I.J. Intelligent Systems and Applications, **2022, 4, 52-65**
Published Online August 2022 in MECS (http://www.mecs-press.org/)
DOI:10.5815/ijisa.2022.04.05

Development of Crop-Weather Models Using Gaussian Process Regression for the Prediction of Paddy Yield in Sri Lanka

Piyal Ekanayake
Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, 60200, Sri Lanka
E-mail: piyalekanayake@wyb.ac.lk

Lasini Wickramasinghe
Department of Electronics, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, 60200, Sri Lanka
E-mail: lasini2013@gmail.com

Jeevani W. Jayasinghe
Department of Electronics, Faculty of Applied Sciences, Wayamba University of Sri Lanka, Kuliyapitiya, 60200, Sri Lanka
E-mail: jeevani@wyb.ac.lk

Received: 10 January 2022; Revised: 22 February 2022; Accepted: 08 March 2022; Published: 08 August 2022

Abstract: This research introduces machine learning models using the Gaussian Process Regression (GPR) depicting the association between paddy yield and weather in Sri Lanka. All major regions in the island with most contribution to the total paddy production were considered in this research. The climatic factors of rainfall, relative humidity, minimum temperature, maximum temperature, average wind speed, evaporation, and sunshine hours were considered as input (independent) variables, while the paddy yield was the output (dependent) variable. The collinearity within each pair of independent and dependent variables was determined using Spearman's and Pearson's correlation coefficients. Data sets corresponding to the two main annual paddy cultivation seasons since 2009 were trained in MATLAB to develop crop-weather models. The most appropriate Kernel function was chosen from among four types of Kernels viz. Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2 based on their degree of coherence in modeling. This approach exploits the full potential of GPR in developing highly accurate crop-weather models. The performance of the crop-weather models was measured by the Correlation Coefficient, Mean Absolute Percentage Error, Mean Squared Error, Root Mean Squared Error Ratio, Nash Number and the BIAS. All the GPR-based models proposed in this paper are highly accurate in terms of the aforementioned evaluation metrics. Accordingly, when the climatic data are known or projected, the paddy yield and thereby the harvest of Sri Lanka can be predicted precisely by using the proposed crop-weather models.

Index Terms: Gaussian Process Regression, Kernel Function, Machine Learning, Modeling, Yield Prediction.

1. Introduction

Ever since the humans engaged in agricultural and livestock farming for raising their living, the consumption and production of rice (Oryza sativa) have continued to increase worldwide along with the growth of global populace and the invent of better farming techniques. Today, rice is the widely consumed staple diet of most people living in South and Southeast Asia. Cultivation of paddy has created multiple opportunities in trade, employment, and manufacturing of agricultural by-products in countries whose economic activities are dominated by domestic agriculture products.

Due to high water consumption, the productivity of paddy is highly dependent on favorable climatic conditions viz. timely rainfall or irrigation, temperature, humidity etc. throughout the farming season. The yield of paddy is conventionally measured only after harvesting ends which information if available in advance would be very useful in ensuring food security and regulating stock. Introducing an accurate harvest prediction mechanism would be immensely beneficial to countries like Sri Lanka where large farming communities live in less income strata. Particularly, such predictions will be useful for the farmers to find better ways in managing their total production with minimum wastage

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

and earning a better income.

Even a reliable information network of the paddy yield forecast in geographically distinct areas could be developed if the relationship between yield and major seasonal climatic conditions affecting the yield is accurately modeled. The paddy yield prediction at country level where rice is the staple food is reported in literature; e.g. China [1], Bangladesh [2], Egypt [3], Korea [4], and South Korea [5]. In some countries, studies on paddy yield prediction were confined only to a specific region, e.g. Ebro Delta in Spain [6]. Moreover, prediction of paddy yield in some states in India like Maharashtra [7], Tamil Nadu [8], and Andhra Pradesh [9] was reported. In Sri Lanka, development of paddy yield prediction models was limited only to a few regions [10] or a geographically small region like North-Western Province [11], Nilwala river basin [12], and Kurunegala district [13] whereas no comprehensive single study has been undertaken covering the entire country. In order to fill this knowledge gap covering the entire island, this study encompassed all main paddy cultivation areas of the country.

Paddy yield forecasting models were augmented with the use of multiple machine learning [14], deep learning [1], data mining, and statistical techniques on climate data and also by using satellite [15] or aerial imagery [16]. Artificial Neural Networks (ANN), Support Vector Machines Regression (SVMR), and Multiple Linear Regression (MLR) have been well employed by researchers on studies conducted using climate data. In this research, the Gaussian Process Regression (GPR) was used to forecast the paddy yield in Sri Lanka, as it has reportedly performed well in developing crop yield prediction models compared to other statistical and machine learning techniques. Besides, it has not yet been used widely in this research field. For example, Wickramasinghe et al. attempted to present the association between paddy yield and several climate variables in a main paddy producing zone (North-western province) in Sri Lanka by applying commonly used machine learning techniques of ANN, SVMR, GPR, and regression-based statistical techniques of MLR, Power Regression, and Robust Regression. They concluded that GPR-based prediction model is the most precise of them [11]. In contrast, this work proffers most suited paddy yield prediction approaches for all major paddy producing regions in Sri Lanka by applying four Kernel functions in GPR modeling.

GPR has been applied to develop crop yield prediction models in some other countries as well. For example, development of paddy yield prediction models for India using GPR, Deep Neural Network Regression (DNNR), Linear Regression, and Lasso Regression is presented in [17]. The relationship between crop yield and elements influencing crop produce viz. cultivation area and monthly climate factors (temperature, humidity, rainfall, wind speed, UV index, sunshine hours, and pressure) was modeled in that study. Though they had used only the Radial Basis Function (RBF) and the Rational Quadratic Kernel, GPR and DNNR were reported to have performed well with R2 > 0.90. According to research on the paddy yield estimation in Nepal, GPR outperformed SVMR, Linear Regression, and Ridge Regression [18]. However, this study was limited to a single Kernel of Squared Exponential.

GPR has been proved effective in estimating not only paddy yield but also the yield of some other crops. Using eight machine learning algorithms and integrating climate, remote sensing, and soil data, Han et al. predicted the wheat yield at country level in China with a coefficient of determination higher than 0.75, and a yield error below 10%. They pointed out Random Forest, GPR, and Support Vector Machine as the most suitable algorithms for wheat forecast in winter [19]. Yet, only the Exponential Kernel function was used to develop GPR-based models in their study. In [20], a Gaussian Process was combined with a Convolutional Neural Network to introduce a country-level forecasting model for Soybean production in the U.S., but only with Squared Exponential Kernel.

The development of GPR-based crop-weather models covering all major paddy producing regions in Sri Lanka is presented in this paper. More importantly, four Kernel functions were used in modeling and the best Kernel has been selected to develop the prediction model. Section 2 of this paper describes the paddy and weather statistics, assessment of collinearity, GPR techniques, and the statistical measures used in the evaluation of crop-weather models. Section 3 presents the resultant collinearity and crop-weather models with a discussion on the GPR-based analyses in comparison to recent studies. The conclusive findings are recapitulated in Section 4.

2. Materials and Methods

2.1. Paddy Statistics

Paddy is cultivated as a wetland crop to different extents in all 25 administrative districts of Sri Lanka over a total estimated area of about 700,000 Hectares at present. There are two main cultivation seasons in the country based on mostly regular annual rainfall received from the monsoons viz. South-west monsoon and North-east monsoon. The yield and harvest data of paddy in Sri Lanka over 16 years from 2004 to 2019 were collected from the Department of Census and Statistics. They were analyzed to identify the regions that account for bulk of the paddy production in the country. The harvest data were available in 27 regions comprising of 25 districts and two special agricultural zones namely Mahaweli ‘H’ and Udawalawa, which are cultivated from a major irrigation project in Sri Lanka.

The extent of annual paddy production in major regions is illustrated in Figure 1 where only 15 regions are shown because paddy production in no other region exceeded 150,000 tons per annum since 2004. Further, paddy production data are depicted in 5-year intervals to make the chart clearer as they provide sufficient information to represent the overall situation during the 16-year period.

As per paddy statistics over the 16 years (2004-2019), the highest contribution comes from Ampara district, which amounts to 15% of the overall production of the country, while Polonnaruwa, Kurunegala, and Anuradhapura districts produced the next highest harvest (Table1). Further, the contribution from each of the five districts viz. Hambantota, Batticaloa, Trincomalee, Badulla, and Monaragala and from each special agricultural zone accounted for more than 3% of the overall paddy harvest of the island. As over 80% of the total paddy production is covered by the above 11 regions, the other regions were not considered in this research study.

Moreover, no rain-fed paddy cultivation was reported from the two special agricultural zones where water is supplied by a major irrigation system in Sri Lanka. As this research focused on modeling the relationship between climatic factors and paddy yield, those special agricultural zones were also excluded. The paddy cultivation areas considered in this study are highlighted in Figure 2a. The corresponding topographical information is presented in Figure 2b demarcating agro-climatic zones and elevation. Based on rainfall, four main agro-climatic regions viz. wet zone, dry zone, intermediate zone, and arid zone have been identified in Sri Lanka. Further, countrywide topography has been distinguished as central highlands (>1500m), the plains (300-1500m), and the coastline (<300m) based on the elevation from the mean sea level.

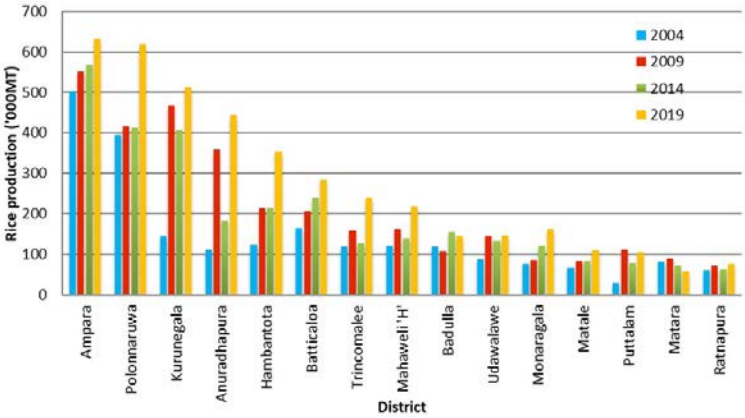
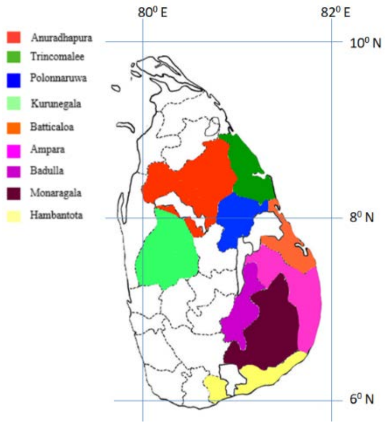


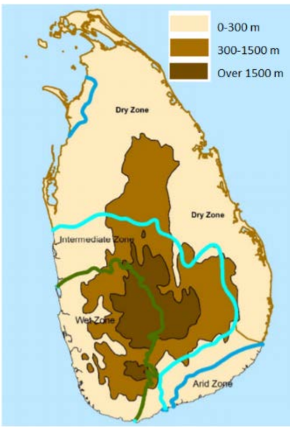
Fig.1. Paddy Statistics in Sri Lanka (2004-2019)

Table 1. Average Paddy Harvest in the Selected Regions (2004-2019) Considered for Modeling.

Administrative Unit	Input to the Overall Paddy Harvest (%)	Climatic Parameters Considered for Modeling
Ampara	15.57	Rainfall (RF), Relative Humidity (RH), Minimum Temperature (T_{min}), Maximum Temperature (T_{max}), Average Wind Speed (WS_a)
Batticaloa	6.17	
Trincomalee	4.45	
Polonnaruwa	10.64	RF, RH, T_{min}, T_{max}, WS_a, Evaporation, Sunshine hours
Kurunegala	10.56	
Anuradhapura	9.07	
Hambantota	6.23	
Badulla	3.62	
Monaragala	3.49	



(a)



(b)

Fig.2. (a) Paddy Cultivation Areas Considered for the Present Study (b) Topography and Main Agro-Climatic Zones in Sri Lanka [21]

Paddy yield data were considered as the dependent output variable in modeling and a summary of their analysis in the selected regions is presented in Table 2. It can be observed that both the mean and median paddy yields vary over a wide interval of 3.1 t/ha and 5.8 t/ha. The highest yield is reported from Hambantota while the lowest is from Batticaloa. However, as the paddy cultivation area in Hambantota is small, it occupies the fifth position when the regions are ranked according to the paddy harvest. The mean and median paddy yield values in most of the regions considered in the present study are approximately equal indicating a symmetrical Normal Distribution of yield data over the past 16 years.

Table 2. Average Paddy Yield in the Selected Regions (2004-2019)

Statistic	**Actual Yield (t/ha)**																	
	Ampara		Polonnaruwa		Kurunegala		Anuradhapura		Hambantota		Batticaloa		Trincomalee		Badulla		Monaragala	
	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n	S_s	S_n
Mean	4.8	4.6	4.9	5.0	3.7	4.2	4.6	4.7	5.2	5.8	4.2	3.1	4.8	4.1	4.6	4.4	4.0	4.2
Median	4.8	4.8	5.0	5.0	3.6	4.0	4.5	4.6	5.1	5.8	4.1	3.1	4.8	4.4	4.5	4.4	4.0	4.2
Range	1.1	2.0	1.3	1.6	0.6	1.5	1.4	1.7	1.6	1.4	1.8	1.9	0.5	2.4	0.9	1.5	1.4	0.9

Note: Ss is the Southwest-monsoon Season and Sn is the Northeast-monsoon Season

2.2. *Weather Statistics*

Eleven years of weather data from 2009 to 2019 were collected from the Department of Meteorology of Sri Lanka for developing the GPR-based prediction models. RF, RH, Tmin, Tmax, WSa, evaporation, and sunshine hours are the climatic parameters used as input variables (Table1). However, evaporation and sunshine data in Trincomalee, Batticaloa, and Ampara districts located in the Eastern province of the island were not available and therefore only five climatic factors could be used for developing models in those districts.

Paddy is cultivated during two main agricultural seasons using rainwater harvested from the monsoons. One of them is based on rain from the South-west monsoon, which spans from May to August. Monthly weather indicators without RF during this period were averaged and the RF was aggregated. The other season irrigated from the North-east monsoon spans from September to March of the ensuing year. Amount of precipitation during the Southwest-monsoon season varies within about 100 mm and over 3000 mm (Figure 3a) [22]. Maximum precipitation during this season occurs in the South-western part of the country (≈3260 mm) and it decreases rapidly towards the central highlands. South-western coastal region experiences over 1000 mm of RF during this period while the lowest figures are reported from Northern and South-eastern parts of the island. During the period of Northeast-monsoon, the highest RF occurs in the Northern and Eastern side of the central highlands (Figure 3b). Maximum RF during this season is reported in the central province (≈1280 mm), while the least occurs in the Western coastline (≈170 mm).

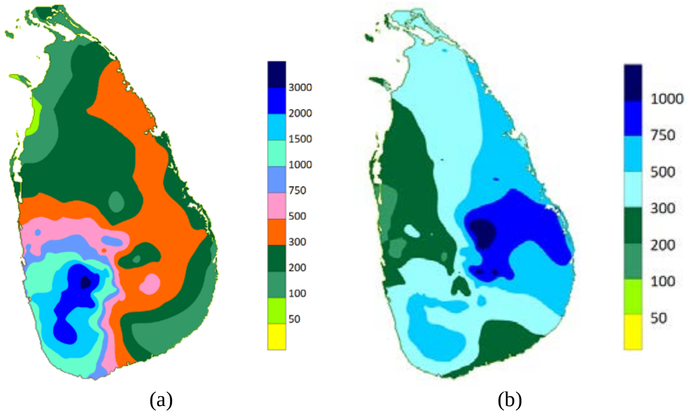


Fig.3. RF in Sri Lanka (a) Total RF from Southwest-monsoons (b) Total RF from Northeast-monsoons (Source: Department of Meteorology, Sri Lanka)

The variation of average yearly temperature in Sri Lanka is shown in Figure 4 [22]. Monthly mean temperature depends on the seasonal transition of the sun and the effect of precipitation. In lowlands, the annual mean temperature oscillates within 26.5-28.5 °C. Temperature falls when the altitude increases and reaches 16 °C in the central highlands. The coldest month is usually January and the warmest months are April and August. Monthly mean T_{max} in paddy growing regions considered in this research varies between 31.3 °C and 34.93 °C, while the monthly mean T_{min} varies between 22.0 °C and 26.2 °C.

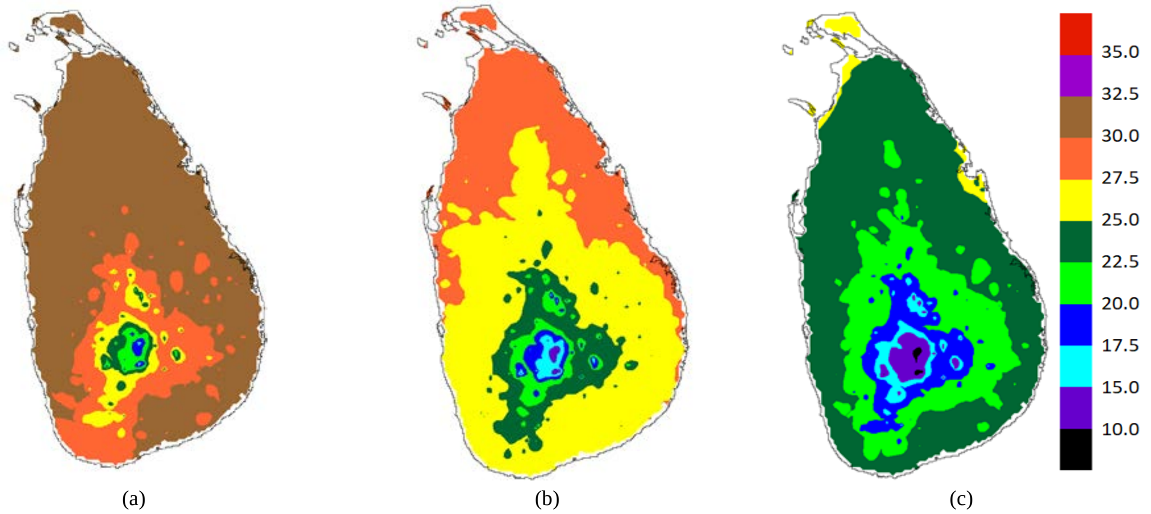


Fig.4. Average Annual Temperature in Sri Lanka (a) Tmax (b) Tmin (c) Mean temperature (Source: Department of Meteorology, Sri Lanka)

Humidity prevails throughout the country with an average RH over 65% [23]. In the coastal areas, it rises to 90% during the monsoon seasons and drops to about 60% in the eastern region from June to September. Times of sunrise and sunset in Sri Lanka do not differ much throughout the year due to its proximity to the equator. The number of sunshine hours per day varies between 5.8 and 8.5 hrs and the average seasonal wind speed remains within the range of 1.7-7.9 km/h.

2.3. Assessment of Collinearity

The pairwise correlations the paddy yield has with each and every weather parameter were calculated to identify whether the influence of that weather index on yield is significant or not. Further, the correlation between each couple of independent variables was inspected to identify the collinearity between them. Pearson's correlation coefficient ρ_p and Spearman's correlation coefficient were computed using R studio (version 1.3.1093), the programming language used in this study ρ_p, which is calculated using the following formula, computes both the magnitude and trend of a pairwise linear association between two variables [24].

$$\rho_p = \frac{\sum_{i=1}^{N}(\mathrm{A}_i - \overline{\mathrm{A}})(\mathrm{P}_i - \overline{\mathrm{P}})}{\sqrt{\sum_{i=1}^{N}(\mathrm{A}_i - \overline{\mathrm{A}})^2 \sum_{i=1}^{N}(\mathrm{P}_i - \overline{\mathrm{P}})^2}} \tag{1}$$

where, A_i and P_i denote the actual and forecast yields respectively while $\overline{A}$ and $\overline{P}$ are the corresponding mean values. As nonlinear associations between the paddy yield and climatic parameters were reported in similar research works [25], the pairwise ρ_s was also calculated from formula (2).

$$\rho_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(\mathrm{N}^2 - 1)} \tag{2}$$

where, d_i is the difference between the ranks of each pair of variabes and N is the number of data points. ρ_p and ρ_s can vary within the range from -1 to +1 [26]. A value of this coefficient in the range of 0 to1 signifies that the change in both variables is towards one direction while minus numbers denotes that variables change towards the reverse directions. Further, coefficients approaching |1| indicate powerful collinearity between the variables, while values hovering around 0 can be assumed to represent no clear relationship [27]. In this research study, strong correlations between two indices were defined if both correlation values were within the interval [±0.75, ±1.0]. Moderate correlations were determined if at least one of the values is within the interval [±0.50, ±0.74] and the other value falls in the upper range.

2.4. Gaussian Process Regression Method

Being nonparametric probabilistic models with a Kernel, Gaussian Process Regressions (GPRs) deal with a limited number of random variables having a multivariate distribution [28]. As all the linear combinations are assumed to be regularly distributed, Gaussian processes are governed by the concept of Normal distribution. The GPR carries out Gaussian processes for regression-related work and being a non-parametric machine learning technique, it is particularly useful in solving non-linear problems with small sample sizes [29]. It projects the response using latent variables. A Gaussian process applied on a training data set, which is defined as $\{(x_g, y_g): g = 1, 2, ..., n\}$ where, $x_g \in R^d$ and $y_g \in R$, is given as follows [30]:

$$y \sim GP(m(x), k(x_g, x_h)) \tag{3}$$

where, $m(x)$ is the mean function denoting the expectation and $k(x_g, y_g)$ is the Kernel function. The Kernel function is symmetrical and denotes the covariance. It is an aggregate of random variables x ∈ X having characteristics of a joint Gaussian distribution. The covariance matrix of a linear regression model in the form of $y = x^T \beta + \varepsilon$ is parameterized as follows [24]:

$$\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{pmatrix} \tag{4}$$

The function k is applied to every pair of locations in X. The covariance function has a set of Kernel parameters or hyperparameters and written as $k(x_g, x_h | \theta)$ indicating the dependence on θ. θ defines the statistical covariance between measurements made at two points, which are d units apart from each other.

In this research study, prediction models for paddy yield were developed from climatic data. The GPR was applied with four Kernel functions of Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2 and modeled in MATLAB (version 9.4.0.813654- R2018a). The 5-fold cross-validation was employed in training and validation of data to avoid overfitting.

Rational Quadratic GPR Kernel facilitates modeling of data varying at multiple scales and is suitable for multivariate statistical analysis performed on metric spaces [31]. The Rational Quadratic Kernel function is written as follows [32]

$$k\left(x_g, x_h | \theta\right) = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\sigma_1^2}\right) \tag{5}$$

where,

$$r = \sqrt{\left(x_g - x_h\right)^T \left(x_g - x_h\right)} \tag{6}$$

θ is the maximum a posteriori estimates, σ_f is the standard deviation, σ_1 is the lengthscale, T is the transpose, and α is the non-negative parameter of the covariance. The covariance depends only on distances between points, which are stationary. Covariance function is isotropic when the gap is Euclidean.

Squared Exponential GPR is a function space expression of a regression model with infinite number of RBFs. It is much similar to the Exponential GPR, faster in replacing inner products of RBFs with Kernels, and the Euclidean distance is squared. The Square Exponential Kernel function is expressed as

$$k\left(x_g, x_h | \theta\right) = \sigma_f^2 \exp\left(-\frac{r^2}{2\sigma_1^2}\right) \tag{7}$$

Matern 5/2 Kernel takes spectral densities of the stationary Kernel and performs Fourier transformation of the RBF Kernel. As sample functions are differentiable $|u-1|$ times, the hyperparameter u can control the degree of smoothness. Its Kernel function is illustrated as follows.

$$k\left(x_g,x_h|\theta\right)=\sigma_f^2\left(1+\frac{\sqrt{3}r}{\sigma_1}\right)\exp\left(-\frac{\sqrt{3}r}{\sigma_1}\right) \tag{8}$$

Exponential GPR replaces inner products of RBFs with Kernels and is identical to the Squared Exponential GPR except that the Euclidean distance of the former is not squared. It handles smooth functions well with minor errors though not good at handling discontinuities. The Kernel function of the median exponential GPR is given as

$$k\left(x_g,x_h|\theta\right)=\sigma_f^2\exp\left(-\frac{\sqrt{3}r}{\sigma_1}\right) \tag{9}$$

2.5. Evaluation of the Crop-Weather Models

The dexterity of GPR-based crop-weather models was evaluated with the undermentioned statistical parameters [31,33]:

BIAS:

$$BIAS=\frac{\sum_{i=1}^{N}\left(P_i-A_i\right)}{N} \tag{10}$$

Mean Absolute Percentage Error:

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{A_i-P_i}{A_i}\right)100\% \tag{11}$$

Mean Squared Error:

$$MSE=\frac{\sum_{i=1}^{N}\left(A_i-P_i\right)^2}{N} \tag{12}$$

Root Mean Squared Error Ratio:

$$RSR=\frac{\sqrt{MSE}}{\sigma_A} \tag{13}$$

Nash Number:

$$NashNumber=1-\left[\frac{\sum_{i=1}^{N}\left(A_i-P_i\right)^2}{\sum_{i=1}^{N}\left(A_i-\overline{A}\right)^2}\right] \tag{14}$$

Relative Root Mean Square Error

$$RRMSE=\sqrt{\frac{1}{N}\sum\left(\frac{A_i-P_i}{A_i}\right)^2} \tag{15}$$

where, A and P are the actual and predicted yields respectively, $\overline{A}$ and $\overline{P}$ are their means, σ_A is the standard deviation of the observed yield data, and N is the total number of data points. According to statistical norms, low values of MSE, MAPE, RSR, and RRMSE denote the most appropriate models for yield estimation. Similarly, the models with ρ_p and Nash number closer to 1 are highly accurate. Further, the model accuracy is vindicated by very small values of BIAS close to zero, while its positive and negative values indicate overestimation and underestimation respectively.

3. Results and Discussion

3.1. Collinearity

As per the correlation values between paddy yield and weather indices presented in the correlation matrix (Figure 5), the RF, maximum humidity, T_{min}, and evening wind are negatively correlated with the yield, while the T_{max}, evaporation, and morning wind are positively correlated with the yield. When the correlation values among weather indices were examined, strong and positive associations could be observed within the pairs of maximum RH vs minimum RH and evaporation vs T_{max}. Further, moderate positive correlations exist within the pairs of RF vs RH, sunshine hours vs T_{max}, sunshine hours vs evaporation, evaporation vs evening WS, and the maximum RH vs morning WS. On the other hand, strong negative correlations can be seen within the pairs of T_{max} vs RH, rainfall vs T_{max}, and maximum RH vs sunshine hours. Moreover, the pairs of maximum RH vs evaporation, RF vs evaporation and RF vs sunshine hours are negatively associated.

	Yield	Minimum RH	T_{max}	Evening WS	Evaporation	Morning WS	RF	Sunshine hours	T_{min}	Maximum RH
Yield	1.00									
Minimum RH	ρ_p = -0.05 ρ_s = 0.01	1.00								
T_{max}	ρ_p = 0.21 ρ_s = 0.18	ρ_p = -0.79 ρ_s = -0.79	1.00							
Evening WS	ρ_p = -0.20 ρ_s =-0.10	ρ_p = 0.15 ρ_s = 0.17	ρ_p = -0.07 ρ_s = -0.09	1.00						
Evaporation	ρ_p = 0.39 ρ_s =0.42	ρ_p = -0.56 ρ_s = -0.44	ρ_p = 0.78 ρ_s = 0.79	ρ_p = 0.75 ρ_s = 0.63	1.00					
Morning WS	ρ_p = 0.22 ρ_s =0.25	ρ_p = 0.03 ρ_s = 0.21	ρ_p = -0.07 ρ_s = -0.18	ρ_p = 0.67 ρ_s = 0.63	ρ_p = 0.68 ρ_s = 0.49	1.00				
RF	ρ_p = -0.14 ρ_s = -0.12	ρ_p = 0.57 ρ_s = 0.62	ρ_p = -0.79 ρ_s = -0.84	ρ_p = -0.22 ρ_s = -0.24	ρ_p = -0.59 ρ_s = -0.67	ρ_p = -0.15 ρ_s = -0.01	1.00			
Sunshine hours	ρ_p = 0.07 ρ_s = 0.03	ρ_p = 0.51 ρ_s = -0.46	ρ_p = 0.58 ρ_s = 0.57	ρ_p = 0.47 ρ_s = 0.38	ρ_p = 0.70 ρ_s = 0.59	ρ_p = 0.51 ρ_s = -0.36	ρ_p = -0.52 ρ_s = -0.53	1.00		
T_{min}	ρ_p = -0.17 ρ_s = -0.13	ρ_p = 0.11 ρ_s = 0.04	ρ_p = 0.27 ρ_s = 0.32	ρ_p = 0.52 ρ_s = 0.53	ρ_p = 0.33 ρ_s = 0.47	ρ_p = 0.17 ρ_s = 0.11	ρ_p = -0.46 ρ_s = -0.56	ρ_p = 0.22 ρ_s = 0.25	1.00	
Maximum RH	ρ_p = -0.30 ρ_s = -0.30	ρ_p = 0.84 ρ_s = 0.80	ρ_p = -0.81 ρ_s = -0.84	ρ_p = 0.60 ρ_s = 0.64	ρ_p = -0.87 ρ_s = -0.68	ρ_p = 0.71 ρ_s =0.80	ρ_p = 0.54 ρ_s = 0.66	ρ_p = -0.93 ρ_s = -0.80	ρ_p = 0.09 ρ_s = -0.07	1.00

Fig.5. Correlation matrix

3.2. Crop-Weather Models

For each geographical region considered in this study, four GPR models were constructed based on four types of Kernel functions namely Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2. The best fitting model for a particular district was selected based on the correlation produced by each Kernel function. For example, the performance of each model when the four Kernel functions were applied on the data of Ampara district is illustrated in Figure 6. The GPR model based on the Matern 5/2 Kernel, which yielded MSE= 0.02 and R=0.97, outperformed the models developed by applying the other three Kernel functions. The Quadratic GPR generated the next impressive model with MSE= 0.04 and R=0.95. The other two estimations were not as accurate as the aforementioned models (MSE>0.05 and R<0.94) due to the presence of deviations from the line of best fit.

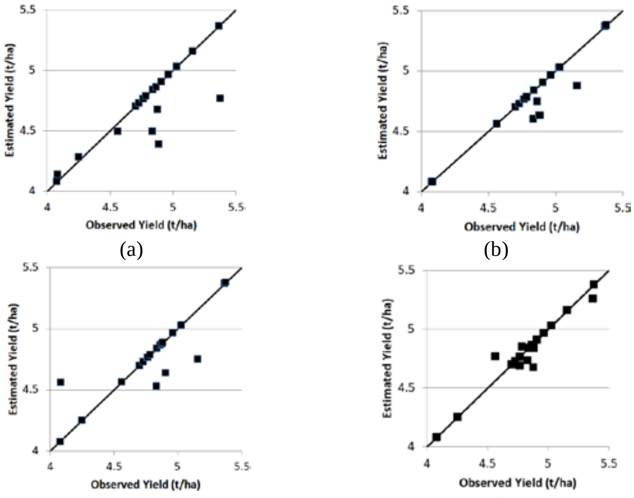


Fig.6. Comparison of the Observed with Predicted Paddy Yield in Ampara district (a) Exponential GPR (b) Quadratic GPR (c) Squared Exponential GPR (d) Matern 5/2 GPR

The crop-weather model developed for Ampara district can be approximated by the following linear relationship:

$$Yield = -10.4 + 0.000079RF + 0.377T_{\max} - 0.091T_{\min} + 0.67RH_{\min} + 0.176WS_m - 0.141WS_e \tag{16}$$

where, RF is rainfall, $T_{\max}$ and $T_{\min}$ represent maximum and minimum temperatures, $RH_{\min}$ is the minimum relative humidity, and WS_m / WS_e are the morning/ evening wind speeds respectively. The variation of paddy yield in Ampara district estimated by the Matern 5/2 GPR model is shown in Figure 7. It provides a visual expression of the model accuracy illustrating that the differences between the estimated and the observed yield values are insignificant.

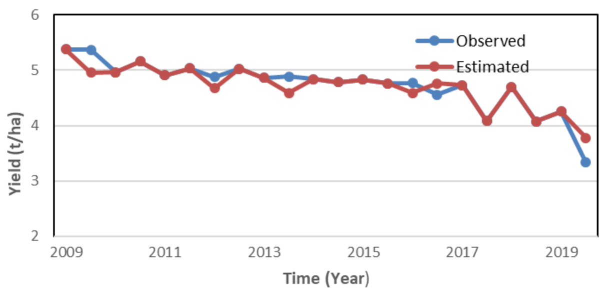


Fig.7. Paddy yield in Ampara district estimated by the Matern 5/2 GPR model

Similarly, GPR-based models were constructed for each district employing all four Kernel functions and their performance was assessed in terms of the statistical measures. The Matern 5/2 Kernel function produced the best model for the districts of Ampara, Hambantota, Batticaloa, Trincomalee, Badulla, and Monaragala, while the Exponential Kernel function was the most appropriate method in modeling data of the other three districts. The variation of the predicted yield in each district against the corresponding observed values produced by the best model is shown in Figure 8. It can be observed that most of the values estimated by using the GPR-based modeling are equal or hover around the observed yield. Further, model proficiency was assessed using the ρ_p, MAPE, MSE, RSR, Nash Number, and the BIAS as presented in Table 3.

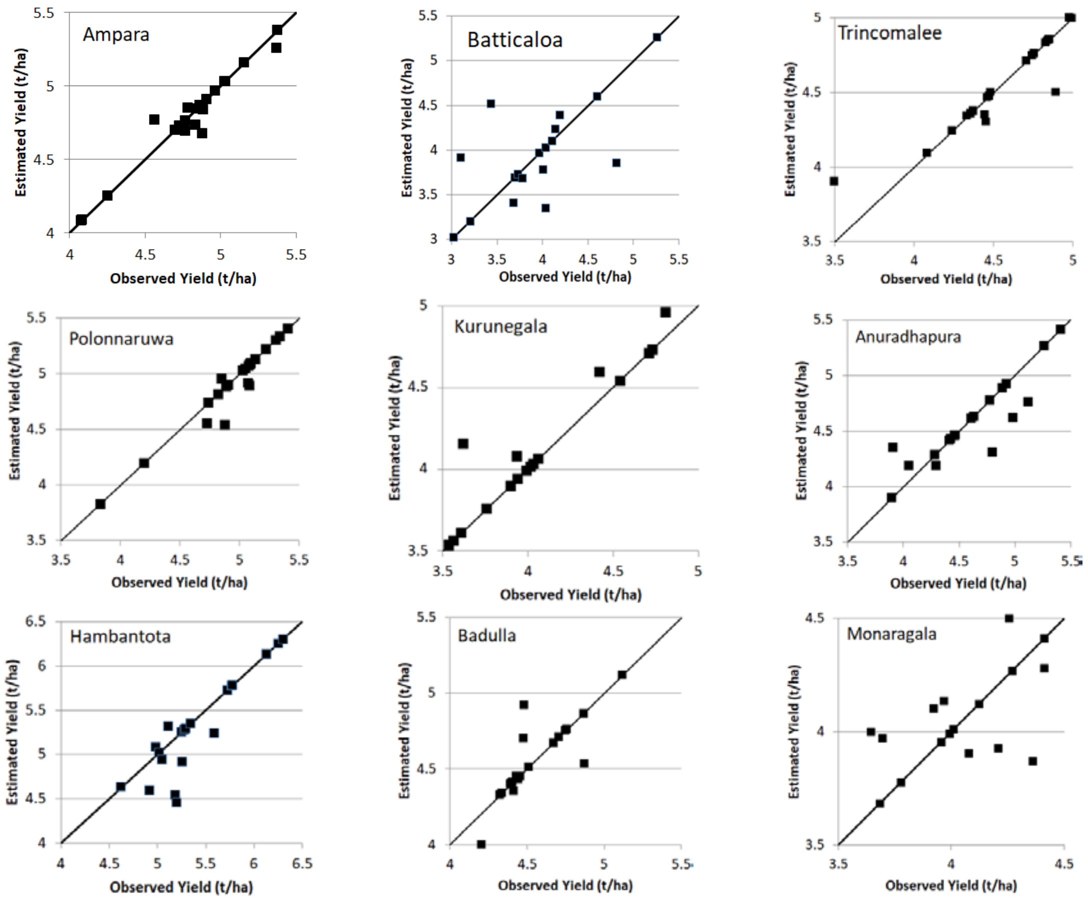


Fig.8. Comparison of the Observed with Predicted Paddy Yield

Table 3. Performance of GPR-based Prediction Models

Statistical Parameter	Ampara	Polonnaruwa	Kurunegala	Anuradhapura	Hambantota	Batticaloa	Trincomalee	Badulla	Monaragala
MSE	0.02	0.01	0.02	0.03	0.06	0.15	0.02	0.02	0.04
ρ_p	0.97	0.97	0.96	0.91	0.94	0.87	0.97	0.92	0.61
MAPE (%)	2.4	0.9	1.1	1.8	2.4	5.3	1.3	1.3	4.1
Nash number	0.79	0.93	0.91	0.83	0.79	0.74	0.95	0.84	0.32
RSR	0.44	0.26	0.29	0.40	0.44	0.49	0.22	0.39	0.80
BIAS	-0.10	-0.03	0.04	-0.03	-0.10	-0.001	-0.01	0.002	0.01
RRMSE	0.05	0.02	0.34	0.04	0.05	0.11	0.03	0.03	0.05

The MSE values of all the models except Batticaloa were less than or equal to 0.06. RRMSE values were less than 0.35 indicating less error of the models and the BIAS values varied between –0.1 and 0.04 ensuring that none of the models overestimate or underestimate the yield. All the models except Monaragala and Batticaloa produced very high values of R greater than 0.91. Further, RSR values were less than 0.5 and MAPE values were less than or equal to 2.4%. Further, the Nash number has also recorded high values above 0.74 except in Monaragala. These exceptions at Monaragala and Batticaloa are consistent with some deviated data points forecast by the model, seen in Figure 8. It is quite natural to come across with some marginal exceptions, due to tough living conditions in underprivileged districts coupled with sparse data collection facilities. Nevertheless, all these statistical indicators are in agreement with the general conclusions reached from this study.

3.3. Discussion

Though paddy yields the widely consumed food item of most people, no research has been conducted so far on the yield prediction at country level by applying the most suitable Kernel function in modeling. The performance of the proposed GPR-based models (Table 3) was compared with that of the prediction models derived in similar studies (Table 4). In related research studies, numerous machine learning, deep learning, and data mining techniques were found to forecast the paddy yield accurately, using climate variables in different parts of the world. It is noteworthy that

GPR has outperformed ANN, SVMR, and statistical regression techniques as per the performance assessed in terms of statistical parameters such as R, R^2, MSE, MAPE and Nash number [17]. Further, GPR was reportedly accurate when predicting the yield of other crops as well at country level. For example, wheat yield was predicted with $0.79 < R^2 < 0.81$, RMSE <750 kg/ha, and MAE <531 kg/ha in China [19].

The models proposed in this research were compared with GPR-based paddy yield prediction models presented in similar research studies, listed in Table 5. Accordingly, the modeling methods introduced in this paper can be applied to forecast rice yield in Sri Lanka accurately. However, all those previous studies were limited to the use of a single Kernel or two. In contrast, the most accurate Kernel for each data set was applied in this research, thus optimizing the performance of the models. If one of the Kernel functions recommended in previous research studies (Exponential/ RBF /Squared Exponential) were applied, the accuracy of the proposed crop-weather models would be reduced. More importantly, the Matern 5/2 Kernel, which was the most suitable function in developing models for 6 regions in this research, was not recommended in previous related research studies [16-19]. The use of Exponential Kernel, which was found to be the best in developing models for the other 3 regions, was also not recommended by other researchers though it was used in previous research conducted by the authors [11]. Further, the use of 5-fold cross-validation in training and validation of data ensured no overfitting, thus adding further strength to the reliability of this work.

When the weather is forecast, it can serve as input for GPR models proposed in this paper for estimating the corresponding paddy yield accurately. Time series methods (e.g. Auto Regressive Integrated Moving Average) can be applied to predict the weather from the data used in this study. In this context, the models developed in this research shall be utilized to assess the expected paddy yield in Sri Lanka accurately. Such information will be beneficial to the decision makers and other stakeholders at national level to ensure uninterrupted supply of paddy and to plan out its export or import needs in advance.

Table 4. Comparison of Performance of the Paddy Yield Prediction Models Developed Using Other Methods

Ref.	Study Area	Climate data	Technique(s)	Performance
[1]	China	Minimum, maximum temperature, precipitation, palmer drought severity index, evapotranspiration, vapor pressure, vapor pressure deficit and land surface temperature	Least Absolute Shrinkage and Selection Operator regression	$0.33 < R^2 < 0.42$ 633 < RMSE < 1231 kg/ha
			Random Forest	$0.76 < R^2 < 0.82$, 366 < RMSE < 723 kg/ha
			Long Short-Term Memory Networks	$0.77 < R^2 < 0.87$, 298 < RMSE < 724 kg/ha
[2]	Bangladesh	Rainfall, temperature and humidity	MLR	RMSE = 0.385
			Adaptive Boosting	RMSE = 0.404
			SVMR	RMSE = 0.424
			Modified Nonlinear Regression	RMSE = 0.369
[5]	South Korea	Minimum, mean and maximum temperatures and sunshine hours	Random Forest	RRMSE < 5%
[7]	Maharashtra state, India	Precipitation, Minimum Temperature, Average Temperature, Maximum Temperature and Reference Crop Evapotranspiration	Multilayer Perceptron Neural Network	Accuracy = 97.5%
[10]	Sri Lanka	Rainfall, morning and evening relative humidity, minimum and maximum temperature, wind speed, evaporation, and sunshine hours	ANN (Levenberg–Marquardt algorithm)	MSE = 0.019 0.78< R<0.96
			ANN (Bayesian Regularization algorithm)	MSE = 0.006 0.56< R<0.92
			ANN (Scaled Conjugated Gradient algorithm)	MSE = 0.074 0.65< R<0.92
[11]	North Western Province, Sri Lanka	Rainfall, temperature (minimum and maximum), evaporation, average wind speed (morning and evening), and sunshine hours	ANN	MSE= 0.04, R= 0.82, MAPE= 3.7%, Nash number= 0.67, RSR= 0.573, BIAS= 0.023
			SVMR	MSE= 0.10, R= 0.24, MAPE= 6.3%, Nash number= 0.16, RSR= 0.906, BIAS= 0.026
[18]	Nepal	rainfall, minimum and maximum temperature, relative humidity	2D- Convolutional Neural Network (CNN)	RMSE= 157
			3D-CNN	RMSE= 107.26
			Linear Regression	RMSE=452.90
			ridge regression	RMSE=473.06
			SVMR	RMSE=345.97

Table 5. Comparison of Performance of Paddy Yield Prediction Models developed using GPR

Ref.	Study Area	Climate data	Kernel Function	Performance
[11]	North Western Province, Sri Lanka	Rainfall, temperature (minimum and maximum), evaporation, average wind speed (morning and evening), and sunshine hours	Exponential	MSE= 0.008, R=0.98
[17]	India	Temperature, humidity, rainfall, wind speed, UV index, sun hours, and pressure	RBF	R^2 = 0.90
			Rational Quadratic	R^2 = 0.87
[18]	Nepal	rainfall, minimum and maximum temperature, relative humidity	Squared Exponential	RMSE=336.34
Proposed	Sri Lanka	Rainfall, Relative Humidity, Minimum Temperature, Maximum Temperature, Average Wind Speed, Evaporation, Sunshine hours	Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2	$0.01<MSE<0.15$ $0.61<R<0.97$

4. Conclusion

In this study, GPR-based crop-weather models were developed for paddy yield prediction in Sri Lanka by using the statistics of the major regions, which supply the bulk of paddy production of the country. The statistical measures of MSE, RRMSE, MAPE, RSR, BIAS, R, and the Nash number properly justified the coherence of predictions generated by the models with the observed values. As the model accuracy has also been verified in comparison to a number of studies conducted in other countries, the proposed models can be used in real-world applications to predict the paddy yield. A reliable model of such accuracy implies that when the climate data during a paddy cultivation season are known, the yield can be inferred by using the proposed GPR models. A notable novelty is extracting the best GPR-based model by applying four different Kernel functions for each data set. A total of 36 models were developed altogether as the Kernel functions of Rational Quadratic, Exponential, Squared Exponential, and Matern 5/2 were applied on weather statistics of each region. The Matern 5/2 Kernel function produced the best model for 6 districts (Ampara, Hambantota, Batticaloa, Trincomalee, Badulla, and Monaragala), while the Exponential Kernel was the most appropriate function in modeling data of the other three districts (Polonnaruwa, Kurunegala, and Anuradhapura). If this research were confined to the application of a widely used Kernel or a few Kernel functions, the accuracy of the models developed for most of the regions would have degraded considerably.

It could also be concluded that maximum temperature, evaporation, and morning wind are positively correlated with the paddy yield. Nevertheless, the RF, maximum RH, Tmin, and evening WS are negatively correlated with the paddy yield. These findings would encourage the researchers to conduct a more comprehensive study on the impact of all the weather indices on paddy yield, which remains to be addressed in the Sri Lankan context. Given further technological improvements, more climate data collection stations in remote regions and taking the area of cultivation into account, the yield and harvest of a forthcoming season can also be predicted. In this context, projecting future climate data and applying them as input variables into the proposed models enable the prediction of paddy yield in Sri Lanka. Such an accurate scientific approach will facilitate the establishment of a proper harvest management system solving a major problem faced by the Sri Lankan farmers. Moreover, the unexpected fluctuations in paddy price in Sri Lanka can be controlled if the proposed crop-weather models are adopted for predictions. The government will also receive benefits as the import or export of rice and the cultivation of minor crops can be planned in advance.

Acknowledgment

Authors are grateful to Department of Census and Statistics and the Department of Meteorology, Sri Lanka for providing past records of paddy harvest, yield and climate data.

Conflicts of Interests

No conflict of interest was declared by the authors.

References

[1] Juan Cao et al., "Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches," *Agricultural and Forest Meteorology,* vol. 297, pp. 108275, 2021, "doi:10.1016/j.agrformet.2020.108275".

[2] U. K. Dey, A. H. Masud, and M. N. Uddin, "Rice Yield Prediction Model using Data Mining," *In 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE),* Cox's Bazar, Bangladesh, 321-326, (2017), "doi:10.1109/ECACE.2017.7912925".

[3] N. A. Noureldin, M. A. Aboelghar, H. S. Saudy, and A. M. Ali, "Rice Yield Forecasting Models using Satellite Imagery in Egypt," *The Egyptian Journal of Remote Sensing and Space Science,* vol. 16, no. 1, pp. 125-131, 2013, "doi:10.1016/j.ejrs.2013.04.005".

[4] S. I. Na, J. H. Park, and J. K. Park, "Development of Korean Paddy Rice Yield Prediction Model (KRPM) using Meteorological Element and MODIS NDVI," *Journal of the Korean Society of Agricultural Engineers*, vol. 54, no. 3, pp. 141-148, 2012, "doi:10.5389/KSAE.2012.54.3.141".

[5] J. Kim et al., "Rice Yield Prediction in South Korea by using Random Forest," *Korean Journal of Agricultural and Forest Meteorology*, vol. 21, no. 2, pp. 75-84, 2019, "doi:10.5532/KJAFM.2019.21.2.75".

[6] D. Casanova, J. Goudriaan, M. C. Forner, and J. C. M. Withagen, "Rice Yield Prediction from Yield Components and Limiting Factors," *European Journal of Agronomy*, vol. 17, no. 1, pp. 41-61, 2002, "doi:10.1016/S1161-0301(01)00137-X".

[7] N. Gandhi, O. Petkar, and L. J. Armstrong, "Rice Crop Yield Prediction using Artificial Neural Networks," *In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, pp. 105-110, 2016, "doi:10.1109/TIAR.2016.7801222".

[8] A. K. Mariappan and J. A. B. Das, "A Paradigm for Rice Yield Prediction in Tamilnadu," *In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, pp. 18-21, 2017, "doi:10.1109/TIAR.2017.8273679".

[9] N. Vadaparthi, G. S. Tejaswini, and N. B. S. Pallavi, "A Novel Approach for Rice Yield Prediction in Andhra Pradesh," *In Advances in Decision Sciences, Image Processing, Security and Computer Vision*, vol. 4, pp. 688-692, 2020, "doi:10.1007/978-3-030-24318-0_78"

[10] V. Amaratunga, L. Wickramasinghe, A. Perera, J. Jayasinghe, and U. Rathnayake, "Artificial Neural Network to Estimate the Paddy Yield Prediction using Climatic Data," *Mathematical Problems in Engineering*, vol. 2020, 2020, "doi:10.1155/2020/8627824".

[11] L. Wickramasinghe, R. Weliwatta, P. Ekanayake, and J. Jayasinghe, "Modeling the Relationship between Rice Yield and Climate Variables Using Statistical and Machine Learning Techniques," *Journal of Mathematics*, vol. 2021, 2021, "doi:10.1155/2021/6646126".

[12] M. P. N. M. Dias, C. M. Navaratne, K. D. N. Weerasinghe, and R. H. A. N. Hettiarachchi, "Application of DSSAT Crop Simulation Model to Identify the Changes of Rice Growth and Yield in Nilwala River Basin for Mid-centuries under Changing Climatic Conditions," *Procedia Food Science*, vol. 6, no. 2016, pp. 159-163, 2016, "doi: 10.1016/j.profoo.2016.02.039".

[13] W. R. S. S.Dharmarathna, S. Herath, and S. B. Weerakoon, "Changing the Planting Date as a Climate Change Adaptation Strategy for Rice Production in Kurunegala District, Sri Lanka," *Sustainability Science*, vol. 9, no. 1, pp. 103-111, 2014, "doi:10.1007/s11625-012-0192-2".

[14] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review," *Computers and Electronics in Agriculture*, vol. 151, no. 2018, pp. 61-69, 2018, "doi: 10.1016/j.compag.2018.05.012".

[15] Y. S. Shiu and Y. C. Chuang, "Yield Estimation of Paddy Rice based on Satellite Imagery: Comparison of Global and Local Regression Models," *Remote Sensing*, vol. 11, no.2, pp. 111, 2019, "doi:10.3390/rs11020111".

[16] M. K. Mosleh, Q. K. Hassan, and E. H. Chowdhury, "Application of Remote Sensors in Mapping Rice Area and Forecasting its Production: A Review," *Sensors*, vol. 15, no. 1, pp. 769-791, 2015, "doi:10.3390/s150100769".

[17] Y. Vijayalata, V. R. Devi, P. Rohit, and G. R. Kiran, "A Suggestive Model for Rice Yield Prediction and Ideal Meteorological Conditions during Crisis," *International Journal of Scientific & Technology Research*, vol. 8, no. 9, pp. 1572-1576, 2019.

[18] R. Fernandez-Beltran, T. Baidar, J. Kang, and F. Pla, "Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3D CNN: A Case Study in Nepal," *Remote Sensing*, vol. 13, no. 7, pp. 1391, 2021, "doi:10.3390/rs13071391".

[19] J. Han et al., "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China," *Remote Sensing*, vol. 12, no. 2, pp. 236, 2020, "doi:10.3390/rs12020236".

[20] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, "Deep Gaussian Process for Crop Yield Prediction based on Remote Sensing Data," *In Thirty-First AAAI conference on artificial intelligence*, San Francisco, California, pp. 4559-4565, 2017.

[21] S. Lanka and M. Depārtamēntuva, "The National Atlas of Sri Lanka," Survey Department, Sri Lanka, 2007.

[22] [Online]. Available: https://meteo.gov.lk/. [Accessed: 15-Dec-2021].

[23] [Online]. Available: http://www.climate.lk/. [Accessed: 20.12.2021].

[24] A. Ly, M. Marsman and E. J. Wagenmakers, "Analytic Posteriors for Pearson's Correlation Coefficient," *Statistica Neerlandica*, vol. 72, no. 1, pp. 4-13, 2018, "doi:10.1111/stan.12111".

[25] V. S. Konduri, T. J. Vandal, S. Ganguly, and A. R. Ganguly, "Data Science for Weather Impacts on Crop Yield," *Frontiers in Sustainable Food Systems*, vol. 4, pp. 52, 2020, "doi:10.3389/fsufs.2020.00052".

[26] J. Hauke and T. Kossowski, "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data," *Quaestiones Geographicae*, vol. 30, no. 2, pp. 87-93, 2011, "doi:10.2478/v10117-011-0021-1".

[27] A. K. Sharma, Text book of Correlations and Regression, New Delhi, India, Discovery Publishing House, (2005).

[28] Joseph Isabona, Divine O. Ojuh," Machine Learning Based on Kernel Function Controlled Gaussian Process Regression Method for In-depth Extrapolative Analysis of Covid-19 Daily Cases Drift Rates ", International Journal of Mathematical Sciences and Computing, Vol.7, No.2, pp. 14-23, 2021.

[29] C. E. Rasmussen, and C. K. Williams, Gaussian Processes for Machine Learning, vol. 1, 2006.

[30] C. K. Williams., and C. E. Rasmussen, Gaussian Processes for Machine Learning, Cambridge, MA: MIT press, vol. 2, no. 3, pp. 4, 2006.

[31] N. Zhang, J. Xiong, J. Zhong, and K. Leatham, "Gaussian Process Regression Method for Classification for High-dimensional Data with Limited Samples," *In 2018 Eighth International Conference on Information Science and Technology (ICIST)*, Cordoba, Granada, and Seville, Spain, pp. 358-363, 2018, "doi:10.1109/ICIST.2018.8426077".

[32] S. Stajkowski, D. Kumar, P. Samui, H. Bonakdari, and B. Gharabaghi, "Genetic-algorithm-optimized Sequential Model for Water Temperature Prediction," *Sustainability*, vol. 12, no. 13, pp. 5374, 2020, "doi:10.3390/su12135374".

[33] A. Gholami et al., "Uncertainty Analysis of Intelligent Model of Hybrid Genetic Algorithm and Particle Swarm Optimization with ANFIS to Predict Threshold Bank Profile Shape based on Digital Laser Approach Sensing," *Measurement*, vol. 121, pp. 294-303, 2018, "doi:10.1016/j.measurement.2018.02.070".

Authors' Profiles

Piyal Ekanayake obtained his first degree, B.Sc. (Hons.) in Mathematics, from the University of Kelaniya, Sri Lanka (1986) and M.Sc. & D.Sc. Degrees from Kyushu University, Japan in 1991 & 1994 respectively. He was a Postgraduate Fellow at Oxford University, United Kingdom (2001 – 2002).

He worked as the Dean of the Faculty of Applied Sciences (2005-2017) and the Vice-Chancellor of the Wayamba University of Sri Lanka (2017-2020) and currently as an Associate Professor of Mathematical Sciences at the Department of Mathematical Sciences of the same University. His research areas include numerical modelling of atmospheric tides and applications of statistical and machine learning techniques. His research contribution amounts to over 50 papers/ abstracts in journals and conference proceedings.

Lasini Wickramasinghe received her B.Sc. (Joint Major) degree from Wayamba University of Sri Lanka, Kuliyapitiya, Sri Lanka in 2020 and reading for MPhil in electrical and electronic engineering at Faculty of Engineering, University of Peradeniya, Sri Lanka.

She is currently working as a Demonstrator at Wayamba University of Sri Lanka. Her research areas include machine learning based modelling and statistical data analysis. She has published more than 10 research articles/abstracts in journals and conference proceedings.

Jeevani W. Jayasinghe received Ph.D. in Engineering and B.Sc. (Hons.) in Engineering from the University of Peradeniya, Sri Lanka in 2014 and 2006 respectively. She was a Visiting Fellow of Macquarie University, Sydney, Australia in 2018 and a Research fellow of Ramon Llull University, Spain in 2013.

She is working as a Professor of the Department of Electronics, Wayamba University of Sri Lanka. She has published more than 80 research articles in journals and conference proceedings.

Prof. Jayasinghe is a Senior Member of IEEE, and an Associate Member of Institute of Engineers Sri Lanka. Her current Google scholar h-index is 13 and Scopus h-index is 11.

How to cite this paper: Piyal Ekanayake, Lasini Wickramasinghe, Jeevani W. Jayasinghe, "Development of Crop-Weather Models Using Gaussian Process Regression for the Prediction of Paddy Yield in Sri Lanka", International Journal of Intelligent Systems and Applications(IJISA), Vol.14, No.4, pp.52-65, 2022. DOI:10.5815/ijisa.2022.04.05