

STOCHASTIC GROWTH MODEL FOR SPATIAL CLUSTER BIRTH AND DEATH PROCESS WITH MIGRATION

Ro'fah Nur Rachmawati and Rokhana Dwi Bekti

Department of Statistics, School of Computer Science, Bina Nusantara University, Indonesia

Received 2013-01-08, Revised 2013-02-18; Accepted 2013-05-09

ABSTRACT

In real life, we can encounter the cluster birth and death process with migration, such as the birth of new individuals and disease in certain place and time. These phenomena can be represented by stochastic model that are considered to be more representatives in capturing natural indications involving the concepts of space and time. In this study, we construct a stochastic model to estimate the population growth in Bogor, West Java, using spatial cluster birth and death methods. It is considered cluster birth and death process, including the effect of migration, with birth and death are Poisson process. In clustering process we use Spatial 'K' luster Analysis by Tree Edge Removal (SKATER), which is view to considering the influence of the location or the relationship between locations. It has been formulated the stochastic growth model for spatial cluster birth and death process with migration and has been applied in predicting population growth in Bogor, West Java. The estimator that we construct can estimate the population growth with the value of birth rate is not too small.

Keywords: Poisson Process, Cluster Birth and Death Process, SKATER

1. INTRODUCTION

Stochastic processes can be grouped into two, namely discrete time stochastic processes and continuous time stochastic processes. One particular form of stochastic process with discrete space and continuous time is Poisson process that counts the number of events that occur at specified intervals. The development of the Poisson process is the process of birth, which is the number of counts that occurred at a specific time interval whose value depends on the number of events in the previous time interval. The birth process that has been widely discussed, namely pure birth process that can be seen in (Taylor and Karlin, 1984; Papoulis, 1984; Cox and Miller, 1987). A birth and death process is a stochastic process in which jumps from a particular state, are only allowed to the neighboring states. This property considerably simplifies the mathematical analysis, but the process remains applicable to numerous real-world systems.

The very beginning of population growth modeling, it had been clear that a most refined analysis must take into account the role of stochastic factors in the evolution of population. Mathematical field inspired by biology is the theory of birth and death process. Novozhilov *et al.* (2006) discuss application of the theory of birth and death processes in biology problems, primarily, those of evolutionary genomics. On the other studies, quasi stationary distribution for birth and death processes has been studied by (Cavender, 1978; Pollet, 1988; Doorn 1991; Schoutens, 2000; Clancy and Pollet, 2003). Calculating quasi stationary distribution for a linear birth and death process by two different methods has been studied by Liu and Zhang (2011).

Births and death is one of benchmarks for assessing the extent of the community welfare achievement in the health development. In addition, it use for monitoring the health situation, the input of count population

Corresponding Author: Ro'fah Nur Rachmawati, Department of Statistics, School of Computer Science, Bina Nusantara University, Indonesia

estimation and to identify groups of people who have a death high risk.

According to the data from Ministry of Health Indonesia, the rank of crude birth rate in Indonesia 2010 in the ASEAN was 5. It was 20 births per 1000 population and the death mortality rate was 30 deaths per 1000 live births. The number of births child in West Java Province 2007 was 822,481 and the number of born child which death was 2,575. In Bogor, West Java Province, there are 88,633 children born which life and the 210 child born was death.

However, the birth process does not adequately describe the real phenomenon that involves many elements of change in the amount of a population. Changes in the population were influenced by death, immigration and emigration effects. In addition, many facts found that events occur simultaneously in specific space and time, for example the birth of new individuals, disease in certain place and time, these situations are a picture of cluster birth process.

The methods for clustering birth process which regard on space and time are spatial cluster. Spatial method is a method to get information of observations influenced by space or location effect (Anselin, 1988). So the spatial cluster which build clustering of region by considering the location of the regions. This method was used because there are many factors which influence of birth and death, such as economics, health facility, education and others. The influence factors of every individual are different, but there are similarities some individuals which live close together or in the close locations.

The cluster method which used was Spatial 'K'luster Analysis by Tree Edge Removal (SKATER). The algorithm used was a strategy for transforming the regionalization problem into a graph partitioning problem (Assuncao *et al.*, 2006). Other spatial cluster methods are Tango's Index and Kulldorf spatial scan statistics (Tango, 2010), Moran's I and Geary's I (Lee and Wong, 2001). These methods were also used in disease cases, such as Brooker *et al.* (2004) and Cox and Miller (1987), which was clustering malaria disease in Western Kenya and by Moreno *et al.* (2010) which was clustering the epidemic Cholera.

Discussion about cluster birth and death with the influence of immigration and emigration already presented (Rachmawati, 2013), but the clustering process using SKATER method has not been discussed and the formula has not been applied to the original data. So this study aims to construct stochastic growth model, especially in the process of cluster birth and death with

the influence of immigration and emigration and apply this model to the population growth in Bogor, West Java Province, Indonesia, with the clustering process use SKATER method.

2. MATERIALS AND METHODS

2.1. Assumptions: Construction of Stochastic Growth Model

Let $x(t, t + h)$ is a random variable that specifies the number of individuals at the time $(t, t + h)$ and satisfy the following assumptions.

$N(t, t + h)$ states the number of clusters at time $(t, t + h)$ and $N(t, t + h)$ is a Poisson process has a mean value $(\lambda + \theta)h$, with λ and $\theta = \theta_1 - \theta_2$ consecutive states average birth and net migration (different between the rate of immigration (θ_1) and emigration (θ_2)).

Each cluster has a random number X_i , which shows the number of events occurring in i -th cluster. Random variables X_i are independent and identically distribution and independent from the random variables $N(t, t + h)$.

Let $\{\lambda_n = \eta\lambda + \theta\}$ and $\{\mu_n = \eta\mu\}$ is a sequence of positive numbers that state the rate of birth and death with the effects of net migration, the initial state is $M(0) = i$ and $M(t) = E[X(t)]$.

Determination of $M(t + h)$ by using conditional expectation value, by requiring $X(t)$ Equation 1:

$$M(t+h) = E[E[X(t+h) | X(t)]] \tag{1}$$

Thus, in the interval $(t, t + h)$, possibility of events that occur in the process of cluster birth and death with the effect of migration are.

There is one or more new individuals are born with probability:

$$\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] X(t) \lambda h + o(h)$$

There is one or more individuals who migrate with probability:

$$\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] \theta h + o(h)$$

There is one or more individuals who die with probability:

$$\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] X(t) \mu h + o(h)$$

There is no single individual is born, migrate or die with probability:

$$1 - \frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)(\lambda + \mu) + \theta]h + o(h)$$

Suppose given a value $X(t)$, then the number of individuals at the time $t + h$ are:

$$X(t + h) = X(t) + \sum_{i=1}^{N(t)} X_i$$

with probability:

$$\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)\lambda + \theta]h + o(h)$$

$$X(t + h) = X(t) - \sum_{i=1}^{N(t)} X_i$$

with probability:

$$\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)\mu]h + o(h)$$

$X(t+h) = X(t)$ is no single individual is born, migrate or die with probability:

$$1 - \frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)(\lambda + \mu) + \theta]h + o(h).$$

2.2. Clustering Process by SKATER

Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) use algorithm which is a strategy for transforming the regionalization problem into a graph partitioning problem. SKATER is done by two steps. Reis *et al.* (2007) explain these steps. First, it creates a Minimal Spanning Tree (MST) from the graph representation for the neighborhood structure of the geographic entities. The cost of an edge represents the similarity of the entities attributes, defined as the Euclidean squared distance between them. The MST represents a statistical summary of the neighborhood graph based on the entities attributes. The second step, SKATER performs a recursive partitioning of the MST.

Building the MST based on Prim's algorithm has been shown in Assuncao *et al.* (2006). These algorithms

give a connectivity graph for a set of vertices and a set of edges on each location. After step on MST partitioning, performs a recursive partitioning of the MST to get contiguous clusters. It considers explicitly the clusters internal homogeneity.

3. RESULTS

3.1. Theorem: Stochastic Growth Model for Cluster Birth and Death Process with Migration

Suppose $X(t, t + h)$ is a random variable that specifies the number of individuals at the time $(t, t + h)$ and satisfy assumptions (i) and (ii), then Equation 1:

$$M(t) = \frac{\theta}{\lambda - \mu} \left[e^{E[X_i] (\lambda - \mu)t} - 1 \right] + ie^{E[X_i] (\lambda - \mu)t} \tag{2}$$

is growth model of cluster birth and death process with the effect of migration.

Proof

We obtain:

$$\begin{aligned} & E[E[X(t+h) | X(t)]] \\ &= \left[X(t) + \sum_{i=1}^{N(t)} X_i \right] \left[\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)\lambda + \theta]h \right] + \\ & \left[X(t) - \sum_{i=1}^{N(t)} X_i \right] \left[\frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)\mu]h \right] + \\ & X(t) \left[1 - \frac{1}{\sum_{i=1}^{N(t)} X_i} E[X_i] [X(t)(\lambda + \mu) + \theta]h \right] + o(h) \end{aligned}$$

Or we can simplify above Equation 3:

$$\begin{aligned} & E[E[X(t+h) | X(t)]] \\ &= \left[E[X_i] X(t)(\lambda - \mu) + E[X_i] \theta \right]h + X(t) + o(h). \end{aligned} \tag{3}$$

From Equation 1, so (3) can be written as:

$$\frac{M(t+h) - M(t)}{h} = E[X_i] M(t)(\lambda - \mu) + E[X_i] \theta + \frac{o(h)}{h}$$

By taking $h \rightarrow 0$, then $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$ so the above equation can be written as:

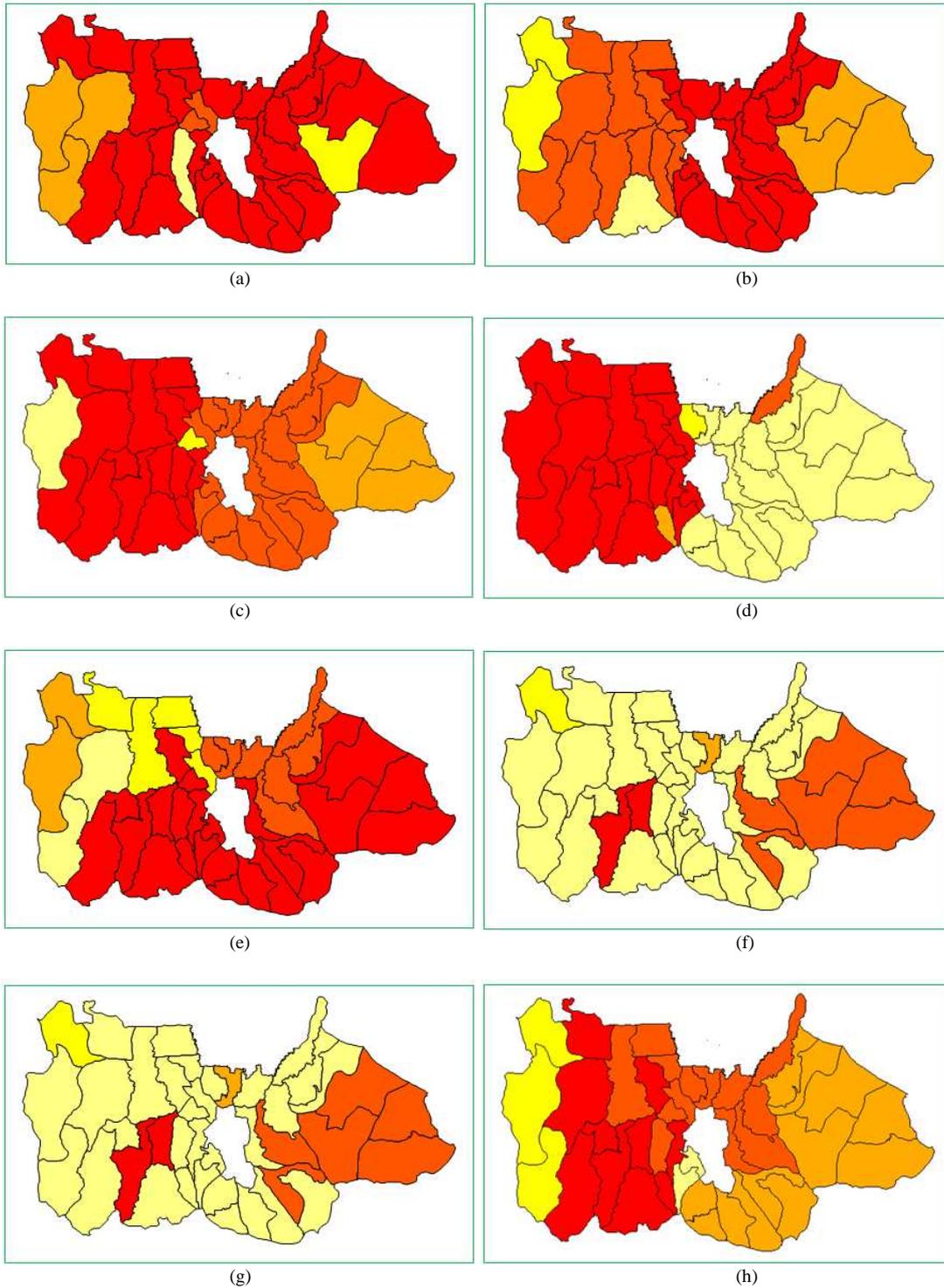


Fig. 2. (a) Bogor, West Java in 2001, (b) Bogor, West Java in 2003, (c) Bogor, West Java in 2005, (d) Bogor, West Java in 2006, (e) Bogor, West Java in 2007, (f) Bogor, West Java in 2008, (g) Bogor, West Java in 2009, (h) Bogor, West Java in 2010

Table 1. Estimation value of population growth

Year	Population growth/day	Estimation of population growth/day
2001	119.09	125.36
2003	75.36	61.54
2005	148.04	187.80
2006	182.71	172.88
2007	107.78	91.19
2008	8.19	-10.18
2009	3.69	-15.32
2010	1.00	-18.00

Table 2. Parameters value

Year	Rate of birth for each district/day (λ)	Rate of death for each district/day (μ)	Rate of migration for each district/day (θ)
2001	1.87914	0.68423	0.00605
2003	1.28924	0.52705	0.00381
2005	2.57390	0.56562	0.00464
2006	1.29658	0.48329	0.01029
2007	1.19459	0.66555	0.00593
2008	0.11308	0.04795	0.00038
2009	0.13178	0.06116	0.00006
2010	0.07479	0.04336	-0.00002

In **Fig. 2a**, from 35 districts can be clustered to five group locations. The first group consists of 28 districts which located in the middle of Bogor. The second group was 2 districts which located in the middle too. The third, four and five groups were 3 districts, 1 district and 1 district. This clustering showed the characteristic distribution of the birth and death ratio from population of each district, the ratio of health workers, hospitals and the percentage of poor families in Bogor 2001. The same interpretation can be done for the characteristics in 2003-2010. In 2010, 40 districts can be clustered to five groups. Each cluster consists of 10 districts, 13 districts, 12 districts, 3 districts and 2 districts. From our stochastic model in (2) we get the estimation value for population growth as in **Table 1**. The parameters are in **Table 2**.

4. DISCUSSION

We can notify in 2001, 2003, 2006 and 2007 we get value of $\lambda > 1$ person per day and we get the estimation value close to the real population growth. But in 2005 the birth rate value, λ , is more than 2 individuals each day. The obtained estimation value using Equation 2 is not quite accurate, this is because the property of exponential function which the resulting population growth will increase more rapidly than the birth rate itself. In the other hand, in 2008-2010 we have $\lambda < 1$

person per day consequently we have the estimation value is not close enough to the real population growth. This is because, for a small value of birth rate for each district/day, $\lambda < 1$, models such as Equation 2 has a value with a slow acceleration, even it is an exponential growth trend.

The results of clustering which shown in **Fig. 2** was perform by SKATER. The steps include establish a Minimum Spanning Tree (MST) Generation and determining the MST Partitioning. This method was build number of group or cluster for difference number of districts as described previously. **Figure 2a-2h** determined that there were five numbers of clusters.

The clustering showed the characteristic distribution of the birth and death ratio from population of each district, the ratio of health workers, hospitals and the percentage of poor families in Bogor 2001-2010. In **Fig. 2a-2h** has shown that some districts which in the same cluster were close together. It means that spatial pattern can be seen through this figure. The characteristics some districts were interrelated with other districts which close together.

The result was appropriate with the first law of geography proposed by Anselin (1988) states that all things are interconnected with each other, but something close more influence than something far away. The SKATER have good performance to support estimation value of population (**Table 1**) which be affected by many factors and include locations or spatial factor. It can be seen from the value of the estimation value close to the real population growth.

For further research, we can expand the scope of our models in Equation 2, by adding the growth rate of the female population, because population growth is also affected by the number of females. We can also apply this formula to model the growth of a disease in an area.

5. CONCLUSION

The construction of stochastic model of cluster birth and death process with the effect of migration can be done using the properties of Poisson process and some mathematical manipulations. By using selected initial values, our model can give very well estimation for population growth with the value of birth rate for each district per day is not too small.

The SKATER have good performance to support estimation value of population which be affected by many factors and include locations or spatial factor. By determined that there were five numbers of clusters, it

perform the clustering of districts which in the same cluster were close together.

6. REFERENCES

- Anselin, L., 1988. Spatial Econometrics: Methods and Models. 1st Edn., Springer, ISBN-10: 9024737354, pp: 304.
- Assuncao, R.M., M.C. Neves, G. Camaras and C.D.A.C. Freitas, 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *Int. J. Geographical Inform. Sci.*, 20: 797-811. DOI: 10.1080/13658810600665111
- Brooker, S., S. Clarke, J.K. Njagi, S. Polack and B. Mugo *et al.*, 2004. Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Tropical Med. Int. Health*, 9: 757-766. PMID: 15228485
- Cavender, J.A., 1978. Quasi-stationary distributions of birth-and-death processes. *Adv. Applied Prob.*, 10: 570-586.
- Clancy, D. and P.K. Pollet, 2003. A note on quasi-stationary distributions of birth-death processes and the SIS logistic epidemic. *J. Applied Prob.*, 40: 821-825.
- Cox, D.R. and H.D. Miller, 1987. *The Theory of Stochastic Processes*. 1st Edn., Chapman and Hall, pp: 398.
- Doorn, E.A.V., 1991. Quasi-Stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. Applied Prob.*, 23: 683-700.
- Lee, J. and D.W.S. Wong, 2001. *Statistical Analysis with ArcView GIS*. 1st Edn., John Wiley and Sons, New York, ISBN-10: 047143776X, pp: 208.
- Liu, W. and H. Zhang, 2011. Quasi-stationary distributions in linear birth and death processes. *J. Math. Res.*, 3: 27-32.
- Moreno, D.R., M. Pascual, M. Emch and M. Yunus, 2010. Spatial clustering in the spatio-temporal dynamics of endemic cholera. *BMC Infect. Dis.*, 10: 51-51. DOI: 10.1186/1471-2334-10-51
- Novozhilov, A.S., G.P. Karev and E.V. Koonin, 2006. Biological applications of the theory of birth-and-death processes. *Brief Bioinform.*, 7: 70-85. PMID: 16761366
- Papoulis, A., 1984. *Probability, Random Variables and Stochastic Processes*. 2nd Edn., Mc-Graw Hill, ISBN-10: 0070484686, pp: 576.
- Pollet, P.K., 1988. Reversibility, invariance and μ -invariance. *Adv. Applied Prob.*, 20: 600-621.
- Rachmawati, R.N., 2013. Mathematical growth model of cluster birth and death process with immigration and emigration. *Int. J. Applied Math. Statist.*, 36: 0973-7545.
- Reis, I.A., G.R. Camara, R. Assuncao and A.M.V. Monteiro, 2007. Data-Aware clustering for geosensor networks data collection. *Proceedings of the Anais XIII Simposio Brasileiro de Sensoriamento Remoto, (BSR' 07) Florianópolis, Brasil*, pp: 6059-6066.
- Schoutens, W., 2000. Birth and death processes, orthogonal polynomials and limiting conditional distributions. *Math. Sci.*, 25: 87-93.
- Tango, T., 2010. *Statistical Methods for Disease Clustering*. 1st Edn., Springer, ISBN-10: 1441915729, pp: 247.
- Taylor, H.M. and S. Karlin, 1984. *An Introduction to Stochastic Modeling*. 3rd Edn., Academic Press, ISBN-10: 0126848807, pp: 399.