Joint Spatial Mapping of Multiple Crime Rates Using Multivariate CAR Model Approach

Rasheed A. Adeyemi, Davies Obaromi & James Mayaki

Biostatistics & Epidemiology Research Team (BERT)
Department of Statistics
Federal University of Technology, Minna

e-mail: rashid.adeyemi@futminna.edu.ng
OR adyras001@myuct.ac.za

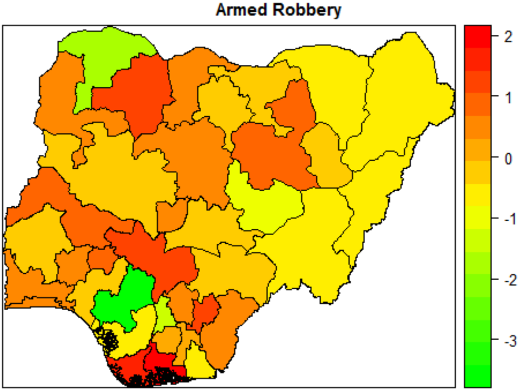
Paper presented at Visual Meeting of 4^{th} Annual Conference of Professional Statisticians Society Nigeria (PSSN), ILORIN 2020

Disease Mapping

- Ecological studies of crime are of great interest to geographers and criminologists and they are used to reveal the geographic pattern of crime risks as well as the relevant risk factors explaining that pattern.
- Crimes are rarely considered a public health problem or investigated using epidemiological methods.
- Broadly speaking, the ecological or neighborhood determinants of health and crime are themselves one in the same, or at least correlated[2, 3]
- Disease mapping:
  - to describe geographical variation of disease
  - to generate hypothesis about the possible causes of differences in risk of disease
- Related Databases
  - National Bureau of Statistics (NBS)
  - Nigeria Demographics and Health Survey (DHS)
  - Surveillance, Epidemiology, and End Results (SEER)
  - Mapping Malaria Risk in Africa (MARA/ARMA collaboration),

Background: single crime mapping

Mapping of raw armed robbery incidence 2017 across 36 states and FCT - Abuja in Nigeria



Background: modeling of a single crime

For rare case, Poisson regression model:

$$Y_i|\mu_i \sim Poisson(E_i \exp(\mu_i)) \quad i = 1, \dots, n,$$

where $\mu_i = x_i'\beta + \phi_i$. The x_i are explanatory, state(district)-level spatial covariates, having parameter coefficients β.

- $E(Y_i) = E_i \exp(\mu_i) \longrightarrow SMR = \exp(\hat{\mu}_i)$, where $\hat{\mu}_i = x_i'\hat{\beta} + \hat{\phi}_i$.
- μ_i represents the log relative risk of departures of the Y_i from the E_i.
- Hierarchical Bayesian modeling:
    - Using Markov chain Monte Carlos (McMC) methods
    - Fist stage: likelihood of the observation data
    - Second stage : prior distribution of the fixed effect and the random effect $\phi = (\phi_1, \phi_2, \dots, \phi_n)'$.

Background: modeling of a single crime cont. . .

- Markov random field (MRF): the conditional distribution of a state's response given the responses of all the other states depends only on the observations in the neighborhood of this site.
- Prob(A site's response | All other sites)=Prob(A site's response | its neighbors
- Mathematically, the Conditionally autoregressive (CAR) prior on $\phi = (\phi_1, \phi_2, \ldots, \phi_n)'$ is given as

$$\phi_i|\phi_j, i \neq j, \sim N\left(\frac{\alpha}{m_i}\sum_{j\sim i}\phi_i, \frac{1}{\tau m_i}\right), \quad i, j = 1, \ldots, n,$$

- where m_i is the number of neighbours of area i and α is smoothing parameter

Background: modeling of a single crime cont. . .

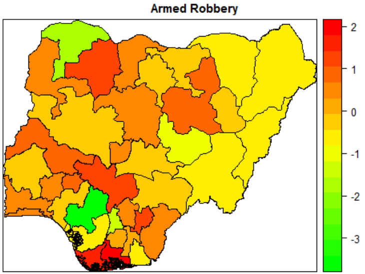
- The following from the equation above, it implies that

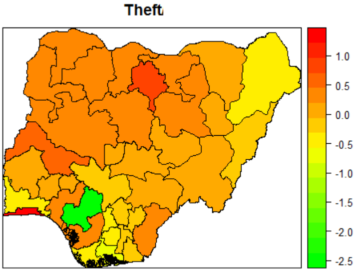
$$\longleftrightarrow \quad \phi \sim \mathsf{N}_n \left(0,\ [\tau(D - \alpha W)]^{-1}\right)$$

where $D = Diag(m_i)$, and W is adjacency matrix of the map i.e. $w_{ii} = 0$ and $w_{ii'} = 1$ if i' is adjacent to i and $\mathbf{0}$ otherwise.

Motivation: mapping of multiple crimes

Mapping of reported armed robbery and theft (stealing) incidence 2017 across 36 states and FCT - Abuja in Nigeria





Motivation: spatial modeling of multiple crimes

- For rare cases, poisson regression model:

$$Y_{ij} \sim Poisson(E \cdot e^{x'_{ij}\beta_j + \phi_{ij}}) \quad i = 1, \ldots, n, \quad j = 1, \ldots, p \tag{1}$$

  where the x_{ij} are explanatory, region(state)-level spatial covariates for crime j having parameter coefficient β_j.
- Correlations in multiple crime data:
  - Spatial correlation for each disease across regions
  - Dependence among multiple crimes within the same region
  - Cross-spatial correlation among multiple crime rates in different regions

Multivariate modeling cont. . .

- In multivariate setting, where $\boldsymbol{\phi} = (\phi_1, \phi_2)'$ is modeled using a multivariate conditional autoregressive prior
- that is $\mathbf{\Phi} \sim \text{MCAR}(1, \Sigma)$, and where Σ is the covariance matrix including correlation.
- where $\beta_{j0}, j = 1, 2$ in equation (1) represents individual specific crime intercept,
- given $\boldsymbol{\phi_i}$ and $\boldsymbol{\phi_i} = (\phi_1, \phi_2)'$ is a 2×1 vector of spatial dependent random effects for the i_{th} region (state)

Model Specifications

Univariate Prior Specification (CAR)

- Consider a vector $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_n)'$ of p components, which follow a multivariate Gaussian distribution with mean zero and variance -covariance matrix $\boldsymbol{Q}^{-1}$,
- the joint pdf of $\boldsymbol{\phi}$ is given by

$$p(\phi) = (2\pi)^{\frac{p}{2}} |\boldsymbol{Q}|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\boldsymbol{\phi}^T \boldsymbol{Q}\boldsymbol{\phi}\right\}$$

where $\boldsymbol{Q}$ is $p \times p$ symmetric and positive definite matrix.

Multivariate Prior Specification (MCAR)

- The development of MCAR model is credited to Mardia1(988) as an extension of Besag (1974)
- Then Φ is an $np \times 1$ vector having a multivariate Gaussian distribution with mean, $\mathbf{0}$ and precision matrix $\mathbf{Q}$, mathematically expressed as

$$p(\boldsymbol{\Phi}) = (2\pi)^{\frac{np}{2}} |\boldsymbol{Q}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\Phi}^T \mathrm{Q} \boldsymbol{\Phi}\right\} \quad (2)$$

Statistical inference

- In a full Bayesian framework, appropriate prior distributions are assigned to all model parameters .
- non-information priors were assigned to the regression coefficients.
- For the each intercept, diffuse priors were assumed, that is, $p(\alpha_k)$
- For the regression coefficients, highly dispersed normal distribution priors are assumed, that is, $p(\beta) \sim \mathsf{N}(0, 10^4)$.
- an inverse Wishart prior is assumed for $\Sigma \sim IW(r, R)$ with R considered to be an identity matrix.
- All model were fitted using WinBUGS software

Posterior Estimates of risk factors of crimes

Table 1: Posterior Estimates of risk factors of covariates and model fit parameters

Parameters	Theft	Armed robbery
Fixed effects	Post. mean (95% CI)	Post. mean (95% CI)
β_0	-0.346 (-0.597 , -0.116)	0.134 (-0.205 0.453)
β_1	-0.352 (-0.753 , 0.034)	-0.114 (-0.689 0.580)
β_2	-0.311 (-0.636 , -0.028)	0.095 (-0.384 0.515)
β_3	0.334 (-0.131 , 0.859)	-0.596 (-1.377 0.254)
β_4	-0.292 (-0.726 , 0.043)	-0.173 (-0.803 0.338)
β_5	0.191 (-0.210 , 0.538)	0.227 (-0.236 0.898)
β_6	0.122 (-0.399 , 0.596)	0.101 (-0.529 0.646)
Random effects		
σ_u^2	0.395 (0.089 , 0.990)	0.679 (0.099 1.950)
σ_v^2	0.829 (0.601 , 1.120)	1.060 (0.686 1.459)
ρ_{12}	0.4654 (-0.224, 0.8785)	
Model fit		
$\bar{D}$	323	256.2
pD	-3119	-645.4
DIC	-2796	-389.2

$\beta_0 =$ overall base risk (intercept), $\beta_1 =$ number of divisional police HQ, $\beta_2 =$ unemployment rate, $\beta_3 =$ population density, $\beta_4 =$ education Index $\beta_5 =$ gross national income (GNI)$\beta_6 =$proportion young adult male per state (age 18-35)

Spatial correlation and conditional variances

- The marginal conditional variances in the geographical prevalence of the crime rates are : armed robbery : σ_u^2: 0.395 95%CI(0.089, 0.990) and stealing 0.679 95%CI(0.099 1.950).
- There is weak positive correlation between the spatial incidence of robbery and stealing : 0.4654 (-0.224, 0.8785).

Predicted maps of multiple crimes

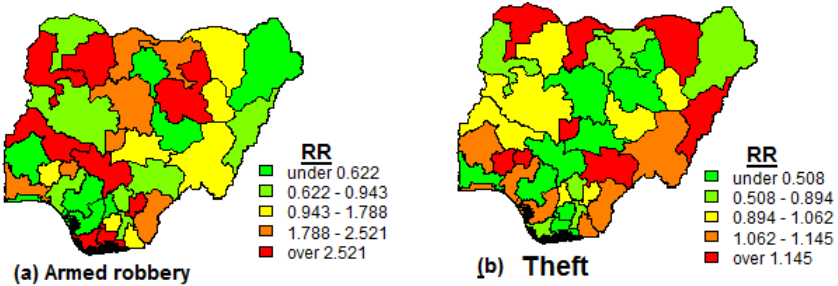


Figure 1: Predicted Risk Surface of crime rates (a) armed robbery (b) theft using convulsion model

Concluding remarks

- The present study expands the methodological strategy by linking the existing criminology literature and spatial modeling approach in a unified manner.
- In contrast to the conventional regression model, the Bayesian spatial model has taken into account neighbourhood effect of crime rates
- Our approach also detected hot spot regions and evaluated the share risk factors of the crime rates.

On Going Research on Multivariate lattice

- Dynamic MCAR models for multivariate spatiotemporal data
- Spatially varying coefficients model
- Spatial factor analysis with p factors
- Linear model of coregionalization (LMC)
- Some other applications of multivariate lattice models

Research Output & Future Direction

- Adeyemi et al(2016) Semi-parametric Mutinomial Ordinal Models to analyze the spatial patterns of child birthweight in Nigeria **published** *Int. J. Environ. Res. Public Health* **2016**, *13, 1145;* doi:10.3390/ijerph13111145
- Adeyemi et al(2016) Bayesian Mutinomial Ordinal Models to analyze the risk factors and spatial patterns of childhood anemia in Tanzania **published** Proceeding of 58^{th} Annual Conference of South African Statistical Association
- Adeyemi et al(2019) Multivariate Spatial Joint Mapping of the risk of Childhood Anemia and Malnutrition in sub-Saharan Africa: A cross-sectional study of small-scale geographical disparities *African Health Sciences*

Acknowledgements

- We acknowledge National Bureau of Statistics for the permission to use the Crime statistics data
- The first author also appreciate the study fellowship/support received from the Federal University of Technology , Minna-NIGERIA for undergoing his postgraduate study in South Africa.

References I

Borrell LN, Roux AVD, et al(2004).
Neighbourhood characteristics and mortality in the atherosclerosis risk in communities study.
Int. J. Epidemiol 33:398–407.

Browning CR, Byron RA, et al (2010).
Commercial density, residential concentration, and crime: land use patterns and violence in neighborhood context.
J Res Crime Deliq. 47:329–57.

Freedman VA, Grafova IB, et al(2011).
Neighborhoods and chronic disease onset in later life.
Am J Public Health. 101:79–86.

Besag, J., York, J., & Molli, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1), 1-20.

References II

Knorr-Held, L., & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. Journal of the Royal Statistical Society: Series A (Statistics in Society), 164(1), 73-85..

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.

Assunção, R. M. and Castro, M. S. (2004).
Multiple cancer sites incidence rates estimation using a multivariate bayesian model.
International journal of epidemiology, 33(3):508–516.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014).
Hierarchical modeling and analysis for spatial data.
Crc Press.

References III

Besag, J., York, J., and Mollié, A. (1991).
Bayesian image restoration, with two applications in spatial statistics.
Annals of the institute of statistical mathematics, 43(1):1–20.

Best, N. and Hansell, A. L. (2009).
Geographic variations in risk: adjusting for unmeasured confounders through joint modeling of multiple diseases.
Epidemiology (Cambridge, Mass.), 20(3):400.

Congdon, P. (2017).
Representing spatial dependence and spatial discontinuity in ecological epidemiology: a scale mixture approach.
Stochastic Environmental Research and Risk Assessment, 31(2):291–304.

Dabney, A. R. and Wakefield, J. C. (2005).
Issues in the mapping of two diseases.
Statistical methods in medical research, 14(1):83–112.

References IV

Dobra, A., Lenkoski, A., and Rodriguez, A. (2011).
Bayesian inference for general gaussian graphical models with application to multivariate lattice data.
Journal of the American Statistical Association, 106(496):1418–1433.

Downing, A., Forman, D., Gilthorpe, M. S., Edwards, K. L., and Manda, S. O. (2008).
Joint disease mapping using six cancers in the yorkshire region of england.
International Journal of Health Geographics, 7(1):41.

Held, L., Natário, I., Fenton, S. E., Rue, H., and Becker, N. (2005).
Towards joint disease mapping.
Statistical methods in medical research, 14(1):61–82.

References V

Ibáñez-Beroiz, B., Librero-López, J., Peiró-Moreno, S., and Bernal-Delgado, E. (2011).
Shared component modelling as an alternative to assess geographical variations in medical practice: gender inequalities in hospital admissions for chronic diseases.
BMC medical research methodology, 11(1):172.

Kinyoki, D. K., Manda, S. O., Moloney, G. M., Odundo, E. O., Berkley, J. A., Noor, A. M., and Kandala, N.-B. (2017).
Modelling the ecological comorbidity of acute respiratory infection, diarrhoea and stunting among children under the age of 5 years in somalia.
International Statistical Review, 85(1):164–176.

Knorr-Held, L. and Best, N. G. (2001).
A shared component model for detecting joint and selective clustering of two diseases.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 164(1):73–85.

References VI

MacNab, Y. C. (2011).
On gaussian markov random fields and bayesian disease mapping.
Statistical Methods in Medical Research, 20(1):49–68.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002).
Bayesian measures of model complexity and fit.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(4):583–639.


Thank you