

PAPER • OPEN ACCESS

Bayesian spatial modeling of poverty risk in Kelantan

To cite this article: S A Nawawi *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **842** 012051

View the [article online](#) for updates and enhancements.

You may also like

- [Application of log-linear models to pneumonia patients: a case study of Jigme Dorji Wangchuck National Referral Hospital \(JDWNRH\) in Bhutan](#)
Sonam Zam and Montip Tiensuwan
- [Robust scaling in fusion science: case study for the L-H power threshold](#)
G. Verdoolaege and J-M. Noterdaeme
- [A nonlinear mixed model approach to predict energy expenditure from heart rate](#)
Lauri Kortelainen, Jouni Helske, Taija Finni et al.

Bayesian spatial modeling of poverty risk in Kelantan

S A Nawawi^{1*}, N M Fauzi¹, A N M Nor¹, N Ibrahim², R M Jamil¹, H A Aziz¹, R Nawawi³, S H Ya'acob¹, N S A Halim¹.

¹Faculty of Earth Science, Universiti Malaysia Kelantan, 17600 Jeli, Malaysia

²Faculty of Bioengineering and Technology, Universiti Malaysia Kelantan, 17600 Jeli, Malaysia

³UiTM Cawangan Kota Bharu, Kota Bharu Kelantan.

*Corresponding author: aisyah.n@umk.edu.my

Abstract. Poverty data vary in rural areas, in certain states or regions, and some urban areas. These areal data tends to have spatial autocorrelation. A Bayesian hierarchical model is commonly used to estimate the risks using a combination of available covariate data and a set of spatial random effects. These random effects are commonly modelled by conditional autoregressive (CAR) prior distributions, a type of Markov random field model. Spatial autocorrelation between the random effects, ϕ in CAR models is induced by a $n \times n$ neighbourhood matrix, \mathbf{W} . However, many studies assumed that the \mathbf{W} is fixed when fitting the model. Therefore, this study evaluates the performance of the Poisson-log linear Leroux Conditional Autoregressive (CAR) model with m -nearest neighbourhood weight matrices using a simulation study. This study creates simulated poverty data for 66 districts of Kelantan with different scenarios that related with random effects and covariate. A Poisson log-linear Leroux CAR model with $m=1, 5$ and 10 nearest neighbours are applied to the simulated poverty data. The performance of the models is evaluated using bias, Root Mean Square of Error (RMSE) and Deviance Information Criterion (DIC). The results show that the choice of $m=1, 5$ and 10 neighbourhood matrices and scenarios do not affect the bias for either the regression parameter β or the risk R_k and RMSE for the risk R_k . Nevertheless, there is the dissimilarity of the performance of the models in the RMSE of regression parameter β . The results suggest that the Poisson log-linear Leroux CAR model with the $m=5$ nearest neighbours performed overall best for simulated poverty data. It consistently gives good results across different strength of spatial autocorrelation of random effects ϕ and covariate. The model also gives the lowest DIC in all the scenarios, indicating a better fitting model than other models. The findings of this study give guidance in choosing the suitable m -nearest neighbourhood matrices to estimate the poverty risk in Kelantan.

1. Introduction

Poverty persists as a global problem that has long plagued humanity regardless of rich or non-developing countries. Hence, broad varying definitions of poverty have emerged. Poverty could be defined as the basis number of households or families having a total income of less than half or two-thirds of the community's average income [1]. Many effects of poverty such as limited or lack of access to education, disease, increased suicide rates, unsafe and degraded environment and social discrimination have been discussed in studies to minimize the effects, particularly on the productive resources to ensure sustainable livelihood [2, 3, 4, 5, 6].



The number of poor households varies from one place to another place. Modelling the data related to the adjacent spatial units is a common problem in several statistical applications. To model these data, one must consider spatial autocorrelation. That is, areas close together have more similar values on average than areas further apart. Most studies overcome this problem by adding a set of autocorrelated random effects ϕ to the linear predictor of the regression model. These random effects are most often modelled by a Conditional Autoregressive (CAR) prior as part of a hierarchical Bayesian model [7, 8, 9]. A $n \times n$ neighbourhood matrix, \mathbf{W} induce spatial autocorrelation in CAR models. However, many studies assumed that the \mathbf{W} is fixed when fitting the model. Therefore, this research will compare the performance of a Poisson log-linear Leroux CAR model with different m -nearest neighbourhood matrix, \mathbf{W} . This study will apply each \mathbf{W} specification to simulated poverty data under various scenarios.

2. Materials and Methods

This section shows the methodology used in this study.

2.1 Distance-based neighbour

The spatial autocorrelation between the random effects ϕ in the Poisson log-linear Leroux CAR model is determined by a binary $n \times n$ neighbourhood matrix, \mathbf{W} . The element w_{kj} represents a measure of closeness between area k and area j for $k=1, \dots, n$ areas. One of the alternatives to defining neighbours is a function of geographical distance. Distance between areas k and j is often measured as the Euclidean distance between their respective centroids. $w_{kj}=1$ if area k is one of the m -nearest neighbours of area j and 0 otherwise. This study will use three different m -nearest neighbours where $m=1, 5$ and 10.

2.2 Poisson log-linear Leroux Conditional Autoregressive (CAR) model

A Bayesian Hierarchical model is adopted to model the data y_k using covariates information $x_k = (x_{1k}, \dots, x_{pk})$ and a random effect ϕ_k . The random effects $\phi = (\phi_1, \dots, \phi_n)$ are included to model any spatial autocorrelation in the data that persist after adjusting for the available covariate information. The random effects are modelled by a CAR prior distribution, a Gaussian Markov Random Field (GMRF) model. The model is determined by a set of n univariate full conditional distributions $f(\phi_k | \phi_{-k})$, where $\phi_{-k} = (\phi_1, \dots, \phi_{k-1}, \phi_{k+1}, \dots, \phi_n)$ for $k = 1, \dots, n$. In this study, the random effects are given the Leroux CAR prior [10].

In this study, the poverty data used are counts data. Thus the Poisson log-linear Leroux CAR model is employed. Therefore, the formulation of the Poisson log-linear Leroux CAR model used in this analysis is shown below:

$$\begin{aligned} Y_k &\sim \text{Poisson}(E_k R_k) \text{ for } k = 1, \dots, n, \\ \ln(R_k) &= x_k^\top \beta + \phi_k, \\ \phi_k | \phi_{-k} &\sim N \left[\frac{\rho \sum_{j=1}^n w_{kj} \phi_j}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{j=1}^n w_{kj} + 1 - \rho} \right]. \end{aligned} \quad (1)$$

In the above equation, R_k is the poverty risk in area k . If $R_k = 1$, then $\mathbb{E}(y_k) = E_k$ is thus the average risk. While if $R_k = 1.2$, then $\mathbb{E}(y_k) = 1.2E_k$ which means 20% more cases than expected. Here ρ is the level of spatial autocorrelation in the random effects, where $\rho = 1$ shows strong spatial autocorrelation between random effect and corresponds to the intrinsic model and $\rho = 0$ corresponds to independence ($\phi_k \sim N(0, \tau^2)$). Finally, τ^2 the conditional variance of $\phi_k | \phi_{-k}$. Inference for this type of model is typically based on Markov Chain Monte-Carlo (MCMC) simulation, using a combination of Gibbs sampling and Metropolis-Hasting steps. The software used for this study is *CARBayes* [11], which is an R package for Bayesian spatial modelling with conditional autoregressive priors.

2.3 Simulation Study

This section provides simulation research that compares the performance of the Poisson log-linear Leroux CAR model with $m=1, 5, 10$ nearest neighbours.

2.3.1. Data generation

Simulated data are generated with irregular lattices comprising a real map of the 66 districts of Kelantan. Simulated poverty counts Y_k are generated from the model (1). For simplicity, a covariate is used and the regression coefficient is fixed at $\beta = 0.10$. The expected numbers of poverty cases, E , are the expected of poverty cases in 66 districts of Kelantan. This study follows data generation from Lee [4] in disease mapping study, which used in their simulation study $n = 271$ intermediate geographies (IG) in the Greater Glasgow health board, the expected numbers of disease cases are the expected of cancer cases for 271 IG, and there are several scenarios regarding the random effects. A covariate (x) and the random effects ϕ are generated from multivariate normal distributions given by $\phi \sim N(0, \tau^2 R)$ and $x \sim N(0, \sigma^2 V)$. Both τ^2 and σ^2 are fixed at 0.10. Two different measures of R are used corresponding to two different structures of the covariate, which are independent and spatially autocorrelated. While two different measures of V are used, they correspond to weakly and strongly spatially autocorrelated random effects ϕ . The poverty data are simulated under each of the following scenarios:

1. Scenario 1: spatial autocorrelation between ϕ are weakly autocorrelated and independent spatial autocorrelation of a covariate at different locations.
2. Scenario 2: spatial autocorrelation between ϕ are strongly autocorrelated and independent spatial autocorrelation of a covariate at different locations.
3. Scenario 3: spatial autocorrelation between ϕ are weakly autocorrelated and strong spatial autocorrelation of a covariate at different locations.
4. Scenario 4: spatial autocorrelation between ϕ are strongly autocorrelated and strong spatial autocorrelation of a covariate at different locations.

In this study, a total number of one hundred simulated data sets are generated under each of the four scenarios. The Poisson log-linear Leroux CAR model with $m=1, 5, 10$ nearest neighbours was applied in each scenario. The bias, root mean square error (RMSE) of the estimated β and poverty risks R_k and Deviance Information Criterion (DIC) are used to evaluate the relative performance of the three models.

3. Results and Discussion

Figures 1 to Figure 3 represent maps of Kelantan and how spatial neighbours are defined among the 66 districts of Kelantan. The maps are produced using maptools in R programming. Every line indicates a connection among nodes. There are apparent differences in the number of neighbours assigned by $m=1, m=5$, and $m=10$ nearest neighbours.

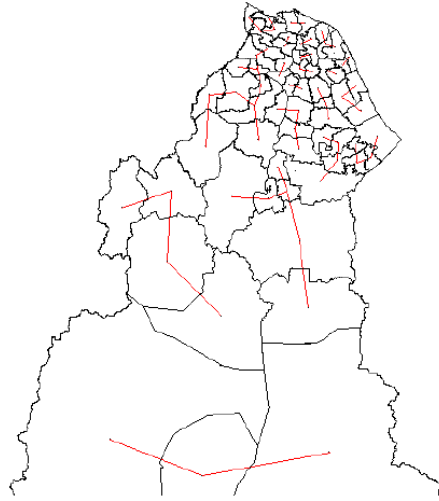


Figure 1. Pictorial representation of the neighbourhood matrices for $m = 1$ -nearest neighbours.

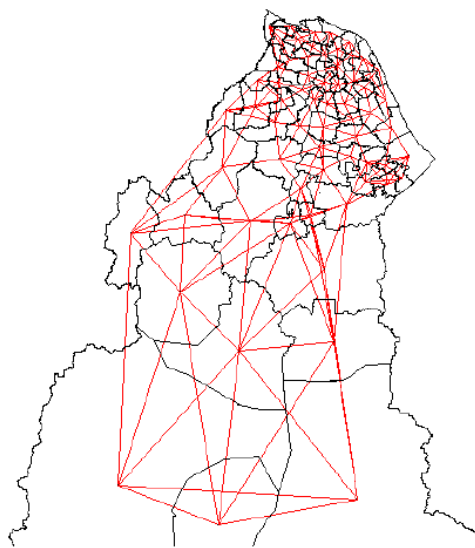


Figure 2. Pictorial representation of the neighbourhood matrices for $m = 5$ -nearest neighbours.

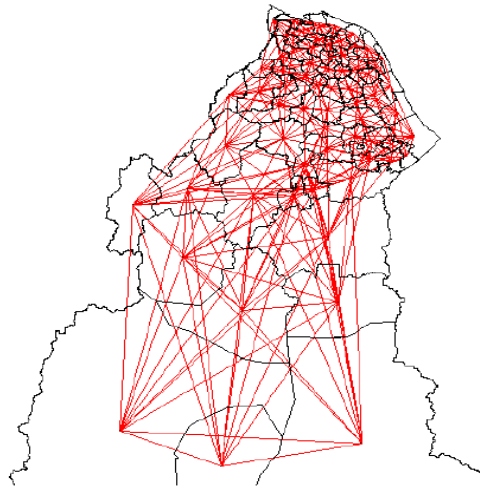


Figure 3. Pictorial representation of the neighbourhood matrices for $m = 10$ -nearest neighbours.

The results of the simulation study are represented in Table 1. The choice of \mathbf{W} matrices and scenarios does not affect the bias for either the regression parameter β or the risk R_k . In all cases, the biases are less than 0.021 (β) and 0.003 (R) in absolute value, which is approximately 0. This shows that there is a small difference between the expected value and the true value of the parameter being estimated. All the model can estimate the true value of β and R . The Poisson log-linear Leroux CAR model with $m=1, 5, 10$ nearest neighbours can estimate the true value of β and R_k .

Overall, all the models produce a similar result in RMSE for poverty risk R_k in all scenarios. However, the performance of the models is different in terms of RMSE for the regression parameter β . In the weakly autocorrelated between ϕ and independent spatial autocorrelation of the covariate, the Poisson log-linear Leroux CAR model with the $m=10$ performs the worst in terms of RMSE for the β with the value of 0.49. If spatial autocorrelation between ϕ is strongly autocorrelated and independent spatial autocorrelation of the covariate, the $m=10$ model also performs the worst with the highest RMSE (0.114). On the other hand, if there are strongly autocorrelated between ϕ and independent spatial autocorrelation of the covariate, the Poisson log-linear Leroux CAR model with $m=1$ nearest neighbours performs the worst with RMSE is 0.229. The model also performed the worst in the presence of strong spatial autocorrelation between ϕ and strong spatial autocorrelation of the covariate.

Finally, in each scenario, the DIC, a measure of how well a model fits a collection of data, was calculated and presented in the bottom part of Table 1. For scenarios 1, 2, 3 and 4, the Poisson log linear Leroux CAR model with $m=5$ nearest neighbours appears to be the best fit with the lowest DIC values of 671.605 and 681.640, 670.160 and 681.752, respectively. The remaining two models ($m=1$ and $m=10$) have high values of DIC.

Table 1. Summary of the Bias, RMSE and DIC of the regression parameter β and R_k from the simulation study relates to the districts of Kelantan study region

Metric	Scenario	$m=1$	$m=5$	$m=10$
Bias (R_k)	1	-0.003	-0.003	-0.003
	2	-0.003	-0.003	-0.003
	3	-0.002	-0.002	-0.003
	4	-0.003	-0.003	-0.003
Bias (β)	1	0.008	0.006	0.006
	2	0.001	0.008	0.006
	3	0.021	0.021	-0.018
	4	0.055	0.062	0.050
RMSE (R_k)	1	0.040	0.040	0.040
	2	0.042	0.042	0.042
	3	0.040	0.039	0.040
	4	0.042	0.042	0.042
RMSE (β)	1	0.047	0.043	0.049
	2	0.110	0.113	0.114
	3	0.229	0.191	0.212
	4	0.374	0.367	0.363
DIC	1	672.606	671.605	672.700
	2	682.020	681.640	681.967
	3	671.288	670.160	671.524
	4	682.557	681.752	681.971

4. Conclusion

This study has evaluated the performance of a Poisson log Linear Leroux CAR model with $m=1$, 5, and 10 nearest neighbours. There were clear differences in the number of neighbours assigned by $m=1$, 5, and 10 nearest neighbours. The performance of these models has been evaluated through simulation, particularly the reliability in estimating β and poverty risk R_k . In all scenarios, all the models produce bias approximately 0 for both β and R_k . There are similarities in the RMSE of R_k regardless of the different strength of spatial autocorrelation of random effects and covariate. The choice of \mathbf{W} matrices for by $m=1$, 5, and 10 nearest neighbours and scenarios does not affect the RMSE for the risk R_k .

Nevertheless, there is dissimilarity in the RMSE of regression parameter β . If the data contains weakly autocorrelated between ϕ and difference strength of spatial autocorrelation of the covariate, the Poisson log-linear Leroux CAR model with the $m=10$ has the highest RMSE. While in the presence of strongly autocorrelated between ϕ and difference strength of spatial autocorrelation of the covariate, the Poisson log-linear Leroux CAR model with the $m=1$ has the highest value of β . For DIC, the model with $m=5$ nearest neighbours gave the low values of the DIC in all the scenarios, which indicate a better fitting model compared with other models.

Overall, these results suggest that the Poisson log-linear Leroux CAR model with the $m=5$ nearest neighbours performed overall about the best for simulated poverty data since it consistently gives good results across different strength of spatial autocorrelation of random effects ϕ and covariate. This study may assist other researchers in conducting studies involving GIS applications and spatial data related to districts of Kelantan, especially in poverty study. Therefore, future works will involve a different measure of spatial closeness, including adjacency and other distance-based neighbourhood matrices.

Acknowledgement

The authors would like to thank Universiti Malaysia Kelantan for the approval given to conduct this research.

References

- [1] Townsend P, 2010 The meaning of poverty *The British journal of sociology* **61** 85–102.
- [2] Claro S, Paunesku D and Dweck C S 2016 Growth mindset tempers the effects of poverty on academic achievement *Proceedings of the National Academy of Sciences* **113(31)** 8664–68.
- [3] Brody G H, Gray J C, Yu T, Barton A W, Beach S R, Galv'an A, MacKillop J, Windle M, Chen E, Miller G E et al 2017 Protective prevention effects on the association of poverty with brain development *JAMA pediatrics* **171(1)** 46–52.
- [4] Kerr W C, Kaplan M S, Huguet N, Caetano R, Giesbrecht N and McFarland B H 2017, Economic recession, alcohol, and suicide rates: comparative effects of poverty, foreclosure, and job loss *American journal of preventive medicine* **52(4)** 469–475.
- [5] Sledge K E 2019 Reviewing the effects of poverty and food scarcity on the prevalence of type ii diabetes mellitus in the metro atlanta area
- [6] Abas M A, Yusoh M P, Sibly S, Mohamed S and Wee S T 2020 Explore the rural community understanding and practices on sustainable lifestyle in Kelantan, Malaysia *IOP Conf. Ser.: Earth Environ. Sci.* **596** 012054.
- [7] Lee D 2011 A comparison of conditional autoregressive models used in bayesian disease mapping *Spatial and Spatio-temporal Epidemiology* **2(2)**, 79–89.
- [8] Aswi A, Cramb S M, Moraga P, Mengersen K 2018 Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review *Epidemiology & Infection*, **147**.
- [9] Saha D, Alluri P, Gan A and Wu W 2018 Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models *Accident Analysis & Prevention*, **118**, 166-177.
- [10] Leroux B G, Lei X and Breslow N 2000 Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical models in epidemiology, the environment, and clinical trials* 179-191 Springer, New York, NY.
- [11] Lee D 2017 Carbayes version 4.6: An r package for spatial areal unit modelling with conditional autoregressive priors. *Glasgow: University of Glasgow*.