Study on Conditional Autoregressive Model of Per Capita Grain Possession in Yellow River Delta

Yujian Yang, Huaijun Ruan, Yan Tang, Wenxiang Zhao, Xueqin Tong

HAL Id: hal-01220902

https://hal.inria.fr/hal-01220902

Submitted on 27 Oct 2015

Study on Conditional Autoregressive Model of Per Capita Grain Possession in Yellow River Delta

Yujian Yang,Huaijun Ruan,Yan Tang, Wenxiang Zhao and Xueqin Tong

S & T Information Institute of Shandong Academy of Agricultural Science
Corresponding author, Address: Yujian Yang, Number 202 Gongye North Road, Licheng District of Jinan, 250100, Shandong Province, P. R. China, Tel: +86-531-83179076, Fax:+86-531-83179821, Email:yyjtshkh@126.com

Abstract. In the paper, we reviewed hierarchical statistical modeling about conditional autoregressive models method for spatial data located in Yellow River Delta. Moreover, we also proposed a new method about the evaluation and prediction of per capita grain possession on county-level in the high-efficiency ecological zone. With the support of MCMC approaches and conditional auto-regressive (CAR) Model, we estimated the posterior distribution through iterative sampling among per capita grain possession and the correlated factors. As a useful tool, the spatial bayesian model appeared to gain insights into per capita grain possession on spatio-temporal variability related to the grain sowing area, efficient irrigation area, agriculture machinery conditions. The study results showed that auto correlation characteristics and the posterior density distribution were described by the quantity method with CAR model, correspondingly the posterior mean of observed value and the posterior mean of predicted value of per capita grain possession were developed in Yellow River Delta. The posterior probability was more evident statistical significance on basis of prior information and sample characteristics in credible interval. Apparently, three factors of the grain sowing area, efficient irrigation area and agriculture machinery resulted in complex random effects on per capita grain possession. The results also indicated that bayesian method was not only the more quantitative evaluation, but also the more estimated accuracy for the per capita grain possession on country level in Yellow River Delta.

Keywords: Yellow River Delta；Per capita grain possession; CAR model, Bayesian

1   Introduction

In the recent decades, hierarchical models had drawn the attention of scientists in many fields and were especially suited to studying the spatial variables. The approach to spatial modeling was fairly effective to evaluate the variables, and the merits of solving the flexibility and stability for high-dimensional spatial model or spatio-temporal model were very evident with bayesian hierarchical model. The bayesian estimation indicated significance for this variable whereas maximum likelihood estimates did not, and the bayesian considered the prior information regarded as the prior distribution for the variables, had the highest posterior density. So the posterior

distribution was given the better precision. Recent development of the spatial bayesian hierarchical framework had gained increasing popularity in many domains. The spatial effects can be explicitly expressed in the model by assuming a prior distribution with its parameters specified from the information of spatial neighbors from spatial bayesian models. Based on a statistical point of view, however, the observations were not independent since proximity in space playing a key role in complicated spatio-temporal model. The importance of space and the problem of parameter heterogeneity were emphasized by Martin Feldkirche(2007), Doppelhofer et al. (2004) and Fernandez et al. (2001), In particular, hierarchical bayesian modeling allowed borrowing of strength across similar (say, geographically adjacent) countries, hence improved estimation prediction and mapping of underlying model features driving the data[1,2,3]. CAR model is more suitable for scientific research. Anselin(2002), Haining(2003) on one hand. From the model perspective, on the other hand, the computerization of Spatial weight matrix was emphasized in CAR model, undoubtly, the quantitative degree was improved in variable computerization process. Moreover, distance range attenuation function was used in spatial weight matrix, which was determined by the correlated prior knowledge, and estimation of semi-variogram plot or correlated plots. Apparently, considering of prior knowledge improved the predication accuracy of model variables [4, 5].

The Yellow River Delta was formed on basis of the yellow river alluvial plain and the adjacent coastal regions in Shandong province, including Dongying city, Binzhou city, Hanting district, Shouguang city, Changyi city in Weifang city, Laoling and Qingyun in Dezhou city, Gaoqing country in Zibo city and Laizhou city in Yantai city. The total area is 26500Km2. In 2009, the construction of high-efficiency ecological zone has been officially christened, in 2011, national science and technology demonstration of modern agriculture in Yellow River Delta was established to enforce the total development of agriculture. Due to the bayesian model provided the flexibility to accommodate that possibility, and the flexible nature of the model also made it possible to consider the inclusion of additional spatial (e.g., grain production) variables when appropriate. The study was carried out to analyze the location and interaction of taking the case of per capita grain possession, conditional autoregressive models of per capita grain possession considering of the typical variable heterogeneity and the better prediction were achieved by the bayesian method.

2   Material and Methodology

2.1   Data Acquisition and preprocessing

It was necessary to preprocess the spatial data before Bayesian inference Using Gibbs Sampling (GeoBUGS) analysis, which attributed the model of WINBUGS software. log transformation on per capita grain possession was used to ensure a constant variance. The map was comprised of 19 country units. The data came from

Shandong Statistical Year Book-2008, Chinese county (city) social economic statistical yearbook-2008[6].

Considered of the special data format for GeoBUGS, firstly the transformation from *.shp format was carried out to ensure the following steps in the study. In general, 4 steps were carried out to data transformation and correlated computerization. The first step, the first line contains the key word "map" followed by a colon and an integer, N, where N was the number of distinct areas in the map. The second steps, the import file is a 2 column list giving: (column 1) the numeric ID of the area, this must be a unique integer between 1 and N; the areas should be labelled in the same order as the corresponding data for that area appears in the model. The third step, it begins with a line containing the key word "regions". The fourth step, the final part of the import file gives the co-ordinates of the polygons. However, they are not used by GeoBUGS, so can be arbitrary. Subsequent rows contain a 2-column list of numbers giving the x- and y-coordinates of the poly. The polygon coordinates can be listed either clockwise or anticlockwise. In the process, polygons should be separated by a line containing the key word END[7].

2.2   Spatial Model Framework of Conditional autoregressive model

In CAR model, the expected value of the response variable was regarded as being conditional on the recorded values at all other locations, we may smooth the observed value of per capita grain possession by fitting a random effect on poisson model allowing for spatial correlation, using the intrinsic CAR prior proposed by Besag(1974), York and Mollie (1991) [8]. For the risk evaluation of per capita grain possession, the expected value of the response variable in this model was regarded as being conditional on the recorded values at all other locations. CAR was also a popular hierachical spatial model for use with areal data, or popular auto-gaussian model, the implementation of the CAR model was convenient in hierarchical bayesian settings because of the explicit conditional structure. The model may be written as:

$$O_i \sim \text{Poisson}(\mu_i) \tag{1}$$

$$\log \mu_i = \log p_i + \alpha_0 + \alpha_1 x_i / 10 + b_i \tag{2}$$

where α_0 is an intercept term representing the baseline (log) relative risk of per capita grain possession across the study region, x_i is the covariate in district i, with associated regression coefficient α_1 and b_i is an area-specific random effect capturing the residual or unexplained (log) relative risk of per capita grain possession in area i. We often think of b_i as representing the effect of latent (unobserved) risk factors. The distribution sets the value of b_i equal to zero for areas i that are islands, hence the posterior relative risks depend on the overall baseline rate 0 and the covariate x_i. To allow for spatial dependence between the random effects b_i in nearby areas [9], we may assume a CAR prior for these terms. Intrinsic CAR model given referred to the following description:

$$\phi_i | \phi_{-i} \sim N(\overline{\phi_i}, 1/(\lambda m_i)) \qquad (3)$$

Where, $\phi_{-i} \equiv (\phi_1, ..., \phi_{i-1}, \phi_{i+1}, ..., \phi_p)^T$, m_i is the number of adjacent regions of region i, and $\overline{\phi_i} = m_i^{-1} \sum_{jadj} \phi_j$, the average of the adjacent regions values. This model corresponded to a multivariate normal distribution for $\phi \equiv (\phi_1, ..., \phi_p)^T$ with a less than full rank covariance matrix. This arose due to the translation invariance clearly visible in formular (3), the prior was a pairwise difference prior, identified only up to an additive constant. This impropriety was typically ignored, since the posterior for ϕ will typically emerge as proper even though this prior is not. In order to clarify the parameterisation and the suitable hyperpriors for the parameters of this model, application of proper CAR model required the user to specify unnormalized weights, must always be used, so the new description was chosen an alternative parameterisation from Stern and Cressie (1999), as the following illustrated. $M_{ii} = 1/E_i$ (the inverse of the expected count in area i, $C_{ij} = (E_j/E_i)^{1/2}$ for neighboring areas i, j and 0 otherwise. So we used the dflat() distribution in WINBUGS developed the prior information for the intercept term in a model including CAR random effects[10].

MCMC procedure was now automated within WinBUGS, its spatial package GeoBUGS can be used to carry out bayesian computations, this software packages now routinely used by practicing biostatisticians and spatial workers. The adjacency matrices were edged from the GeoBUGS module including an option to produce a data file containing the adjacency matrix for any map loaded on the system. We edit ed the adjacency map of Yellow River Delta to include these areas as neighbors. In the article, the sum of 80 neighboring areas was given by spatial CAR models to the risk evaluation of per capita grain possession on country-level in Yellow River Delta.

3   Results and Analysis

3.1 Bayesian posterior analysis of CAR Model parameters

There was a large and growing literatures on bayesian analysis and Markov chain Monte Carlo (MCMC) method. MCMC method offered an iterative computational method suitable for solving this problem. We obtained posterior estimation of key fixed effects, smoothed maps of both frailties and spatially varying coefficients. MCMC also provided a mechanism for taking spatially dependent samples from probability distributions in situations, where the usual sampling was difficult, if not impossible. Many auto-models fall into this category, particularly because the normalizing constants for their joint or posterior probability distributions were either too difficult to calculate or analytically intractable. MCMC approaches can be utilized

to estimate the posterior distribution through iterative sampling. The series of results were detailed given in the contexts.

There was no evidence of drift in the study, the mixing was quite good with the full range of each parameter covered in relatively few simulations. The plot also showed the median and 95% credible interval for each parameter. Trace plots can be usefully displayed individually, as shown here. Well-behaved MCMC output was characterized by parallel-chain trace plots showing that each sampler in a set of parallel simulations (started at over-dispersed points in parameter space) has moved freely around a common parameter space, with no sign of being stuck within a localized region, and with no systematic trends. The iteration trace of alpha0, alpha1, tau and the trace plots about sigma parameter were analyzed in the article. The iteration results indicated that the markov chain based gibbs sampler of model parameters presented the convergence, including the two groups of initial values of sigma parameter according to the trace plot after 10000 iterations [11,12]. After that the gibbs sampler, we also developed the correlogram plots the autocorrelations of the chain for each parameters of interests (alpha0, alpha1, sigma and tau) to understand the convergence of the parameters in detail.
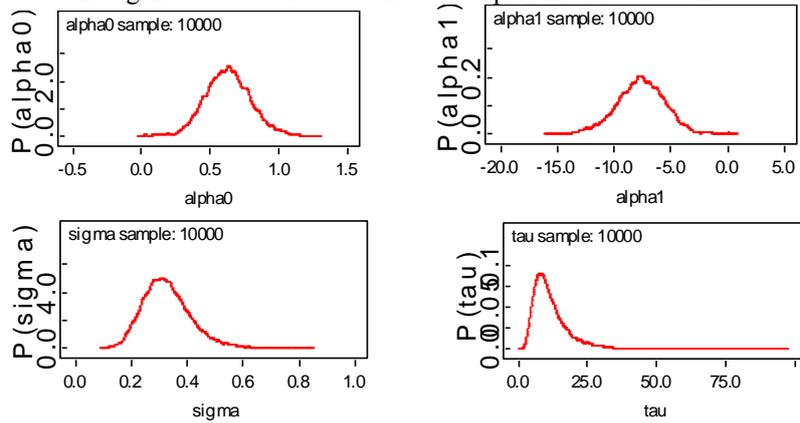

Figure1. Parameters posterior kernel density of CAR Model

Figure1 was produced by computerization and was showed by the densities of sections of the chains for the parameters alpha0, alpha1, sigma and tau, and was produced an approximate visual kernel estimate of the posterior density or probability function. The posterior summary was estimated by the MCMC output: mean, standard deviation, MC (Monte Carlo) error, and selected percentiles (default values are 2.5%, 50%, and 97.5%). It also provided iterations finally kept for estimation. MC error was estimated using the batch mean method, the posterior mean table of variables was showed inTable1.

Table1. Posterior mean table of variables

	sd	MC_error	val2.5pc	median	val97.5pc	start	sample	mean
sigma	0.08675	0.0021	0.1828	0.3174	0.5231	1	10000	0.3259
alpha0	0.1725	0.0072	0.2977	0.624	0.9819	1	10000	0.625
alpha1	2.181	0.0919	-12.15	-7.633	-3.518	1	10000	-7.651
tau	7.334	0.1717	3.662	9.925	29.93	1	10000	11.71

Table 1 summarized the results for the key variables. The density estimated of the posterior distributions for the prior, the bayesian estimation of 4 parameters(alpha0, alpha1, sigma and tau) was 0.625, -7.651, 0.3259 and 11.71, the confidence intervals of the parameters were (0.2977, 0.9819),(-12.15, -3.518),(0.1828, 0.5231), respectively[12].The literature reviews showed that MCMC was an effective method for bayesian computerization on the premise of the Markov convergence, so it was significant for estimation of simulation sampler, the iterations times to reject the influence of the initial values, it was suitable for 10000 iterations to estimate the parameters of bayesian model in the study considering of these factors, correspondingly, the posterior mean plots were given, the high sustainable status of model were resulted from the data transformation, or policy and the micro-structure of market usually caused the results[13].

3.2 The posterior mean of per capita grain possession

Posterior density plots of parameters were estimated from posterior sample using gibbs sampler, the results generally confirmed our previous conclusion, the bayesian analysis of the coefficient alpha0, alpha1, sigma and tau was given independent non informative normal priors while the standard deviation of the random effect, a spatial bayesian hierarchical approach was employed to simulate the relationships of 3 factors to per capita grain possession. In the study, according to the series of random numbers relevant to the uncertainty of per capita grain possession, the results was expressed to estimate the risk, maps of posterior mean of observed value and predicted value for the models in the paper.
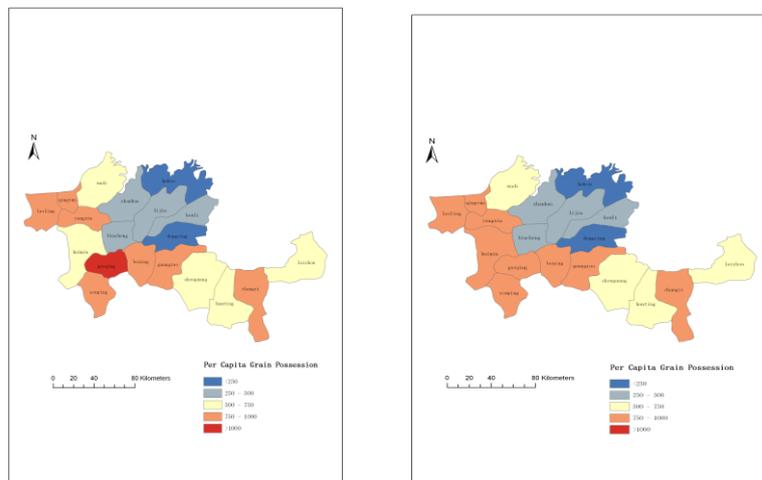

Figure2 The posterior mean of observed value per capita grain possession on country level in Yellow River Delta
Figure3 The posterior mean of predicted value of per capita grain possession on country level in Yellow River Delta

The posterior mean of observed value per capita grain possession observation in Yellow River Delta, respectively. Hekou district and Dongying district (<250 kg per capita grain possession), Zhanhua country, Lijin country, Kenli country and Bincheng district(250 kg -500 kg per capita grain possession), Huimin country, Wudi country, Shouguang city, Hanting district and Laizhou city (500 kg -750 kg per capita grain possession),Qingyun country, Laoling city, Yangxin country, Zouping city, Boxing country, Guangrao country Changyi city (750kg -1000kg per capita grain possession), Gaoqing country(>1000 kg per capita grain possession), as illustrated in Figure2.

The posterior mean of predicted value about per capita grain possession was computerized on country level in Yellow River Delta, respectively. Hekou district and Dongying district (<250kg per capita grain possession), Zhanhua country, Lijin country, Kenli country and Bincheng district(250kg-500kg per capita grain possession), Wudi country, Shouguang city, Hanting district and Laizhou city (500kg -750kg per capita grain possession), Huimin country, Qingyun country, Laoling city, Yangxin country, Zouping city, Boxing country, Guangrao country, Changyi city and Gaoqing country (750kg -1000kg per capita grain possession), as shown in Figure 3.

These results indicated that the observed values were consistent with the predicted values from the posterior predictive distribution and the spatial maps of CAR model, and the parameters responded to the characteristics of model computerization. three types of the grain sowing area, efficient irrigation area and agriculture machinery resulted in complex random effects on per capita grain possession was given to estimate the CAR model, the proposed model gave a quantitative result. Our results generally agreed with previous research findings in the literature, referred to GWR results in another paper[14].

4   Conclusions and Discussion

As mentioned above, we had shown the application of spatial conditionally autoregressive model to examine the random effects on per grain possession, and we had also estimated per capita grain possession a set of 19 heterogeneous countries in Yellow River Delta. The Bayesian hierarchical approach distinguished itself by being able to explicitly model the spatial association of grain production data, the bayesian results showed the observed value and predicted value of per capita grain possession can be fitted using MCMC based on CAR model, the flexibility and stability estimation offered by model established were improved. 4 parameters in the model (including alpha0, alpha1, sigma and tau) showed that the posterior probability was more evident statistical significance on basis of prior information and sample characteristics in confidence interval. So CAR of per capita grain possession on country–level in Yellow River Delta was not only the more estimated accuracy, but also the more quantitative evaluation depending on bayesian method. The Bayesian model provided the flexibility to accommodate that possibility. The flexible nature of the model also made it possible to consider the inclusion of additional spatial (e.g., soil or topographic) variables when appropriate.

References

1   Martin Feldkircher. A Spatial CAR Model applied to a Cross-Country Growth Regression. Institute for Advanced Studies Applied Empirical Homework,2007
2   Doppelhofer, G., Miller, R. I., and Sala-i Martin, X. (2004). Determinants of long-term growth: A bayesian averaging of classical estimates (bace) approach. American Economic Review, 94:813–835
3   Fernandez, C., Ley, E., and Steel, M. F. (2001). Model uncertainty in cross-country growth regressions. Journal of Applied Econometrics, 16:563–576.
4   Anselin L under the hood:issues in the specification and interpretation of spatial regression models,agricultural economics, 2002,17(3),247-267
5   Robert Haining.Spatial data analysis theory and practice, Cambridge University Press,1993,328-350
6   Statistical Bureau of Shandong Province, Shandong Statistical Yearbook 2008 [M], Beijing: China Statistics Press(in Chinese)
7   GeoBUGS User Manual, Version 3.2.1, March 2011
8   Joanna Horabik, Zbigniew Nahorski, Statistical Spatial Modeling of Gridded Air Pollution Data,101-105, 2nd International Workshop on Uncertainty in Greenhouse Gas Inventories, International Institute for Applied Systems Analysis   A-2361 Laxenburg, Austria ,27–28 September 2007
9   Ali Arab, Mevin B. Hooten, Christopher K. Wikle. Hierarchical Spatial Models, June 2006, Department of Statistics,University of Missouri-Columbia.
10 Pingping Jiang Æ Zhuoqiong He Æ Newell R. Kitchen Æ Kenneth A. Sudduth. Bayesian analysis of within-field variability of corn yield using a spatial hierarchical model Precision Agric (2009) 10:111–127
11 Martin D. King, Fernando Calamante, Chris A. Clark and David G. Gadian. Markov Chain Monte Carlo Random Eects Modeling in Magnetic Resonance Image Processing Using the BRugs Interface to WinBUGS. Journal of Statistical Software,2011,14(2):1-20
12 Zhu Huiming,Lin Jing. The economical model of Bayesian computation. Beijing: Science Pressing,2009(in Chinese)
13 Haijun Ma, Beth A. Virinig, and Bradley P. Crain. Spatial Methods in Geographic Administrative Data Analysis. August 16,2005
14 Yujian Yang, Xueqin Tong, Jianhua Zhu.A geographically weighted model of the regression between grain production and typical factors for the Yellow River Delta. Mathematical and Computer Modelling. 2013, 58: 582–587