

## 공간데이터 크리깅 적용을 위한 공간상관함수 추정 Estimation of Spatial Coherency Functions for Kriging of Spatial Data

배태석<sup>1)</sup>  
Bae, Tae-Suk

### Abstract

In order to apply Kriging methods for geostatistics of spatial data, an estimation of spatial coherency functions is required priorly based on the spatial distance between measurement points. In the study, the typical coherency functions, such as semi-variogram, homeogram, and covariance function, were estimated using the national geoid model. The test area consisting of  $2^\circ \times 2^\circ$  and the Unified Control Points (UCPs) within the area were chosen as sampling measurements of the geoid. Based on the distance between the control points, a total of 100 sampling points were grouped into distinct pairs and assigned into a bin. Empirical values, which were calculated with each of the spatial coherency functions, resulted out as a wave model of a semi-variogram for the best quality of fit. Both of homeogram and covariance functions were better fitted into the exponential model. In the future, the methods of various Kriging and the functions of estimated spatial coherency need to be studied to verify the prediction accuracy and to calculate the Mean Squared Prediction Error (MSPE).

Keywords : Random Process, Spatial Coherency, Geoid, Kriging

### 초 록

지구통계학적인 공간분석의 대표적인 방법인 크리깅(kriging)을 적용하기 위해서는 두 관측점 사이의 거리에 기반한 상관성을 나타내는 공간상관함수의 추정이 우선적으로 이루어져야 한다. 본 연구에서는 다양한 크리깅에 적용할 수 있는 대표적인 상관함수인 semi-variogram, homeogram, covariance function에 대하여 국가지오이드 모델을 기반으로 추정하였다. 경위도 각각  $2^\circ$ 의 대상지역 내 통합기준점의 지오이드고를 이용하였으며, 선형모델을 이용하여 공간적인 편향성을 제거하였다. 전체 100개의 샘플 포인트에 대해서 중복되지 않은 두 점 간의 거리를 기준으로 구간을 나누고, 각 함수에 대한 경험적인 값을 계산하였다. 공간상관함수의 경험적인 값은 각각 두 개의 모델에 최소제곱조정 방법으로 피팅한 결과 semi-variogram의 wave 모델 적합도가 가장 높았으며, homeogram과 covariance function은 exponential 모델이 상대적으로 좋은 피팅 결과를 보였다. 본 연구에서 결정한 공간상관함수는 추후 다양한 크리깅 방법을 통해 임의의 지점에서의 예측값에 대한 정확도 검증과 이에 대한 평균제곱예측오차(Mean Squared Prediction Error, MSPE)를 계산함으로써 각 함수의 활용성에 대한 추가적인 연구가 수행되어야 한다.

핵심어 : 랜덤프로세스, 공간상관성, 지오이드, 크리깅

Received 2016. 02. 03, Revised 2016. 02. 15, Accepted 2016. 02. 23

1) Member, Dept. Geoinformation Engineering, Sejong University (E-mail: baezae@sejong.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

1980년대 초 지구통계학적 분석방법이 도입된 이후로, 공학, 지질학, 수학, 통계학 등 다양한 분야에서 학제간 연구가 활발히 진행되어 왔다(Cressie, 1993). 지구통계학적 분석은 특정 위치에서 관측된 값을 바탕으로 공간적인 상관성을 결정하고, 원하는 지점에서의 값을 내삽 또는 예측하는 것을 말한다. 그러나 현실 세계에서 취득되는 대부분의 데이터는 결정론적 성분만으로 모델링할 수가 없으므로 확률적인 요소를 도입할 필요가 있다(Jekeli, 2001). 측지학의 응용분야를 포함하여 대부분의 분야에서는 관측자료를 이용하여 임의 지점에서 정확한 예측값을 구할 필요가 있으며, 이를 위해 다양한 공간예측 방법이 연구되고 있다. 공간예측은 위치를 알고 있는 지점에서 관측한 랜덤값의 공간적인 경향성에 대한 분석을 기반으로 도메인 내부의 새로운 위치에서의 개별적인 성분을 공간적으로 예측하는 방법이다(Schaffrin, 2000). 크리깅은 프로세스의 2차 모멘트 성질에 기반한 공간예측 방법으로서 최소평균제곱근오차를 결정한다(Cressie, 1993). 일반적으로 공간분석을 위해서는 OK(Ordinary Kriging) 방법을 많이 사용해 왔으며, 이는 바이어스를 포함한 대량의 데이터를 처리하는데 유용한 것으로 알려져 있다. 한편 OK의 불편추정성(unbiasedness) 대신 향상된 평균제곱예측오차(Mean Squared Prediction Error, MSPE)를 얻을 수 있는 것으로 알려진 OBK(Optimal Biased Kriging) 방법도 제시되었다(Schaffrin, 1993). OBK는 homBLIP(Best homogeneously Linear Prediction)으로도 알려져 있으며, 추정값 계산을 위한 수식이 단순하기 때문에 OK에 대한 대안으로 사용가능하다.

지구통계학적 공간분석 결과에 대한 평가는 정확도, 바이어스, 계산의 복잡성 등의 기준을 적용할 수 있다. 측지학에서 가장 많이 알려진 랜덤신호는 중력이상값을 나타내는 것으로서, 이를 기반으로 최소제곱 콜로케이션 이론을 적용하여 지오이드 모델을 만든다. 랜덤신호의 상관함수는 신호를 분석하고 예측하기 위한 가장 좋은 방법을 제공하므로 이를 우선적으로 결정해야 한다(Jekeli, 2003). 이러한 공간적인 상관성을 측정하기 위한 다양한 방법이 제시되었는데, 그 중에서 semi-variogram, homeogram, covariance function 등이 대표적이며, 이를 다양한 크리깅 방식에 통합적으로 적용하기 위한 연구도 많이 수행되었다(Schaffrin, 2000, 2013). 본 논문에서는 크리깅 적용을 위해 필요한 관측값 사이의 공간적인 상관성을 나타내기 위해 대표적인 세 가지 공간상관함수를 추정하였으며, 모델별 적합도에 대한 분석을 수행하였다.

## 2. 랜덤프로세스

랜덤프로세스는 이산 또는 연속 확률변수의 집합으로서 일반적으로 시간 또는 공간좌표로 표현되는 결정론적 매개변수와 관련되어 있다(Jekeli, 2001). 시간 또는 공간 도메인의 각 점에서의 프로세스는 미리 가정한 확률밀도함수(probability density function)를 가지는 확률변수이다. 실제 공간 데이터는 확률프로세스의 구현이 한번만 가능한 경우가 많으므로, 이 때 프로세스는 시간 또는 공간에 대한 함수라고 가정할 수 있다. Ergodicity는 평균, 분산 등 프로세스의 통계정보가 시공간 기반의 평균값으로부터 유도된 통계값에 해당하는 경우, 즉 프로세스의 단일 구현에 의한 결과와 동일한 경우를 나타낸다. 따라서 ergodicity 성질에 기반하여 한번의 구현으로부터 프로세스의 확률적 특성을 분석할 수 있다(Jekeli, 2003). Schaffrin(1993)은 구(sphere)에서의 homogeneous-isotropic 프로세스에 대한 연구를 수행하였으며, 이 때 프로세스는 가우시안(Gaussian)과 ergodic을 동시에 만족할 수 없음을 보였다.

일반적으로 시·공간 신호는 랜덤프로세스이므로 확률변수 간 결합확률을 고려해야 한다. 랜덤프로세스의 확률이 시간 또는 공간에 따라 변하지 않는 경우를 정적프로세스라고 하는데, 이 경우 결합확률밀도함수는 시간 원점에 무관하며, 평균, 분산 및 자기상관(autocorrelation) 구조는 시간에 따라 변하지 않는다. 따라서 상관함수는 프로세스의 두 확률변수 사이의 시간 간격에만 의존한다(Jekeli, 2001).

측지학에서 다루는 대부분의 데이터에 대한 공간적인 경향성은 ergodicity와 isotropy의 두 가지 성질로 나타낼 수 있다. Ergodicity는 모든 프로세스의 구현을 공간적인 평균값으로 대체할 수 있다는 것을 의미하며, 점  $s$ 에서의 확률변수  $x(s)$ 의 기대값을 수식으로 표현하면 Eq. (1)과 같다.

$$E\{x(s)\} := \int_{-\infty}^{\infty} x(s)f(x(s))dx = \frac{1}{|S|} \int_S x(s)ds = \mu \quad \text{for all } s \in S \quad (1)$$

where  $f(\cdot)$  represents the probability density function (pdf), and  $\mu$  is the mean (or expectation) of the random variable  $x(s)$ .

Ergodicity는 homogeneity 또는 stationarity를 의미하는데, 만약 프로세스가 isotropy를 만족한다면 임의의 두 점 사이의 거리를 사용할 수 있다(Schaffrin, 2000). 이 경우 프로세스는 시간(또는 공간)  $s$ 의 함수가 아니며, 두 확률변수  $x(s)$ 와  $x(s')$ 의 자기공분산(auto-covariance)은 Eq. (2)와 같이 두 점  $s$ 와  $s'$  사이의 거리에만 전적으로 의존하게 된다. 따라서,

$$\begin{aligned}
 C\{x(s), x(s')\} &:= \int_{-\infty}^{\infty} (x(s) - \mu)(x(s') - \mu) f(x(s)) f(x(s')) dx \\
 &= \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}_h} (x(s) - \mu)(x(s+h) - \mu) ds = \mu \quad \text{for } h := s' - s \quad (2) \\
 &= C\{x(s), x(s+h)\} = C_x(|h|)
 \end{aligned}$$

이를 정리하면 Eq. (3)에서 보는 바와 같이 유사성의 정도를 나타내는 homeogram  $\eta_x(h)$  를 정의함으로써 비중심 공분산함수로 나타낼 수 있다.

$$\begin{aligned}
 C_x(s, s') &= C_x(|h|) \\
 &:= E\{(x(s) - \mu)(x(s') - \mu)\} \\
 &= E\{x(s) \cdot x(s')\} - \mu^2 \\
 &:= \eta_x(|h|) - \mu^2
 \end{aligned} \quad (3)$$

한편, semi-variogram은 비유사성을 측정하는 지표로서 Eq. (4)와 같이 정의된다.

$$\begin{aligned}
 \gamma_x(s, s') &= \gamma_x(|h|) \\
 &:= \frac{1}{2} D\{x(s) - x(s')\} = \frac{1}{2} E\{(x(s) - x(s'))^2\} \\
 &= \sigma_x^2 - C_x(|h|) \\
 &= \eta_x(0) - \eta_x(|h|)
 \end{aligned} \quad (4)$$

Isotropy는 구면상에서 관측된 데이터 사이의 공분산이 방향에 무관함을 나타낸다. 따라서 공간예측은 주어진 관측값  $y := [y_1, y_2, \dots, y_n]^T$ 로부터 공간상관함수를 이용한 예측값  $\tilde{x}(s)$ 와 평균제곱예측오차(MSPE)를 찾아내는 것이다. 여기서 Eq. (4)에서 알 수 있듯이 semi-variogram은 covariance function과 역대칭 함수로 표현할 수 있다.

구면상의 모든 점  $s \in \mathcal{S}$  에서 스칼라 값으로 이루어진 프로세스가 정의되었다고 가정할 때, 공간상의 한 점  $s$ 에서의 관측값  $x(s)$ 는 실수집합  $\mathcal{R}$ 에서 확률변수이며, 평균과 분산은 Eq. (5)와 같이 정의된다.

$$\begin{aligned}
 E\{x(s)\} &= \mu(s) \\
 D\{x(s)\} &= \sigma_x^2(s)
 \end{aligned} \quad (5)$$

공간프로세스  $x(s)$ 가 특정 위치  $s_i$  for  $i \in \{1, \dots, n\}$ 에서 관측되었다고 할 때, 관측값은 랜덤 노이즈 성분을 고려하여 Eq. (6)과 같이 나타낼 수 있다.

$$\begin{aligned}
 y_i &= x(s_i) + e_i, \quad e_i \sim (0, \sigma_e^2) \quad (i=1, \dots, n) \\
 C\{e_i, e_j\} &= 0 \quad \text{for } i \neq j
 \end{aligned} \quad (6)$$

관측값 오차에 대한 분산은 일반적으로 모든 위치에서 동일하다고 가정할 수 있으며, 또한 프로세스  $x(s)$ 와 관측값 오차의 상관성은 없다고 가정할 수 있다(즉,  $C\{x(s_i), e_j\} = 0$ ). 따

라서 공분산 함수는 프로세스가 isotropy라는 가정하에 Eq. (7)과 같이 주어진다(Schaffrin, 2000).

$$\begin{aligned}
 C_y(0) &:= \sigma_y^2 = \sigma_x^2 + \sigma_e^2 \quad \text{for all } s \in \mathcal{S} \\
 C_y(|h|) &= C_x(|h|) \quad \text{for all } s \neq s'
 \end{aligned} \quad (7)$$

공간프로세스에서는 공간상관함수를 얼마나 정확하게 추정하느냐의 문제와 서로 다른 공간상관함수에 기반하여 MSPE를 추정할 때 결과값의 일관성을 비교할 필요가 있다.

### 3. 공간상관함수의 추정

본 연구에서는 공간데이터로부터 추정한 다양한 공간상관함수에 대한 비교, 분석을 수행하였다. 국토지리정보원에서는 2011년부터 육지-해양 통합지오이드 모델 개발 연구를 시작하였으며, 2013년에는 지리산, 속리산 등 산악지역의 정확도를 개선한 KNGeoid13 모델을 개발하였다. 2014년에 구축된 합성지오이드 모델(KNGeoid14)은 우리나라 전역에서 약 3.31 cm의 정밀도를 나타낸다(NGII, 2016). 본 연구에 사용한 국가지오이드 모델 KNGeoid13은 다양한 데이터(항공/지상/선상 중력 자료, GPS/levelling 자료 등)를 기반으로 최소제곱 콜로케이션 방법으로 추정되었다. 공간상관함수 추정을 위해서 산악지역과 평지를 포함하고 있는 경위도 각 2°에 해당하는 영역을 선택함으로써 다양한 분포의 지오이드고에 기반한 공간상관함수 추정이 가능하도록 하였다(Fig. 1).

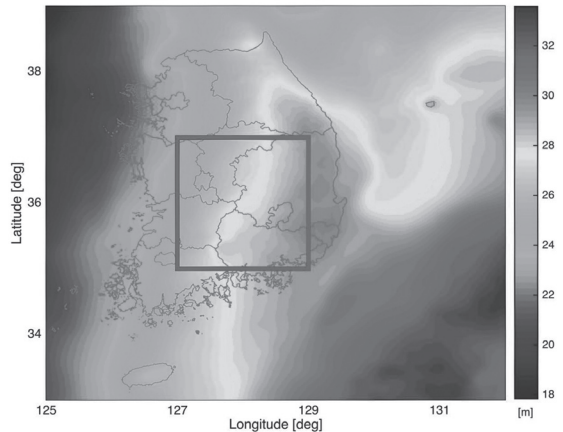


Fig. 1. Test area for spatial coherency functions (Bae and Schaffrin, 2015)

우리나라 전국 통합기준점(Unified Control Points, UCPs)은 타원체고와 정표고 값을 가지고 있다. Fig. 2에서 보는 것처럼 본 연구에서 사용하는 영역 내 통합기준점 중 균일한 공간적인 분포를 가지도록 100개의 좌표를 선정하여 공간상관함수의 경험적인 값을 생성하는데 사용하였다. 전체 100개의 통합기준점에서 GPS/leveling 자료와 KNGeoid13 모델의 지오이드고 차이는 평균제곱근오차(RMSE) 기준으로 약  $\pm 2.9$  cm 수준이다. 그러나 일부 산악지역의 경우 큰 차이를 보이는 경우가 있으므로 지오이드 오차를 대략 10 cm 수준으로 가정하는 것이 바람직 할 것으로 판단되어 본 연구에서는 관측오차의 분산을  $\sigma_a^2 = (0.1m)^2$  로 가정하였다.

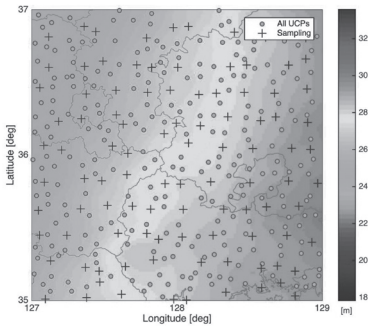


Fig. 2. Location of UCPs and sampling points

총 100개의 샘플 데이터의 측지좌표를 평면투영(TM)한 결과 두 점 사이의 평균적인 거리는 약 105 km이며, 표준편차는 50 km 수준이다. Fig. 3은 샘플 데이터 사이의 거리(lag 거리)에 대한 히스토그램으로서 210 km 이상 구간은 표본수가 적어서 제외했다. 그림에서 보는 것처럼 전체적으로 각 구간별로 고른 분포를 보이고 있으므로 경험적인 공간상관함수를 계산하기에 적절할 것으로 판단된다. 각 구간별 거리는 대략 15.6 km이며, 이 간격을 기준으로 구간을 나누어 경험적인 공간상관함수를 추정하였다.

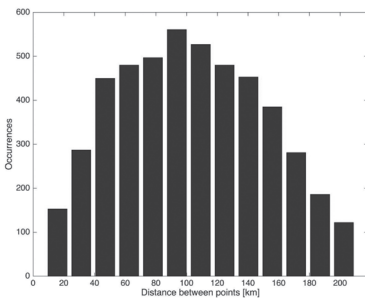


Fig. 3. Histogram of lag distances between sample points

테스트 지역에 존재하는 지오이드의 경향성을 제거한 후 잔차 지오이드를 이용하여 공간상관함수를 추정하기 위해 지오이드고를 위도, 경도에 따른 일차식으로 정의한 평면모델에 피팅하였다(Eq. (8)).

$$N(\varphi, \lambda) = a(\lambda - \lambda_0) + b(\varphi - \varphi_0) \tag{8}$$

where  $N$  = geoid,  $\varphi$  = latitude,  $\lambda$  = longitude,  $\varphi_0$  = latitude origin,  $\lambda_0$  = longitude origin ( $\lambda_0$  and  $\varphi_0$  are the longitude and latitude of the center of the test area, respectively).

지오이드고의 평면 모델을 추정할 때 바이어스 항은 추정하지 않았으며, 이는 바이어스 정보가 필요한 covariance function과 바이어스에 대한 사전 정보가 필요없는 homeogram의 차이를 보여주기 위해서다. 추정된 파라미터는  $[a, b] = [2.3852, -0.8953]$  [m/deg]이며, Fig. 4는 선형의 경향성을 제외한 잔차 지오이드(123x123)와 표본점의 위치를 나타낸 것이다. 그림에서 보는 바와 같이 좌하단의 산악지역에서의 지오이드가 크게 나타나고 있으며, 우하단에서 상대적으로 작은 값을 보여준다. 본 연구에서 사용한 100개의 표본점에서의 선형 경향성을 제거한 지오이드는 Fig. 5와 같다.

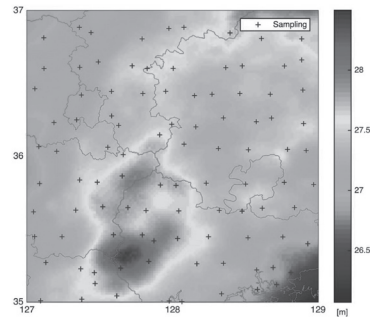


Fig. 4. Sampling points along with the de-trended geoid

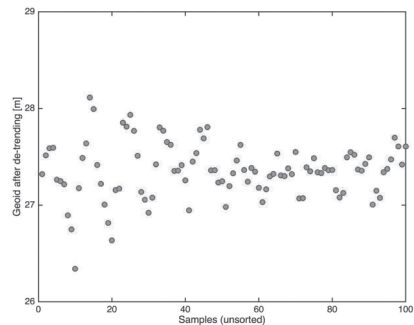


Fig. 5. Geoid after de-trending at 100 sampled UCPs (unsorted)

경향성을 제거한 잔차 지오이드를 이용하여 경험적인 공간 상관함수를 추정하기 위한 방법은 Table 1에 요약되어 있다 (Cressie, 1993). 본 연구에서는 세 종류의 공간상관함수를 추정하였으며, 각각을 결정하기 위해서는 Fig. 3의 히스토그램과 같이 구간을 나눈 후, 각 구간 내에서 하나의 경험적인 값을 결정한다. Semi-variogram은 각 구간내에서 중복되지 않

도록 두 점을 하나의 쌍으로 설정하고, 두 점의 지오이드 값 차이의 제곱에 대한 평균을 계산한다. Covariance function은 일반적인 공분산 계산과 유사하게 각 쌍별로 구간 내에서 지오이드 평균값과의 차이를 곱한 후 평균하는 방식이다. 이에 반해 homeogram은 평균값을 제거하지 않고 지오이드 값 자체를 이용하여 기대값을 구하는 방식이다.

**Table 1. Summary of spatial coherency functions**

Functions	Descriptions
Semi-variogram	$\hat{\gamma}(h) = \frac{1}{ d(h) } \sum_{d(h)} (N(s_i) - N(s_j))^2$ $d(h) := \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \dots, n\}$ $N(s_i) \text{ is the random effect of the actual process (geoid) at site } s_i$ $ d(h)  \text{ is the number of distinct pairs in } d(h)$
Homeogram	$\eta(h) = E\{x(s) \cdot x(s+h)\} = \frac{1}{ d(h) } \sum_{d(h)} N(s_i)N(s_j)$
Covariance function	$\hat{C}(h) = \frac{1}{ d(h) } \sum_{d(h)} (N(s_i) - \bar{N})(N(s_j) - \bar{N})$ $\bar{N} = \frac{1}{n} \sum_{i=1}^n N(s_i)$

**Table 2. Spatial coherency models used in this study (Cressie, 1993)**

Functions	Models
Semi-variogram	<p><u>Spherical</u>:</p> $\gamma_y(h) = \begin{cases} c_0 + c_s \left\{ \frac{3 h }{2a_s} - \frac{1}{2} \left( \frac{ h }{a_s} \right)^3 \right\}, & 0 \leq  h  \leq a_s \\ c_0 + c_s, &  h  \geq a_s \end{cases}$ <p>where <math>c_0 \geq 0</math>, <math>c_s \geq 0</math>, and <math>a_s \geq 0</math></p> <p><u>Wave</u>:</p> $\gamma_y(h) = c_0 + c_w \left\{ 1 - \frac{a_w}{ h } \sin \left( \frac{ h }{a_w} \right) \right\}$ <p>where <math>c_0 \geq 0</math>, <math>c_w \geq 0</math>, and <math>a_w \geq 0</math></p>
Homeogram	<p><u>Exponential</u>:</p> $\eta_y(h) = c_0 + c_w \left\{ 1 - \frac{a_w}{ h } \sin \left( \frac{ h }{a_w} \right) \right\}$ <p>where <math>c_0 \geq 0</math>, <math>c_w \geq 0</math>, and <math>a_w \geq 0</math></p> <p><u>Wave</u>:</p> $\eta_y(h) = \begin{cases} \sigma_y^2 + \mu^2, & h = 0 \\ c_0 - c_h \left\{ 1 - \frac{a_h}{ h } \sin \left( \frac{ h }{a_h} \right) \right\}, & h \neq 0 \end{cases}$



Functions	Models	
Covariance function	<p><b>Exponential:</b></p> $C_y(h) = c_0 \cdot e^{-a^2 h ^2} \quad h \neq 0$ $C_y(0) = \sigma_y^2 = c_0 + \sigma_e^2 \quad h = 0$	<p><b>Wave:</b></p> $C_y(h) = \begin{cases} \sigma_y^2, & h = 0 \\ c_0 - c_w \left\{ 1 - \frac{a_w}{ h } \sin\left(\frac{ h }{a_w}\right) \right\}, & h \neq 0 \end{cases}$

각각의 공간상관함수 모델을 이용하여 경험적으로 계산된 값을 Table 2에 제시한 모델함수에 피팅함으로써 경험적인 공간상관함수를 추정하게 된다. 함수 모델은 다양한 형태가 제시되어 있으나, 본 연구에서는 각 공간상관함수별 특성에 가장 적합한 두 개의 함수만을 적용하여 비교 분석하였다.

공간상관함수의 파라미터 추정을 위해서는 최소제곱 방법을 적용했으며, 가중값은 거리의 제곱에 반비례하도록 설정하였다. Lag 거리(h)가 0인 경우는 함수 피팅을 위해 관측오차의 분산값인  $\sigma_e^2$ 을 적용하였으며, 가중값이 발산하는 것을 방지하기 위해 h는 매우 작은 값(0.01 km)으로 설정했다. Fig. 6은 경험적인 semi-variogram을 wave 모델에 적용한 것으로서, lag 거리가 약 100 km 이내에서는 모델 적합도가 우수하다. Lag 거리가 0인 경우에는 Table 1의 정의에 따라 0의 값을 가지지만 함수의 수렴값은  $\sigma_e^2$ 이 된다.

으므로, 계산된 경험적 homeogram은 특정 샘플링 포인트에 과도한 영향을 받을 가능성이 존재한다. 따라서 이 문제를 해결하기 위해서는 별도의 가중값 부여 방법을 고안할 필요가 있으며, 각 구간에서 반영 횟수가 다르기 때문에 가중값을 부여하는 것이 매우 복잡하다. 따라서 본 연구에서는 원래의 샘플링 포인트(즉, h=0) 값을 각 구간에서 출현 횟수에 따라 스케일링 하였다. 예를 들어, 어떤 구간에서 점 A가 p 회, 점 B가 q 회 나타났다고 하면, 스케일링 된 homeogram은 Eq. (9)와 같이 계산한다.

$$\gamma_y(h) = \frac{p \cdot A^2 + q \cdot B^2}{p + q} \tag{9}$$

따라서 lag 거리가 d ( $d \neq 0$ )인 모든 경험적인 homeogram은 스케일 된 homeogram과의 상대적인 값의 크기인 상대비율에 기반하여 최종 경험적 homeogram을 계산한다(Fig. 7). Covariance function을 경험값으로 추정된 결과는 Fig. 8과 같으며, homeogram과 마찬가지로 h=0에서 관측값 오차의 분산값에 해당하는 크기의 불연속이 발생한다.

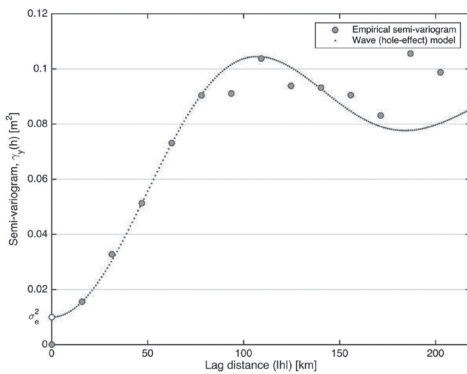


Fig. 6. The empirical semi-variogram and wave model fitting (Bae and Schaffrin, 2015)

경험적인 homeogram은 공간상의 두 점을 하나의 쌍으로 조합하는데, 구간별 조합 숫자가 다르므로 결과적으로 경험값 계산에 반영되는 각 샘플 포인트의 비중이 서로 다르다. 따라서 각 샘플링 포인트는 서로 다른 지오이드 값을 가지고 있

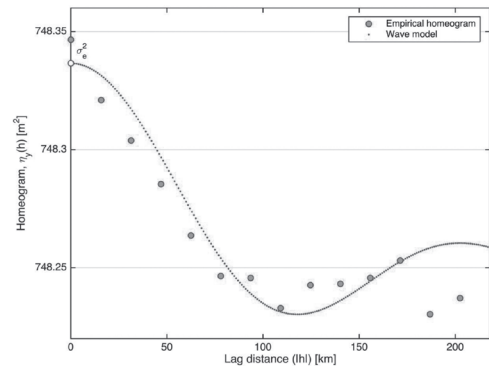


Fig. 7. The empirical homeogram of y and wave model fitting. The homeogram is offset by  $\mu^2$  from the covariance function. There is a jump by  $\sigma_e^2$  at h=0 (Bae and Schaffrin, 2015)

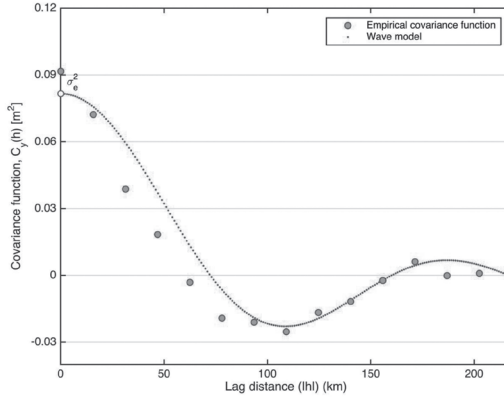


Fig. 8. The empirical covariance function of  $y$  and wave model fitting. There is a jump by  $\sigma_c^2$  at  $h=0$  (Bae and Schaffrin, 2015)

Table 3. Quality of fit for each spatial coherency function

Semi-variogram		Homeogram		Covariance function	
Spherical	Wave	Exponential	Wave	Exponential	Wave
1.5201e-5	<b>0.5616e-7</b>	<b>1.7889e-7</b>	6.3970e-7	<b>2.4630e-7</b>	4.7590e-7

Note: A large number is assigned to the case of lag distance 0 instead of (stochastic) constraints.

각 공간상관함수에 대한 모델 적합도는 최소제곱조정 후 계산한 잔차를 이용하여 표현할 수 있다. Table 3에서 보는 것처럼 공간상관함수에 따라서 최적 모델이 다르며, 특히 동일한 wave 모델을 적용했을 때 semi-variogram의 모델 적합도가 가장 좋은 것으로 나타났다. 그러나 모델 적합도는 사용한 지오이드에 기반하여 계산한 경험값에 의존하므로, 다른 영역에서도 일관된 경향을 보이는지에 대해서는 추가적인 분석이 필요하다. 또한 공간상관함수에 대한 모델 적합도가 높다고 하더라도 크리깅 방식에 따라 평균제곱오차의 특성이 동일하지 않을 수 있으므로, 향후 연구에서는 이에 대한 분석이 필요하다.

#### 4. 요약 및 결론

랜덤프로세스에 기반한 공간 데이터의 분석 및 예측을 위해서 많이 활용되고 있는 크리깅을 적용하기 위해서는 관측된 데이터를 기반으로 공간적인 상관성을 나타내는 공간상관함수를 추정할 필요가 있다. 본 연구에서는 다양한 크리깅 방법에 공통적으로 적용할 수 있는 공간상관함수를 결정하고 비교하였으며, 대상 모델은 semi-variogram, homeogram, covariance function이다. 통합기준점에서의 GPS/leveling 자

료와 정표고 자료를 이용하여 계산한 지오이드와 국가지오이드 모델을 비교한 지오이드 관측오차는 약  $\pm 2.9$  cm이며, 이를 기반으로 공간상관함수의 관측오차를 10 cm로 설정하였다. 공간적인 편향성을 제거하기 위해 대상 지역의 지오이드를 선형접합하여 경향성을 제거하고, 잔차 지오이드를 이용하여 공간상관함수를 추정하는데 이용하였다. 경험적인 값을 계산하기 위해 본 연구에서 선정한 100개의 샘플 포인트를 이용하여 두 점 사이의 공간적인 거리에 따라 약 16 km 간격으로 13개 구간을 나누어 모델별 경험값을 계산하였다. 각 함수별로 특성에 잘 부합하는 두 개의 모델을 선택하여 적용하였으며, 각 함수별 직접적인 비교를 위해 wave 모델을 공통적으로 사용하였다. Lag 거리의 역수를 가중값으로 하여 최소제곱조정한 결과 semi-variogram의 wave 모델의 적합도가 가장 우수한 것으로 나타났다. 또한 homeogram과 covariance function의 경우 wave 모델 보다는 exponential 모델에 유사한 것으로 판단된다. Homeogram의 경우 평균값을 고려하지 않고 공간상관함수를 추정하기 때문에 특정 위치에서의 지오이드 값에 과도한 영향을 받을 가능성이 있다. 따라서 본 연구에서는 경험적 homeogram 계산을 위해 각 구간에서의 homeogram을 스케일링하여 상대비율에 기반한 homeogram 계산 방법을 제안하였다. 그러나 최종적으로 결

정된 공간 프로세스의 추정값 및 평균제곱예측오차는 선형적으로 변하지 않으므로(Bae and Schaffrin, 2015), 공간상관 함수의 모델 적합도를 판정하기 위해서는 이에 대한 추가적인 분석이 필요하다.

## 감사의 글

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2015R1D1A1A01059295)

## References

- Bae, T.S. and Schaffrin, B. (2015), On various Kriging predictors for geoid densification: a comparison, *SIAM Conference on Mathematical and Computational Issues in the Geosciences*, SIAM, 29 June – 2 July, Stanford University, California, USA.
- Cressie, N. (1993), *Statistics for Spatial Data (2nd)*, Wiley, New York, N.Y.
- Jekeli, C. (2001), *Inertial Navigation Systems with Geodetic Applications*, Walter de Gruyter, Berlin, New York.
- Jekeli, C. (2003), *Fourier Geodesy*, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, Ohio.
- NGII (2016), Current geoid models of Korea, *National Geographic Information Institute*, Suwon, Korea, [http://ngii.go.kr/geoid/intro/geoid\\_condition.do](http://ngii.go.kr/geoid/intro/geoid_condition.do) (last date accessed: 15 January 2016). (in Korean)
- Schaffrin, B. (1993), Biased Kriging on The Sphere?, In: Soares, A. (ed.), *Geostatistics Tróia '92*, Springer Netherlands, Vol. 1, pp. 121-131.
- Schaffrin, B. (2000), Establishing equivalent systems for universal Kriging, *9<sup>th</sup> International Workshop on Matrices and Statistics*, December 9-13, Hyderabad, India.
- Schaffrin, B. (2013), Towards unified computational schemes for various forms of Kriging, *2013 Spatial Statistics Conference*, Columbus, Ohio.