

## 베이지안 시공간 모형을 이용한 한반도 오존농도 예측\*

박다인<sup>1</sup>, 윤상후<sup>2</sup>

### 요 약

한반도의 대기오염 상태는 계절과 지역에 따라 다르게 나타나기 때문에 한반도 오존농도 분포를 시공간적 분포로 나타낼 수 있다. 본 연구에서는 한반도 오존농도를 계층적 베이지안 시공간모형을 이용하여 예측하고 시각화 하였다. 베이지안 시공간모형의 확률과정으로 독립가우시안 모형(GP model)과 자기상관모형(AR model)을 이용하였다. 연구에 사용된 자료는 우리나라 대기오염 관측소와 기상관측소에서 관측된 2014년부터 2016년까지 3년간 오존농도자료와 기상요인자료이다. 기상요인은 최대기온, 평균풍속, 상대습도, 일사량을 사용하였다. 대기오염 관측소의 위치와 기상관측소의 위치가 상이하여 기상관측소로부터 5km이내에 있는 33개의 대기오염 관측소만 이용하여 분석을 수행하였다. 베이지안 시공간 모형에서는 공간상관성을 설명하는 모수인  $\phi$ 의 사전분포를 설정하는 것이 중요하다. 그러나  $\phi$ 의 사후분포가 수렴하지 않아 한반도 오존농도의 RMSE가 가장 낮은 최적의  $\phi$ 값을 찾았다. Geweke 진단법을 이용하여 베이지안 모형의 MCMC 수렴도를 평가하였고 RMSE, MAE, MAPE, BIAS를 이용하여 GP모형과 AR모형을 비교분석 하였다. 마지막으로 97개소 ASOS 기상데이터를 이용하여 울릉도, 제주도, 백령도 등 대기오염 관측소가 없는 지역까지 한반도의 전역의 오존농도를 시각화 하였다.

주요용어 : 가우시안확률과정, 베이지안, 시공간모형, 오존농도, 자기회귀.

### 1. 서론

오존은 성층권에서 태양의 자외선을 막아주는 지구 생태계 유지를 위해 꼭 필요한 대기물질이지만 고농도의 오존이 지표면에서 발생하면 동물과 식물에 악영향을 미친다. (NRC, 1992; Choi, Park, Yang, 2017). Tilton(1989)에 따르면 기관지를 통해 인체 내부로 유입된 오존은 만성 호흡기 질환, 폐렴, 폐활량 감소, 생체 면역능력 감소 등 기관지 질환을 유발한다. Choi, Park, Yang(2017)에서는 오존농도 1% 증가가 호흡기계 질환으로 인한 내원 확률을 증가시키는 것을 입증하였다.

한반도의 경우 1990년 초 대도시를 중심으로 오존 문제가 제기되어 1995년부터 오존 경보제를 실시하였다. 오존의 피해를 줄이기 위한 노력에도 불구하고 지표면의 고농도 오존 발생 빈도는 증가 추세에 있으며 관심 지역 또한 확대되고 있다.

환경부에서는 오존농도가 일정 기준치 이상이 되면 오존 경보제를 내려 국민들의 생활환경에

\*이 논문은 제1저자 박다인의 석사학위논문의 축약본입니다.

<sup>1</sup>38453 부산시 해운대구 센텀7로 12, APEC기후센터 기후분석과, 원급연구원. E-mail : dain105@apcc21.org

<sup>2</sup>(교신저자) 38453 경상북도 경산시 진량읍 대구대로 201, 대구대학교 과학생명융합대학 수리빅데이터학부, 조교수. E-mail : statstar@daegu.ac.kr

[접수 2020년 4월 28일; 수정 2020년 6월 10일; 게재확정 2020년 6월 11일]

오존이 미치는 영향을 최소화하고자 노력하고 있다. 오존 경보제는 주의보, 경보, 중대경보 3단계로 발령되며 1시간 오존 평균 농도의 기준은 각각 0.12, 0.3, 0.5 ppm이다. 1ppm은 100만 분의 1을 표시하는 농도 단위로 1톤당 1g씩 나타나는 물질의 농도를 의미한다. ppb는 ppm의 1,000배이며 본 연구에서는 ppb를 기본단위로 사용하였다.

지표오존은 강한 햇빛 아래에서 질소산화물과 휘발성 유기화합물질이 반응하여 생성된다. 자동차와 공장 등에서 주로 배출되는 질소산화물은 일산화질소로 주로 대기 중 오존 또는 발생기 산소와 결합하여 이산화질소로 변환된다. 이산화질소는 강한 햇빛에 노출되면 일산화질소와 산소원자로 다시 광분 되고 생성된 산소 원자는 대기 중 산소와 결합하여 오존으로 변환한다. 생성된 오존은 일산화질소와 결합하여 다시 이산화질소를 생성시키는 순환을 되풀이한다.

2차 오염물질인 오존의 생성은 기상조건과도 밀접한 관련이 있다. Vukovich(1995)는 오존이 기온, 일사량, 풍향, 그리고 풍속과 같은 기상조건에 영향을 받는다고 하였다. Jeon, Woo(1999)는 기온, 일사량, 풍속은 오존과 양의 상관관계를 보이고 상대습도는 오존과 음의 상관성이 있음을 보였다.

기상대기는 시공간적 특성을 지니고 있으므로 오존농도를 예측하기 위해선 시간과 공간의 특성이 모두 반영되어야한다(Oh, Kim, 2002). 다양한 연구에서 대기오염 물질의 시공간적 특성을 반영하여 연구를 수행하였다(Kim, Yoon, 2016; Jang, Lee, Lee, 2012; Oh, Kim, 2017; Moon, 2019; Kim, Kim, Choi, 2016).

본 연구는 한반도에 설치 운영 중인 대기오염 관측소와 기상관측소에서 관측된 2014년부터 2016년까지 3년간 오존농도자료와 기상요인자료를 이용하였다. 공간적 범위는 한반도 전역을 대상으로 설정하였고 시간 해상도는 일 단위이다. 시공간 모형에 사용된 대기오염 관측소의 위치와 기상관측소의 위치가 상이하여 기상관측소로부터 5km 이내에 있는 33개의 대기오염 관측소만 이용하여 분석을 수행하였다.

## 2. 연구방법론

### 2.1. 베이지안 시공간모형

베이지안 시공간모형은 두 단계의 계층적 구조로 설명된다(Gelfand, 2012). 첫 번째 단계는 관측자료 기반 확률과정을 나타내고, 두 번째 단계는 시공간적 랜덤 효과를 나타낸다.

$$1 \text{ 단계: } Z_t = O_t + \epsilon_t \text{ [data | process, parameter]}, \quad (1)$$

$$2 \text{ 단계: } O_t = X_t \beta + \eta_t \text{ [process | parameter]}, \quad (2)$$

$$3 \text{ 단계: prior distribution [parameter]}. \quad (3)$$

1단계는 실제 관측치  $Z_t$ 를 시공간확률과정  $O_t$ 와 오차항  $\epsilon_t$ 로 설명한다. 공간  $s_i (i=1, \dots, n)$ 과 시간  $(l, t)$ 이 고려된 종속변수  $Z_l(s_i, t)$ 는  $Z_l = (Z_l(s_1, t), \dots, Z_l(s_n, t))'$ ,  $O_t = O(O_l(s_1, t), \dots, O_l(s_n, t))'$ 인 확률과정과 오차항  $\epsilon_t = (\epsilon_l(s_1, t), \dots, \epsilon_l(s_n, t))'$ 로 구성된다.  $t$ 는 장시간  $l(\text{year}, l, \dots, r)$ 과 단시간  $t(\text{day}, t=1, \dots, T)$ 를 포함하고 있으며  $\epsilon_t$ 는 정규분포  $N(0, \sigma_\epsilon^2 I_n)$ 을 따른다. 2단계는 시공간확률과정  $O_t$ 로 시공간 특성이 반영된 디자인행렬  $X_t$ 와 회귀계수  $\beta$  그리고 시공간 랜덤효과  $\eta_t$ 로 표현할 수 있다.

여기서  $\eta_{lt} = (\eta_l(s_1, t), \dots, \eta_l(s_n, t))'$ 는 시간에 따라 독립성  $N(0, \Sigma_\eta)$ 을 가정한다. 분산 공분산행렬  $\Sigma_\eta$ 은  $\sigma_\eta^2 S_\eta$ 으로 나타낸다.  $\sigma_\eta^2$ 는 공간에 따라 변하지 않는 분산이고,  $S_\eta$ 는 공간 상관행렬로 본 연구에서는 지수상관함수를 가정하였다.

$$k(s_i, s_j, \phi) = \exp(-\phi|s_i - s_j|), \quad \phi > 0. \quad (4)$$

여기서 모수  $\phi$ 는 거리  $|s_i - s_j|$ 에 따라 공간 감쇠율을 의미한다(Cressie, 1992). 3단계는 베이지안 추론을 위한 사전분포로 2.4절에서 다룬다.

## 2.2. 독립 가우시안 모형(Independent Gaussian process model, GPmodel)

독립 가우시안 회귀 식은 아래와 같이  $Z_{lt}$ 와  $O_{lt}$ 로 정의된다.

$$Z_{lt} = O_{lt} + \epsilon_{lt}, \quad (5)$$

$$O_{lt} = X_{lt}\beta + \eta_{lt}. \quad (6)$$

여기서  $\epsilon_{lt}$ 와  $\eta_{lt}$ 는 서로 독립인 가우시안 확률과정을 따른다. 디자인행렬인  $X_{lt}$ 는  $n \times p$  행렬이며 공변량  $p$ 가 존재한다.  $\beta = (\beta_1, \dots, \beta_p)'$ 는 회귀계수로  $p \times 1$  벡터이다.  $\theta = (\beta, \sigma_\epsilon^2, \sigma_\eta^2, \phi)$ 를 GP모형을 구성하는 모수라 하자.  $\pi(\theta)$ 가 사전분포라 하면 GP모형의 결측값과 모수들의 결합사후분포에 로그를 취한 식은 아래와 같다.

$$\begin{aligned} \log \pi(\theta, O, z^*|z) \propto & -\frac{N}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{l=1}^r \sum_{t=1}^{T_l} (Z_{lt} - O_{lt})' (Z_{lt} - O_{lt}) - \frac{1}{2} \sum_{l=1}^r T_l \log |\sigma_\eta^2 S_\eta| \\ & - \frac{1}{2\sigma_\eta^2} \sum_{l=1}^r \sum_{t=1}^{T_l} (O_{lt} - X_{lt}\beta)' S_\eta^{-1} (O_{lt} - X_{lt}\beta) + \log \pi(\theta). \end{aligned} \quad (7)$$

여기서  $z^*$ 는 결측값을 의미한다. 오존 농도의 결측값은 김스샘플링의 조건부분포로부터 추정하였고 조건부분포는 초깃값에 영향을 받는 3,000개를 번인(burn-in)한 13,000번 반복시행을 통해 구했다. 김스샘플링을 위해 R 패키지 spTimer의 spT.Gibbs함수가 사용되었다(Bakar와 Sahu, 2015). GP모형의 완전 조건부 분포는 Bakar, Sahu(2015)에 제시되어있다.

## 2.3. 자기상관모형(Autoregressive models, AR model)

AR모형은 Sahu, Mardia(2005)가 제안하였다. AR모형에서도 GP모형과 마찬가지로 첫 번째 계층은 실제자료기반 확률과정을 나타낸다.

$$Z_{lt} = O_{lt} + \epsilon_{lt}, \quad (8)$$

$$O_{lt} = \rho O_{lt-1} + X_{lt}\beta + \eta_{lt}. \quad (9)$$

GP모형과 달리 AR모형은 확률과정  $O_{lt}$ 의 자기상관성  $\rho$ 가 반영되어 있고  $\rho$ 의 범위는 (-1,1)이다. 초기 조건은  $O_{l0} = \gamma_1 + \mu \mathbf{1}$ ,  $l = 1, \dots, r$  이고  $\gamma_l = (\gamma_l(s_1), \dots, \gamma_l(s_n))'$ 는  $l = 1, \dots, n$ 에 대해 독립인 가우시안 확률과정을 따른다. AR모형에서  $\rho = 0$ 일 때 GP모형과 같게 된다.  $\pi(\theta)$ 가 사전분포를 나타낼 때 AR모형의 결측치와 모수들의 결합사후분포에 로그를 취한 식은 아래와 같다.

$$\begin{aligned}
\log \pi(\theta, \mathbf{O}, \mathbf{z}^* | \mathbf{z}) \propto & -\frac{N}{2} \log \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{l=1}^r \sum_{t=1}^{T_l} (\mathbf{z}_{lt} - \mathbf{O}_{lt})' (\mathbf{z}_{lt} - \mathbf{O}_{lt}) - \frac{1}{2} \sum_{l=1}^r T_l \log |\sigma_\eta^2 S_\eta| \\
& - \frac{1}{2\sigma_\eta^2} \sum_{l=1}^r \sum_{t=1}^{T_l} (\mathbf{O}_{lt} - \rho \mathbf{O}_{lt-1} - \mathbf{X}_{lt} \beta)' S_\eta^{-1} (\mathbf{O}_{lt} - \rho \mathbf{O}_{lt-1} - \mathbf{X}_{lt} \beta) \\
& - \frac{1}{2} \sum_{l=1}^r \log |\sigma_l^2 S_0| - \frac{1}{2} \sum_{l=1}^r \frac{1}{\sigma_l^2} (\mathbf{O}_{l0} - \mu_l)' S_0,
\end{aligned} \quad (10)$$

여기서  $\mathbf{z}^*$ 는 결측값이다. GP모형과 같이 AR모형의 오존 농도 결측값은 13,000번 반복 시행 하여 초깃값에 영향을 받는 3,000개를 번인(burn-in)한 깃스샘플링의 조건부분포로부터 추출하였다. AR모형의 완전 조건부 분포는 Bakar, Sahu(2012)에 제시되어있다.

## 2.4. 사전분포

베이지안 추론에서 사전분포는 데이터를 얻기 전 모수에 대한 정보이므로 사전분포의 설정이 중요하다. 모수에 관한 사전정보가 미흡하거나 확신이 없을 경우에는 무정보적 사전분포를 설정한다. 무정보적 사전분포는 우도함수에 비해 상대적으로 평평하여 사후분포에 최소한의 임무만 수행한다. 본 연구에서는 각 모수의 사전분포를 선행연구를 참조하여 설정하였다(Sahu, Baker, 2012; Kim, Yoon, 2016; Pi, Hwang, 2018).  $\beta$ 의 사전분포로 정규 분포  $\sim N(0, 10^4)$ 를 고려하였고, 시공간적 랜덤효과에 대한 오차항  $\sigma_\eta^2$ 과 통계적 랜덤효과에 대한 오차항  $\sigma_\epsilon^2$ 의 사전분포로 역감마 분포  $\sim IG(2, 1)$ 를 고려하였으며, 공간모수  $\phi$ 의 사전분포로 균등분포, 정규분포, 감마분포를 고려하였다.

## 2.5. 오존농도 예측

본 연구에서 종속변수는 일 최대 8시간 평균 오존 농도이며 독립변수는 최고기온, 바람세기, 상대습도, 일조량이다. 위치  $s'$ 와 시간  $t'$ 에서 공간 예측은 종속변수  $Z_l(s', t')$ 에 대한 사후분포로부터 계산된다. 종속변수  $Z_l(s_0, t')$ 에 대한 사후분포는 다음식이다(Bakar, Sahu, 2015).

$$\pi(Z_l(s', t') | \mathbf{z}) = \int \pi(Z_l(s', t') | Q_l(s_0, t'), \sigma_\epsilon^2) \pi(Q_l(s', t') | \theta, \mathbf{z}^*) \pi(\theta | \mathbf{z}) dQ_l(s', t') d\theta dz^*. \quad (11)$$

먼저 사후분포  $\pi(\theta, \mathbf{z}^* | \mathbf{z})$ 로부터 깃스샘플링을 이용하여 랜덤샘플  $\theta^{(j)}$ 를 추출한다. 베이지안 크리깅을 적용하여  $Q_l(s', t')$ 의 조건부분포에서  $Q_l^{(j)}(s', t')$  샘플을 얻는다. 깃스샘플링의 마지막 단계에서 베이지안 크리깅을 이용하여  $Z_l^{(j)}(s', t'), j=1, \dots, J$ 을 예측하고 평균, 중앙값, 95% 신용구간을 사용하여 깃스샘플링 예측치를 요약 한다. 베이지안 크리깅은 지구 통계학적 기법으로 이미 알고 있는 주위의 값들의 선형 조합으로 관심 있는 위치의 값을 예측하는 방법이다.

## 2.6. 예측성능비교

베이지안 시공간모형의 성능을 평가하기 위해 예측치와 실제 관측치의 차이에 대한 통계적 평가방법인 평균제곱오차(RMSE), 평균절대오차(MAE), 평균절대백분율오차(MAPE) 그리고 편의(bias)를 이용하였다. RMSE는 오차의 제곱을 평균하여 제곱근 취한 값이다. MAE는 오차의 절대값의 평

균이고 **MAPE**는 실제값에 대한 **MAE** 비율이다. 편의(**bias**)는 예측된 값이 편향된 정도를 의미한다. 각 예측적도는 예측치와 관측치에 대한 정밀도를 나타내기 때문에 값이 작을수록 좋다.

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad \text{MAE} = \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|} \\ \text{MAPE} &= \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|, \quad \text{BIAS} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i). \end{aligned} \quad (12\sim 15)$$

여기서  $n$ 은 전체 관측치의 수이고,  $y_i$ 는 관측치를 나타내며,  $\hat{y}_i$ 은  $i$ 번째 예측치를 나타낸다.

### 3. 연구자료

지표면 오존농도는 기상요소인 일 최고기온, 바람세기, 상대습도 그리고 일조시간에 영향을 받는다(Jeon, Woo, 1999). 기상청에서 운영 중인 지상기상관측장비는 종관기상관측장비와 자동기상관측장비가 있다. 상대습도와 일조시간은 종관기상관측장비에서만 관측 가능하기 때문에 본 연구에서는 종관기상관측장비로부터 수집된 일 최고기온, 바람세기, 상대습도, 일조량을 이용하였다. 연구기간은 2014년부터 2016년까지 총 3년이다.

Figure 1은 연구에 사용된 대기오염관측소와 종관관측소의 위치를 시각화한 그림이다. 대기오염관측소는 총 348개소이고, 종관기상관측장비는 총 97개소이다. 대부분의 대기오염관측소는 경기도와 남해지역에 밀집되어 있으며 중부내륙지방은 밀도가 상대적으로 낮았다. 하지만 기상관측소 위치와 대기오염관측소가 상이하므로 공간적 매칭을 위해 대기오염 관측소와 기상관측소 거리가 5km 이내인 33개의 종관기상관측소만 이용하여 베이지안 시공간모형을 세웠다.

환경부는 4월 15일부터 10월 15일까지 오존농도가 높은 184일간 오존예보를 시행하고 있다. 따라서 본 연구의 총 관측치는  $3\text{년} \times 184\text{일} \times 33\text{개소} = 18,216\text{개}$ 이다. 전체 데이터에서 결측값 비율은 5.75%이고 총 1,047개이다.

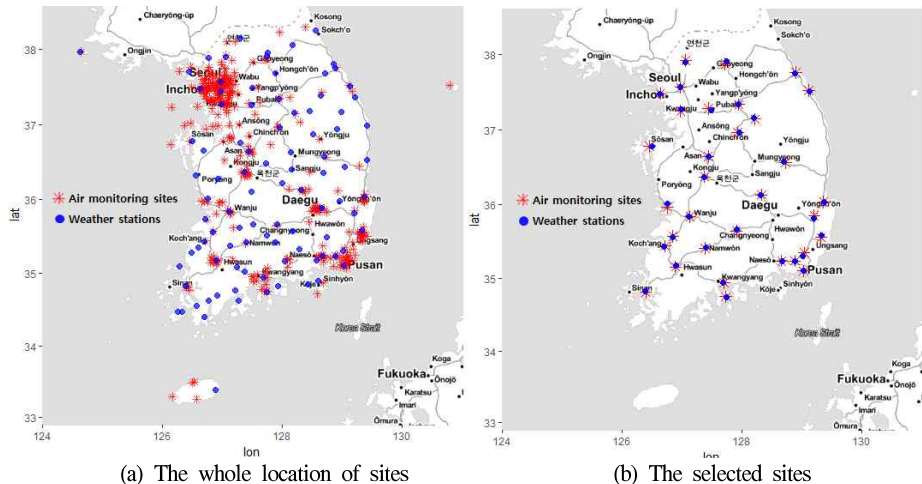


Figure 1. The location of weather stations and air monitoring sites

환경부는 일 최대 8시간 오존농도를 이용하여 오존예보를 시행하고 있다. 한반도 오존농도의 8시간 이동평균에 대한 boxplot은 Figure 2이다. 분석결과 11시에서 19시까지 오존평균농도가 0.038ppm으로 가장 높았다. 따라서 본 연구에서는 일 평균오존농도가 아닌 일 최대 8시간 오존농도(11시~19시)의 평균을 사용하였다.

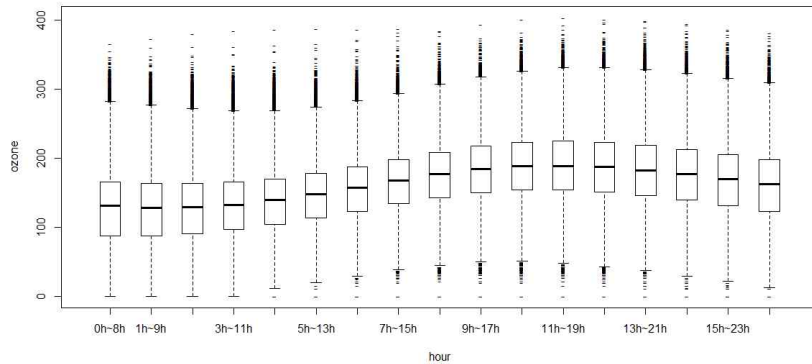


Figure 2. The moving average of 8 hour ozone concentration level

본 연구에서는 가우시안 확률과정을 기반으로 베이지안 시공간모형을 세우므로 오존농도의 분포는 정규분포와 비슷하게 대칭이면 좋다(Sahu, Baker, 2012). Figure 3은 수집된 오존농도 분포의 히스토그램이다. 제곱근으로 변수변환을 시켰을 때, 오존농도자료가 정규분포에 근사함을 Q-Q 그림으로 확인할 수 있다.

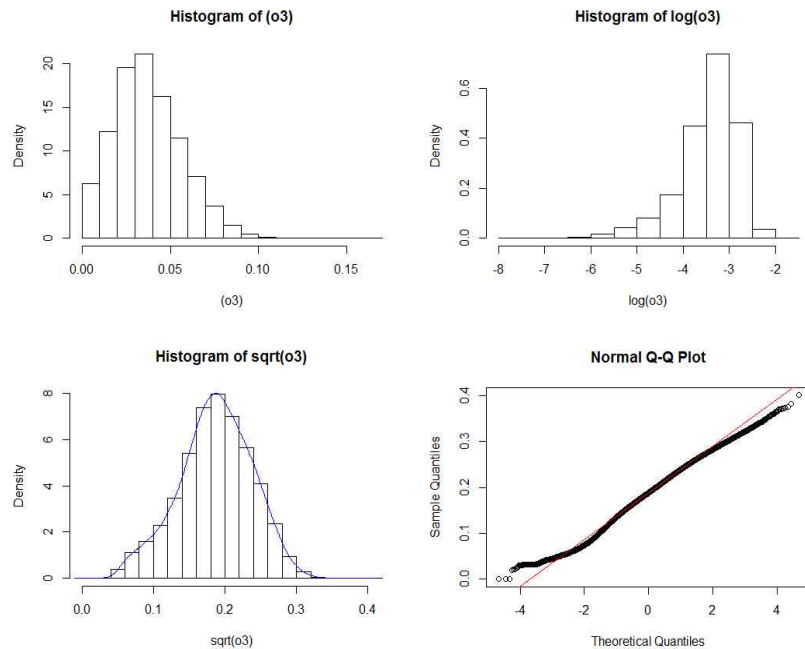


Figure 3. The distribution of transformed ozone concentration level

## 4. 연구결과

### 4.1. 베이지안 시공간 모형의 트레이스 그림

본 연구에서는 13,000번 반복 시행하여 3,000개를 버린 후 10,000개의 사후분포를 이용하여 모수를 추정하였다. 사전분포에 따른 사후분포의 trace plot은 Figure 4이다. GP모형과 AR모형의 시공간 랜덤효과  $\sigma_\eta^2$ 와 공간모수  $\phi$ 에 대한 트레이스 그림을 살펴보면  $\phi$ 에 따라  $\sigma_\eta^2$ 가 변하고 있으며  $\phi$ 가 잘 수렴되지 않았다. 이 외의 트레이스 그림은 잘 수렴되어 본 연구에서는 결과를 생략했다.

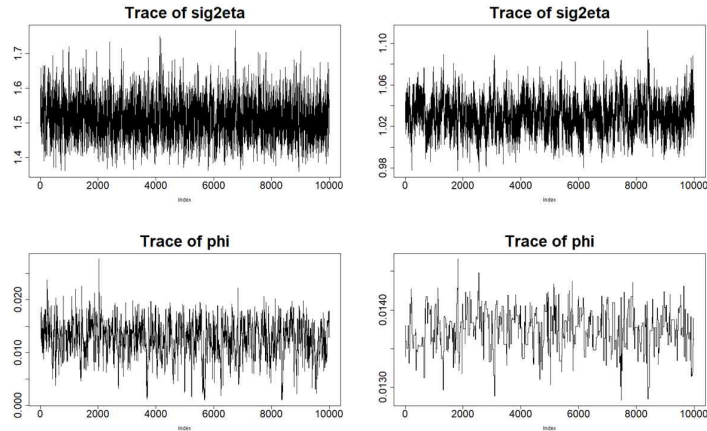


Figure 4. The trace plot of selected parameters in GPmodel(left) and ARmodel(right)

### 4.2. 최적공간모수 $\phi$ 추정

최적공간모수  $\phi$ 를 찾기 위해  $\phi$ 의 값을 0.01씩 옮겨가며 오존농도 예측치와 관측치의 RMSE가 최소가 되는  $\phi$ 를 찾았다.  $\phi$ 에 따른 RMSE는 Figure 5와 같이 아래로 볼록한 그림이므로 수치적 최적화를 통해 최적  $\phi$ 를 구할 수 있다. 분석결과 RMSE가 최소가 되는 공간모수  $\phi$ 는 GP모형과 AR모형 모두 0.088이었다.

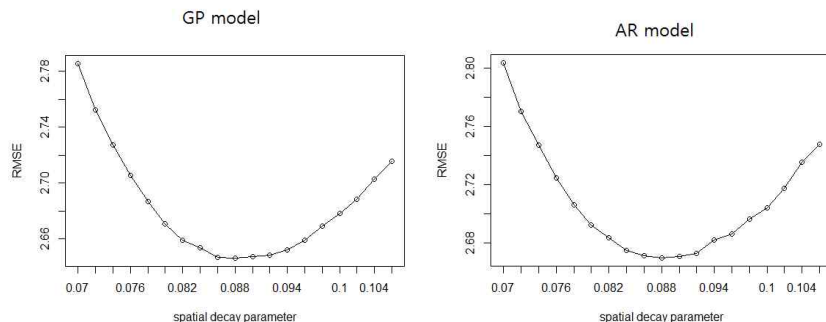


Figure 5. The RMSE of spatial decay parameter  $\phi$  (left : GPmodel, right : ARmodel)

#### 4.3. 최적공간모수 고정 후 트레이스 그림

Figure 6은 최적공간모수  $\phi$ 를 0.088로 고정했을 때 GP모형과 AR모형의 트레이스 그림이다. 모든 모수들이 잘 수렴하였다. 깃스샘플링에서 각 모수들의 수렴도를 알아보기 위해 Geweke 진단법을 이용하였다(Geweke, 1991). Geweke 진단법은 처음 10%의 평균과 마지막 50% 평균이 다른지를 t-test로 판단하여 모수의 수렴 여부를 확인하는 방법이다. GP모형과 AR모형 모두 모든 모수들의 Z 값이 -2에서 2사이에 분포하고 있어 유의수준 0.05에서 수렴하였다.

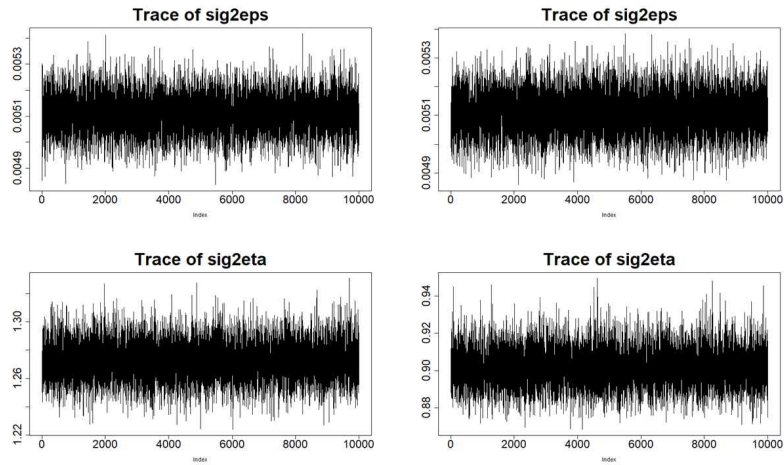


Figure 6. The trace plot of selected parameters in GP model (left) and AR model (right) after  $\phi$  fixed

#### 4.4. 모수추정결과

한반도 오존농도에 대한 GP모형과 AR모형의 추정결과는 Table 1이다. 모수에 대한 사후분포의 평균, 중앙값, 표준편차, 95% 신용구간이 제시되었다. 최고기온과 일조량은 오존농도와 양의 상관관계를 보였고, 바람세기와 상대습도는 음의 상관관계를 보였다. 순수오차분산인  $\sigma_\epsilon^2$ 보다 시공간분산인  $\sigma_\eta^2$ 가 더 크므로 시공간 모형이 한반도 오존농도를 잘 설명하였다고 할 수 있다.

Table 1. The posterior probability estimates of parameters

	GP model				AR model			
	mean	median	sd	95% CI	mean	median	sd	95% CI
(Intercept)	7.312	7.312	0.085	(7.149 7.480)	3.104	3.103	0.086	(2.935 3.274)
max.temp	0.053	0.053	0.002	(0.049 0.057)	0.045	0.045	0.002	(0.042 0.049)
wind	-0.106	-0.106	0.009	(-0.124 -0.089)	-0.024	-0.024	0.008	(-0.039 -0.010)
humid	-0.032	-0.032	0.001	(-0.034 -0.031)	-0.019	-0.019	0.001	(-0.020 -0.018)
sun	0.044	0.044	0.002	(0.041 0.047)	0.041	0.041	0.001	(0.038 0.043)
$\sigma_\epsilon^2$	0.005	0.005	0.000	(0.005 0.005)	0.005	0.005	0.000	(0.005 0.005)
$\sigma_\eta^2$	1.279	1.279	0.014	(1.252 1.308)	0.906	0.906	0.010	(0.887 0.927)
$\rho$					0.488	0.488	0.006	(0.477 0.499)



AR모형의  $\rho = 0.488$ 이고 95% 신용구간에 0이 포함되어 있지 않으므로 오존농도의 자기상관성이 확인된다. 하지만 GP모형의 시공간 분산이 AR모형의 시공간 분산보다 크므로 GP모형이 오존농도의 시공간성을 상대적으로 잘 설명하고 있다. 추가로 AR모형에서 초기조건인  $t$ 가 0일 때의  $O_0$ 의  $\mu_0$ 과  $\sigma_0^2$ 은 Table 2이다.

Table 2. The posterior probability estimates of initial parameters in ARmodel

		mean	median	sd	95% CI	
$\mu_0$	2014	8.268	8.268	0.815	(7.134	8.913)
	2015	8.312	8.312	0.797	(7.391	8.945)
	2016	8.314	8.314	0.841	(7.389	8.961)
$\sigma_0^2$	2014	0.059	0.059	0.015	(0.048	0.068)
	2015	0.059	0.059	0.015	(0.049	0.067)
	2016	0.059	0.059	0.015	(0.049	0.067)

#### 4.5. 예측성능비교

베이지안 시공간 모형의 예측성능을 비교하기 위해 연구에 사용된 33개 관측소 중에서 임의로 27개 관측소만 랜덤하게 선택하여 훈련자료로 사용하였고 나머지 6개의 관측소를 검증자료로 사용하였다. 관측소의 선정에 따라 예측성능에 차이를 보일 수 있으므로 이를 100회 반복하였다. Table 3은 GP모형과 AR모형의 예측성능을 비교한 표이다. GP모형의 RMSE, MAE, MAPE가 AR모형에 비해 상대적으로 낮고 산포가 작으므로 GP모형이 AR모형에 비해 한반도 오존농도의 시공간성을 잘 예측하고 있다.

Table 3. The comparison of prediction performance of GPmodel and ARmodel

	RMSE		MAE		MAPE		BIAS	
	mean	sd	mean	sd	mean	sd	mean	sd
GP model	15.051	0.989	11.962	0.768	32.260	2.949	-0.447	2.781
AR model	15.500	1.092	12.225	0.877	32.893	3.301	-0.145	3.210

#### 4.6. 한반도 오존 농도의 시공간성 예측

Figure 7은 97개소 기상관측소의 최대온도, 상대습도, 바람세기, 일조량을 이용한 한반도 오존 농도의 시공간예측 결과이다. 기상관측소는 한반도 전역에 설치되어 있으므로 울릉도, 제주도, 백령도 등 대기오염 관측소가 운영되지 않는 지역까지 기상요소를 이용하여 오존 농도를 예측할 수 있다. 관측소가 없는 나머지 지역은 역거리가중법(inverse distance weighted method; IDW)을 이용하여 보간 하고 시각화하였다. 역거리가중법은 관측지점과 예측지점 사이에 거리의 역수를 가중치로 하는 가중평균 방법이다(Cho, Ji, Yoon, 2018). 2014년의 일 한반도 오존농도는 충주와 원주 그리고 연천지역이 위치한 한반도 북쪽이 높았으나 시간이 지날수록 대구, 익산, 고창 남원이 있는 남쪽이 높아지고 있다.

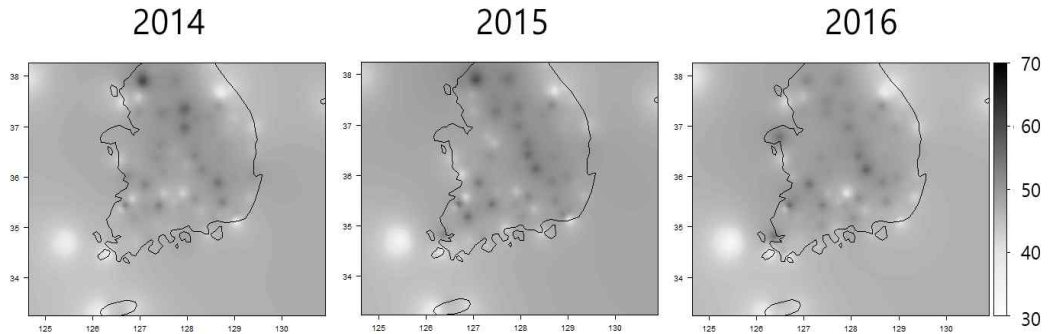


Figure 7. The spatial distribution of annual 8 hour ozone concentration level from 2014 to 2016

Figure 8은 월별 변동성을 살펴보기 위해 4월부터 8월까지의 오존농도를 시각화한 그림이다. 베이지안 시공간모형에서 온도 및 일조량과 오존 농도 사이에 양의 상관성이 있어 봄보다 여름이 오존위험이 높을 것이라 판단되었지만, 분석결과 봄철인 4월부터 6월까지의 한반도 오존농도가 높으며 특히 5월의 오존농도는 위험한 수준이다. 한반도의 복잡한 지형이 풍속의 변화와 잦은 강수를 유발하여 광범위하고 장기적인 오존 고농도현상을 막기 때문이다. Oh, Kim, Hwang(2005)도 여름철 강수량이 오존 농도를 낮추므로 상대적으로 봄과 가을에 고농도의 오존 현상이 발생한다고 하였다. 추가적으로 5월은 편서풍의 영향으로 중국의 남동지역에서 미세먼지가 한반도로 불어와 오존 농도 증가에 영향을 미치는 것으로 판단된다.

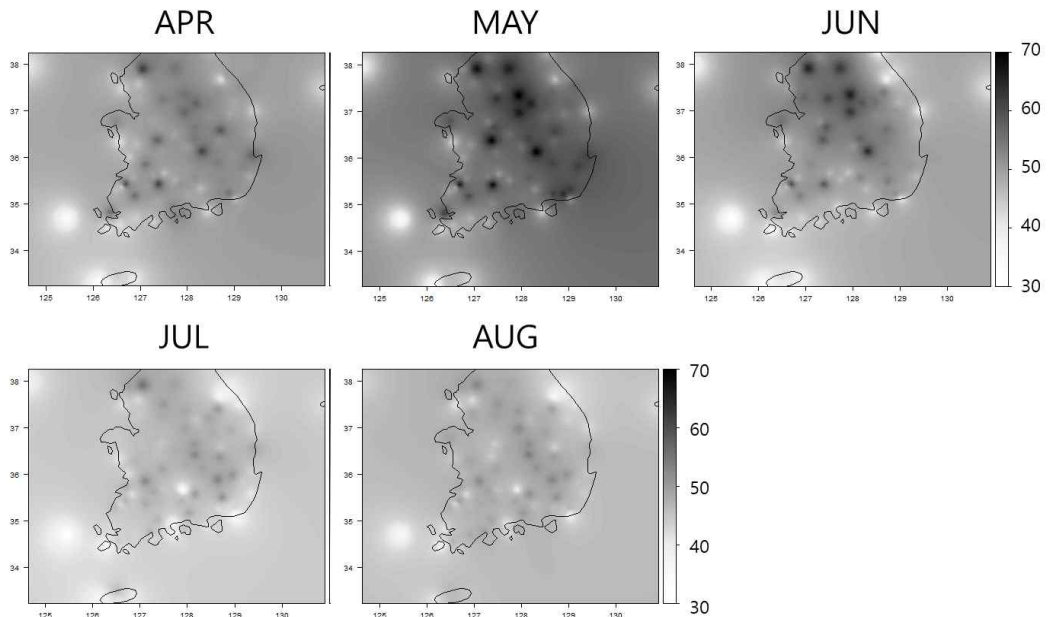


Figure 8. The monthly spatial distribution of annual 8 hour ozone concentration

## 5. 결론

오존으로 인한 환경오염 문제로 인해 중앙정부 및 지방자치단체는 이에 대한 대책을 수립하여 시행하고 있다. 오존 관련 정책을 효율적으로 세우기 위해선 한반도 오존 농도의 시공간적 특징에 대한 분석이 필요하다. 본 연구는 베이지안 시공간모형을 이용하여 한반도 오존 농도의 시공간적 특징을 분석하였다. 베이지안 시공간모형으로 GP모형과 AR모형을 고려하였다. 분석결과 AR모형에 비해 GP모형이 한반도 오존농도의 시공간성을 잘 반영하고 있어 GP모형을 이용하여 연도별, 월별 오존농도의 공간적 분포를 살펴보았다. 그 결과 한반도의 오존농도 분포를 쉽게 살펴 볼 수 있었다. 그러나 오존의 공간적 분포는 위도와 경도에 따라 일정하게 움직이는 것은 아니다. 따라서 33개 자동기상관측망의 위도와 경도 정보를 이용한 97개 자동기상관측망의 예측정보를 일반화하기 위해 해당지역의 지형과 지리를 한번 더 고려해 볼 필요가 있다. GP모형과 AR 모형 외에 베이지안 시공간모형으로 임의의 매듭(knots)을 베이지안 크리깅으로 예측하여 AR모형을 개선시킨 Gaussian predictive processes모형도 있다. 이 모형의 경우  $\phi$ 의 수렴과 knots의 설정 등에 어려움이 있어 논문에 반영하지 못했지만, 후속연구를 통해 GP모형 및 AR과 비교분석 된다면 한반도 오존 농도의 시공간성을 파악하는데 많은 도움이 될 것으로 판단된다.

## References

- Bakar, K. S., Sahu, S. K. (2015). spTimer: Spatio-temporal bayesian modelling using R, *Journal of Statistical Software*, 63(15), 1-32.
- Cho, J., Ji, G., Yoon, S. (2018). The visualization for real-time road weather risk index, *Journal of the Korean Data Analysis Society*, 20(5), 2355-2365. (in Korean).
- Choi, J.-I., Park H.-S., Yang, H.-S. (2017). Estimation of health benefits by reducing ozone concentration through VOCs reduction, *Journal of the Korean Data Analysis Society*, 19(2), 755-767. (in Korean).
- Cressie, N., Biele, J. (1994). A sample-size-optimal bayesian procedure for sequential pharmaceutical trials, *Biometrics*, 700-711.
- Gelfand, A. E. (2012). Hierarchical modeling for spatial data problems. *Spatial Statistics*, 1, 30-39.
- Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *Minneapolis, MN: Federal Reserve Bank of Minneapolis, Research Department*, 196.
- Jang, J. H., Lee, H. W., Lee, S. H. (2012). Spatial and temporal features of PM10 evolution cycle in the Korean peninsula, *Journal of Korean Society for Atmospheric Environment*, 21(2), 189-202.
- Jeon, E. C., Woo, J. H. (1999). Statistical analysis of the meteorological elements for ozone and development of the simplified model for ozone concentration, *Journal of Korean Society for Atmospheric Environment*, 15(3), 257-266.
- Kim, K. W., Kim, J. W., Choi, B. S. (2016). A model construction for probability of human-caused wildfire using geographical information, *Journal of the Korean Data Analysis Society*, 18(6), 2987-2999. (in Korean).
- Moon, S. H. (2009). An Exploratory Analysis of the Korean Ozone Data Using Spatial Principal Component, *Journal of the Korean Data Analysis Society*, 11(1), 371-379.
- National Research Council. (1992). *Rethinking the ozone problem in urban and regional air pollution*, National Academies Press.
- Oh, I. B., Kim, Y. K. (2002). Surface ozone in the major cities of Korea: Trend, diurnal and seasonal variations, and horizontal distributions, *Journal of Korean Society for Atmospheric Environment*, 18(4), 253-264. (in Korean).

- Oh, I.-B., Kim, Y.-K., Hwang, M.-K. (2005). Ozone Pollution patterns and the relation to meteorological conditions in the greater Seoul area, *Journal of Korean Society for Atmospheric Environment*, 21(3), 357-365.
- Oh, S., Kim, M. S. (2019). Predicting onion production by weather and spatial time series model, *Journal of the Korean Data Analysis Society*, 19(5), 2447-2456. (in Korean).
- Pi, S., Hwang, J. (2018). Comparison of bayesian model selection criteria in Fay-Herriot model, *Journal of the Korean Data Analysis Society*, 20(5), 2333-2342. (in Korean).
- Sahu, S. K., Bakar, K. S. (2012). Hierarchical bayesian autoregressive models for large space - time data with applications to ozone concentration modelling, *Applied Stochastic Models in Business and Industry*, 28(5), 395-415.
- Sahu, S. K., Mardia, K. V. (2005). Recent trends in modeling spatio-temporal data, *Proceedings of the special meeting on Statistics and Environment*.
- Tilton, B. E. (1989). Health effects of tropospheric ozone, *Environmental Science & Technology*, 23(3), 257-263.
- Vukovich, F. M. (1995). Regional-scale boundary layer ozone variations in the eastern United States and their association with meteorological variations, *Atmospheric Environment*, 29(17), 2259-2273.
- Yoon, S., Kim, M. (2016). Spatio-temporal models for generating a map of high resolution NO<sub>2</sub> level, *Journal of the Korean Data and Information Science Society*, 27(3), 803-814. (in Korean).

## Bayesian Spatio-Temporal Models for Ozone Concentration Level in South Korea

*Dain Park<sup>1</sup>, Sanghoo Yoon<sup>2</sup>*

### Abstract

Environmental has a property of space-time. Having both the spatial and temporal dimensions makes substantial complexity to environmental data analysis. In this study we built models using daily maximum 8-hour ozone concentration data obtained from 33 sites in South Korea for analysis between 2014 and 2016. Maximum temperature, relative humidity, and wind speed were considered as covariates. The data on the square root scale seems most good in terms of both stabilizing and symmetry the variance. We specified Independent Gaussian process model and Autoregressive model as the probability process of hierarchical Bayesian framework and Markov Chain Monte Carlo techniques. These spatio-temporal models allow spatial prediction of a temporally aggregated ozone summary along with its uncertainty. Although the autoregressive parameter was statistically significant, Independent Gaussian process model well reflected the spatio-temporality of ozone concentration level on the Korean Peninsula.

*Keywords* : Autoregressive model, Bayesian inference, Gaussian process model, Spaital-temporal model.

---

<sup>1</sup>Researcher, Climate analytics department, APEC climate center, 12 centum 7-ro, haeundae-gu Busan 48058, Korea. E-mail : dain105@apcc21.org

<sup>2</sup>(Corresponding Author) Assistant professor, Division of mathematics and big data science, Daegu University, Gyeongbuk, 38453, Korea. E-mail : statstar@daegu.ac.kr

[Received 28 April 2020; Revised 11 June 2020; Accepted 12 June 2020]