Endemic Outbreaks of Brown Planthopper (*Nilaparvata lugens* Stal.) in Indonesia using Exploratory Spatial Data Analysis

Sri Yulianto Joko Prasetyo¹, Subanar², Edi Winarko³ and Budi Setiadi Daryono⁴

¹Faculty of Information Technology, Satya Wacana Christian University
Salatiga, Central of Java, Indonesia

²Faculty of Mathematic and Natural Science, Gadjah Mada University
Yogyakarta, Indonesia

³Faculty of Mathematic and Natural Science, Gadjah Mada University
Yogyakarta, Indonesia

⁴Faculty of Biology, Gadjah Mada University, Yogyakarta

**Abstract**

This research attempts to compare the method of analysis and the mapping of endemic outbreaks of brown planthopper on the staples commodity and horticulture using spatial autocorrelation method. The research was done through three steps; those are (1) data preprocessing (2) endemic prototype of BPH (3) the analysis of the increase of attacked stem brown planthopper parameter, the rainfall intensity, and the density of natural enemy. GISA, LISA, and Getis Ord Statistic were used in the endemic of BPH modeling. The result of this research shows that hotspot pattern in 37 subdistrict area and coldspot pattern in 13 subdistrict areas in 2001 - 2010 can be classified by using these methods. From the comparison of Local Moran’s experiment map and Getis Ord BPH experiment map in 2001, 2006 and 2010, it is found that the indication of hotspot on Local Moran’s is the same as clustering indication on Getis Ord based on the value of Z(Gi)>2. The distribution of BPH attack is determined by the spatial object connectivity pattern of BPH population, rainfall and the territorial position.

**Keywords:** Spatial Autocorrelation, Getis Ord, Brown Planthoppers, Endemic Pattern.

1. **Introduction**

Two or more connected spatial objects are always involved in the epidemiological process. The first spatial object is called as the source of disease outbreak and the second is the side which suffered from the disease. The pattern of spatial object connectivity between the source and the outbreak target is termed as the interspatial variable. It is influenced by several factors including euclidean distance, neighborhood, and interaction factor[1]. It can also be used to describe spatial tendency, hierarchical effect, neighborhood effect, and the territorial risk of occurrence. The decision on the degree of spatial object connectivity can be done using Spatial Autocorrelation (SA) method [1][2]. SA constitutes a method used for exploring and analyzing spatial data or Exploratory Spatial Data Analysis (ESDA). Many scientists have oftenly used it to make a modeling and simulation of natural phenomena in a real world. The modeling and simulation are needed because some of the causal natural phenomena cannot be identified. The ESDA method is a set of techniques for the process of spatially distributed visualization, the identification of atypical location or outliers, the representation of spatial association pattern, the identification of spatial cluster or hotspot, coldspot, and spatial regime as spatial heterogeneity [3][4][5][6]. ESDA method has been used for surveillance and epidemiology. Surveillance and epidemiology themselves are for (1) the mapping of disease attack occurrence, (2) the analysis of disease dynamics, and (3) the prediction of disease attack/outbreak in a particular area [7].

In Indonesia, it is observation center for plant pest organisms agricultural department (BBPOPT) that works on surveillance activities and the technology of pest prognostication development. The operational standard procedure of surveillance and attack prognostication BBPOPT has not provided spatial structural information, guideline of correlation between spatial object, attack occurrence, and
information about the dynamics of occurrence process.

One of the pest types having endemically attribute is Brown Planthopper (Nilaparvata lugens Stal.)(BPH). This plant disease attacks rice plants commodity in Asia, Oceania, including Australia, New Zealand and some islands around the Pacific Ocean. Asia comprises Bangladesh, Brunei, Burma (Myanmar), China, Hong Kong, India, Japan, Cambodia, Korea, Laos, Malaysia, Nepal, Pakistan, Philippines, Singapore, Sri Lanka, Taiwan, Thailand, and Vietnam [8][9]. One of Asian Countries with high BPH attack endemic is Indonesia. The greatest BPH attack happened in 1961 – 1970 spread in 52,000 Ha, 3,093,593 Ha in 1971 – 1980, 458,038 Ha in 1981 – 1990, 312,610 Ha in 1991 – 2000 and 351,748 Ha in 2001 – 2010 [10] [11]. The increase of stem brown planthopper attack in planting season during 2009 – 2010 was 126,840 Ha, and Central Java, which reached the number of 33,425 Ha, became the widest attacked area [12]. According to data provided by Agricultural Department RI on Staples and Horticulture in Central Java, BPH attack for the last 10 years has spread in 28 Counties/cities, and the highest attack are in Klaten, Sukoharjo, Karanganyar, Sragen, Boyolali, and Wonogiri.

All this time, analytical method and the mapping of pest endemic area is based on the historical attack occurrence data in that area regardless of the neighboring areas (independent data). Analysis parameter comprises: (1) the frequency of attack, (2) the average widespread attack, (3) the average of puso, (4) puso ratio [13]. In this research, the decision on pest endemically stratification will be proposed based on the historical attack occurrence data and the correlation between attacked area and its neighboring areas. Basic idea in this research is that pest attack data is not independent but influenced by attack dynamics of neighboring area (dependent data). In order to prove that idea, spatial autocorrelation method is used. This method is hoped to provide information about dynamics of attack process, the factor related to the attack occurrence, and the prediction of attack process based on the association inter spatial object. The research was done in the areas with high endemics in Indonesia comprising the county of Klaten, Sukoharjo, Karanganyar, Sragen, Boyolali, and Wonogiri in Central Java.

This paper is organized as follows. Section 2 describes some related ESDA researches used as the reference for model development. The section 3 is the theoretical background, the concept of Moran’s I, G statistic, and Exponential Smoothing prediction method. The section 4 is the method proposed, describes architectural model proposed in this research. The section 5 is describes discussion the proposed experiment ESDA. The section 6 is conclusion and future work.

2. Related Works

In 1988, the research of SA modeling especially LISA (Local Indicator Spatial Association) was started by Luc Anselin in spatial econometrica model although the use of this model had become popular in 1995 [14]. Furthermore, the focus of LISA research area or hotspot analysis was developed by Getis and Ord [15]. The most popular function used for spatial connectivity modeling was Moran’s I, the experiment of Mantel Test and Correlogram. The spatial connectivity constituted representation of (1) the degree of correlation inter location, (2) the parameter of the best connectivity for the different situation, (3) the procedure for finding out landscape configuration change (4) connectivity change as the result of organism dissemination [16].

In the spatial connectivity, the resemblance of distance value constituted important indicator to determine three things, they were (1) spatial pattern, (2) spatial structure, and (3) spatial process. Spatial structure could be used to describe dependency of spatial quantification per distance inter classes or lag through its structural function, including:(1) correlogram, (2) variogram and (3) periodogram [17].

3. Theoretical Background

3.1. Spatial Autocorrelation

SA is the correlations inter observed area in the form of spatial pattern (distance, time, and spatial pattern) [18]. The criterion of SA phenomenon occurrence is when the distribution of one observed variable value follows the particular pattern systematically. The use
of SA method for determining the degree of connectivity inter spatial object was done through two approaches, they were: (1) Global Indicator Spatial Association (GISA) and (2) Local Indicator Spatial Association (LISA). The Moran’s method worked is comparing particular variable value in each area with the value in all observed areas [19]. GISA is the analysis of spatial associated pattern on a broader scale to see data distribution, whether clustering was formed or not, dispersed, random in one space [20]. GISA was defined on the equation (1).

\[
I = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}
\]  

(1)

Notation \( n \) is the total observed case data, \( w_{ij} \) is spatial weight matrix element, \( x_i \) is observed value i, \( x_j \) is the value of neighbor observation j, \( \overline{x} \) was the average value x. GISA value is interpreted by the use of Moran’s I \( I \) ranging from \(-1 \) until \(+1 \). Index I that had value of -1 represents data spread or averaged objects (uniform), I that had value of 0 represented random spread and independently. I that had value of +1 represented data/objects that were alike to form clustering [21]. LISA is the analysis for SA quantification in the smaller area which produces highly statically significance (hotspots), lower statically significance (colds spots) and outlier. LISA was defined on the equation (2).

\[
I = \frac{z_i \sum z_j w_{ij}}{\sum w_{ij}}
\]

(2)

\( z_j \) and \( z_i \) are the amounts of value of I deviation.

LISA was interpreted as follows, (1) if the variable in one observed location is the same with its neighbor and has high value, it will be called as HH (high-high) or hotspots, (2) if the variable in one observed location is the same with its neighbor and has low value, it will be called as LL (low-low) or coldspots, (3) if the variable in one observed location has high value while its neighbor has low value, it will be called as HL (high-low) or outlier, (4) if the variable in one observed location has low value while its neighbor has high value, it will be called as LH (low-high) or outlier.

3.2. Exponential Smoothing

Generally, Exponential Smoothing method is used to predict surveilant data because it is simpler, easier to be interpreted, adaptive and easier to develop in automatization compared to other methods [22]. Empirically, in surveilant data analysis, Exponential Smoothing method has an pestimal accuaverage prediction for analysis of tendency (trend analysis) and is more effective for classifying spatial data which has seasonal attributes on the various level of scale [22][23][24][25]. Food and Agriculture Organization (FAO) recommended the use of Exponential Smoothing method to replace linier regression model which has been used in prediction. This method is effective in the short-time prediction (prediction for 1 – 2 period of time in the future)[26]. The appropriate method for tendency pattern formation (trend) is Double Exponential Smoothing which has been formulated in the equation (3) until (7).

\[
F_t = a_t + b_t(m)
\]

(3)

\[
a_t = S_t' + (S_t' - S_t''), \quad S_t' = 2S_t' - S_t''
\]

(4)

\[
b_t = \frac{1}{1-\delta}(S_t' - S_t'')
\]

(5)

\[
S_t' = \left(\frac{\delta}{n}\right)x_t + (1 - \left(\frac{\delta}{n}\right))S_{t-1}'
\]

(6)

\[
S_t'' = \left(\frac{\delta}{1-\delta}\right)S_{t-1}'' + (1 - \left(\frac{\delta}{1-\delta}\right))S_{t-1}''
\]

(7)

Notation \( \delta \) is constant, notated with the value of between 0 and 1. \( \left[1 - \left(\frac{\delta}{n}\right)\right] \) is the actual value of time progression, while \( F_t \) is the result of prediction on time, and \( t \) and \( m \) is the data period that is going to be predicted.

3.3. G Statistic

Getis and Ord popularized LISA method known as G and G * statistic [27]. This method was used to evaluate hotspots and coldspots spatial pattern from the observed data. The value of G would be positive if the clustering happened in one observed data result or that value is more than mean and the value of G would be negative if the clustering value which is less than mean happened in one observed data result[28]. The equation of G statistic could be defined as in the equation (8).

\[
G = \sum w_{ij} (d) z_j / \sum z_j
\]

(8)

Based on this equation, \( w_{ij} \) is the element of spatial weight matrix and \( d \) is the element of distance inter spatial object [29].

4. Method Proposed

Architectural model, proposed in this research, can be seen in (Fig.1.) The experiment consist of four steps. The step 1 is data preprocessing. The data are used for research, attack potential prediction, the fluctuation of BPH attack, the rainfall on the
The worst BPH plant disease attack in the last 10 years happened in six counties which later on became study focus. Those counties were Klaten, Sukoharjo, Wonogiri, Sragen, Karanganyar and Boyolali. The outbreak of BPH attack was 19.181 Ha. It was 57.38% out of the outbreak of BPH attack in the whole Central Java province, which was 33.425 Ha [28]. The comparison of ricefields width in the same year in those six counties based on Central Java data in 2010 was 172.745 Ha, the width of the unproductive fields because of the BPH attack was 11.10%. Based on the BPH attack data in 2001 – 2010, it could be known that the culminating point of BPH attack in the last 10 years happened in 2010 (Fig.2.).

5. Experiment

5.1. Data Experiment

The modeling of endemicity and parameter analysis used R (http://cran-r-project.org). Exponential Smoothing prediction was done by using package tseries and forecast and spatial analysis using package spdep, rgdal, sp, classInt, RColorBrewer, maptools, rgdal and maps. Visualization and representation of analysis result was done by using (1) Choropleth map, (2) Moran’s Scatterplot, and (3) LISA (Moran’sI and G statistic) map. The research was done in the areas that have high BPH attack comprising 124 subdistricts in seven counties in Central Java.
on June 2010 with 9658 Ha in width and then went down to 1995 Ha in July 2010 (Fig.3.). The high rainfalls, which were 31.353 mm (starting on March 2010), 100.680 mm on April 2010, and 24.269 mm on May 2010, took role in reaching this culminating point (Fig.4.). The high rainfall on March – May 2010 made the grain cluster and weeds around them being washed away. Thus, the rain water brought the eggs, larvae and BPH spreading to its surroundings. When the rainfall went down to the lowest level but with high humidity, the BPH outbreak happened on June 2010.

5.4. The Mapping of Data Attack

Attack frequency (F) was the number of planting time reported hit by the BPH attack (the value is not 0) in all planting seasons (20). F value was determined by counting classification of frequency interval and predicted attack class. The result of F value counting was mapped by using choropleth as the BPH endemicity classification (Fig.5(a)). This endemicity classification was only determined by counting on historical BPH attack data in 10 years regardless the spatial correlation attack occurrence inter counties. As the comparison, the mapping of BPH attack frequency using LISA method was done to see endemicity stratification based on spatial association inter county areas (Fig.5(b)).

We can see in (Fig.5(a).) that the high attack frequency occurred in the very vast areas, that was 66 sub districts or 53% out of the whole observed area. However, if we go a little bit further, attack frequency value has wide interval, it was 3 for the lowest attack per planting season (PS) and 19 for the highest attack frequency value per PS. If we compared to LISA attack frequency map, we could see that there was a number of area which could be categorized as cluster and hotspots. The cluster areas comprised 21 subdistricts or 17 % out of total observed areas that had spatial correlation. The decision on the average of widespread attack BPH counted in three steps, they were (1) the average of each observed hit area fitted to analyzed planting season counting, (2) approximated hit area classification determination, and (3) determination of class hit area average. The average of widespread attack BPH in Region V area Surakarta in 2001, 2006 and 2010 based on data in 2001 – 2010 is in (Fig.6.)

Unique phenomenon in spatial analysis was the possibility in visualizing BPH outbreaks based on the particular period. In 2001, there were 10 subdistricts or 8 % out of total observed areas included in high attack category. On the next 5 years (2006), it increased to 24 subdistricts or 19% out of total
observed areas included in high attack category. And in 2010, there were 52 subdistricts or 42% out of total observed areas included in high attack category.

5.5. GISA and LISA Modeling

According to GISA analysis in table.1., the experiment of index Moran’s in 2001 - 2010 had positive value, it means that BPH attack in region V Surakarta area (Boyolali, Klaten, Sukoharjo, Sragen, Karanganyar and Wonogiri) formed the clustering pattern (Fig.7. – Fig.9.). The result of Moran’s Scatterplot analysis showed that there were 22 subdistricts in Quadrant 1 (High-High) in between 2001 - 2010. The areas in Quadrant 1 constituted high BPH attacked endemicity and surrounded by areas with the same high BPH attacked endemicity.

Table 1. The result of Index Moran’s experiment in observed area in 2001 - 2010

<table>
<thead>
<tr>
<th>Year</th>
<th>Index Moran’s</th>
<th>Year</th>
<th>Index Moran’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.3006806</td>
<td>2006</td>
<td>0.24234210</td>
</tr>
<tr>
<td>2002</td>
<td>0.5352283</td>
<td>2007</td>
<td>0.22915600</td>
</tr>
<tr>
<td>2003</td>
<td>0.2797728</td>
<td>2008</td>
<td>0.02557908</td>
</tr>
<tr>
<td>2004</td>
<td>0.1575799</td>
<td>2009</td>
<td>0.23961580</td>
</tr>
<tr>
<td>2005</td>
<td>0.1106040</td>
<td>2010</td>
<td>0.46613220</td>
</tr>
</tbody>
</table>

The specific attribute of this phenomenon was that the area with high LTS BPH attack was surrounded by the areas with high LTS BPH attack. The 12 subdistrict areas in Quadrant 2 (High – Low) constituted high BPH attack endemicity areas surrounded by areas with low endemicity. Generally, the areas in Quadrant 2 were potential enough to be Quadrant 1 if the improvement of cultivation and pest control was not done immediately (Fig.7.).

The result of Local Moran’s’s experiment was represented in the form of LISA map, aimed to quantify spatial autocorrelation in smaller areas and produced high statistically significance (hotspots), low statistically significance (coldspots) and outlier. HH area was known as hot spots, LL was cold spots, HL and LH was outliers. Based on LISA analysis, spatial hotspots phenomenon in seven subdistricts in 2001 became 9 subdistricts in 2006, and finally, 21 subdistricts in 2010. Besides, in 2010, spatial cold spots phenomenon occurred and could be potentially changed into hotspots if the management of pest was not done immediately.

5.6. G Statistic Modeling

Based on the result of Global and Local Moran’s experiment, it could be concluded that in the range of 2001 – 2010, there were 37 subdistrict areas becoming hot spots and 13 subdistrict areas becoming cold spots. Both hot spots area and cold spots area constituted the source of BPH attack in the whole studied areas (124 subdistricts). The outbreaks of BPH attack in studied areas were possible to happen because of the spatial connectivity factor. It could be in the form of transportation and irrigation network correlation (open area that allows BPH migration and the similarity of grain variety to occur.
Beside Global and Local Moran’s experiment, the evaluation of tested research result was done by using Getis Ord method. It was one of numeric methods providing the mechanism of autocorrelation comparison in different spatial environmental characteristic [29]. The analysis result of BPH attack in 2001 – 2010 showed the tested value of Getis Ord as shown in (Fig.9. – Fig.11.) The interpretation of Getis Ord tested value based on the indicator Z(Gi) was (-2.0 < Z(Gi) < 2.0). The value of (Z(Gi) < -2.0) represented the data distribution of outlier and random, mean while, (Z(Gi) > 2.0) represented central data distribution (cluster). By using indicator Z(Gi), it could be concluded that all results had positive value and spatial pattern tendency aiming at clustering. From the comparison of Local Moran’s experiment result map and Getis Ord BPH map in 2001, 2006 and 2010, it could be known that the indication of clustering attack occurrence or hot spots or HH in Local Moran’s was the same with the clustering indication in Getis Ord based on the value of Z(Gi) > 2 (Fig.10. and Fig.11.). The strength of Moran’s analysis could be seen on the comparison of Local Moran’s maps (Fig.8.) and Getis Ord experiment results maps (Fig. 9 – Fig.11.). Based on that comparison, we concluded that both hot spots (High – High) phenomenon and cold spots (Low – Low) phenomenon could be visually and obviously seen and it was easier to be interpreted. While on both hot spots (High – High) Getis Ord phenomenon and cold spots (Low – Low) could only be seen from the value of Z(Gi) appearing in each subdistrict.

Generally, the fluctuation dynamics of the outbreak of BPH attack area follows its environmental dynamics with four components, they were : (1) climate, (2) the availability of food supply, (3) predator population, (4) the condition where that population exists and (5) spraying conduct. The most influencing climate component on the population dynamics was the rainfall, followed by other components (air temperature and humidity) [30]. Some studies on BPH attack were related to climate factor especially the rainfall and humidity. In the high rainfall, most of BPH, larvas, and eggs were washed away by the rain water and taken along through the stream flow. However, the level of humidity in the rainy season was high. As a result, it could be the incentive factors of the increase of BPH reproduction. This can be seen on Figure 14 showing the increase of rainfall in 2010 with the high BPH widespread (Fig.13. and Fig.15.).
The average of yearly rainfall in the whole studied areas was 1172 mm and the culminating point of the rainfall happened in Tawangsari subdistrict in Sukoharjo county (5869 mm). The rainfall prediction using Holt Winter method in 2014 described the periodic increasing followed by the increase of BPH population in studied areas (Fig. 16.) [31].

6. Conclusions and Future Work

Based on this research, it can be concluded that BPH attack will happen in a very vast area. The culminating point of attack will happen in June 2012 as wide as 3932 Ha and in 2013 as wide as 4076 Ha. The high frequency of attack will happen in a very vast area (66 subdistricts or 53% out of the whole observed area). However, that value of attack frequency has very wide range. Those are the three occurrences’s lowest value and the 19 occurrence’s highest value per PS. Compared to LISA map; there were some areas that could be categorized as cluster. They comprised 21 subdistricts or 17% out of total observed areas that have spatial autocorrelation.

In the range time of 2001 – 2010, there were 37 subdistrict areas becoming hot spots and 13 subdistrict areas as cold spots. Both hot spots and cold spots areas constituted the source of BPH attack in the whole studied areas (124 subdistricts). The outbreak of BPH attack in studied areas was possible to happen because of spatial connectivity factor. It could be in the form of transportation and irrigation network correlation (open area that allows BPH migration and the similarity of grain variety to occur). From the comparison of Local Moran’s experiment result map and Getis Ord BPH map in 2001, 2006 and 2010, it could be known that the indication of clustering attack occurrence (cluster) or hot spots or HH in Local Moran’s was the same with the clustering indication in Getis Ord based on the value of Z(Gi) > 2.
The future research is modeling of BPH endemicity was done by using GISA and LISA method from result predicted with Exponential Smoothing method.

References


[22] Shmueli G., dan Fienberg S.E., Current and Potential Statistical Methods for Monitoring Multiple Data Streams for Bio-Surveillance, National Institute of


Sri Yulianto J.P., Received is B.Sc at Biology from the Duta Wacana Christian University Yogyakarta Indonesia in 1995. Received is Master degree at Computer Science at Faculty of Mathematics and Natural Sciences from the Gadjah Mada University Yogyakarta Indonesia in 2002. He is student Doctoral Program Computer Science at Faculty of Mathematics and Natural Sciences from the Gadjah Mada University Yogyakarta Indonesia. He is currently a lecturer in the Informatic Engineering Department, Faculty of Information Technology, Satya Wacana Christian University Salatiga Indonesia. His current research interests include spatial statistic, GIS, simulation and modeling, data mining and their applications.

Subanar, Received the M.Sc. degree in Statistics from Faculty of Mathematics and Natural Sciences Gadjah Mada University Yogyakarta Indonesia in 1971. He received his Ph.D. degree in Statistics from the University of Wisconsin United States in 1976. He is currently a lecturer and professor in the Faculty of Mathematics and Natural Sciences from the Gadjah Mada University Yogyakarta Indonesia. His current research interests include Mathematical Statistics, Neural Network and Bootstrap Method.

Edi Winarko, Received the B.Sc at Statistics from Faculty of Mathematics and Natural Sciences Gadjah Mada University Yogyakarta Indonesia in 2001, M.Sc. degree in Computer Science from the Queen’s University Canada, United States in 2002. He received his Ph.D. degree in Computer Science from the Flinders University of South Australia in 2007 He is currently a lecturer in the Faculty of Mathematics and Natural Science from the Gadjah Mada University Yogyakarta Indonesia. His current research interests include data warehousing, data mining, information retrieval.

Budi Setiadi Daryono, Received the B.Sc at Biology from Faculty of Biology Gadjah Mada University Yogyakarta Indonesia in 2001, M.Agr.Sc degree from the Tokyo University of Agriculture in 2002.He received his Ph.D. degree from Tokyo University of Agriculture in 2005. He is currently a lecturer in the Faculty of Biology Gadjah Mada University Yogyakarta Indonesia. His current research interests include Molecular Genetics and Phylogenetic.

Copyright (c) 2012 International Journal of Computer Science Issues. All Rights Reserved.