# Clustering and Modelling Beef Cattle Population in Probolinggo District with Spatial Autoregressive Method (SAR)

Henny Pramoedyo[1], Vita Dewi Islami[2]
{vitabubble17@gmail.com[2]}

Universitas Brawijaya, Indonesia[1, 2]

**Abstract**. Probolinggo is an area in East Java Province that has a large potential of livestock resources for the development of beef cattle breeding business. National demand for meat increases in line with the rate of economic growth which is getting better, the rate of population growth, and the increase of awareness in the importance of consuming nutrients from livestock. For this reason, the government puts a target in achieving the program of accelerating beef self-sufficiency (P2SDS) in Indonesia. The purpose of this research is to clustering beef cattle population in Probolinggo Regency and to obtain a model of beef cattle population distribution using SAR models. Beef cattle population (y), area of corn plant (x) are used as research variables. According to Spatial Clustering of beef cattle population conducted by Morans' I and Geary's, it finds out that Tongas and Wonomerto sub districts are categorized as the highest beef cattle population which are located around areas which has high number of beef cattle population. On the other hand, Maroon and Pajarakan Sub-districts have low beef cattle populations and are located around areas having low beef cattle as well. This research using the analysis of Spatial Autoregressive (SAR) shows that there is spatial dependence among districts. All independent variables are significant at the 5% level with the value of $R^2$ as many as 41,9%.

**Keywords:** Spatial Clustering, Spatial autoregressive (SAR).

## 1 Introduction

Regression analysis is one of the statistical analysis that studies the relationship between response variables and predictor variables, but along with the research developments, researchers are faced with the fact that space or spatial effects cannot be ignored, so it takes a method that can consider the spatial effect. If there is data with spatial effect then the analysis is used spatial regression analysis [1]. In spatial modelling, a matrix of spatial weights is developed to determine the spatial relationship that occurs between regions one and another. The spatial weighting matrix is obtained with regard to the alignment between one region to other regions in accordance with the Regional division map [2]. One of the models of spatial regression is the Spatial Autoregressive (SAR) that combines linear regression models with spatial lag in response variables using cross section data and is assumed that the autoregressive process only Occurs in the response variable [1]. SAR is one of the spatial modelling related to the area approach.

IISS 2019, October 23-31, Malang, Indonesia

DOI 10.4108/eai.23-10-2019.2293022

This spatial regression is widely used in various areas and one of them is in area of farm. One of the efforts that the Government can do in increasing the number of cattle population is the main target to be implemented in order to support the acceleration of beef self-sufficiency (P2SDS) program in Indonesia [3]. The need for national meet demand is increasing as the rate of economic growth improves the rate of population growth, and increased awareness of the importance of consuming nutrients from livestock [4]. In previous studies, Winarso *et al*, showed that the utilization of corn crop as feed material was one of the efforts in raising beef cattle in East Java [5].

Probolinggo is one of the areas in East Java province that has a potential for a large farm resource for the development of cattle farming business [6]. Researchers suspect that the geographical condition in Probolinggo district affects the amount of beef cattle population, so in analyzing the case uses spatial regression. The spatial cluster as a group of events is geographically restricted to the region [7]. There are several methods that can be used for the analysis of spatial clusters among others Moran's I and Geary's C. The purpose of this research is to clustering the beef cattle population according to sub-districts in Probolinggo, and obtain a model of cattle population distribution. Based on the research objectives above the government is expected to be able to take appropriate steps in achieving the target and handle the cow waste to be produced

## 2 Research Method

The data used in this research in is the form of secondary data obtained from the central Statistics Agency (BPS) of Probolinggo Regency and it is the data on the number of cattle population in 2018 [8]. The observation unit (termination) used in this research is 24 sub-districts in Probolinggo district. Map of the Division of Administrative territory of the subdistrict in Probolinggo presented in Figure 2.

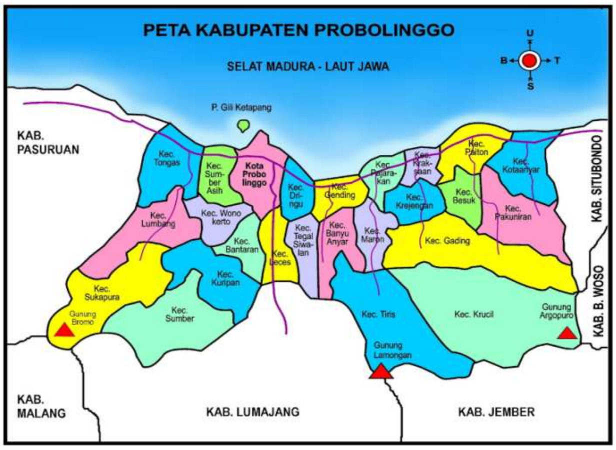

**Fig. 1.** Map of Probolinggo Regency.

The variables are used in the study includes dependent variables and independent variables. Dependent variable (Y) is the population of beef cattle. The independent variable is comprised of an area of corn plant (X). The analysis is used spatial clustering to map the distribution patterns of beef cattle populations and spatial regression to the modelling of beef cattle populations.

The Software is used in this research is GeoDa. The method and stages of analysis are used in this study are as follows [9]:

a) Conducting descriptive data analysis or exploration of beef cattle population data in Probolinggo district.

b) Specifying a spatial weight matrix (W) i.e. by using the Rook Contiguity method.
Create a spatial weighting matrix based on the interaction of Rook Contiguity, which is a matrix is used to describe inter-location relationships based on the interside contact of the thematic map. Common form of spatial weighting matrix is:

$$W_{ij} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n1} & \cdots & W_{nn} \end{bmatrix} \quad (1)$$

The value of $W_{ij}$ will be worth 0 when I and J are not mutually-side and if I and j are mutually-determined then $W_{ij}$ will be worth 1 [10].

c) Investigate the presence of autocorrelation between regions by doing the Moran'I test.
Hypothesis of *Moran'I test*:
$H_0$: I = 0 (no spatial autocorrelation)
$H_1$: I ≠ 0 (there is a spatial autocorrelation)
The statistics of Moran's I can be measured by the following formula [11]:

$$\boldsymbol{I} = \frac{\boldsymbol{N}}{\boldsymbol{S_0}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W_0} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (2)$$

Where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} W_0$.
If the index value I is approaching 1 indicates that it occurs grouping a region with the same characteristics or can be said to occur positive spatial autocorrelation.

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \quad (3)$$

Acquired rejection decision $H_0$ if $|Z(I)| > Z_{\alpha/2}$ is interpreted as a spatial autocorrelation between regions. If the index value I is approaching 1 indicates that it occurs grouping a region with the same characteristics or can be said to occur positive spatial autocorrelation [12].

d) Conducted global test of Moran's I and Geary's C to map the spread of beef cattle populations.
Scatter plot distribution of Moran's I as the tool for exploratory spatial data analysis (ESDA) assess spatial patterns and classifications [13].

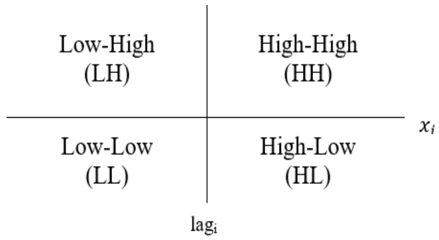


**Fig. 2.** The Moran Scatter Plot.

a) The quadrant I is called High-High (HH), indicating the area with high observation value is surrounded by areas with high observation values.
b) Quadrant II is called Low-High (LH), indicating an area with a low observation but is surrounded by areas with high observation value.
c) Quadrant III is called Low-Low (LL), indicating areas with low observational values and is surrounded by areas that also have low observation values.
d) Quadrant IV is called High-Low (HL), indicating areas with high observation values are surrounded by areas with low observation values.

Moran's Scatterplot which put many observations in the HH quadrant and LL Quadrant will likely have a positive spatial autocorrelation value (clusters). Whereas Moran's Scatterplot put many observations on the HL and LH quadrants will likely have a negative spatial autocorrelation value [13].

Geary's C is an alternative spatial grouping measurement that takes on the form of a familiar cross-product equation [14]. The similarity of two locations is quantifiable as the difference between the values in each squared location.

Hypothesis:

$\mathbf{H_0}$: c = 1 (no spatial autocorrelation)

$\mathbf{H_1}$: c ≠ 1 (there is a spatial autocorrelation)

Standard index and test statistics [15]:

$$\mathbf{c} = \frac{(\mathbf{n}-\mathbf{1})}{\mathbf{2S_0}} \frac{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \mathbf{W_0}\left(\mathbf{x_i}-\mathbf{x_j}\right)^2}{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} (\mathbf{x_i}-\bar{\mathbf{x}})^2} \tag{4}$$

$$\boldsymbol{Z(C)} = \frac{\boldsymbol{I-E(C)}}{\sqrt{\boldsymbol{Var(C)}}} \tag{5}$$

If the c < 1 is a positive spatial autocorrelation indicating that adjacent locations have similar or likely group values *(Clustered)* and c > 1 negative spatial autocorrelation that adjacent locations have different values.

Identify spatial effects with the Lagrange Multiplier (LM) test. The spatial dependency test at lag can be used by the test Lagrange Multiplier (LM).

Hypotheses of the LM test are:

$H_0$: ρ = 0 (there is no spatial dependency on lag)

$H_1$: ρ ≠ 0 (there is spatial dependency on lag)

LM's test statistics are [16]:

$$LM = \frac{(\varepsilon' Wy/\hat{\sigma}^2)^2}{D+T} \sim \chi^2(1) \quad (6)$$

The criterion for decision-making is to reject H0 if $LM > \chi^2(1)$.

e) Modelling spatial regression with SAR or SEM as appropriate from the LM test results.
Spatial autoregressive (SAR) is a combination of a simple regression model with spatial lag that uses data cross section [10]. General SAR models are:

$$y = \rho W_1 y + X\beta + \varepsilon \quad (7)$$

Where $\varepsilon \sim N(0, \sigma^2 I)$.
Description:
Y : variable vector dependent
X : Variable Matrix Independent
β : vector parameter regression coefficient
ρ : different spatial coefficient of kala (lag) dependent variable
ε : vector error
$W_1$: Spatial weighting matrix.

f) Interpreting and concluding the results obtained.

## 3 Result and Discussion

### 3.1 Descriptive Analysis

The number of cattle populations in Probolinggo district in 2018 is presented in Figure 3.

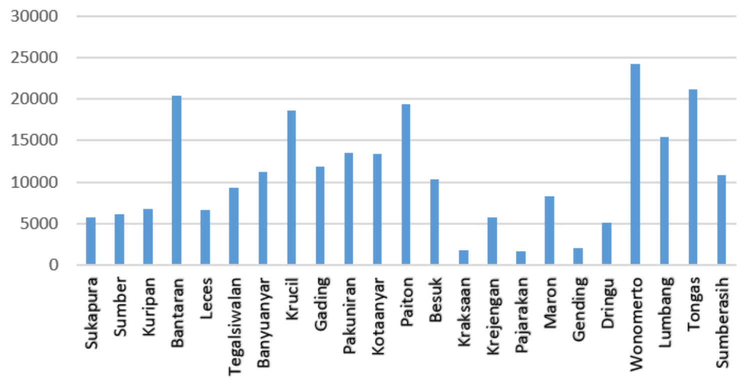


**Fig. 3.** Histogram of beef cattle population

Based on Figure 3 it shows that the number of beef cattle population in Probolinggo district in 2018 was the highest in the Wonomerto subdistrict as much as 24203, while the lowest beef cattle population was in Pajarakan subdistrict as many as 1661. The total number of beef cattle populations in Probolinggo district in 2018 was 249857.

### 3.2 Spatial Autocorrelation Test

To identify whether there is spatial autocorrelation on the number of beef cattle populations in the regency of Probolinggo in 2018 using global statistics Moran's I and Moran's scatter plot.

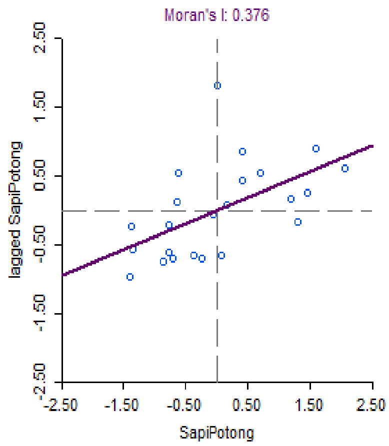


**Fig. 4.** Results of Moran's Scatter Plot.

Based on Figure 4 above, many observation points are placed in the HH quadrant and the LL Quadrant will likely have a positive spatial autocorrelation value or form a cluster pattern. In addition, Moran's index value of 0.376 is positive, so there can be positive autocorrelation. Then the number of cattle populations in Probolinggo District in 2018 is grouping or cluster. So, it can be concluded that the beef population in Probolinggo district in 2018 has the influence of its closest neighbors and has cattle the pattern of the group.

### 3.3 Spatial Cluster

To know the most regional groups of beef cattle population use Moran's I and Geary's C Test. Areas based on grouping can be seen in Figure 5.

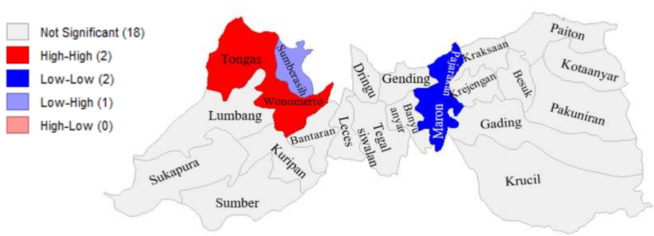


**Fig. 5.** Grouping Beef Populations with Moran's I.

Based on Figure 5 shows that the grouping region has the highest beef cattle population and is around the area that has high cattle cows as well, there are in the sub-districts of Tongas and Wonomerto amounting to 21,129 cows and 24,203 cows. The two areas are in the High-High quadrant (HH).

Sumberasih Sub-district is located on the Low-High quadrant (LH) which means that in this location has a slight number of cattle populations but is around the area that has a lot of beef cattle of 10,851 cows. Maroon and Pajarakan sub-districts are in the Low-Low quadrant (LL) meaning that the sub-district based on grouping region has a slight cattle cow population and is around the area that has a cow abbatoir. Maroon subdistrict was 8321 and Pajarakan Sub-district of 1661 cows. Grouping spatial beef done with Test Geary's C can be seen on figure 6.

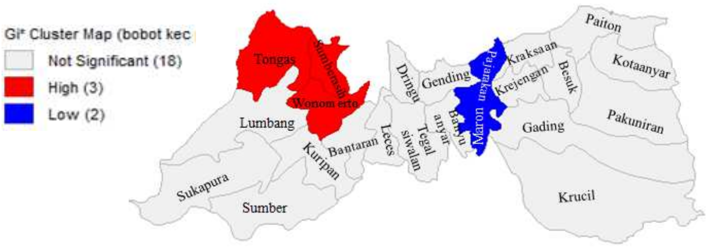


**Fig. 6.** Grouping Beef Populations with Geary's C Test

Based on Figure 6 shows that the number of cattle population in Probolinggo sub-district, based on grouping region with the Geary's C method which has the highest beef cattle population and is around the area that has beef cattle High, there are in the sub-district of Tongas, Wonomerto and Sumberasih with an average population of 18,728 cowss. While the sub-district based on grouping region has a slight beef population and is around the area that has a little beef cattle also in Maroon and Pajarakan sub-districts with an average beef population of 4,991 Cows.

### 3.4 Spatial Regression Modelling with SAR Methods

#### 3.4.1 Spatial Dependency Test

The spatial dependency test indicates that the probability or P-value of LM (lag) is 0.034 < 0.05, so the decision taken is to reject the meaningful $H_0$, and the appropriate model to use is SAR.

**Table 2.** *Spatial Dependency Test*

| Spatial Dependency Test | Score | Prob |
|---|---|---|
| Lagrange Multiplier (Lag) | 4.493 | 0.03 |
| Lagrange Multiplier (Error) | 3.032 | 0.08 |
| $\alpha = 0.05$ | | |

According to table 2, that value-p Lagrange Multiplier lag of 0.035 is significant with then there is a dependency of spatial lag. While the Lagrange Multiplier error is not significant then there is no dependency of spatial error. Therefore, it is more appropriate to use a spatial lag or SAR regression by incorporating a spatial dependency effect into a model.

**3.4.2 SAR Modelling**

In SAR modelling, parameter estimation is done with Maximum Likelihood Estimation (MLE). The result of the estimation of the (Y) beef population SAR parameter with an area of corn plant (x) is presented in table 3.

**Table** 3. Estimation Parameters of SAR

| Parameters | Estimated Results | $z - value$ | Probability | $R^2$ |
|---|---|---|---|---|
| constant | 4216.48 | 1.82 | 0.06 | 41,9% |
| $\rho$ | 0.42 | 2.14 | 0.03 | |
| $\beta$ | 1.2 | 2.8 | 0.007 | |

Based on table 3 that the area of corn plant (X) has significant effect on the number of beef cattle population (Y) with α = 5%. The coefficient of the area variable of corn plant is positively marked, which illustrates that if the area of corn plant expanded it will increase the number of cattle populations in Probolinggo district. In addition to the independent variable's coefficient, there is a new coefficient of ρ because it uses SAR models. The regression model formed on the number of cattle populations in Probolinggo district uses the Spatial Autoregressive model (SAR) is as follows.

$$\hat{y} = 4216.48 + 0.42W_y + 1.2\ \mathrm{X}$$

## 4 Conclusion

Clustering spatial beef cattle population are carried out with Moran's I and Gear's C tests obtained almost the same results, that sub-districts have the highest beef cattle population and are around areas that have high beef cattle as well, In the sub-district of Tongas and Wonomerto. Meanwhile, the sub-district has a low number of beef populations but is located around the area that has a lot of beef cattle in Sumberasih sub-district. Maroon and Pajarakan sub-districts have a low beef cattle population and are located around the area that has low beef cattle as well. The number of beef cattle populations in Probolinggo district can be modelled with the SAR model. The area of corn plants (X) affects the number of cattle (Y) populations which means that if the area of corn plant is expanded it will increase the number of cattle populations in Probolinggo district.

### 4.1 Suggestions

The advices that can be given is due to data limitation so it is expected in subsequent research can be done addition of other variables. Government Program that must be fulfilled in the target of beef self-sufficiency in Indonesia, the government should also consider the waste is produced. Bad waste treatment can cause poor consequences such as decreased environmental quality, odor that is not tasty and can cause health problems in humans. So, that the areas that will be dug deeper about the potential resources of these farms should be more concerned about the techniques of waste treatment and the level of hygiene in the environment.

## References


[1] L. Anselin, *Spatial Econometrics*. Dallas: University of Texas, 1999.

[2] N. M. S. Jayanti, I. W. Sumarjaya, and M. Susilawati, "Pemodelan Penyebaran Kasus Demam Berdarah Dengue (DBD) di Kota Denpasar Dengan Metode Spatial Autoregressive (SAR)," *E-Jurnal Mat.*, vol. 6, no. 1, pp. 37–46.

[3] Kementerian Pertanian, *Lolit Sapi Potong*. Jakarta: Kementerian Pertanian, 2018.

[4] S. Umela, "Spatial Patterns of Carrying Capacity Development for Beef Cattle Production in Pohuwato Regency Gorontalo Sulawesi," *J. Disertasi Geogr.*, vol. 4, no. 2, pp. 13–25, 2013.

[5] B. Winarso, R. Sajuti, and C. Muslim, "Tinjauan ekonomi ternak sapi potong di Jawa Timur," in *Forum Penelitian Agro Ekonomi*, 2016, vol. 23, no. 1, pp. 61–71.

[6] Kementerian pertanian, *Potensi Swasembada Daging Sapi, Balitbangtan Dampingi Peternak Probolinggo*. Jakarta: Badan Penelitian dan Pengembangan Pertanian, 2019.

[7] P. L. Knox, S. A. Marston, and M. Imort, *Human geography: Places and regions in global context*. Pearson New York, 2016.

[8] BPS, *Probolinggo Dalam Angka 2018*. Probolinggo: Badan Pusat Statistik, 2018.

[9] M. D. Ward and K. S. Gleditsch, *Spatial regression models*, vol. 155. Sage Publications, 2018.

[10] J. P. LeSage, "The theory and practice of spatial econometrics," *Univ. Toledo. Toledo, Ohio*, vol. 28, no. 11, 1999.

[11] J. Lee and D. W. S. Wong, *Statistical analysis with ArcView GIS*. John Wiley & Sons, 2001.

[12] J. R. Anselin, Luc; Sergio, *Perspective on Spatial Data Analysis*. New York: Springer, 2009.

[13] M. M. Fischer and A. Getis, *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Science & Business Media, 2009.

[14] R. C. Geary, "The contiguity ratio and statistical mapping," *Inc. Stat.*, vol. 5, no. 3, pp. 115–146, 1954.

[15] A. J. Lembo, "Spatial autocorrelation," *Cornell Univ.*, 2006.

[16] L. Anselin, *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media, 2013.