



# Project summary

Name: 권나연

Date: 2022. 11. 18

## Title

A Bayesian Localized Conditional Autoregressive Model for Estimating the Health Effects of Air Pollution

## Authors

Duncan Lee, Alastair Rushworth, and Sujit K. Sahu

## Keywords

Air pollution and health; Conditional autoregressive models; Spatial autocorrelation.

## Research Goal

- Spatial small-area의 disease incidence data를 이용하여 대기오염이 장기적으로 건강에 미치는 효과를 추정하는 것

## Problem Definition

- 기존의 conditional autoregressive model (CAR)을 이용하여 random effect를 모델링할 경우 CAR prior와 다른 covariate (e.g., 대기 오염) 간의 collinearity가 존재하며, 이는 각 covariate의 효과에 대한 estimation의 성능을 저하시킴

- 특히, intrinsic CAR 모델은 random effect가 지리적 인접에 의해 결정되는 단일 전체 수준의 공간적인 smoothness를 보이도록 강제하며, 이는 residual spatial autocorrelation에 존재하는 복잡하고 국소적인 (localized) 구조를 포착하기에 유연하지 못한 모델임
  - Residual spatial autocorrelation는 특정 지역에서 매우 강하게 (smoothness), 또는 약하게 (급격한 step change) 나타날 수 있음
- 이에 대한 해결방안으로 localized conditional autoregressive (LCAR) 모델을 고안함

Notes: 여기서 step change란 지리적으로 인접한 지역이 매우 다른 값을 가지는 것을 의미함

## LCAR prior

- LCAR prior는 spatial smoothness 또는 뚜렷한 step change를 포착할 수 있을 정도로 유연한 residual spatial autocorrelation을 모델링하기 위해 고안된 prior
- 이는 neighborhood matrix  $\mathbf{W}$ 를 random quantity로 여김
  - 많은 수의 partial correlation parameter 추정을 필요로 하며, 이 문제는 prior 정보를 사용함으로써 극복할 수 있음
  - Prior 정보의 사용이  $\mathbf{W}$ 에 대한 표본 공간의 크기와 차원을 크게 줄여줌
- 이 모형을 통해 residual autocorrelation structure와 fixed effect 간의 collinearity 문제 해결 가능 ← Covariate effect가 제거된 후에 prior 정보로부터 이를 알아내기 때문
- LCAR prior는 확장된 random effect의 집합  $\tilde{\phi}$  와 edge의 집합  $W$ 에 대한 joint prior distribution으로 구성됨. 즉, 이는 다음과 같이 분해될 수 있음

$$f(\tilde{\phi}, W) = f(\tilde{\phi}|W)f(W)$$

- LCAR prior는  $\tilde{\phi}$  에 대한 수정된 CAR prior와  $W$ 에 대한 discrete uniform distribution, 그리고 Poisson likelihood를 결합하여 full Bayesian hierarchical model을 구성함

## Results

### 1 Simulation study

#### 데이터 생성 및 연구 설계

- Motivating study의 지역을 포함하기 위하여 271개의 행정구역 (IG)에 대한 시뮬레이션 데이터를 생성함
- LCAR 모델과 BYM (Besag et al., 1991), LM (Lee and Mitchell, 2013), 그리고 HH (Hughes and Haran, 2013) 모델의 성능을 비교함
- $M = 0.5, 1, 1.5$  와  $E_k \in [10, 25], [50, 100], [150, 200]$ 의 조합으로 구성된 9개의 시나리오를 비교
  - $M$ : 값이 클수록 더 국소적인 구조를 띠 (the extent of local rather than global residual autocorrelation)
  - $E$ 의 길이: 유병률을 정량화

## 모델 평가 기준

- 추정된 회귀 모수  $\beta$ 에 대한 RMSE와 95% credible interval
- 추정된 회귀 모수  $\beta$ 의 95% credible interval에 대한 percentage coverage와 평균 길이

## 결과

- LCAR 모델이 covariate effect 추정에 있어서 가장 좋은 성능을 보임
- LCAR 모델의 RMSE가 가장 작았으며, 대부분의 시나리오에서 coverage가 90% 이상이며, BYM 모델보다 더 좁은 credible interval을 가짐
- Residual spatial autocorrelation에 대한 부적절한 통제는 fixed effect 추정을 방해하므로, 신중한 모델링이 매우 중요

## 2 Motivating study

### 연구 설계

- 연구 지역: 스코틀랜드의 항구도시인 Glasgow와 Clyde 강 어귀
  - 2011년 기준 거주 인구 수는 120만명 이하 → spatial small-area에 해당
  - 지역은 intermediate geographies (IG)라고 불리는 행정구역으로 구분되며 ( $n = 271$ ), 각 행정구역은 평균적으로 4,000명의 인구를 포함함
- 반응변수: 2011년 각 IG 내 주요 호흡기질환 진단을 위해 병원 (정신건강의학과, 산부인과 제외) 방문 횟수

- Covariates:
  - 대기오염 관련: 일산화탄소 (CO) 농도, 이산화질소 (NO<sub>2</sub>) 농도, 이산화황 (SO<sub>2</sub>) 농도, 미세먼지 (PM10) 농도, 초미세먼지 (PM2.5) 농도
  - 그 외: 사회/경제적 빈곤 측정, 실직 상태이며 2010년 구직 상황에서 재정적 지원을 받는 인구 수 (JSA), 인종, health care 접근성

Notes: 대기오염에 대한 노출은 반응변수 이전에 일어나야하므로, 2010년의 대기오염 농도를 사용

## 모델 평가 기준

- DIC와 Moran's I
- 추정된 covariate effect를 비교

## 결과

- Greater Glasgow 내의 미세먼지와 이산화질소는 호흡기 건강에 악영향을 미침
- 연간 평균 농도가 1 표준 편차만큼 증가할 경우 시민의 질병에 대한 부담이 약 4% 증가할 것으로 추정됨
- 하지만 이는 observational study의 결과이므로, 개인 수준에서의 원인과 결과로 해석될 수 없음

## Conclusions

- LCAR 모형을 통해 기존의 모형들이 축소해왔던 미세먼지와 이산화질소의 건강에 대한 영향이 상당히 크다는 것을 밝혀냄

## References

Lee, D., Rushworth, A., & Sahu, S. K. (2014). A Bayesian localized conditional autoregressive model for estimating the health effects of air pollution. *Biometrics*, 70(2), 419-429.