

Inflated Modified Kumaraswamy Regression Model for Invasive Plants Detection in NDVI Imagery

Aline Armanini Stefanan, Bruna G. Palm, *Member, IEEE*, Fábio M. Bayer, *Member, IEEE*, Simon Hallösta, and Mats I. Pettersson, *Senior Member, IEEE*

Abstract—This study proposes the inflated modified Kumaraswamy (iMK) distribution, a flexible probability model defined on the unit interval $[0, 1]$. It captures asymmetric behaviors while accommodating inflation at zero, one, or both boundaries, as commonly observed in normalized difference vegetation index (NDVI) data. Based on the iMK distribution, we develop a new regression model (iMKreg) suitable for double-bounded responses. From this model, we derive a detection tool for invasive plant species, particularly applicable to NDVI imagery. Model performance was evaluated using synthetic NDVI data, with further assessment of predictive accuracy and detection efficacy conducted on real-world measured NDVI image. The application to detecting black-grass (*Alopecurus myosuroides*) in wheat crops in southern Sweden shows that the iMKreg model outperforms both standard Gaussian-based linear regression and existing inflated Kumaraswamy regression models.

Index Terms—Ground type detection, Inflated modified Kumaraswamy distribution, Regression model

I. INTRODUCTION

Normalized difference vegetation index (NDVI) is a useful tool to monitor vegetation health by measuring the difference between near-infrared (NIR) and red light reflectance (RED), normalized by their sum, given by $NDVI = \frac{(NIR-RED)}{(NIR+RED)}$. With values inherently limited to the range of $[-1, 1]$, NDVI frequently presents a skew toward higher values, indicating vigorous, healthy, and dense vegetation, which absorbs most of the red light and reflects a large portion of the NIR light. In contrast, unhealthy, sparse, and stressed vegetation absorbs more NIR light and reflects more red light, leading to lower NDVI values, with a zero value meaning no vegetation [1].

Recent global warming, particularly extreme heat events, significantly exacerbates the threat of invasive plants in crop plantations by favoring their spread and competitive advantage over native crops, while posing new challenges for their

detection and management. In this context, NDVI images have become a key component in agricultural applications, particularly for the early detection and monitoring of invasive plants [2]. These plants outcompete crops for critical resources such as light, water, and soil nutrients, and their early detection protects crop yields, enabling precision herbicide use, which prevents excessive pollution of soil and water while slowing the development of chemical resistance in weeds.

From a statistical point of view, NDVI values do not follow the symmetric bell-shaped curve of a Gaussian distribution, thereby requiring more flexible distributions to model their skewness and bounded nature. For asymmetric bounded data defined on $[-1, 1]$, or, without loss of generality, on the standard unit interval $[0, 1]$, the modified Kumaraswamy (MK) [3] distribution arises as a suitable and flexible statistical model. It is a two-parameter probability model, which accommodates density shapes that the classical beta and Kumaraswamy distributions cannot capture, such as increasing–decreasing–increasing patterns. However, such a model does not handle the presence of zeros and/or ones, which occurs with non-null probability in standardized NDVI data. By mixing continuous and discrete distributions, it is possible to accommodate the presence of 0 and 1 in unit-bounded data. This approach, commonly referred to as inflating the distribution, allows the model to capture both the continuous behavior within the open interval and the point-mass probabilities at the boundaries [4], [5]. Furthermore, image modeling often relies on the assumption of a constant parameter distribution across all evaluated pixels [6], [7]. An alternative approach is to adopt a regression model framework for pixel-by-pixel prediction. In this context, regression models have been used for vegetation parameter estimation [8], change detection [9], image filtering [10], and radar data interpretation [11], [12].

In this letter, our goal is three-fold. First, we propose the inflated modified Kumaraswamy (iMK) distribution, suitable for modeling $[0, 1)$, $(0, 1]$, and $[0, 1]$ data. Second, we propose a regression model for the iMK distribution. For the proposed model, we introduce parameter estimation, large data record inference, and goodness-of-fit measures. Third, we propose a detector based on the asymptotic properties of the estimators, focusing on detecting invasive plants in NDVI imagery. In particular, the study targets the identification of black-grass (*Alopecurus myosuroides*), a resistant invasive species native to Eurasia that is currently affecting farmland in southern Sweden and is known to significantly reduce crop yields, in some cases by up to half.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, grant number 88881.690009/2022-01, and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - FAPERGS, the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, grant numbers 308578/2023-6 and 200831/2024-0, Swedish-Brazilian Research and Innovation Centre (CISB), and Saab AB.

Aline Armanini Stefanan is with the Industrial Engineering Department, Federal University of Santa Maria (UFSM), Brazil. (e-mail: aline.armanini@acad.ufsm.br).

Bruna G. Palm, Simon Hallösta, and Mats I. Pettersson are with the Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, Sweden (e-mail: bruna.palm,mats.pettersson,simon.hallosta@bth.se).

Fábio M. Bayer (M'20) is with the Department of Statistics and LACESM, Universidade Federal de Santa Maria, Brazil (e-mail: bayer@ufsm.br).

Manuscript received xx; revised xx.

II. THE PROPOSED MODELS

This section introduces the median-based iMK distribution, and proposes regression models suitable for modeling the median of MK and iMK distributed data.

A. The iMK Distribution

Based on a mixed continuous–discrete distribution, we propose the iMK distribution, which combines the unit MK distribution with a Bernoulli component that assigns probability mass at zero, one, or both boundaries. The probability density function (pdf) of a random variable Y , which follows an iMK distribution is given by:

$$\bar{f}(y; \boldsymbol{\vartheta}, \alpha) = \begin{cases} \lambda(1-p), & \text{if } y = 0, \\ (1-\lambda)f(y; \mu, \alpha), & \text{if } 0 < y < 1, \\ \lambda p, & \text{if } y = 1, \end{cases} \quad (1)$$

being $0 \leq y \leq 1$ the observed signal value, $\boldsymbol{\vartheta} = (\mu, \lambda, p)^\top$, $0 < \mu < 1$ the median parameter of the MK distribution, $0 \leq \lambda \leq 1$ is the mixture parameter, representing the probability that Y is inflated, $0 \leq p \leq 1$ is the probability that a Bernoulli-distributed random variable equals one, and $f(y; \mu, \alpha)$ is the median-based MK pdf given by:

$$f(y; \mu, \alpha) = \frac{\alpha \log(0.5) e^{\alpha - \frac{\alpha}{y}} \left(1 - e^{\alpha - \frac{\alpha}{y}}\right)^{\frac{\log(0.5)}{\mu^*} - 1}}{\mu^* y^2},$$

where $\mu^* = \log\left(1 - e^{-\frac{\alpha}{\mu}}\right)$ and $\alpha > 0$ is a shape parameter. The iMK cumulative distribution function (cdf) is given by:

$$\bar{F}(y; \boldsymbol{\vartheta}, \alpha) = \lambda(1-p) + (\lambda p) \mathbb{I}_{\{1\}}(y) + (1-\lambda)F(y; \mu, \alpha),$$

where $F(y; \mu, \alpha) = 1 - \left(1 - e^{\alpha - \frac{\alpha}{y}}\right)^{\frac{\log(0.5)}{\mu^*}}$ is the median-based MK cdf and $\mathbb{I}_{\{1\}}(y)$ is an indicator function that equals one if $y = 1$, and zero otherwise. The inverse of $\bar{F}(y; \boldsymbol{\vartheta}, \alpha)$ is the quantile function used to generate random observations of the iMK distribution, and is given by:

$$\bar{F}^{-1}(u) = \begin{cases} 0, & \text{if } u \leq \lambda(1-p), \\ F^{-1}\left(\frac{u - \lambda(1-p)}{1-\lambda}\right), & \text{if } \lambda(1-p) < u < (1-\lambda p), \\ 1, & \text{if } u \geq (1-\lambda p). \end{cases}$$

Notice that, when $p = 0$, we obtain the iMK distribution inflated at zero, and when $p = 1$, we obtain the iMK distribution inflated at one, as particular cases. Additionally, when $p = \lambda = 0$, Y follows the MK law [3].

B. Regression Models

Consider a set of N independent random samples, denoted as $Y[1], \dots, Y[N]$, where each $Y[n]$, $1 \leq n \leq N$, assumes values $y[n]$ and follows the iMK distribution with pdf in (1). The proposed iMK regression (iMKreg) model is established by incorporating three regression structures for the mixture parameters and the median of $Y[n]$. The formulation is as follows:

$$g_1(\mu[n]) = \sum_{j=1}^k x_j[n] \beta_j, \quad g_2(\lambda[n]) = \sum_{i=1}^c z_i[n] \nu_i,$$

$$g_3(p[n]) = \sum_{b=1}^m a_b[n] \rho_b,$$

where $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ are strictly monotone and twice differentiable link functions, such that $g_d : (0, 1) \rightarrow \mathbb{R}$ for $d = 1, 2, 3$. For example, g_1 , g_2 , and g_3 may be chosen as logit or probit link functions. Additionally, \mathbf{X} , \mathbf{Z} , and \mathbf{A} are the $(N \times k)$, $(N \times c)$, and $(N \times m)$ dimensional full column rank matrices where $x_j[n]$ denotes the (n, j) th element of \mathbf{X} , $z_i[n]$ denotes the (n, i) th element of \mathbf{Z} , and $a_b[n]$ denotes the (n, b) th element of \mathbf{A} ; $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^\top$, $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_c)^\top$, and $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_m)^\top$ are the k , c , and m dimensional vectors of unknown parameters associated to \mathbf{X} , \mathbf{Z} , and \mathbf{A} , respectively. Usually, $x_1[n] = z_1[n] = a_1[n] = 1 \forall n$, so that β_1 , ν_1 , and ρ_1 are the intercept parameters for the iMK regression. When the response only takes values in the open interval $(0, 1)$, the MK regression model arises given by just the $\mu[n]$ submodel.

The maximum likelihood estimators (MLEs) of the iMK regression model parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\nu}^\top, \boldsymbol{\rho}^\top, \alpha)^\top$, can be obtained by maximizing the logarithm of the likelihood function, i.e., where the score function is equal to zero. The score vector for $\boldsymbol{\theta}$, $\mathbf{U}(\boldsymbol{\theta})$, is presented in the Appendix A. The MLEs of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\nu}}^\top, \hat{\boldsymbol{\rho}}^\top, \hat{\alpha})^\top$ is given by solving $\mathbf{U}(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{0}$ is the $(k + c + m + 1)$ -dimensional vector or zeros. As the solution has no closed-form, the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) method with analytic first derivatives was adopted. Computational implementations of iMKreg model are available [13].

C. Large Data Record Inference

Under mild regularity conditions, the MLEs are consistent and asymptotically normal as $N \rightarrow \infty$. Hence, for large samples,

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{K}(\boldsymbol{\theta})^{-1}), \quad (2)$$

approximately, where $\mathbf{K}(\boldsymbol{\theta})$ is the observed information matrix and \mathcal{N} denotes the multivariate normal distribution. The observed information matrix is presented in the Appendix B. This asymptotic result is the basis for constructing confidence intervals and performing hypothesis tests (detection) [14], [15].

The $(1 - \kappa)100\%$ confidence interval (CI) for the i -th parameter θ_i is given by:

$$\left[\hat{\theta}_i - z_{1-\kappa/2} \sqrt{[\mathbf{K}^{-1}(\hat{\boldsymbol{\theta}})]_{ii}}, \hat{\theta}_i + z_{1-\kappa/2} \sqrt{[\mathbf{K}^{-1}(\hat{\boldsymbol{\theta}})]_{ii}} \right],$$

where $\hat{\theta}_i$ is the i -th component of $\boldsymbol{\theta}$, $[\mathbf{K}^{-1}(\hat{\boldsymbol{\theta}})]_{ii}$ is the i -th diagonal element of the inverse of \mathbf{K} evaluated at $\hat{\boldsymbol{\theta}}$, and $z_{(\cdot)}$ denotes the quantile of the standard normal distribution.

The Wald test is a common tool for making inferences about regression parameters. Let the parameter vector $\boldsymbol{\theta}$ be partitioned into: (i) a vector of parameters of interest $\boldsymbol{\theta}_I$ of dimension ω , and (ii) a vector of nuisance parameters $\boldsymbol{\theta}_M$ of dimension $k + c + m + 1 - \omega$. The null and alternative hypotheses are:

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{\theta}_I = \boldsymbol{\theta}_{I0}, \\ \mathcal{H}_1 : \boldsymbol{\theta}_I \neq \boldsymbol{\theta}_{I0}, \end{cases}$$

where θ_{I0} is a fixed column vector of length ω . The Wald statistic is given by [14]:

$$T_W = (\hat{\theta}_{I1} - \theta_{I0})^\top \left(\left[\mathbf{K}^{-1}(\hat{\theta}_1) \right]_{\theta_I \theta_I} \right)^{-1} (\hat{\theta}_{I1} - \theta_{I0}),$$

where $\hat{\theta}_1 = (\hat{\theta}_{I1}^\top, \hat{\theta}_{M1}^\top)^\top$ are the unrestricted MLEs, $\left[\mathbf{K}^{-1}(\hat{\theta}) \right]_{\theta_I \theta_I}$ denotes the block of the inverse information matrix corresponding to θ_I . From the consistency of the MLEs and from Equation (2), T_W is asymptotically distributed as χ_ω^2 .

Hypothesis tests can be applied for ground-type detection, as discussed in [16]. In this study, we assume that the parameter $\mu[n]$ of an iMK-distributed signal varies according to the ground type. Under this assumption, the Wald test can be used for ground-type classification by comparing the computed T_W statistic with a threshold value γ , determined from the χ_ω^2 distribution and a predefined probability of false alarm. This approach adapts the methodology previously proposed for Rayleigh-based regression models [11] to the iMK regression framework, leveraging the median $\mu[n]$ as a robust central measure for double-bounded NDVI data.

D. Diagnostic Analysis and Model Selection

Randomized quantile residuals [17] of the fitted model are calculated for diagnostic checks and goodness-of-fit tests. The residuals are given by:

$$\hat{e}^q[n] = \begin{cases} F_{N(0,1)}^{-1}(u^0[n]), & \text{if } y[n] = 0, \\ F_{N(0,1)}^{-1}(\hat{F}(y[n]; \hat{\lambda}[n], \hat{p}[n], \hat{\mu}[n], \hat{\alpha})), & \text{if } 0 < y[n] < 1, \\ F_{N(0,1)}^{-1}(u^1[n]), & \text{if } y[n] = 1, \end{cases}$$

where $F_{N(0,1)}^{-1}$ is the standard normal quantile function and $u^0[n]$ and $u^1[n]$ are uniform random variables in the intervals $(0, \hat{\lambda}[n](1 - \hat{p}[n]))$ and $(1 - \hat{\lambda}[n]\hat{p}[n], 1)$, respectively. Additionally, $\hat{\lambda}[n]$, $\hat{p}[n]$, $\hat{\mu}[n]$, $\hat{\alpha}$ are the MLEs of $\lambda[n]$, $p[n]$, $\mu[n]$, and α , respectively.

To check the assumptions of non-autocorrelated and Gaussian distributed residuals, the Durbin-Watson test and the Shapiro-Wilk test, respectively, are suggested. Those assumptions lead to reliable statistical inferences.

For model selection purposes, a set of candidate models with different covariates can be fitted. Among those that pass the diagnostic analysis, the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) can be used for comparison. The selected model is the one that minimizes the AIC or BIC. The Vuong test [18] is also suggested for model selection. AIC and BIC estimate the relative information loss when approximating the true model from a candidate set, while the Vuong test statistically compares competing models based on their relative closeness to the truth.

III. NUMERICAL EXPERIMENTS

In this section, we evaluate the MLEs of the iMK regression model parameters and the performance of the proposed ground-type detector. Two numerical experiments are conducted: (i) a simulation study to assess the finite sample properties of the estimators and (ii) an application to real NDVI data. Without loss of generality, we adopt a linear transformation of NDVI values from the $[-1, 1]$ range to the unit interval $[0, 1]$ in our analysis.

TABLE I

SIMULATION RESULTS ON POINT AND INTERVAL ESTIMATION OF THE iMKREG PARAMETER ESTIMATORS FOR A UNIT VARIABLE IN THE INTERVAL $[0, 1]$, CONSIDERING A SIGNIFICANCE LEVEL OF 0.05. PARAMETER VALUES ARE PRESENTED INSIDE THE PARENTHESES

Measures	$\hat{\beta}_1$ (0.6812)	$\hat{\beta}_2$ (0.1722)	$\hat{\beta}_3$ (-0.7170)	$\hat{\alpha}$ (11.3206)	$\hat{\nu}_1$ (-2.6390)
$N = 50$					
RB	0.0053	0.5985	0.0391	5.8021	4.0383
MSE	0.0025	0.0049	0.0029	2.6060	0.3442
CR	0.9350	0.9332	0.9356	0.9402	0.9576
$N = 100$					
RB	0.0064	0.4648	0.0207	2.8181	3.5039
MSE	0.0012	0.0024	0.0013	1.0800	0.2162
CR	0.9378	0.9420	0.9382	0.9440	0.9688
$N = 300$					
RB	0.0291	0.0298	0.0391	0.8485	0.8290
MSE	0.0004	0.0008	0.0004	0.3089	0.0572
CR	0.9492	0.9512	0.9498	0.9482	0.9538

A. Synthetic Data Modeling

Monte Carlo simulations were performed to evaluate the finite sample performance of the MLEs on a synthetic unit variable in the interval $[0, 1]$. Synthetic NDVI data were generated under the scenario $\beta_1 = 0.6812$, $\beta_2 = 0.1722$, $\beta_3 = -0.7170$, $\alpha = 11.3206$, and $\nu_1 = -2.6390$. The scenario and parameter values are based on the model fitted in the empirical application presented in Section III-B. For this purpose, two dummy covariates were considered to distinguish different ground-type, $x_2 = 1$ for $\frac{2N}{3} + 1 \leq n \leq N$ and zero otherwise, and $x_3 = 1$ for $1 \leq n \leq N - \frac{2N}{3}$ and zero otherwise. The logit link function was employed for the $\mu[n]$ and $\lambda[n]$ submodels. We ran 5000 replications of unit variables generated via the inversion method for three different sample sizes: $N \in \{50, 100, 300\}$. In order to numerically evaluate the point estimators, we computed the percentage relative bias (RB), and the mean square error (MSE). To evaluate the interval estimators, we computed the coverage rate (CR) of the CI with a significance level of 5%.

Table I presents the simulation results. In general, the MLEs presented small values of RB and MSE. As expected, increasing the signal length (N), the RB and the MSE present lower values, which match the consistency of the MLEs. The CR is close to the nominal level of 95% for all evaluated signal lengths. It is noted that $\hat{\alpha}$ and $\hat{\nu}_1$ exhibit a higher RB and MSE than the others, while $\hat{\beta}_1$ presents the lowest RB.

B. Analysis with Measured NDVI Data

Table II summarizes the measured NDVI dataset (acquisition and preprocessing). The acquisition objective was to support detection of black-grass (*Alopecurus myosuroides*) in agricultural fields. Black-grass and wheat are morphologically similar, which complicates visual identification in RGB imagery. NDVI was computed as $\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$ after multispectral band co-registration [19].

The employed image is displayed in Fig. 1. In this image, pixels with zero values are due to image framing issues on the right and bottom edges and correspond to 0.0330% of the image pixels. The sampled NDVI values show medians

TABLE II
NDVI DATA ACQUISITION AND PREPROCESSING SUMMARY.

Item	Value
Platform / sensor	DJI Phantom 4 Multispectral (FC6360)
Date / coordinates	2023-03-27; 56.09811°N, 12.96359°E
Altitude / view	~1 m AGL; near-nadir
Bands (nm)	450±16 / 560±16 / 650±16 / 730±16 / 840±26
Image / sampling	16-bit TIFF, 1600×1300; ≈0.52 mm/pixel (center)
Radiometric / atmospheric	None (no reflectance calibration; no atmospheric correction)
NDVI scaling	Linear transform from [-1,1] to [0,1]

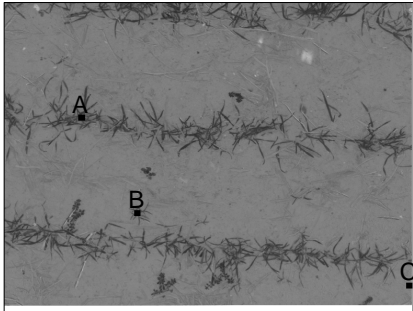


Fig. 1. NDVI image with selected areas, where A is the plant, B is the pest, and C is the bare soil.

of 0.7196 for plants, 0.6373 for invasive plants, and 0.4941 for bare soil. The close proximity between the plant and invasive plant medians compared to the bare soil highlights the challenge in distinguishing these classes.

To evaluate the proposed ground-type detection tool, ten NDVI pixels were sampled from each of the selected (26×26)-pixel regions representing plant (A), invasive plant (B), and bare soil (C), using a fixed random seed of 3821. The bare soil sample contained two zero-valued pixels. In this formulation, the vectorized NDVI pixel values of the three sampled regions constitute the response variable $y[n]$, while $x_2[n] = 1$ and $x_3[n] = 1$ indicate plant and bare soil pixels, respectively, and are set to zero otherwise. Given that the presence of zero-valued pixels is independent of the ground type, the inflation parameter λ was estimated considering only the intercept. The model section follows the methodology suggested in Section II-D. The estimated parameters of the fitted iMKreg model and the results of the diagnostic tests are presented in Table III. None of the residual-based tests reject the null hypothesis, indicating that the iMKreg model provides an adequate fit for the NDVI image pixels.

The ground-type discrimination is carried out by testing the null hypothesis $\mathcal{H}_0 : \beta_j = 0$, for $j = 2, 3$. A ground type is considered distinct when the null hypothesis is rejected. In this context, $\beta_j \neq 0$ indicates that the median pixel values of the evaluated regions exhibit significantly different behavior. For comparative purposes, a Gaussian-based regression model, as in [20] for NDVI modeling, and an inflated Kumaraswamy regression (iKreg) model were also fitted, and the same

TABLE III
FITTED iMKREG FOR THE NDVI IMAGE PIXELS. MLE COEFFICIENTS, CONFIDENCE INTERVALS, STATISTICS, AND P-VALUE OF THE WALD TEST, AND THE STATISTICS AND P-VALUES OF DURBIN-WATSON AND SHAPIRO-WILK TESTS

Parameter	MLE	Lower bound	Upper bound	Wald's Statistic	p-value
β_1	0.6812	0.5553	0.8071	112.4434	< 0.0001
β_2	0.1722	0.0018	0.3426	3.9240	0.0476
β_3	-0.7170	-0.8509	-0.5832	110.2145	< 0.0001
α	11.3206	7.8642	14.7770	41.2091	< 0.0001
ν_1	-2.6390	-4.0735	-1.2045	13.0007	0.0003
Durbin-Watson test statistic = 2.5748 (p-value = 0.1479); Shapiro-Wilk test statistic = 0.9730 (p-value = 0.6235).					

TABLE IV
A SIGNIFICANT TEST, AIC, BIC, AND MdRAE MEASURE FOR iMKREG MODEL AND COMPETITORS iKREG AND GAUSSIAN-BASED REGRESSION MODELS FIT OF THE NDVI IMAGE PIXELS DATA. P-VALUES IN BOLD INDICATE THAT THE NULL HYPOTHESIS IS REJECTED

Model	Parameter	MLE	p-value	AIC	BIC	MdRAE
iMKreg	β_1	0.6812	< 0.0001	-76.6715	-69.6656	0.5481
	β_2	0.1722	0.0476			
	β_3	-0.7170	< 0.0001			
iKreg	β_1	0.6978	< 0.0001	-69.4742	-59.6658	0.5693
	β_2	0.1664	0.0515			
	β_3	-0.7454	< 0.0001			
Gaussian-based regression	β_1	0.6349	< 0.0001	-32.7197	-27.1149	0.6682
	β_2	0.0737	0.2135			
	β_3	-0.2372	0.0003			

detection procedure was applied.

The p-values of the Wald test for the β_2 and β_3 parameters, shown in Table III, indicate that the iMKreg model successfully distinguishes plant and bare soil type areas from pest-affected areas at the 5% significance level; the plant area has a larger median NDVI pixel value than the pest area; and bare soil area has a smaller median NDVI pixel value than the pest area. The plant, pest, and bare soil detection results of the competing regression models are presented in Table IV. It can be seen that iKreg and Gaussian-based regression models fail to detect the plant type area change from the pest type area at the 5% significance level. Additionally, the iMKreg model presents a lower AIC, BIC, and median relative absolute error (MdRAE) than the competitor models. The Vuong test also evidences the superiority of iMKreg over the Gaussian-based regression (statistic = 4.3679, p-value < 0.0001) and the iKreg (statistic = 2.1645, p-value = 0.0152) models.

IV. CONCLUSION

We introduced the inflated modified Kumaraswamy (iMK) distribution, which accommodates asymmetry and has double-bounded support, and proposed an iMK-based regression model. In summary, the iMKreg model proved effective in analyzing double-bounded data with inflation, showing reliable estimation and superior performance in detecting vegetation-pest transitions in NDVI images when compared to benchmark models.

APPENDIX A SCORE VECTOR

In matrix form, the score vector for ϑ can be written as $U(\vartheta_d) = \mathbf{R}_d^T \cdot \mathbf{T}_d \cdot \mathbf{v}_d$, where \mathbf{R}_d are the \mathbf{X} , \mathbf{Z} , and \mathbf{A} matrices

for $d = 1, 2, 3$, respectively. The diagonal matrices are $\mathbf{T}_d = \text{diag} \left\{ \frac{1}{g_d'(\vartheta_d[1])}, \dots, \frac{1}{g_d'(\vartheta_d[N])} \right\}$. The working vectors are $\mathbf{v}_1 = (c_\mu[1] \mathbb{I}_{\{(0,1)\}}(y[1]), \dots, c_\mu[N] \mathbb{I}_{\{(0,1)\}}(y[N]))^\top$, with $c_\mu[n] = \frac{\alpha e^\alpha (\mu^*[n] + y^\# [n])}{\mu^2[n] \left[e^{\frac{\alpha}{\mu^*[n]}} - e^\alpha \right]} (\mu^*[n])^2$,

$\log(0.5) y^*[n]$, and $y^*[n] = \log \left(1 - e^{\alpha - \frac{\alpha}{y^*[n]}} \right)$,

$\mathbf{v}_2 = \left(\frac{(1-\lambda[1]) \mathbb{I}_{\{(0,1)\}}(y[1]) - \lambda[1] \mathbb{I}_{\{(0,1)\}}(y[1])}{\lambda[1](1-\lambda[1])}, \dots, \frac{(1-\lambda[N]) \mathbb{I}_{\{(0,1)\}}(y[N]) - \lambda[N] \mathbb{I}_{\{(0,1)\}}(y[N])}{\lambda[N](1-\lambda[N])} \right)^\top$, and $\mathbf{v}_3 = \left(\frac{(1-p[1]) \mathbb{I}_{\{1\}}(y[1]) - p[1] \mathbb{I}_{\{0\}}(y[1])}{p[1](1-p[1])}, \dots, \frac{(1-p[N]) \mathbb{I}_{\{1\}}(y[N]) - p[N] \mathbb{I}_{\{0\}}(y[N])}{p[N](1-p[N])} \right)^\top$. The score component

for α is $\mathbf{U}(\alpha) = \sum_{n=1}^N c_\alpha[n] \mathbb{I}_{\{(0,1)\}}(y[n])$, where $c_\alpha[n] = 1 + \frac{1}{\alpha} - \frac{(y[n]-1) \left[\frac{\log(0.5)}{\mu^*[n]} - 1 \right] e^{\alpha - \frac{\alpha}{y^*[n]}}}{y^*[n] - y[n] e^{\alpha - \frac{\alpha}{y^*[n]}}} + \frac{\varphi y^\# [n]}{\mu^*[n]} + \varphi - \frac{1}{y^*[n]}$, and $\varphi = \frac{(\mu[n]-1) e^{\frac{\alpha}{\mu^*[n]}}}{\mu[n] \left[e^{\frac{\alpha}{\mu^*[n]}} - e^\alpha \right]} \mu^*[n]$.

APPENDIX B

OBSERVED INFORMATION MATRIX

The observed information matrix \mathbf{K} is expressed as:

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{bmatrix} K_{(\beta, \beta)} & K_{(\beta, \alpha)} & 0 & 0 \\ K_{(\alpha, \beta)} & K_{(\alpha, \alpha)} & 0 & 0 \\ 0 & 0 & K_{(\nu, \nu)} & 0 \\ 0 & 0 & 0 & K_{(\rho, \rho)} \end{bmatrix},$$

where $\mathbf{K}_{(\vartheta_d, \vartheta_d)} = \mathbf{R}_d^\top \cdot \mathbf{v}_d' \cdot (\mathbf{T}_d)^2 \cdot \mathbf{R}_d + \mathbf{R}_d^\top \cdot \mathbf{v}_d \cdot g_d'' \cdot \mathbf{R}_d$,

$\mathbf{v}_1' = (c_{\mu\mu}[1] \mathbb{I}_{\{(0,1)\}}(y[1]), \dots, c_{\mu\mu}[N] \mathbb{I}_{\{(0,1)\}}(y[N]))^\top$,

with $c_{\mu\mu}[n] = \frac{2\alpha^2 e^{2\alpha} y^\# [n]}{\mu[n]^4 (\mu^\Delta[n])^2 (\mu^*[n])^3} +$

$\frac{\alpha e^\alpha \left\{ \left[2\mu[n] e^\alpha + (\alpha - 2\mu[n]) e^{\frac{\alpha}{\mu^*[n]}} \right] y^\# [n] + \alpha e^\alpha \right\}}{\mu[n]^4 (\mu^\Delta[n])^2 (\mu^*[n])^2} +$

$\frac{\alpha e^\alpha \left[2\mu[n] e^\alpha + (\alpha - 2\mu[n]) e^{\frac{\alpha}{\mu^*[n]}} \right]}{\mu[n]^4 (\mu^\Delta[n])^2 \mu^*[n]}$, $\mu^\Delta[n] = e^\alpha - e^{\frac{\alpha}{\mu^*[n]}}$ and $y^\Delta[n] =$

$e^\alpha - e^{\frac{\alpha}{y^*[n]}}$, $\mathbf{v}_2' = \left(-\frac{(\lambda[1]-1)^2 \mathbb{I}_{\{(0,1)\}}(y[1]) + \lambda[1]^2 \mathbb{I}_{\{(0,1)\}}(y[1])}{\lambda[1]^2 (\lambda[1]-1)^2}, \dots, \right.$

$\left. -\frac{(\lambda[N]-1)^2 \mathbb{I}_{\{(0,1)\}}(y[N]) + \lambda[N]^2 \mathbb{I}_{\{(0,1)\}}(y[N])}{\lambda[N]^2 (\lambda[N]-1)^2} \right)^\top$, $\mathbf{v}_3' =$

$\left(-\frac{(p[1]-1)^2 \mathbb{I}_{\{1\}}(y[1]) + p[1]^2 \mathbb{I}_{\{0\}}(y[1])}{p[1]^2 (p[1]-1)^2}, \dots, \right.$

$\left. -\frac{(p[N]-1)^2 \mathbb{I}_{\{1\}}(y[N]) + p[N]^2 \mathbb{I}_{\{0\}}(y[N])}{p[N]^2 (p[N]-1)^2} \right)^\top$, $g_d'' = -\frac{g_d'(\vartheta_d)}{(g_d'(\vartheta_d))^3}$,

$\mathbf{K}_{(\beta, \alpha)} = \mathbf{K}_{(\alpha, \beta)}^\top = -\mathbf{X}^\top \frac{\partial}{\partial \alpha} (\mathbf{v}_1) \mathbf{T}_1 \mathbb{1}$, $\frac{\partial}{\partial \alpha} (\mathbf{v}_1) =$

$(c_{\alpha\mu}[1] \mathbb{I}_{\{(0,1)\}}(y[1]), \dots, c_{\alpha\mu}[N] \mathbb{I}_{\{(0,1)\}}(y[N]))^\top$,

with $c_{\alpha\mu}[n] = \frac{e^\alpha \log(0.5) [\mu[n] (-e^\alpha)^\alpha (y[n]-1) \mu^\Delta[n]}{\mu[n]^3 y[n] (\mu^\Delta[n])^2 y^\Delta[n] (\mu^*[n])^2} -$

$\frac{e^\alpha \left[\mu[n] e^\alpha - (\mu[n] \alpha + \mu[n] - \alpha) e^{\frac{\alpha}{\mu^*[n]}} \right]}{\mu[n]^3 (\mu^\Delta[n])^2 \mu^*[n]} + \frac{(\mu[n]-1) \alpha e^{2\alpha}}{\mu[n]^3 (\mu^\Delta[n])^2 (\mu^*[n])^2} -$

$\frac{e^\alpha y^\# [n] \left[\mu[n] e^\alpha - (\mu[n] \alpha + \mu[n] - \alpha) e^{\frac{\alpha}{\mu^*[n]}} \right]}{\mu[n]^3 (\mu^\Delta[n])^2 (\mu^*[n])^2} + \frac{2\alpha e^{2\alpha} (\mu[n]-1) y^\# [n]}{\mu[n]^3 (\mu^\Delta[n])^2 (\mu^*[n])^3}$,

$\mathbb{1}$ is the $N \times 1$ vector of ones, $\mathbf{K}_{(\alpha, \alpha)} = -\mathbf{U}'(\alpha) =$

$-\frac{2(\mu[n]-1) e^{2\alpha} (y[n]-1) \log(0.5)}{\mu[n] y[n] \mu^\Delta[n] \left(e^{\frac{\alpha}{\mu^*[n]}} - e^\alpha \right) (\mu^*[n])^2} + \frac{2(\mu[n]-1)^2 e^{2\alpha} y^\# [n]}{\mu^2[n] (\mu^\Delta[n])^2 (\mu^*[n])^3} +$

$$\frac{\varphi (\mu[n]-1) y^\# [n]}{\mu[n] \mu^*[n]} + \frac{(\mu[n]-1)^2 e^{2\alpha} y^\# [n]}{\mu^2[n] (\mu^\Delta[n])^2 (\mu^*[n])^2} + \frac{(\mu[n]-1)^2 e^{2\alpha}}{\mu^2[n] (\mu^\Delta[n])^2 (\mu^*[n])^2} + \frac{(\mu[n]-1)^2 e^{2\alpha}}{\mu^2[n] (\mu^\Delta[n])^2 \mu^*[n]} - \frac{(y[n]-1)^2 e^{\alpha - \frac{\alpha}{y^*[n]}} \left(\frac{\log(0.5)}{\mu^*[n]} - 1 \right)}{y^2[n] \left(1 - e^{\alpha - \frac{\alpha}{y^*[n]}} \right)} - \frac{(y[n]-1)^2 e^{2\alpha - 2 \frac{\alpha}{y^*[n]}} \left(\frac{\log(0.5)}{\mu^*[n]} - 1 \right)}{y^2[n] \left(e^{\alpha - \frac{\alpha}{y^*[n]}} - 1 \right)^2} - \frac{1}{\alpha^2} + \frac{\varphi (\mu[n]-1)}{\mu[n]}.$$

REFERENCES

- [1] NASA, *Normalized Difference Vegetation Index (NDVI)*, 2000. [Online]. Available: https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php
- [2] K. K. Singh, T. D. Surasinghe, and A. E. Frazier, "Systematic review and best practices for drone remote sensing of invasive plants," *Methods in Ecology and Evolution*, vol. 15, no. 6, pp. 998–1015, 2024.
- [3] M. Sagrillo, R. R. Guerra, and F. M. Bayer, "Modified Kumaraswamy distributions for double bounded hydro-environmental data," *Journal of Hydrology*, vol. 603, p. 127021, dec 2021.
- [4] R. Ospina and S. L. P. Ferrari, "Inflated beta distributions," *Statistical Papers*, vol. 51, no. 1, pp. 111–126, mar 2010.
- [5] F. Cribari-Neto and J. Santos, "Inflated Kumaraswamy distributions," *Anais da Academia Brasileira de Ciências*, vol. 91, no. 2, 2019.
- [6] M. Sagrillo, R. R. Guerra, F. M. Bayer, and R. Machado, "The LA distribution: An approximation of the GOA distribution for amplitude SAR image modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, p. 5205408, 2023.
- [7] A. Frery, H.-J. Muller, C. Yanasse, and S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 648–659, May 1997.
- [8] H. Wang and K. Ouchi, "Accuracy of the K-distribution regression model for forest biomass estimation by high-resolution polarimetric SAR: Comparison of model estimation and field data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1058–1064, 2008.
- [9] M. İlsever and C. Ünsalan, *Two-dimensional change detection methods: remote sensing applications*. Springer Science & Business Media, 2012.
- [10] M. Yahia, T.-A. Hamrouni, and R. Abdelfattah, "Infinite number of looks prediction in SAR filtering by linear regression," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2205–2209, 2017.
- [11] B. G. Palm, F. M. Bayer, R. J. Cintra, M. I. Pettersson, and R. Machado, "Rayleigh regression model for ground type detection in SAR imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1660–1664, 2019.
- [12] A. Nascimento, P. Almeida-Junior, J. Vasconcelos, and A. Borges-Junior, "K-Bessel regression model for speckled data," *Journal of Applied Statistics*, pp. 1–24, 2022.
- [13] A. A. Stefanan, B. G. Palm, and F. M. Bayer, "iMKreg model to fit double-bounded unitary data with inflation at zero, one, or both," 2025. [Online]. Available: <https://github.com/alinestefanan/iMKreg>
- [14] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*, A. V. Oppenheim, Ed. Prentice Hall, 1998, vol. II.
- [15] Y. Pawitan, *In all likelihood statistical modelling and inference using likelihood*. Oxford: Oxford University Press, 2001.
- [16] R. J. Cintra, A. C. Frery, and A. D. Nascimento, "Parametric and nonparametric tests for speckled imagery," *Pattern Analysis and Applications*, vol. 16, no. 2, pp. 141–161, 2013.
- [17] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, p. 236, sep 1996.
- [18] Q. H. Vuong, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica*, vol. 57, no. 2, pp. 307–333, 1989.
- [19] S. Hallösta, S. Javadi, M. Dahl, and M. I. Pettersson, "Multispectral image registration and sensor calibration for low-altitude agricultural drones," in *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 6209–6213.
- [20] D. Fan, "Research on the establishment of NDVI long-term data set based on a novel method," *Scientific Reports*, vol. 13, no. 1, p. 9838, 06 2023.