III. METHODOLOGY

3.1 Time and Location

This study was conducted from December 2010 to May 2011. Data processing and analysis was accomplished at Remote Sensing and GIS Laboratory of Master of Science in Information Technology (MIT) Bogor Agricultural University, while data collection and field survey were conducted in South Kalimantan province.

Geographically, South Kalimantan province located in southern of Borneo Island at 114° 19" – 116° 33" E and 1° 21" – 1° 10" S. It is bounded by East Kalimantan in Northern, Central Kalimantan in Western, Makassar Strait in eastern and Java Sea in southern. The total area is 37,377,53 km² and mostly covered by forest and peat land. South Kalimantan Province consists of 151 sub districts in 11 districts and two municipalities (PEMPROV KALSEL 2007). The annual mean minimum and maximum temperature is 21.5° C and 34.7° C with average 27.0° C, while the annual mean humidity is 82.2% (BPS 2010).

![Figure 3. Study Area](image-url)
3.2 Data and Tools

The cattle population and JD serological data used in this study were obtained from Animal Husbandry Office of South Kalimantan Province. Cattle population was based on the 2009 annual report, and JD serological data was based on the surveillance activities from suspected animal during 2008 to 2010 in endemic area and in area with higher of cattle population. All samples were transported to the laboratory and screened with Enzyme-Linked Immunosorbent Assay (ELISA) or Polymerase Chain Reaction (PCR). To conduct the spatial analysis of JD, only samples that have been examined by PCR or ELISA followed by PCR diagnostic assays were involved in this study to determine the JD seropositive. Hence, JD case was defined as the samples that positive to PCR assay.

The sub district level polygon base map at 1:250,000 scale was obtained. The geographical coordinates of the central points of each village and sub district was created. JD seropositive was assumed to be located at the centre of each village. JD seropositive and cattle population were imported to the base map and converted into shape file and visualize using ArcGIS 9.2 (ESRI Inc.). All vector data were geo-referenced to the Universal Transverse Mercator (UTM) coordinate system zone 50S with World Geodetic System (WGS) 1984 Datum. Digital Elevation Model (DEM) 90 meter also downloaded from Consultative Group on International Agricultural Research (CGIAR).

<table>
<thead>
<tr>
<th>No</th>
<th>Data</th>
<th>Source</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Base Map (RBI)</td>
<td>Bapeda Tanah Laut</td>
<td>2007</td>
</tr>
<tr>
<td>2</td>
<td>Elevation</td>
<td>SRTM-DEM 90m</td>
<td>2009</td>
</tr>
<tr>
<td>3</td>
<td>Jembrana Disease Serological Data</td>
<td>Disnak Kalsel*</td>
<td>2008 – 2010</td>
</tr>
<tr>
<td>4</td>
<td>Cattle Population</td>
<td>Disnak Kalsel*</td>
<td>2009</td>
</tr>
</tbody>
</table>

* Animal Husbandry Office of South Kalimantan Province

In order to carry out the data processing and analysis, software used in this study are Microsoft Office 2007 to make final report, ArcGIS version 9.2 with extension Spatial Analyst and Geostatistical analyst (ESRI Inc.) to analyze data
and visualize the distribution of JD, and SaTScan version 9.0. (available at http://www.satscan.org) to identify JD cluster.

3.3 Research Framework

The main objective of this study is conducted a spatial analysis of JD in South Kalimantan province. This study consists of four main activities i.e. exploratory analysis to describe the spatial distribution of JD, identification of JD cluster, produce the spatial JD map based on the ordinary kriging analysis, and identifying factors associated with the spatial distribution of JD. Figure 4 depicts the general methodology of this study.

Figure 4. The Study Framework
3.4 Spatial Distribution of Jembrana Disease

To conduct the GIS analysis on the spatial distribution of JD, an endemic area is defined as district that having at least one seropositive of JD based on the PCR assay during 2008 to 2010. The aggregate data on district level were used to mapping the JD seropositive and seroprevalence in South Kalimantan province.

Based on the JD seropositive data, all districts were grouped into four classes: non endemic area, low endemic area with JD seropositive between 1 – 10 cases, medium endemic area with the JD seropositive between 11 – 30 cases, and high endemic area with JD seropositive more than 30 cases.

A hazard map of JD was produced to assess the risk of JD in South Kalimantan Province. The hazard map represents the JD seroprevalence at each district. The crude prevalence of JD was defined as the number of positive samples (JD seropositive) over the total number of samples and was calculated by district. Districts grouped into four categories: free area, low risk area with the seroprevalence less than 10%, medium risk area with seroprevalence between 10 to 20%, and high risk area with the seroprevalence more than 20%.

3.5 Jembrana Disease Mapping

Disease mapping is considered as exploratory analysis used to get an impression of the geographical or spatial distribution of disease or the corresponding risk (Berke 2004). Disease mapping usually chooses certain spatial interpolation method, and then creates a continuous surface of disease distribution according to geographically distributed sampling data of disease (Zhong et al. 2005).

The ordinary kriging is used to generate spatial continuous map of JD in South Kalimantan province. Kriging is an interpolation technique in which the surrounding measured values are weighted to derive a predicted value for an unmeasured location (Wittich 2007). It is determines the spatial distribution at unsampled points and provides optimal unbiased estimates with known estimation variances based on a model of the spatial variation (Fischer et al. 1996).
The basic model for kriging is based on the following equation:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i)$$

where $Z(s_i)$ is the measured value at the $i$th point, $I = 1,2, \ldots, N$, $\lambda_i$ an unknown weight for the measured value at the $i$th point, $s_0$ the prediction location, and $N$ the number of observed points (Pfeiffer et al. 2008).

Kriging is determined by the semivariogram, the distance to the prediction location and the spatial relationship between measurements around the prediction location. The empirical semivariogram shows the spatial dependence in the variable of interest as a scatterplot. Distance (spatial lag) is presented on the $x$-axis and semivariance on the $y$-axis. The semivariance is calculated as follows (Pfeiffer et al. 2008):

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} [Z(S_i) - Z(S_j)]^2$$

where $Z(S_i)$ is the measured value at the $i$th location, $N(h)$ is the set of distinct pairs of values separated by distance $i - j = h$, $|N(h)|$ is the number of distinct pairs in $N(h)$, $z_i$ and $z_j$ are data values at locations $i$ and $j$, respectively.

The semivariogram, a graph of semivariance plotted against separation distance $h$, conveys information about the continuity and spatial variability of the process. If observations close together are more alike than those farther apart, the semivariance increases as the separation distance increases, reflecting the decline of spatial autocorrelation with the distance (Chen et al. 2007).

The semivariogram is characterized by particular parameter i.e. sill ($c_0$), range ($\alpha_0$), and nugget ($c_n$). The sill is the semivariogram value where the model first flattens or upper bound of semivariogram. The range is the distance at which the semivariogram reaches the sill value (SAS Institute 2011). Sample locations that are separated by distance closer than the range are spatially correlated, while the distance of sample location farther than the range are not. In theory, the semivariogram value at the origin (0 lag) should be zero. If it is significantly different from zero for lags very close to zero, then this semivariogram value is
referred to as the nugget. The nugget represents variability at distances smaller than the typical sample spacing measurement error (Bohling 2005).

![Graph of Semivariogram](image)

Figure 5. The Graph of Semivariogram (SAS Institute 2011)

The semivariogram is central to geostatistic and essential for most geostatistical analysis. It compares the similarity pairs of points a given distance and direction apart (the lag) (Fischer et al. 1996). The range identified the maximum distance at which spatial autocorrelation was detected among the sampled counties. The nugget quantified the minimum variability at a lag distance of zero, whereas the sill quantified maximum variability among spatially independent samples (Yabsley et al. 2005). The nugget describes the spatially uncorrelated variation in the data. The larger of nugget value show the less spatial dependence (Pfeiffer et al. 2008).

The module of ArcGIS 9.2 is used to carry out the ordinary kriging analysis and developed the prediction map based on the number of JD seropositive. The kriging method was ordinary, semivariogram model was spherical, search radius type was variable, and the number of points was 12. The analysis extent was same as South Kalimantan province, and for advanced parameters was setting based on the semivariogram result.
3.6 Spatial Cluster Analysis

The aim of spatial cluster is to detect JD cluster in South Kalimantan province. It is to perform the geographical surveillance of disease, to test whether the spatial distribution of JD was homogeneously distributed over the region or clustered in space, and to evaluate any identified spatial disease cluster for statistical significance.

The SaTScan software version 9.0 was used to identify and locate significant spatial clusters of JD in South Kalimantan province. SaTScan is a free software that analyzes spatial, temporal, and space-time data using the spatial, temporal, or space-time scan statistics. It is designed for any of the following interrelated purposes: 1) perform geographical surveillance of disease, to detect spatial or space-time disease clusters, and to see if they are statistically significant; 2) test whether a disease is randomly distributed over space, over time or over space and time; 3) evaluate the statistical significance of disease clusters alarms; and 4) perform prospective real-time or time-periodic disease surveillance for the early detection of disease outbreaks. The spatial scan statistic imposes a circular window on the map. The circle is centered on each of the points (Kulldorff 2010).

SaTScan uses a likelihood test for the number of cases found in the study region population (the null hypothesis) to a model that has different disease risk depending on being inside or outside a circular zone. The test can be applied to both point and area data (Stevenson 2009). The program tests the hypothesis that animals within a particular window have the same risk of being seropositive as animals outside the window. Primary cluster is cluster with the largest likelihood ratio, while secondary cluster is clusters with smaller likelihood ratio (Jaime et al. 2005).

To conduct JD cluster analysis, JD cases are defined as the number of JD seropositive in the sub district level and the population at risk is the total cattle number per sub district. Data were assumed to be located at the center of each sub district. Data were created and imported to the SaTScan. Data consist of case files (sub district, number of cases), population files (sub district, year, number of cattle population), and coordinate files (sub district, x-coordinate, y-coordinate).
In this study, the type of analysis was purely spatial; probability model was Discrete Poisson Model, and scan for areas with high rates. With the discrete Poisson model, the number of cases in each location is Poisson-distributed. Under the null hypothesis, and when there are no covariates, the expected number of cases in each area is proportional to its population size (Kulldorff 2010). Cluster analysis was set with the maximum cluster size of <50% of the population at risk with a circular window space. Significant clusters were identified using Monte Carlo simulation with maximum number of replication was 999. Primary and secondary cluster were searched and described by center and radius of the cluster. No overlapping of the circles was allowed. Clusters were imported and mapped using ArcGIS software (ESRI Inc).

3.7 Factors Associated with the Spatial Distribution of Jembrana Disease

JD usually occurs at the end of dry season or at the beginning of the wet season. However, it is important to identify the relationship between JD seropositive and its related factors. A multiple regression analysis was used to identify factors that influence the risk of JD being present or absent at specific location using binary data i.e. positive (disease present) and negative (disease absent).

The analysis used regression by Pezzulo (2011) version 05.07.20, free statistic software for research development and education. The program generates the coefficients of a prediction formula (standard error of estimate and significance levels) and odd ratios with 95% confidence levels. The odd of an event is defined as the possibility of the outcome event occurring divided by the probability of the event not occurring. The odd ratio for a predictor tells the relative amount by which the odd of the outcome increase (OR greater than 1.0) or decrease (OR less than 1.0) when the value of the predictor value is increase by 1.0 unit (Pezzulo 2011).

There are one non-geographical variable (cattle density) and two geographical variables (elevation and distance to the river) were included in multiple regression. The map of elevation and cattle density were overlay with the location of JD seropositive to investigate the association between JD seropositive
and its related factors. A buffer of area was also created around main rivers in
South Kalimantan to determine distance sample location to river.

Cattle density is defined as the number of cattle in kilometer square. It was
grouped into five classes: less than 5 heads / km², 5 – 10 heads / km², 10 – 20
heads / km², 20 – 30 heads / km², and more than 30 heads / km². Elevation data
were obtained as digital elevation model and grouped into five classes: less than
25 meters, 25 – 50 meters, 50 – 100 meters, 100 – 300 meters, and more than 300
meters above sea level. To determine the distance to the river, a buffering was
created with 500 meters, 1000 meters, 5000 meters, and 10,000 meters around the
main rivers.