

공간보간기법을 이용한 환경자료의 지도화

Mapping of Environmental Data Using Spatial Interpolation Methods

조흥래*, 정종철

Hong-lae Cho*, Jong-chul Jeong

서울대학교 박사수료, 남서울대학교 지리정보공학과 부교수
dshift1@snu.ac.kr, jjc1017@paran.com

요약

환경분야에서 사용되는 대부분의 자료는 공간상 모든 위치에 그 값이 존재하나 모든 지점에서 자료를 획득하는 것이 불가능하므로 몇 개의 대표 지점에서 필요로 하는 자료를 수집한 후 이를 미관측 지역까지 확장하여 사용하게 된다. 관측된 자료를 이용하여 미관측 지점의 값을 예측하는 과정에는 공간보간 기법이 사용되는데, 본 논문에서는 지역경향면 모델, IDW, RBF, 크리깅 등의 공간보간 기법을 서울시의 미세먼지(PM10) 연평균 농도 공간보간에 적용하고 그 정확성을 살펴보았다. 정확성 평가를 위하여 예측값의 범위, RMSE, 평균오차 등을 살펴보았으며 이로부터 크리깅, RBF 기법의 예측 정확도가 높은 것으로 분석되었다.

1. 서론

강수량, 기온, 고도, 오염농도와 같이 환경분야에서 많은 사용되는 데이터는 공간상 연속적으로 분포하는 특징을 갖는데, 이렇게 공간상 모든 지점에서 존재하는 데이터를 지구통계자료(geostatistical data)라 한다. 지구통계자료가 필요할 때, 가장 좋은 방법은 모든 지점에서 필요로 하는 자료를 획득하는 것이겠으나 비용과 시간을 고려할 때 모든 지점에서 원하는 값을 얻는 것은 현실적으로 불가능하다. 따라서 특정 지점을 선정하여 관측값을 얻은 후 이를 이용하여 미관측지점의 값을 예측하는 방법이 사용되는데, 이를 공간보간(spatial interpolation)이라 한다. 지구통계자료에 대한 공간보간을 수행하기 위해서는 먼저 대상 공간자료의 분포 특성에 대한 분석과 사용될 공간보간 기법

의 특징 및 적용범위에 대한 이해가 필요하다. 본 논문에서는 이를 위하여 공간보간 기법에 대한 검토와 실제 공간자료에 대한 적용 가능성을 살펴보았다. 구체적인 공간자료로는 서울시 및 인근지역에서 측정된 연평균 미세먼지(PM10) 농도를 이용하였다.

2. 공간보간

본 연구에서는 공간보간 기법으로 지역경향면 모델(local trend surface model), IDW(inverse distance weighted), RBF(radial basis function), 크리깅(kriging) 방법을 사용하였다.

경향면 모델은 측정지점의 좌표값을 독립변수로 하는 다중회귀모델을 구성하여 미관측 지점의 값을 예측하는 공간보간 기법이다. 이때 다중회귀모델을 구성

하는 함수의 차원에 따라 경향면 모델은 1차원(linear), 2차원(quadratic), 3차원(cubic)으로 구분된다. 지역경향면 모델(local trend surface model)은 경향면 모델과 동일하게 관측지점의 좌표값을 이용하여 회귀식을 구성하여 미관측 지점의 값을 예측하는 방법이나, 경향면 모델이 전체 관측자료를 모두 이용하여 회귀식을 생성하는 것과 달리 지역경향면 모델은 각 예측지점으로부터 가깝게 위치하고 있는 관측지점만을 이용하여 회귀식을 작성하게 된다. 따라서 지역경향면 분석 과정에는 회귀식 구성에 사용될 관측지점 수, 탐색반경, 가중치 등의 지정이 필요하다.

IDW은 공간적으로 인접한 지점 사이의 값은 유사성을 갖는 반면, 두지점 사이의 거리가 멀어질수록 이러한 유사성이 상대적으로 감소하게 된다고 가정한다. 이를 수식으로 표현하면 식 (1)과 같다.

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i) \quad \text{식 (1)}$$

여기서 $\hat{Z}(s_0)$ 는 위치 s_0 에서 우리가 예측하고자 하는 값을 나타내며, N 은 관측지점의 수, λ_i 는 관측지점 s_i 에 사용할 가중치, $Z(s_i)$ 는 관측지점 s_i 에서의 관측값을 나타낸다. 식 (1)에서 가중치 λ_i 는 식 (2)와 같이 거리가 증가함에 따라 그 값이 줄어들게 된다.

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p} \quad \sum_{i=1}^N \lambda_i = 1 \quad \text{식 (2)}$$

위의 식 (2)에서 d_{i0} 는 관측지점 s_i 와 예측지점 s_0 사이의 거리를 나타내며, 지수에 포함된 p 는 거리에 따른 가중치의 변화 정도를 결정한다.

스플라인(spline)이라 불리기도 하는 RBF는 예측된 Z 값이 구성하는 표면의 곡률 총합이 최소가 되도록 미관측지점의

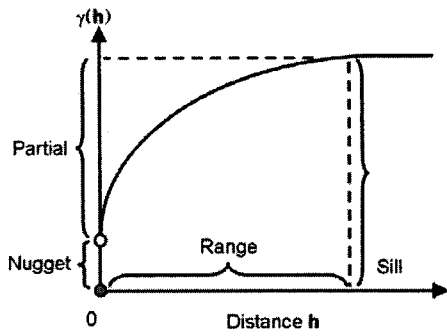
값을 예측하는 보간법이다. 앞에서 설명한 IDW와 비교할 때 RBF의 특징은 IDW가 예측 값이 관측값의 범위를 벗어날 수 없는 것에 반하여 RBF는 관측된 최대값, 최소값보다 예측값이 크거나 작게 나타날 수 있다 (Johnston et al 2001).

크리깅은 미관측 지점의 값을 이미 알고 있는 주위의 값들을 가중선형조합으로 예측하는 기법으로, 기본적인 수식 표현은 앞의 식 (1)과 동일하다. 여기서 가중치는 거리에 따른 함수로 예측값과 참값 사이의 오차가 최소가 되도록 결정된다. 크리깅에서는 공간적 자기상관성을 나타내는 척도인 베리오그램(variogram)을 이용하게 되는데, 베리오그램은 식 (3)과 같이 표현된다.

$$2\gamma(h) = E[(z(x) - z(x+h))^2] \quad \text{식 (3)}$$

여기서 h 는 지연거리 또는 분리거리(lag distance)라 하여 두 자료간의 거리를 나타낸다. 따라서 베리오그램은 분리거리 h 만큼 떨어져 있는 두 지점 사이의 자료값 차이를 제공한 것에 대한 기대값이라 할 수 있다. 베리오그램의 반에 해당하는 값을 반베리오그램(semivariogram)이라 하며 분리거리 h 만큼 떨어진 자료의 개수가 n 개 일 때, 반베리오그램은 식 (4)와 같이 계산된다. 식 (4)에서 계산된 반베리오그램을 분리거리에 따른 그래프로 나타내면 다음 [그림 1]과 같다.

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2 \quad \text{식 (4)}$$



[그림 1] 반베리오그램

관측된 자료로부터 계산된 베리오그램은 일반적으로 경험적 베리오그램이라 하는데, 이 경험적 베리오그램은 관측된 자료들만을 대상으로 계산되기 때문에 이산되어 있다. 그러나 공간보간을 위해서는 관측되지 않은 모든 미지의 지점 s_0 에 대한 베리오그램이 필요하며, 따라서 경험적 베리오그램을 함수화한 이론적 베리오그램을 추정하게 된다. 이론적 베리오그램은 대표적으로 구형모델, 지수모델, 가우스모델 등으로 구분할 수 있으며, 각각 식 (5)~(7)과 같이 정의된다. 각 식에서 C_0 는 문턱값, a 는 상관거리, h 는 분리거리를 나타낸다 (최종근 2002).

구형모델: 식 (5)

$$\gamma(h) = C_0 Sph_a(h) = \begin{cases} C_0 [1.5(h/a) - 0.5(h/a)^3], & (h \leq a) \\ C_0, & (h > a) \end{cases}$$

지수모델: 식 (6)

$$\gamma(h) = C_0 Exp_a(h) = C_0 [1 - \exp(-3h/a)]$$

가우스 모델: 식 (7)

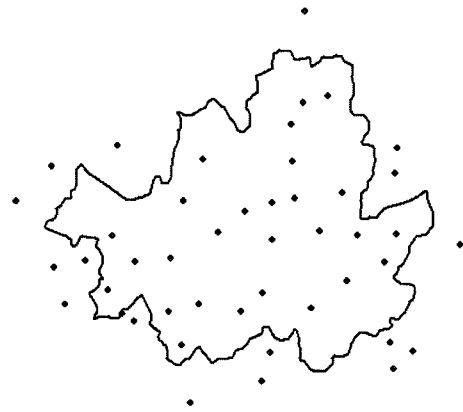
$$\gamma(h) = C_0 Gauss_a(h) = C_0 [1 - \exp(-3(h/a)^2)]$$

이론적 베리오그램이 최종 예측 결과에 직접적인 영향을 미치기 때문에 실험적 베리오그램으로부터 이론적 베리오그램을 찾는 작업은 크리깅 기법에서 매우

중요한 의미를 갖는다.

3. 적용

앞에서 설명한 공간보간 기법을 미세먼지 PM10 연평균농도의 공간보간에 적용하였다. PM10은 공기역학적 직경 10 μm 이하의 입자를 나타내며, 이 크기의 미세먼지는 폐에 침투하는 효율이 가장 높아 건강에 악영향이 큰 것으로 알려져 있다 (김동영, 2006). 공간보간에 사용된 미세먼지 자동측정 지점은 서울시 27곳과 서울 인근지역 18곳을 포함한 총 45개 지점으로, 공간적 분포는 [그림 2]와 같다.



[그림 2] 공간보간에 사용된 미세먼지 PM10 자동측정지점의 공간적 분포

각 공간보간 기법의 정확성 평가와 최적 파라미터 선정 과정에는 2005년 연평균농도를 이용하였다. 사용된 2005년 연평균 PM10 관측값은 최고 88 $\mu g/m^3$, 최저 43 $\mu g/m^3$, 평균 60 $\mu g/m^3$ 의 값을 나타내고 있다. 정확성 평가는 교차검증(cross validation)을 이용하였다.

적용방법을 살펴보면 지역경향면 모델의 경우 2차식의 다중회귀모델을 적용하였다. 이때 거리에 따른 가중치는 지수함수에 의해 부여하였으며, 관측지점의 수를 6~12까지 순차적으로 증가시켜 관측지점수의 변화에 따른 예측치의 정확성을 평가하였다. 미세먼지 자동관측지점이 연구대상지역에 비교적 균일하게 분포하고 있

다고 판단하여 관측지점 선정 시 탐색거리에 제한을 설정하지 않았으며, 가중치 변화 파라미터는 교차검증 과정에서 RMS 에러가 최소가 되는 값을 사용하여 별도로 지정하지 않았다.

IDW는 거리에 따라 가중치의 변화 정도를 나타내는 p 값을 1, 1.5, 2로, 관측지점의 수를 4~14까지 변화하여 적용하였다. 관측지점 탐색 반경은 지역경향면 모델에서처럼 제한을 두지 않았다.

RBF는 적용하는 ϕ 함수에 따라 regularized spline 함수와 tension 함수를 사용하는 경우로 구분하여 적용하였다. 이때 사용되는 관측지점 수는 4~20까지 순차적으로 증가시켜 정확성을 평가하였다. 관측지점 탐색 반경은 제한을 두지 않았으며, ϕ 함수에 포함된 텐션 파라미터는 교차검증 과정에서 RMS 에러가 최소가 되는 값을 사용하였다.

크리깅의 경우 이론적 베리오그램 산정을 위해 구형, 지수형, 가우스형 베리오그램 함수를 각각 적용하였다. 이때 분리거리를 각각 2.0 km, 2.5 km, 3.0 km, 3.5 km로 구분하여 분리거리 크기에 따른 예측 정확도를 살펴보았다. 각 분리거리별 개수는 분리거리 크기와 개수의 곱이 연구범위 내 관측소가 서로 떨어진 최대 거리인 40 km를 넘지 않으면서 최대거리의 반인 20 km를 초과하는 범위에서 선정하였다 (Johnston et al. 2001). 이에 따라 분리거리가 2.0 km인 경우 10~20 개, 2.5 km인 경우 8~16 개, 3.0 km인 경우 7~14 개, 3.5 km의 경우 7~12 개의 분리거리를 사용하였다. 크리깅에 사용되는 관측지점 수는 20 곳이 되도록 하며, 탐색반경은 분리거리 크기*분리거리 개수로 지정하여 반베리오그램 값을 계산하였다. 일반 크리깅 기법을 사용하기 위해서는

먼저 대상지역 전체에 걸친 미세먼지의 추세를 제거할 필요가 있으며, 이를 위하여 x, y 위치에 의한 2차식을 적용하여 대상지역에 전체에 걸친 경향을 제거한 후 일반 크리깅을 적용하였다. 각 보간기법의 적용방법과 평가회수를 정리하면 [표 1]과 같다.

4. 분석결과

각 공간보간 기법에 포함된 파라미터 값을 변화하여 PM10에 대한 공간보간을 실시한 후 각 기법의 정확성을 평가하였다. 각 공간보간 기법별로 가장 높은 정확성을 보인 경우에 대한 결과와 예측값에 대한 통계치는 [그림 3]과 [표 2]와 같다. [표 2]에서 MIN, MAX 값은 서울시를 대상으로 한 값이며 RMSE, 평균오차는 서울 인근까지 포함한 전체 보간 대상지역에 대한 결과이다.

[표 2] 공간보간 기법별 가장 높은 정확성을 보인 방법에서의 예측결과 통계치

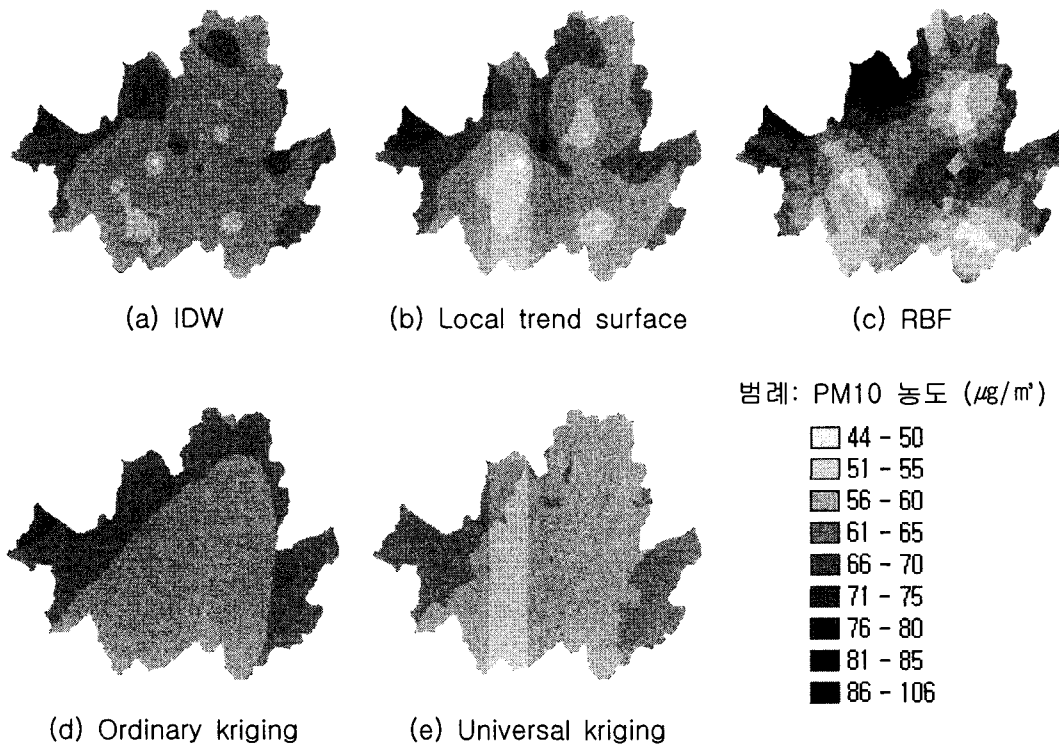
구분	MIN	MAX	RMSE	평균오차
관측값	43	74		-
IDW	44	74	9.73	-1.00
RBF	43	74	9.88	-0.35
LTS	44	106	13.89	0.97
OK	56	67	9.44	-0.14
UK	56	65	9.36	-0.47

RMSE를 기준으로 살펴볼 경우 크리깅 기법이 여타 다른 공간보간 기법에 비하여 좋은 결과를 보여주었다. 일반 크리깅의 경우 RMSE가 9.42~9.36의 범위를 보였으며, 정규 크리깅의 경우에는 9.76~9.46 사이의 범위를 나타내어 일반 크리깅이 정규크리깅에 비하여 우세한 결과를 보여주었다.

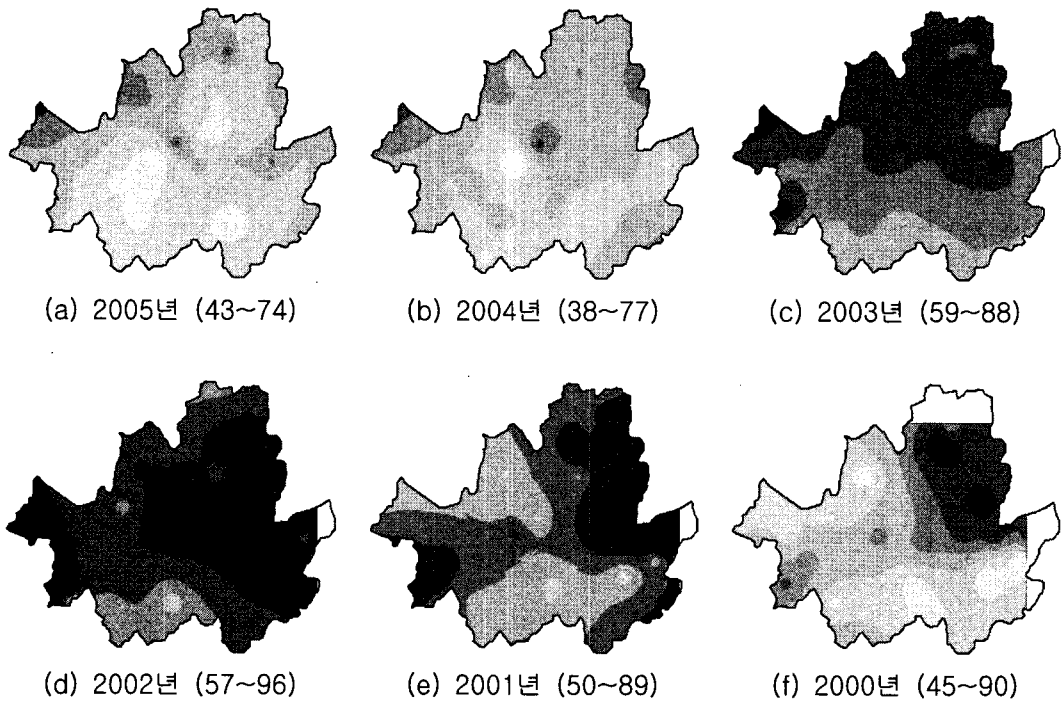
[표 1] 공간보간 기법의 적용방법

공간보간 기법	파라미터	회수
1. Local Trend Surface	samples: 6~12	7
2. IDW	samples: 4~14 / Power: 1, 1.5, 2	33
3. RBF	samples: 4~20 / ϕ 함수: 2 종류*	34
4. Ordinary Kriging	베리오그램: Spherical, Exponential, Gaussian 분리거리(개수): 2.0 km(10~20), 2.5 km(8~16) 3.0 km(7~17), 3.5 km(7~12)	105
5. Universal Kriging		105

*Completely regularized spline, Spline with tension



[그림 3] 각 공간보간 기법에서 가장 높은 예측 정확성을 보인 결과



[그림 4] 연도별 서울시 PM10 연평균농도($\mu\text{g}/\text{m}^3$)의 공간적 분포 (최소값~최대값)
 (범례: 38 - 45 46 - 50 51 - 55 56 - 60 61 - 65 66 - 70 71 - 75 76 - 80 81 - 85 86 - 96)

이론적 베리오그램에 따른 정확성 평가의 경우 일반 크리깅, 정규 크리깅 모두 Gaussian, Spherical, Exponential 모델 순으로 정확성이 높은 것으로 분석되었으며, 베리오그램 모형 선택이 분리거리 크기와 개수의 변화에 비하여 정확성 변화에 크게 기여하는 것으로 분석되었다. RMSE를 기준으로 볼 때, 베리오그램 함수로는 가우스 함수를 사용하며 분리거리 3.0 km, 분리거리개수로 14 개를 지정한 일반 크리깅 기법에서 가장 높은 정확성을 보인 것으로 분석되었다.

IDW의 경우 p 값으로 1, 관측점수로 13개를 사용한 경우가 RMSE 9.73로 가장 높은 값을 보여주었으며, p값과 관측지점수의 변화가 정확도 변화에 미치는 정도가 다른 보간기법에 비하여 작은 것을 보여주었다. IDW가 보여준 RMSE 9.73은 다른 기법과 비교할 때 비교적 높은 정확도를 나타내나, 평균오차가 -1.0을 기록하여 다른 모델에 비하여 편향되어

있을 가능성을 내포하고 있다. 지역경향면 모델의 경우 관측점수로 11 개를 사용한 2차모델이 RMSE 13.89로 가장 좋은 정확성을 보여주었으나, 다른 기법에 비해 예측 정확성이 크게 떨어지는 것으로 분석되었다.

RBF의 경우 tension 함수를 사용한 경우가 regularized spline 함수를 사용한 경우에 비하여 PM10 예측에 효과적인 것으로 분석되었으며, 샘플수로 18 개를 사용한 경우가 RMSE 9.88로 가장 높은 정확도를 보여주었다. RBF의 경우 RMSE와 평균오차를 함께 고려할 때 비교적 높은 예측 정확성을 보여준다.

보간에 사용된 관측값의 범위와 예측값 사이의 관계를 살펴보면 크리깅의 경우 관측값의 범위에 비하여 예측값의 범위가 매우 줄어들어 예측값이 일반화 되어 있음을 확인할 수 있으며, 지역경향면 모델의 경우에는 회귀식 작성에 사용된 샘플수에 민감하게 반응하여 관측점수가

적은 지역의 경우 그 값이 관측값의 범위를 크게 벗어나 예측되는 것으로 분석되었다.

종합적으로 평가할 때, 미세먼지의 미관측지점 값을 예측함에 있어 크리깅 기법과 RBF 기법이 우수한 공간보간 결과를 보여주는 것으로 판단된다. 특히 RBF의 경우 크리깅의 예측값이 일반화되는 것에 비하여 예측 결과가 관측값 범위내에서 고르게 분포하는 것을 보여주며, 이는 PM10 농도의 공간적, 시간적 분포 특성을 파악하는데 큰 도움이 될 것으로 보인다. [그림 4]는 2000~2005년 사이의 연평균 PM10 농도에 대해 RBF 공간보간을 적용한 결과로 이로부터 시공간 변화에 따른 PM10 농도 분포의 경향을 파악함에 있어 RBF를 이용한 공간보간의 적용가능성을 확인할 수 있다.

5. 결론

본 연구에서는 공간보간 기법을 이용하여 PM10 연평균 농도에 대한 공간보간을 실시하였으며, 이로부터 크리깅과 RBF 기법이 미관측 지점의 PM10 농도를 예측함에 있어 효과적임을 살펴보았다. 향후 본 연구에서 사용한 PM10 연평균 농도 이외에 계절별, 일별, 시간별 농도변화에 대한 공간보간을 통하여 미세먼지의 시공간적 변화의 경향성을 파악할 수 있을 것이다. 또한 본 연구에서 사용된 공간보간 기법의 적용과정은 미세먼지만이 아니라 다른 지구통계자료에도 사용될 수 있을 것으로 판단된다.

[참고 문헌]

- 김동영, 2006, 수도권 미세먼지 모델링 연구, 경기개발연구원
최종근, 2002, 공간정보 모델링-크리깅과 최적화 기법, 구미서관
Johnston, K., Ver Hoef, J.M., Krivoruchko, K., Lucas, N. 2001, Using ArcGIS Spatial Analyst, ESRI Press