Estimation of Paddy Productivity at Subdistrict Level using Geoadditive Small Area Estimation Model in Ponorogo Regency

Arswenda Putra Maulana[1], Rindang Bangun Prasetyo[2*]
[1]Computational Statistics, STIS Polytechnic of Statistics, Jakarta, Indonesia
[2]Computational Statistics, STIS Polytechnic of Statistics, Jakarta, Indonesia
*Corresponding author: rindang@stis.ac.id.

Received: 31 May 2024
Revised: 23 September 2024
Accepted: 31 October 2024

ABSTRACT – **Paddy is the most important food crop in the world and it is the source of food needed by more than half of the population on a global scale. However, the world is experiencing the threat of a food crisis, so the Indonesian government continues to be committed to increasing national paddy production and ensuring food sufficiency in the country by implementing food self-sufficiency programs in each region. Paddy productivity data can be used as one of the government's benchmarks to assess the success of the food self-sufficiency program, but BPS-Statistics Indonesia only provides data on paddy productivity up to the district/cities level. Therefore, this study aims to estimate paddy productivity at sub-district level using the Geo-SAE method. Based on the research results, the estimation of the average paddy productivity in Ponorogo Regency in 2022 using Geo-SAE was obtained at 5.8 tons/ha and resulted in a smaller RSE value compared to the direct estimation at sub-district level. This indicates that the Geo-SAE method has better precision than the direct estimation method. Meanwhile, additional result from estimation of paddy productivity shows that in Ponorogo Regency in 2022 there is a large rice surplus. Therefore, it can be said that Ponorogo Regency is experiencing a very good food sufficiency condition.**

Keywords– **Geo-SAE, paddy, paddy productivity, food self-sufficiency.**

I. INTRODUCTION

Paddy is the most important food crop in the world compared to other food crops as it is the main source of food and provides more than 20 percent of the calories needed by approximately half of the population on a global scale, especially in Asia and South America [1]. This is evident as global rice consumption has reached 521.37 million tons during the survey period from 2022 to mid-2023, according to a report by the United States Department of Agriculture (USDA) [2]. Meanwhile, according to data from the USDA's official website in 2022, global rice production reached 502.97 million tons [3]. Indonesia ranks fourth in terms of global rice production and consumption [3]. Based on SUSENAS 2022 results, 98.35% of households in Indonesia consume rice with an average per capita rice consumption of 82.87 kg per year [4]. However, based on the publication of the Executive Summary of Expenditure and Consumption of the Indonesian Population from 2017 to 2022, the amount of rice consumption in Indonesia shows an increasing trend [5]. Meanwhile, the amount of rice production shows a downward trend in the same period.

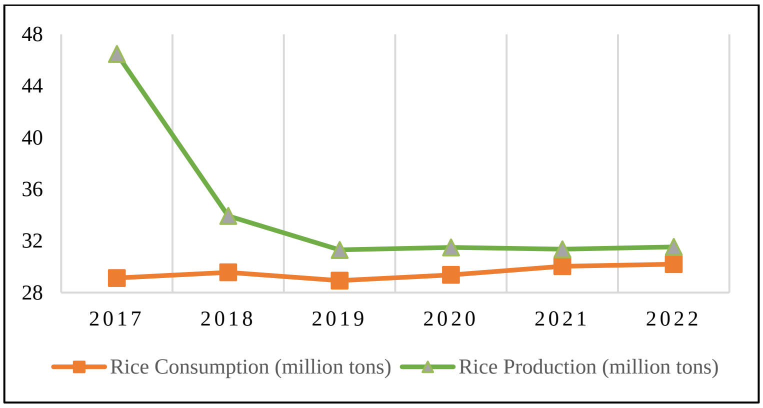

Figure 1. Rice production and rice consumption of the Republic of Indonesia

As one of the world's leading consumers and producers of paddy or rice, the Indonesian government continues to be committed to increasing national production to ensure domestic food sufficiency. However, based on data released by BPS in 2022, Indonesia's rice imports reached an amount of 429 thousand tons, an increase compared to the previous year which amounted to 407.7 thousand tons [5]. President Jokowi stated that Indonesia needs to import rice because it is difficult to achieve self-sufficiency and to maintain national rice reserves, especially with the increasing population growth and their need for rice. In addition, according to the President, rice imports are carried out because domestic rice production continues to decline [6].

DOI: 10.12962/j27213862.v7i3.20505

In Indonesia, the food self-sufficiency program has been a top priority in every government period involving various policies and initiatives to achieve sufficient food availability for the people of Indonesia and one of the programs is to increase rice production, which is the main food commodity in the country. According to Law No.18/2012 Article 12 Paragraph 2 on Food states that local governments are responsible for food availability in the region and the development of local food production in the region. Therefore, Indonesia's food self-sufficiency program requires each local government to be responsible for food availability and the development of local food production [8]. Data related to food crops such as paddy productivity data is very important because it can be a measure of success in the food self-sufficiency program which can provide a comprehensive picture of the development of production and problems that occur in food crops in each region. In addition, paddy productivity data can also be used to see how the condition of regional food fulfillment or in other words, regional food surplus and deficit, so that local governments really need support from paddy productivity data to the sub-district level [9]. With this data, local governments can identify the condition of food sufficiency in each region, so that they can easily find out which areas are prone to rice deficits.

BPS-Statistics Indonesia does not provide data on paddy production down to the sub-district level because the sample size used in the Ubinan Survey is not representative enough to estimate directly down to the sub-district level. The lack of sample size can be overcome with two solutions, namely, direct estimation and indirect estimation. One of the indirect estimation methods that can be used is small area estimation. Small area estimations utilize additional statistical information that has the property of borrowing information strength from the relationship between the variable that is concerned and the participating variable [10]. The SAE method can be used to estimate indirectly in small areas of paddy productivity data as has been done by several researchers, one of which is Ardiansyah (2018) and Santoso (2022) using the geoadditive small area model method which provides a relatively small error value [9] [11]. Geoadditive small area model is a model that combines the geoadditive model with SAE, namely by adding geo-spatial information to the classic SAE model in the form of a linear mixed model [9] [11]. Violations related to the assumptions of linearity and normality in the data can be handled with the geoadditive model [11]. Geoadditive models also have the ability to change the shape of the relationship between the response variable (y) and the predictor or accompanying variable (x) with high flexibility because the model is a form of smoothing function [12]. In this research, the estimation of paddy productivity was conducted in one of the cities in Java, namely Ponorogo Regency, East Java Province.

Based on BPS data, East Java Province is the province with the largest production in Indonesia, reaching 9.5 million tons in 2022 [13]. One of the regency/cities in East Java Province that contributes to the amount of paddy production and harvest area is Ponorogo Regency with paddy production reaching 370 thousand tons or if converted into rice, reaching around 214 thousand tons [14]. However, the amount of paddy production and paddy harvest area in 2022 in Ponorogo Regency has decreased compared to the previous year, this is in line with the decline in paddy production and harvest area in East Java Province. Meanwhile, the productivity value of the paddy crop in Ponorogo Regency is also still quite low when compared to other regency/cities [15]. This indicates that the strategies implemented so far have been less effective so that policy evaluation and improvement of the programs that have been launched are needed. Therefore, an identification step is needed through calculating paddy productivity data with a lower administrative level so that decision making, monitoring, and evaluation can be carried out more precisely and accurately.

In this study, the estimation of sub-district level paddy productivity data is carried out using the Geo-SAE model based on unit level. The results of the estimation with this method will be evaluated by estimating the RSE value using the bootstrap method. This method treats the sample as a population and then performs a random resampling process with returns. Furthermore, the results of estimating rice productivity using the Geo-SAE method will be used to calculate rice production. The goal is to identify food sufficiency conditions in each sub-district in Ponorogo Regency in 2022.

II. LITERATURE REVIEW

A. Paddy Productivity

Paddy productivity is a value that shows the average yield of paddy crops per unit of crop area in the reporting period. The estimation of paddy productivity figures is obtained from the Ubinan Survey. Since 2018, BPS has used the results of the KSA Survey in the determination of ubinan samples [16]. The use of KSA (Sampling Frame Area) basis in determining the sample of the sample is aimed at reducing the risk of missed harvests (non-response) so that the calculation becomes more accurate [16]. Suppose the weight of the results in kilogram (kg), the area of the sampled plot used is 6.25 m2, then the productivity of paddy plants is described as follows:

$$Paddy\ Productivity\ (GKP) = the\ weight\ of\ ubinan \times \frac{1\ hektare}{ubinan\ area} \tag{1}$$

BPS-Statistics Indonesia uses data on paddy productivity in the form of milled dry grain (GKG) as one of the indicators to measure paddy production in a region. However, since most paddy yields in the field are measured in the form of harvested dry grain (GKP), it is necessary to convert the data into GKG. The calculation of the conversion of grain to rice involves a conversion value from dry harvested grain (GKP) to milled dry grain (GKG) and a convertion value of milled dry grain (GKG) to rice. In 2018, BPS updated these two values through the implementation of the Grain to Rice Conversion Survey (SKGB) in two different seasonal periods, detailing data based on provinces. It aims to obtain conversion rates specific to each province. Here is a presentation of the process of conversion from straw to rice used as a resident food for the East Java Province.

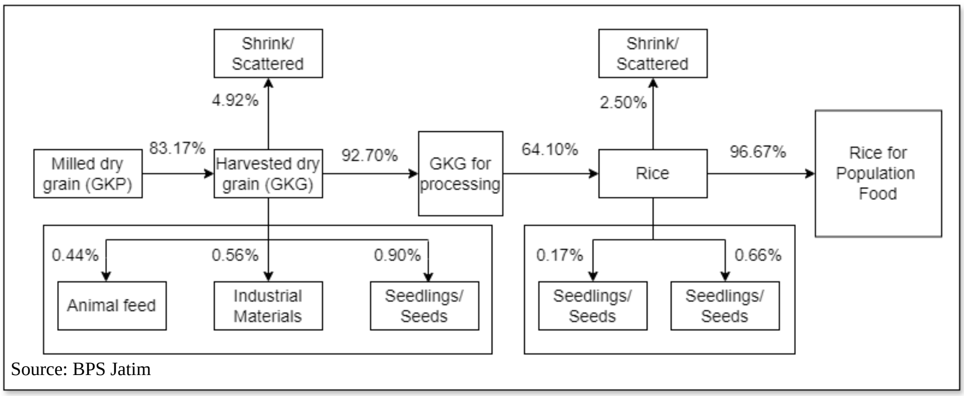

Figure 2. A stream of conversion of grain into rice in the East Java region

B. Geoadditive Small Area Estimation Model

Small Area Estimation (SAE) is a statistical technique for estimating subpopulation parameters on variables of interest with small sample sizes and insufficient precision [17]. The main advantage of small area estimation is that it improves the efficiency of a limited sample size by borrowing the strength obtained from inside and outside the area [9]. Presented as a linear mixed model, the geoadditive model is a technique of merging the kriging model and the additive model [18]. Geoadditive model offers a flexible regression connection when the interaction between the variables is complicated and cannot be readily handled by specific linear or nonlinear functions [19]. The Geo-SAE model is a model that combines the geoadditive model with SAE, namely by adding geo-spatial information to the classic SAE model in the form of a linear mixed model [9]. Violations related to the assumptions of linearity and normality in the data can be handled with the geoadditive model [9]. Geoadditive models also have the ability to change the shape of the relationship between the response variable (y) and the predictor or accompanying variable (x) with high flexibility because the model is a form of smoothing function [12]. The geoadditive model can be formulated in the form of a mixed linear model written as follows.

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{e} \tag{2}$$

with, $\mathbf{X} = [1, r_i, t_i, \boldsymbol{s}_i^T]_{1 \le i \le n}$ where r_i and t_i are continuous predictors of y_i at spatial location s_i , $\boldsymbol{\beta} = [\beta_0, \beta_r, \beta_t, \boldsymbol{\beta}_s^T]$, $\boldsymbol{v} = [v_1^S, \dots, v_{Ks}^S]$, $\boldsymbol{\beta}$ and $\boldsymbol{v}$ are parameter vectors, $\mathbf{Z}$ contains the joint matrix of the spline basis functions, $\mathrm{K}_1^s, \dots, \mathrm{K}_{Ks}^s$ are spatial location nodes or commonly called knots [20]. Since our next stage estimate may be framed as a penalized spline regression, Shan Yu, et al. (2019) conclude that estimation with geoadditive models is computationally quick and efficient [19].

Knots are focal points along a spline line or curve where two distinct segments meet. In other words, knots are points where the polynomials in the spline change [9] [11]. Knots are used to connect and shape curves from observational data. The standard for most individual observations is about 4-5 observations at each knot. If the number of datasets gets larger, the number of knots will also increase, so it is recommended to consider about 20-40 knots as the maximum number [21]. One of the methods used for selecting the optimum knots point is by using GCV (Generalized Cross Validation). The definition of GCV can be formulated as follows.

$$GCV\ (\boldsymbol{K}) = \frac{MSE\ (\boldsymbol{K})}{[n^{-1} trace(I - \boldsymbol{A})]^2} \tag{3}$$

With, $MSE(K) = n^{-1}\boldsymbol{y}^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{A}_{(K)}\right)^{\mathrm{T}}\left(\boldsymbol{I} - \boldsymbol{A}_{(K)}\right)\boldsymbol{y}$, where $\boldsymbol{K} = \mathrm{K}_1^s, \dots, \mathrm{K}_{Ks}^s$ are knots and $\boldsymbol{A}_{(K)} = \boldsymbol{C}^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \lambda\boldsymbol{D})^{-1}\boldsymbol{C}^{\mathrm{T}}$with $\boldsymbol{C} = [\boldsymbol{X}, \boldsymbol{Z}]$, $D = diag(\boldsymbol{0}_{\boldsymbol{p+1}}, \boldsymbol{1}_{\boldsymbol{K}})$, $\boldsymbol{A}_{(K)}$ called smoothing matrix [9]. To determine the optimum number of knots, this study uses the fixed selected method which can be formulated as follows [21].

$$\mathrm{K} = min\ (\tfrac{1}{4} \times a\ unique\ number\ of\ x_i, 35) \tag{4}$$

Meanwhile, the determination of the location of knots uses the x_k th quantile of the unique x_i. The following is the formula for determining the location of knots according to Ruppert, Wand, & Carrol in 2003 [21].

$$K_k = (\tfrac{K+1}{K+2}) \tag{5}$$

The main purpose of the knots selection method is to ensure that the number of knots (K) is large enough so that it is more flexible when controlling the smoothness of the estimated curve with smoothing parameters and so that the calculation time is not too long and the MSE is small [22].

Suppose there are m small areas to be estimated and y_{ij} denotes the response variable (y) where *i = 1,2, ...,m* and *j = 1,2, ..., n_i* which indicates that m is the number of small areas in a region and n_i is the sample size in each small area. Suppose there is also x_{ij} is a vector of accompanying variables (x) corresponding to the response unit, then the classic SAE model can be formulated as follows [17].

$$y_{ij} = \boldsymbol{x}_{ij}^{T}\boldsymbol{\beta} + u_i + e_{ij} \tag{6}$$

With, u_i is the *random effect* area which is normally distributed $u_i \overset{iid}{\sim} N(0, \sigma_u^2)$, e_{ij} is the individual level random error in the small area (sub-district)-i and sample unit (farmer)-j which is distributed $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, and u_i and e_{ij} are mutually independent both between individuals and between regions [17]. If the matrix is defined $\mathbf{D} = [d_{ij}]$, where:

$$\boldsymbol{d_{ij}} = \begin{cases} 1\ ; \ if\ the\ sample\ j\ is\ within\ a\ small\ area\ of\ the - i \\ 0\ ;\ for\ others \end{cases} \tag{7}$$

with, $\boldsymbol{y} = [y_{ij}]$, $\boldsymbol{X} = [\boldsymbol{x}_{ij}^{\mathrm{T}}]$, and $\boldsymbol{e} = [e_{ij}]$, then the matrix form of equation (6) is:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Du} + \boldsymbol{e} \tag{8}$$

Where, $E\begin{bmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$, $Cov\begin{bmatrix}u\\ e\end{bmatrix} = \begin{bmatrix}\sigma_u^2\boldsymbol{I}_m & 0\\ 0 & \sigma_u^2\boldsymbol{I}_n\end{bmatrix}$. Therefore, the variance matrix of **y** is $\boldsymbol{\Omega} = \sigma_u^2\boldsymbol{DD}^{\mathrm{T}} + \sigma_e^2\boldsymbol{I_n}$ [20].

So, equation (8) of the classic SAE model with the *Geoadditive* model in equation (2) can be used to estimate small areas called Geo-SAE. Since both *spline* models and SAE models can be viewed as models with random effects, it is possible to combine the two concepts in nonparametric small area estimation based on mixed linear models [23]. Equation (8) combines the *Geoadditive* and SAE models (Geo-SAE model) with two components of random effects.

Geo-SAE Model
$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zv} + \boldsymbol{Du} + e$

SAE Model
$\boldsymbol{y} = \boldsymbol{X_1\beta} + \boldsymbol{Du} + e$

Geoadditive Model
$\boldsymbol{y} = \boldsymbol{X_2\beta} + \boldsymbol{Zv} + e$

(10)

Because there are two types of accompanying variables in the geoadditive SAE model, the accompanying variable matrix becomes $\mathbf{X} = [1, \mathbf{x}_{ij}^{\mathrm{T}}, \mathbf{s}_{ij}^{\mathrm{T}}]_{1\le i\le m;\, 1\le j\le n_i}$ which is a fixed effect matrix of size $n \times (p+1)$, p is the number of accompanying variables (x) or fixed effects, $\boldsymbol{\beta}$ is a vector of fixed influence coefficients of size $(p+1)\times 1$, $\mathbf{Z}$ is the matrix of the spline-2 (thine-plate spline) of size $n \times K$ which is based on the radial basis function (modulus), namely $\mathbf{Z} = [C(s_{ij} - K_k)]_{1\le i\le m;\, 1\le j\le n_i;\, 1\le k\le K}$ with $C(\boldsymbol{s}) = \|\boldsymbol{s}\|^2\ log\|\boldsymbol{s}\|$, where K_k are knots, s_{ij} are spatial location coordinates, $\|\cdot\|$ is the notation of Euclidean norm, **v** is the spline-2 coefficient vector of spatial variables as spatial random effects of size $K \times 1$, **D** is a matrix of small area random effects of size $n \times m$ which is defined in equation (7), **u** is the coefficient vector of the area specific random effect of size $m \times 1$, and **e** is the individual level random error [9] [11].

The unknown variance component or variance can be estimated using REML (*Restricted Maximum Likelihood*) so as to obtain $\hat{\sigma}_v^2, \hat{\sigma}_u^2,\ dan\ \hat{\sigma}_e^2$ [17]. The estimated covariance matrix of y is as follows.

$$\widehat{\boldsymbol{\Omega}} = \hat{\sigma}_v^2\mathbf{ZZ}^{\mathrm{T}} + \hat{\sigma}_u^2\boldsymbol{DD}^{\boldsymbol{T}} + \hat{\sigma}_e^2\boldsymbol{I_n} \tag{11}$$

After that, the EBLUP estimator for $\beta, v, and\ u$ is as follows.

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X^T}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X^T}\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y} \tag{12}$$

$$\widehat{\boldsymbol{v}} = \hat{\sigma}_v^2\boldsymbol{Z^T}\widehat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{13}$$

$$\widehat{\boldsymbol{u}} = \hat{\sigma}_u^2\boldsymbol{D^T}\widehat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \tag{14}$$

The EBLUP estimator for the mean of the response variable (y) can be formulated as follows.

$$\bar{y}_i = \overline{\boldsymbol{X}_i}\widehat{\boldsymbol{\beta}} + \bar{\boldsymbol{z}}_i\widehat{\boldsymbol{v}} + \boldsymbol{d}_i\widehat{\boldsymbol{u}} \tag{15}$$

Where the vector $\boldsymbol{d_i}$ will be value 1 if within the area-i and value 0 if it is outside area-i [24].

C. MSE Parametric Bootstrap

MSE (Mean Square Error) is one way to show the level of precision of an estimator. In addition, the use of MSE is as a tool to compare between methods used to get the best and most accurate method [25]. The parametric bootstrap method will be used to calculate the MSE value so that the precision level of the estimation of paddy productivity data can be known. The bootstrap approach is the calculation of MSE by utilizing resampling techniques [17].

The parametric bootstrap rule is basically done by generating bootstrap sample data, namely $\{(y_{ij}^*, x_{ij});\ i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n_i$ [8]. The bootstrap sample data can be formulated as follows.

$$y_{ij}^* = \boldsymbol{x}_{ij}^T \widehat{\boldsymbol{\beta}} + \bar{\boldsymbol{z}}_i v_i^* + u_i^* + e_{ij}^* \tag{15}$$

Where the random influence v_i^* is generated from $N(0,\sigma_v^2)$, u_i^* generated from $N(0,\sigma_u^2)$, and e_{ij}^* generated from $N(0,\sigma_e^2)$ [9]. Suppose $\mu_i^* = \bar{\boldsymbol{X}}_i^T \widehat{\boldsymbol{\beta}} + \bar{\boldsymbol{z}}_i v_i^* + u_i^*$ be the bootstrap version of the target parameter $\mu_i = \bar{\boldsymbol{X}}_i^T \widehat{\boldsymbol{\beta}} + \bar{\boldsymbol{z}}_i v_i + u_i$. $\bar{X}_i$ is the parameter value of the accompanying variable at the i-th small area level, while $\bar{\boldsymbol{z}}_i$ is the spline-2 value at the i-th small area level. Using bootstrap data, the bootstrap version of the SAE geoadditive estimator $\hat{\mu}_i^G$ which is:

$$\hat{\mu}_i^{*G} = \bar{\boldsymbol{X}}_i^T \widehat{\boldsymbol{\beta}}^* + \bar{\boldsymbol{z}}_i \hat{v}_i^* + \hat{u}_i^* \tag{16}$$

Where $\widehat{\boldsymbol{\beta}}^*$, $\hat{v}_i^*$, and $\hat{u}_i^*$ are calculated as $\widehat{\boldsymbol{\beta}}, \hat{v}_i, dan\ \ \hat{u}_i$ in equations (11), (12), and (13) but using bootstrap generation samples. Theoretically, the bootstrap MSE of the Geo-SAE estimator repeated B times is formulated as follows.

$$MSE_B(\hat{\mu}_i^G) = E_*(\hat{\mu}_i^{*G} - \hat{\mu}_i^*)^2 = \frac{1}{B}\sum_{b=1}^{B}(\hat{\mu}_i^{*G}(b) - \hat{\mu}_i^*(b))^2 \tag{17}$$

RMSE (Root Mean Square Error) and RSE (Relative Standard Error) for each i-th small area (sub-district) can be calculated using the following formula.

$$RMSE(\hat{\mu}_i^G) = \sqrt{MSE_B(\hat{\mu}_i^G)} \tag{18}$$

$$RSE(\hat{\mu}_i^G) = \frac{RMSE(\hat{\mu}_i^G)}{\hat{\mu}_i^G} \text{ x } 100\% \tag{19}$$

III. METHODOLOGY

A. Variables and Data Source

The data used in this study comes from secondary data, namely in the form of raw data (raw data) of the 2022 Cropping Survey from BPS-Statistics Indonesia and data from the Ponorogo Regency Agriculture Office & National Disaster Mitigation Agency (BNPB) to estimate the level of productivity of paddy crops at the sub-district level in Ponorogo Regency. The Ubinan Survey is a survey conducted routinely every year by BPS-Statistics Indonesia in three periods which aims to obtain the value of food crop productivity [16]. The survey is designed to present data up to the regency level so that direct estimation at the sub-district level is not possible. Data on paddy harvest areas are obtained through reports collected by the KCD (Head of Service Branch) or Mantri Tani in each sub-district from the Agriculture Office. The collection of data on paddy harvest areas is conducted in a complete manner (census) through an area approach in all sub-districts which is reported monthly by the Agriculture Office. Meanwhile, National Disaster Mitigation Agency (BNPB) collects data on natural disasters such as floods and droughts by conducting direct monitoring and evaluation activities or 'picking up the ball' by the PDSI Division team which is then verified and validated. These activities are carried out over a period of six months, from January to June. The selection of research variables used is derived from a literature review of the Ubinan Survey and previous research and the availability of available data. The following table lists the variables used in this study.

Table 1. Research variable

Code	*Variable*	*Source*	*Year*
Y	Paddy productivity	Ubinan Survey	2022
X1	Proportion of paddy harvest area	Ubinan Survey & Agriculture Office	2022
X2	Climate impact	Ubinan Survey & BNPB	2022
X3	Latitude	Ubinan Survey & *G-Maps*	2022
X4	Longitude	Ubinan Survey & *G-Maps*	2022

B. Stages of Analysis

In this research, data processing will be carried out using software tools such as, Ms. Excel 2016 and R Studio. Meanwhile, the results of the data processing process will be analyzed through an approach, namely inferential analysis using the geoadditive small area estimation model method based on the unit level to estimate sub-district level paddy productivity in Ponorogo Regency. The following are the stages of analysis used in this research.

1). Data preparation stage
   a. Perform *data cleaning* process on 2022 Ubinan Survey data
   b. Determine the accompanying variables used
   c. Form paddy productivity variables using formula (1)
   d. Conversion of paddy productivity values from GKP to GKG based on figure (2)

2). Modeling stage for indirect estimation with geoadditive-SAE
   a. Determining the optimum number of knots using the fixed selection method with formula (4)
   b. Calculating the GCV value for each knot or node with formula (3)
   c. Calculating the spline-2 Z-matrix with formula (10)
   d. Estimating the coefficient vector value $\boldsymbol{\beta}, \boldsymbol{v}, dan\ \boldsymbol{u}$ with formulas (12), (13), and (14)
   e. Modeling and estimating paddy productivity with Geo-SAE
   f. Estimating the average paddy productivity in each sub-district
   g. Calculating MSE value with parametric bootstrap
   h. Calculate the RMSE and RSE values of paddy productivity estimation results with formulas (18) and (19)

3). Advanced analysis stage of indirect estimation results
   a. Calculating paddy production value based on Geo-SAE estimation results
   b. Calculating the value of rice production and rice consumption
   c. Identifying food sufficiency conditions at sub-district level

IV. RESULTS AND DISCUSSIONS

1) Overview of Research Variables

Ponorogo Regency is a regency located in East Java Province and directly adjacent to Central Java Province. Geographically, Ponorogo Regency is located between 7° 49' to 8° 20' South latitude and 111° 17' to 111° 52' East longitude with an area of 1,371.78 km². Based on data from the Ponorogo Regency Agriculture Office in 2022, the paddy harvest area reached 74,130 hectares. The following map presents the distribution of paddy harvest areas, spatial data, and climate impacts in Ponorogo Regency in 2022.

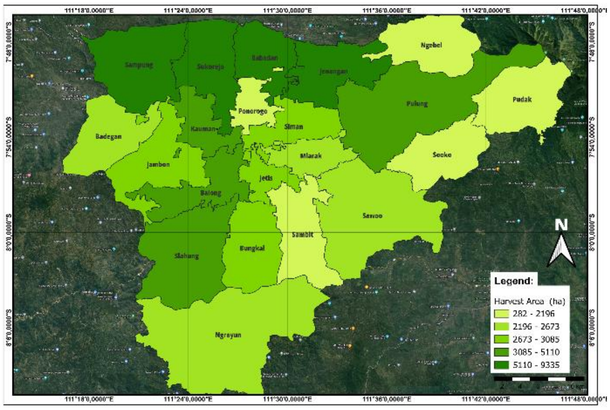

Figure 3. Distribution map of paddy harvest area

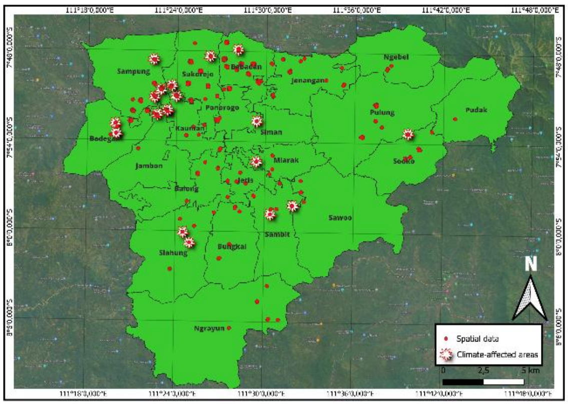

Figure 4. Distribution map of Spatial data and climate impacts

Based on Figure 3 above, it shows that the northern sub-districts tend to have a very large harvest area compared to the other sub-districts. The northern sub-districts include Jenangan, Babadan, Sukorejo, Sampung. Tobler (1970) [26] states the first law of geography which reads: *"Everything is related to everything else, but near things are more related than distant things"*. This concept is the basis for regional scientific studies, where spatial effects often occur between one region and another. Meanwhile, based on Figure 4 above, the red dots show the location of latitude and longitude based on sample data obtained from the Ubinan Survey in Ponorogo Regency in 2022. Sub-districts located on the north side such as Babadan, Sukorejo, and Sampung tend to have a large number of points, indicating that these sub-districts have a large number of samples. Meanwhile, the white-colored points indicate that the paddy fields contained at that point are affected by climate change in the form of floods or droughts.

2) Exploration of Relationships on Spatial Variables

Indirect estimation begins with an initial step that involves tracing the relationship pattern between the variable to be measured and the associated variables. The following figure presents a scatterplot to identify the relationship between the response variable and the accompanying variables. This exploration phase aims to investigate the relationship between the variables carefully, identify patterns that may be nonlinear, and identify additional variables that may potentially affect the response variable. This process is an important step in analysis as it allows a deeper understanding of the dynamics and interactions between variables, which in turn can help in formulating a more accurate and effective estimate model. The following is a scatterplot image to identify the relationship between the response variable and the accompanying variable.

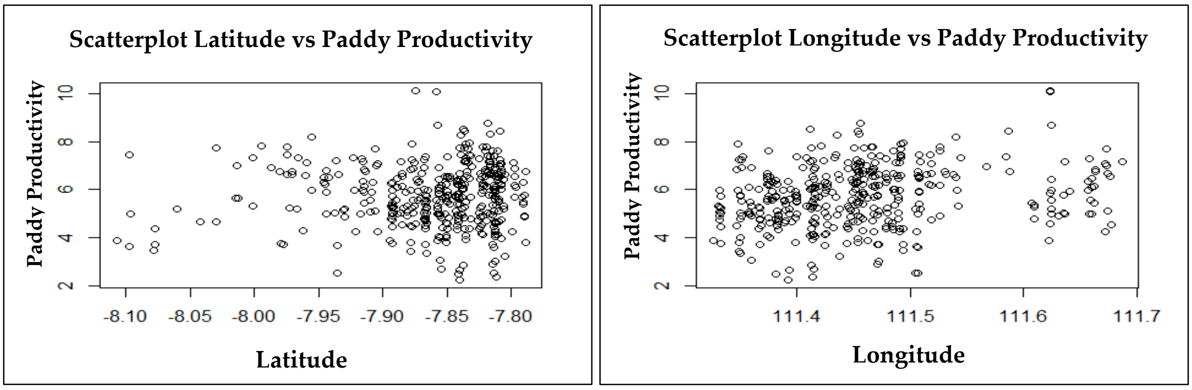

Figure 5. Scatterplot between latitude and longitude with paddy productivity

Based on Figure 5 above, it can be concluded that the relationship pattern formed between the spatial variables of latitude and longitude to paddy productivity does not follow a linear relationship pattern or in other words, the relationship pattern between these variables is non-linear which shows the behavior of changing data patterns. Therefore, spatial variables in the form of latitude and longitude are nonparametric variables. Therefore, this study uses the Geo-SAE method because it is nonparametric so that it can handle the problem of linearity assumptions on the spatial variables used.

3) Determining the Optimum Knot Point

To start modeling small area estimation using the Geo-SAE model, the first step is to determine the optimal number of vertices based on the minimum value of Generalized Cross-Validation (GCV). The smoothness of the spatial estimation surface curve is affected by the number of vertices applied to its radial basis function. Based on the predetermined selection method, the maximum number of knots used in this study is 35 knots. The following table presents the GCV value of each knot.

Table 2. Number of knot points based on minimum GCV value

Number of knot points	*GCV for latitude*	*GCV for longitude*
K=8	1.583	1.519
K=16	1.569	1.522
K=24	1.563	1.521
K=32	1.573	1.522

. Based on table 2 above, the minimum GCV calculation results show that the optimum number of knots used is 24 knots with a GCV value for latitude of 1.563 and a GCV value for longitude of 1.521. The following figure displays the pattern of the relationship between latitude and longitude variables with paddy productivity when approximated using spline-2 with 24 knots.

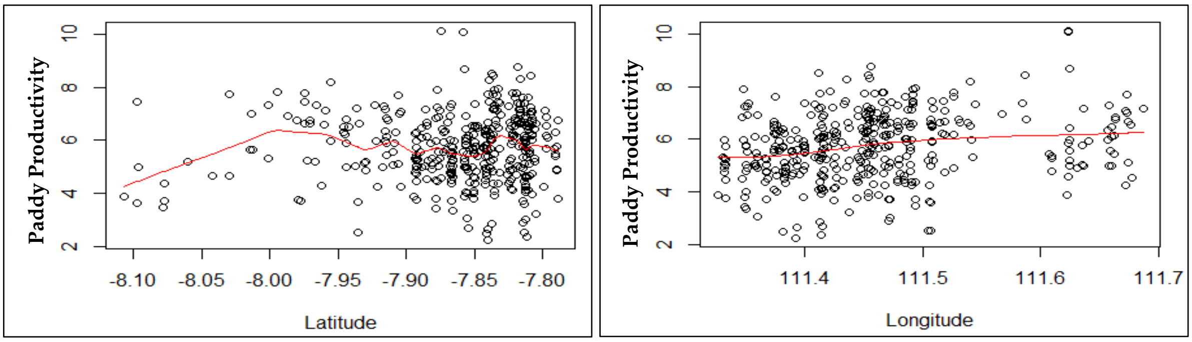

Figure 6. Plot between latitude and longitude with paddy productivity

4) Estimating Geo-SAE Model Parameters

The model is based on unit-level Ubinan Survey data, which is then used to generate parameter estimates in the mixed linear model equations, known as the Geo-SAE model. The Geo-SAE model consists of fixed effects as well as two random effects, namely the spatial random effect of the spline-2 modulus and the small area random effect. The table below shows the results of parameter estimation in the Geo-SAE model using four accompanying variables.

Table 3. Estimation results of geoadditive-SAE model parameters

Area	*Sub-district*	*Fixed Effects*	*Random Effects*	
			$\bar{z}_i\hat{v}$	$d_i\hat{u}$
010	Ngrayun		-0.319	-1.080 x 10^-9
020	Slahung		-0.272	5.935 x 10^-10
030	Bungkal		-0.192	4.917 x 10^-10
040	Sambit		-0.099	9.136 x 10^-10
050	Sawoo		-0.126	9.337 x 10^-10
060	Sooko		-0.217	-9.064 x 10^-11
061	Pudak		-0.110	6.956 x 10^-11
070	Pulung	β_0 = -420.814 β_1 = -1.169 β_2 = 0.217 β_3 = -0.134 β_4 = 3.818	-0.110	-7.530 x 10^-10
080	Mlarak		-0.055	-5.455 x 10^-10
090	Siman		-0.019	-4.693 x 10^-11
100	Jetis		-0.065	5.758 x 10^-10
110	Balong		-0.133	-1.179 x 10^-10
120	Kauman		-0.102	-6.988 x 10^-11
130	Jambon		-0.164	8.410 x 10^-10
140	Badegan		-0.247	-1.035 x 10^-9
150	Sampung		-0.226	1.786 x 10^-12
160	Sukorejo		-0.086	-2.929 x 10^-9
170	Ponorogo		-0.024	5.891 x 10^-10
180	Babadan		-0.039	1.007 x 10^-9
190	Jenangan		-0.038	1.705 x 10^-9
200	Ngebel		-0.169	-8.125 x 10^-11

Based on table 3 above, it can be seen that in the Geo-SAE model there are fixed effects and two random effects where $\bar{z}_i\hat{v}$ is the spatial random effect of spline-2 modulus, whereas $d_i\hat{u}$ is the small area random effect. It is important to note that the random effect values for each sub-district have varying estimates, indicating differences in the contribution of each sub-district to the variability in paddy productivity. Estimators of $\hat{v}$ and $\hat{u}$ are the empirical best linear unbiased predictor (EBLUP) of SAE.

Estimation is also done to estimate the variance of the random effect variance. There are two random effect variances estimated using the restricted maximum likelihood (REML) method, namely the variance of the spatial spline-2 random effect (σ_v^2) and the variance of the small area random effect (σ_u^2). The estimated value of the variance of the spline-2 spatial random effect is obtained as follows $\sigma_v^2 = 2.457$, while for the estimated value of the variance of the random influence of small area $\sigma_u^2 = 1.574 \times 10^{-10}$. From these values, it can be concluded that there is a significant variation in the random effects on the estimation results. The presence of variance values of random effects that are not zero indicates that random factors do have an influential impact on the estimation results, so modeling using the SAE method is feasible to take into account relevant spatial effects in the analysis.

5) Estimation Result of Paddy Productivity with Geoadditive-SAE Model

The estimation of paddy productivity at the sub-district level is obtained through the estimation of model parameters that have been compiled using the Geo-SAE method previously. The results of the calculation of sub-district level paddy productivity estimates in Ponorogo Regency in 2022 using Geo-SAE will be presented in the form of visualization in the form of bar charts.

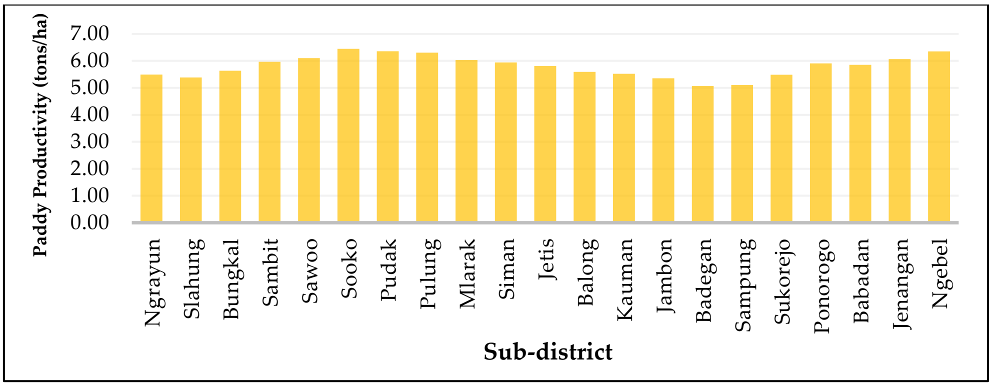

Figure 7. Estimation results of paddy productivity (GKG) with geoadditive-SAE model

Figure 7 above shows that the highest value of paddy productivity was produced by Sooko sub-district with a value of 6.447 tons/ha. Meanwhile, Badegan sub-district has the lowest productivity value, which is 5.074

DOI: 10.12962/j27213862.v7i3.20505

tons/ha. When viewed in Figure 3, the paddy harvest area, Sooko sub-district is classified as a sub-district with a low harvest area category. However, interestingly, Sooko sub-district actually recorded the highest paddy productivity value. This shows that Sooko sub-district has better farm business management compared to other sub-districts, as based on the Ubinan Survey, 70% of farmers in Sooko sub-district are members of farmers group. Meanwhile, Badegan sub-district has the lowest rice productivity due to the relatively small harvest area, coupled with rice plants that have suffered crop failure or damage due to climate impacts such as flooding or drought. According to the 2022 Ubinan Survey, there are several areas in Badegan sub-district that have indeed been affected by climate impacts such as flooding or drought. Overall, the average value of paddy productivity in Ponorogo Regency in 2022 reached 5.801 tons/ha.

6) Model goodness evaluation

Evaluation of the model quality is done by examining the Relative Standard Error (RSE) value in each sub-district. In this study, for indirect estimation, the MSE is calculated using the parametric bootstrap method, where the sample is considered as the population and observations are randomly extracted repeatedly as B=300 with replacement. At each iteration, the parameters are estimated and compared with the Geo-SAE estimates on the bootstrap data. The results are summed, squared, and then divided by the number of iterations to obtain the RMSE value. The following visualization images are presented in the form of graphs and tables of RSE calculation results.

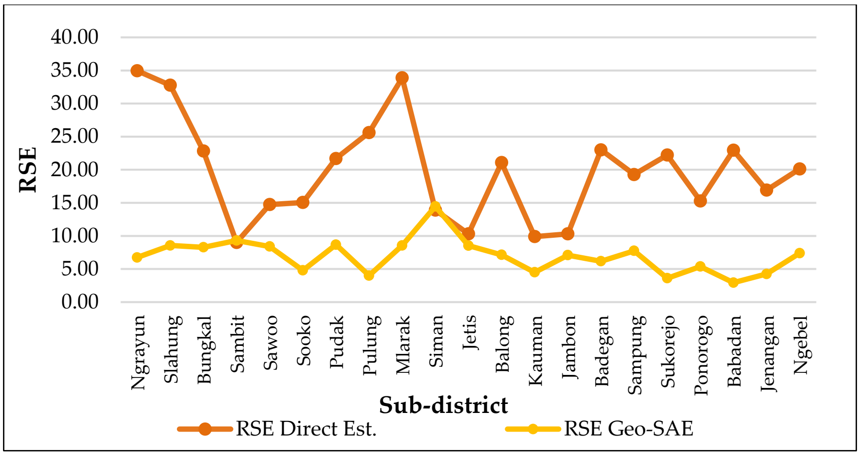

Figure 8. Comparison of RSE values between direct estimation and Geo-SAE

Table 4. Descriptive statistics of RSE values

Descriptive Statistics	*Direct Est.*	*Geo-SAE*
Range	25.936	11.525
Minimum	9.017	2.941
Maximum	34.953	14.466
Average	19.802	6.980

Based on Figure 8 above, it can be seen that the value of RSE results using the Geo-SAE model has a smaller value than using direct estimation. This shows that the results of the estimation of sub-district level paddy productivity in Ponorogo Regency in 2022 using the Geo-SAE model provide a relatively small error rate. The smallest RSE value when viewed from the graph and table above is in Sukorejo sub-district and Babadan sub-district, where the sub-district has a large enough sample size. This statement indicates that the parameter estimation method using Geo-SAE meets the requirements of a consistent estimator, where if the sample size is large, the variation of the estimate will be smaller. Therefore, the Geo-SAE model has a better level of precision compared to direct estimation. Meanwhile, when viewed also the RSE value of each sub-district using Geo-SAE also has an RSE value that is less than 25 percent so that it indicates that the results have met the eligibility standards of the estimates used by BPS-Statistics Indonesia. Meanwhile, based on Table 4 of the descriptive statistics above, it can be observed that the Geo-SAE model shows differences in the mean, maximum, and minimum values of the Relative Standard Error (RSE). It was found that the mean, minimum, and maximum RSE values using Geo-SAE were smaller than those using direct estimation.

7) Comparison Between Direct Estimation and Geo-SAE

Based on the results of the calculation of paddy productivity estimates at the sub-district level in the Ponorogo Regency 2022 using direct erimation and Geo-SAE, the following are presented visualization images in the form of line graphs.

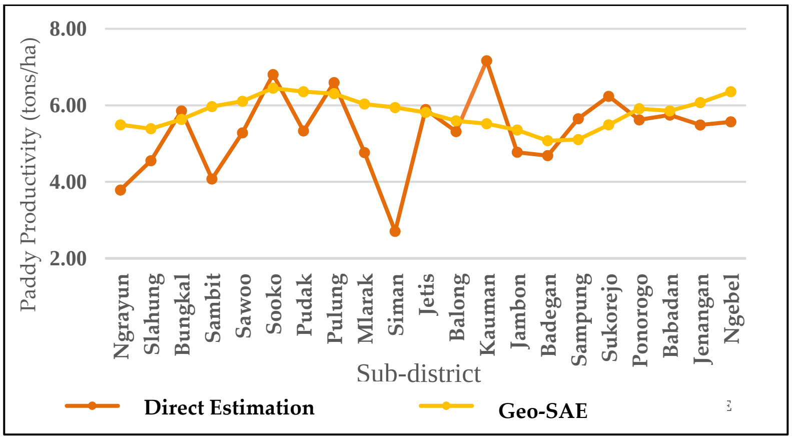

Figure 9. Comparison between direct estimation and Geo-SAE

Based on Figure 9 above shows that the results of paddy productivity estimates at sub-district level in Ponorogo region 2022 using Geo-SAE have a pattern that tends to be smoother compared to the direct estimates. In the Siman sub-district, the value of paddy productivity indicates that the output of the estimates has a low number of yields, because the number of samples present in the precipitation is so small that it does not present information about the location for all regions covered in the census data. On the contrary, using the Geo -SAE model, the yield value in the Simans sub-district has a larger value. This indicates the advantage of the Geos -SAE model in overcoming the relatively small size of the sample.

8) Identification of Food Sufficiency

To find out the condition of food sufficiency, namely rice, in each sub-district in Ponorogo Regency in 2022, it is necessary to calculate the amount of rice production for food for the population and the amount of rice consumption in each sub-district. The result of the multiplication between productivity and paddy harvest area in each sub-district will produce milled dry grain (GKG) paddy production data. Then, to produce the amount of rice production in each sub-district, it is done by using the calculation and conversion flow of East Java Province paddy production (GKG) into rice production for food for the population presented in Figure 1 above. Meanwhile, the rice consumption figures for each sub-district were calculated by using the per capita rice consumption figures for East Java Province multiplied by the total population.

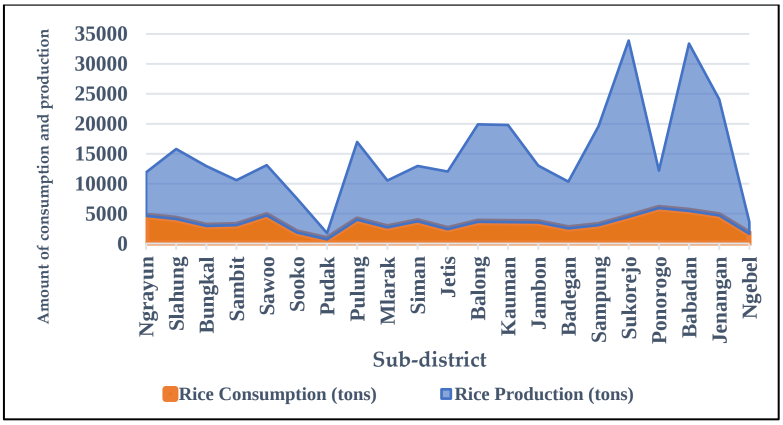

Figure 10. Comparison the amount of rice consumption and production

Table 5. Estimated rice consumption and production (tons)

Sub-district	*Rice Consumption*	*Rice Production*	*Ket.*
Ngrayun	4620.34	7293.58	Surplus
Slahung	4123.54	11730.44	Surplus
Bungkal	2948.86	9996.16	Surplus
Sambit	3062.22	7504.16	Surplus
Sawoo	4702.50	8370.88	Surplus
Sooko	1848.74	5675.58	Surplus
Pudak	708.30	1029.09	Surplus
Pulung	3985.83	12995.65	Surplus
Mlarak	2708.45	7785.25	Surplus
Siman	3713.46	9241.25	Surplus
Jetis	2412.50	9621.88	Surplus
Balong	3635.86	16247.97	Surplus

Kauman	3587.17	16175.83	Surplus
Jambon	3536.96	9475.15	Surplus
Badegan	2550.96	7834.56	Surplus
Sampung	3078.96	16584.96	Surplus
Sukorejo	4449.92	29543.99	Surplus
Ponorogo	5926.63	6213.31	Surplus
Babadan	5441.24	28003.68	Surplus
Jenangan	4689.57	19394.44	Surplus
Ngebel	1628.11	1980.31	Surplus

A sub-district is considered to have a rice surplus when the amount of rice production in the sub-district is greater than its rice consumption. Conversely, a sub-district is considered to have a rice deficit when the amount of rice production in the sub-district is less than its rice consumption [8]. Based on the table above, it can be seen that the amount of rice production in each sub-district has a value that is greater than the amount of rice consumption. It can be concluded that all sub-districts in Ponorogo Regency in 2022 experienced a rice surplus condition, so it is expected to be allocated to other districts/cities that experience a rice deficit. The following figure presents a map of the distribution of rice surplus in each sub-district in Ponorogo Regency.

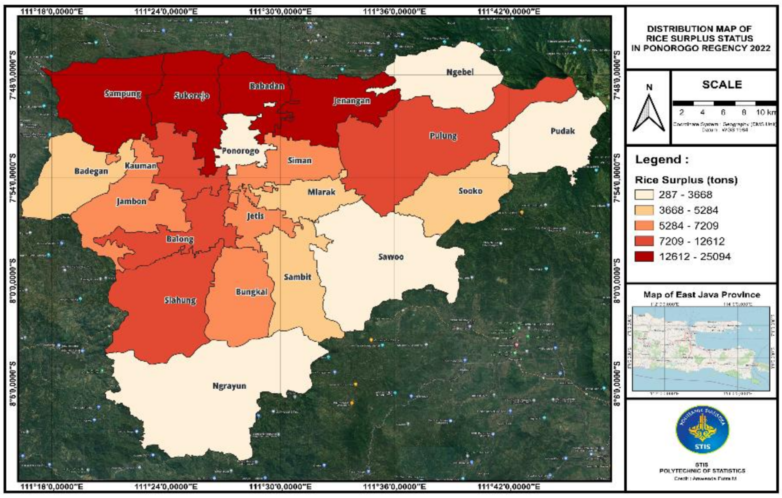

Figure 11. Distribution map of rice surplus

Based on the figure 11 above, it can be seen that the area in the northern part of Ponorogo Regency has a fairly dark color. This indicates that the region has a large amount of rice surplus compared to other regions that are light in color. The lighter the color in the region can be interpreted that the smaller the amount of rice surplus produced in the region. Meanwhile, the areas marked in white and located in the center represent urban regions, indicating that these areas have a slight rice surplus. This is due to several main factors. First, the rice harvest area in this urban region is relatively small because most of the land is used for residential housing and other urban infrastructure. Second, the productivity of rice in this urban area is not very high, which also affects the total amount of rice produced. Overall, the results of the estimation of rice production and consumption show that in 2022 Ponorogo Regency will experience a rice surplus of 169.16 thousand tons. Therefore, it can be said that Ponorogo Regency is experiencing a very good food sufficiency condition.

V. CONCLUSIONS AND SUGGESTIONS

Based on the results of the research, the estimation of paddy productivity using the Geo-SAE method shows that the average paddy productivity in Ponorogo Regency in 2022 was obtained at 5.801 tons/ha. The highest paddy productivity value was produced by Sooko sub-district with a value of 6.447 tons/ha. Meanwhile, Badegan sub-district has the lowest productivity value, which is 5.074 tons/ha. Besides that, the results of the identification of food sufficiency conditions by sub-district in Ponorogo Regency in 2022 showed that overall Ponorogo Regency experienced a rice surplus condition of 169.16 thousand tons and all sub-districts in Ponorogo Regency also experienced a rice surplus condition or in other words, each sub-district was able to meet the food needs of its people in the form of rice. Based on the results of the rice surplus mapping, it shows that the northern sub-district area has a very large rice surplus condition. Based on the evaluation of the goodness of the model, the Relative Standard Error (RSE) value using the Geo-SAE model has a smaller value than using direct estimation, indicating that the Geo-SAE model provides a relatively small error rate. The average RSE value of direct estimation results is 19.802 percent, but there are still several sub-districts that have an RSE value of more than 25 percent, while the average RSE value of indirect estimation results using Geo-SAE is 12.466 percent.

Meanwhile, the RSE value in each sub-district using the Geo-SAE model has a value of less than 25 percent, so it has met the feasibility standards of the estimation used by BPS-Statistics Indonesia.

There are several advantages to estimating paddy productivity using the Geo-SAE model, including the ability to present paddy productivity data down to the sub-district level, while direct estimation can only provide data at the district level. It has the capability to adjust for samples that do not accurately represent the population, and it has been proven to improve the quality of direct estimates of paddy productivity at the sub-district level, resulting in more accurate and usable outcomes. The suggestion of this study is that the Geo-SAE model has proven to be effective and can be considered a reliable method to estimate a variety of indicators affected by spatial variables up to a small area. In this context, BPS-Statistics Indonesia may consider using the geo-SAE method in estimating the productivity of the paddy up to the level of accuracy. In addition, this method also has great potential to be applied in the estimation of other indicators whose data are not available to a smaller area, thereby improving the accurability and depth of the statistical analysis carried out by the BPS or other researchers.

REFERENCES

[1] Dinas Ketahanan Pangan, Tanaman Pangan dan Hortikultura Provinsi Lampung, "Jenis-Jenis Padi," Dinas Ketahanan Pangan, Tanaman Pangan dan Hortikultura Provinsi Lampung, 2024.

[2] USDA, *Rice Outlook 2022*. Washington D.C.: United States Department of Agriculture, 2023.

[3] USDA, *Rice Crop Production 2022*. Washington D.C.: United States Department of Agriculture, 2022.

[4] M. A. Rizaty, "Konsumsi Beras Penduduk Asia Tertinggi di Dunia," Databoks, Jakarta, Feb. 12, 2021.

[5] BPS RI, *Ringkasan Pengeluaran untuk Konsumsi Penduduk Indonesia 2022.* Jakarta: Badan Pusat Statistik, 2023.

[6] R. A. Nugroho, "RI Ternyata Bergantung Beras ke 4 Negara Ini," CNBC Indonesia, Jakarta, Jan. 15, 2024.

[7] E. Yanwardhana, "Kenapa RI Masih Impor Beras?" CNBC Indonesia, Jakarta, Jan. 02, 2024.

[8] Kementan RI, *Kebijakan Penyelamat Swasembada Pangan.* Jakarta: Kementerian Pertanian, 2018.

[9] M. Ardiansyah, A. Djuraidah, and A. Kurnia, "Pendugaan Area Kecil Data Produktivitas Padi dengan Geoadditive Small Area Model," *Jurnal Penelitian Pertanian Tanaman Pangan*, 2(2), 101. https://doi.org/10.21082/jpptp.v2n2.2018.p101-110, 2018.

[10] N. Chand, and C. H. Alexander. *Using Administrative Records for Small Area Estimation in the American Community Survey*. Maryland, United States: U.S. Bureau of the Cencus, 1995.

[11] G. F. Santoso and S. Muchlisoh, "Estimasi Produktivitas Padi Level Kecamatan di Kabupaten Tulungagung Menggunakan Geoadditive SAE," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, no. 3, pp. 23–36, 2022, doi: https://doi.org/10.34123/jurnalasks.v14i1.385.

[12] N. H. Pusponegoro, A. Djuraidah, A. Fitrianto, and I. M. Sumertajaya, "Geo-additive Models in Small Area Estimation of Poverty". *Journal of Data Science and Its Applications*, 2(1), 59–67. https://doi.org/10.21108/jdsa.2019.2.15, 2019.

[13] BPS RI, *Luas Panen dan Produksi Padi di Indonesia 2022*. Jakarta: Badan Pusat Statistik, 2023.

[14] BPS Jawa Timur, *Ringkasan Eksekutif Luas Panen dan Produksi Padi di Jawa Timur 2022*. Surabaya: Badan Pusat Statistik Provinsi Jawa Timur, 2023.

[15] BPS Ponorogo, *Kabupaten Ponorogo Dalam Angka 2022*. Ponorogo: Badan Pusat Statistik Kabupaten Ponorogo, 2023.

[16] BPS RI, *Buku Pedoman Pengumpulan Data Survei Ubinan Tanaman Pangan 2018*. Jakarta: Badan Pusat Statistik, 2018.

[17] Rao, and I. Molina, *Small Area Estimation Second Edition.* New Jersey: John Wiley & Sons, Inc., 2015.

[18] E. E. Kammann, & M. P. Wand. Geoadditive Models. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 52(1), 1–18, 2003, doi.org/10.1111/1467-9876.00385.

[19] Shan Yu, G. Wang, Li Wang, C. Liu & L. Yang, "Estimation and Inference for Generalized Geoadditive Models," *Journal of the American Statistical Association*, 2019, doi: https://doi.org/10.1080//01621459.2019.1574584.

[20] C. Bocci, *Geoadditive Models for Data with Spatial Information*. Florence: University of Florence, 2009.

[21] D. Ruppert, M. P. Wand, and R. J. Carroll, *Semiparametric Regression.* New York: Cambridge University Press, 2003.

[22] F. Apriani, "Pemodelan Pengeluaran per Kapita Menggunakan Small Area Estimation dengan Pendekatan Semiparametrik Penalized Spline," Thesis, Institut Teknologi Sepuluh Nopember, Surabaya, 2017.

[23] A. Kurnia, "Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di dalam Pendugaan Area Kecil dengan Penerapan pada Data Susenas," Thesis, Institut Pertanian Bogor, Bogor, 2009.

[24] A. Petrucci and M. Pratesi, "Spatial Models in Small Area Estimation in the Context of Official Statistics". *Statistica Applicata - Italian Journal of Applied Statistics*, 24(1), 9-27, 2014.

[25] Biemer, P. P., & Lyberg, L. E, *Introduction to Survey Quality*. Hoboken: John Wiley & Sons, Inc., 2003.

[26] H. J. Miller, "Tobler's First Law and Spatial Analysis." *Annals of the Association of American Geographers*, vol. 94, no. 2, June 2004, pp. 284–289, https://doi.org/10.1111/j.1467-8306.2004.09402005.x.