ISSN: 0067-2904

Spatio-Temporal Mixture Model for Identifying Risk Levels of COVID-19 Pandemic in Iraq

Sadeq A. Kadhim*, Safaa K. Kadhem[2]
[1]*Higer Education and Scientific Research Ministry, Iraq*
[2]*College of Administration and Economics, AL Muthanna University, AL Muthanna, Iraq*

Received: 20/5/2022 Accepted: 7/9/2022 Published: 30/6/2023

Abstract

This paper focuses on choosing a spatial mixture model with implicitly includes the time to represent the relative risks of COVID-19 pandemic using an appropriate model selection criterion. For this purpose, a more recent criterion so-called the widely Akaike information criterion (WAIC) is used which we believe that its use so limitedly in the context of relative risk modelling. In addition, a graphical method is adopted that is based on a spatial-temporal predictive posterior distribution to select the best model yielding the best predictive accuracy. By applying this model selection criterion, we seek to identify the levels of relative risk, which implicitly represents the determination of the number of the model components of all regions over independent time periods. The estimation of parameters and the model selection are both performed in a Bayesian framework. Also, the means of estimated relative risk for the selected mixture model are mapped to give a clearer picture of distributing the disease risks in each district.

Keywords: Mixture model, Widely Akaike information criterion, Relative risk, COVID-19, Bayesian framework

نموذج خليط مكاني- زماني لتحديد مستويات خطورة وباء كوفيد-19في العراق

صادق عواد كاظم[1]*, صفاء كريم كاظم[2]
[1]وزارة التعليم العالي والبحث العلمي, بغداد، العراق
[2]كلية الادارة والاقتصاد, جامعة المثنى,المثنى، العراق

الخلاصة

يركز هذا البحث على اختيار نموذج خليط مكاني مع الاخذ بنظر الاعتبار الزمن ضمنيا لتمثيل الخطورات النسبية لوباء كوفيد-19 باستخدام معيار اختيار نموذج ملائم. لهذا الغرض,تم استخدام معيار اكثر حداثة يسمى معيار معلومات اكياكي الموسع والذي نعتقد بان استخدامه في سياق نمذجة الخطورة النسبية هو محدود جدا. بالاضافة لذلك, تم تبني طريقة رسومية تستند على توزيع لاحق تنبوئي مكاني-زماني لاختيار افضل نموذج يعطي افضل دقة تنبؤي. بتطبيق معيار اختيار النموذج, نسعى من خلال هذا البحث لتحديد مستويات الخطورة النسبية, والتي تمثل ضمنيا تحديد عدد مكونات النموذج, لكافة المناطق عبر فترات زمنية مستقلة. تم اجراء تقدير المعلمات واختيار النموذج باستخدام اطار بيزي. ايضا, تم رسم متوسطات الخطورة النسبية المقدرة على الخريطة لتعطي صورة واضحة لتوزيع خطورة المرض في كل منطقة.

*Email: sadiq2061@gmail.com

1. Introduction

The health data about the spreading of a certain disease is typically reported as a statistical summary or information that differs from one region to region. This variation in data forms is called spatial variation or heterogeneity. One of the challenges is how to accommodate or capture this heterogeneity in data that can be not easy to modelling using standard statistical models. On the other side, the spatial modelling of heterogeneity maybe not enough to describe the evolution of disease over time. Hence, inserting time, as another dimension with the space, can form an important role in analyzing precisely the spread of the disease. Besides, mapping of this spatial-temporal heterogeneity can appear to the underlying structure of scattered infections data [1].

There have been many literatures focusing on the development of spatial and temporal modeling methods to accommodate this type of data heterogeneity. In the literature on relative risk, the measure of standardized mortality ratio (SMR) is the most often used as an epidemiological approach to map relative risk. However, this path has some cons. This measure gives unstable results, especially in small areas [2]. In addition, this approach cannot identify the high or low levels of risk of disease in infected regions under the research [3]. Despite of existence some traditional methods to classify health areas according to their relative risk such as the standardized mortality ratio (SMR) [2] and the Gamma–Poisson model which has been introduced abundantly by [4, 5, 6]. But, those methods have some disadvantages in describing the spatial distribution of the risk of the disease. To overcome these limitations. A flexible methodology called mixture modeling has long been used to treat such heterogeneity in data [7]. They have been used in many articles introduced to analyse the disease nature. For instance, to reveal the extreme mortality risks of breast cancer and Scottish lip cancer, a comparison using two types of mixture models was implemented by [8] to how well each of these models determines the districts which have high risks. In [9], the authors developed a spatial mixture model to analyse the mortality of cancer of the larynx data among females in France. K. C. Flórez et al. [10] used a Poisson mixture model to estimate the relative risk and cluster detection for the reported cases of varicella disease in Spain. S. K. Kadhem and S. A. Kadhim [11] Adopted a Markov model to analyse the absorbing case (death) of COVID-19 patients as a criterion to measure the risk rate in the Iraqi population. The contribution of this paper is to develop spatial modelling for the evolution of the COVID-19 infections in Iraq through independent fixed times. Actually, this work is developed from a previous work introduced by [12], who shows that spatial modelling using mixture models is an appropriate approach to analyse the COVID-19 infections. In other words, they modelled the relative risk of this pandemic through different fixed periods and each period has its own mixture model. The issue of determination of levels of relative risk, i.e. model selection has special importance in identifying and evaluating the disease risk which can form a key point in the working of health systems and also controls diseases [1]. So, this paper also contributes by introducing a model selection methodology using two approaches. The first approach is the model selection criterion which is so-called the widely Akaike information criterion (WAIC) introduced by [13]. The properties of this criterion were investigated by [14] and [15] using an application that included a hidden mixture modelling. The second approach is confined to supporting the model selected in the first approach by checking its predictive accuracy using the predictive posterior distribution. At each period, the model is selected that gives the best goodness of fit to the data using the above two tools. The aim of conducting such a study is to stimulate health organizations in revising their future plans for such infectious diseases. The results of this study give a warning to health organizations by determining the most affected regions.

The structure of this paper is distributed as follows. In Section 2, the definition of the proposed model and approaches to model selection under the Bayesian principle is introduced. Section 3 includes the description of infection data of COVID19. In section 4, all results regarding the model estimation and selection are presented. Finally, section 5 shows the conclusions.

2. Material and methods

2.1 Bayesian spatial-temporal mixture model

Assume that there is a disease occurring within n regions and is also observed within T successive fixed time periods. The resulting observed infections count from that disease can define as y_{it} where $i = 1,2,\dots,n;\ t = 1,2,\dots,T$. For relatively rare diseases, the observed infections count yit is often modelled as a standard Poisson model so that

$$y_{it} \sim Pois(\theta_{it}), \tag{1}$$

where θ_{it} is the mean parameter and its probability mass function is given by [16]:

$$f(y_{it};\theta_{it}) = \frac{{\theta_{it}}^{y_{it}} e^{-\theta_{it}}}{y_{it}!}. \tag{2}$$

Since disease occurs within a population at risk for the disease, the mean θ_{it} has to be modified by a population effect in some way. This modification is often done by two indicators such that

$$\theta_{it} = E_{it}\lambda_{it}. \tag{3}$$

The first indicator E_{it} refers to the population at risk and is often computed as the expected count. The E_{it} is estimated based on a population number [9]:

$$E_{it} = N_{it} \times \frac{\sum_{j=1}^{L}\sum_{t=1}^{T} y_{jt}}{\sum_{j=1}^{L}\sum_{t=1}^{T} N_{jt}}, \tag{4}$$

where N_{it} is the size of the population of i^{th}region that observed at time t (we here assume that population size for each region is fixed over all time periods) and L is the number of all regions under study. The second indicator λ_{it} is known as a relative risk which is computed as $\lambda_{it} = \frac{y_{it}}{E_{it}}$. Since the values of y_{it} and E_{it} are not fixed and can change from one region to another region and from one time to another time, the values of relative risk as a result are affected and heterogeneous. Our focus is on modelling the heterogeneity in the values of relative risk which can be accommodated by a mixture model that takes into account the spatial trend as well the temporal of disease. Our focus is on modelling the heterogeneity in the values of relative risk which can be accommodated by a mixture model that takes into account the spatial trend as well as the temporal disease. In this case, a finite spatial-temporal mixture with K components will be developed to accommodate the heterogeneity in the relative risks. That means, the model in equation (2) is extended such that the mixture density, which is the weighted sum of Poisson densities for each region i and each time point t, is obtained as follows [17]:

$$\boldsymbol{f(y_{it}, \Omega, E_{it}) = \sum_{j=1}^{K} w_{jt} f(y_{it}, \lambda_{jt}, E_{it})}, \tag{5}$$

With $0 < w_{jt} < 1$ and $\sum_{j=1}^{K} w_{jt} = 1$ for all t,

where $\Omega = (\boldsymbol{\lambda}, \boldsymbol{w})$ represents a collection of the model parameters and w_{it} refers to the mixing weighing. In the context of latent class models, a process called the hidden allocation or data augmentation [18] is often included to interpretation purposes and to simplify the computational complexities of such models [19]. Let $\mathbf{z} = \{z_{ijt}\}, i = 1,2,\dots,n;\ t = 1,2,\dots,T;\ j = 1,2,\dots,K$ be an allocation vector that indicate the component to which y_{it} belongs, where K

refers to the components or groups number in the model, so that $z_{ijt} \in \{0,1\}$ and $\sum_{j=1}^{K} z_{ijt} = 1$. The pmf of the complete data $(y_{it};\ z_{ijt})$ is

$$f(y_{it}, z_{ijt}|\lambda, w, E) = \prod_{j=1}^{K} w_{jt}^{z_{ijt}}\, p[y_{it}|\lambda_{jt}, E_{it}]^{z_{ijt}}. \quad (6)$$

The joint complete posterior distribution for the model parameters can be given by:

$$f(\lambda, w, z|y, E) \propto f(y, z|\lambda, w, E) f(z|w) f(\lambda) f(w), \quad (7)$$

where the first term in equation (7) represents the complete data likelihood, the second term represents the posterior of allocation variables, the third term represents the prior on the relative risk parameter and the fourth term represents the prior on the mixing weights. For completing the specification of the Bayesian model, the priors of unknown model parameters have to be specified. The prior on the mixing weights w are given independently Dirichlet distribution with hyper-parameter η:

$$f(w|y, z) \propto \prod_{j=1}^{K} w_{jt}^{R_{jt}+\eta_{jt}+1} = Dirichelt\left(R_{jt} + \eta_{jt}\right), \quad \text{for each } t \quad (8)$$

where $\eta_{jt} > 0$ for each t is hyper-parameters of the Dirichlet distribution and $R_{jt} = \sum_{l=1}^{n} \mathbb{I}_{zlt=j}, j = 1,2, \ldots, K,$ denote the allocation sizes. The prior on the component-specific relative risk parameter λ, we independently assume a Gamma distribution as a conjugate prior [20], on each distinct relative risk parameter $\lambda_{zit} \equiv \lambda_j$ such that

$$f(\lambda_j|\alpha, \beta) \sim Gamma(\alpha, \beta),$$

$$= \frac{\lambda_j^{\alpha-1} e^{-\beta\lambda_j} \beta^{\alpha}}{\Gamma(\alpha)}, \qquad \lambda_j > 0;\ \alpha > 0, \qquad \beta > 0,$$

where α and β are hyper-parameters that represent the shape and rate or inverse scale parameters of the Gamma distribution respectively. The Gamma prior density above has mean α/β and variance α/β^2. Sampling from the relative risk parameter and the parameter of mixing weights are implemented using the Gibbs sampler, an MCMC method that is based on sampling from conditional posterior distribution instead of the joint posterior. Given the observations,y_{it}, the fully conditional posterior distribution of allocation probability z_{ijt} can be obtained as

$$f(z_{it} = j|y_{it}, \lambda, w, E) \propto w_{jt} Pois\left(y_{it}|E_{jt}, \lambda_{jt}\right) = \frac{w_{jt}\, Pois(y_{jt}|E_{jt}\lambda_{jt})}{\sum_{l=1}^{K} w_{lt} Pois\,(y_{lt}|E_{lt}\lambda_{lt})} \quad (9)$$

2.2Data source

This article focuses only on the observed number of infected people for the period from March to December 2021 which has the highest infections in Iraq. All information about the population of each province, infections and relative risk (the parameter of interest) are attached in the appendix. The number of infected people was directly obtained from the world health organization (WHO) (World Health Organization, 2021)[21]. According to the central statistical organization of Iraq (Central Statistical Organization of Iraq, 2019) (CSO)[22], the estimated population denoted $by\ n_i$ was provided based on the estimated census of 2019. The parameter of relative risk E_{it} in equation 4 is computed based on the number of observed infections and populations for each province also provided in the appendix

.

2.3Determining the relative risk levels

This section is interested in determining for the spatial levels of relative risk of pandemic over independent times. For this purpose, two approaches have been employed. In the first approach, the WAIC model selection criterion is used to choose the best model to fit the infection data among several candidate models. In the second approach, a graphical tool which is represented by the predictive posterior distribution of infection data is adopted. This former

step is to examine the predictive accuracy of the model selected in the first approach.

2.3.1The WAIC

Assume that a sequence of count data set $y = \{y_{it}\}, i = 1,2,\dots,n$ and $t = 1,2,\dots,T$ that follows a mixture Poisson distribution with the parameters collection: (λ, w, E), and a sequence of latent variables $z = \{z_{ijt}\}, i = 1,2,\dots,n; j = 1,2,\dots,K;\ t = 1,2,\dots,T$ in which each z_{ijt} is specified for each corresponding observation y_{it}, the integrated pointwise predictive density (ilppd) can be defined as follows:

$$ilppd_y = \log \prod_{i=1}^{n} \prod_{t=1}^{T} Pois_{post}(y_{it}),$$

$$= \sum_{i=1}^{n} \sum_{t=1}^{T} \log E_{\{z,\lambda,w\}}\left[Pois\left(y_{ijt}\middle|z_{it},\lambda,w\right)|\boldsymbol{y}\right],$$

$$= \sum_{i=1}^{n} \sum_{t=1}^{T} \int_z \int_\lambda \int_w Pois\left(y_{it}\middle|z_{ijt},\lambda,w\right) Pois\,(\boldsymbol{z},\boldsymbol{\lambda},\boldsymbol{w}|\boldsymbol{y}) dz\ d\lambda\ dw \quad (10)$$

The integrals in equation 10 can be approximated by the posterior draws of the model parameters (z^m, λ^m, w^m) that were obtained over the Gibbs sampling. Hence, the approximated ilppd can be then given as follows:

$$\widehat{ilppd}_y = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{n} \sum_{t=1}^{T} \log Pois\,(y_{it}\,|\boldsymbol{\lambda}^{(m)}_{z^{(m)}_{ijt}}, \boldsymbol{w}^{(m)}_{z^{(m)}_{ijt}}, \quad (11)$$

Where $\lambda^{(m)}_{z^{(m)}_{ijt}}$ and $w^{(m)}_{z^{(m)}_{ijt}}$ represent the m^{th} state-based relative risk and risk weight, respectively. To avoid the bias, Gelman et al. (2014) [21] proposed adding a correction term or so-called effect number of parameters p_{WAIC} which is based on computed the variance of individual terms in the ilppd, which is defined as follows:

$$p_{WAIC} = \sum_{i=1}^{n} \sum_{t=1}^{T} V_{z,\lambda,w}\,[\log Pois(y_{it}\,|z,\lambda,w)],$$

$$= \sum_{m=1}^{M} \sum_{i=1}^{n} \sum_{t=1}^{T} V_{z,\lambda,w} \left[\log Pois\,\left(y_{it}\middle|\lambda^{(m)}_{z^{(m)}_{ijt}}, w^{(m)}_{z^{(m)}_{ijt}}\right)\right]. \quad (12)$$

The WAIC is constructed as follows:

$$WAIC = -2\,\widehat{ilppd}_y + 2p_{WAIC}. \quad (13)$$

2.3.2The predictive posterior distribution

After the implementation of the model selection step using the criterion presented in the previous section, the posterior predictive distribution (PPD [21] can be used as a graphical tool to assess the adequacy of the selected model. Given the estimation of the model parameters sampling through an MCMC, $(w_{jt}^{(m)}, \lambda_{jt}^{(m)}; m = 1,2,\dots,M)$ and observed infections data $\boldsymbol{y} = (y_{1t}, y_{2t}, \dots, y_{nt})$ for each region i at the time t, the PPD for predicted infections, $y_{it}^*;\ i = 1,2,\dots,n;\ t = 1,2,\dots,T$ of the Poisson mixture model can be defined as follows:

$$Pr(\,y_{it}^*|\boldsymbol{y}) = \int_\lambda \int_z Pois(y_{it}^*\,|\lambda,w,E,z) Pois_{post}(\lambda,w,E,z|y) dz d\lambda, \quad (14)$$

where $Pois_{post}(\lambda,w,E,z|y)$ represents the joint complete posterior distribution. Given samples of the relative risk parameter, $\lambda_{jt}^{(m)}$, and latent variables, $z_{jt}^{(m)}$ are drawn from an MCMC run, the predictive data of a Poisson mixture model can be approximated as

$$y_{it}^* \sim Pois\left(E_{it}\lambda^{(m)}_{z^{(m)}_{ijt}}\right); i = 1,2,\dots,n; t = 1,2,\dots,T. \quad (15)$$

2.4 The model implementation

For each time period, one model is specified. The estimation of the model parameters for each time period is obtained separately by running the Gibbs sampler. For the relative risk parameter λ a non-informative Gamma hyperprior is used with hyperparameters: $\alpha = 0.001$ and $\beta = 0.001$ and for the allocation probabilities w the Dirichlet distribution is considered: $w \sim Dir(w_j, \eta)$ as a prior distribution where $\eta = 1 \forall\, j$. The number of components K is assumed to be fixed but unknown number of components. In other words, a maximum term on the number of components which is $K_{max} = 6$, i. e. $K = 2, 3, 4, 5, K_{max}$ of models for each time period that being fitted for our data. Selecting the level of relative risk for all provinces at each time period t is identified using the WAIC criterion. After that, a graphical checking by the posterior predictive distribution to examine the goodness of fit of the model selected by the criterion is done. The sampler is run for 30000 iterations and the last 15000 iterations were adopted for inference. The problem of the identifiability in model was addressed by imposing artificial constraints on the relative risk parameter such that: $\lambda_1 < \lambda_2 < \ldots < \lambda_K$. The convergence of the posterior distributions is verified by running three different chains for every parameter with different initial values. The convergence of the model is checked using the Gelman-Rubin statistic. For stability, the precision of the posterior mean of parameters is verified by checking the Monte Carlo error. Three different chains were run together to check the convergence for all parameters of the model using the Gelman-Rubin statistic. For checking the precision of the posterior mean for all parameters, the Monte Carlo error is employed for this purpose.

3. Discussion of experimental results

The data of COVID19 infections for the period from March-December 2021 based on a Bayesian spatio-temporal Poisson mixture model is analysed. In the first stage, the results of model selection are presented in Table 1 which illustrates that the provinces take different choices with respect to the levels (components) of the relative risk. These different levels of relative risk suggesting to the heterogeneity in infections through the different time periods. As we can see from Table 1 that two levels of relative risk are enough to accommodate the heterogeneity in the number of infected people in the months: March, April and June, while three levels of relative risk are representing the heterogeneity in infections for all rest of the months. Two levels of relative risk can be interpreted as the more stable for infections in concerning months of March, April and June throughout Iraq which are probably classified as low and high risk. On the other hand, the relative risk is classified as three levels in the rest months that indicate to unstable infection numbers which can be classified as low, medium and high. The distribution of the spatio- temporal areal allocation to the levels (components) of the relative risk over time is given in Figure 2 where each map is allocated to a certain time period. This different classification in levels of relative risk at different time periods can be important for the health organizations in Iraq to diagnose in which month the increasing in levels of risk is. The goodness of fit is also examined by the posterior predictive distribution for the models selected by the WAIC. The 95% prediction interval obtained by the posterior predictive distributions and the mean of observed infections for each province and each time period were drawn as shown in Figure 1 which appears somewhat match. The estimation of the model parameters selected by WAIC for each time period is presented in Table 2. From the same table, in March 2021, it can be seen that about 55% of provinces were assigned to a low level of relative risk of COVID-19 and about 45% were assigned to a high level. In April 2021, 60% were assigned to a low level equal to 0.801 and 40% of provinces were assigned to a high level with a relative risk average equal to 1.779. In May 2021, with an average of 1.878, about 44% of provinces were assigned to a high level of risk. About 60% of provinces also were assigned

to a high level with an average equaling 1.710 in June 2021. In July 2021, most provinces were assigned to a medium level of risk with an average equal to 1.285. From August to December 2021, provinces with high proportions were assigned to a low level of relative risk of COVID-19 with averages: 0.888, 0.988, 0.971, 0.765 and 0.723 respectively.

Table 1: The results of the model selection over the period from March to December 2021.The numbers in bold font represent the smallest value introduced by the WAIC among values of the models fitted to the data.

K	March		April		May		June		July	
	WAIC	p_{WAIC}	WAIC	p_{WAIC}	WAIC	p_{WAIC}	WAIC	p_{WAIC}	WAIC	p_{WAIC}
2	**187.775**	5.443	**183.737**	5.099	187.766	4.644	**179.556**	6.323	185.655	5.258
3	189.056	7.911	185.036	6.971	**187.001**	5.099	181.00	6.788	**184.111**	5.638
4	192.112	9.776	187.555	8.226	188.221	6.222	182.00	7.256	185.665	6.551
5	194.665	9.714	187.906	8.881	189.763	7.889	184.00	9.121	187.298	8.467
6	195.512	10.116	188.099	9.151	190.141	9.031	186.00	9.202	188.712	9.323
K	August		September		October		November		December	
	WAIC	P_{WAIC}	WAIC	P_{WAIC}	WAIC	P_{WAIC}	WAIC	P_{WAIC}	WAIC	P_{WAIC}
2	184.877	4.012	187.332	4.441	185.761	3.996	194.985	5.774	190.985	4.816
3	**183.596**	5.776	**186.665**	6.661	**183.221**	5.744	**193.087**	7.358	**188.074**	5.916
4	185.239	8.161	188.654	8.139	185.873	7.404	204.826	9.225	191.573	8.664
5	187.166	9.674	190.120	9.887	187.111	7.955	210.471	11.176	194.363	11.129
6	188.971	11.017	191.306	10.443	190.344	10.532	213.829	15.851	198.855	13.397

Table 2: Results of the estimation of the model parameters selected by the model selection criterion of all provinces over periods: March-December 2021.

March		April		May		June		July	
$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$
0.553	0.789	0.592	0.801	0.284	0.344	0.388	0.901	0.112	0.436
0.447	1.872	0.408	1.799	0.273	0.952	0.612	1.710	0.611	1.285
---	---			0.443	1.878	---	---	0.277	2.055

August		September		October		November		December	
$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$	$\hat{w}j$	$\hat{\lambda}j$
0.521	0.888	0.544	0.988	0.667	0.971	0.556	0.765	0.502	0.723
0.288	1.389	0.333	1.266	0.112	1.411	0.166	1.208	0.277	1.395
0.191	2.321	0.123	2.117	0.221	2.993	0.278	2.687	0.221	2.916

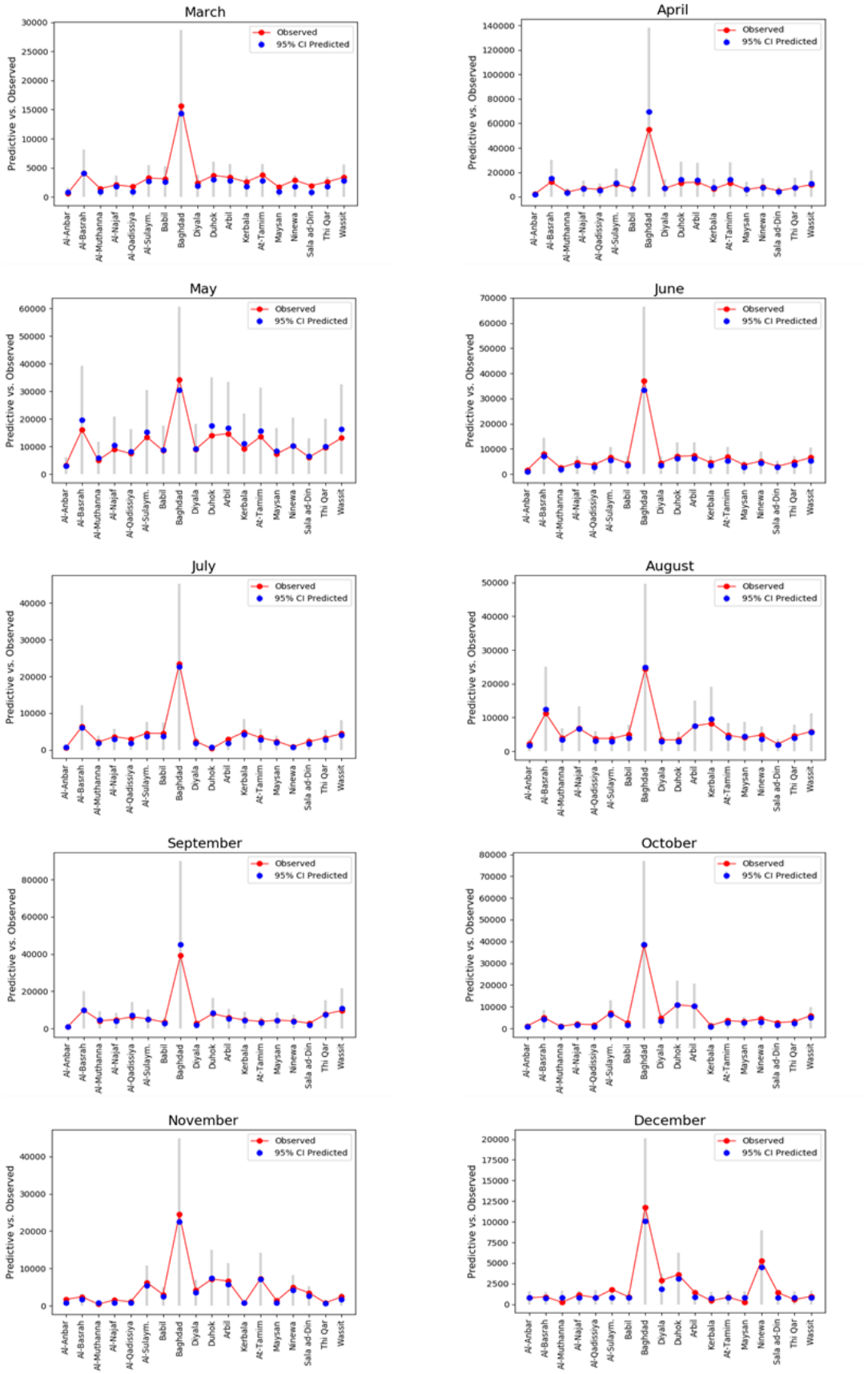

Figure 1: The best spatial predictive performance for the models selected by the WAIC over period from March-December 2021, represented by 95\% posterior predictive intervals of the model.

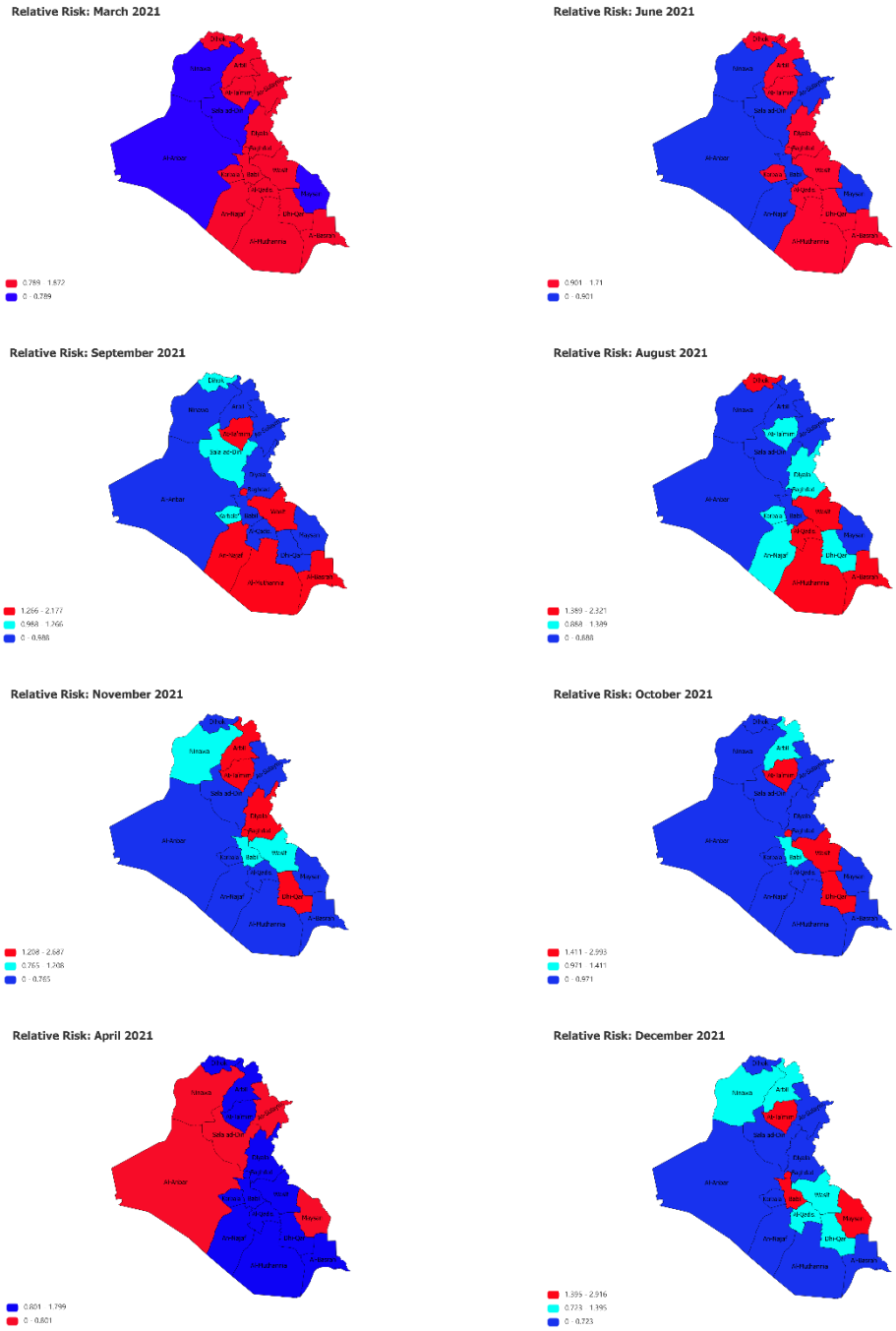

Figure 2: Relative risk–based spatial classification of all provinces on the basis of the best model fits the infections data over the period from March to December 2021.

4. Conclusion

This paper introduced a Bayesian spatio-temporal mixture model that classifies the provinces of Iraq infected by COVID-19 pandemic into levels of relative risk over fixed time periods. In other words, we developed a spatio-temporal mixture model in which the model is based on an unknown fixed number of risk levels (components) and all provinces under study has been assigned to range of relative risk, over different time periods, by means of independent allocation variables. The identification of levels of time-specific spatial relative risk is implemented by using WAIC criterion. Those identified levels of risk can have special importance for governments to review their health systems and also controls diseases. At the beginning of a pandemic, especially in March, April and June the situation in Iraq is classified

into two cases: low and high levels of risk with a high tendency to have a low risk. After that, with increasing the number of infections, the situation in the next months became more complicated due to appear high heterogeneity in the infections, hence, increasing the classification of levels of relative risk. Although, all provinces are tended to have a lower level of risk indicating a stable health case. In comparison, our proposed methodology, which is based on the classification principle is superior to the standardized mortality ratio (SMR) method introduced by [2], as the former is able to identify the high or low levels of the risk of disease in infected regions.

In the development of our methodology, we seek to include as many as covariates in the model in order to obtain a more accurate estimation. Hence, the challenge that can be encountered in this case is to increase the dimensions of the model which needs to estimation method more developed.

References

[1] A. B. Lawson, R. Carroll, C. Faes, R. S. Kirby, M. Aregay, and K. Watjou, "Spatiotemporal multivariate mixture models for Bayesian model selection in disease mapping", Environmetrics, Vol. 28, No. 8, e2465, 2017.

[2] P. Schlattmann and D. Böhning, "Mixture models and disease mapping," Statistics in medicine, Vol. 12, No.19-20, pp. 1943-1950, 1993.

[3] D. Böhning, E. Dietz, and P. Schlattmann, "Space-time mixture modelling of public health data", Statistics in Medicine. Vol. 19, No. 17-18, pp. 2333-2344, 2000.

[4] A. Militino, M. Ugarte, and C. Dean, "The use of mixture models for identifying high risks in disease mapping," Statistics in Medicine, Vol. 20, No. 13, pp. 2035-2049, 2001.

[5] D. Catelan, C. Lagazio, and A. Biggeri. "A hierarchical Bayesian approach to multiple testing in disease mapping", Biometrical Journal. Vol. 52, No. 6, pp. 784-797, 2010.

[6] A. B. Lawson, Bayesian disease mapping: hierarchical modeling in spatial epidemiology, CRC press, 2013.

[7] R. L. Wolpert and K. Ickstadt, "Poisson/gamma random field models for spatial statistics," Biometrika, Vol. 85, No. 2, pp. 251-267, 1998.

[8] S. Watanabe and M. Opper, "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory," Journal of machine learning research, Vol. 11, No. 12, 2010.

[9] "Spatiotemporal multivariate mixture models for Bayesian model selection in disease mapping," Environmetrics, Vol. 28, No. 8, e2465, 2017.

[10] G. J. McLachlan and D. Peel, Finite mixture models, John Wiley & Sons, 2004.

[11] C. Fernández, and P. J. Green, "Modelling spatially correlated data via mixtures: a Bayesian approach". Journal of the royal statistical society: series B (Statistical methodology). Vol. 64, No. 4, pp. 805-826, 2002.

[12] K. C. Flórez, A. Corberán-Vallet, A. Iftimi, and J. D. Bermúdez," A Bayesian unified framework for risk estimation and cluster identification in small area health data analysis". PloS one. Vol. 15, No. 5, e0231935, 2020.

[13] S. K. Kadhem and S. A. Kadhim, "Absorbing Markov chains for analyzing covid-19 infections," Asian-European Journal of Mathematics, Vol. 15, No. 01, 2250008, 2022.

[14] S. K. Kadhem, "Spatial identification of component-based relative risks," Model Assisted Statistics and Applications, Vol. 16, No. 1, pp. 65-72, 2021.

[15] M. Tanner and H. D. Wong, "The Calculation of Posterior Distributions by Data Augmentation", Journal of the American statistical Association, Vol. 82, No. 398, pp. 528-540, 1987.

[16] S. K. Kadhem, P. Hewson, and I. Kaimi, "Using hidden Markov models to model spatial dependence in a network," Australian & New Zealand Journal of Statistics, Vol. 60, No. 4, pp. 423-446, 2018.

[17] S. Fruhwirth-Schnatter, Finite Mixture and Markov Switching Models. Springer Series in Statistics. Springer, New York, 2006.

[18] S. K. Kadhem and S. A. Kadhim, "Widely applicable information criterion for estimating the order in a hidden Markov model," Songklanakarin Journal of Science and Technology, Vol.43, No.3, pp. 824-833, 2020.

[19] B. P. Carlin and T. A. Louis, Bayesian methods for data analysis, CRC Press, 2008.

[20] World Health Organization. Coronavirus disease (covid2019) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/. Accessed July 31, 2020.

[21] Central Statistical Organization of Iraq. Iraq population estimation. http://cosit.gov.iq/ar/?option=com_content&view=article&layout=edit&id=174. Accessed 2019.

[22] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criterion for Bayesian models," Statistics and Computing, Vol. 24, No. 6, pp. 997-1016, 2014.