

기상자료의 통계내삽

이 동 규¹⁾

요 약

불규칙 분포의 관측자료를 격자점자료로 내삽하는 방법의 하나인 통계내삽방법(Statistical interpolation)과 이 방법이 필요로 하는 통계적 상관함수(Correlation function)를 구하는 방법을 소개하고, 극동 아시아 지역의 기상변수들의 상관관계를 계산한다. 극동 아시아 지역 34개의 관측점에서 1977년 12월부터 1980년 2월까지 겨울철 12, 1, 2월의 9개월에 걸쳐 관측된 500mb면 지오폠펀셀고도 및 바람, 850mb면 온도 및 혼합비에 대하여 (34*33)/2개의 상관함수가 계산되었다. 관측점 사이의 거리가 증가함에 따라 지오폠펀셀고도와 온도의 자기상관함수는 천천히 감소하는 상관관계를, 바람과 혼합비의 자기상관함수는 비교적 빨리 감소하는 상관관계를 나타내었다. 지오폠펀셀고도와 바람에 대하여 계산된 자기 및 교차상관함수들의 수평분포는 겨울철 종관특징을 잘 표현하였다.

1. 서론

기상 분야에서는 불규칙 분포의 기상 관측 자료를 객관 분석하여 격자점 자료를 구하여 기상 변화의 특징을 이해하고 예측하는 데 이용한다. 객관 분석 방법에는 여러 가지가 있으나 이들 모두 각각 크든 작든 내삽오차를 가지고 있으며 이를 경감하려는 데 목적이 있다. 이 방법들 중에는 어느 특정한 지역과 기간에 대하여 기상변수의 거리에 대한 상관함수를 구하여 이를 내삽하는 데 사용하여 오차를 최소화하는 통계내삽법 (statistical interpolation method)이 있다. 이 방법은 내삽하는 과정에서 대기 운동 및 성질의 공간 변화성을 이용하므로 다른 방법보다 우월성이 있다.

Gandin(1963), Gandin과 Kuzenetsora(1965)은 구조함수와 상관함수를 정의하여 최적내삽(optimal interpolation)하는 과정에서 내삽오차를 최소화하는 방법을 제시하였다. 이어 Thieboux(1975), Schlatter(1975), Boyle(1981)등은 객관분석에 이용할 상관함수를 계산하여 내삽의 하중인자(weighting factor), 주사길이(scan length)를 결정하는 데 사용하였다. 이 최적내삽은 일명 통계내삽이라고도 불리우며, 내삽하는 과정에서 사용된 자료의 기간 및 지역, 대기의 운동 규모등을 고려하여야 하므로 많은 연구

1) (151-742) 서울특별시 관악구 신림동 서울대학교 대기과학과, 교수
Internet : e1291dk@cray2s.seri.re.kr

가 지속되고 있으며 그 이용도 점차 확대 되고 있다.

이 논문에서는 통계내삽과 상관함수를 소개하고 극동 아시아 지역에서 1977년부터 1980년 사이 겨울철 3개월의 자료를 사용하여 종관규모 운동계의 상관함수를 계산한 이동규(1985), 이희상(1985)의 결과를 토의한다. 극동 아시아 지역은 세계적으로 경압성이 강한 지역이므로 이 지역의 기상자료를 내삽할 때 변수의 특징을 고려함이 바람직하다. 다음절에서는 통계내삽이론을 소개하고, 3절에서는 상관함수를 정의하며, 4절에서 극동 아시아 지역에서 계산된 상관함수의 특징을 토의하고, 5절에서 결론을 내린다.

2. 통계내삽이론

격자점 주위에 관측값들을 격자점에 내삽하는 방법으로서 통계내삽방법은 n 개의 관측치, $f_i^o (i=1, \dots, n)$ 를 사용하여 주어진 격자점에서의 내삽된 값(이하 분석값이라 부름) f_s^A 는 이 격자점에서의 추정값 f_s^p 와 이 추정값으로부터의 관측값 편차 $f_i^o - f_i^p$ 이 선형조합으로 이루어지도록 하는 것이다. 즉,

$$f_s^A = f_s^p + \sum_{i=1}^n p_i (f_i^o - f_i^p) \quad (2.1)$$

로 나타낸다. 여기서 $p_i (i=1, \dots, n)$ 는 내삽 가중치이다. 추정값 f_s^p, f_i^p 등은 제공되는 값으로서 보통 기후자료 또는 수치예보모델자료를 사용한다. 식(2-1)에서 내삽 가중치 p_i 를 구하는 것이 문제인데, 격자점에서의 참값 f_s (일반적으로 알려지지 않음)와 분석값 f_s^A 사이의 내삽오차의 제곱평균이 최소가 되도록 하여 결정할 수 있다. 즉, 오차

$$E = \overline{(f_s - f_s^A)^2} \quad (2.2)$$

를 최소화하는 것이다. 식(2-1)을 식(2-2)에 넣고 정리하면

$$E = \overline{f_s'^2} + \sum_{i=1}^n \sum_{j=1}^n \{ \overline{f_i f_j} + \overline{\Delta f f_s'} + \overline{f_i \Delta f_j} + \overline{\Delta f_i \Delta f_j} \} p_i p_j - 2 \sum_{i=1}^n \{ \overline{f_s f_i} + \overline{f_s' \Delta f_i} \} p_i \quad (2.3)$$

이 된다. 여기서 관측점에서의 참값을 f_i (일반적으로 알려지지 않음)로 하면 추정오차는 $f_i' = f_i^p - f_i$ 되고, 관측오차는 $\Delta f_i = f_i^o - f_i$ 이 된다. 관측오차는 보통 분석값에 포함될 필요가 없는 작은 규모의 기상변화도 포함한다.

식(2-3)에는 추정오차와 관측오차 사이의 교차 공분산인 $\overline{\Delta f f_j'}$, $\overline{f_i \Delta f_j}$ 이 포함되어 있는데, 보통은 추정오차와 관측오차는 각각 다른 근원에서 발생된 것이므로 이 오차들 사이에 종속관계가 존재하지 않는다고 가정할 수 있다. 따라서,

$$f_s' \Delta f_i = \Delta f_i f_j' = f_i \Delta f_j = 0 \quad (2.4)$$

로 들 수 있고, 식(2-3)은

$$E = \overline{f_s^2} + \sum_{i=1}^n \sum_{j=1}^n \{ \overline{f_i f_j} + \overline{\Delta f_i \Delta f_j} \} p_i p_j - 2 \sum_{i=1}^n \overline{f_s f_i} p_i \quad (2.5)$$

된다. 식(2-2)의 의미에 의하여 E를 최소화하기 위하여 p에 대하여 미분하면

$$-\frac{\partial E}{\partial p_k} = \left\{ \sum_{i=1}^n (\overline{f_i f_k} + \overline{\Delta f_i \Delta f_k}) p_i - \overline{f_s f_k} \right\} \quad (2.6)$$

이 되고, 이것의 필요조건은

$$-\frac{\partial E}{\partial p_k} = 0 \quad (2.7)$$

된다. 즉,

$$\sum_{i=1}^n (\overline{f_i f_k} + \overline{\Delta f_i \Delta f_k}) p_i = \overline{f_s f_k}. \quad (2.8)$$

여기서 $k=1, \dots, n$ 이다. 만일 식(2-8)에서 추정오차 $\overline{f_i f_k}$, $\overline{f_s f_k}$ 와 관측오차 $\overline{\Delta f_i \Delta f_k}$ 의 공분산이 알려지면, 식(2-8)은 n개의 선형방정식으로 이루어지므로 이를 풀면 내삽가중치 p_i 를 얻는다. p_i 를 알면 식(2-1)로부터 분석치 f_g^A 를 계산하여 얻는다.

실제 문제에 있어서 식(2-8)을 적용할 때 몇가지 반드시 고려되어야 할 사항이 있다. 정리하면,

i) 모든 격자점과 관측점에서 추정값이 제공되어야 한다.

ii) 추정오차의 공분산($\overline{f_i f_k}$, $\overline{f_s f_k}$) 및 관측오차의 공분산($\overline{\Delta f_i \Delta f_k}$)이 주어야 한다.

iii) p_k 를 얻기 위하여 선형 방정식계 식(2-8)을 풀어야 한다.

추정오차와 관측오차의 공분산을 구하는 문제는 3절에서 설명될 것이며 식(2-8)은 많은 계산을 요구한다. 통계내삽방법의 장점은 많은 자료와 장기간의 자료를 이용함으로써 내삽오차를 최소화할 뿐만 아니라 이 자료 속에는 대기의 특징이 포함되어 있는 데 있다.

3. 상관함수

추정오차와 관측오차의 공분산을 구하기 위하여 Gandin(1963)은 한 쌍의 관측점 또는 격자점과 관측점 사이에서 통계적 구조 (statistical structure)로 볼 수 있는 상관함수 (correlation function)를 도입한다. 관측점 \vec{r}_1 과 \vec{r}_2 사이의 기상변수 f 의 공분산(covariance)은

$$COV(\vec{r}_1, \vec{r}_2) = \overline{\{f(\vec{r}_1) \cdot f(\vec{r}_2)\}}^2 \tag{3.1}$$

로 주어진다. 여기서 \vec{r}_1 과 \vec{r}_2 는 관측점들의 위치 벡터이고, 윗줄(—)은 시간평균을 의미한다. 이와같이 상관함수, 공분산(covariance), 분산은 두 관측점 사이의 시간적인 평균값 공간적 변화와 상호 관련성을 이해하는 정보를 제공한다. 즉, 어떤 대기운동의 지배하에서 두 점사이의 이 대기운동의 공간적 관계를 통계적으로 나타낸 것이다.

시간 평균값 $\bar{f}(\vec{r}_1)$, $\bar{f}(\vec{r}_2)$ 로 부터 편차,

$$\begin{aligned} f'(\vec{r}_1) &= f(\vec{r}_1) - \bar{f}(\vec{r}_1) \\ f'(\vec{r}_2) &= f(\vec{r}_2) - \bar{f}(\vec{r}_2) \end{aligned} \tag{3.2}$$

을 이용하면

$$COV(\vec{r}_1, \vec{r}_2) = \overline{\{f'(\vec{r}_1) \cdot f'(\vec{r}_2)\}}^2 \tag{3.3}$$

로 표현할 수 있다.

기상 변수 f 가 수평으로 균질(homogeneous)하고 등방(isentropy)하다고 가정 하고, \vec{r}_1 과 \vec{r}_2 점에 위치하는 두 관측점 사이의 거리를, $|\vec{r}_1 - \vec{r}_2| = d$ 라 놓고 식(3-3)을 다시 쓰면

$$COV(d) = \overline{\{f'(0) \cdot f'(d)\}}^2 \tag{3.4}$$

가 된다. 공분산을 구함에 있어서 기상변수가 수평으로 균질하다면 공분산의 값은 단지 두 관측점에서의 거리와 방향에 의하여 결정되고 그 두 관측점 한쌍의 위치에 무관하다. 더 나아가서 등방하게 되면 두 관측점을 연결하는 선의 중앙점에서 어느 방향으로나 공분산의 값이 동일하게 되어 공분산의 값은 완전히 관측점들 사이의 스칼라거리(d)에 의해서 결정된다. 일반적으로 어느 지역을 지배하는 대기의 운동 또는 성질은 균질하지도 않으며 등방이지도 않다. 그러나 대기의 속성에 의하여 인접한 두 점 사이에서 균질성과 등방성의 관계는 커지며 평균적 또는 기후적 대기에서는 더욱 그러하다.

만일, $\vec{r}_1 = \vec{r}_2 = \vec{r}$ 이면 공분산은 분산(variance)이

$$COV(0) = \overline{\{f'(\vec{r})\}}^2 \tag{3.5}$$

되고, 상관함수 CF는 다음과 같이 정의한다.

$$CF(d) = \frac{Cov(d)}{Cov(0)} \tag{3.6}$$

비슷한 방법으로 두 개의 다른 기상변수 f 와 g 사이에 대하여 교차공분산(Cross-covar

iance)이

$$CCOV(d) = \overline{\{f'(0) \cdot g'(d)\}} \quad (3.7)$$

로 정의될 수 있고, 교차상관함수(Cross-Correlation function) CCF는 다음과 같이 정의된다.

$$CCF(d) = \frac{CCOV(d)}{\{f \text{의 } COV(0) \cdot g \text{의 } COV(0)\}^{1/2}} \quad (3.8)$$

4. 상관함수의 계산

상관함수는 장기간 많은 관측점에서의 자료를 이용하면 그 특성이 보다 두드러진 값을 가지게 된다. 이 연구에서는 1977년 12월부터 1980년 2월까지 겨울철 12, 1, 2월의 9개월에 걸쳐 매일 한 번씩 극동 아시아에서 관측된 500mb 지오폠텔셀고도 및 바람, 850mb면 온도 및 혼합비를 사용한다.

사용된 관측점들은 경도 106°E에서 146°E, 위도 20°N에서 54°N내에 있는 극동 아시아 지역의 57개 레디오존데 관측점 중에서 34개 관측점들만 선정하였다. 이 선정에서 일본과 중국의 관측점들 중에서 매일 관측하지 않은 지점들은 제외하고, 황해, 동해, 일본 남쪽 서태평양의 관측점이 희박한 점을 고려하여 일본지역에서 조밀하게 분포된 몇 개의 관측점을 제외한 것이다. 이 34개의 관측점들의 평균거리는 약 740km, 가장 근접한 두 관측점사이의 거리는 약 250km, 가장 먼 두 관측점 사이의 거리는 약 4,000km이다. 관측점들의 위치를 그림1에 나타내었다.

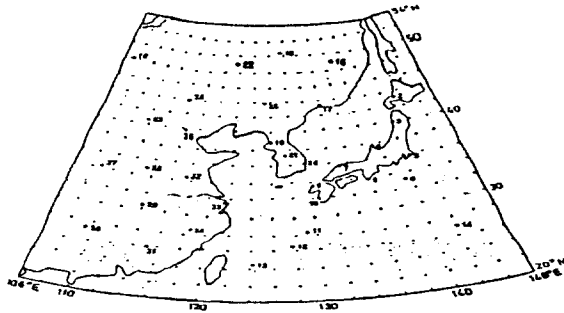


그림 1) 관측점의 위치

그림2는 500mb 지오폠텔셀고도와 풍속의 거리에 대한 상관함수를 보인 것이다. 이 상관함수 값들의 갯수는 $(34 \times 33) / 2 = 56$ 이다. 지오폠텔셀고도의 상관함수는 약 2,500 km 근처에 이르러 아주 작은 상관관계를 보이고, 풍속의 상관함수는 1,000km 보다

짧은 거리에서 상관관계가 음으로 나타남을 보인다. 지오폠텔셜고도와 바람사이에 공간적 변화성이 크게 다르다.

그림3은 850mb면 온도와 혼합비의 거리에 대한 상관함수를 보인 것이다. 온도의 상관함수는 약 2,000km 근처에서 음의 상관관계가 나타나기 시작하고, 혼합비의 그것은 500km 근처에서 음의 상관관계가 나타나기 시작한다. 온도의 상관관계가 혼합비의 그것보다 훨씬 높게 나타난다.

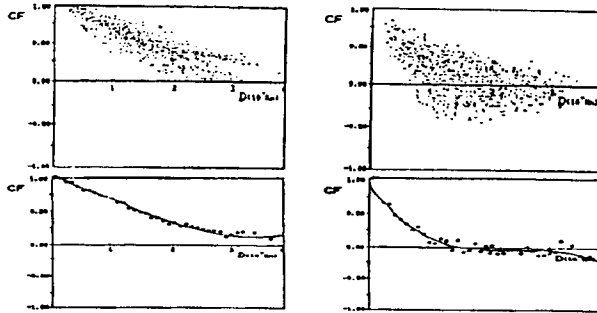


그림 2) 500 mb 면 지오폠텔셜고도(왼쪽)와 풍속(오른쪽)의 상관함수

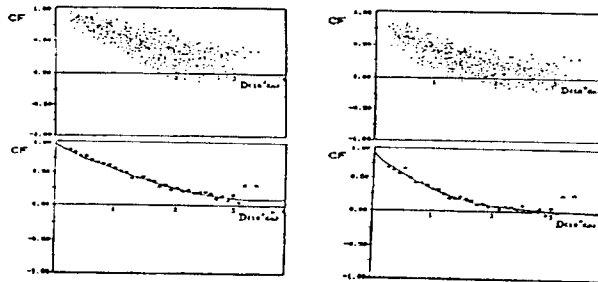


그림 3) 850mb면 온도(왼쪽)와 혼합비(오른쪽)의 상관함수

거리를 100km씩 구간으로 나누고 각 구간에 들어오는 상관함수의 값들을 평균한 값을 거리에 대하여 나타내고(그림2와 3), 3차 다항식과 지수식을 적용하여 최소제곱법에 의한 회귀방정식을 얻는다. 여기서 100km씩 구간을 나누어 평균값을 만든 이유

는 회귀방정식을 구할 때 많은 상관값을 사용하는 것을 피하기 위한 것이며, 실제로 이 결과들 사이에 큰 차이가 없다. 또, 3차 다항식을 선정함에는 특별한 배경이 없다. 이 회귀곡선의 함수를 식(2-8)의 공분산으로 하여 가중인자들을 구하는데 사용하게 된다.

3차 다항식에 의한 회귀곡선들을 살펴보면 운동장과 풍속이 거리에 대한 상관관계가 가장 작게 나타나므로 약 1,500km까지의 관계만 이용한다. 그러나 지오포텐셜고도, 온도, 혼합비와 같은 스칼라양들은 질량과 크게 관계가 있어 그 상관관계가 3,000km이상까지 높게 나타나므로 자료들 사이의 상관관계를 이 거리까지 이용할 수 있다

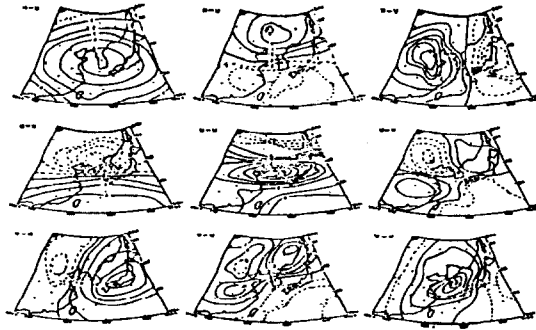


그림 4) 기상변수 h, u, v 의 교차상관 분포

극동 아시아 지역 34개 관측점 중에서 오산 관측점을 중심으로 500mb면 지오포텐셜고도(h)와 바람(u,v) 사이의 교차상관함수를 계산하였다. 그림4는 변수 h, u, v 사이에 3개의 자기상관함수와 6개의 교차상관함수들의 500mb면 분포를 보인 것이다. h, u, v의 상관함수 및 교차상관함수 분포는 그 변수의 특성과 변수사이의 관계를 잘 드러낸다. 이 상관함수 분포는 해안에 관측점의 수가 적고 극동 아시아 지역의 기상현상 특징에 의하여 이상적인 대칭을 나타내지는 못 한다. 질량장은 균질성 및 등방성이 크기 때문에 지오포텐셜고도의 자기상관함수가 대칭성을 보이며 바람은 편서풍이므로 등방성이 작아 교차상관함수의 분포가 대칭에서 벗어난다.

5. 결론

1977년부터 1980년 겨울철 동안의 500mb면 지오포텐셜고도 및 풍속, 850mb면 온도 및 혼합비의 거리의 함수로서 상관관계는 서로 크게 다르다. 지오포텐셜고도와 온도는 자기상관관계가 상대적으로 큰 반면 풍속과 혼합비는 작다. 지오포텐셜고도, 바람 사이의 교차상관관계는 변수들 사이의 관계와 극동 아시아 지역의 해양과 대륙에

의하여 나타나는 기상 특성을 잘 나타낸다. 상관함수를 사용하여 통계내삽을 할 경우 내삽할 변수, 고도, 시기 등에 따라 이 상관함수의 특성이 고려되는 것이 바람직하다.

6. 감사의 말

한국통계학회 1992년 춘계학술논문발표회에서 이 논문을 발표할 기회를 가진데 감사하며, 또, 전명식 교수와 조신섭 교수의 토론에 감사한다.

< 참고 문헌 >

- [1] 이동규, 박순용 (1985), "극동 아시아 지역에서의 기상변수의 구조및 상관함수," 한국기상학회지, 21(1), 15-33.
- [2] 이희상 (1985), "겨울철(1977-1980) 극동아시아 지역에서의 500mb등압면 바람과 지오폠폠텔셜 고도의 상관적 특성에 관하여," 서울대학교 대기과학과 석사논문, 41pp.
- [3] Boyle, J. S. (1981), "Autocorrelations of moisture parameters on isentropic surfaces," *Monthly Weather Review*, 2109, 2401-2404.
- [4] Gandin, L.S. and T. I. Kuzenetsova (1965), "On the three-dimensional statistical structure of the geopotential field," *Trudy Glavnoi Geofizicheskoi Observatorii(GGO)*, Imeni A. I. Voeikova, Leningrad, No. 168, 84-93.
- [5] Gandin, L.S. (1963), *Objective Analysis of Meteorological Field*, Gidrometerolo-gichesKoe Izdatel'stvo, Leningrad. (Translated form Russian, Israel Program for Scientific Translated from Russian, Jerusalem, 1965, 242pp).
- [6] Schlatter, T.W. (1975), "Some experiments with a multivariate statistical objective analysis scheme," *Monthly Weather Review*, 103, 246-257.
- [7] Thiebaux, H.J. (1975), "Experiments with correlation representations for objective analysis," *Monthly Weather Review*, 103, 617-627.

Statistical Interpolation of Meteorological Data

Dong-kyou Lee¹⁾

Abstract

Statistical interpolation which uses past experience about the behaviour of the atmosphere(correlation function) to interpolate the observations irregularly distributed in space and time to regular grids is discussed. Correlation functions are computed in the Far East Asian region during the winter periods of December 1977 to February 1980.

Results show from the computation of correlation functions that there exists a large difference in the autocorrelation functions by slowly decreasing geopotential height and temperature, and rapidly decreasing wind and mixing ratio with increasing data-separation distance. The horizontal distribution of auto-correlation and cross-correlation functions between geopotential height and wind are well corresponded to persistence of wintertime synoptic features.

1) Atmospheric Science Department, Seoul National University, Shillim-dong, Kwanak-gu, Seoul, 151-742, Korea
Internet : e1291dk@cray2s.seri.re.kr