*The Journal of GIS Association of Korea, Vol. 14, No. 1, pp. 29-41, April 2006*

# 강우자료에 대한 공간보간 기법의 적용

조홍래* · 정종철**

# Application of Spatial Interpolation to Rainfall Data

Hong-lae Cho* · Jong-Chul Jeong**

## 요 약

공간통계자료는 비록 공간상 모든 위치에 그 값이 존재하나, 실제 우리는 모든 지점에서 공간자료를 수집하지 못한다. 따라서 몇 개의 대표 지점에서 자료를 수집한 후 이를 미관측 지역까지 확장하여 사용하게 되는데, 이때 공간보간 기법이 적용된다. 본 논문에서는 다양한 분야에서 매우 중요하게 사용되는 강우자료에 대해 공간보간을 실시하고 그 정확성을 검증하였다. 기상청에서 운영하는 총 373 개의 강우관측소의 2004 년도 연 총강우량에 대해 공간보간을 실시하였으며, 그로부터 이론적 베리오그램으로 가우시안 함수를 사용한 일반 크리깅이 가장 높은 정확성을 보인 것으로 확인되었다. 다른 보간기법에 비해 크리깅 기법은 비교적 정확성이 높은 것으로 분석되었으며, 결과에 있어서는 예측값의 분포 범위가 좁은 일반화의 특성을 갖는 것으로 나타났다. 지역경향면 분석(local trend surface analysis), IDW, RBF의 경우 강우자료에 대한 예측값의 범위가 크리깅에 비해 매우 넓게 분포하는 것으로 나타났으며, 공간적으로는 인접한 지역의 예측값이 크리깅에 비해 급변하는 것으로 분석되었다.

주요어 : 공간 보간, 크리깅, IDW, RBF, 경향면 분석

**ABSTRACT** : Geostatistical data are obtained only at selected sites even though they are potentially available at any location in a continuous surface. Therefore it is necessary to estimate the unknown values at unsampled locations based on observations. In this study we compared the accuracy of 5 spatial interpolation methods: local trend surface, IDW, RBF, ordinary kriging, universal kriging. These interpolation methods were applied to annual rainfall

*서울대학교 환경대학원 박사과정(dshift1@snu.ac.kr)
**남서울대학교 지리정보공학과 부교수(jjc1017@paran.com)

data. As the results of validation tests, universal kriging with gaussian variogram model showed the best accuracy in comparison with other interpolation methods. In the case of kriging, the predicted values were more accurate and within a more narrow range than other methods. In contrast with kriging, local trend surface analysis, IDW and RBF showed the wide range of predicted values and abrupt changes between neighbors.

**Keywords** : Spatial Interpolation, Kriging, IDW, RBF, Trend Surface Analysis

## 1. 서 론

환경 분야에서 사용되는 많은 데이터는 공간상 연속적으로 분포하는 특징을 갖는다. 여기서 공간상 연속적으로 분포한다는 것은 공간상의 모든 지점에서 그 값이 존재한다는 의미를 나타내며 강수량, 기온, 고도, 오염농도 등이 그 구체적인 예라고 할 수 있다. 공간상 연속적으로 존재하는 이러한 데이터를 공간분석 분야에서는 공간통계자료(geostatistical data)라 호칭한다. 공간통계자료가 필요할 때, 가장 좋은 방법은 모든 지점에서 필요로 하는 자료를 획득하는 것이겠으나 비용과 시간 문제로 인하여 모든 지점에서 원하는 값을 얻는 것은 현실적으로 불가능하다. 따라서 많은 경우 특정 지점을 선정하여 관측값을 얻은 후 이를 이용하여 미관측지점의 값을 예측하는 방법이 사용되며, 이러한 작업과정에 공간보간(spatial interpolation)이 이용된다. 대표적인 공간보간 기법으로는 IDW(Inverse Distance Weighted), 스플라인(spline), 크리깅(kriging), 경향면 분석(trend surface model) 등을 예로 들 수가 있을 것이다. 지금까지 환경분야의 자료에 대해서도 이상의 기법들을 이용하여 미관측지점의 값을 예측하려는 다양한 연구들이 수행되었다. 대표적인 공간보간 연구로 대기오염농도 및 기온에 대한 연구를 제시할 수 있다(Kurtzman and Kadmon, 1999; 최용규, 2001; Jang, 2003).

공간통계자료에 대한 보간을 수행하기 위해서는 먼저 대상 공간데이터의 분포특성에 대한 면밀한 분석과 사용될 공간보간 기법의 특징 및 적용범위에 대한 명확한 이해가 필요하다. 그러나 지금까지 환경 분야에서 적용된 공간분석 기법에 대한 연구는 공간보간에 사용될 기법들에 대한 명확한 이해에 기초하지 않고 단순히 공간보간 기능을 제공하는 프로그램의 기본 설정을 이용하여 미관측지점의 값을 예측하는 미흡한 모습을 보여주었다. 따라서 본 논문에서는 공간보간 기법에 대한 면밀한 검토를 수행하고 이를 바탕으로 실제 공간데이터에 대한 적용을 살펴보도록 한다. 구체적인 공간데이터로는 강우 데이터를 사용하였다. 본 논문에서 강우 자료를 공간보간의 대상 데이터로 선정한 이유는 강우 자료가 수문, 수질, 대기, 생태 등 다양한 분야에서 매우 중요한 인자로 사용되기 때문이다.

## 2. 공간 보간

공간 보간에 사용되는 방법들은 크게 ① 미관측지점의 값을 추정하기 위하여 모든 관측 지점의 값을 사용하는 전역적(global) 방법과 인접한 주변 지역의 값만을 사용하는 지역적(local) 방법, ② 보간 결과 관측 지점에서의 예측값이 원 측정값과 동일하게 나타나는 정확(exact) 방법과 원 조사값과 다르게 나타나는 비정확(inexact) 방법, ③ 예측값의 오차를 평가할 수 있는 통계적(stochastic) 방법과 그렇지 않은 결정론적(deterministic) 방법, ④ 인접 지점 사이의 예측값의 변화 정도에 따라 급변적(abrupt) 방법과 완변적(smooth) 방법으로 구분할 수 있다. 공간 보간에 많이 사용되는 몇 개의 보간 방법을 위의 4 가지 구분 방법에 따라 정리하면 다음 <표 1>과 같다.

### 2.1 공간모형

s를 공간상의 위치벡터라 하고, Z(s)를 공간 위치 s에서 강우량, 고도, 기온 등과 같이 우리가 관심 있어 하는 현상을 유발하는 공간 확률과정이라 할 때, 이 공간 확률과정은 식 (1)과 같이 4 가지의 요소로 구성된다(Cressie, 1993).

$$Z(s)=\mu(s)+w(s)+\eta(s)+\varepsilon(s) \qquad 식(1)$$

위 식 (1)에서 $\mu(s)$는 대규모변동(largescale variation)으로 전지역적 추세를 나타내는 결정적 요소를 의미하며 $w(s)$는 소규모변동(small scale variation)으로 결정적 추세 주변의 확률적 변동을 나타낸다. $\eta(s)$는 미세규모변동으로(microscale variation) 수집된 거리보다 더 작은 거리에서의 변동을 나타내며 $\varepsilon(s)$는 측정오차(measurement error)를 나타낸다(Lee, 2003). 위 식을 전지역적 추세를 나타내는 대규모변동과 나머지 오차항의 결합 형태로 재구성할 경우 식 (2)와 같이 표현할 수 있다(Cressie, 1993).

$$Z(s)=\mu(s)+\delta(s) \qquad 식(2)$$

### 2.2 경향면 분석

경향면 분석(trend surface analysis)은 측정지점의 좌표값을 이용하여 다중회귀모델을 구성한 후 구성된 다중회귀모델을 이용하여 미관측 지점의 값을 예측하는 공간보간 방법이다.

<표 1> 공간보간 기법의 분류

| Global | | Local | |
|---|---|---|---|
| Deterministic | Stochastic | Deterministic | Stochastic |
| Trend surface (inexact) | Transfer Function (inexact) | Local trend surface (inexact)<br>Thiessen (exact)<br>Inverse distance weighted (exact)<br>Splines (exact) | Kriging (exact) |

경향면 분석은 공간 확률과정을 나타내는 식 (2)에서 전지역적 추세를 나타내는 $\mu(s)$를 식 (3)과 같이 (x, y) 좌표값의 함수로 표현한다.

$$Z(s_i) = \mu(s_i) + \delta(s_i) = \mu(x_i, y_i) + \delta(s_i) \quad \text{식(3)}$$

이때 μ를 구성하는 함수의 차원에 따라 경향면 분석은 1차원(linear), 2차원(quadratic), 3차원(cubic)으로 구분할 수 있으며, 각 회귀식을 구성하는 계수는 최소자승법을 사용하여 추정된다.

## 2.3 지역경향면 분석

지역경향면 분석(local trend surface analysis)은 경향면 분석과 동일하게 관측지점의 좌표값을 이용하여 회귀식을 구성한 후 미관측 지점의 값을 예측한다. 그러나 전체 관측데이터를 모두 이용하여 회귀식을 생성하는 경향면 분석과 달리 지역경향면 분석은 각 지점에 대하여 가깝게 이웃하고 있는 관측지점만을 대상으로 회귀식을 작성하게 된다. 따라서 전체지역에 대해 단일한 회귀식을 생성하지 않는 대신, 지역경향면 분석은 각 지점에 대한 개별적이 회귀식을 생성한다. 지역경향면 분석은 지수함수를 비롯한 다양한 형태의 함수를 이용하여 예측지점과 관측지점 사이의 인접 정도에 따라 서로 다른 가중치를 부여할 수 있다. 이렇게 관측지점별 가중치를 산정한 후 회귀식을 구성하는 계수는 가중최소자승법(WLS; Weighted Least Squares) 방법을 이용하여 찾아내게 된다.

## 2.4 IDW

IDW(Inverse Distance Weighted)의 기본 가정은 공간적으로 인접한 지점 사이의 값은 공통된 위치요인으로 인하여 유사성을 갖게 되는 반면에 두 지점 사이의 거리가 증가할수록 이러한 유사성은 상대적으로 감소하게 된다는 것에 기초한다. IDW에 대한 기본 수식은 식 (4)와 같이 표현된다.

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \quad \text{식(4)}$$

여기서 $\hat{Z}(s_0)$는 위치 so에서 우리가 예측하고자 하는 값을 나타내며, N은 관측지점의 수, λi는 관측지점 si에 사용할 가중치, Z(si)는 관측지점 si에서의 관측값을 나타낸다. 위 식 (4)에서 가중치 λi는 식 (5)와 같이 거리가 증가함에 따라 그 값이 줄어들게 된다.

$$\lambda_i = d_{io}^{-p} / \sum_{i=1}^{N} d_{io}^{-p} \qquad \sum_{i=1}^{N} \lambda_i = 1 \quad \text{식(5)}$$

식 (5)에서 dio는 관측지점 si와 예측지점 so사이의 거리를 나타내며, 지수에 포함된 p는 거리에 따른 가중치의 변화 정도를 결정한다.

## 2.5 RBF

RBF(Radial Basis Function)는 예측된 Z값이 생성하는 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. RBF는 고도 데이터와 같이 공간상 완만하게 변화하는 데이터의 보간에 적합한 것으로 알려져 있다(Johnston et al., 2001). RBF 보간법을 식으로 표현하면 다음 식 (6)과 같다.

$$\hat{Z}(S_0)=\sum_{i=1}^{N} w_i \phi(\| s_i - s_o \|) + w_{n+1} \qquad \text{식(6)}$$

여기서 $\phi$는 radial basis function을 나타내며, $\|s_i - s_0\|$는 예측지점 so와 각 관측지점 si와의 거리 r을 나타낸다. {ωi: I = 1,2, …,n+1}는 각 관측값에 대한 가중치를 나타내며, ωn+1은 평향(bias) 파라미터를 뜻 하다. RBF는 식 (6)에 포함된 radial basis function $\phi$를 어떻게 구성하는 가에 따라 completely regularized spline function, spline with tension function, multiquadric function, inverse multiquadric function, thinplate spline function 등으로 구분된다 (Johnston et al., 2001).

## 2.6 크리깅

크리깅(kriging)은 미관측 지점의 값을 이미 알고 있는 주위의 값들을 가중선형조합으로 예측하는 기법으로, 기본적인 수식 표현은 앞의 식 (4)와 동일하다. 여기서 가중치는 거리에 따른 함수로 예측값과 참값 사이의 오차가 최소가 되도록 결정되며 많은 경우에 추정값이 편향되지 않아야 한다는 조건을 추가하여 그 값을 결정한다(최종근, 2002). 크리깅 기법에는 단순 크리깅, 정규 크리깅, 일반 크리깅, 구역 크리깅, 공동 크리깅 등 다양한 종류가 있다.

### 2.6.1 기본가정과 개념

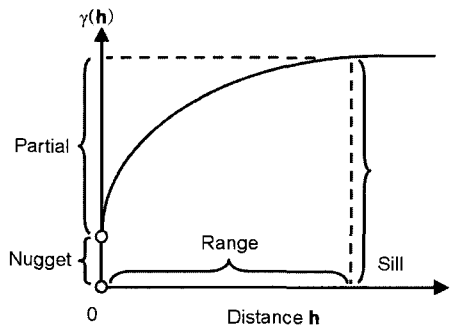
시계열 데이터가 일반적으로 시간적으로 서로 상관성을 갖는 것과 마찬가지로 공간상에서 수집된 데이터 또한 공간적 상관성을 갖게 된다. 따라서 공간자료에 대한 분석을 위해서는 이러한 공간적 상관성에 대한 이해가 필요하게 되는데, 이에 대한 대표적인 척도로 베리오그램이 사용된다. 베리오그램은 식 (7)과 같이 정의된다.

$$2\gamma(h) = E[(Z(s) - Z(s+h))^2] \qquad \text{식(7)}$$

여기서 h는 지연거리 또는 분리거리(lag distance)라 하여 두 자료간의 거리를 나타낸다. 따라서 베리오그램은 분리거리 h만큼 떨어져 있는 두 지점 사이의 자료값 차이를 제곱한 것에 대한 기대값이라 할 수 있다. 베리오그램의 반에 해당하는 값을 반베리오그램(semivariogram)이라 하며 분리거리 h만큼 떨어진 자료의 개수가 n개 일 때. 반베리오그램은 식 (8)과 같이 계산된다.

$$r(h) = \frac{1}{2n}\sum_{i=1}^{n}[Z(s_i) - Z(s_i+h)]^2 \qquad \text{식(8)}$$

식 (8)에서 계산된 반베리오그램을 분리거리에 따른 그래프로 나타내면 다음 [그림 1]과 같다(Johnston et al., 2001).




[그림 1] 반베리오그램

[그림 1]에서 분리거리가 증가하여 일정한 거리 이상이 되면 반베리오그램이 일정한 값을 나타내게 되는데 이때의 분리거리를 상관거리(range), 반베리오그램 값을 문턱값(sill)이라 한다. 문턱값은 자료들 사이에 아무런 관계가 없음을 나타내는 것으로 자료의 분산을 뜻한다(최종근, 2002). 분리거리가 0인 경우 반베리오그램은 이론적으로 0으로 나타나야 하지만, 실제 반베리오그램 그림에서 분리거리 0일 때의 값을 유추할 경우 분리거리 0에서도 반베리오그램은 0이 아닌 상수값을 나타내게 된다. 이를 너깃(nugget)이라 하며, 미세규모변동이나 측정오차를 나타낸다.

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 s0에 대한 베리오그램이 필요하며, 따라서 경험적 베리오그램을 함수화한 이론적 베리오그램을 추정하게 된다. 이론적 베리오그램은 대표적으로 구형모델, 지수모델, 가우스모델 등으로 구분할 수 있으며, 각각 식 (9)~(11)와 같이 정의된다(최종근, 2002). 각 식에서 C0는 문턱값, a는 상관거리, h는 분리거리를 나타낸다.

• 구형모델: 식(9)

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

• 지수모델: 식(10)

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

• 가우스 모델: 식(11)

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

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

### 2.6.2 크리깅의 종류

□ 단순 크리깅 (Simple Kriging)

단순 크리깅은 공간적 확률과정이 2차 정상성(stationarity)을 만족하며, 평균값은 우리가 이미 알고 있다고 가정한다. 여기서 2차 정상성은 평균이 위치에 무관하게 상수로 일정하다는 것과 공분산이 자료의 위치에 의존하지 않고 단지 주어진 자료 사이의 거리에 따라 변화하는 함수임을 의미한다. 이

상의 가정을 전제로 단순 크리깅은 예측오차를 최소로 하는 가중치를 구하여 공간보간을 수행하게 된다. 예측오차를 최소로 하는 가중치는 오차분산을 이용하여 계산할 수 있다. 단순 크리깅은 크리깅 기법 중 가장 간편한 방법이라 할 수 있다. 그러나 평균을 이미 알고 있을 때 적용 가능한 방법으로 많은 경우 사전에 평균을 알 수 있는 경우가 드물어 적용에 한계를 갖는다.

□ 정규 크리깅 (Ordinary Kriging)

정규 크리깅은 모집단의 평균을 사전에 알 수 없지만, 그 값은 고정되어 있다는 것으로 가정한다. 단순 크리깅은 추정식이 편향되어 있어 추정식의 평균이 모집단의 평균과 일치하는 않는 문제점이 있다(최종근, 2002). 이를 극복하기 위해 정규 크리깅은 추정식이 평향되지 않기 위한 제한조건을 부여한다. 평향은 모집단의 평균과 추정식의 평균 차이로 정의되며 이 차이가 없을 평향되어 있지 않다고 할 수 있다. 이로부터 평향되어 있지 않기 위해서는 가중치의 총합이 1이 되어야 한다는 제약조건이 유도되며 이 제약조건 하에서 오차분산이 최소가 되도록 가중치를 구하고 이를 이용하여 미지의 값을 예측하게 된다. 제약조건 하에서 최소·최대를 구하는 문제는 라그랑제 인자법(Lagrange parameter method)에 의하여 계산될 수 있다.

□ 일반 크리깅 (Universal Kriging)

2차 정상성은 모든 공간 프로세스를 유발하는 확률과정에서 만족되는 것은 아니다. 2차 정상성은 모집단의 평균이 위치에 상관없이 같은 값을 갖는다고 가정하나 실세계의 자료들의 평균은 공간적 추세가 존재하는 경우가 많다. 이럴 경우 해당 자료는 2차 정상성을 더 이상 만족하지 않으며, 공간적 추세를 무시하고 크리깅을 기법을 적용할 경우 예측값이 잘못될 가능성이 커지게 된다. 이러한 문제점을 극복하기 위하여 일반 크리깅에서는 평균이 갖는 공간적 추세를 제거한 잔차를 이용하여 베리오그램을 추정한 후 이를 이용하여 미관측 지점의 값을 예측하게 된다.

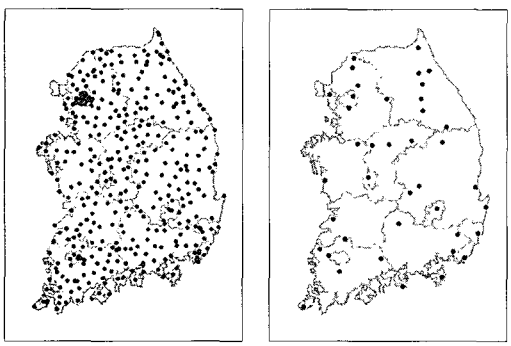
## 3. 적 용

그 동안 유역 내의 총강우량을 산정하기 위해 산술평균법, 티센(Thiessen) 기법, 등우선법 등이 사용되었다. 그러나 산술평균법은 유역 내의 강우관측소의 정보를 단순히 산술평균하는 방법으로 대상유역 및 호우의 물리적 특성을 전혀 반영하지 않아 타 기법에 비해 강우를 과소 또는 과대 평가하는 경향이 있으며, 티센 기법의 경우 관측소의 지배면적에 민감하게 반응하고 등우선법은 일반적으로 유역경계선상에서 등우선이 걸치는 경우 경계처리가 임의적일 수밖에 없는 단점을 갖는다(최용규, 2001). 따라서 미관측 지점의 강우량을 계산하기 위해 합리적인 방법이 요구되며, 본 논문에서는 이를 위하여 공간분석 분야에서 활발하게 사용되고 있는 공간보간 기법을 평가하였다.

### 3.1 연구범위

본 연구에서는 기상청에서 운영하는 강

우관측소 중 내륙에 위치한 총 413 개를 이용하여 공간보간을 실시하였다. 413 개의 관측소 중 10%에 해당하는 40 개의 관측소는 공간보간의 정확성 검정에 사용하였으며, 실제 공간보간 과정에는 총 373 개의 관측소가 사용되었다. 사용된 강우관측소의 위치 분포는 [그림 2]와 같다.



[그림 2] 강우관측소 분포(좌: 공간보간에 사용된 강우관측소, 우: 검증에 사용된 강우관측소)

[그림 2]에서 왼쪽 그림은 미관측 지점의 연 총강우량을 예측하는데 사용된 강우관측소를 나타내며, 오른쪽은 예측 정확성을 검정하기 위해 사용된 강우관측소를 나타낸다. 예측 정확성을 검증하기 위해 사용된 강우관측소는 기상청에서 운용하고 있는 413 개의 강우관측소 중 40 개를 무작위로 선정하여 구성하였다. 공간보간 대상이 되는 강우자료로는 2004 년도 연 총강우량을 사용하였으며, 사용된 강우자료의 통계값은 <표 2>와 같다. 강우 예측에 사용된 총 373 강우관측소에서 측정된 2004 년도 연 총강우량은 최소 765 ㎜, 최대 2,547 ㎜, 평균 1,448 ㎜인 것으로 나타났으며, 정확성 검증에 사용된 40 개의 강우관측소는 최소 887 ㎜, 최대 1,838 ㎜, 평균 1,381 ㎜인 것으로 조사되었다.

<표 2> 2004년도 연강우량 통계치

(단위 : ㎜)

| Dataset | Num. | Min | Max | Mean | Std Dev. |
|---|---|---|---|---|---|
| Training | 373 | 765 | 2547 | 1446 | 225 |
| Test | 40 | 887 | 1838 | 1381 | 197 |

### 3.2 방법

앞에서 살펴본 공간보간 기법 중 지역경향면 분석, IDW, RBF, 정규 크리깅, 일반 크리깅 기법을 적용하였다. 단순 크리깅의 경우 평균값을 알고 있다는 가정이 강우자료에 적용하기에 현실성이 결여되어 적용 기법에서 제외하였다.

□ 지역경향면 분석

2차식의 다중회귀모델을 적용하였다. 이때 거리에 따른 가중치는 지수함수에 의해 부여하였으며, 관측지점의 수를 8, 10, 12, 15, 20으로 변환하여 관측지점의 변화에 따른 예측치의 정확성을 평가하였다. 강우관측지점이 연구 대상지역에 비교적 균일하게 분포하고 있다고 판단하여 관측지점 선정 시 탐색거리에 제한을 설정하지 않고 주어진 관측지점의 수를 모두 포함하여 공간보간을 실시하였으며, 가중치 변화 파라미터는 교차검증 과정에서 RMS 에러가 최소가 되는 값을 사용하여 별도로 지정하지 않았다.

□ IDW

거리에 따라 가중치의 변화정도를 나타

내는 p 값(식 5)과 관측점 수를 각각 1, 1.5, 2 및 8, 10, 12, 15, 20으로 구분하여 총 15 개(3×5)의 방법을 적용하였다. 관측지점 탐색 반경은 지역추세모델에서처럼 제한을 두지 않았다.

□ RBF

RBF는 적용하는 $\phi$함수(식 6)에 따라 'completely regularized spline'과 'spline with tension' 함수로 구분하여 평가하였다. 이때 사용되는 관측지점 수는 각각 8, 10, 12, 15, 20으로 구분하여 관측지점수 변화에 따른 정확성 변동을 살펴보았다. 관측지점 탐색 반경은 제한을 두지 않았으며, $\phi$함수에 포함된 텐센 파라미터는 교차검증 과정에서 RMS 에러가 최소가 되는 값을 사용하였다.

□ 크리깅

이론적 베리오그램 산정을 위해 구형, 지수형, 가우스형 베리오그램 모델을 각각 적용하였다. 이때 분리거리(lag size)를 각각 20 km, 25 km, 30 km, 35 km로 구분하여 분리거리 크기에 따른 예측 정확도를 살펴보았다. 각 분리거리별 래그 개수는 분리거리 크기와 개수의 곱이 연구범위 내 강우관측소가 서로 떨어진 최대 거리인 400 km를 넘지 않으면서 최대거리의 반인 200 km를 초과하는 범위에서 선정하였다(Johnston et al., 2001). 이에 따라 분리거리가 20 km인 경우 래그 개수를 15 개, 25 km인 경우 12 개, 30 km인 경우 10 개, 35 km인 경우 8 개를 사용하였다. 크리깅에 사용되는 관측지점 수는 20 곳이 되도록 하며, 탐색반경은 분리거리 크기x분리거리 개수로 지정하여 반베리오그램 값을 계산하였다. 일반 크리깅 기법을 사용하기 위해서는 먼저 유역전체에 걸친 강우사상의 추세를 제거할 필요가 있으며, 이를 위하여 위치(x, y)에 의한 2차식을 적용하여 유역 전반에 걸친 경향을 제거한 후 일반 크리깅을 적용하였다. 이상의 적용 기법을 정리하면 <표 3>과 같다.

## 4. 분석결과

<표 3>과 같이 각 공간보간 기법에 포함된 파라미터를 변화하여 보간을 실시한 후 보간 과정에 사용되지 않은 40 개의 강우관측소의 자료를 이용하여 각 기법의 정확성을 검증 하였다. 정확성 검증 결과는 <표 4>와 같다.

전체적으로 살펴볼 경우 크리깅 기법이 여타 다른 공간보간 기법에 비하여 월등히 좋은 결과를 보여주었으며, 이 중 분리거리가 35 km, 베리오그램 함수로 가우스 함수를 사용한 일반 크리깅 기법이 가장 정확한 보간을 수행한 것으로 나타났다(<표 4>). 정규 크리깅과 일반 크리깅을 비교할 경우 대체로 일반 크리깅이 미세하나 좋을 결과를 보여주었다. 일반 크리깅이 정규 크리깅에 비하여 좋은 결과를 보인 이유는 정규 크리깅이 2차 정상성 가정에 의하여 모집단의 평균이 공간적 위치에 관계없이 일정하다고 전제하나, 실제 강우자료가 그러한 특성을 갖지 않기 때문인 것으로 판단된다. 즉 강우자료의 경우 위치에 따른 공간적 추세를 갖게 되며, 따라서 이러한 추세를 제거한 후

<표 3> 각 공간보간 기법의 적용방법

| Method | | Parameters | Number of neighbor samples |
|---|---|---|---|
| 1. Local Trend Surface | | - | 8, 10, 12, 15, 20 |
| 2. IDW | | Power: 1, 1.5, 2 | |
| 3. RBF | Completely regularized spline | - | |
| | Spline with tension | - | |
| Method | | Variogram function | Lag size (km) |
| 4. Ordinary Kriging | | Spherical, Exponential, Gaussian | 20, 25, 30, 35 |
| 5. Universal Kriging | | | |

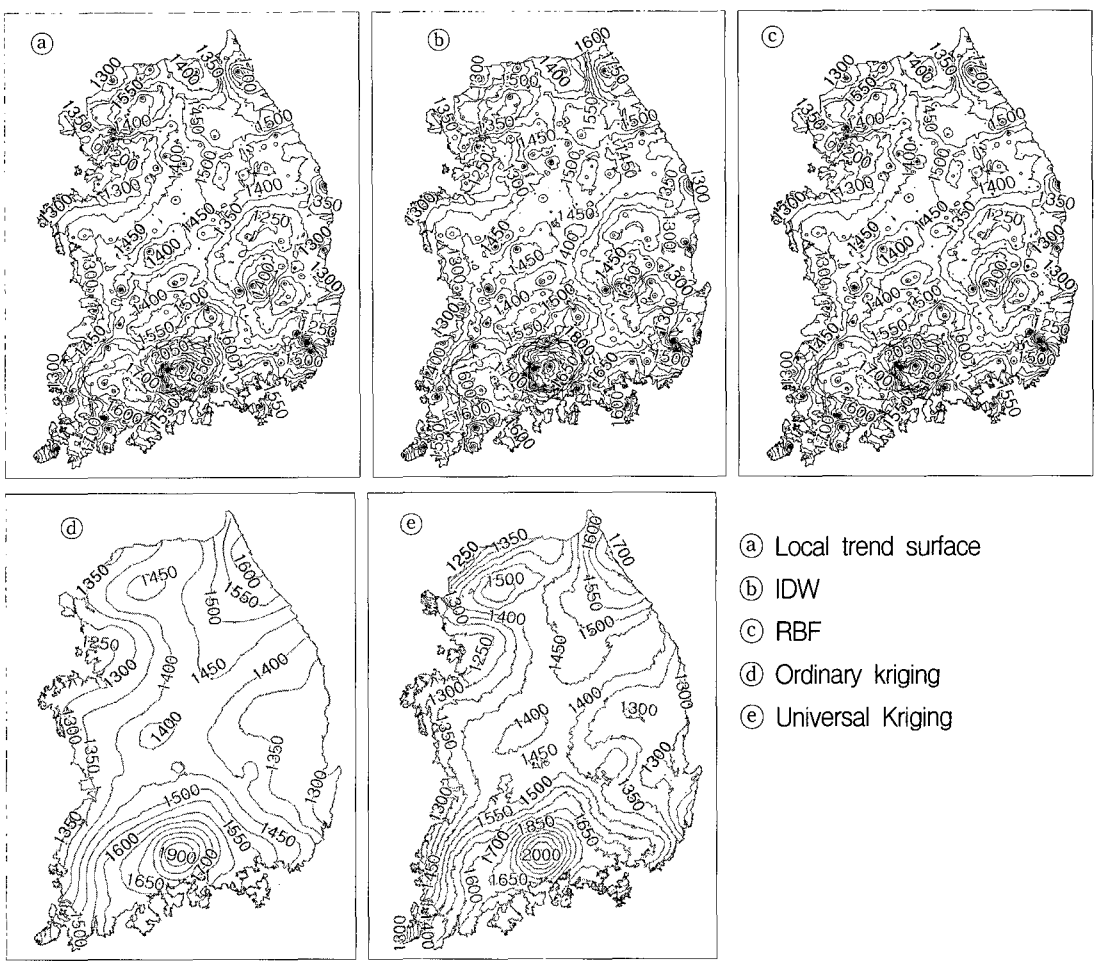
<표 4> 공간보간 방법별 RMS 에러

| Method | | Number of neighbors | | | | |
|---|---|---|---|---|---|---|
| | | 8 | 10 | 12 | 15 | 20 |
| Local trend surface | | 611.0 | 302.3 | 305.8 | 277.7 | 244.7 |
| IDW | power=1.0 | 218.6 | 220.0 | 220.0 | 220.2 | 219.2 |
| | power=1.5 | 222.3 | 222.8 | 222.1 | 222.0 | 220.6 |
| | power=2.0 | 227.8 | 227.8 | 226.8 | 226.6 | 225.1 |
| RBF | Regularized spline | 214.1 | 215.1 | 215.4 | 216.3 | 215.0 |
| | With tension | 216.1 | 215.1 | 215.4 | 216.3 | 215.0 |
| Method | Variogram function | Lag size (km) | | | | |
| | | 20 | 25 | 30 | 35 | |
| Ordinary Kriging | Spherical | 208.7 | 208.4 | 208.4 | 210.9 | |
| | Exponential | 208.8 | 208.8 | 208.8 | 208.9 | |
| | Gaussian | 210.4 | 210.4 | 210.4 | 210.1 | |
| Universal Kriging | Spherical | 208.4 | 209.4 | 208.4 | 207.1 | |
| | Exponential | 210.7 | 210.3 | 209.6 | 209.3 | |
| | Gaussian | 208.1 | 207.5 | 207.0 | 206.8 | |

크리깅 기법을 적용한 일반 크리깅이 좀 더 정확한 결과를 보인 것이다. 크리깅 기법의 정확성은 <표 4>에서 보여주는 것과 같이 베리오그램 함수에 따라 그 값이 영향을 받는 것으로 나타났다. 이에 반하여 분리거리의 변화는 크리깅에 있어 예측 정확성에 큰 영향을 미치지 않는 것으로 분석되었다. 따라서 강우자료에 대해 크리깅 기법을 적용할 경우 분리거리에 대한 정의보다 베리오그램 함수 선정에 더욱 신중을 기해야 할 것으로 판단된다. 지역경향면 분석의 경우 샘플수에 매우

큰 영향을 받는 것으로 분석되어 8개의 샘플을 사용했을 경우 검증 결과 611의 RMS 에러를 보인 것에 반하여 20 개의 샘플을 사용할 경우 244.7의 RMS 에러로 그 정확성이 급격하게 변동하는 것으로 나타났다. IDW의 경우 샘플수보다는 거리에 따른 가중치 파라미터 p(식 5)가 어떤 값을 갖는가하는 점이 정확성에 큰 영향을 미치는 것으로 나타났으며, 그 값이 작을 수로 정확성이 향상되어 본 연구에서는 1인 값을 부여할 때 가장 좋은 결과를 나타내는 것으로 분석되었다. RBF의 경우에는 샘플수에 비하여 적용 기법에 큰 영향을 받는 것으로 분석되었으며, $\phi$ 함수(식 6)로 spline with tension 함수보다 completely regularized spline 함수를 사용하는 것이 좋은 예측 결과를 보여주는 것으로 나타났다.

각 기법별 가장 정확한 예측 결과를 등우선도 표현하면 [그림 3]과 같다. [그




[그림 3] 각 공간기법에서 가장 높은 예측 정확성을 보인 결과에 대한 등우선도

림 3]은 각 보간기법의 예측 결과를 50m 간격의 등우선을 이용하여 나타낸 것으로 모든 결과에서 남해안 일대 및 강원도 동해안 지역 특히, 지리산 부근에서 높은 강우량이 예측된 것을 확인할 수 있다. [그림 3]에서 각 보간기법의 결과를 살펴보면 지역경향면 분석, IDW, RBF는 인근 지역 사이의 예측값이 비교적 급변하는 것으로 나타났으며, 정규 크리깅과 일반 크리깅에서는 인접 지역 사이의 예측값 변화 정도가 비교적 완만한 것으로 분석되었다. 이 결과를 예측값의 최소, 최대치를 통하여 살펴보면 지역경향면 분석, IDW, RBF의 경우 예측값의 최소값이 900 ㎜ 이하, 최대값이 2300 ㎜ 이상인 것에 반하여 정규 크리깅과 일반 크리깅의 경우 최소값이 1000 ㎜, 최대값이 2000 ㎜ 사이에 분포하는 것으로 나타났다(<표 5>).

<표 5> 각 공간보간 기법별 가장 높은 정확성을 보인 방법에서의 예측결과 통계치

(단위 : ㎜)

| Method | Min | Max | Mean | Std dev. |
|---|---|---|---|---|
| LTS* | 191 | 2,374 | 1,445 | 169 |
| IDW | 851 | 2,491 | 1,450 | 147 |
| RBF | 825 | 2,523 | 1,448 | 146 |
| OK** | 1,221 | 1,902 | 1,445 | 116 |
| UK*** | 1,079 | 2,003 | 1,445 | 134 |

*LTS : Local Trend Surface Analysis
**OK : Ordinary Kriging, ***UK: Universal Kriging

따라서 크리깅 기법이 다른 공간보간 기법에 비하여 예측 결과를 일반화하고 있음을 확인할 수 있다. 또한 크리깅 기법 안에서는 일반화 정도를 살펴보면 정규 크리깅이 일반 크리깅에 비하여 더욱 예측값의 범위가 좁혀져 있는 것으로 분석되었다.

<표 5>를 보면 지역경향면 분석의 경우 최소값이 다른 기법과 확연히 차이 나는 191 ㎜을 보여주고 있다. 지역경향면 분석에서 이렇게 낮은 값이 예측된 지역은 강화도 일대로 이 지역은 여타 다른 지역에 비하여 강우관측소가 매우 멀리 떨어져 있어이러한 결과를 나타낸 것으로 판단된다. 따라서 지역경향면 분석의 경우 관측지점이 고르게 분포하지 않은 지점에서 국지적으로 이상치가 예측될 수 있으므로, 이에 대한 고려가 필요할 것으로 보인다.

## 6. 결 론

공간통계자료는 비록 공간상 모든 위치에 그 값이 존재하나, 실제 우리는 모든 지점에서 공간자료를 수집하지 못한다. 따라서 몇 개의 대표 지점에서 자료를 수집한 후 미관측 지역까지 확장한 후 사용하게 되는데, 이때 공간보간 기법이 이용된다. 본 논문에서는 다양한 분야에서 매우 중요한 자료로 사용되는 강우자료에 대해 공간보간을 실시하고 그 정확성을 검증하였다. 기상청에서 운영하는 총 373개의 강우관측소의 2004 년도 연강우량에 대해 공간보간을 실시하였으며, 그로부터 이론적 베리오그램으로 가우시안 함수를 사용한 일반 크리깅이 가장 높은 정확성을 보인 것으로 확인되었다. 다른 보간기

법에 비해 크리깅 기법이 비교적 정확성이 높은 것으로 분석되었으며, 공간보간 결과에 있어서는 예측값의 분포 범위가 좁은 일반화의 특성을 갖는 것으로 분석되었다. 지역경향면 분석, IDW, RBF의 경우 강우자료에 대한 예측값의 범위가 크리깅에 비해 매우 넓게 분포하는 것으로 나타났으며, 공간적으로는 인접 지역 사이의 예측값이 크리깅에 비해 급변하는 것으로 분석되었다.

공간통계자료는 관측시점의 길이에 따라 다양한 시계열적 자료의 형태를 갖는다. 예를 들어, 본 논문에서는 연강우량에 대해서만 공간보간을 실시하였으나, 강우자료는 관측시점의 길이에 따라 월자료, 일자료, 시자료 등 매우 다양한 형태로 수집될 수 있다. 따라서 강우자료의 시공간적 분포특성에 대한 정확한 이해를 위해서는 향후 본 논문에서 사용한 연강우량 뿐만 아니라 월별, 일별, 시간별 강우자료에 대한 공간보간 분석이 필요할 것이다.

# 참 고 문 헌


최종근, 2002, 공간정보 모델링 크리깅과 최적화 기법, 구미서관

최용규, 2001, "Kriging 기법을 이용한 면적평균 강우량 산정", 충남대학교 석사학위 논문

Cressie, N.A.C., 1993, Statistics for Spatial Data, John Wiley & Sons

Jang, J., Hong, T., Namkung, P., 2003, "Spatial Data Analysis Using the Kring Method", The Korean communications in Statistics, Vol. 10, No. 2, pp.423-432

Johnston, K., Ver Hoef, J.M., Krivoruchko, K., Lucas, N., 2001, Using ArcGIS Spatial Analyst, ESRI Press

Kurtzman, D. and Kadmon, R., 1999, "Mapping of Temperature Variables in Israel: a Comparison of Different Interpolation Methods", Climate Research, Vol. 13, pp.33-43

Lee, E., 2003, "A Space Model to Annual Rainfall in South Korea", The Korean communications in Statistics, Vol. 10, No. 2, pp.445-456