Adaptive Estimation for the Distribution Model of Golden Apple Snail (Pomacea canaliculata (Lamarck)) Pests Using Kernel and Spline Smoothers with Goldenshluger-Lepski Method

Zulfikar[1,*], Mohamad Nasirudin[2], Ambar Susanti[3] Agus Sifaunajah[4]

[1]*Department of Informatics Science, Faculty of Informatics Technology, Universitas KH. A. Wahab Hasbullah, Jombang, Indonesia*

[2]*Department of Agroecotechnology, Agricultural Faculty, Universitas KH. A. Wahab Hasbullah, Jombang, Indonesia*

[3]*Department of Informatics System, Faculty of Informatics Technology, Universitas KH. A. Wahab Hasbullah, Jombang, Indonesia*

(Received: December 3, 2024; Revised: January 4, 2025; Accepted: March 26, 2025; Available online: April 15, 2025)

Abstract

The accuracy of the golden apple snail pest distribution model estimation is very much needed by farmers in dealing with pest attacks, especially in the rainy season. This research aimed to obtain the best distribution model of golden apple snail pests with kernel estimators and spline smoothing through the Goldenshluger-Lepski adaptive bandwidth selection method with an estimation error rate below 10%. The parameters measured were population density 7-42 days after planting, Morisita index, and environmental correlation. The results showed that the population density of golden apple snail pests from four research locations differed significantly in both the juvenile phase (Pr>F = 0.00161), pre-adult (Pr>F = 0.000872), and adult (Pr>F = 0.019122). The highest density was found in Bandar Kedungmulyo District (9.23 individuals.m-2), while the lowest was found in Megaluh District (6.37 individuals.m-2). The population pattern is evenly distributed with a Morisita index of less than one and the highest index (Id = 0.469) was recorded in Megaluh District. The best population distribution model was obtained using the optimum h(7) kernel smoothing estimator, with the lowest Mean Square Error (0.001), and Mean Absolute Square Error (0.032) values in Megaluh District. Furthermore, the best distribution model was obtained using the natural cubic spline smoother with the lowest Mean Square Error (0.055), and Mean Absolute Square Error (0.020) values in Tembeleng District. In conclusion, the best golden apple snail pest distribution model was obtained using the adaptive kernel smoothing estimator of the Goldenshluger-Lepsky model approach, which produced the lowest estimation error rate compared to the spline smoother. This research contributes to developing the best distribution model for golden snail pests, which can strengthen the information technology database for monitoring, controlling, and utilizing the potential of golden snail pests.

Keywords: Adaptive Estimation, Distribution Model, Golden Apple Snail, Kernel and Spline Smoothers, Goldenshluger-Lepski

1. Introduction

Golden apple snail (Pomacea canaliculata (Lamarck)) is a large freshwater snail native to tropical and subtropical South America. It was originally introduced to Taiwan from South America in 1980 for local food consumption and export[1]. The snail can be processed into a biostimulant to stimulate plant growth because the extract contains phytohormone IAA (Indole Acetic Acid) [2]. Additionally, golden apple snail protein contains flavonoids that function as antioxidants and immunomodulators [3].

According to previous research, golden apple snail is naturally distributed in tropical and subtropical freshwater ecosystems in the world, can reproduce continuously, reach maturity faster, complete more generations per year, as well as maintain high young and adult populations in the year in tropical areas [4]. Several models have been developed to describe the potential range and performance in different geographic locations, using various methods and parameter combinations to predict distribution, density, and population dynamics[5], [6]. Distribution models have also been developed for other invasive golden apple snails including the congeneric species P. maculata [7], [8]. Using the 'climate matching' method through MaxEnt to evaluate the risk of invasion is very suitable to prioritize areas for future

***Corresponding author: author (zulfikardia@gmail.com)**
DOI: https://doi.org/10.47738/jads.v6i2.578

surveys by estimating potential distribution [6]. However, this method still has some limitations where the variables used for modeling only represent a small subset of the possible environmental variables that can affect the overall distribution of golden apple snail. MaxEnt modeling cannot consider other biotic and abiotic interactions such as competition and predator-prey interactions that are not included in the environmental variables used to predict the species distribution [9].

SDMs models can be problematic for species with few occurrence records, such as P. canaliculata and P. diffusa. Ironically, species rarely documented are often the ones that need predictive models the most [10]. Elith et al. [11] stated that the estimation of golden apple snail distribution model with extended SDMs obtained 15% as the suitability of the habitat. The value remains relatively high because it exceeds 10% to fulfill the requirements for a safe estimate of risk and has quite high speculation [12]. Another fundamental problem is that the estimation of golden apple snail distribution model over a long period becomes very weak. GARCH (1,1) method for volatility measurements cannot show the up and down movements clearly [13]. The weakness of this method is requiring a long period to form the noise effect as a criterion for time series data analysis which tends to increase the risk.

Adaptive estimation of kernel and spline smoothers can reduce the difference in estimated values to below 10% as a safe requirement. Therefore, when the analysis is carried out, the movement of long-term estimated values can be overcome and the long-term memory effect is quickly detected [11]. The advantage of Goldenshluger-Lepski model is the ability to obtain the best estimator in nonparametric regression through the adaptive bandwidth method with increased estimation efficiency [14]. In the case of density estimation with kernel estimator, Goldenshluger-Lepski model can overcome failure in estimation when the variance term is too small [15]. This model extends Lepski method to adapt to several parameters, specifically in nonparametric regression [16].

Based on these problems, research needs to be conducted to obtain the best estimator of golden apple snail pest distribution model using the adaptive estimation of kernel and spline smoothers in Goldenshluger-Lepski model. Nonparametric estimators can build estimates entirely based on data without making strong assumptions. This research is expected to obtain a valid distribution model to support the information system database for monitoring, controlling, and using the potential of golden apple snail as amino acid and flavonoid materials.

2. Literature Review

2.1. Kernel Smoother

Kernel is the best smoothing method for handling static and random models. It is sometimes called weight or window function, continuous or symmetric function, and has an integral that is the same as an integer one when (bandwidth) is very small. The smoother regression formula is as follows [17].

$$\widehat{m_h}(x) = \frac{\sum_{i=1}^{n} y_i k(x-X_i)/h}{\sum_{i=1}^{n} k(x-X_i)/h}, \tag{5}$$

$$w_i(x) = \frac{\frac{k(x-X_i)}{h}}{\sum_{i=1}^{n} k(x-X_i)/h} \tag{6}$$

Where $\sum_{i=1}^{n} k(x-X_i)/h$ indicates the endodermic function, $w_i(x)$ represents the weight function and one of the conditions is positive. Meanwhile, h represents the smoothing parameter (bandwidth) in the bbb estimator. When the value is large, then the function is smooth, but when the value is small, the function is not smooth. The general form of the estimation function $m(x_i)$ using the Gaussian kernel smoother is as follows:

$$K(x) = \frac{1}{h\sqrt{2\pi}} e^{-0,5\left(\frac{x-x_i}{h}\right)^2} \tag{7}$$

2.2. Spline Smoother

Nonparametric spline regression smoother relies on the sum of squared errors and is used when the regression line is divided into some parts. Considering the explanatory variable x with the period (a, b) is divided and the cut line is called a sliding knot, sliding smoothing overcomes the problem of knot selection. The identification of new knots or modification of existing ones is divided into cubic spline (SPC) and natural cubic spline (NSPC) [18], [19]:

$$S(m) = \sum_{i=1}^{n}(y_i - \widehat{m}(x_i))^2 + \lambda \int_a^b [\widehat{m}(x)]^2\, dx \quad \lambda > 0 \tag{8}$$

$\sum_{i=1}^{n}(y_i - \widehat{m}(x_i))^2$ is the sum of squared errors, $\widehat{m}(x_i)$ denotes the second derivative of the bootstrap function, λ is the penalty factor indicating the width of the fit quality package by $\sum_{i=1}^{n}(y_i - \widehat{m}(x_i))^2$ and the smoothing value is indicated by $\int_a^b [\widehat{m}(x)]^2\, dx$.

2.3. Goldenshluger-Lepsky Method

Goldenshluger-Lepski adaptive bandwidth extends Lepski method to perform adaptation across multiple parameters. This method has been used in different contexts after being first applied in multidimensional white noise models. Due to the wide applications in recent nonparametric estimation research, the idea of this method for adaptive nonparametric estimation is to select an estimator that reduces the number of unknown variance bias factors [20], [21]. The Goldenshluger-Lepsky method is used because it can provide a decrease in the estimation error value for the bandwidth selection of the two nonparametric estimators used. Goldenshluger-Lepski formula is as follows [22]:

$$\hat{h}(x_i) = argmin_{h\epsilon H_n}\{\hat{A}(h, x_i) + \hat{V}(h, x_i)\} \tag{9}$$

$$\hat{A}(h, x_i) = max_{h'\epsilon H_n}([\widehat{m}_{h'}(x_i) - \widehat{m}_{hvh'}(x_i)]^2 - V(h', x_i)) \tag{10}$$

$$\hat{V}(h, x_i) = k\sigma^2 \frac{\ln n}{n\hat{\varphi}(h)}, h \neq 0 \tag{11}$$

Where K is a constant that does not depend on h, $\widehat{m}_{h'}(x_i)$ is the function estimator, Hn represents a set of smoothing parameters (bandwidth), $\hat{V}(h, x_i)$ represents the empirical variance analog, and $\hat{A}(h, x_i)$ represents the estimated bias term. To estimate the regression curve, the criteria used are Mean Absolute Error Squares (MAE), Roots Mean Squares Error (RMSE), and Mean Squared Error (MSE) using the adaptive bandwidth of Goldenshluger-Lepski model on the experimental side [19], [23].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{m}(x)| \tag{12}$$

$$RMSE = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{m}(x))^2} \tag{13}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{m}(x))^2. \tag{14}$$

The selection of the smoothing parameter λ can be based on unexceptional cross-validation, namely minimizing the cross-validation score. General Cross Validation (GCV) criterion uses the average leverage value:

$$GCV(\lambda) = \frac{1}{n}\sum\left(\frac{y_i - \widehat{m}_n(x_i, \lambda)}{1 - n^{-1}trace(A(\lambda))}\right)^2 \tag{15}$$

$\hat{m}(x, p = 0, h) = A(h)y$ [24]. It is also carried out to minimize the penalized residual sum of squares:

$$RSS(f, \lambda) = \sum_{i=1}^{N}\{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt \tag{16}$$

λ is a smoothing parameter [25].

3. Analysis Method

3.1. Data Collection

This research was conducted using a survey method based on differences in location and cultivation system in rice plants from 7 to 42 days after planting (DAP) to determine the population density, distribution, and structure of golden apple snail. The determination of the research location used the purposive random sampling method by taking 4 predetermined sub-districts, where each sub-district was assigned 3 sample villages.The location determination was based on the results of a preliminary survey, namely rice fields in Bandar Kedungmulyo Sub-district at an altitude of ± 35 meters above sea level (masl), Megaluh Sub-district at ± 38 masl, Kesamben Sub-district at ±34 masl, and Tembelang Sub-district at ±41 masl.

The sampling stage was conducted using a quadrant of PVC pipe measuring 50 cm x 50 cm. The sampling method was carried out by dividing the land area into seven plots with each plot measuring 2 m apart. Each plot was sampled using a quadrant three times, and golden apple snail taken as samples were not based on a certain size. Therefore, all small to large snails were taken as samples. The number of golden apple snail was obtained from the sampling results and calculated in each plot, including those on the water surface, the water, and attached to the rice plants, taken using a fishing net (0.5 mesh). In the next stage, the collected snails were cleaned from mud or rice field soil with running water, placed into a plastic bag (ziplock), stored in a cool box, and taken to the laboratory for identification.

3.2. Observation

In the laboratory, the samples were sorted and stored in small collection bottles containing 70% alcohol. Furthermore, individual golden apple snails were counted using a hand counter and the diameter was measured with a caliper. Snails measuring less than 5 mm were observed using a microscope and those measuring more than 5 mm were observed directly. The identification stage used the Field Guide of Freshwater Invertebrates of North America [26].

Snails in each rice field were calculated and converted into units of individuals/m² using the Brower and Zar formula [27].

$$D = {}^{N}\!/_{A} \tag{17}$$

where: D (Number of individuals per square meter (individuals/m²)), N (Total number of individuals), and A (Area of the square plot (m²)). The distribution pattern is to take and count all individuals in each plot, then analyze using the Morisita Index (Id) formula [28], [29]:

$$\mathrm{Id} = n\frac{\left(\sum x_i^2 - \sum x_i\right)}{(\sum x_i)^2 - \sum x_i} \tag{18}$$

Id = 1 (The distribution pattern is random), Id> 1 (The distribution pattern is clumped), and Id <1 (The distribution pattern is uniform). The technique for collecting data on golden snail pests in paddy fields based on their growth phase is shown in figure 1.

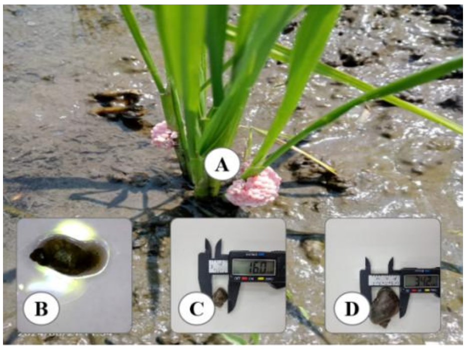


Figure 1. Population structure of golden apple snail observed in the egg (A), juvenile (5-10 mm) (B), pre-adult (10-25 mm) (C), and adult phases (25-40 mm) (D)

Observations were made based on the shell diameter of golden apple snail divided into three, namely juvenile (5-10 mm), pre-adult (10-25 mm), and adult (25-40 mm). This was performed directly by measuring the shell diameter using a caliper [30].

Data on population density, distribution, and structure of golden apple snails were acquired by calculating the total number obtained at each sampling. Adaptive estimation was processed in a nonparametric regression method through the use of kernel and spline smoothers using the R Program application version 4.4.1. The nonparametric regression formula is as follows:

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1,2,3,\dots,n, \quad \varepsilon \sim N(0,\sigma^2) \tag{19}$$

ISSN 2723-6471

Where $y_i$is the response variable, $m(x_i)$ is the unknown function to be estimated, xi is the explanatory variable: ε_i is the value of the random variable, referring to a normally distributed white noise. The adaptive estimator for the parameter vector is as follows [31]:

$$\hat{\theta}(\mathrm{x}) = \hat{\theta}_{\mathrm{k}}(\mathrm{x}) = \left(\theta_{\mathrm{k}}^{1}(\mathrm{x}), \dots \dots, \theta_{\mathrm{k}}^{\mathrm{p}}(\mathrm{x})\right)^{\mathrm{T}} \tag{20}$$

$k = 1, \dots, p \quad \theta_k^1, \dots \dots, \theta_k^p$: unknown parameters, θ is estimated based on sample observations (x_i, y_i).

4. Results and Discussion

Based on the observations, collection, and analysis of data on the distribution of golden apple snail pest population from four research locations, the presentation of results and discussion includes a description of the population density, the relationship with the environment, as well as distribution models with kernel and spline smoothers. The population density of golden apple snail produced a significant difference in all observed growth phases, both juvenile, pre-adult, and adult. The highest growth was obtained in the juvenile phase while the distribution model was even and not clustered. The even pest distribution model is difficult to assess, hence, it is necessary to estimate the model using a nonparametric regression method with kernel and spline smoothers through the selection of smoothing parameters in Goldenshluger-Lepski model. This is based on the criteria for the best model by selecting the smallest value of GCV, MSE, RMSE, and MAE. A detailed description of the results is as follows:

4.1. Population Density of Golden Apple Snail

The observation results in three growth phases observed from 7 to 42 DAP show that Bandar Kedungmulyo Sub-district has the highest population density of golden snail in all phases with a juvenile of 6.25 individuals, pre-adult of 2.09 individuals and adult of 0.98 individuals per m2, as shown in figure 2. Meanwhile, the other three sub-districts, namely Megaluh, Tembelang, and Kesamben, have almost the same number in all growth phases. Megaluh has a juvenile phase of 4.74, pre-adult of 1.42, and adult of 0.48 per m2, Tembelang has a juvenile phase of 4.67, pre-adult of 1.41, and adult of 0.48 per m2, while Kesamben has a juvenile phase of 4.80, pre-adult of 1.44, and adult of 0.47 per m2. The total population density is 9.23 for Bandar Kedungmulyo, 6.37 for Megaluh, 6.56 for Tembelang, and 6.71 individuals per m2 for Kesamben.

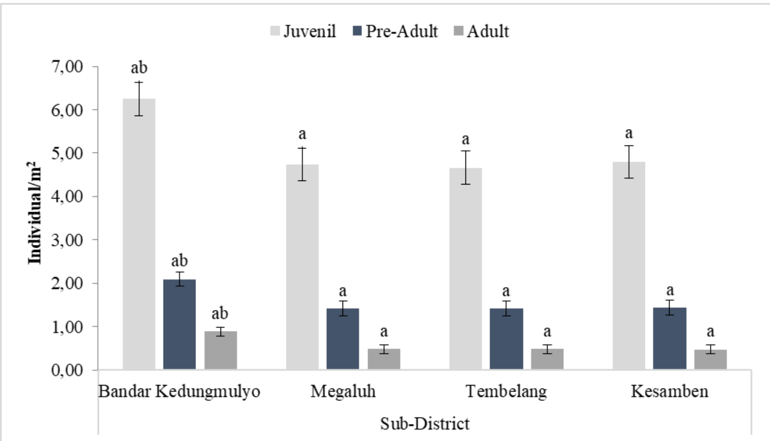


Figure 2. Average population density of golden apple snail based on their growth phases per square meter from 4 research locations during observations between 7 and 42 DAP on rice plants

The high growth of juvenile phase at the research locations is due to the ability to reproduce continuously, reach maturity faster, complete more generations per year, as well as maintain a high population of young and adult snails in the year in tropical areas [4]. The growth of golden apple snail population damages various plants, specifically rice seedlings, where one adult snail can eat 5–24 rice seedlings per day [32]. Therefore, the high growth rate leads to a significant decrease in rice yields (~14%) and net income from rice (~60%) [33]. Three-way ANOVA test results in Table 1 below show that Bandar Kedungmulyo has a significant difference in average population density both in juvenile (0.01) (Pr>F = 0.00161), pre-adult (0.001) (Pr>F = 0.000872), and adult phases (0.1) (Pr>F = 0.019122) compared to the other three sub-districts.

Table 1. Results of Three-Way ANOVA Test of Population Density in 3 growth phases of the Golden Snail Pest in 4 research locations.

	Df	Sum Sq.	Mean Sq.	F Value	Pr (>F)
Juvenile	1	2.682	2.682	21.75	0.001617**
Pre. -adult	1	3.275	3.275	26.55	0.000872***
Adult	1	1.056	1.056	8.56	0.019122*
Residual	8	0.987	0.123		
Signif. Codes:		*0 '***'*	*0.001 '**'* *0.01 '*'*	*0.05 '.'*	*0.1 ' ' 1*

Although the growth of juvenile phase was relatively high, the distribution tends to be uniform as indicated by the Morisita index with a value of less than one. The distribution model based on index measurement is indicated by a value of less than one (Id <1) (figure 3) forming an even/uniform distribution [21].

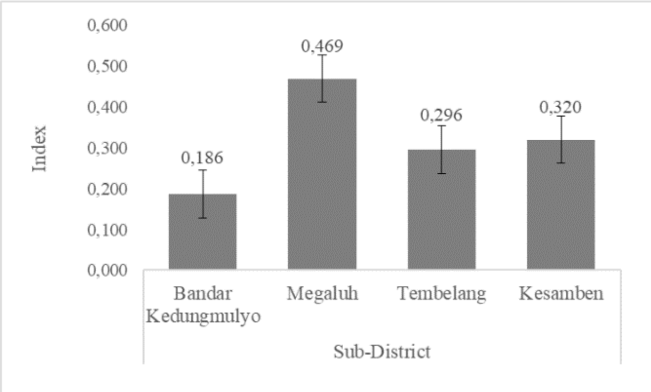


Figure 3. Morisita index of golden apple snail pest growth with a value of less than one (Id < 1) indicates that the pest distribution is uniform/even.

Morisita index from each research location shows Bandar Kedungmulyo at 0.186, Mageluh at 0.470, Tembelang at 0.296, and Kesamben at 0.365 with the largest index occurring in Megaluh, and the smallest in Bandar Kedungmulyo. Although Bandar Kedungmulyo was found to have the highest population density in all growth phases, it also had the lowest Morisita index indicating that the four research locations produced a uniform distribution model. This is presumably because the first river irrigation flow enters Bandar Kedungmulyo area, then goes to the three other sub-districts as the first entry points for golden apple snail from Brantas River. Additionally, the invasive characteristics and nature of the growth are relatively rapid.

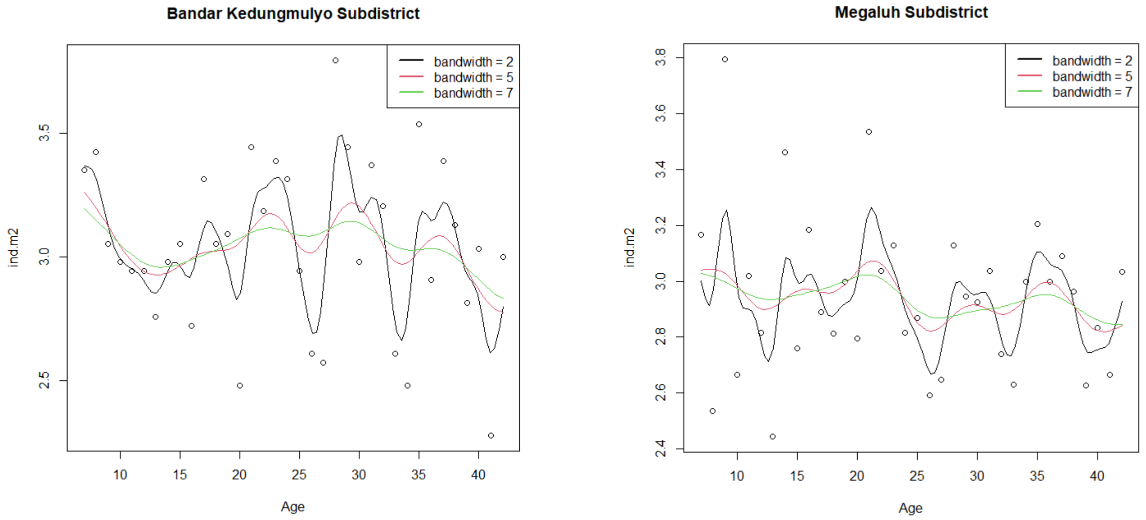
The distribution over time shows that golden apple snail spreads gradually by human activities, water flow, global warming, and other environmental factors [34]. Golden apple snail moves from one place to another by attaching to an object following the water flow and sometimes are deliberately carried by humans who admire the beautiful shape with a golden yellow color. Considering the even distribution across all research locations, estimating the distribution model will be difficult where the results of the skater plot between population density and observation time form a random pattern. The nonparametric regression estimation model method is very appropriate to use when the initial model conditions are difficult to determine the form [19].

4.2. Distribution Model with Kernel Smoother

The application of golden apple snail distribution model estimation used a nonparametric regression equation: $y_i = m(x_i) + \varepsilon_i, \quad i = 1,2,3,\dots,n, \quad \varepsilon \sim N(0,\sigma^2)$, Where y_i: population density of golden apple snail per m2, $m(x_i)$: unknown function to be estimated, xi: growth age of rice plants (days), and ε_i : normally distributed random variable value. Furthermore, the estimation function $m(x_i)$ used Gaussian kernel smoother. The analysis of research data from

 ISSN 2723-6471

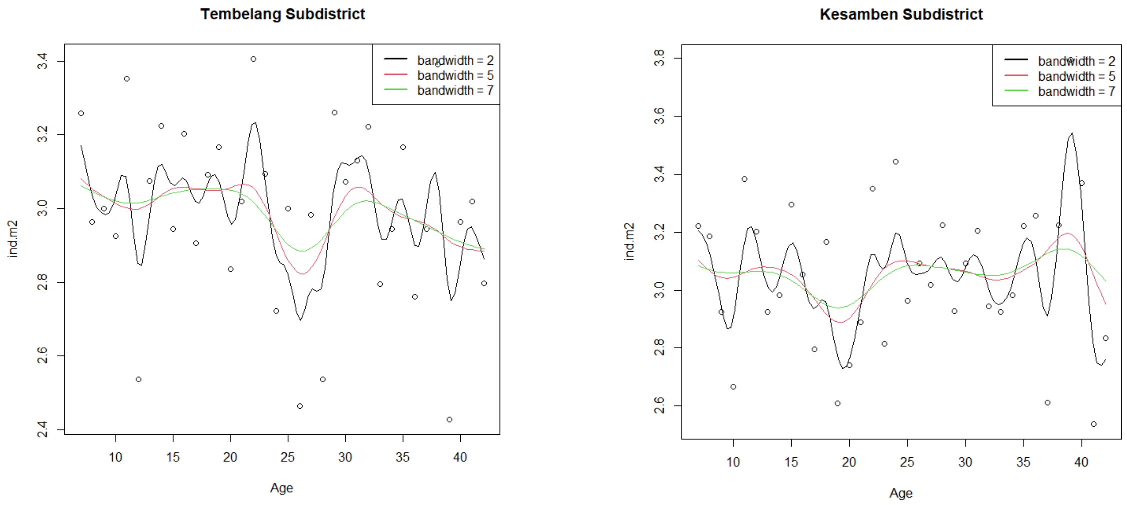
four sub-districts with kernel bandwidth smoother (h) of 2, 5, and 7 as well as observation time between 7 to 42 DAP obtains golden apple snail distribution model shown in figure 4 and figure 5.



(a) (b)

Figure 4. Golden Apple Snail Distribution Model using Adaptive Kernel Smoother with Goldenshluger-Lepski Method in Bandar Kedungmulyo (a) and Megaluh (b) Sub-districts with bandwidth selection (h) of 2, 5, and 7

Based on figure 4 in Bandar Kedungmulya and Megaluh, kernel smoother regression curve is smoother when selecting bandwidth (h=7) (green curve), compared to h(5) (red curve) and h(2) (black curve) which appears rough and rougher, respectively. In figure 6(a), kernel smoother curve model h(2) shows a sharp up and down movement during observations on 20 to 40 DAP. However, in general, the movement of golden apple snail pest distribution model decreases up to the last observation (42 DAP) as shown in h(5) and h(7). In figure 6(b), the movement of h(2) kernel smoother model curve tends to be stable, although there is a rather extreme movement in observations from 10 to 25 DAP. Generally, there is a downward movement of the curve as shown in the h(5) and h(7) kernel smoothers.



(a) (b)

Figure 5. Golden Apple Snail Distribution Model using Adaptive Kernel Smoother with Goldenshluger-Lepski Model in Tembelang (a) and Kesamben (b) Sub-districts with bandwidth selection (h) of 2, 5, and 7

As shown in figure 5, h (7) tends to provide a smoother graph compared to h (5) and h (2) which appear rough and very rough, respectively. Figure 7(a) shows that the distribution model with the h (2) kernel smoother has a sharp up-and-down movement of the curve in observations from 20 to 30 DAP. In general, there is a decline in kernel smoother models h(5) and h(7). Different conditions are shown in figure 7(b), where there is a sharp decline in the observation of 20 DAP and a sharp increase in the observation of 40 DAP. When viewed in kernel smoother curve models h(5) and h(7), there is a downward movement even though the decline appears to be small. Three main factors influencing the fluctuation of golden apple snail pests include natality, and mortality [35]. The natality factor is relatively high, hence, the population increases rapidly. The mortality factor is weak because the presence of natural enemies is not yet balanced, hence, the development of golden apple snail is fast[36].

The best model criteria from various bandwidth values (h) of Goldenshluger-Lepski adaptive model are based on the selection of the lowest GCV, MSE, MAE, and RMSE values as shown in table 2. The results showed that the best model for estimating the distribution of golden apple snail population with kernel smoother is shown in h(7) at Bandar Kedungmulyo, Megaluh, Tembelang, and Kesamben Sub-districts. The lowest MSE (0.001), MAE (0.032), and RMSE (0.246) values were found in Megaluh, hence, this sub-district was considered to have the best distribution model. The adaptive estimator method of kernel smoother with bandwidth selection using Goldenshluger-Lepski is the best solution for building a valid distribution model [21].

Table 2. Criteria for the Best Model Based on GCV, MSE, MAE, and RMSE Values in Adaptive Kernel Smoother of Goldenshluger-Lepski Model

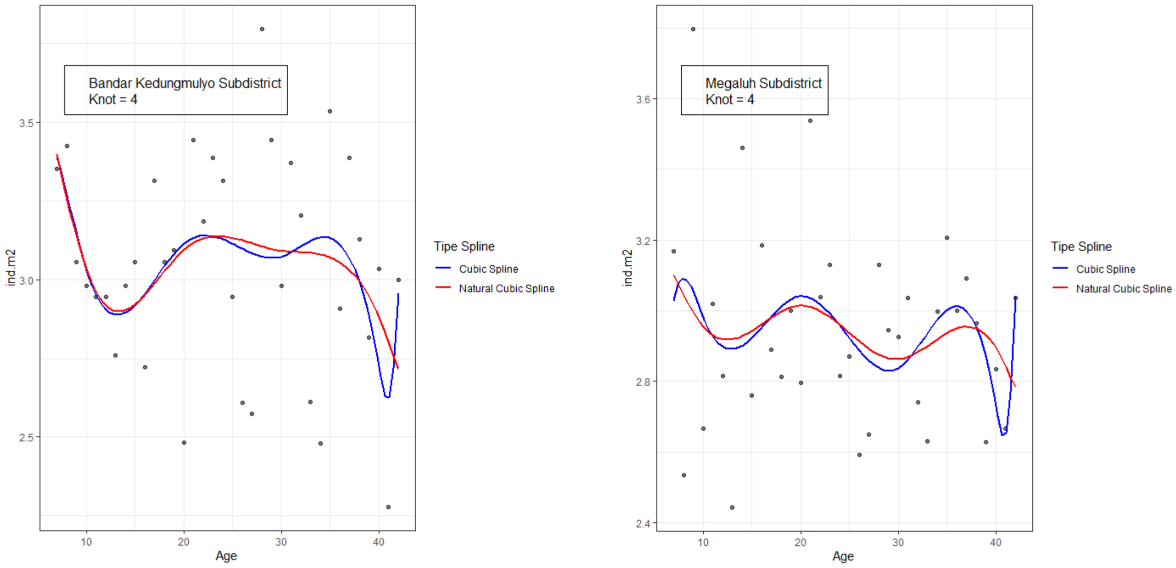
Sub-District	Bandwidth (h)	GCV	MSE	MAE	RMSE
	2	0.1542	0.042	0.178	0.398
Bandar Kedungmulyo	5	0.1375	0.010	0.084	0.349
	7	0.1358*	0.005	0.061	0.340
	2	0.1699	0.018	0.110	0.395
Megaluh	5	0.1101	0.004	0.048	0.289
	7	0.0734*	0.001	0.032	0.246
	2	0.1174	0.018	0.091	0.263
Tembelang	5	0.0788	0.004	0.047	0.248
	7	0.0736*	0.002	0.033	0.247
	2	0.0951	0.023	0.112	0.289
Kesamben	5	0.0945	0.004	0.049	0.261
	7	0.0905*	0.002	0.035	0.257

*Description: * Minimum GCV as a determinant of the optimum bandwidth value (h)*

Estimation using kernel smoother provides the best model for selecting the optimum bandwidth (h=7). The best distribution model was obtained with kernel smoother based on the selection of the lowest MSE, MAE, and RMSE values in selecting the optimal bandwidth [18]. According to Maharani and Saputro[24], GCV method obtains a minimum value that determines how well the smoothing parameters indicated by the estimator do not change significantly even though the amount and position of the bandwidth vary.

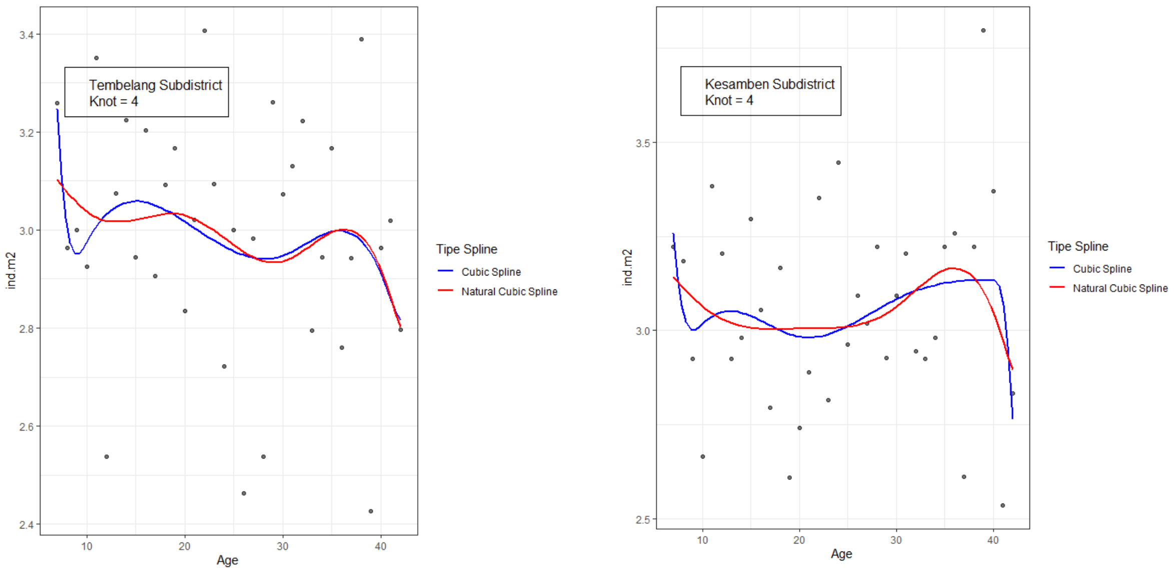
4.3. Distribution Model with Spline Smoother

Nonparametric spline regression smoother is applied with variable x as the growth age of rice plants and y representing the population density of golden apple snail during the identification of 4 knots and modeling using SPC and NSPC [37]. The distribution model with linear and SPC smoothers is shown in figure 6 and figure 7. Specifically, figure 6(a) shows that the spline curve experiences a sharp downward movement at knots 10 and 40, both SPC (red lines) and NSPC (blue lines). In figure 6(b), the sharp downward movement of the curve is observed at knot 40. The movement of the curve in figure 6 shows a decrease in population density up to the 42nd DAP observation.



(a) (b)

Figure 6. Golden Apple Snail Distribution Model with Adaptive Spline Smoother of Goldenshluger-Lepski Model in Bandar Kedungmulyo (a) and Megaluh (b) Sub-districts with the selection of 4 knots (10, 20, 30 and 40)



(a) (b)

Figure 7. Golden Apple Snail Distribution Model with Adaptive Spline Smoother of Goldenshluger-Lepski Model in Tembelang (a) and Kesamben (b) Sub-districts with the selection of 4 knots (10, 20, 30 and 40).

Figure 7(a) shows a sharp downward movement in SPC knot 10, while figure 7(b) shows a sharp downward movement after knot 40 until the end of observation (42 DAP). A similar result is shown in Figure 7, indicating an NSPC compared to the SPC spline smoother. However, there is still a decrease in golden apple snail population density to the end of observation. The best model criteria were obtained by calculating the lowest GCV, MSE, and MAE values, as shown in table 3. Distribution modeling with linear and SPC smoothers at the same knots shows that Tembelang Sub-district had the lowest RSS, GCV, MSE, and MAE values, suggesting the spline curve has good criteria. The best model with a spline smoother was shown by the lowest RSS (2.060), GCV (0.054), MSE (0.057), and MAE (0.020).

The best distribution model of golden apple snail pests used SPC smoother as indicated by lower RSS, GCV, MSE, and MAE values than linear spline as shown in table 3. Tembelang produced the best distribution model with the lowest

RSS (2.0311), GCV ((0.0498), MSE * 0.0554), and MAE (0.0200) values compared to other sub-districts using SPC smoother. The adaptive estimator model method of spline smoother using Goldenshluger-Lepski is the best solution for building a valid distribution model [21].

Table 3. Criteria for the Best Model based on MSE, MAE, and RMSE values in Adaptive Spline Smoother of Goldenshluger-Lepski model

Spline	Sub-District	Penalized Criterion (RSS)	Generalized Cros-Validation (GCV)	MSE	MAE
cubic spline (SPC)	Bandar Kedungmulyo	3.9702	0.1236	0.1103	0.2671
	Megaluh	2.7501	0.0856	0.0764	0.0216
	Tembelang	2.0600	0.0542	0.0572	0.0201
	Kesamben	2.4949	0.0777	0.0693	0.0214
natural cubic spline (NSPC).	Bandar Kedungmulyo	3.7789	0.1157	0.1100	0.2598
	Megaluh	2.5615	0.0776	0.0687	0.0215
	Tembelang	2.0311	0.0498	0.0554	0.0200
	Kesamben	2.4560	0.0664	0.0634	0.0209

Smoothing parameter spar = 4.0, λ = 4095.998, Equivalent Degrees of Freedom (Df) = 2.000001

The best distribution model with a spline smoother was based on the selection of 4 knots while the lowest GCV, MSE, and MAE values were obtained on SPC. GCV method obtains the minimum value determining how well the smoothing parameters indicated by the estimator do not change significantly even though the number and position of knots are different [24]. This research shows that using spline smoother estimator produced the best model for NSPC compared to SPC because it provides lower GCV, MSE, and MAE values. According to Krivobokova and Kauermann [38], using MSE and MAE to estimate smoothing parameters outperforms other methods, such as GCV or the Akaike criterion, specifically when the error correlation structure is determined less appropriately. Between the two nonparametric estimators that show the best golden apple snail pest distribution model, kernel smoother estimator can provide lower GCV, MSE, and MAE values than spline smoother. Another advantage of this kernel smoother estimator is the ability to estimate data models in detail and be more valid on data distributions that tend to be uniform [39]. This estimator is a solution to overcome the difficulty of building a data distribution model in a field closer to the real data.

5. Conclusion

In conclusion, golden apple snail pest in Jombang Regency was evenly distributed, with the highest population density being in the juvenile phase. The best distribution model was obtained through optimal h(7) kernel smoother estimation and could overcome the difficulty in estimating a uniform/even pest distribution model. Although the population decreased toward the last observation, the distribution estimation model with kernel smoother method could still show detailed fluctuations. This distribution model is very helpful in providing additional information for monitoring and control based on data recorded in real time.

6. Declarations

6.1. Author Contributions

Conceptualization: Z., M.N., A.S., and A.S.; Methodology: A.S.; Software: Z.; Validation: Z., A.S., and A.S.; Formal Analysis: Z., A.S., and A.S.; Investigation: Z.; Resources: A.S.; Data Curation: A.S.; Writing Original Draft Preparation: Z., A.S., and A.S.; Writing Review and Editing: A.S., Z., and A.S.; Visualization: Z. All authors have read and agreed to the published version of the manuscript.

ISSN 2723-6471

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This article was financially supported by a regular fundamental research grant in 2024 from the Directorate of Research, Technology and Community Service, Ministry of Education, Culture, Research and Technology, Indonesia.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Rusli, Hidrayani and R. E. Putra, "Golden apple snail management (Pomacea canaliculata Lamarck) using integrated pest management (IPM) and nonintegrated pest management (non-IPM) on paddy cultivation". *IOP Conf. Ser.: Earth Environ. Sci.* vol. 1160, no. 1 pp. 1-9, Nov. 2023. doi: 10.1088/1755-1315/1160/1/012048

[2] S. L. Asmono, Rahmawati, N.B. Arifiana, and Suharjono, "Phytohormone and amino acid content in golden apple snail (Pomaceae canaliculata L.) extract which fermented at several effective microorganism 4 (EM4) concentrations". *IOP Conf. Ser.: Earth Environ. Sci.* vol. 1338. no. 11 pp. 1-6. Nov. 2024. doi:10.1088/1755-1315/1338/1/012008

[3] W.C. Ferdyana, Y. Widiyastuti, B. Pujiasmanto, A. T. Sakya, and A. Yunus, Short Communication: "Morphological diversity and the addition of golden snail protein: Its effect on flavonoid content on Echinacea purpurea". *Biodiversitas* vol. 23 no.1, pp. 62-66, Jan. 2022. doi: org/10.13057/biodiv/d230108

[4] D. Djeddour, C. Pratt, F. Makale, I. Rwomushana and R. Day, "The apple snail, Pomacea canaliculata: an evidence note on invasiveness and potential economic impacts for East Africa". *CABI Working Paper*, vol. 21, no. 5, pp.1-77, May. 2021. doi: org/10.1079/CABICOMM-62-8149.

[5] J. Lei, L. Chen and H. Li, "Using ensemble forecasting to examine how climate change promotes worldwide invasion of the golden apple snail (*Pomacea canaliculata*)". *Environ Monit Assess* vol. 189. no. 4040. pp. 130-141, Jul. 2017. doi: org/10.1007/s10661-017-6124-y.

[6] F. Makale, A. M. Muvea, I. Mugambi, D. Chacha, E. A. Finch and I. Rwomushana, "Current and potential distribution of the invasive apple snail, Pomacea canaliculata in Eastern Africa: evidence from delimiting surveys and modeling studies", *CABI Agr and Biosci*, vol. 5, No. 92, pp. 1-13, Oct. 2024. doi:10.1186/s43170-024-00301-7.

[7] C. P. Cretini, and K. A. Galloway, "Acidic Apple Snails: Effects of Climate Change on the Mechanical Properties of an Invasive Gastropod", *Integr and Comp Bio*, vol. 64, no. 2, pp. 270-278. August 2024, doi:org/10.1093/icb/icae013

[8] D. Barbitta, C. Clavijo and A. Carranza, "Ecoregional-level assessment of the potential distribution of the invasive apple snail Pomacea maculata Perry, 1810 (Gastropoda: Ampullariidae): setting geographically explicit priorities for the management of the invasion". *Russ J of Bio Inv*, vol. 11, no. 4, pp.172–181, April. 2020. doi: org/10.1134/S2075111720020022

[9] K. López-Reyes, L. Osorio-Olvera, O. Rojas-Soto, X. Chiappa-Carrara, C. Patrón-Rivero and C. Yáñez-Arenas, "*An* exhaustive evaluation of modeling ecological niches above species level to predict marine biological invasions", *Mar Envir Res*, vol.186, no. 4, pp.1-21. April. 2023, doi.org/10.1016/j.marenvres.2023.105926.

[10] F. Goudarzi, M. R. Hemami, M. Malekian, S. Fakheran and F. M. Freiría, "Species versus within-species niches: a multi-modeling approach to assess range size of a spring-dwelling amphibian". *Sci Rep* vol.11, no. 597, pp. 1-11. Jan. 2021. doi.org/10.1038/s41598-020-79783-0

[11] S. Jeremy, L. Dieter, L. J. Corina, and C. Nancy, "The role of climate change and niche shifts in divergent range dynamics of a sister-species pair". *Peer Comm J*, vol. 3, no. e23. pp.1-28, Sep. 2023. doi : 10.24072/pcjournal.248.

[12] A. L. Singleton, C. K. Glidden, A. J. Chamberlin, R. T. Raquel, G. S. Palasio, A. Pinter, R. L. Caldeira, C. L. F. Mendonça, O. S. Carvalho, M. V. Monteiro, T. S. Athni, S. H. Sokolow, E. A. Mordecai, and G. A. De Leo, "Species distribution modeling for disease ecology: a multi-scale case study for schistosomiasis host snails in Brazil", *medRxiv* vol..07. no.10.23292488, pp. 1-51, Mar. 2023

[13] E. Kim, J. Moon, J. Shim and E. Hwang, "Predicting invasive species distributions using incremental ensemble-based pseudo-labeling", *Ecol Inf*, vol. 79, no. 102407, pp. 1-11, Mar. 2024

[14] N. S. Hassan and S. H. Z. Al-Thalabi, “Using simulations to estimate adaptive parameters in a nonparametric regression”. AIP Publishing. *AIP Conf. Proc*. vol. 2977 no. 1. pp. 22 Dec. 2023. doi: org/10.1063/5.0182267

[15] G. Chagny, T. Laloë, and R. Servien, "Multivariate adaptive warped kernel estimation". *Electron J Statist*, vol.13, no.1, pp.1759-1789. Jun. 2019. doi: 10.1214/19-EJS1565.

[16] S. Page, and S. Gr¨unew¨alder, “Ivanov-regularised least-squares estimators over large RKHSs and their interpolation spaces”. *J of Mach Learn Res*, vol 20, no.120, pp.1–49. Jun. 2019.

[17] Z. Zulfikar, M. Munawarah, A. Susanti, “Combination of high order kernel estimation for estimation of fruit trees biomass on critical land”. *J. Phys. Conf. Ser*. Vol. 2157 no. 8. pp. 1-12, August 2022. doi:10.1088/1742-6596/2157/1/012035

[18] B. Lestari, Fatmawati, I N Budiantara and N. Chamidah, "Smoothing parameter selection method for multiresponse nonparametric regression model using smoothing spline and Kernel estimators approaches", *J. Phys. Conf. Ser*. vol. 1397 no. pp. 1-8. 2019. 012064.

[19] N. S. Hassan and S. H. Z. Al-Thalabi, “Nonparametric adaptive smoothing with practical application”. *Euro J. of Res Dev and Sus* (EJRDS), vol. 3 No. 9 pp.69-76. Nov. 2022.

[20] J. Lederer, “Statistical guarantees for sparse deep learning”. *AStA Adv Stat Anal*, vol. 108, no. January, pp. 231–258. 2024. doi.org/10.1007/s10182-022-00467-3.

[21] S. Page and S. Grünewälder, "The Goldenshluger–Lepski method for constrained least-squares estimators over RKHSs." *Bernoulli* vol. 27, no. 4, pp. 2241 - 2266, Nov. 2021. doi.org/10.3150/20-BEJ1307

[22] L. Liu, R. Mukherjee, J. Robins and E. T. Tchetgen, "Adaptive estimation of nonparametric functionals", *The J. of Mach Learn Res*. vol. 22. no. 1. pp. 1-66, January 2021. JMLR.org.

[23] S. Piradl, A. Shadrokh and M. Yarmohammadi, “A robust estimation method for the linear regression model parameters with correlated error terms and outliers”. *J Appl Stat*. vol. 49, no. 7 pp.1663-1676. Feb. 4. 2021. doi: 10.1080/02664763.2021.1881454. PMID: 35707565; PMCID: PMC9042062.

[24] R. Agoune and F. Merahi, “Estimation in nonparametric functional-on regression models with real responses”, *A Can J. of appl math, com sci and stat*, vol. 120, no. 1, pp. 1954–1962, May 2024.

[25] G. James, D. Witten, T. Hastie and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R", Second Edition, Springer-Crossmark. 597, pp. 1-14. Jun. 2021.

[26] G. Kriska, "Identification Key: Freshwater Invertebrates. In: Freshwater Invertebrates in Central Europe". *Springer, Cham.* no. 1. pp. 1-20. Jan. 2022.

[27] J. E. Brower, J. H. Zar, and C. von Ende, “Field and Laboratory Methods for General Ecology”. WCB Publishers, Dubuque. Vol. 1990, no.1, pp.3-7. Feb. 1990.

[28] M. Morisita, “Application of I_{δ}-Index to sampling techniques”. *Res. Popul. Ecol.*, vol. 6, no. 12. pp.43–53. Dec. 1964. doi.org/10.1007/BF02518805

[29] J. J. Hayes and O. Castillo, “A New Approach for Interpreting the Morisita Index of Aggregation through Quadrat Size”. ISPRS Inter J of Geo-Inf, vol.6. no. 10. 296 pp. Sep. 2017. doi.org/10.3390/ijgi6100296.

[30] M. A. J. Torres, L. R. T. Amarillo, R. Joshi, L. Sebastian, Z. Baoanan and C. G. Demayo, “Intra- and Inter-Population Variability in the Golden Apple Snail Pomacea canaliculata”. *Aust. J. Basic and Appl. Sci.* vol. 5. no.6. pp.789-800. Jun. 2011.

[31] B. Kargoll, M. Omidalizarandi, I. Loth, J.A. Paffenholz and H. Alkhatib, “An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations”. *J Geod*, vol. 92. No. 1, pp. 271–297. Mar. 2018.

[32] X. Xu, J. Wu, Y. Yang, R. Zhu and S. Yan, "Operculum of a Water Snail is a Hydrodynamic Lubrication Sheet". *J Bionic Eng* vol. 15, no. 1, pp. 471–480. May. 2018.

[33] K. L. Constantine, F. Makale, I. Mugambi, D. Chacha, H. Rware, A. Muvea, V. K. Kipngetich, J. Tambo and A. Ogunmodede, “Assessment of the socio-economic impacts associated with the arrival of apple snail (*Pomacea canaliculata*) in Mwea irrigation scheme, Kenya”. *Pest Manag Sci*. vol.79, no. pp.4343–4356. Jul. 2023. doi.org/10.1002/ps.7638

[34] Y. X. Yang, Y. C. Hu, X. H. Li, X. J. Wang, X. D. Mu and H.M. Song, “Historical invasion, expansion process and harm investigation of Pomacea canaliculata in China”. *Chin Agric Sci Bull.* vol.26: no.5. pp.245–50. Mar. 2020.

[35] F. X. Wagiman, Harisma, B. Triman and J. B. Ariani, “Feeding Strategy and Feeding Capacity of duck (Anas platyrhyncos domesticus) as a Predator of Golden Snail (Pomacea canaliculata).” *Inter J. of Adv in Pharm, Bio, and Chem (IJAPBC)*, vol. 4, no. 2, pp.491-495. Apr-Jun. 2015.

[36] J. A. Bunga, N. J. Lapinangga, and J. H. H. Sonbai,”Tumbuhan inang dan daya makan keong mas (*Pomacea canaliculata*) pada beberapa varietas padi di kabupaten malaka”. *Partner. v*ol. 23, no. 2, pp. 822-831. Nov. 2023. doi.org/10.35726/jp.v23i2.324

[37] Y. Wang, "Smoothing Splines: Methods and Applications". Paperback, Series: Chapman and Hall/CRC. *Mon on Stat and Appl Prob*. vol. 394, no. 1, pp. 1-13. Apr. 2023.

[38] N. Dai, "Inference for penalized spline regression: Improving confidence intervals by reducing the penalty". *arXiv: Method.* vol. 1706. 00865v1.[stat.ME]. no. 1, pp. 1-22. Jun. 2017.

[39] M. Y. Mustafa and Z. Y. Algamal, "Smoothing parameter selection in kernel nonparametric regression using bat optimization algorithm", *J. Phys. Conf. Ser* vol. 1897 no. 012010 pp. 1-9. Nov. 2021. doi:10.1088/1742-6596/1897/1/012010