Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Four parameter beta GLMM Bayesian inference approach: Improving paddy productivity predictions for area yield index crop insurance

Dian Kusumaningrum[a,b,d*], Hari Wijayanto[b*], Khairil Anwar Notodiputro[b], Anang Kurnia[b], Muhlis Adrainsyah[c] and Islam MD Parvez[d]

[a]*Business Mathematics Program Study Prasetiya Mulya University, BSD, Tangerang, Indonesia*
[b]*Statistics and Data Science Department, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, West Java, Indonesia*
[c]*BPS-Statistics of Kotawaringin Timur, Central Kalimantan, Indonesia*
[d]*Department of Biomechanical Systems, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama 790-8566, Japan*

C H R O N I C L E

Received May 23, 2025
Received in revised format June 28, 2025
Accepted August 28 2025
Available online August 28 2025

Keywords:
Area Yield Index Crop Insurance
Paddy Productivity Prediction
Four Parameter Beta Distribution
Generalized Linear Mixed Model (GLMM)
Bayesian Approach

A B S T R A C T

Paddy is a staple crop and a vital component of Indonesia's agriculture, significantly contributing to food security and rural livelihood. Nevertheless, paddy cultivation is highly vulnerable to risks such as pests, diseases, extreme weather, and natural disasters, which can lead to significant productivity losses. Thus, schemes like Area Yield Index (AYI) insurance have a critical role in mitigating these risks because. It provides financial protection to farmers by compensating them for losses due to area-wide productivity shortfalls. Hence, accurate predictions of paddy productivity are essential for setting fair and precise AYI premiums. Therefore, this study proposes an innovative framework by developing a four-parameter beta distribution Generalized Linear Mixed Model (GLMM) based on a Bayesian approach for predicting paddy productivity. The approach is motivated by the model's ability to apply a four-parameter beta distribution that effectively models the bounded nature of paddy productivity and ensures that predictions remain within realistic range value. The inclusion of random effects also accounts for variability of paddy productivity across areas, which is commonly found in Indonesia and other countries. Meanwhile, the Bayesian framework further enhances robustness by integrating prior knowledge and providing probabilistic predictions. Based on the proposed approach, we then design an enhanced AYI policy based on district and sub district conditions. The framework is first developed through simulation studies designed to replicate real paddy productivity conditions. Comparative testing of the Stan and BRMS packages in R reveals that the proposed four-parameter beta GLMM implemented in Stan is more flexible and accurate. The methodology is then applied to an empirical case study predicting paddy productivity in Central Kalimantan (2020), using farmer survey data and lagged values of Sentinel-2A satellite indices (bands 4, 8, and NDVI) as covariates. Results show that agronomic practices such as pest management and current and historical satellite data enhance prediction accuracy, demonstrating the model's potential to predict productivity with high precision, proving that the proposed method is well-suited for calculating premiums and risks under AYI crop insurance policies. The estimated pure AYI premium ranges from IDR 300.000 to 410.000. Unlike conventional premium calculations based on average historical yields, the proposed GLMM approach provides a nuanced, data-driven alternative that accounts for various productivity factors, ensuring greater adaptability, accuracy, and responsiveness to changes in agricultural conditions, including those driven by climate change.

© 2026 by the authors; licensee Growing Science, Canada.

1. Introduction

Paddy is a staple crop and a basis of Indonesia's agricultural sector because it is the primary source of food for around 284 million Indonesians and it is also a crucial component in ensuring food security. It plays a vital role in the national economy by supporting rural livelihoods, generating employment, and providing a stable income for farmers across Indonesia's diverse

* Corresponding author
E-mail address dian.kusumaningrum@prasetiyamulya.ac.id (D. Kusumaningrum) hari@apps.ipb.ac.id (H. Wijayanto)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)
© 2026 by the authors; licensee Growing Science, Canada.
doi: 10.5267/j.ijdns.2025.9.003

archipelago. Given its substantial contribution to household food consumption and national stability, the Government of Indonesia (GoI) has implemented strategies to enhance paddy production, focusing on improving productivity through technological advancements, pest and disease control, and business-based food production (Zainul et al., 2020). However, Indonesia's diverse agroclimatic conditions, limited resources, pest outbreaks, and high input costs pose significant challenges to these efforts. Given these vulnerabilities, crop insurance plays a vital role in mitigating risks, protecting farmers' livelihoods, and supporting sustainable paddy productivity.

Area Yield Index (AYI) crop insurance policy is a productivity-based index insurance policy designed to provide financial protection for farmers. In this policy, indemnities are calculated based on the average productivity of an area level. Payouts will be given when the actual average yield falls below the benchmark area level set by the insurers. In Indonesia, AYI is considered as an alternative promising policy that can be offered to farmers. Thus, a pilot study conducted by BAPPENAS has shown benefits, such as reduced moral hazard, minimal adverse selection, and lower administrative costs. However, challenges still exist, particularly the lack of reliable historical data needed to calculate benchmark productivity at area levels and insufficient estimates of monthly or seasonal productivity at lower area levels. Statistics Indonesia (SI) has conducted monthly paddy productivity estimates based on Crop Cutting Experiments (CCEs). Nevertheless, the accuracy and consistency of predictions need to be enhanced to ensure reliable benchmark productivity calculation. Current estimates may not sufficiently capture productivity variations at sub-districts or villages, which are critical for precise indemnity calculations (Shynkarenko et al., 2019). Addressing these challenges will make paddy productivity predictions more robust and effective, enhancing the implementation and success of AYI insurance schemes. Previous techniques used to predict paddy productivity such as Exponential Smoothing, ARIMA, Regression and Machine Learning models have been developed and shown to be promising. Nevertheless, current models do not account for specific paddy productivity distribution that is bounded to a to a specific maximum and minimum value and regional heterogeneity. Therefore, we propose an innovative framework by developing a four-parameter beta distribution Generalized Linear Mixed Model (GLMM) based on a Bayesian approach.

A four-parameter beta distribution is a distribution bound to a specific maximum and minimum value. Although it is not as common as beta distribution, the four-parameter beta distribution is more flexible because it is not restricted to values between 0 and 1. It can also accommodate a wide range of shapes, skewness and even heavy tails, allowing econometric models to capture different patterns in the data. This flexibility is valuable when dealing with different economic phenomena that have different characteristics. Therefore, a model developed based on a four-parameter beta distribution can be used as predicting paddy productivity for calculating the risks and premium of AYI. This type of distribution can be valuable for capturing the variability and characteristics of crop productivity, which is critical for developing effective insurance models such as AYI. Further on, this distribution can also be applied in other economic and business domains, such as modeling inflation, growth, index, rates, etc. For example, the average growth of Gross Domestic Product (GDP) in a country or region over the last ten years has a minimum value and can only reach a maximum value. Hence, the versatility of this distribution is potential for analyzing and predicting patterns across various fields, fostering better decision- making and strategic planning.

In contrast to unbounded data, it is known that for bounded data, central tendency measures, skewness, and other characteristics of the underlying distribution are reliant on the distribution's support or range of values for which the Probability Density Function (PDF) is non-zero. As a result, greater attention must be taken when drawing conclusions for these features based on bounded data, especially when the support parameter is unknown (Zhou & Huang, 2022). Nevertheless, existing approaches for analyzing the four-parameter beta distribution are limited. Most references have opted to transform the data into a beta distribution that has a prefixed support of (0, 1) and applied a beta mean regression (Ferrari & Cribari-Neto, 2004). Furthermore, the model has been expanded to permit covariates to affect the precision parameter (Smithson and Verkuilen, 2006). On CRAN, you can find the betareg R package (Cribari-Neto and Zeileis, 2010; Grün et al., 2012) that can be used to fit the beta mean regression model with different levels of precision and perform model diagnostics. However, we must keep in mind that in many cases, the transformation process can lead to biases in the parameter estimates and complications in the interpretation of coefficient values (Kusumaningrum et al., 2024a,b).

Recently, Zhou and Huang (2022) developed the mean and mode four parameter beta regression model based on a Bayesian approach. They also provided the R package betaBayes, which estimates the parameters and provides model evaluations. Nevertheless, the inclusion of a random effects component to a model is often very important. As in econometrics or paddy productivity predictions, researchers often analyze panel data and cross- sectional data. In panel data, observations are collected over time for multiple individuals or units. Therefore, random effects can capture individual-specific effects that vary over time and unobserved heterogeneity across units. In contrast for cross sectional data, where data are collected by observing multiple units at a single point in time, the use of random effects is less common than for panel data. Nevertheless, there are many cases in which heterogeneity across units occur. Hence, we further developed Zhou and Huang’s regression model by including a random effects component and introducing the four-parameter beta mean Generalized Linear Mixed Model (GLMM). This model was developed based on a Bayesian inference by block sampling Markov Chain Monte Carlo (MCMC) approach. This gives us the advantages of (1) handling uncertainty within the specific complex model framework, (2) making inferences for all unknown parameters of the model, including the fixed effects, random effects, precision, and support bounds (minimum and maximum value), (3) obtaining probabilistic predictions, prediction intervals, and information about the distribution of possible outcomes, and (4) appropriately selecting a prior distribution for setting parameter restrictions.

To evaluate the developed model, we conducted a simulation study. The analysis of the accuracy and efficiency of the estimated parameter as well as the accuracy of the predictions are the focus of the evaluation process. Subsequently, this model is applied to predict paddy productivity in the Indonesian province of Central Kalimantan. This province was chosen because it is one of the provinces in Indonesia that has been selected for Indonesia's food estate program and the level of productivity among areas in this province also tends to be diverse. Paddy productivity itself has a bounded distribution with a certain minimum and maximum value. Paddy productivity conditions usually vary from area to area as they depend on agroclimatic conditions. Nevertheless, accurate predictions of paddy productivity in Indonesia are very important. Not only because rice is a staple food in Indonesia, but also because Indonesia has started to conduct pilot studies and introduce an alternative Area Yield Index crop insurance policy from 2022 (Sanyu Consultants Inc. and Sompo Risk Management Inc, 2023). This policy is a productivity-based index agricultural insurance policy that is highly dependent on accurate paddy productivity predictions in various areas.

The remainder of this article is organized as follows. Section 2 describes the materials and methodology used to develop the four-parameter beta GLMM in this paper. We begin with an introduction to the four-parameter beta distribution and then explain the four-parameter beta regression and its further development. We then explain the Bayesian inference in subsection 2.3. Here, we have described the model specifications, the selection of priors, the estimation and evaluation of parameter models, the model selection criteria, and the predictions. In subsection 3, the simulation process is described to demonstrate the accuracy of the inference results compared to the relevant and current techniques. Section 4 illustrates the proposed methodology and results in the empirical study of paddy productivity predictions for crop insurance policy. Finally, Section 5 presents the conclusions, summarizing the results of our research and outlining potential future research.

2. Proposed Method Framework

2.1 Four parameter beta distribution and four parameter beta regression model

The four-parameter beta distribution is one of the continuous variable distributions that have an interval of (a, b) and has the following probability density function:

$$f(y;\alpha,\beta,a,b)=\frac{\Gamma(\alpha+\beta)(y-a)^{\alpha-1}(b-y)^{\beta-1}}{\Gamma\alpha\Gamma\beta B(\alpha,\beta)(b-a)^{\alpha+\beta-1}},a\le y\le b,\alpha>0,\beta>0 \tag{1}$$

where α is the location parameter and α-β is the scale parameter. Meanwhile, a is the minimum value and b is the maximum value. Furthermore, the mean of a four-parameter beta distribution is $\mu=a+\frac{\alpha}{\alpha+\beta}(b-a)$, while the variance of the distribution is $\sigma^2=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)}(b-a)^2$. The four-parameter beta distribution is highly flexible and can model a wide range of shapes, including symmetric and skewed distributions. This flexibility makes it suitable for fitting a variety of data patterns. Modeling data that has a four-parameter beta distribution is not commonly found. Previously, most references suggested to conduct a transformation of the response variable from a four-parameter beta distribution into a beta distribution (Gray & Alava, 2018) and apply a Beta GLMM (Bonat et al., 2015) that can be formulated as

$$\boldsymbol{\eta}_i=\mathbf{X}_i^T\boldsymbol{\beta}+\mathbf{Z}_i^T\mathbf{b}_\mathbf{i} \tag{2}$$

In this model we assume that $Y\sim B(\mu,\varphi)$ and it is known that $\mathbf{X}_i^T$ and $\mathbf{Z}_i^T$ are fixed effect matrix design and random effect matrix design. The fixed effect and random effect estimators in the model are denoted as $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$. It is assumed that $\mathbf{b}_i\sim \text{iid}\, N(\mathbf{0},\boldsymbol{\sigma}_i^2)$ or denoted as $[\mathbf{b}_\mathbf{i}\mid\boldsymbol{\Sigma}]\sim N(\mathbf{0},\boldsymbol{\Sigma})$. The logit link function is often used in GLMM, where $\boldsymbol{\eta}_\mathbf{i}=g\left(E\left(\mathbf{Y}_i\right)\right)=g\left(\boldsymbol{\mu}_\mathbf{i}\right)=\log\left(\frac{\boldsymbol{\mu}_\mathbf{i}}{\mathbf{1}-\boldsymbol{\mu}_\mathbf{i}}\right)$. The parameters of this model are estimated by maximizing the marginal likelihood function obtained by integrating the combined probability density function [Y, b]. Nevertheless, transformation process tends to lead to bias in parameter estimates and it is also sometimes complicated to interpret the results. Zhou and Huang (2022) show that if it is known that the response variable has a four-parameter beta distribution or $Y\sim B(\alpha,\beta,a,b)$, then $W=\frac{Y-a}{b-a}$ will also have a beta distribution with shape parameters α and β. Zhou and Huang (2022) then developed the mean four parameter beta mean regression model that can be formulated as follows:

$$g\left[\frac{\mathbf{mean}\left[\mathbf{y}_\mathbf{i}\mid\mathbf{x}_\mathbf{i}\right]-a}{b-a}\right]=\mathbf{X}_\mathbf{i}^\mathbf{T}\boldsymbol{\beta} \tag{3}$$

The left side of the equation can be interpreted as the quantile scores for the position of the mean $\left[\mathbf{y_i} \mid \mathbf{x_i}\right]$ in the range (a, b). As known previously, $\mathbf{X}_i^T$ and $\boldsymbol{\beta}$ are fixed effect matrix design and fixed effect parameters, where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_K)$.

2.2 Four parameter beta mean generalized linear mixed model (GLMM) with Bayesian approach

In this paper we will introduce an advancement of the four-parameter regression model, where in this model we assume that the response variable has a four-parameter beta distribution $Y_i \sim B(\alpha, \beta, a, b)$ that is measured at a certain entities/groups/areas and i=1, ..., q. It is also assumed that the response variable does not only depend on the independent variables $X_i = (X_{1i}, ..., X_{Ki})$ but it is also affected by the variance between entities/groups/areas (McCullagh & Nelder 1989). Therefore, the advanced and novel four parameter beta mean GLMM can be formulated as:

$$g\left[\frac{\mathbf{mean}\left[\mathbf{y_i} \mid \mathbf{x_i}\right] - a}{b - a}\right] = \mathbf{X_i^T}\boldsymbol{\beta} + \mathbf{Z_i^T}\mathbf{b_i} \tag{4}$$

The random effect matrix design is denoted by $\mathbf{Z}_i^T$, while $\boldsymbol{b_i}$. represents the random effect estimators in a particular entity/group/area. Following Bonat et al. (2015), we assumed that $\boldsymbol{b_i}$~iid N(0, σ^2) and applied the logit link function. In this model a greater mean value results from a higher quantile-score. The quantile-score for the mean value increases by β_j units for every unit increase in x_{ij}, where j = 1,..., K. Meanwhile, $\boldsymbol{b_i}$ represents the difference in the quantile-score for the mean value that can be attributed to entities/groups/areas that are not directly included in the model. These factors may include unmeasured individual characteristics or time-varying confounding variables that are not accounted for in the fixed effects of the model.

2.3 Bayesian inference for the four-parameter beta mean generalized linear mixed model (GLMM)

Inferences of the developed model is based on Bayesian method. There are three main aspects of Bayesian approach in developing a four-parameter beta GLMM, which includes 1) model specification, 2) prior distribution selection, 3) parameter model estimation and evaluation, 4) Model Selection and predictions. Each aspect will be explained in more detail below.

To specify the four-parameter beta mean GLMM, as described in Zhou and Huang (2022), we must first set $W = \frac{Y - a}{b - a}$.

Let μ_W and μ_Y denote the mean for W and Y, respectively. Thus $\mu_W = \frac{\alpha}{\alpha + \beta}$ and $\mu_Y = \alpha + \frac{\alpha}{\alpha + \beta}(b - \alpha) = \alpha + \mu_W(b - a)$.

Next, we can set $\alpha = \varphi m$ and $\beta = \varphi(a - m)$, for $0 < m < 1$ and $\varphi > 0$, where φ is the precision parameter. A larger precision value will lead to a lower variance of the four-parameter beta distribution. The parameters α and β in the beta distribution control the shape of the distribution. By introducing m, you can directly link m to the mean of the four-parameter beta distribution because this reparameterization process will set the value of μ_W to be equivalent to m. Through this reparameterization we will also directly link φ to a scaling factor of the four-parameter beta distribution. In most econometric applications it will be more meaningful to express the four-parameter beta distribution in terms of mean. Furthermore, in analyzing the results of the model, it will be more interpretable to analyze the mean of the distribution rather than analyzing the distribution solely in terms of α and β. To estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{b_i}$, alongside the minimum (a) and maximum (b) value, and the precision value (φ) of the four-parameter beta mean GLMM a Bayesian approach will be used, which is specified below

$$\begin{aligned} & y_i \mid m_i, \varphi, a, b \sim Beta4\left(\varphi m_i, \varphi(1 - m_i), a, b\right) \\ & \mathbf{m_i} \equiv m(\mathbf{x_i}) = g^{-1}\left(\mathbf{X_i^T}\boldsymbol{\beta} + \mathbf{Z_i^T}\mathbf{b_i}\right) \\ & p\left(\boldsymbol{\beta}, \mathbf{b_i}, \varphi, a, b\right) = p(\boldsymbol{\beta})p(\mathbf{b_i})p(a)p(\varphi) \end{aligned} \tag{5}$$

where $g(.)$ is a link function, commonly logit, but also can be probit, and log-log. While $p(.)$ represents a prior density. After defining the model specifications of the four-parameter beta mean GLMM we need to define the prior distributions for $\theta = (\beta, b_i, a, b, \varphi)$ by specifying the parameters probability distributions (e.g. normal, uniform, gamma, etc.). The main aim of Bayesian statistics is to find an appropriate prior distribution of model parameters with a probability distribution and use the current data to update the prior to create a posterior distribution with reduced uncertainty (Lynch, 2007). Nevertheless, selecting the right choice of prior tends to be challenging. In general, there are three categories of prior distribution, which are: non-informative, informative, and weakly informative priors. A non-informative or flat prior gives each value an equal probability and does not consider any previous information. Alternatively, informative prior typically characterizes a particular range of parameter values and expresses a stronger pre-existing belief because it is commonly derived from previous empirical research. Meanwhile the weakly informative priors aim to balance between the first two by including "weak" parameter information that covers all potential "real-world" values without assigning any value with an excessively high probability (Johnston et al., 2023).

Referring to Zhou and Huang (2022) it is suggested that the prior's distribution should be specified as:

i. **Prior for Minimum (a) and Maximum Value (b)** is a uniform distribution $a \sim \text{unif}(a_1, b_1)$ and $b \sim \text{unif}(a_2, b_2)$. If we have the information of the previous conditions we can set $y_{(1)}$ ($a_1 < y_{(1)} < b_1$) and $y_{(k)}$ ($a_2 < y_{(k)} < b_2$). Where $y_{(k)}$ refers to the k-th order statistic $\{y_{(1)}, .., y_{(k)}\}$. Nonetheless, if we do not have the previous knowledge of these boundaries, we can set $a_1 = y_{(1)} - \Delta$, $b_1 = Y_{(1)} - 10^{-15}$, $a_2 = Y_{(n)} + 10^{-15}$, $b_2 = Y_{(n)} + \Delta$, where $\Delta > 10^{-15}$ or $\Delta = 2s_y$, where s_y is the sample standard deviation of y_i's

ii. **Prior for the Precision Parameter (φ)** can be inferred from the observed data. Nevertheless, when there is limited information, priors can be set to a gamma distribution, where $\Gamma(a_\varphi, b_\varphi)$ and $a_\varphi = b_\varphi = 0.001$ are the defaults. Therefore, we believe that the precision parameter is small and does not heavily influence the posterior.

iii. **Prior for the Fixed Effect Parameters (β)** must take into attention the prior information while adjusting for covariates. For this purpose, Zhou and Huang (2022) considered the g-prior (Zellner, 1986) where $\beta \sim N_{p+1}\left(be_1, gn(X'X)^{-1}\right)$, $e_1 = (1, 0, ..., 0)'$ is of length $p + 1$, b is a prior mean for the intercept, and $g > 0$ is a scaling constant. In our purposed model we can also consider using parameter estimates of a four-parameter beta regression model as the prior for the fixed effect (**β**) because this model is an advance extension of the four-parameter beta regression.

iv. **Prior for Random Effect Parameters($\boldsymbol{b_i}$)** is commonly based on the normal distribution. Therefore, we can set $\boldsymbol{b_i} \sim iid\ N(0, \boldsymbol{\Sigma})$ and the variance in the matrix variance covariance ($\boldsymbol{\Sigma}$) should be positive definite. In this paper, we have applied informative prior and weakly informative priors because often non-informative priors will lead to bias parameter estimates, especially when the sample size is small (LeMoine, 2019). Besides that, in selecting the priors we have considered relevant research studies. Incorporating the priors and the likelihood function allows us to obtain Bayesian estimation of the parameters of a four parameter beta mean GLMM through the posterior distributions. Given data $D = \{(y_i, x_i), i = 1,2,..,n\}$, the likelihood function of parameters $\theta = (a, b, \varphi, \boldsymbol{\beta}, \boldsymbol{b_i})$ associated with the four parameter beta mean GLMM can be given as

$$L(\theta) = \frac{\Gamma(\alpha_i + \beta_i)(y_i - \alpha)^{\alpha_i - 1}(b - y_i)^{\beta_i - 1}}{\Gamma(\alpha_i)\Gamma(\beta_i)(b - a)^{\alpha_i + \beta_i - 1}} \tag{6}$$

where $\alpha = \varphi m_i$ and $\beta = \varphi(1 - m_i)$, for $0 < m_i < 1$ and $\varphi > 0$. The posterior density than can be calculated as

$$p(\boldsymbol{\beta}, \mathbf{b}_i, \varphi, a, b, | D) \propto L(\theta) \times \exp\left\{\frac{-1}{2gn}(\boldsymbol{\beta} - \mathbf{be_1})\mathbf{X'X}(\boldsymbol{\beta} - \mathbf{be_1})\right\} \times \exp\left(\frac{-1}{2}\mathbf{b_i}\boldsymbol{\Sigma}^{-1}\mathbf{b_i}\right) \times I(a_1 < a < b_1) I(a_2 < b < b_2) \times \varphi^a \varphi^{-1}(-b_\varphi \varphi)$$

Posterior sampling is than conducted based on adaptive Metropolis samplers (Haario et al., 2001**)**. There was a consideration that central tendency measures of bounded data typically entangle with the support of the underlying distribution. Thus, the posterior distribution of $\boldsymbol{\beta}$ and (a,b) are often highly correlated. Zhou and Huang (2022) have considered this problem and updated it into a single block to effectively eliminate problematic MCMC mixing. It has been well documented in the literature that block sampling can improve MCMC efficiency relative to updating each parameter independently (Liu et al.,1994**;** Roberts and Sahu, 1997; Sargent et al., 2000). The algorithm for posterior sampling that incorporates block sampling can be seen in Zhou and Huang (2022).

Prediction in the four-parameter beta mean GLMM implies estimating the response variable based on fixed and random effects while accounting for the four parameter beta distribution's constraints $(a\ and\ b)$ and the link function. As we know that the response is modeled between bounds $a\ and\ b$ and it depends on shape parameters α and β that are derived from the predicted mean μ and a precision parameter ϕ. The stages of obtaining a prediction are:

1. We have $\boldsymbol{\theta}^{(s)} = \left(a^{(s)}, b^{(s)}, \varphi^{(s)}, \boldsymbol{\beta}^{(s)}, \mathbf{b}_i^{(s)}\right)$ after generating posterior samples using the methods explained above. Here, (s) refers to the index of posterior samples obtained from MCMC.

2. We calculate the linear predictor ($\boldsymbol{\eta}_i$) by using the posterior samples of $\boldsymbol{\beta}^{(s)}$ and $\mathbf{b}_i^{(s)}$. First, recall the modelling framework $\boldsymbol{\eta}_i^s = g(\boldsymbol{\mu}_w^s) = \mathbf{X}_i^T\boldsymbol{\beta}^s + \mathbf{Z}_i^T\mathbf{b}_i^s$ where g is the logit link function, solve for $\boldsymbol{\mu}_w^s$, then $\boldsymbol{\mu}_w^s = g^{-1}\left(\mathbf{X}_i^T\boldsymbol{\beta}^s + \mathbf{Z}_i^T\mathbf{b}_i^s\right)$.

3. Next, apply the inverse logit function to obtain the predicted mean of the response variable It is known that $g^{-1}(z) = \frac{\exp(z)}{1 + \exp(z)}$ hence, $\boldsymbol{\mu}_w^s = \frac{\exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}{1 + \exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}$. Due to the reparameterization described in section 2.2, it is known that $\boldsymbol{\mu}_l^s = \frac{\text{mean}\left[\mathbf{y_i} \mid \mathbf{x_i}\right]^s - a^s}{b^s - a^s} = \frac{\exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}{1 + \exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}$ and $\text{mean}\left[\mathbf{y_i} \mid \mathbf{x_i}\right]^s = a^s + \left(b^s - a^s\right)\frac{\exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}{1 + \exp\left(\mathbf{X_i^T}\boldsymbol{\beta}^s + \mathbf{Z_i^T}\mathbf{b_i^s}\right)}$. This gives the predicted mean of the response variable in the four-parameter beta distribution for each sample (s).

4. Drive the beta four parameter distribution parameters a^s and b^s for each MCMC sample. We have obtained the predicted mean of the response variable for each MCMC sample. Nevertheless, the distribution of the response variable also consists of the shape parameters α and β. These parameters indicate how skewed or peaked the distribution is and most importantly the variance of the distribution. The precision parameter (φ) indicates the variance and is related to both α and β. Zou and Huang (2022) explained that α^s and β^s can be driven by $\alpha^s = \varphi^s m$ and $\beta^s = \varphi^s(1-m)$, where $m \equiv \mu_w^s$ and $0<m<1$ and $\varphi < 1$.

5. Generate the predicted response $Y_i^s \sim B\left(\alpha^s, \beta^s, a^s, b^s\right)$ based on the estimated α^s and β^s obtained from step 4 and the estimated a^s and b^s obtained from step 1.

6. Summarize the predictions that have been given in step 5. We have $\mathbf{Y}_i^s$ for $s = 1, ..., S$. Therefore we can calculate summary statistics, such as:

a. The posterior predictive mean can be calculated based on $E\left[\mathbf{Y}_i\right] = \frac{1}{S}\sum_{s=1}^{S}\mathbf{Y}_i^s$

b. The posterior predictive median: Sort the simulated predictions: $\left\{Y_i^{(1)}, ..., Y_i^{(s)}\right\}$. Afterwards, we can compute the median by taking the middle value based on the number of samples (S) , when S is odd. If S is even the median is the average of the two middle values.

c. The posterior variance and other relevant statistical values

The posterior variance can be calculated based on $Var\left[\mathbf{Y}_i\right] = \frac{1}{S}\sum_{s=1}^{S}\left(\mathbf{Y}_i^s - E\left[\mathbf{Y_i}\right]\right)^2$ while other relevant statistics can be also calculated based on the formulation examples given.

d. Construct credible interval (α) for the predicted responses. If α =0.05, then we will have the 95% credible intervals for the predicted responses. Like calculating the median, we first need to sort the simulated predictions: $(Y_i^1, ..., Y_i^s)$. Next, we need to identify the lower and upper bound, where the lower bound CI_{low} and upper bound CI_{high} are formulated as $CI_{low} = Y_i^{\left(\frac{\alpha}{2}S\right)}$ and $CI_{high} = Y_i^{\left(\left(1-\frac{\alpha}{2}\right)S\right)}$ where $\lceil x \rceil$ is the ceiling function that rounds up to the nearest integer. The final credible interval is $\left[Y_i^{low}, Y_i^{high}\right] = \left[CI_{low}, CI_{high}\right]$. Suppose we S=10,000 and α =0.05, then we can take the 2.5th percentiles and 97.5th percentiles of the sorted predictions. The final credible interval is $\left[Y_i^{[250]}, Y_i^{9750}\right]$.

These measures summarize the posterior predictive distribution of the response variable and provide insights into its central tendency, spread, and uncertainty.

3. Simulation study

3.1 Simulation design and process

We construct a simulation study to evaluate the effectiveness of Bayesian approach that was applied to develop the novel four parameter beta mean GLMM. In general, there are ten main steps in the simulation, which can be seen in Fig. 1.

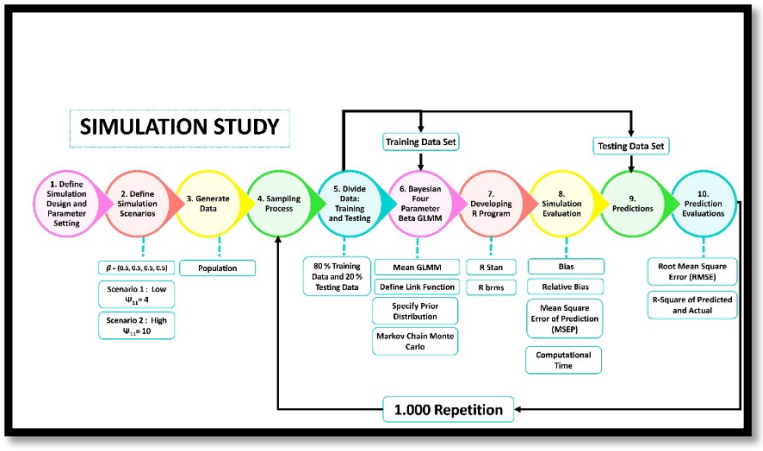


Fig. 1. Simulation Process Flowchart

The pseudo population was designed to mimic the characteristics of a real population of interest. Therefore, two scenarios were developed in the simulation. A more detailed explanation of the stages of the simulation is given below.

1. Define Simulation Design and Parameter Setting

In this step, we must define the parameters and simulation design below:

i. Determine the number of areas (Area = 10) and the total population in each area (N_{Area} =20.000)
ii. Determine K independent variables ($X_1, \ldots, X_K$) with its distribution and the random effect with its distribution. In our study K = 3, thus
- $X_1 \sim N(0, 5)$ is the first independent variable that influences the response variable
- $X_2 \sim Bin(n, 0.5)$ is the second independent variable that influences the response variable
- $X_3 \sim N(0, 5)$ is the third independent variable that does not influence the response variable
- $\boldsymbol{b_i} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ is the random effect of area, where $\boldsymbol{\Sigma} = \boldsymbol{\psi_{11}}$ for random intercepts in the GLMM model

iii. Note that the assumptions underlying the simulation design were made based on previous relevant studies and empirical data provided by Statistics Indonesia (SI) that will also be used in the case study.

2. Define Simulation Scenarios

i. Determine the fixed effect $\boldsymbol{\beta}$, the random effect $\boldsymbol{b_i}$, and $\boldsymbol{\psi_{11}}$ based on the scenarios given below

Table 1
Data Generating Scenarios in the Simulation Process

Scenario	Random Effect	Ψ_{11}	Fixed Effect
1	Low	4	$\boldsymbol{\beta}$ = (0.5, 0.5, 0.5, 0.5)
2	High	10	$\boldsymbol{\beta}$ = (0.5, 0.5, 0.5, 0.5)

ii. The considerations in developing the scenarios include:

- In this study we limit the scope to only develop a four-parameter beta mean GLMM that has random intercepts.
- The fixed effect parameter ($\boldsymbol{\beta}$) was set to 0.5.

While designing the random effects, researchers considered two specifications (low and high) for the covariance matrix. As we have known that the variability of productivity among areas in Indonesia is a common problem. Thus, considering different magnitude levels of inter-group variability will be a beneficiary.

3. Generate Data

i. Generate the independent variables (X_1, X_2., X_3) and the random effect based on the simulation design explained above.
ii. Generate the response variables $Y_i \sim B(\alpha, \beta_i, a, b)$ based on the initialization and scenario above. Thus, we must determine m_i, α_i, β_i, a, *and* b based on empirical data provided by Statistics Indonesia (SI). Then, we will have:

$$\mathbf{m}_i \equiv m(\mathbf{x}_i) = g^{-1}\left(\mathbf{X_i^T}\boldsymbol{\beta} + \mathbf{Z_i^T}\mathbf{b_i}\right),\ \alpha_i = \varphi m_i \text{ and } \beta_i = \varphi(1-m_i)\ \ \alpha \sim unif\left(a_1, b_a\right) \text{ and } b \sim unif\left(a_2, b_2\right).$$

iii. ExtDist package in *R* was used to generate the four-parameter beta distribution response variable (Y_i).

4. Sampling Process

i. Determine the sample size (n_i) for each area i =1, 2,…I.
ii. The sample size for each area will be calculated based on a stratified proportional sampling and the sampling because we will take a sample for each area and the sample sizes of each area will depend on the size of the population itself.
iii. Select data randomly for each area based on the calculated sample sizes.
iv. Repeat the Sampling process (B = 1000 times).
v. Repeat Steps 1-10 for each scenario.

5. Divide Data: Training and Testing

i. Determine the proportion of training data (80%) and testing data (20%).

ii. Divide the data into training and testing data according to predetermined proportions.
iii. Training Data will be used to build the novel four parameter beta mean GLMM model. In this step we will obtain the parameter estimates of a four-parameter beta mean GLMM $\boldsymbol{\theta} = (a,b,\varphi, \boldsymbol{\beta}, \boldsymbol{b_i})$.
iv. Testing Data will be used for predictions and evaluation of prediction results from the novel four parameter beta mean GLMM. In this step we will obtain predictions of the developed model, $y_{i.}$

6. Bayesian Four Parameter Beta GLMM

i. Specify Model based on Eq. (4).
ii. Define the prior distributions.
iii. Estimate posterior distribution based on a Bayesian Approach, where the likelihood function is shown in equation (6).
iv. Evaluate parameter estimates of the $\hat{\boldsymbol{\theta}} = \left(\hat{a},\hat{b},\hat{\varphi},\hat{\boldsymbol{\beta}},\widehat{\mathbf{b}_i}\right)$.

7. Developing R Program

i. For the computational process of developing the advanced and novel four parameter beta mean GLMM and the simulation process we will use R software and have utilized it.
ii. First, we will utilize R-Stan, which is a well-known and robust Bayesian analysis platform available for a variety of analytical programming languages (Stan Development Team, 2022). Stan has the benefit of a strong development and support community. Stan is also the foundation for many Bayesian-oriented R packages, such as "RStan" (Stan Development Team, 2021). These packages provide comparable functionality. Although it offers the utmost flexibility, nevertheless it is not user-friendly for novice users having minimum programming skills.
iii. Next, we will use the R "brms" package (Bürkner, 2017), which enables models to be written in a standard, user-friendly formula notation. Nevertheless, it is less flexible than the RStan. As for the four-parameter beta regression model we can use the betaBayes package in R that has been developed by Zhou and Huang (2022).

8. Simulation Evaluation

i. Simulation evaluation of the developed model is based on the training data set.
ii. In the evaluation process we focus on examining two main aspects in the simulation study:

Accuracy and efficiency of the estimated parameters. are shown by the bias and relative bias that can be formulated as:

$$Bias(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}}), E(\hat{\boldsymbol{\theta}}) = \frac{\sum_{b=1}^{B}\hat{\boldsymbol{\theta}}}{B}, \text{ Relative Bias } (\hat{\boldsymbol{\theta}}) = \frac{Bias(\hat{\boldsymbol{\theta}})}{\boldsymbol{\theta}}$$

Computational time of R-Stan and R "brms". The computational time will be based on the average time needed to complete a parameter estimation process of 1000 time repetition.

iii. The Bias and relative bias will be calculated for all parameters in the four-parameter beta mean GLMM, which are a,b,φ, $\boldsymbol{\beta}$, $\boldsymbol{b_i}$.

iv. We will do the evaluation for each scenario and draw a conclusion.

9. Predictions

i. Predictions are conducted on the testing data set.
ii. Prediction will be made based on the developed model.
iii. Predictions will be made for all scenarios.

10. Prediction Evaluations

i. The accuracy of the predictions is shown by the Mean Square Error of Prediction (MSEP) of parameter estimates as follows.

$$MSEP(\hat{\boldsymbol{\theta}}) = \frac{1}{B}\sum_{b=1}^{B}(\boldsymbol{\theta} - E(\hat{\boldsymbol{\theta}}))^2$$

ii. The MSEP will be calculated for all scenarios.
iii. We will do an evaluation of the predictions accuracy for each scenario and draw a conclusion.

3.2 Simulation Results and Discussion

As mentioned before, we constructed a simulation study to analyze the parameter estimate's performance of our proposed Beta 4 Parameter Mean GLMM with Bayesian approach with four scenarios. We examine several things in the simulation study: (1) Bias, (2). Relative Bias, and (3). MSEP. Computationally wise, we considered two approaches in developing the model. First, we have used the Stan package in R that allows more flexibility thus we can estimate the minimum and maximum value (see 4P Beta GLMM (1) in Table 2). Second, we have used the brms package in R, that is more user friendly but does not have much flexibility. Thus, we cannot estimate the minimum and maximum value (see 4P Beta GLMM (2) in Table 2). Utilizing an Intel core i7 desktop computer and R Server to perform the simulation, CPU time of the brms package was much faster, it took around 1.4 – 2 minutes to complete a parameter estimation process. While for the stan program, CPU time is 3 to 4 times longer. Moving forward, we can enhance the processing time by defining the priors for the model in advance. In this study we have used the results of a Beta GLM parameter estimate for the priors in our model.

Table 2
Simulation Results for Four Parameter Beta Mean GLMM with Stan (4P Beta GLMM (1)) and for Four Parameter Beta Mean GLMM with brms (4P Beta GLMM (2))

Scenario	Parameter Estimates	4P Beta GLMM (1)				4P Beta GLMM (2)			
		Mean Estimate	Bias	Relative Bias	MSEP	Mean Estimate	Bias	Relative Bias	MSEP
Scenario 1	b_0	0.062	0.438	0.876	0.191	0.96	-0.46	-0.92	0.212
	b_1	0.501	-0.001	-0.002	0	0.512	-0.012	-0.024	0
	b_2	0.502	-0.002	-0.004	0	0.509	-0.009	-0.018	0
	b_3	0.501	-0.001	-0.002	0	0.5	0	0	0.25
	Random Effect	0.438	-0.438	-	0.198	0.514	-0.514	-	1.944
	Phi	13.572	-3.572	-0.357	15.551	17.042	-7.042	-0.704	
	a	0.001	-0.001	-	0				
	b	1.962	0.038	0.019	0.001				
Scenario 2	b_0	0.054	0.446	0.892	0.198	0.883	-0.383	-0.766	0.146
	b_1	0.501	-0.001	-0.002	0	0.506	-0.006	-0.012	0
	b_2	0.503	-0.003	-0.006	0	0.504	-0.004	-0.008	0
	b_3	0.501	-0.001	-0.003	0	0.5	0	0	0.25
	Random Effect	0.428	-0.428	-	0.188	0.5	-0.5	-	
	Phi	16.993	-6.993	-0.699	54.386	22.247	-12.247	-1.225	
	a	0	0	-	0				
	b	1.98	0.02	0.01	0				

The results of the estimation of fixed effect and random effect parameters of the four-parameter beta mean GLMM with stan and brms package can be seen in Table 2. Based on the table, for scenario 1 $\hat{\boldsymbol{\beta}} = (0.062,\ 0501,\ 0.502,\ 0.501)$. It is known that the initial value of the fixed effect parameter is $\boldsymbol{\beta}$ = (0.5, 0.5, 0.5, 0.5). It appears that $\hat{\boldsymbol{\beta}}$ for scenario 1 is almost unbiased, except for the intercept. This is the same case for scenario 2, where $\hat{\boldsymbol{\beta}} = (0.054, 0501, 0.503, 0.501)$ and $\boldsymbol{\beta}$ = (0.5, 0.5, 0.5, 0.5). Meanwhile, the mean estimate of the random value is like the biasness of the intercept. Next, the estimates of minimum $\hat{a}$ and maximum $\hat{b}$ value of the four-parameter beta distribution for both scenarios also have minimum bias. Nonetheless, the bias, relative bias, and MSEP for the estimates of precision parameter $\hat{\varphi}$ for Scenario 1 and Scenario 2 are the highest compared to other estimates. Even though, 4P Beta GLMM (2) has faster computational processing, nevertheless 4P Beta GLMM (1) in general outperforms 4P Beta GLMM (2) in terms of bias, relative bias, and MSEP for almost all parameters of the four-parameter beta mean GLMM in each scenario. The largest difference is apparent in the bias and relative bias of the precision parameter. As an example, we can see that the difference is almost double when the covariance matrix is high in Scenario 2. Not only does 4P Beta GLMM (1) tend to have better parameter estimates, as mentioned earlier on, the 4P Beta GLMM (1) also has an upper hand in its flexibility to estimate the minimum and maximum value. Thus, we can conclude from the simulation that 4P Beta GLMM (1) based on Stan package in R is more flexible and accurate, while using 4P Beta GLMM (2) based on brms package in R is faster.

4. Predicting Paddy Productivity with Satellite Data

4.1 Predicting paddy productivity process

In this study we have specifically applied the novel four parameter beta mean GLMM to predict the paddy productivity of an area. The process consists of five main steps, starting from theory development up to defining the AYI pure premium and VaR (Fig. 2). We will assess the performance of the paddy productivity predictions in Central Kalimantan (2020).

The theory development process of a four-parameter beta mean GLMM has been explained in detail in Section 2, while the simulation process and results have been explained in Section 3. Next, for the data Collection and data preparation Statistics Indonesia (SI) has collected 534 CCEs in Central Kalimantan at that period, which indicates the actual paddy productivity at each rice field. The CCEs are also followed by farmer interviews to gather information related to climate, agriculture conditions, and farming business. In total there were 37 items asked in the farmer survey. The Variable Selection Process was based on preliminary research done by Kusumaningrum et al. (2024). There were seven selected explanatory variables that will be used in the model that were considered significant. Additionally, there will be the current and a three-month period lag in the Sentinel 2A satellite imagery data for each plot coordinate surveyed. Researchers have determined that band 4 (red band), band 8 (infra-red band), and a computed Normalized Difference Vegetation Index (NDVI) will also be incorporated as the model's explanatory variables. These variables were also proven to have a significant effect on paddy productivity in Central Kalimantan (Kusumaningrum, 2024). In total there are fourteen explanatory variables used in the model, see Table 2.

Table 2
Variables Used in the Proposed four parameter beta Mean GLMM

Variable	Name	Unit measurement/category	Data source
Paddy Productivity	Y	Tons/Ha	Crop Cutting Survey (SI)
Pest Attacks This Year	X_1	1 = Heavy 2 = Medium 3 = Light 4 = Not Affected	Crop Cutting Farmer Survey (SI)
Pest Attacks Last Year	X_2	1 = Heavy 2 = Medium 3 = Light 4 = Not Affected	Crop Cutting Farmer Survey (SI)
Impact of Climate Change This Year	X_3	1 = Affected 2 = Not Affected	Crop Cutting Farmer Survey (SI)
Impact of Climate Change Last Year	X_4	1 = Affected 2 = Not Affected	Crop Cutting Farmer Survey (SI)
Water Sufficiency This Year	X_5	1 = Not Enough 2 = Sufficient 3 = More than Enough	Crop Cutting Farmer Survey (SI)
Water Sufficiency Last Year	X_6	1 = Not Enough 2 = Sufficient 3 = More than Enough	Crop Cutting Farmer Survey (SI)
How to Handle Pest	X_7	0 = No Actions 1 = Agronomist 2 = Mechanical 3 = Biological 4 = Chemical	Crop Cutting Farmer Survey (SI)
Band 4	X_8	Mm	Sentinel-2A Imagery
Band 8	X_9	Mm	Sentinel-2A Imagery
NDVI	X_{10}	Index	Sentinel-2A Imagery
Number of seeds used/area of similar plants	X_{11}	Plants/Ha	Crop Cutting Farmer Survey (SI)
KCL	X_{12}	Kg/Ha	Crop Cutting Farmer Survey (SI)
Solid Organic Fertilizer /Compost	X_{13}	Kg/Ha	Crop Cutting Farmer Survey (SI)
Liquid Organic Fertilizer	X_{14}	Kg/Ha	Crop Cutting Farmer Survey (SI)

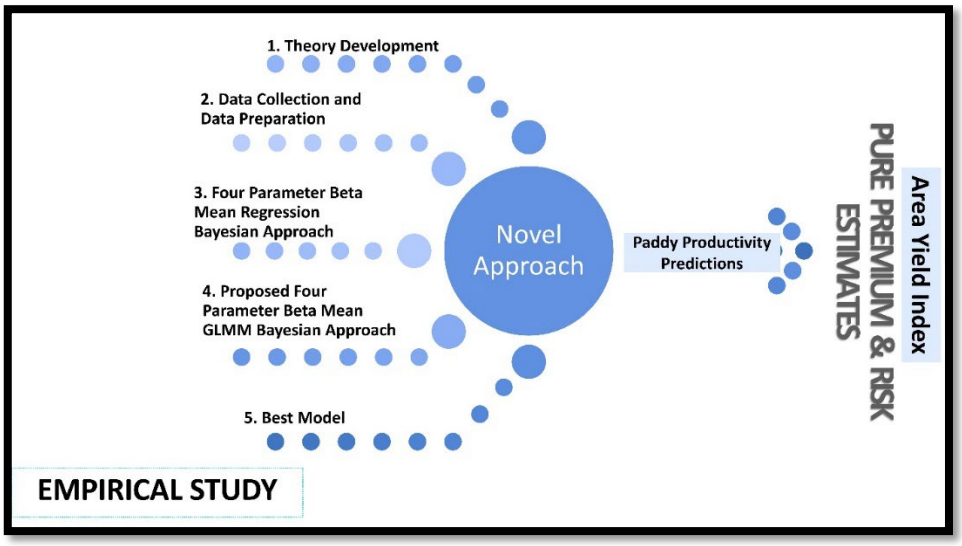


Fig. 2. Empirical Study Process Flowchart

Such as in the simulation process, we will split the data into training data (80%) and testing data (20%). The selected variables shown in Table 2 for the training data set will be used in the Four Parameter Beta Mean Regression shown in Eq. (2). We will also apply this data set to the proposed Four Parameter Beta Mean GLMM with the model shown in Eq. (4). Both models will apply the Bayesian approach in the process of estimating the parameters. To compare the goodness of fit of the two different models, we will use Watanabe- Akaike information criterion (WAIC) that can be calculated as

$$WAIC = -2\sum_{i=1}^{n}\log\left\{\frac{1}{L}\sum_{i=1}^{L}\log L_i\left(D_i \mid \Omega^l\right)\right\} + 2p_w, \text{ with } p_w = \sum_{i=1}^{n}\left[\frac{1}{L-1}\sum_{i=1}^{L}\left\{\log L_i(\Omega^l) - \frac{1}{L}\sum_{k=1}^{L}\log L_i(D_i \mid \Omega^k)\right\}^2\right]$$

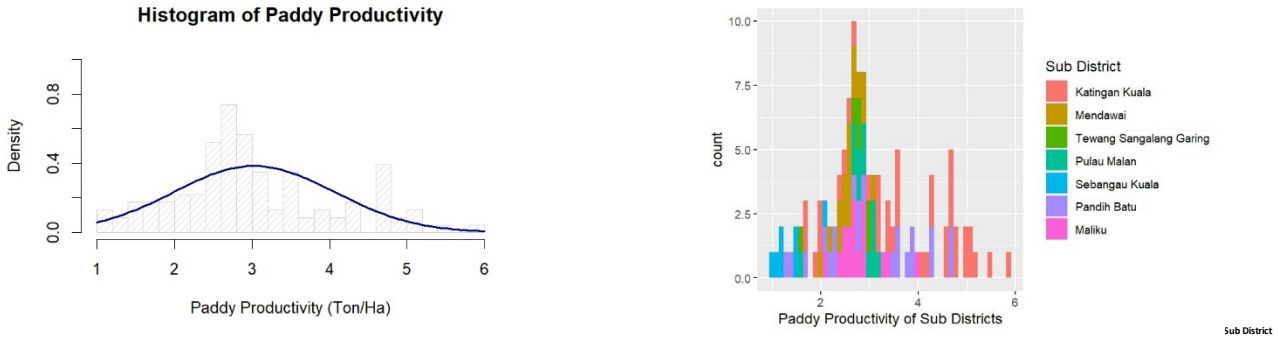
WAIC has gained popularity in recent years due to its stability compared to DIC (Watanabe, 2010). A smaller value of WAIC indicates a better fit of the model. After we find the best fit model, it hence becomes efficient to obtain specified predictions $\hat{y}_i$ based on the testing data set. The predicted are the by-products of our MCMC sampling approach. The predictions, will then be evaluated based on RMSE value, which is the standard deviation of the prediction errors formulated as

$$RMSE = srql\left[\frac{\sum_{l=1}^{n}(\hat{y} - y_i)}{n}\right].$$

Here $\hat{y}$ denotes the fitted/predicted value, while y_i represents the actual value, and n is the total number of observations or data points in the testing data set. To assess the results of parameter estimates of the selected model, we can evaluate the convergence of the posterior distribution to ensure that the parameter estimates of the four-parameter beta mean GLMM are accurate. Diagnostic checks that help assess for convergence include trace plots, the R ("Rhat") statistic, and Effective Sample Size (ESS) that was suggested in Reis et al. (2023). A Trace plot will show whether the MCMC chain converges or not. Meanwhile consistency of model parameter estimates is proven if the Rhat value is lower than 1. Meanwhile the ESS value should be at least 100 (approximately) to show reliability of the parameter estimates (Vehtari et al., 2021; Gelman, 2014).

4.2 Paddy productivity conditions in Central Kalimantan

In Central Kalimantan (2020), the minimum and maximum values of paddy productivity are bounded and vary slightly with the seasons. The productivity of paddy varies from 0.691 to 5.890 tons per hectare for each planting season in 2020. According to the distribution fit test, a four-parameter beta distribution has shown to have the best fit indicated by the lowest AIC and AICc value when compared to the Normal and Laplace distributions. Based on the parameter estimates of a four-parameter beta distribution, the variance of paddy productivity is 1.070 tons per hectare and the average paddy productivity is estimated to be 3.03. We also know that the shape parameters α =3.067 and β=6.091. Thus, it is indicated that the distribution is skewed to the right, because β is relatively higher than α which can also be seen in Fig. 3. Having an indication that the distribution is slightly skewed to the right means that the distribution is asymmetric, and the average value will be more than the median. Therefore, in designing AYI, we need to keep in mind that most of the farmer's actual paddy productivity tends to be lower than the average. Insurers need to carefully assess the risk associated with low paddy productivity, as these events may have significant financial implications for the insurance product.



(a) Overall Paddy Productivity (b) Paddy Productivity per Sub District

Fig. 3. Overall and Per Sub District Paddy Productivity in Central Kalimantan

Regarding the predictions of the proposed novel model to be applied to calculate the premium of AYI, in this study we have decided to specifically use the first planting season (January – March of 2020) data in Central Kalimantan for the case study. An AYI crop insurance is commonly provided for a specific planting season and not for the whole year because premiums will vary depending on the predicted productivity of each planting season. In this season there were two districts surveyed, which are Katingan and Pulang Pisau. Meanwhile, there are seven sub districts that were selected by SI as the targeted survey area from the two districts. Fig. 3(b) shows that paddy productivity among sub districts differs. Relevantly, district and sub district random effect variables will be applied in the models.

4.3 Applying the four-parameter beta GLMM and Evaluations

We will apply the novel four parameter beta mean GLMM not only to predict paddy productivity but also indicate significant factors that affect paddy productivity. We have known that predicting paddy productivity is very crucial in the calculations of the premium. Meanwhile, identifying significant variables affecting paddy productivity are important to enhance precise risk assessment associated with paddy productivity fluctuations in a specific area and develop a comprehensive risk management strategy to mitigate various risks.

In general, this process starts with data preparation and exploration to ensure data quality and relevancy. Next, benchmark models are applied to establish a baseline for comparison within the study. The researchers then apply the novel four parameter beta mean GLMM and conduct evaluations comparing the fit of the proposed and benchmark models. In specifying the model, we begin by using (1) farmer survey data, (2) satellite data along with the 3 months lag of the satellite data, and (3) farmer survey data combined with all the satellite data with its 3 months lagged data. Thus, there will be three main conditions to choose from, allowing flexibility for policymakers and insures to select which condition is most suitable with the data provided. A full model of all the three conditions was done and variable selection was conducted subsequently. The nominated best models can be seen in Table 3 where assessments were made based on WAIC. Afterwards, we can identify key factors affecting paddy productivity and analyze the results of the best-chosen model. Lastly, researchers will predict paddy productivity. Further on, we will demonstrate how these predictions can be used to estimate the pure premium and VaR of the AYI crop insurance policy in Central Kalimantan.

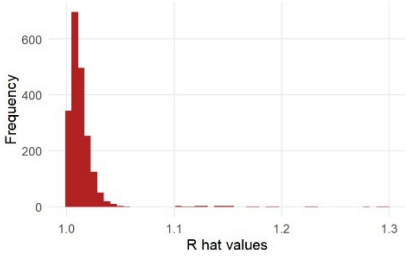
To apply the novel four parameter beta mean GLMM based on the Bayesian inference, we must first define prior distributions as follows (1). $a \sim U(-1e-15, 0.1790452)$ dan $b \sim U(1e-15,\ 1)$, (2) $\phi \sim \Gamma(0.01,0.01)$, (3) fixed effect ($\boldsymbol{\beta}$) was given by the parameter estimates of a four-parameter beta GLM, and (4) random effect $\boldsymbol{b_i}$~N($\mathbf{0}$, $\boldsymbol{\Sigma}$). The initial values given were based on the initial data explorations and benchmark modeling. Next, we will estimate the posterior through a Markov Chain Monte Carlo (MCMC) simulation with a Burn-in of 1000 replications, and the iterations of simulations MCMC is set to10.000. Aside from setting a prior distribution for the random effects (b_i), the settings were the similar for conducting the four-parameter beta mean regression models (Zhou & Huang, 2022).

Table 3 proves that the proposed four parameter beta mean GLMM has a much better model fit compared to the four-parameter beta mean regression model in this case study. The proposed model exceeds the benchmark in all three conditions indicating that the random effects improve the model accuracy. Hence, accounting for the variability of a district or sub district will also lead to a better understanding of the data conditions. Overall, the best fit model is the model where we both combine the farmer survey data and satellite Sentinel 2A data (Model 12). The random effect in the model was a sub district level, where the fixed effect variables included Pest Attacks Last Year, How to Handle Pest, Band 4, Band 8, NDVI, KCL, and Solid Organic Fertilizer /Compost. Complete results of parameter estimates, and the 95% credible intervals of this model can be seen in Table 4.

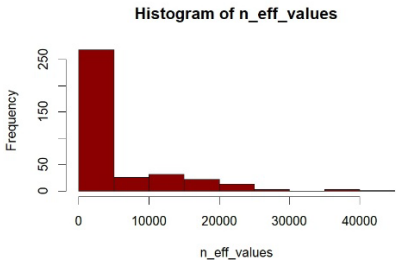
Table 3
Model Evaluation for Beta 4 Parameter Mean Regression and Beta 4 Parameter Mean GLMM

Model	Name of model	Variables	Fixed effect	Random Effect	WAIC
Four Parameter Beta Mean Regression (Zou, 2023)	Model 1	All Survey Data	X_1-X_{14}	-	-126.020
	Model 2	Selected Survey Data	X_2, X_7, X_{12}, X_{13}	-	-115.000
	Model 3	Satellite Data Lag 3 Month	X_8, X_9, X_{10} and Lags of X_8, X_9, X_{10} up to the past 3 months	-	-91.620
	Model 4	Selected Survey Data and Satellite Data 3 Months	X_2, X_7, X_{12}, X_{13}, X_8, X_9 X_{10} and Lags of X_8, X_9, X_{10} up to the past 3 months	-	-127.000
	Model 5	Selected Survey Data and Satellite Data Current Month	X_2, X_7, X_{12}, X_{13}, X_8, X_9, X_{10}	-	-124.000
Proposed Four Parameter Beta Mean GLMM	Model 6	Selected Survey Data	X_2, X_7, X_{12}, X_{13}	District Area	-181.666
	Model 7	Satellite Data Lag 3 months	X_8, X_9, X_{10} and Lags of X_8, X_9 X_{10} up to the past 3 months	District Area	-203.887
	Model 8	Selected Survey Data and Satellite Data 3 Months	X_2, X_7, X_{23}, X_{24}, X_8, X_9 X_{10} and Lags of X_8, X_9 X_{10} up to the past 3 month	District Area	-132.385
	Model 9	Selected Survey Data and Satellite Data Current Month	X_2, X_7, X_{12}, X_{13}, X_8, X_9, X_{10}	District Area	-188.180
	Model 10	Selected Survey Data	X_2, X_7, X_{12}, X_{13}	Sub District Area	-202.609
	Model 11	Satellite Data Lag 3 Months	X_8, X_9, X_{10} and Lags of X_8, X_9 X_{10} up to the past 3 months	Sub District Area	187.735
	Model 12	Selected Survey Data and Satellite Data Current Month	X_2, X_7, X_{12}, X_{13}, X_8, X_9, X_{10}	Sub District Area	-206.745

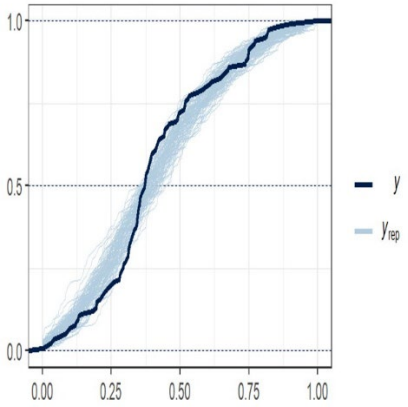
Evaluating the Bayesian performance based on diagnostic checks of the MCMC iterations is essential. Trace plots in Fig. 4 (c) show that all the parameters estimate posterior distribution convergence to a certain value. The average of R-hat values is 1.004 and from the histogram of R-hat shown in Fig. 4 (a), where all the values are concentrated at an R-hat value less than 1.01 ensuring that the parameter estimates are consistent. Meanwhile the mean ESS value (Fig. 4 (b)) is above 100, specifically the mean of ESS is 1124.0, suggesting that the parameter estimates are reliable. We have also checked the proposed four parameter beta mean GLMM ability to predict paddy productivity (Fig. 4 (c)) and results show a good fit indicated by a high correlation of 0.797 between the actual and predicted values. Based on these diagnostic checks of the MCMC iterations and the predictions, we can conclude that the performance of Bayesian methods in estimating the parameters of the proposed four parameter beta mean GLMM and predicting paddy productivity in Central Kalimantan was appropriate. Therefore, we will define the key factors affecting paddy productivity based on this model.



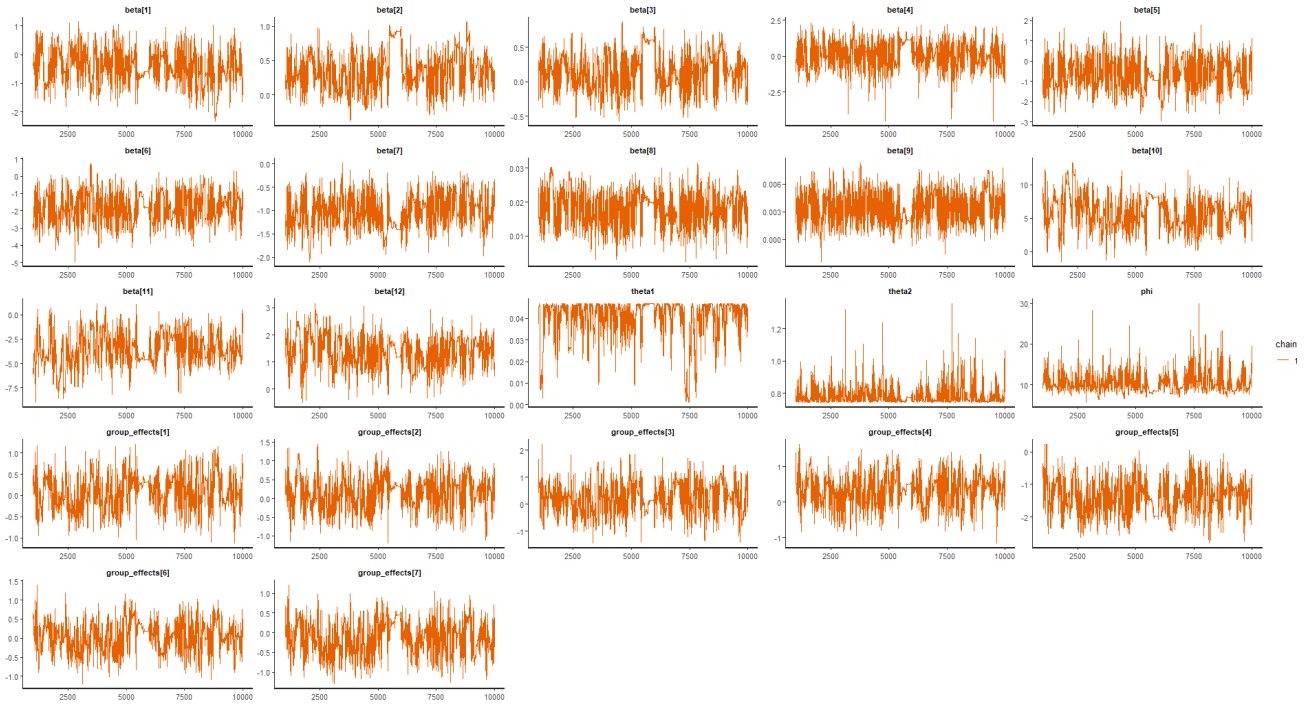
(a) R-hat Histogram



(b) Effective Sample Size (ESS)



(c) Plot of y_i and $\hat{y}_i$



(d) Trace plots of parameter estimates

Fig. 4. Evaluations on Bayesian approach on a four-parameter beta Mean GLMM

Summary statistics of Bayesian parameter estimates of Model 12 are shown in Table 4. First, we have stronger indications to support our belief that the random effect will significantly deviate from the average effect when the 95% credible interval does not overlap with zero, as indicated by Sebangau Kuala subdistrict. Confidence intervals for this subdistrict are less than zero, indicating low average paddy productivity and significant variations among the districts. Variables such as How to Handle Pest (X_7), KCL (X_{12}), Solid Organic Fertilizer /Compost (X_{13}), Band 4, Band 8, and NDVI have a significant effect on paddy productivity. Pest handling through agronomist and mechanical has a potential to decrease paddy productivity. Meanwhile, using the right dosage of KCL and Solid Organic Fertilizer /Compost are in line with paddy productivity. For the upcoming season, it is estimated that the minimum value of paddy productivity predictions of the next season in Central Kalimantan is 1.037 while the maximum is 2.

Table 4
The mean, standard deviation (Stdev), and the 95% credible interval estimate for best fit four parameter beta GLMM

Variables	Notations of the Parameter Estimate	Mean Estimate	Stdev of Estimate	The 95% Credible Lower Bound	The 95% Credible Upper Bound
(Intercept)	beta[1]	-0.424	0.523	-1.521	0.601
X_2 (Moderate)	beta[2]	0.353	0.249	-0.072	0.919
X_2 (Mild)	beta[3]	0.145	0.237	-0.303	0.61
X_2 (Not Affected)	beta[4]	0.061	0.822	-1.618	1.527
X_7 (Mechanical)	beta[5]	-0.584	0.681	-1.99	0.795
X_7 (Agronomist)	**beta[6]**	**-1.824**	**0.752**	**-3.418**	**-0.415**
X_7 (Mechanical)	**beta[7]**	**-1.029**	**0.316**	**-1.661**	**-0.443**
X_{23}	**beta[8]**	**0.018**	**0.004**	**0.01**	**0.027**
X_{24}	**beta[9]**	**0.004**	**0.001**	**0.001**	**0.006**
X_8	**beta[10]**	**5.946**	**2.212**	**1.677**	**10.765**
X_9	**beta[11]**	**-3.563**	**1.54**	**-7.131**	**-0.778**
X_{10}	**beta[12]**	**1.394**	**0.533**	**0.372**	**2.519**
a	**theta1**	**0.04**	**0.01**	**0.008**	**0.047**
b	**theta2**	**0.786**	**0.054**	**0.745**	**0.955**
φ	**Phi**	**10.414**	**2.11**	**7.709**	**15.626**
Katingan Kuala	group_effects[1]	0.387	0.358	-0.606	0.783
Mendawai	group_effects[2]	0.429	0.377	-0.582	0.907
Tewang Sangalang Garing	group_effects[3]	0.641	0.501	-0.771	1.181
Pulau Malan	group_effects[4]	-0.155	0.377	-0.448	1.094
Sebangau Kuala	**group_effects[5]**	**-0.8**	**0.457**	**-2.337**	**-0.516**
Pandih Batu	group_effects[6]	0.325	0.355	-0.636	0.707
Maliku	group_effects[7]	0.115	0.384	-0.852	0.557

In this model it is also shown that satellite data doesn't only have a significant effect but also enhances accuracy. Current and previous satellite data have potential for predicting paddy productivity. The results are in line with previous research showing that NDVI, Band 4, and Band 8 values correlate with paddy productivity. There is a positive shift during growth and higher values indicating greater productivity near harvest in main

growing seasons. Meanwhile weaker patterns and lower values are apparent during dry season conditions (Kusumaningrum, 2024). Model 7 and Model 11, also have low WAIC and therefore show potential for further development. The challenge will be to define the length of lag. As for the chosen Model 12, results show that the RMSE of predicted values is 0.020, which has the lowest value among all proposed four beta parameters mean GLMM. Thus, we will use the predicted results to calculate the premium and VaR of AYI.

4.4 Estimating Pure Premium and Value at Risk (VaR) of the Area Yield Index (AYI) Crop Insurance Policy

AYI is a productivity-based index agricultural insurance policy intended to cover systemic risks like floods, droughts, pests, and diseases, which could affect most farmers simultaneously in the same areas. The benchmark yield (yc) is established at area level based on the average of 7-years historical yield data and actual average yield (yi) is calculated through a field survey such as the Crop Cutting Experiments (CCEs). When the benchmark yield yc is exceeded by the actual average yield (yi) an indemnity is paid to all insured farmers in the same area, regardless of their individual losses, when yi is less than the. data collected. Therefore, the indemnity function per hectare (Ha) can be written as $indemnity = \max(y_c - \overline{y}_i)SI$. Where, Sum Insured (SI) is set at 6,00,000 IDR per Ha, which is based on the current compensation given by the existing crop insurance product. To estimate the indemnity distribution of the alternative AYI policy Monte Carlo Simulations was applied and will lead to the claim amount of each area. Henceforward, we can give further an example on how to evaluate the AYI policy based on the premium (π (X)) and Premium Value at Risk (VaR).

Table 5
AYI Premium and VaR based on four parameter beta GLMM Estimates (in Rupiah)

District	Sub District	Average Total Claim in IDR	Average Productivity (Tones/Ha)	Pure Premium in IDR π (X)	VaR
Katingan	a). All Sub District	40.726.190	3.09	407.262	901,445
	b). Katingan Kuala	-	3.60		
	c). Other Sub Districts	81.452.379	2.59		
Pulang Pisau	a). All Sub District	30.127.475	2.58	301.275	382,115
	b). Pandih Batu	30.926.505	2.57		
	c). Other Sub Districts	29.328.444	2.59		

District level was used in the simulation for an area level in AYI., In each district we simulated the premium and VaR. For simplicity in the calculation process of the average total claim we have selected the sub district that has the highest average productivity based on historical data, which are Katingan Kuala and Pandih Batu. It is expected that these sub districts will have low or even zero average total claims because these sub districts tend to have high average paddy productivity. Meanwhile, the all-sub district average summarizes the total claim in the selected district and in other districts.

Results show that the characteristics of Katingan district differs from Pulang Pisau district. For the next upcoming season, it is expected that Pulang Pisau district tends to have lower paddy productivity variability compared to Katingan district, which is shown by the average total claim for Pandih Batu and other sub districts. In contrast to Katingan district, there are no claims, and all the claims come from other sub districts. The average total claim from other sub districts in Kitingan has been the highest among all and reaches up to almost 81.5 million rupiahs. The high average total claim will lead to higher VaR, which will potentially lead to higher losses that will be faced by the insurer in the upcoming planting seasons. As for the pure premium, the difference between the two districts is not apparent. The estimated AYI pure premium for both sub districts are above the current multiperil crop insurance policy, which is 180.000 IDR. Nevertheless, farmers will benefit from this policy because of the simpler claim process, protection against area wide losses, and lower basis risks.

Commonly, AYI premiums are calculated only based on average historical productivity values. The use of average historical data has a lack of flexibility because it does not account for variations in climate, pest outbreaks, or other dynamic factors that impact crop yield. Historical averages also do not incorporate new information or data that may become available after the initial calculation was made. In contrast, developing a model such as the four parameter beta mean GLMM model for predicting paddy productivity allows for a more nuanced and data- driven approach, considering various factors that influence crop yield. It can offer better adaptability, accuracy, and responsiveness to changes in agricultural conditions. However, it's important to note that the effectiveness of this model depends on the quality and relevance of the data used for training and validations.

Furthermore, discussions with experts in the field of agriculture conducted throughout this research, emphasized the importance of recognizing recent advancements in multi-task AI, which have demonstrated notable success in classification and detection tasks through optimized architectures, robust feature recognition, and enhanced performance. These innovations provide a solid foundation for real-time crop monitoring, particularly for tomatoes, and hold significant promise for advancing precision agriculture (Islam & Hattou, 2024). Crop insurance plays a critical role in mitigating risks, providing initial capital, and strengthening food security resilience. Addressing challenges in this domain necessitates integrated solutions. Future efforts should prioritize the incorporation of data collected across the entire crop lifecycle, beginning at the planting stage rather than solely before harvest. Additionally, leveraging modern technologies such as remote sensing and combining diverse data sources—including satellite imagery, drone-based data, and farmer surveys—will enhance predictive accuracy. This comprehensive approach has the potential to generate more reliable predictions, thereby supporting improved decision-making for stakeholders engaged in agricultural insurance and food security initiatives.

5. Conclusion

In this study we focus on the development of four parameter beta mean GLMM using a Bayesian approach. This model was an extension of the four-parameter beta regression model developed by Zhou (2021), which incorporates random effects into the model. Using a simulation study, we showed that the developed four parameter beta mean GLMM are considered relatively unbiased for the estimated parameters $\theta = (\hat{a}, \hat{b}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b_t}})$. However, estimates of the precision parameter ($\hat{\varphi}$) was still biased. Further research is suggested to increase the precision and accuracy of this parameter. The simulation results also showed that the development of this model using Stan package in R software does not have faster computation time compared to the brms package in R software, but it is more flexible and accurate.

The empirical case study shows that the paddy productivity in Central Kalimantan has a four-parameter beta distribution that is skewed to the right. Hence, it makes more sense to develop the paddy productivity prediction model based on this distribution. Especially when the predictions are used as the basis to calculate the premium and Value at Risk (VaR) of the alternative Area Yield Index (AYI) crop insurance policy. Insurers need to carefully assess the risk associated with low paddy productivity as it has a significant financial implication for the insurance product. We have also shown that the proposed four parameter beta mean GLMM predictions are more accurate than four parameter beta regression models. The key factors affecting paddy productivity based on this model include how to manage pests (X_7), the right dosage of KCL (X_{12}) and solid organic fertilizer /compost (X_{13}), Band 4, Band 8, and NDVI. This model also demonstrates that satellite data not only has a significant effect but also improves accuracy. Satellite data from the past and present can be used to predict paddy productivity. As a result, we will calculate the premium and VaR of AYI based on the prediction results.

To enhance the effectiveness of Area-Yield Insurance (AYI) in Central Kalimantan, it is suggested that district-level data is used as the area level. Next, it is also suggested to transition from traditional premium calculations based on historical paddy productivity averages to a more advanced, data-driven approach. The proposed four-parameter Beta Mean GLMM model enables the incorporation of dynamic factors such as climate variability, pest outbreaks, and other conditions that influence paddy productivity. This model delivers greater precision, responsiveness, and flexibility compared to traditional methods. While the estimated pure AYI premium (IDR 300–410 thousand) is expected to be higher than the current multiperil crop insurance policy, it provides significant benefits to farmers, including easier claims settlement, protection against regional losses, and lower basis risk. Furthermore, leveraging advancements in multi-task AI and integrating modern technologies such as remote sensing, satellite imagery, drone-based data, and farmer surveys can further enhance the accuracy of predictions. Collecting data throughout the entire crop lifecycle, starting from planting, is crucial for building a robust and adaptive insurance framework. These innovations will not only improve premium calculations but also support better decision-making for stakeholders, strengthening agricultural resilience and food security.

Credit Authorship Contribution Statement

Dian Kusumaningrum: Writing – original draft, Investigation, Formal analysis, Validation, Software. **Hari Wijayanto**: Supervision, Methodology, Conceptualization. **Khairil Anwar Notodiputro**: Supervision, Methodology, Conceptualization. **Muhlis Adrainsyah**: Data – collection and Data Pre-processing. **Islam MD Parvez**: Supervision, Review, and Editing. **Anang Kurnia**: Management, Software, and Review. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

I also declare that there is no conflict of interest. Please address all correspondence concerning this manuscript to me by e-mail (dian.kusumaningrum@prasetiyamulya.ac.id).

Data Availability Statement

Data will be available upon completion of investigation and publication of the report upon request to the corresponding author. Any request will be reviewed and approved by the Statistics Indonesia (SI) based on the absence of competing interest. Once approved, data can be transferred after signing of a data access agreement and confidentiality agreement.

Acknowledgment

First, authors would like to express our gratitude for the research funding provided by the Indonesian Ministry of Education, Culture, Research, and Technology Contract Number 22038/IT3.D10/PT.01.03/P/B/2024. Second, authors would also like to acknowledge the role of Southeast Asian Regional Center for Graduate Study and Research in Agriculture (SEARCA) for providing a full PhD scholarship and research support alongside Six University Initiative Japan Indonesia (SUIJI) Japan for the opportunity to do a Collaborative Research Program. Next, authors are also very grateful to Statistics Indonesia for sharing the CCEs and farmer survey data in Central Kalimantan. Last, the authors would like to express their deep appreciation to Prasetiya Mulya niversity Indonesia, IPB University Indonesia, and The United Graduate School of Agricultural Sciences (UGAS) Ehime University Japan for the provision and assistance during this research.

Funding

This research receives a Fundamental Research Grant scheme (Contract Number 22038/IT3.D10/PT.01.03/P/B/2024) provided by Indonesian Ministry of Education, Culture, Research, and Technology, ressearch scholar grant from Southeast Asian Regional Center for Graduate Study and Research in Agriculture (SEARCA) Ref No GBG20-3281, and Prasetiya Mulya University Publication Grant.

References

Bonat, W.H., Ribeiro Jr, P.J., & Zeviani, W.M. (2015). Likelihood Analysis for a Class of Beta Mixed Models. *Journal of Applied Statistics, 42*(2), 252-266. DOI: 10.1080/02664763.2014.947248

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*, 1-28. 10.18637/jss.v080.i01.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of statistical software*, *34*(1), 1-24.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, *31*(7), 799-815.

Gray, L. A., & Alava, M. H. (2018). A command for fitting mixture regression models for bounded dependent variables using the beta distribution. *The Stata Journal*, *18*(1), 51-75.

Grün, B., Kosmidis, I., & Zeileis, A. (2012). Extended beta regression in R: Shaken, stirred, mixed, and partitioned. *Journal of Statistical Software*, *48*, 1-25.

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli, 7*(2), 223–242.

Islam, M. P., & Hatou, K. (2024). Artificial intelligence assisted tomato plant monitoring system–An experimental approach based on universal multi-branch general-purpose convolutional neural network. *Computers and Electronics in Agriculture*, *224*, 109201. https://doi.org/10.1016/j.compag.2024.109201.

Johnston, C. K., Waterhouse, T., Wiens, M., Mondick, J., French, J., & Gillespie, W. R. (2024). Bayesian estimation in NONMEM. *CPT: Pharmacometrics & Systems Pharmacology*, *13*(2), 192-207. doi:10.1002/psp4.13088

Kusumaningrum, D., Wijayanto, H., Kurnia, A., Notodiputro, K. A., Ardiansyah, M., & Islam, M.P. (2024a). Beta Four Parameter Mixed Models Based on Survey and Satellite Data for Paddy Productivity Predictions. *Smart Agricultural Technology, 9*. https://doi.org/10.1016/j.atech.2024.100525

Kusumaningrum, D., Wijayanto, H., Notodiputro, K. A., Ardiansyah, M., & Kurnia, A. (2024b). Beta four parameter GLMM approach to evaluate paddy productivity. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1356, No. 1, p. 012094). IOP Publishing. DOI 10.1088/1755-1315/1356/1/012094

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912-928. DOI: 10.1111/oik.05985

Liu, J. S., Wong, W. H., & Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 27-40.

Lynch, S. (2007). *Introduction to applied Bayesian statistics and estimation for social scientists*. New York, NY: Springer

McCullagh, P. (2019). *Generalized linear models*. Routledge.

Reis, D. J., Kaizer, A. M., Kinney, A. R., Bahraini, N. H., Holliday, R., Forster, J. E., & Brenner, L. A. (2023). A practical guide to random-effects Bayesian meta-analyses with application to the psychological trauma and suicide literature. *Psychological trauma: theory, research, practice, and policy*, *15*(1), 121. doi: 10.1037/tra0001316.

Roberts, G. O., & Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *59*(2), 291-317.

Sanyu Consultants Inc. and Sompo Risk Management Inc. (2023). Area Yield Index Insurance Product Design. National Development Planning Agency (BAPPENAS) and Japan International Cooperation Agency (JICA)

Sargent, D. J., Hodges, J. S., & Carlin, B. P. (2000). Structured markov chain monte carlo. *Journal of Computational and Graphical Statistics*, *9*(2), 217-234.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods*, *11*(1), 54.

Stan Development Team. (2020_. RStan: the R interface to Stan. R package version 2.21.2. Stan Development Team. 2021. Stan User's Guide. Version 2.26. http://mc-stan.org/.

Vehtari, A., & Gelman, A. (2014). *WAIC and cross-validation in Stan*.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. C. (2021). Rank-normalization, folding, and localization: An improved Rˆ for assessing convergence of MCMC (with discussion). *Bayesian analysis*, *16*(2), 667-718. arXiv preprint arXiv:1903.08008

Zainul, A., Hanani, N., Kustiono, D., Syafrial, S., & Asmara, R. (2021). Forecasting the basic conditions of Indonesia's rice economy 2019-2045. *Agricultural Socio-Economics Journal*, *21*(2), 111-120.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*. North-Holland/Elsevier, pp. 233–243.

Zhou, H., & Huang, X. (2022). Bayesian beta regression for bounded responses with unknown supports. *Computational Statistics & Data Analysis*, *167*, 107345. https://doi.org/10.1016/j.csda.2021.107345

© 2026 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).