

# 공간 자기상관함수의 이해 및 국내 기온 예측

유범주, 정성준, 한만취, 홍상혁

Department of Mathematics, Hanyang University, Seoul 133-791, Republic of Korea

## 1. 서론

본 연구에서는 공간 자기상관함수(Spatial Autocovariance function)의 이론적 기반을 탐구한다. 이를 활용한 Kriging 예측 모델을 소개하며 우리나라의 기온을 예측하고 공간 자기상관함수들 간의 성능을 비교한다. 추가로 Kriging 외에 기온을 예측할 수 있는 다른 예측 모델들을 소개하고 그들의 성능을 확인한다. 그리고 기존 모델의 성능 향상을 위한 데이터 전처리 방법을 추가로 고안해본다. 모든 비교는 k-겹 교차 검증(k-fold cross validation)을 통해 진행되며 Mean Squared Error (MSE)를 비교하여 모델의 우수성을 판단한다.

## 2. 데이터 소개

본 연구에서는 2007년부터 2022년까지 5년 간격 계절별 기온의 평균값 자료를 이용하였다. 기상청<sup>1</sup>에서 제공하는 월 평균 기온의 종관(ASOS) 및 방재(AWS) 자료를 이용하였다. 2022년 가을 기준 운영되고 있는 관측소의 개수는 종관 기상 관측소 96개, 방재 기상 관측소 525개이다.

그림 1은 2022년 가을 기준 국내 관측소의 위치 분포를 나타낸 그림이다.

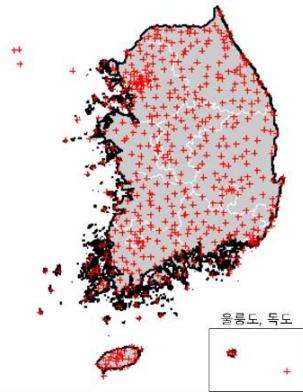


그림 1. 국내 관측소 위치

데이터가 측정된 년도를 기준으로 해당년도 3월~5월을 봄, 6월~8월을 여름, 9월~11월을 가을, 12월~이듬해 2월을 겨울로 계절을 분류하였다. 한 계절 내에 위치를 옮긴 관측소 데이터는 제외하고 진행하였다.<sup>2</sup> 아래 표 1은 기온의 기초통계량, 그림 2, 3은 데이터의 Density plot 및 2022년 가을 Variogram Fitting이다.

표 1. 2022년 기온(°C) 데이터 기초 통계량							
	Min.	Q1.	Q2.	Mean	Q3	Max.	SD.
봄	3.23	11.63	12.77	12.53	13.73	16.07	1.72
여름	15.07	23.47	24.37	24.02	25.07	27.17	1.60
가을	6.20	13.23	14.83	14.67	16.10	21.10	2.22
겨울	-10.00	-1.95	0.03	0.09	2.10	9.30	3.09

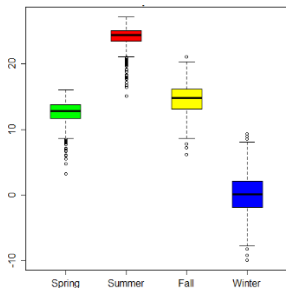


그림 2. Density Plot

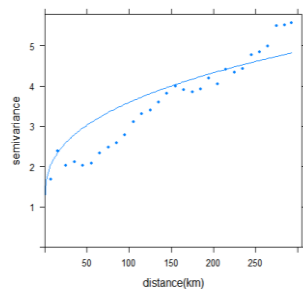


그림 3. Fall Variogram Fitting

## 3. Study 1: Spatial Autocovariance function

공간 자기상관함수를 바꿔가며 기온 예측 정확도의 차이를 관찰하였다. 한 계절의 데이터를 10개의 partition으로 나누어 9개의 데이터로 나머지 하나의 기온을 예측하고 MSE의 평균값을 구하였다. 자세한 식은 다음과 같다.

$$(1) \text{MSE}_i = \sum_{s \in P_i} \frac{(\hat{Z}(s) - Z(s))^2}{n(P_i)} \quad (2) \text{MSE} = \frac{1}{10} \sum_{i=1}^{10} \text{MSE}_i$$

식 (1)은 한 partition의 MSE, 식 (2)는 (1)에서 구한 MSE들의 평균으로 한 계절의 MSE이다. 여기서,  $\hat{Z}(s)$ 는  $s$ 에서의 기온 예측값,  $Z(s)$ 는  $s$ 에서의 기온 측정(실제)값,  $n(P_i)$ 는  $i$ 번째 partition( $P_i$ )의 데이터의 수이다.

예측은 Kriging으로 진행하였으며, Kriging Estimator는 다음과 같다.

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

아래 그림 4는 MSE 비교 결과이다. 모두 Gaussian Process를 기반으로 진행하였으며 Model 1(a), (b), (c)는 각각 Exponential, Matérn, Gaussian Autocovariance function을 사용하였다.

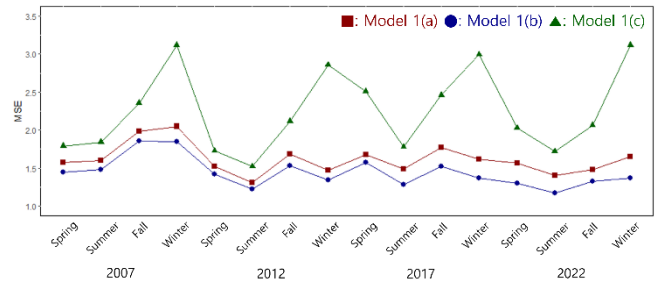


그림 4. Comparing Autocovariance functions

Model 1(b)에서 가장 낮은 MSE가 나오는 것을 확인하였고 Model 1(a), Model 1(c)순으로 MSE가 낮게 나왔다. Model 1(b)에 사용된 자기상관함수는 다른 모델보다 parameter를 하나 더 사용하며, parameter 값을 조정하면 다른 두 자기상관함수를 표현할 수 있기에 성능이 제일 뛰어날 것으로 예측하였고, 실제로 그 결과가 옳은 것을 확인하였다.

## 4. Study 2: Prediction models

다양한 공간 예측 모델을 선택하여 그 성능을 비교해보았다. 성능은 위와 동일하게 MSE로 비교하였다. Model 1, 2, 3, 4는 각각 Study 1의 Model 1(b), Maximum Likelihood Estimator (MLE) Method, Inverse Distance Weighted (IDW), Thin Plate Spline (TPS)를 나타낸다.

MLE Method는 자기상관함수의 Parameter의 MLE를 구하여 Kriging에 적용한 방법으로, 예측 방법은 Model 1과 동일하다. IDW는 가중치를 거리의 역수로 설정하여 예측하는 Model이고, TPS는 Spline에 기반을 둔 예측 Model이다.

<sup>1</sup> <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36>

<sup>2</sup> 삭제된 데이터의 개수는 204 개이다.

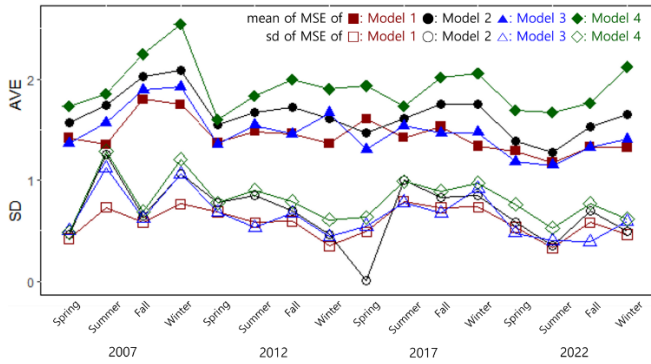


그림 5. Comparing Autocovariance functions

위 그림은 교차검증 partition들의 MSE 평균과 표준편차를 나타낸다. 전반적으로 Model 1이 제일 우수한 성능을 보였고, 그 다음으로는 Model 3, 2, 4 순으로 좋은 성능임을 확인할 수 있었다. MLE Method에서는 Parameter의 MLE를 추정하기에 예측 성능이 향상될 것을 기대하였으나, 오히려 MSE는 더 높게 나오는 것을 확인하였다. Model 3도 MSE가 Model 1, 2와 큰 차이가 나지 않았는데, 이는 대한민국의 지역이 그렇게 크지 않고 크기 대비 많은 관측소가 존재하여 비교적 괜찮은 성능을 보인 것으로 추측된다. 마지막으로 수치해석적 방법인 Model 4는 MSE가 제일 크게 나오는 것으로 보아 공간 데이터 분석에는 공간 통계 기법이 사용되어야 함을 확인하였다.

아래 그림 6은 각 모델로 그려본 국내 기온 지도이다.

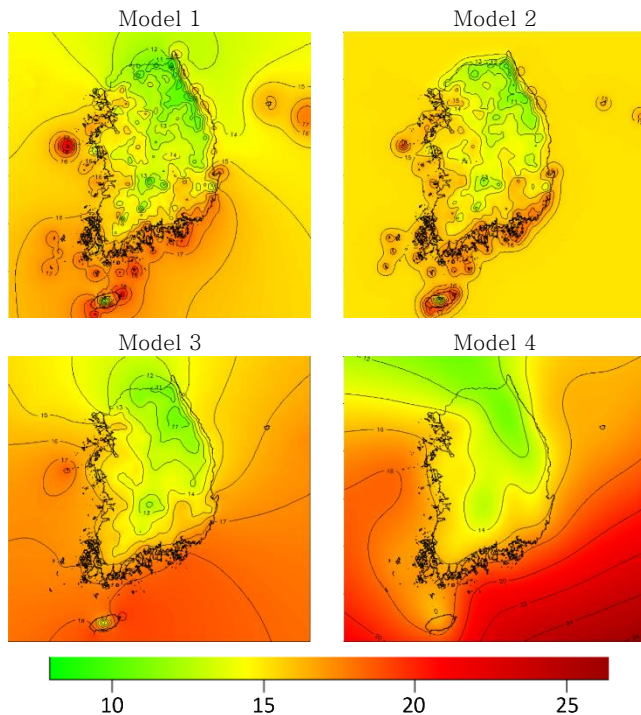


그림 6. 2022년 가을 국내 기온(°C) 지도

### 5. Study 3: Data adjustment

지리적 인자 중 기후요소의 시·공간 변동에 가장 큰 영향을 미치는 인자는 해발고도이다. (박노옥, 장동호, 2008, p. 1005). Study 1, 2는 해발고도를 고려하지 않고 진행된 연구이기에 해발고도를 조정하여 추가 연구를 진행하였다.<sup>3</sup> 추가로 섬의 기온은 그 주변 바다의 영향을 받기에, 섬에 위치한 관측소들도 제외한 후 연구를 진행하였다.<sup>4</sup> 기존 데이터를 Data 1, 해발고도만 조정된 데이터를 Data 2, 섬을 제

외한 데이터를 Data 3, 두 보정을 모두 적용한 데이터를 Data 4로 명명하였다.

보정 후 변화된 2022년 가을 데이터의 기초 통계량은 다음 표 2에, Model 1을 이용해 MSE를 비교해본 결과는 표 3에 나타나 있다.

Data	Min.	Q1.	Q2.	Mean	Q3	Max.	SD.
1	6.20	13.10	14.80	14.61	16.09	21.10	2.24
2	11.33	14.28	15.29	15.50	16.53	21.73	1.67
3	6.20	13.03	14.68	14.42	15.83	18.80	2.06
4	11.33	14.18	15.15	15.25	16.19	19.21	1.44

	봄	여름	가을	겨울
Data 1	1.332	1.175	1.428	1.413
Data 2	0.387	0.199	0.633	0.546
Data 3	1.173	1.104	1.137	1.173
Data 4	0.396	0.187	0.586	0.544

해발고도 조정은 유의미한 결과 향상을 가져왔다. 그에 반해 섬을 제외하는 경우에는 결과에 큰 변화는 없었다. 기존 데이터를 활용하더라도 거리가 먼 경우 낮은 가중치를 부여하였기 때문에, 섬을 제외해도 내륙지방 기온 예측에 끼치는 영향은 적을 것으로 추측된다.

### 5. 결론 및 향후 과제 제언

Gaussian process에 사용되는 Autocovariance function과, 공간 데이터 분석에 사용되는 Kriging을 포함한 여러가지 공간 통계 기법에 대한 분석을 진행하였다. Autocovariance function에는 여러 종류가 있지만, 본 연구에서 사용한 세 가지의 함수 중에서는 Matérn autocovariance function이 제일 좋은 성능을 보임을 확인하였고, 예측 방법에서는 Model 1이 제일 좋은 성능을 보임을 확인하였다. 추가로 Model 4의 성능이 제일 낮은 것으로 보아 공간 데이터 분석에는 공간 통계 모델이 사용되어야 함을 알 수 있었다.

이번 연구에서는 Autocovariance function의 Isotropic을 가정하고 진행하였다는 점, 관측소와 비교적 거리가 멀리 떨어진 바다 위의 기온 예측은 유의미하다고 보기 힘들다는 점 등의 한계이 있다. 추후 Isotropic하지 않을 경우 예측 모델의 변화와 바다 위에서의 기온을 예측하는 대에 적절한 모델에 대한 연구를 진행할 계획이다.

### 6. 참고문헌

- [1] 박노옥, 장동호. (2008). 수치표고모델과 다변량 크리깅을 이용한 기온 및 강수 분포도 작성. 대한지리학회지, 43(6), 1002-1015.
- [2] 장승민, 김성수, 최영찬, 김수강. (2006). 제주도 기온과 주변해역 해수면 온도와의 상관관계에 관한 연구. 한국해양환경·에너지학회지, 9(1), 55-62.
- [3] 정재홍. "Geostatistical Data:Spatial Big Data Analysis with GIS" Korean Statistical Society Winter School, 2023년 2월 24일.
- [4] Cressie, Noel, & Wikle, Christopher K. (n.d.). *Statistics for Spatio-Temporal Data* (pp. 125-145). n.p.: John Wiley & Sons Inc.
- [5] Chris Brunsdon, & Lex Comber. (n.d.). *R for spatial analysis & mapping second edition* (pp. 13-107). n.p.: SAGE.

<sup>3</sup> 해발고도 100m 상승 시 기온은 0.6 °C 감소한다고 가정하였다.

<sup>4</sup> 경도 126° ~ 129.5°, 위도 34.1° ~ 43° 사이에 속하지 않는 관측소를 제외하였고, 제외된 관측소의 개수는 60개이다.