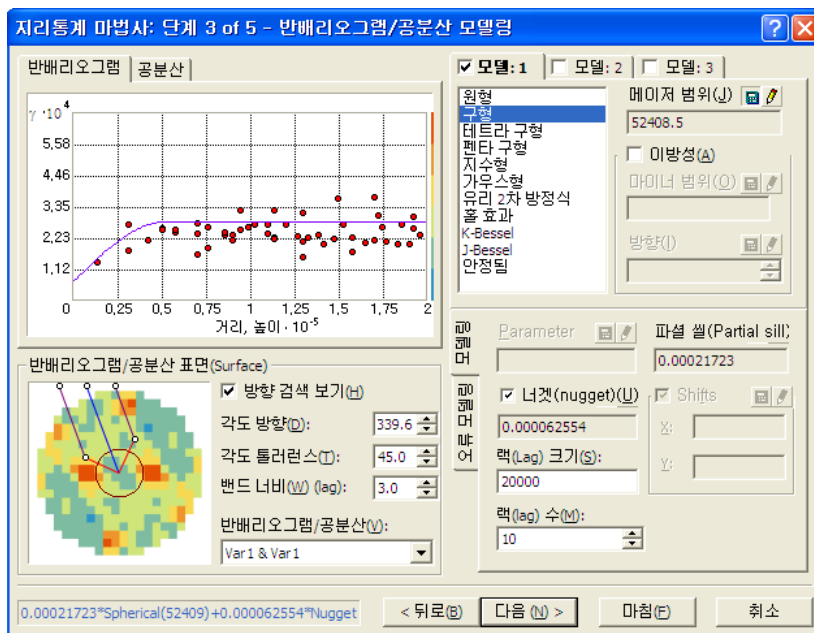
0. 지리통계 마법사 시작 전 지리통계 레이어를 작성하기 위해 방법과 입력데이터를 입력



1. 적절한 지리통계 방법 선택. 어떠한 방법을 선택하느냐에 따라 4단계에서 5단계로 변경



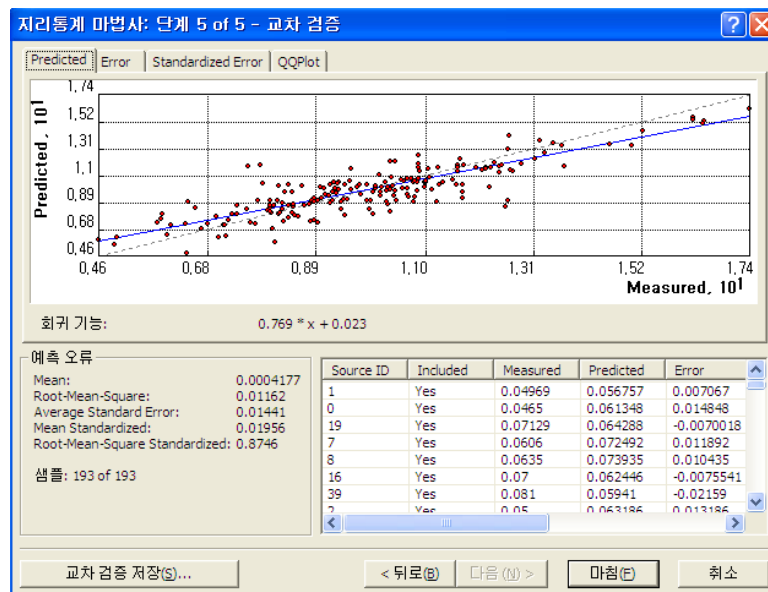
2. 추세제거중 - 검은 선은 남서에서 북서 방향으로 글로벌 추세선, 빨강 선은 북서에서 남서 방향으로 글로벌 추세선 또한, 고급 옵션을 통해 연속면을 변경하고 파워지수 설정이 가능



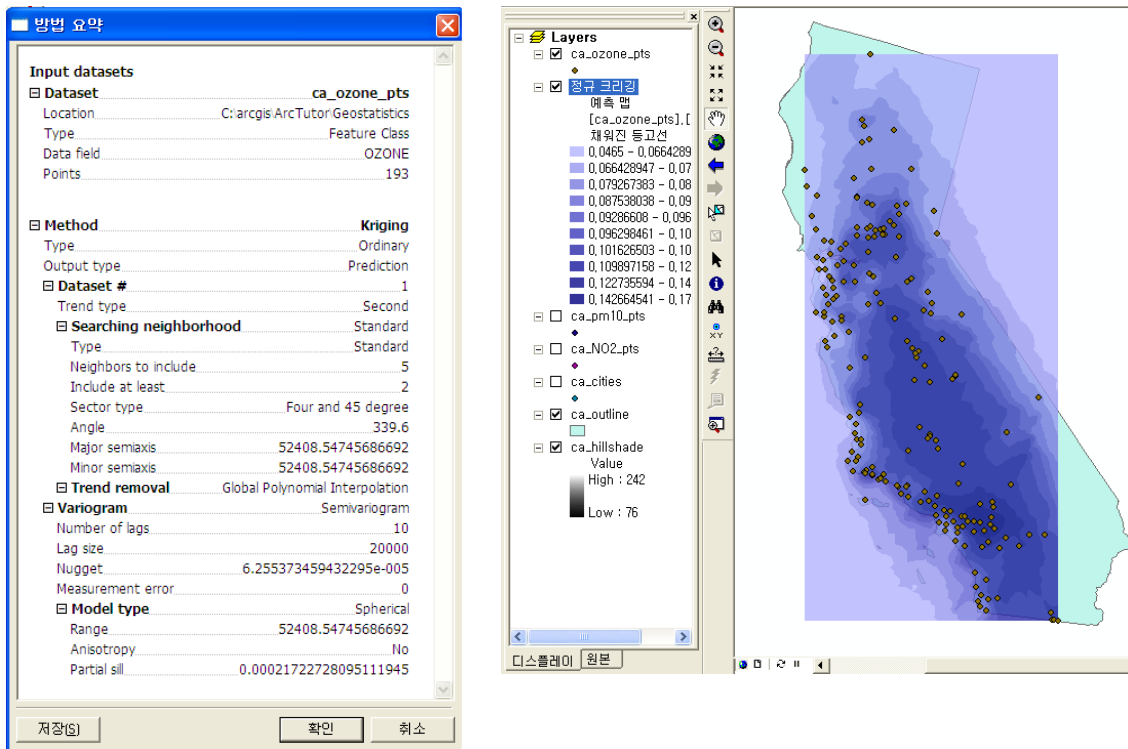
3. 반배리오그램/공분산 모델링



4. 네이버후드 검색



5. 교차검증 - 서브 셋 생성하여 유효성 검사를 통해 만들었던 Included, Predicted, Error 값 생성 가능



☞ 지리통계 마법사 5단계를 통해 모든 설정이 끝나면 방법 요약이라는 창에 설정 수치와 예측 수치가 요약 정리되어 표시, 확인하면 결과가 데이터 뷰에 출력되며 완료

예측 맵과 예측 표준 오류 맵 간의 모델 비교

지리통계 마법사를 통해 생성 예측 맵의 레이어 메뉴에서 예측 표준 오류 맵 생성(Create Prediction Standard Error Map)을 클릭하여 예측 표준 오류 맵 생성 - 반대의 경우도 가능
 둘 중 한 군데 레이어 메뉴에서 비교(compare)를 클릭 - 이 두 개의 맵이 모두 있어야 비교 가능

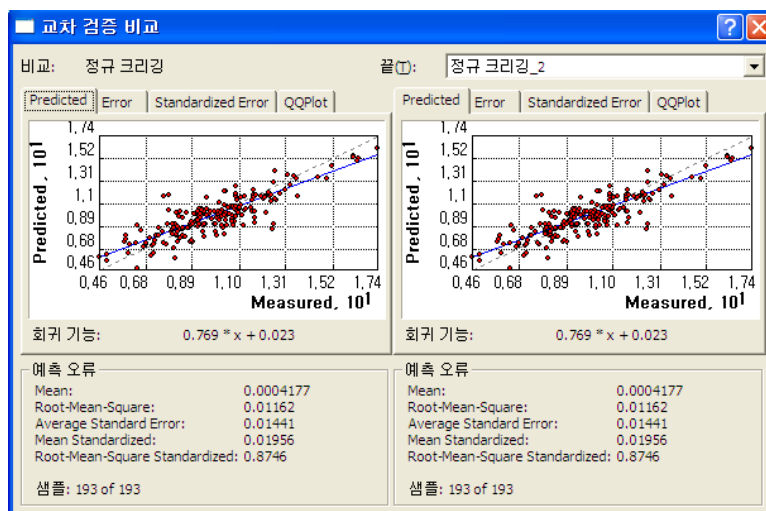


그림 12 교차 검증 비교