A Cellular Automata Modeling for Visualizing and Predicting Spreading Patterns of Dengue Fever

Article *in* TELKOMNIKA Indonesian Journal of Electrical Engineering · December 2016

DOI: 10.12928/TELKOMNIKA.v13i2.xxxx

3 authors:

Puspa Eosina
Universitas Ibn Khaldun Bogor
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Taufik Djatna
Bogor Agricultural University
81 PUBLICATIONS **31** CITATIONS
SEE PROFILE

Helda Khusun
University of Indonesia
7 PUBLICATIONS **54** CITATIONS
SEE PROFILE

TELKOMNIKA, Vol.xx, No.x, June 20xx, pp. 125~132
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v13i2.xxxx

A Cellular Automata Modeling for Visualizing and Predicting Spreading Patterns of Dengue Fever

Puspa Eosina*[1], Taufik Djatna*[2], Helda Khusun[3]
[1]Department of Computer Science, Bogor Agriculture University, Indonesia
(Tel: +62-812-809-83487; E-mail: puspa_eos@apps.ipb.ac.id)
[2]Department of Agro-industrial Technology, Bogor Agriculture University, Indonesia
(Tel: +62-813-197-26435; E-mail: taufikdjatna@ipb.ac.id)
[3]SEAMEO-TROPMED Regional Center for Community Nutrition, University of Indonesia
(Tel: +62-812-858-6856; E-mail: hkhusun@seameo-recfon.org)

Abstract

A Cellular Automata (CA) model is used for visualizing and predicting spreading pattern of the disease. The main problem of this model is how to find a function that represents an update rule that changes the state of a cell in time steps affected by neighborhood. This research aims to develop visualization and prediction model of the spreading patterns of Dengue Hemorrhagic Fever. The contribution of our study is to introduce a new approach in defining a probabilistic function that represens CA transmission rule by employing Von Neumann neighborhood and the Hidden Markov Model (HMM). In this study, we only considered an infective state which dedicated particular attention to the spatial distribution of infected areas. We devided infected data into four catagories and change the definition of a cell as an area. *The evaluation was conducted by comparing the results of the proposed model to that of one yielded by a Susceptible-Infected-Recovered (SIR) model. The evaluation result showed that the CA model was capable of generating patterns that similar to the patterns generated by SIR models with a similarities value of 0.95.*

Keywords*: Cellular Automata, Dengue Fever, HMM, Neighborhood, SIR*

1. Introduction

Modeling is a simplification of a real problem, aiming to study and understand the phenomena in the real world. In epidemiology, system modeling approach is commonly used for viewing the epidemic process [1]. Most of the models for epidemics simulations are based on Ordinary Differential Equations (ODE) or statistical model [2-4]. Moreover, visualization is required as the first step in epidemiological analysis to understand the spatial characteristics of a dataset [2]. Pfeiffer (2008) also mentioned that visualization is needed for identifying the epidemiology of disease pattern in a given geographical area, predicting the spreading pattern of disease in the next period, and creating awareness to the target stakeholders based on the prediction results, hence help clinical management of disease. Unfortunately, ODE or statistical models are unable to elaborate spatial patterns and interactions such as in visualizating and predicting spreading disease [3].

To overcome these limitations, researchers used Celluar Automata (CA) models for involving time and space in epidemic process analysis [5]. Some studies has been conducted such as developing a mathematical model of disease spread and its simulation using Cellular Automata (CA) [6], analyzing some scenarios of disease spread [7], applying the CA approach to the Susceptible-Infective-Recovered (SIR) model of disease spread by considering birth and death factors and the changes of rules for each state in the dynamic CA [8], and analyzing the complex spatiotemporal patterns observed in transmission of vector infectious disease [9].

Basically, CA is one of the dynamic system approaches that implementing discretization of time and space [3, 5, 10]. CA consists of cells, called cellular space, a local connection of to other cells, and boundary conditions [3]. Each cell, representing a state, can change at every time-step using local transmission rules which will generate a new state based on the previous

Received February xx, 2014; Revised May xx, 20xx; Accepted June xx, 20xx

state of the cell and its neighborhood. Therefore, the concept of neighborhoods is very important. Santos (2011) showed the effects of neighborhood structures on diseases spreading by using the susceptible-infected (SI) epidemics CA-model [5]. Moreover, Hagoort (2008) described the rule of neighborhood in determining the model interacts [11].

The other important aspect that determines the accuracy of CA model is the trasmition rule *f.* This rule was able to be represented as a deterministic or probabilistic function [9-10]. Many methods to find function *f* as rule of the CA model have been introduced such as using Markov Chain [12], the differential equations of the classical model [13], and the Genetic Algorithm [14]. In this research we used the Hidden Markov Model (HMM) to find a probabilistic function that represented the CA transmission rule, which has not been used by researchers yet. HMM is a probabilistic model that is suitable for solving the problem related to the data sequential-temporal [15]. To show the effectiveness of the proposed model, we implemented this approach to the Dengue Fever case.

The reason of using the Dengue Fever case is because it includes as one of the deadly and infectious pandemic diseases in Indonesia. This disease, also called Dengue Hemorrhagic Fever (DHF), is caused by the Dengue virus and is transmitted by the *Aedes aegypti* mosquito as a vector. Several studies related to the monitoring DHF in Indonesia have been conducted, such as the studies that aimed to see the trend of dengue outbreak in the future by Saragi (2011) and Octora (2010) [16-17]. Saragi (2011) used the Time Series method for showing the trend of dengue outbreak [16]. The study predicted the number of dengue fever patients for next four years based on DHF patient data in the province of North Sumatra from 2005 to 2009. Octora (2010) compared the Autoregressive Integrated Moving Average (ARIMA) and the Winter approach to predict the number of DHF cases in the next six months [17]. This research used DHF cases data from Surabaya from January 2005 - June 2010. In this study, Octora (2010) applied four models of the Winter method and three models of the ARIMA method.

This paper explained how to develop a spreading pattern model of DHF on CA model that was used for visualizing and predicting spreading pattern of DHF. In this study, we especially focused on determining a probabilistic function using HMM. We used dataset from a limited area such as West Bogor in the period of 2013 and defined the state criteria from these dataset. Moreover, we only considered an invective state which dedicated particular attention to the spatial distribution of infected areas.The evaluation was conducted by comparing the results of the proposed model to that of one yielded by the SIR method, as a classical approach.

2. Research Method

To achieve the research objective, several stages were done, including: collecting datasets, defining the model CA, constructing the data, predicting the spread of disease using an obtained model, and evaluating the model.

2.1. Dataset

In collecting the datasets, we did some steps as follows: identification of geographical study area, conducting field study for data collection, deciding sample used in this research, determining the source of the data. We decided to use dataset collected from Dinas Kesehatan Kota Bogor (DKK-Bogor). The data was collected using an interview technique. We did an interview with the DKK-Bogor Data Officer on July 16, 2014. Table 1 show the dataset that contains the Dengue Fever cases that occurred in West Bogor in 2013.

2.2. Defining of CA Model

A Cellular Automata (CA) is a discrete model consisting of points or identical cells that each in one certain state, from a set of limited state values at the time. The State of a cell changes according to a local transition rule at the next time in time-step [3], [18]. Those cells are arranged uniformly in cellular space that can be one-dimensional, two-dimensional or three-dimensional. The state condition of a one cell at the next time, *t+1*, depens on the states of the other cells surrounding, called its neighborhood, at the time, *t*. Mathematically the CA model is defined as a 4-uplet (C, S, V, *f*). C represents a cellular space. S represents a set of possible state values for each cell in the cellular space. V is a set of neighborhoods around a focus cell. Function *f* defines a local transition function that represents an update rule for each state change of each cell [6]. There are four steps for defining the CA model, such as: defining a

cellular space, defining neighborhood used in a cellular space, defining the criteria of the possible state values, and finding some probabilistic functions *f* that represent the CA rule. Function *f* is required to obtain a spreading pattern of Dengue Fever.

Table 1 Number of Dengue Cases in West Bogor in 2013
(Source: Dinas Kesehatan Kota Bogor)

Id of cell	Region	Number of DHF cases per period											
		1	2	3	4	5	6	7	8	9	10	11	12
1.	Situ Gede	1	0	0	0	0	1	0	0	0	0	0	0
2.	Bubulak	2	1	4	0	1	0	3	0	0	0	0	0
3.	Curug	7	2	2	3	3	0	1	0	1	0	3	3
4.	Curug Mekar	1	0	3	0	1	0	0	0	0	0	2	0
5.	Balumbang Jaya	0	0	0	0	1	0	0	0	0	0	0	3
6.	Sindang Barang	2	2	4	0	1	2	6	0	2	2	0	0
7.	Semplak	2	0	1	0	1	0	0	0	1	0	3	1
8.	Cilendek Timur	5	2	0	0	1	1	4	0	0	0	1	0
9.	Margajaya	1	0	0	2	0	1	0	0	0	0	0	0
10.	Menteng	2	1	2	1	0	2	6	0	0	1	2	0
11.	Cilendek Barat	1	4	3	0	2	1	4	0	4	0	2	0
12.	Pasir Jaya	1	0	0	1	1	0	0	0	0	0	0	0
13.	Gunung Batu	7	3	1	2	0	1	4	0	1	0	1	1
14.	Loji	0	1	1	0	1	0	0	0	0	0	6	0
15.	Pasir Mulya	1	1	0	0	0	0	2	0	0	0	0	0
16.	Pasir Kuda	2	3	1	1	0	0	1	0	1	0	2	1

In this research, we defined 16 cells in two-dimensional cellular space (Figure 1.), which represent 16 regions in West Bogor (Table 1). Each cell represents a region according to *id of cell* listed in Table 1; for instance, the first cell in Figure 1 represents a region of Situ Gede (a region with id 1 in Table 1). Each cell defines ununiformed objects and describes the number of dengue cases that occurred in the region. The number of cell in the cellular spaces actually has not always to be the same as the number of the observed regions. For instance, we are able to define 20 or 25 cellular spaces for the 16 observed regions by adding the definition of boundary regions (the regions which are not included into the 16 observed regions) [3]. In addition, in this study, we assumed null boundary conditions for the proposed model. We also used 4-neighborhoods from Von Neumann, a collection of five cells in which the middle cell is a focus of attention as shown in Figure 2 [6]. The remaining cells are cells that affect the state change of a cell in subsequent periods. The research that related to the state changes of a cell in two-dimensional has performed by Djatna (2008) [19]. His idea is a database state changes were co-related to the shape change of the geometry each pair of attributes data which lead to compute two-dimensional rules affecting to the target attributes. In the proposed model, we defined a state changes based on data content on location. First, we defined the categories value in four categories and setting color for each category. Next, the state changes seen as a sell color change in cellular space. In HMM the state chages described as the state transition diagram. The four colors and its criteria of states are shown in Table 2. The next step is how to determine a Function *f* that represents the CA Rule based on parameters that have been defined.

.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Figure 1. An Example of a Cellular Space Construction

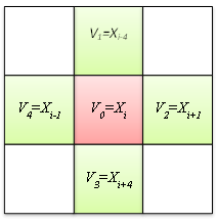

Figure 2. Von Neumann-Neighborhoods

In detail, our method was described as follows. Firstly, from the dataset that consist of 16 regions, we defined the two-dimensional space and put each region into a one cell and set an index for each cell, then we defined an array variable to represent the 16 cells in which each cell has an index. Next, we put the number of data cases for each region (number of infected)

into an array variable in which data in one period were stored into an array variable, running as time-step. Finally, with a set of states criterias (Table 2.), we replaced the data cases with the data states.

Table 2. State Definition of Infected Area

State	State Definition	State Colour
S_1 :	all peoples have been recovered, or no one was infected	
S_2 :	1-2 peoples were infected	
S_3 :	3-5 peoples were infected	
S_4 :	> five peoples were infected	

The main problem of this research is how to find the function that represents a proper CA's rule. Many methods to find function *f* as rule of CA model have been conducted, and in this research we used HMM, a method that has not been used by researchers yet. By ignoring the death factor and the birth factor, and by assuming that the probability of an infected cell is affected by surrounding cells that are considered an effective influence, then the HMM approach was suitable to be used to determine a probabilistic function *f*.

The CA characteristic was represented as a Markov process [20]. Since the dataset was able to be classified as a time series dataset, it was proper to use a probabilistic function that can be found using HMM. HMM is a probabilistic model that is suitable for solving the problem related to the data sequential-temporal [12]. Mathematically, the HMM is written as $\lambda = (T, E, \pi)$, where λ is the HMM model, T is a matrix of Transition Probabilities, E is a Matrix of Emission Probabilities, and π is a *Prior* Matrix [15]. In the CA model that has defined, the state change of a cell to another state can be described as a State Transition Diagram. The State Transition Diagram was able to express HMM model as T. The state change probabilities of a certain area affected by its neighborhoods are emission probabilities was able to express HMM model as E.

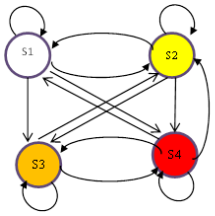

Figure 3. State Transition Diagram – Ergodic HMM

Table 3. Transition Probabilities Matrix

Period n-1	Period n				
	S_1	**S_2**	**S_3**	**S_4**	**Σ**
S_1	$P(S_1\|S_1)$	$P(S_1\|S_2)$	$P(S_1\|S_3)$	$P(S_1\|S_4)$	**1**
S_2	$P(S_2\|S_1)$	$P(S_2\|S_2)$	$P(S_2\|S_3)$	$P(S_2\|S_4)$	**1**
S_3	$P(S_3\|S_1)$	$P(S_3\|S_2)$	$P(S_3\|S_3)$	$P(S_3\|S_4)$	**1**
S_4	$P(S_4\|S_1)$	$P(S_4\|S_2)$	$P(S_4\|S_3)$	$P(S_4\|S_4)$	**1**

Table 4. Emission Probabilities Matrix

Main Object (C)	Observed Object (V_i)			
	S_1	**S_2**	**S_3**	**S_4**
S_1	$P(C=S_1\|X_i=S_1)$	$P(C=S_1\|X_i=S_2)$	$P(C=S_1\|X_i=S_3)$	$P(C=S_1\|X_i=S_4)$
S_2	$P(C=S_2\|X_i=S_1)$	$P(C=S_2\|X_i=S_2)$	$P(C=S_2\|X_i=S_3)$	$P(C=S_2\|X_i=S_4)$
S_3	$P(C=S_3\|X_i=S_1)$	$P(C=S_3\|X_i=S_2)$	$P(C=S_3\|X_i=S_3)$	$P(C=S_3\|X_i=S_4)$
S_4	$P(C=S_4\|X_i=S_1)$	$P(C=S_4\|X_i=S_2)$	$P(C=S_4\|X_i=S_3)$	$P(C=S_4\|X_i=S_4)$
Σ	**1**	**1**	**1**	**1**

Based on the states criteria (Figure 3), Ergodic Hidden Markov Models (Ergodic-HMM) was applied to get a probabilistic function [21]. Each arrow in the state diagram represents a probability value of an object to change the value of state from one period to the next one. The values were stored in *T* as shown in Table 3. The Emission Probabilities were stored in *E* as shown in Table 4. In this study, we assumed that the initial state of a cell at the beginning of the period had the same probability of possible states values. Thus, the prior probabilities were ignored.

The probability of an object (C_i) to change its state was calculated using Bayes theorem as follows:

$$P(\mathrm{C}_i \mid X_i) = \frac{P(X_i \mid \mathrm{C}_i).P(\mathrm{C}_i)}{P(X_i)} \quad (1)$$

In general, it was written as

$$P(C_1, C_2, ..., C_n \mid X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid C_i).\prod_{i=1}^{n} P(C_i \mid C_{i-1}) \quad (2)$$

The Transition Probability value was calculated by the formula:

$$P(Si \mid Sj) = \frac{S_{ij}}{\sum S_i} \quad (3)$$

Moreover, the Emission Probabilities value was calculated by the formula:

$$P(V_i \mid V_0 = \mathrm{S}_j) = \frac{S_{Vi}^{j}}{\sum S_{Vi}} \quad (4)$$

2.3. Data Construction for Spreading Pattern

Data attributes used in this research are: “name of a region” and “number of Dengue cases”. The cellular space is defined as a two-dimensional space in which each cell represents a region with some Dengue cases in each period. The total region in West Bogor is 16 regions. Thus, we defined 16 cells. Each cell contained some un-uniformed objects that described some Dengue cases that occurred in a region for the certain period. Each cell was defined as a one-dimensional variable $X = \{X_i / i = 1, 2, .., 16\}$. Variable X_i represented a region as shown in Figure 1. Cell neighborhoods were defined as a one-dimensional array variable $V = \{V_j \mid j = 0, 1, 2, 3, 4, 5\}$. The Neighborhood frame moved in the cellular space with the equation:

$$V_0 = X_i ; V_1 = X_{i-4} ; V_2 = X_{i+1} ; V_3 = X_{i+4} ; V_4 = X_{i+4} \quad (5)$$

The Neighborhood frame as indicated in Figure 2 moved to each cell in the cellular space. Whenever moving, the initial condition states of each cell were checked. Next, we calculated the maximum probability value of the focus cell to changes the state value for the next period. States value was represented by an array, with the array variable of $S = \{S_1, S_2, S_3, S_4\}$. To build a simulation model, we used Excel spreadsheets as a tool to find a probabilistic function, and used Scipy module in Python 3.4 as tool for evaluating the proposed model.

2.4. Predict the Disease Spread Patterns using an Proposed Model

In order to predict the spreading pattern of Dengue Fever, we applied the CA model to a new dataset. To initialize the simulation, we used the John von Neumann-Random Number Generator based on the CA rule, that is equivalent to a two-dimensional space for generates the j^{th} cell in i^{th}-row by taking cells from the previous row [22] as follows:

$$X_j^{(i)} \equiv \left(X_{j-1}^{(i-1)} + X_j^{(i-1)} + X_{j+1}^{(i-1)} + X_j^{(i-1)} X_{j+1}^{(i-1)} \right) \bmod 4 \quad (6)$$

where mod 4 indicates the number of states that might, $X_j^{(i)}$ describes the j^{th} cell in i^{th}-row which can take on the value 0,1,2, and 3. The values of the cells in the first row, which range from 0 to 3, were randomly assigned by Simple Linear Congruential Generators.

2.5. Verification and Validation

We evaluated the proposed models by conducting verification and validation. Model verification aims to ensure that the CA model has been implemented correctly. Moreover, the purpose of validation is to determine whether the theory and assumptions underlying the preparation of this model are correct [23]. The SIR-model (Susceptible-Infected-Recovered) is a simple mathematical model based on ODE that has been proven to be an acceptable model in epidemic fields [24]. The SIR model was represented as shown:

(7)

(8)

(9)

where **S** = number of susceptible, **I** = number of infectious, and **R** = number of recovered. Case β represents the transmission probability of the disease and γ represents the period of infection.

We verified the model by comparing the tendency of graphs yielded by the proposed model, and the trend graphs yielded by the SIR model. Sequentially we validated the model by measuring the proximity of the simulation results of the proposed model and those of the SIR-model using a correlation coefficient measure to compute similarity [25]:

$$Corr(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)\sigma_X \sigma_Y} \quad (10)$$

where X_i and $\overline{X}$ declared time-series data and the average generated by the proposed model, respectively, Y_i and $\overline{Y}$ declared time-series data and the average generated by the SIR model, respectively. σ_X and σ_Y represented standard deviation of variable *X* and variable *Y*. The similarity value lies between 0.5 - 1. The closer it gets to 1, two time-series data can be said to be similar [26].

3. Results and Analysis

3.1. Probabilistic Functions obtained as Rule on CA Model

The spreading pattern and prediction model were represented by the probabilistic function that represents the CA rule. The probabilistic function obtained in this research is described as follows:

$$f \equiv \max_{i=1}^{4}\left\{\prod_{j=1}^{4} P(V_j \mid V_0^n = S_i).P(V_0^n = S_i \mid V_0^{n-1})\right\} \quad (11)$$

where V_0^n represents state value of cell V_0 in n-th period, and $P(V_o^n = S_i \mid V_0^{n-1})$ is the probability of V_0 that is at S_i in the next period. This function allows us to choose the maximum value of the probability of the state change. It means that the change of a state in the cell of V_0 in the next period depends on the maximum probability value obtained from equations (12-15). These probabilities consist of four possibility values that are defined as follows.

$$P(V_0^n = S_1 \mid V_1, V_2, V_3, V_4) = \prod_{i=1}^{4} P(V_i \mid V_0^n = S_1).P(V_0^n = S_1 \mid V_0^{n-1}) \quad (12)$$

$$P(V_0^n = S_2 \mid V_1, V_2, V_3, V_4) = \prod_{i=1}^{4} P(V_i \mid V_0^n = S_2).P(V_0^n = S_2 \mid V_0^{n-1}) \quad (13)$$

$$P(V_0^n = S_3 \mid V_1, V_2, V_3, V_4) = \prod_{i=1}^{4} P(V_i \mid V_0^n = S_3).P(V_0^n = S_3 \mid V_0^{n-1}) \quad (14)$$

$$P(V_0^n = S_4 \mid V_1, V_2, V_3, V_4) = \prod_{i=1}^{4} P(V_i \mid V_0^n = S_4).P(V_0^n = S_4 \mid V_0^{n-1}) \quad (15)$$

The values of $P(V_i \mid V_0^n = S_1)$; $P(V_i \mid V_0^n = S_2)$; $P(V_i \mid V_0^n = S_3)$; $P(V_i \mid V_0^n = S_4)$ were obtained from THE Emission Probabilities Matrix using the equations (15-18). Moreover, the values of $P(V_0^n = S_1 \mid V_0^{n-1})$; $P(V_0^n = S_2 \mid V_0^{n-1})$; $P(V_0^n = S_3 \mid V_0^{n-1})$; $P(V_0^n = S_4 \mid V_0^{n-1})$ were obtained from *T*,

$P(V_0^n \mid V_0^{n-1})$ was obtained based on the formula described in Table 3 with the results of probability values as follows (Eq. 16):

$$P(V_0^n \mid V_0^{n-1}) = \begin{pmatrix} 0.60 & 0.32 & 0.06 & 0.01 \\ 0.54 & 0.30 & 0.13 & 0.03 \\ 0.59 & 0.24 & 0.18 & 0.00 \\ 0.60 & 0.20 & 0.20 & 0.00 \end{pmatrix} \quad (16)$$

In this matrix, we saw that the change from state S_4 to S_1 has the highest probability value. From the matrix, we also saw that the possibility of a cell's state change next period from S_3 to S_4 was very small or never occurred. Moreover, if the condition of a cell was in the state of S_4, the state tent to change to the better condition because the probability of the cell's state to keep its state was very small or never occurred. It means that there were always the preventive actions to stop the spreading of Dengue Fever diseases in this area.

In this research, *E* is a metric for representing the state change probability of a certain area affected by its neighborhood one. There were four matrices *E* shown in Equation (17-20) as follows.

$$P(V_1 \mid V_0) = \begin{pmatrix} 0.6579 & 0.2500 & 0.0789 & 0.0132 \\ 0.4423 & 0.4038 & 0.1346 & 0.0192 \\ 0.5833 & 0.2500 & 0.1667 & 0.0000 \\ 0.0000 & 0.5000 & 0.2500 & 0.2500 \end{pmatrix} \quad (17)$$

E (Eq. 17) described the probability of a state change of cell V_1 on the next period that is affected by the state change of cell V_0. It was shown that the probability of V_1's state change to S_4 was very small or never occurred, while V_0 was in S_3. Moreover, this matrix also showed that the change from the state of area V_1 to S_1, while V_0 was in S_4 was never occurred.

$$P(V_2 \mid V_0) = \begin{pmatrix} 0.6538 & 0.2692 & 0.0769 & 0.0000 \\ 0.4255 & 0.4255 & 0.1277 & 0.0213 \\ 0.4667 & 0.2667 & 0.2000 & 0.0667 \\ 0.2500 & 0.5000 & 0.2500 & 0.0000 \end{pmatrix} \quad (18)$$

From the equation (19), we saw that the state change of V_2 to S_4, affected by V_0, will never occur while the state of V_0 is in the condition of S_1 or S_4.

$$P(V_3 \mid V_0) = \begin{pmatrix} 0.6250 & 0.2875 & 0.0875 & 0.0000 \\ 0.4222 & 0.4667 & 0.0667 & 0.0444 \\ 0.3750 & 0.4375 & 0.1250 & 0.0625 \\ 0.3333 & 0.3333 & 0.0000 & 0.3333 \end{pmatrix} \quad (19)$$

Equation (19) described the probability of a state change of cell V_3 on the next period that was affected by the state condition of V_0. Equation (20) described the probability of a state change of cell V_4 on the next period that was affected by the V_0 state condition. From the four matrices above, it was concluded that the extreme change conditions of the neighborhoods to S_4, affected by the focus area, were very rare.

$$P(V_4 \mid V_0) = \begin{pmatrix} 0.6456 & 0.2532 & 0.0886 & 0.0127 \\ 0.4468 & 0.4255 & 0.0851 & 0.0426 \\ 0.3750 & 0.3750 & 0.1875 & 0.0625 \\ 0.0000 & 0.5000 & 0.5000 & 0.0000 \end{pmatrix} \quad (20)$$

3.2. Prediction Model Results that Obtained Using CA Model

An example of the simulation results was shown in Figure 4. The pattern was able to obtained using equation (11). The inputs to this equation were the Odds Transition Matrix (Eq. 16), the Odds Emissions Matrix (Eq. 17-20), and the randomized data initialization that was

obtained by equation (6). The visualization of the obtained pattern results indicated that the spread of Dengue disease occurred on average in seven to eight periods.

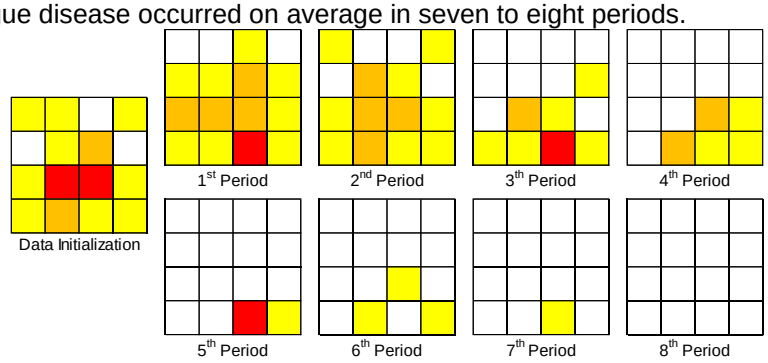

Figure 4. The Prediction Results of Dengue Spreading Pattern on The CA Model

From several computational simulations, it was seen that if the disease began to spread simultaneously in cells 11 and 16, the pandemic would subside in a longer period. For example, cells 11 and 16, respectively, represented the area of Cilendek Barat and Pasir Kuda. It also appears that cell 15, representing the area Pasir Mulya, is the most vulnerable cell to the spread of the disease. It was shown with the color indicator in that area.

3.3. Evaluation

Verification was conducted by comparing the prediction results yielded by the proposed model (using the CA approach) to those of one yielded by SIR model (a classic and popular approach). The verification results showed that the slope of the infected area-period graph (Figure 5) that represented the CA model (the proposed model) was similar to that of the SIR model.

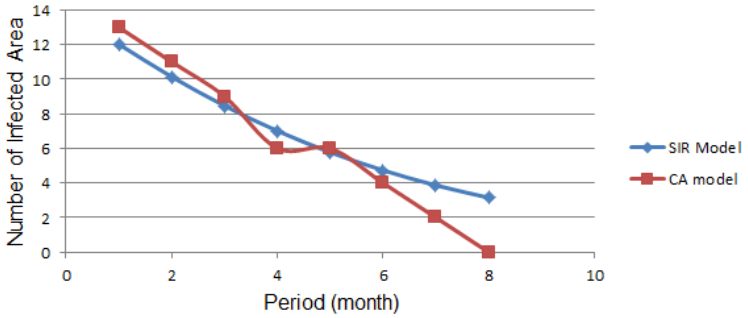

Figure 5. The Tendency of Graph Number of Infected Area in Bogor Barat

In addition, validation was conducted by calculating the similarity between the resulting prediction using the CA model and that of one obtained by the SIR model using Equation 10. The validation result indicated that the resulting prediction using CA had similarities to the SIR model of 0.95. Thus, based on the verification and validation results, it was able to be stated that the proposed approach using CA had been implemented correctly, and the assumptions underlying this model are correct. Thus, the visualization of the spreading pattern yielded by this model was able to be used for understanding and predicting the spread of Dengue disease. This prediction is required for helping prevent the spread of Dengue disease in prone regions. However, the proposed model still has a limitation in that it did not consider the behavior of people. Thus, this model is only valid to a relatively static society.

4. Conclusion

From the results, we concluded that we succeeded in developing a spreading pattern model of the Dengue Fever based on the CA approach by setting four parameters supported by HMM and using Von Neumann neighborhood. The proposed model was able to predict the spread of Dengue disease and provided us the information of which area should be observed carefully. Moreover, the evaluation result showed that the CA model was capable of generating patterns similar to that of one generated by SIR models with a similarity value of 0.95.

5. Recommendation

In this research we did not consider the behavior of people. We realized that the spread of the Dengue fever depends on the population of *Aides aegypti* mosquito, which is related to the behavior of people. Thus, the system dynamic approach can be considered for modeling the spreading pattern as a future work. Moreover, we also could use the other technique for substituting HMM in order to find a probabilistic function that represents the CA rule.

References

[1] Cuesta H. Practical Data Analysis. Packt Publishing Ltd, Birmingham-Mumbai.2013: 153-173.
[2] Pfeiffer D. Spatial Analysis in Epidemiology. Oxford University Press. 2008.
[3] White SH, Rey AMD, Sänchez GR. Modeling Epidemic Using Cellular Automata. *Applied Mathematics and Computations*. 2007; 186:193-202.
[4] Nishiura H. Mathematical and Statistical Analyses of the Spread of Dengue. Dengue Bulletin. 2006; 30: 51-57.
[5] Santos LBL, Maretto RV, Medeiros LCC, Feitosa FF, Monteiro AMV. *A Susceptible-Infected Model for Exploring the Effects of Neighborhood Structures on Epidemic Processes- A Segrwgation Analysis*. Proceedings XII Geoinfo. Campus do Jordio Brasil. 2011, 12. 85-96.
[6] White SH. Using Cellular Automata to Simulate Epidemic Diseases. *Applied Mathematical Sciences. 2009;* 3(20): 959-968.
[7] López L, Burguener G, Giovanini L, Baldomenico P. *A Cellular Automata to Model Epidemics*. Proceedings of the JAIIO – Congreso Argentino de Informatica y Salud (CAIS). 2013; 42: 150-161.
[8] Athithan S, Shukla VP, Biradar SR. Dynamic Cellular Automata Based Epidemic Spread Model for Population in Patches with Movement. *Journal of Computational Environmental Sciences.* 2014. Article ID 518053. Hindawi Publishing Corporation. http://dx.doi.org/10.1155/2014/518053.
[9] Santos LBL, Costa MC, Pinho STR, Andrade RFS, Barreto FR, Teixeira MG, Barreto ML. *Periodic Forcing in a three-level Cellular Automata Model for a Vector-transmitted Disease*. Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics (Print). 2009; 80: 016102.
[10] Elsayed WM, .El-Bassiouny AH,. Radwan EF. Applying Inhomogeneous Probabilistic Cellular Automata Rules on Epidemic Model. *International Journal of Advanced Research in Artificial Intelligence*. 2013; 2(4).
[11] Hagoort M, Geertman S, Ottens H. *Spatial Externalities, Neighborhood Rules and CA Land-use Modeling.* In: The Ammals of Regional Science. 2008; 42(1).
[12] Jyh-Ying P, John ADA, Cheng-Yuan L. Modeling Time Series and Sequences Using Markov Chain Embedded Finite Automata. *International Journal of Innovative Computing, Information, and Control*. 2011; 7(1): 407-431.
[13] German B, Leonardo L, Leonardo G. *Modelling Population Heterogeneity in Epidemics using Cellular Automata.* Proceeding of Asociation Argentina de Mecanica Computacional. Rosario. 2011; 30: 3501-3514.
[14] Mitchell M. *Evolving Cellular Automata with Genetic Algorithms: A Review of Recent Work*. Proceedings of the First International Conference on Evolutionary Computation and Its applications. Moscow, Russia. 1996.
[15] Dugat R. *A Tutorial on Hidden Markov Models*. Indian Institute of Technology. Report number: SPANN-96.1. 1996.
[16] Saragi D. Trend Analisis dengan Metode Time Series untuk Meramalkan Penderita Demam Berdarah Tahun 2010-2014 Berdasarkan Data Penderita Demam Berdarah Tahun 2005-2009 di Propinsi Sumatra Utara. Master Thesis. Medan: Fakultas Kesehatan USU. 2011.
[17] Octora M. Perbandingan Metode ARIMA (Box Jenkins) dan Metode Winter dalam Peramalan Jumlah Kasus DBD. Master Thesis. Surabaya: Fakultas Kesehatan Masyarakat, Unair. 2010.
[18] Wang Z. Public Evacuation Process Modeling and Simulatiaon based on Cellular Automata. *TELKOMNIKA*. 2013; 11(11):. 6468~6476.
[19] Djatna T, Morimoto Y. Attribute Selection for Numerical Databases that Contain Correlations. *Int. J. Software and Informatics*. 2008: 2(2): 125-139.

[20] Knutson JD. A Survey of the Use of Cellular Automata and Cellular Automata-Like Models for Simulating a Population of Biological Cells. Master Thesis. Iowa: Iowa State University. 2011.
[21] Terry L. Ergodic Hidden Markov Models for Visual-Only Isolated Digit Recognition. PhD Dissertation. Evanston Illinois. 2007.
[22] Gentle JE. Random Number Generation and Monte Carlo Methods. In: Chambers J, Eddy W, Härdle W, Sheather S, Tierney L. Editors. Statistic and Computing. New York: Springer Science + Business Media, Inc. 2003.
[23] Sargent RG. *Verification and Validation of Simulation Models*. Proceedings of the 2007 Winter Simulation Conference. IEEE. 2007: 124-137.
[24] Bootsma MCJ, Ferguson NM. *The effect of Public Health Measures on the 1918 Influenza Pandemic in US cities.* Proceedings of the. National Academy of Science. US: The National Academy of Sciences. 2007; 104(18): 7588-7593.10.1073/pnas.0611071104..
[25] Sung-Hyuk C. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2007; 1(4): 300-307.
[26] Bajracharya K, Duboz R. *Comparison of Three Agent Based Platforms on the Basis of a Simple Epidemiological Model (WIP)*. Proceedings of The Symposium on Theory of Modeling & Simulation. 2013.