MAKING BAYESIAN DISEASE MAPPING EASY AND INTERACTIVE: AN R SHINY APPLICATION

Aswi Aswi[1], Muhammad Arif Tiro[1], Sudarmin[1], Sukarna[2], Awi[2], Nurwan[1], Susanna Cramb[3]
[1] Statistics Department, Universitas Negeri Makassar, Indonesia
[2] Mathematics Department, Universitas Negeri Makassar, Indonesia
[3] Australian Centre for Health Services Innovation & Centre for Healthcare Transformation, Queensland University of Technology, Australia

e-mail: *aswi@unm.ac.id*

DOI: 10.14710/medstat.16.2.148-159

Abstract: Spatial analysis of count data is important in epidemiology and other domains to identify spatial patterns. While Bayesian spatial models are a popular approach, they do require detailed knowledge of the process for model fitting, checking, and visualising results. Although a number of R packages are available to simplify running the model, there are still complexities when checking the model. This paper aims to provide a user-friendly and interactive R Shiny web application for the analysis of spatial data using Bayesian spatial Conditional Autoregressive Leroux models. The web application is built with the integration of the R packages shiny and CARBayes. The required data are the number of cases, population, and optionally some covariates for each region. In this case, we used Covid-19 data in 2021 in South Sulawesi province, Indonesia. This application enables fitting a Bayesian spatial CAR Leroux model under several hyperpriors and selecting the most appropriate through comparing several goodness of fit measures. The application also enables checking convergence, plus obtaining and visualising in an interactive map the relative risk of disease for each region.

1. INTRODUCTION

Disease mapping has long been a part of epidemiology and public health. Spatial analysis is important in epidemiology and other fields to reveal spatial patterns. Spatial data are divided into three kinds, namely point patterns, geostatistical data, and areal data (lattice data) (Cressie, 1993). Areal data are suitable when the data are aggregated over the sub-location of the study. Due to the need to protect patient confidentiality, areal data is more readily available than other types of spatial data.

While a common measure of small-area disease risk is the Standardised Incidence Ratio (SIR), Bayesian spatial models are often preferred when data are sparse or additional adjusting for covariates is desired. Many Bayesian real models can now be run quite easily through using R-packages such as CARBayes (Duncan, 2013), R-INLA (Finn & Håvard, 2015), and inlabru package (Bachl et al., 2019). R-the Integrated Nested Laplace Approximation (R-INLA) is formatted for latent Gaussian models and a wide range of spatial and spatio-temporal models. CARbayes package is designed for both univariate and

multivariate Bayesian spatial models using Markov chain Monte Carlo (MCMC) simulation where the spatial dependence is assigned by a range of random effects that are implemented by a conditional autoregressive (CAR) prior distribution. Even so, detailed knowledge and model checking is required to ensure results are valid. Specifically, it is always important to check model convergence (when models are computed using MCMC), model goodness-of-fit, and to run a sensitivity analysis to check on the influence of priors. The fundamental benefit of the Bayesian method for modelling spatial structures is that it can easily define spatial information in prior distributions and can account for uncertainty in the estimates or predictions.

Shiny apps are increasingly used to help people run analyses they are less familiar with. For instance, the SpatialEpiApp package (Moraga, 2017) and SSTCDapp package (Adin et al., 2019) help users run Bayesian spatial and spatio-temporal models. These applications focus on the integration of the R package shiny and R-INLA. Here we present a new Shiny app "CARAMapp" (CAR Analysis & Mapping app) to help two groups of researchers: those who wish to simplify the Bayesian spatial model running and checking, and/or those who wish to create interactive visualisations easily. Currently, the app is designed to model, check and map Covid-19 incidence data for one time period (2021) in South Sulawesi province, Indonesia. This user-friendly and interactive R Shiny web application package (Shalabh, 2021; Sievert, 2020) simplifies fitting an appropriate Bayesian spatial CAR Leroux model and compares multiple goodnesses of fit measures for several hyperprior distributions, namely Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), Watanabe Akaike Information Criterion (WAIC) (Watanabe, 2010), and Modified Moran's I (MMI) (Aswi, A. et al., 2021; Carrijo & Da Silva, 2017) for residuals. The Leroux CAR model offers an advantage in addressing identifiability issues by employing a single set of random effects. The Leroux model, can also offer a simpler and more parsimonious representation compared to other global spatial smoothing models such as the BYM model. If the structure and assumptions of the Leroux model align well with the characteristics of the spatial data, it might be preferred for its simplicity.

2. LITERATURE REVIEW
### 2.1. Bayesian Disease Mapping

Disease mapping encompasses various interpretations. In contemporary contexts such as public health and disease geography, it involves the disease incidence mapping and relative risks at a small geographical scale. This is often done as part of research on the spatial distribution of diseases within a specific population (MacNab, 2022). This field of study can be categorized into four main areas: estimating relative risk (RR), identifying disease clusters, conducting ecological analyses, and implementing surveillance measures (Rao et al., 2017).

2.2. R Shiny

Shiny is a web-based framework that uses the basic R language. Shiny is easy to use to develop and create more interactive web-based applications package (Chang et al., 2022). R-Shiny web applications are divided into two main categories: User Interface (UI) and Server. The purpose of the UI is to make it easier for web application users to see the content and provide input for the relevant online application. In contrast, a server is an element of a web application that has a specific function to launch the relevant algorithms that have been incorporated into it by the user interface. The building blocks of the shiny package are based on reactive programming package (Wan & Hudak, 2000). Reactive programming consists

of three basic building blocks namely, reactive values, expression, and R-shiny. The main task of a web application is to input and generate output. The programming language is intentionally designed to ensure that any modification made to the input, whether it is data or hyperpriors, will have a direct impact on the final outcome. These changes will be promptly displayed through text, tables, or images. This characteristic proves advantageous for web applications that heavily depend on user input. Shiny procedures enable the generation of varied outputs without needing to refresh the web page.

Inside the shiny package, controllers or widgets are generally included to facilitate the usability for application developers. Numerous procedures, such as file uploads and automatic refreshing of the page for generating new plots and tables, are provided as built-in features. These actions are executed based on the previously created output widget. WebSocket packages enable efficient communication between the web user interface and R. Websockets play a crucial role, particularly in scenarios involving continuous and interactive dialogue or data exchange between the client and server. In general, programming components with R-Shiny are given in Figure 1.
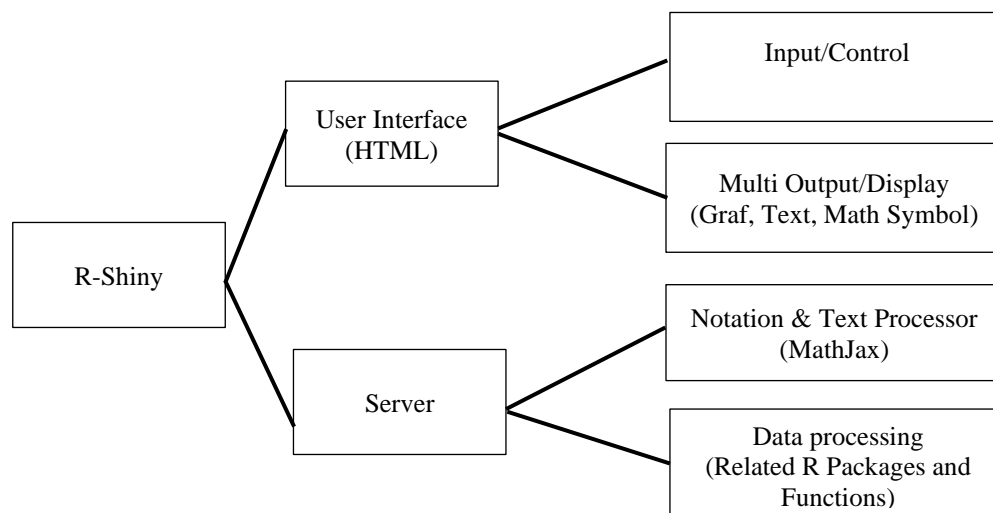
Figure 1. R-Shiny Program Components

3. MATERIAL AND METHOD
3.1. Study Area and Data

The South Sulawesi Province is situated within the longitudinal range of 116º 48' to 112º 36' East and the latitudinal range of 0°12' and 8° South. It has 21 districts and three cities. Covering an area of 46,717.48 square kilometres, the province had an estimated population of approximately 9.139 million people in 2021 (Badan Pusat Statistik, 2021).

This study utilized aggregated data from the official website of the "Ministry of Health of the Republic of Indonesia" at https://infeksiemerging.kemkes.go.id/ to obtain the number of confirmed Covid-19 cases in 24 districts (1 January 2021 -31 December 2021) The calculation of expected count data also involved population data obtained from the "Badan Pusat Statistik" (Badan Pusat Statistik, 2022).

3.2. Bayesian Spatial Models

In this research, the Bayesian spatial using CAR Leroux (Leroux et al., 2000) was used to estimate the RR of Covid-19 in South Sulawesi Province, Indonesia. The model incorporates a random effect called u_i which represents a single spatial structure and

accounts for the level of spatial autocorrelation through a constant value ρ. The number of Covid-19 cases (y_i) is modelled using the Poisson distribution as follows.

$$y_i \sim \text{Poisson}(E_i \theta_i) \tag{1}$$

where y_i and θ_i are the number of cases and the relative risk of Covid-19 in the ith area, respectively. The Poisson distribution is commonly used for modeling count data, where the outcome is a non-negative integer representing the number of events occurring in a fixed interval. The Poisson distribution can still be effectively employed for modeling count data within a Bayesian framework, and it can be adapted to account for overdispersion through the use of informative priors or specific modeling strategies. Since age information is not available in this data, here E_i is calculated in a simple way as the crude overall rate multiplied by the population (n_i) in each area i (Aswi, 2020; Moraga, 2020):

$$E_i = \frac{\sum_i y_i}{\sum_i n_i} n_i \tag{2}$$

The linear predictor is then:

$$\log(\theta_i) = \beta + u_i \tag{3}$$

where β is an overall constant rate and u_i is modelled by a CAR Leroux prior as follows.

$$\left(u_i \middle| u_j, i \neq j, \tau_u^2\right) \sim N\left(\frac{\rho \sum_j u_j w_{ij}}{\rho \sum_j w_{ij} + 1 - \rho}, \frac{\tau_u^2}{\rho \sum_j w_{ij} + 1 - \rho}\right) \tag{4}$$

for $j = 1, \ldots, n$ areas. The spatial neighborhood matrix w_{ij} is constructed a using binary matrix that incorporates first-order adjacency weight. The prior for $\rho \sim$ Uniform (0,1) and $\beta \sim N(0, 100000)$ (Lee, 2013).

As part of a sensitivity analysis, five different hyperpriors were utilized to assess the variance component τ_u^2. The Inverse Gamma distribution is commonly used as a hyperprior for the variance parameter in Bayesian spatial models. Its flexibility allows it to capture a wide range of shapes, which is particularly useful when dealing with spatial data. Some hyperpriors included: the default hyperprior of CARBayes, which is an Inverse-Gamma distribution with parameters IG(1, 0.01) (Duncan, 2013), as well as IG (0.1, 0.1), IG (1, 0.1), IG (0.5, 0.0005), and IG (0.5, 0.05). The estimation of the posterior distribution for the parameter was conducted using MCMC methods. These methods generate samples from the joint posterior distribution of the parameters. The process involved 100,000 iterations, resulting in 110,000 MCMC samples after discarding the initial 10,000 samples as a burn-in period. The convergence was checked by using the visualisation of the trace and density plots. All models were analysed using the CARBayes package version 5.3 package (Lee, 2013). The best model was selected by several criteria, namely DIC (Spiegelhalter et al., 2002), WAIC (Watanabe, 2010), and MMI for residuals which are appropriate to use even for a few areas (Aswi, A. et al., 2021; Carrijo & Da Silva, 2017) in the software R version 4.1.2 package (R Core Team, 2019).

3.3. R Shiny for Bayesian Spatial CAR Leroux Models

The R-Shiny web application is built to analyze the spatial patterns of Covid-19 in South Sulawesi Province using a Bayesian spatial CAR Leroux Model. The software used is R version 4.1.2. The packages used are shiny, shinydashboard, rgdal, spdep, CARBayes, readxl, leaflet, DT, and coda. The data needed is a shapefile file package that contains polygons of 24 districts in South Sulawesi Province and their variables (count, population, and any additional area-level covariates). The R-Shiny web application allows users to

upload files, select cases and populations, and choose hyperpriors. The structure of the R-Shiny Web Application System is given in Figure 2.
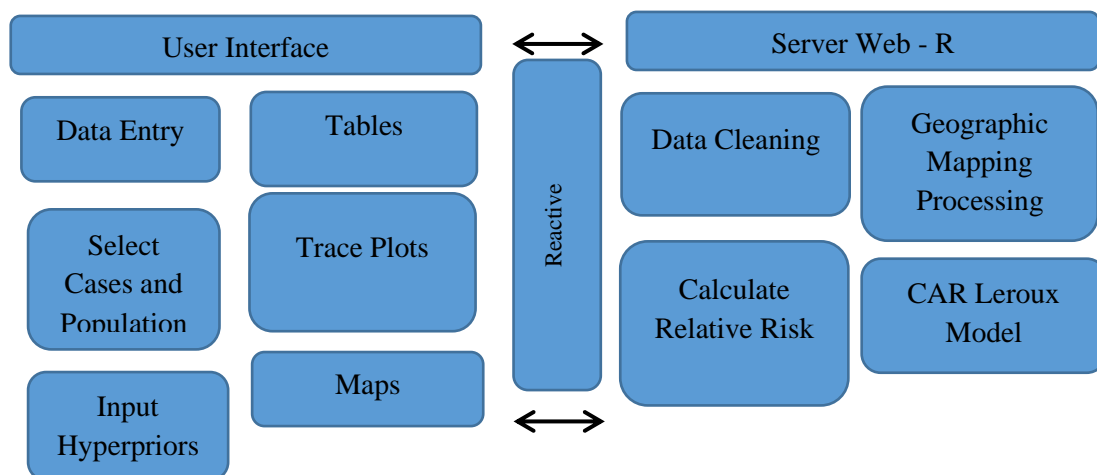
Figure 2. Structure of the R-Shiny Web Application System

3.3.1. Data Menu

The "Data menu" is the first menu that we see when the web application of "SPATIAL ANALYSIS" is opened. This menu consists of a "Browser" button to upload data and a table to display the inputted data. First, we need to upload a data file. The data file must be in .xlsx format and can be uploaded by clicking the "Browser" button and selecting a file from local storage. The data file are with the following columns:
a. Districts: a specific identifier of the region.
b. Cases: number of cases in each area.
c. Population: population in each area.

In our example, first, we upload the data file. The data consists of the number of confirmed positive cases of Covid-19 and the population for each district in South Sulawesi Province in 2021. The columns of the data file are given as follows.
a. District: district name.
b. Cases Covid-19: the number of confirmed positive cases of Covid-19.
c. Population: total population in each district.

3.3.2. Model Comparison Menu

This menu consists of a drop-down sub-menu (Cases and Population columns) and a table containing the DIC, WAIC, and MMI for residual values for the CAR Leroux model for each hyperprior (IG (1, 0.01), IG (1, 0.1), IG (0.1, 0.1), IG (0.5, 0.05), and IG (0.5, 0.0005)). After uploading the data file, we need to determine the best model based on all the criteria used (the smallest DIC, WAIC values, and MMI residuals close to zero). First, press the "Model Comparison" menu on the left panel. Then we need to specify variable cases and populations in the drop-down sub-menu. In our example, we select Cases as "Cases Covid-19" and Population as "Population".

3.3.3. Spatial Model Menu

This menu consists of a sub-menu for entering the selected Hyperprior values based on the best model and a table containing the results of the model (Mean, 2.5%, 97.5%, n.sample, % accept, n.effective, Geweke.diag) for all parameters. After obtaining the model with the most appropriate hyperprior, the user can display the model description by clicking

on the Spatial Model menu in the left panel. Then, enter the hyperprior value and the results will appear automatically on the menu.

3.3.4. "Convergence" Test Menu

The Convergence Test menu automatically displays the trace plot after entering the hyperprior value in the Spatial Model menu. The trace plot of the model sample parameters displayed includes tau2, rho samples, beta samples, and Phi samples.

3.3.5. Map of "Relative Risk" Menu

The Relative risk map menu includes visualisations that display the data and results: a map and a table. The map displays the relative risk values corresponding to the selected case variable in the drop-down menu located in the model comparison menu. The map can be enlarged to allow for better inspection of the areas, and the RR values for each area. When the mouse floats over any area, the area is highlighted and we can see the area's name and the RR values come up to the left or right of the mouse.

The table exhibits information for all areas on the map. Users can filter information by searching for words in a column or across tables. It is also possible to sort the rows in descending or ascending order of the values of any columns.

4. RESULTS

4.1. Data

The program syntax can be found on the link
https://github.com/aswistatunm/RShiny_Spasial-CAR-Leroux-Sulsel/blob/main/app.R

The program can be accessed by running the following R syntax.

```
install.packages(c("shinydashboard","rgdal","spdep","CARBayes","readx
l","leaflet","DT","coda"))

library(shiny)

runGitHub(repo    =    "RShiny_Spasial-CAR-Leroux-Sulsel",username    =
"aswistatunm",ref = "main")
```

This data menu allows the user to upload files from the user's local device storage. Figure 3 shows the input data for Covid-19 cases in 2021. Observations from the data consist of 24 districts in South Sulawesi Province.

Figure 3. Data Menu Display

4.2. Compare Bayesian Spatial CAR Leroux Model Versions

After uploading the data, the next step is to determine the most appropriate hyperprior. To get results from several models using this application, we need to select the Model Comparison menu, then select which data columns represent cases and population. The analysis results using five different hyperpriors will automatically appear as shown in Figure 4.


Figure 4. Display of Model Comparison Menu

Each model contains DIC, WAIC, and MMI values. Based on the results in Figure 4, the smallest DIC value (326.34), is the model with hyperprior IG (1, 0.01) where this model also has the smallest WAIC value (872). The model that has the MMI residual value closest to zero (-0.06) is the CAR Leroux model with hyperprior IG (1, 0.1), but this is reasonably similar to the MMI residual (-0.08) for the model using hyperprior IG (1, 0.01). Therefore,

the CAR Leroux model using hyperprior IG (1, 0.01) is the preferred model in estimating the RR value of Covid-19 cases.

After obtaining the best model, users can access the model description through the Spatial Model menu, which displays parameter estimates and goodness-of-fit measures. Figure 5 illustrates the outcomes of parameter estimates as well as goodness-of-fit measures using the CAR Leroux model with the hyperprior IG(1, 0.01).
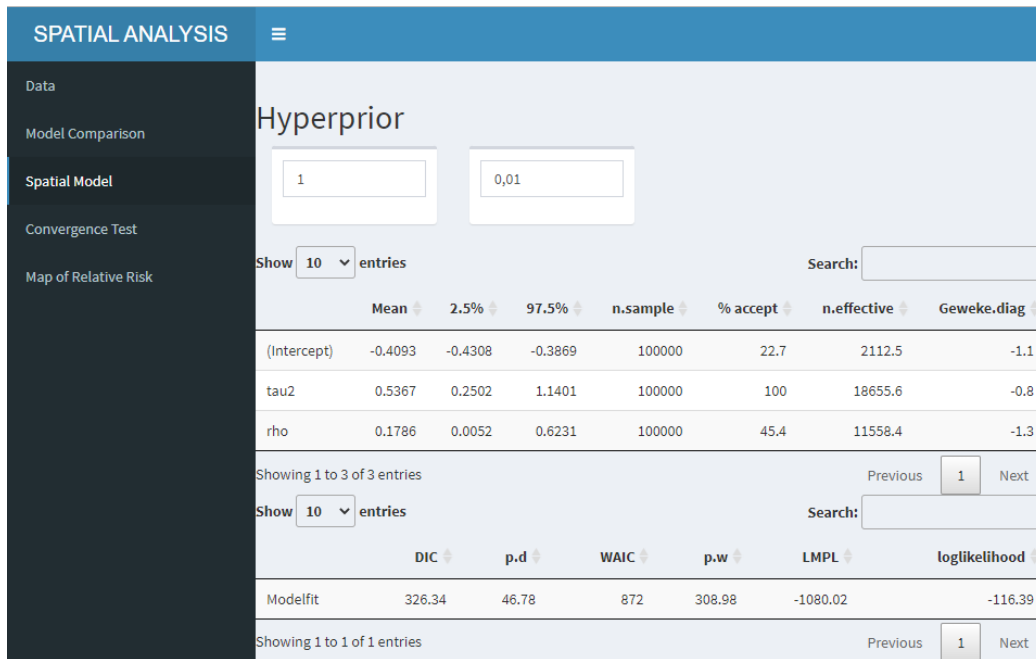

Figure 5. Spatial Model Menu Display

4.3. Convergence Check

The Convergence Test menu contains a trace plot of the parameters using the best model. A trace plot is used to observe the convergence of posterior parameters. The Convergence Test menu automatically displays a trace plot after entering a hyperprior value in the Spatial Model menu. Figures 6,7,8,9 show a trace plot of parameter values using the best model.
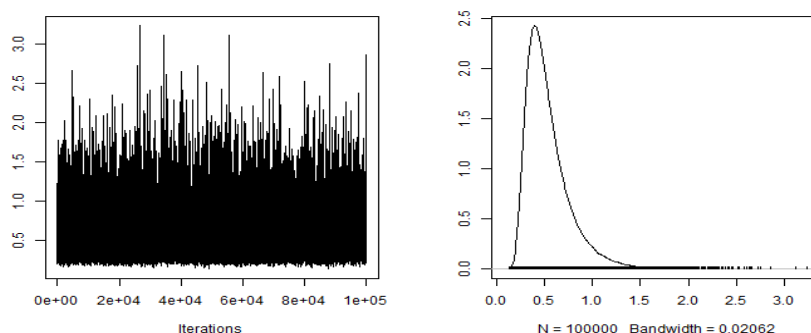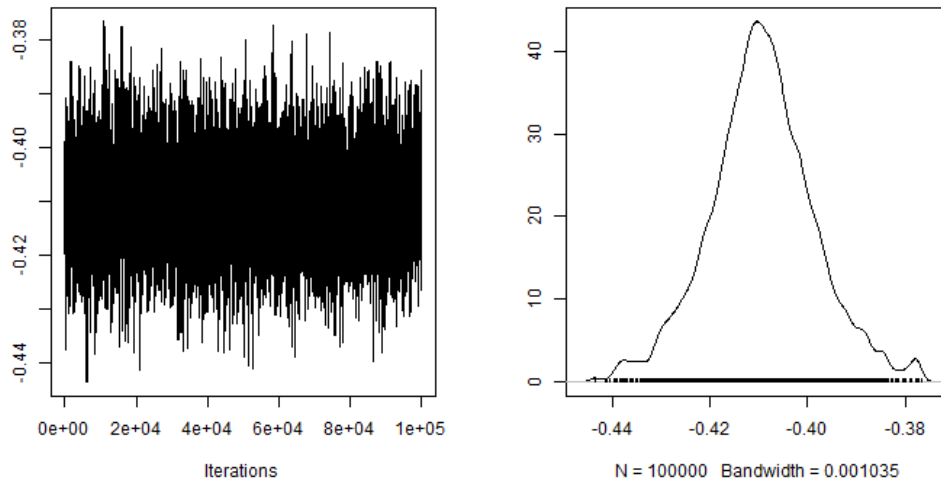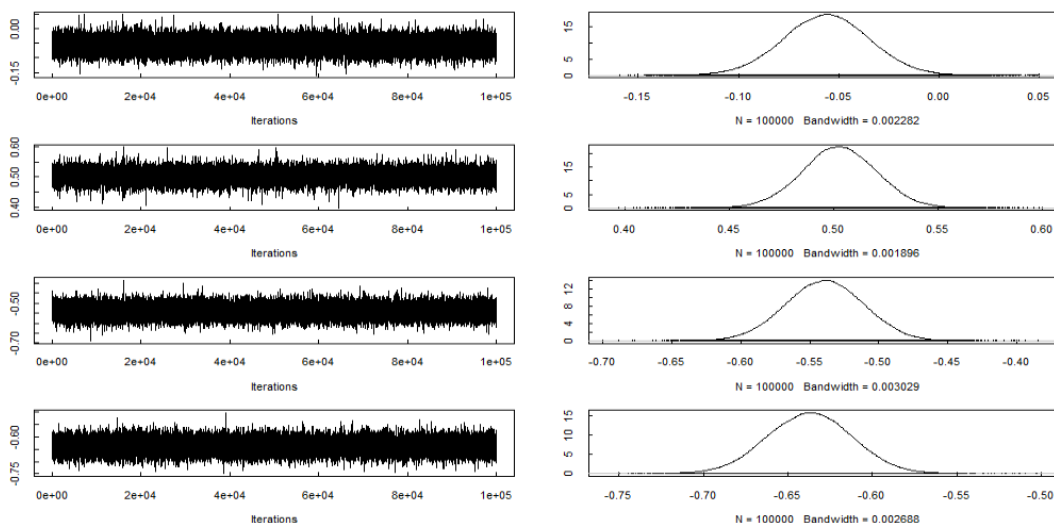

Figure 6. Trace Plot for tau2 Samples
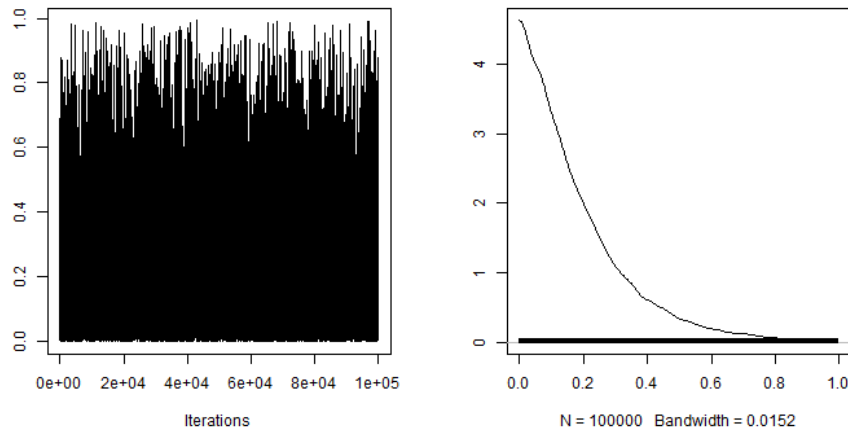
**Figure 7.** Trace Plot for rho Samples

Figure 8. Trace Plot for beta Samples

Figure 9. Trace Plot for phi Samples

4.4. Relative Risk Mapping

The last menu of this application is the Map of RR menu which contains thematic maps and relative risk results (Figure 10). The results of this menu will automatically appear after using the previous menus. Figure 10 displays the thematic map and data containing the estimated relative risk value based on the selected model.
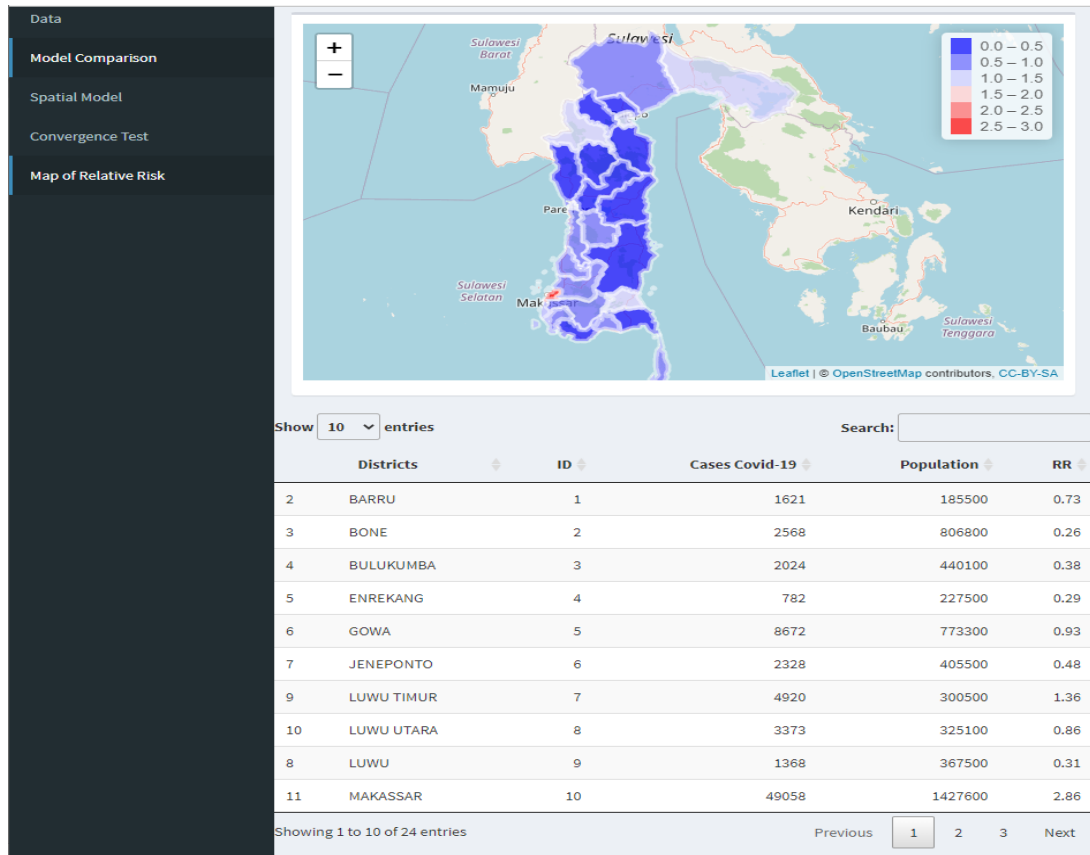
Figure 10. Relative Risk Map Menu Display

Table 1. The RR Value of Covid-19 Cases for Each District/City Based on The Preferred Model

ID	Districts/Cities	RR	ID	Districts/Cities	RR
10	Makassar	2.86	13	Pangkep	0.70
14	Pare-Pare	1.47	20	Takalar	0.63
8	Luwu Timur	1.36	24	Bantaeng	0.54
12	Palopo	1.24	6	Jeneponto	0.48
18	Sinjai	1.20	17	Sidrap	0.41
21	Toraja	1.10	22	Toraja Utara	0.39
5	Gowa	0.93	3	Bulukumba	0.38
9	Luwu Utara	0.86	23	Wajo	0.35
11	Maros	0.86	15	Pinrang	0.33
19	Soppeng	0.85	7	Luwu	0.31
16	Selayar	0.74	4	Enrekang	0.29
1	Barru	0.73	2	Bone	0.26

Table 1 presents the RR values of Covid-19 cases for each district, as determined by the chosen model. Based on Figure 10 as well as Table 1, it is known that the district with the highest RR value is Makassar City (RR=2.86). Makassar city has the RR value (2.86),

indicating that the risk of testing positive for Covid-19 in Makassar is 186% higher than the average. In simpler terms, the probability of a positive Covid-19 confirmation in Makassar is nearly three times higher than that in the general population. Areas with an RR value greater than 1 indicate that the districts/cities are high-risk areas, namely Pare-Pare (1.47), Luwu Timur (1.36), Palopo (1.24), Sinjai (1.20), and Toraja (1.10). Regions with values below 1 are indicative of low risk. Among the areas analyzed, Bone exhibits the lowest RR value of 0.26, followed by Enrekang with 0.29 and Luwu with 0.31.

5. CONCLUSION

Overall, it can be concluded that the R-Shiny application CARAMapp can make it easier for users to analyse disease data at the district level in the Province of South Sulawesi using the CAR Leroux model. By using this web application, the most suitable model for estimating the RR of Covid-19 in South Sulawesi Province in 2021 is the spatial CAR Leroux model with hyperprior IG (1, 0.01). Based on this model, the area with the highest and the lowest RR value is Makassar City (RR = 2.86) and Bone (RR = 0.26), respectively.

There are some important limitations to the current version. No adjustment for age is incorporated into the expected count calculation, so age can influence the mapped patterns. Also, this web application provides only one Bayesian spatial model with different hyperpriors. Considering other Bayesian spatial CAR models as well as Bayesian spatio-temporal CAR models could be possible future work, and allowing users to upload their own shapefiles for other regions is also planned.

ACKNOWLEDGMENT

REFERENCES

Adin, A., Goicoa, T., & Ugarte, M. D. (2019). Online Relative Risks/Rates Estimation in Spatial and Spatio-Temporal Disease Mapping. *Computer Methods and Programs in Biomedicine, 172*, 103-116. doi:10.1016/j.cmpb.2019.02.014

Aswi. (2020). *Bayesian Spatio-Temporal Modelling of Small Areas: Dengue Fever in Makassar Indonesia.* Queensland University of Technology.

Aswi, A., Cramb, S., Duncan, E., & Mengersen, K. (2021). Detecting Spatial Autocorrelation for a Small Number of Areas: a Practical Example. *Journal of Physics. Conference series, 1899*(1), 12098. doi:10.1088/1742-6596/1899/1/012098

Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., & Freckleton, R. (2019). inlabru: an R package for Bayesian Spatial Modelling from Ecological Survey Data. *Methods in Ecology and Evolution, 10*(6), 760-766. doi:10.1111/2041-210X.13168

Badan Pusat Statistik. (2021). *Sulawesi Selatan dalam Angka 2021*.

Badan Pusat Statistik. (2022). Sulawesi Selatan dalam Angka 2022.

Carrijo, T. B., & Da Silva, A. R. (2017). Modified Moran's I for Small Samples. *Geographical Analysis, 49*(4), 451-467. doi:10.1111/gean.12130

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke , B., Xie, Y., McPherson, J, Dipert, A., Borges, G. (2022). shiny: Web Application Framework for R. R Package Version 1.7.2. Retrieved from https://cran.r-project.org/web/packages/shiny/index.html

Cressie, N. A. C. (1993). *Statistics for Spatial Data* (Rev. ed. ed.). New York: Wiley.

Duncan, L. (2013). CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. *Journal of Statistical Software, 55*(1), 1-24. doi:10.18637/jss.v055.i13

Finn, L., & Håvard, R. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software, 63*(1), 1-25. doi:10.18637/jss.v063.i19

Lee, D. (2013). CARBayes: An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors. *Journal of Statistical Software, 55*(1), 1-24.

Leroux, B. G., Lei, X., & 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, 116*, 179-191.

MacNab, Y. C. (2022). Bayesian Disease Mapping: Past, Present, and Future. *Spatial Statistics, 50*, 100593-100593. doi:10.1016/j.spasta.2022.100593

Moraga, P. (2017). *SpatialEpiApp :* A Shiny Web Application for the Analysis of Spatial and Spatio-Temporal Disease Data. *Spat Spatiotemporal Epidemiology,* Nov:23:47-57. doi: 10.1016/j.sste.2017.08.001.

Moraga, P. (2020). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Boca Raton: CRC Press.

R Core Team. (2019). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Rao, A. S. R. S., Payne, S., & Rao, C. R. (2017). *Disease Modelling and Public Health. Part A*. Amsterdam, Netherlands: North Holland is an imprint of Elsevier.

Shalabh. (2021). Interactive Web-based Data Visualization with R, Plotly, and shinyCarsonSievert; Chapman & Hall/CRC; 2020; ISBN 978-1-138-33145-7. *Journal of the Royal Statistical Society. Series A, Statistics in society, 184*(3), 1150-1150. doi:10.1111/rssa.12692

Sievert, C. (2020). *Interactive Web-based Data Visualization with R, Plotly, and Shiny*. Boca Raton, Florida ;: CRC Press.

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. doi:10.1111/1467-9868.00353

Wan, Z., & Hudak, P. (2000). *Functional Reactive Programming from First Principles.* Paper presented at the Conference on Programming Language Design and Implementation.

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research, 11*, 3571-3594.