International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

A Soft Computing Model to Predict the Rice Production in India

Surjeet Kumar, Manas Kumar Sanyal

Abstract*: India, which has the most rice tillage area in the world, is one of the massive cultivators of this white crop. Besides, rice is the main staple food of many Indians. The main purpose of this study is to develop a predictive model on Indian rice production. In this, we have used different types of soft computing models like Fuzzy Logic, Statistical Equations, Artificial Neural Network (ANN) and Genetic Algorithm (GA) and developed a hybrid model to get the optimum result. The vital aspect of this predictive model is the accuracy of the future data prediction on the basis of past time series data. The Prediction performance has been assessed by using error finding equations like Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Average Error.

Keywords: Fuzzy Logical Relationships, Statistical Equations, Artificial Neural Network (ANN), Genetic Algorithm (GA), and Error Calculations.

I. INTRODUCTION

In agricultural and populated country like India, where to maintain the economic growth through cultivation is an inescapable challenge, there to estimate a fact-based rice production is a significant requirement of this economy. Indian agricultural sector contributes the most towards the country's GDP (Gross Domestic Product) calculation than the other two sectors (such that Industrial sector and Services sector). The Time Series data prediction is often conducted by soft computing technique for trend analysis and the Soft Computing technique is widely use in data prediction for all the sectors like agriculture, finance, weather, populations and so on. But here, the aim of our study is to generate a hybrid model to enhance the accuracy of the prediction and forecast of rice yield data. This new approach is applied on the secondary data of rice production [7]. Here, rice cultivation has been noted down with respect to quintal per hectare. In this, Fuzzy logic, Statistical Equations, Artificial Neural Network (ANN) and Genetic Algorithm (GA) are applied for predicting the rice production. In fuzzy logic, first order fuzzy logical relationship is used to predict the data set. Statistical equations like linear, Exponential, Fourier, Gaussian, Polynomial and sum of sine are also applied on rice production estimation. Based on minimum error we have selected one equation and then modified this equation by using Artificial Neural Network (ANN). Finally, Genetic Algorithm (GA) has applied on this data set and prediction performances have been assessed by using error calculating technique. The accuracy and efficiency of this forecast depends on time series data. We have developed a hybrid model which gives more accurate data prediction but it may not be 100%. In agriculture, rice cultivation is majorly depending on weather factor. If unseasonable natural conditions or any natural calamities happen then production may be decreased and the accuracy of prediction also affected. At that condition we can ignore it once.

The rest of this material has been arranged as: In section 2, literature review of soft computing technique. In section 3, Methodology used for prediction. In section 4, Results of the developed model have been explained. In section 5, Conclusion and opportunities. Finally, the reference has been described in section 6.

II. LITERATURE REVIEW

Idea of the fuzzy time series, able to handle the inexplicit and indistinct facts represented as semantic variables has been generated by Song and Chissom [1] by applying set theories and semantic variables given by Zadeh [2,3] Further, he has expanded his theory of fuzzy time series to make it more able to handle the quantitative data by presenting the idea of fuzzification and defuzzification and used it to estimate the pupil enrolments of university of Alabama [4,5] and solved the issue of huge mathematical needs of Song and Chissom method of calculating the fuzzy relationships using max-min composition by substituting it with more feasible numeric function. Fuzzy time series forecasting techniques generally based on the successive values. Other evidence to improve the fuzzy time series prediction arisen from the time variant models by implementing the higher degree procedures of fuzzy time series prediction [6]. This forecasting process provides assistance to reach the accurate and elicited management in farming. Artificial Neural Network (ANN) is a commonly used method of forecasting. Though ANNs have been used to solve various prognostication issues [8], but they have many limitations. The studying algorithm, like back-propagation algorithm, balances the network's weights and biases by computing gradient of the error and next forecasts the error backward by the network to adjust the weights and biases [9]. Many researchers have undertaken in-depth study on the topics including neural networks in various disciplines; but very few have performed the study on neural networks to predict the agronomical cultivation [10-14]. Arif Hepbasli determined agronomical electricity demand through Genetic Algorithm (GA) [15].

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Surjeet Kumar*, Business Administration, University of Kalyani, West Bengal, India.
Manas Kumar Sanyal, Business Administration, University of Kalyani, West Bengal, India.

At recent, some of the researchers have determined the transport energy intake though Genetic Algorithm (GA) [16,17].

III. METHODOLOGY

A. Fuzzy Logic

Fuzzy logic is a logic based mathematical method to solve the prediction problem depends on past time series data. The basic concept of fuzzy time series model has been discussed here [1,4,5].

A fuzzy set is a group of entities with a continuity of membership's gradation. Say, S be the Universe of discourse. $S = \{s_1, s_2, s_3 \ldots s_n\}$, where s_i are feasible semantic values of S, after that a fuzzy set of semantic variables Z_i of S is explained as–

$$Z_i = \mu Z_i(s_1)/(s_1) + \mu Z_i(s_2)/(s_2) + \mu Z_i(s_3)/(s_3) + \ldots + \mu Z_i(s_n)/(s_n) \quad (1)$$

Here, μZ_i is the membership function of the fuzzy set Z_i, in such a way μZ_i: $S \rightarrow [0, 1]$. If sj is the member of Z_i and then μZ_i (s_j) is the degree of belonging of s_j to Z_i.

Above, it has been defined that universe of discourse S as $[W_{min} - W_1$ and $W_{max} + W_2]$ to contain historical data, where W_{min} and W_{max} are the minimum and maximum data from rice cultivation of India respectively. $W_{min} = 3219$ and $W_{max} = 4554$. Here, W_1 and W_2 are used as a positive number, which are appropriately chosen by the users. Let, $W_1 = 3219 - 19$ and $W_2 = 4554 + 46$ for the purpose of dividing the range. We are showing the range of universe of discourse, S= [3200-4600], minimum and maximum range. The universes of discourse are divided in seven equal length denoted as s_1, s_2, ..., s_7 and the interval between W_{min} to W_{max} is 200 shown in table1.

Table 1. Showing The Range Of The Production And Quality Remark

Universe of Discourse	Range	Fuzzy Set	Production Remark
S_1	(3200-3400)	Z_1	Poor
S_2	(3400-3600)	Z_2	Below average
S_3	(3600-3800)	Z_3	Average
S_4	(3800-4000)	Z_4	Good
S_5	(4000-4200)	Z_5	Very good
S_6	(4200-4400)	Z_6	Excellent
S_7	(4400-4600)	Z_7	Outstanding

In the above table it has been mentioned that seven fuzzy sets of semantic variables are Z_1, Z_2..., Z_7 and also indicated that quality of rice production in India. The grades of the membership to these fuzzy sets of semantic variables are explained as:

$Z_1 = 1/S_1 + 0.5/S_2 + 0/S_3 + 0/S_4 + 0/S_5 + 0/S_6 + 0/S_7$

$Z_2 = 0.5/S_1 + 1/S_2 + 0.5/S_3 + 0/S_4 + 0/S_5 + 0/S_6 + 0/S_7$

$Z_3 = 0/S_1 + 0.5/S_2 + 1/S_3 + 0.5/S_4 + 0/S_5 + 0/S_6 + 0/S_7$

$Z_4 = 0/S_1 + 0/S_2 + 0.5/S_3 + 1/S_4 + 0.5/S_5 + 0/S_6 + 0/S_7$

$Z_5 = 0/S_1 + 0/S_2 + 0/S_3 + 0.5/S_4 + 1/S_5 + 0.5/S_6 + 0/S_7$

$Z_6 = 0/S_1 + 0/S_2 + 0/S_3 + 0/S_4 + 0.5/S_5 + 1/S_6 + 0.5/S_7$

$Z_7 = 0/S_1 + 0/S_2 + 0/S_3 + 0/S_4 + 0/S_5 + 0.5/S_6 + 1/S_7$

Fuzzification is a process of identifying associations between time series data and the fuzzy sets and range are also explained in the table1. Each production data is fuzzified as to its highest degree of participation. With these fuzzified values, we can create the fuzzy logical relationship of different orders like ($order_1$, order2, order3, order4, and so on) but in this study we have used the first order logic to predict the data set. Rice production data from 1981 to 2003 have been fuzzified on the basis of Gaussian function, has been specified in Table 2.

Table 2. Showing the Actual data and fuzzified data

Actual value	Fuzzy Set	Z1	Z2	Z3	Z4	Z5	Z6	Z7
3552	Z2	0.24	1	0	0	0	0	0
4177	Z5	0	0	0	0.115	1	0.885	0
3372	Z1	1	0.86	0	0	0	0	0
3455	Z2	0.725	1	0.275	0	0	0	0
3702	Z3	0	0.49	1	0.51	0	0	0
3670	Z3	0	0.65	0.675	0.35	0	0	0
3865	Z4	0	0	0.96	1	0.325	0	0
3592	Z2	0.04	1	0	0	0	0	0
3222	Z1	1	0.89	1	0	0	0	0
3750	Z3	0	0.25	0.745	0.75	0	0	0
3851	Z4	0	0	0	1	0.255	0	0
3231	Z1	1	0.845	0	0	0	0	0
4170	Z5	0	0	0	0.15	1	0.85	0
4554	Z7	0	0	0.64	0	0	0.23	1
3872	Z4	0	0	0	1	0.36	0	0
4439	Z7	0	0	0	0	0	0.805	1
4266	Z6	0	0	0	0	0.67	1	0.33
3219	Z1	1	0.905	0	0	0	0	0
4305	Z6	0	0	0.36	0	0.475	1	0.525
3928	Z4	0	0	0.11	1	0.64	0	0
3978	Z4	0	0	0.65	1	0.89	0	0
3870	Z4	0	0	1	1	0.35	0	0
3727	Z3	0	0.365		0.635	0	0	0

Prediction data = the value of fuzzy order $(Z_i) - (Z_{ij})$, where $Z_j \rightarrow Z_i$, the above table gives the relational matrix of Z_i and Z_j.

B. Statistical Equations

Statistical equation is a part of soft computing technique. Statistical equations like linear, exponential, gaussian, polynomial, Fourier series and sum of sine are used to predict the data set. Curve Fitting tool gives an adjustable interface where you can reciprocally fit curves and surfaces to the data and can see the plots. First, we check which equation is fitted for rice production. After that, we do error analysis. Here, we see that Fourier equation gives the minimum error. The constant value, obtains from the Fourier series, which are applied to predict the future data according to the actual data.

Fourier series equation:

$$f(y) = m_0 + m_1 * \cos(y*v) + n_1 * \sin(y*v) \quad (2)$$

Co-efficient (with 95% confidence bounds) is m0 = 3813 (3670, 3956), m1 = -41.55 (-2753, 2670), n1 = 325.2 (-90.06, 740.4), v = 2.169(2.078, 2.259).

Where m0, m1, n1 is constant, v is the fundamental frequency.

C. Artificial Neural Network

Neural Network is a computational model whose layered structure simulates the networked structure of the neurons. Layers are connected with nodes. Neural network learns from data, hence it can be taught to identify the patterns, grouped the data, and estimate the forthcoming events. Three basic things are used in neural network: first one is Input layer, second one is hidden layer and third one is output layer. In statistical equation, constant value obtains from the Fourier equation. In figure 1, two constant values are used as an input layer, three neurons are used as hidden layers, and after training we get two optimum constant values in output layer. In neural network, weights are automatically adjusted during training until it gives optimum result. This constant value gives more accurately predicted data set. Each iteration shows that error of prediction data. Neural network also gives the forecast data. But the accuracy of forecast data is depending on the prediction data.

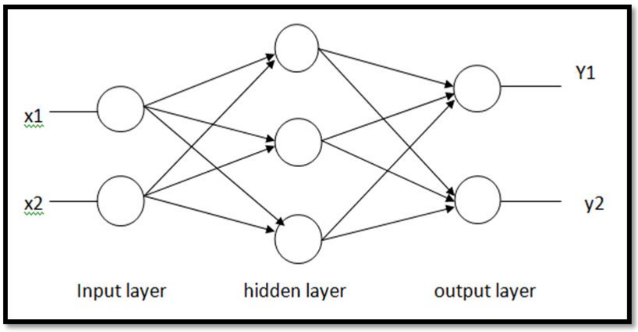


Figure 1. Architecture of Artificial Neural Network

D. Genetic Algorithm (GA)

Genetic Algorithm (GA) is as same as the Natural evolution mechanism. Genetic Algorithm (GA) is a process of finding solution for both constrained and unconstrained optimized problems which is based on natural selection; the method is driven from the biological evolution. Genetic Algorithm (GA) chooses the individuals randomly from the present population to be represented and used them to give birth of a child for next generation [18, 19]. Three basic rules follow:

- Selection – A method to select the individual's chromosomes, called parents, those parents' chromosomes donate to the population at next generation.
- Crossover –This method depends on the Combinations of two parents to form a child for the next generation.
- Mutation –This procedure is used for random changes of the parents' chromosomes to form a child.

Genetic Algorithm (GA) is based on Artificial Neural Network (ANN) for predicting the Indian rice production. This data set is a secondary data set. The prediction accuracy is further improved using genetic algorithm. The mostly used notations in Genetic Algorithm are binary codes (0, 1). So, the data have to convert into binary form and that is called chromosomes. When the conversion of decimal to binary happens, we have to manage the equality of bit for each data set. Chromosomes are represented into decision variables' domain. Hence, it is feasible to evaluate the performances, or fitness, of the independent members of a population.

1. *Selection:* Selection process involves in searching the member chromosomes, as stated the fitness of the members is based on fitness values. In selection procedure, Roulette Wheel method is applied on the given data set. Roulette wheel is a simple and randomly determined selection method coined by Holland [14]. It is classified under proportional selection method as it chooses the individuals on the basis of proportionate probability of the fitness. The fundamental of Roulette Selection is a linear search by a Roulette Wheel with the slits in the wheel weighted in proportion to the individual's fitness values. All the chromosomes of the population are putted on the Roulette Wheel as to their fitness value [19]. There are more chances of high fitness selections. Roulette Wheel Selection applies the exploitation method in its application. The average fitness of population for i^{th} generation in Roulette Wheel Selection is computed by-

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \quad (3)$$

Where f_i indicates that fitness of respective i^{th} population and N indicates the maximum limit to the population.

2. *Cross-over-* Crossover is an important method which allows mutual interchange of the designing features between two copulates. First, selecting two copulates and then from which randomly select two chromosomes in a string and the strings between these two chromosomes of the copulates are interchanged. In crossover procedure, Selection is randomly used but may be chose one chromosome or more than one chromosome from a given population set which is totally depends on copulates' characteristics. Showing crossover procedure below:

Parent1 = 1010101110
Parent2 = 1000011111
Child1 = 1010001111
Child2 = 1000111110

The crossover function is used with a probability of Pc and Crossover chromosomes have to underline.

3. *Mutation-* Mutation operation is another important process in GA. Mutation is a process of changing of the chromosome from 0 to 1 and vice versa. Let, a small mutation probability p_x. The chromosome mutation is executed as bit by bit through flipping a coin with the probability of p_x. This process acts at a random place of the chromosomes with a low probability of p_x [2].
4. *Fitness Function-* The fitness function may not the only function which is correlated closely with the designer's objectives. It must also be calculated quickly. A fitness function has a specific type of objective function that is applied to summarise, as a single figure of merit, how close a given design

solution is to achieve the set objectives. Thus, in each generation individual with minimum difference must be returned. In this study, Mean Absolute Percentage Error (MAPE) is used as fitness function. The general equation of MAPE is as follow:

MAPE = 1/n ∑ [(│Actual Value – Forecasted Value│)/ Actual Value] * 100.

Where n is number of fitted point of data set.

E. Error Calculation Technique

To check the accuracy of the performance of the various soft computing techniques, we are using these statistical parameters.

Mean Square Error (MSE) = $\sum_{i=1}^{n}$(Actual Value – Forecast Value)^2/n (4)

Mean Absolute Percentage error = $\sum_{i=1}^{n}$[mod (E_i-F_i) / E_i] *100/ n (5)

Where E_i=Actual value and F_i = forecast value.

Residual Analysis –

Absolute Residual = mod [Actual value – Forecast value]. (6)

Mean Absolute Residual = mod [Actual value – Forecast value]/Actual value.

F. Actual Data Used

Rice cultivation of India is predicted on the basis of past time series data. This data is a secondary data. The graphical presentation of this data set has been given in figure 2.

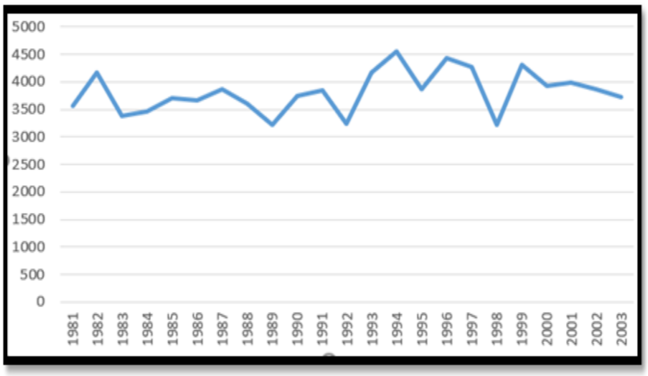


Figure 2. Showing The Actual Time Series Data

IV. RESULT AND DISCUSSION

A. In this paper, we have used Soft Computing Techniques like Fuzzy Logic, Statistical Equations, Artificial Neural Network (ANN) and Genetic Algorithm (GA). We have structured a hybrid model which gives more accurate and suitable results. Here we have executed Actual Data, Fuzzy Logic Predicted Data, Predicted Data of Fourier series, Modified Data of Artificial Neural Network Prediction and finally used Genetic Algorithm to get more accurate predictive data set, which is shown on the bellow table [3].

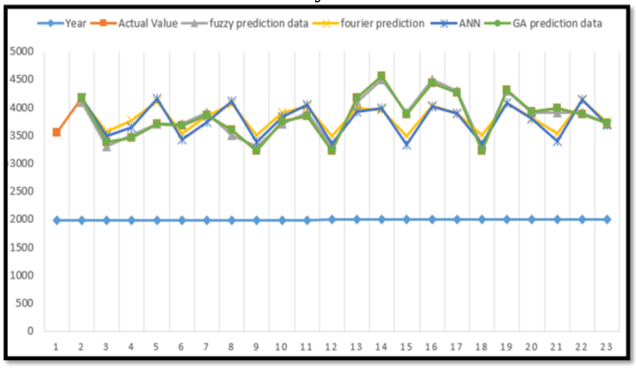
Table 3. Prediction data using soft computing technique

Year	Actual Data	Fuzzy Data	Fourier Data	ANN Data	GA Data
1981	3552				
1982	4177	4100	4132	4175	4177.6
1983	3372	3300	3571	3494	3371.8
1984	3455	3500	3766	3640	3455.9
1985	3702	3700	4107	4156	3702.7
1986	3670	3700	3528	3426	3671.5
1987	3865	3900	3839	3736	3867
1988	3592	3500	4068	4115	3597.1
1989	3222	3300	3499	3376	3226.5
1990	3750	3700	3911	3833	3754.5
1991	3851	3900	4016	4054	3852.4
1992	3231	3300	3486	3346	3230.9
1993	4170	4100	3978	3927	4171.3
1994	4554	4500	3954	3978	4559
1995	3872	3900	3489	3340	3872.3
1996	4439	4500	4036	4011	4440.1
1997	4266	4300	3884	3889	4266.1
1998	3219	3300	3508	3356	3220.2
1999	4305	4300	4084	4081	4304.9
2000	3928	3900	3812	3794	3928.6
2001	3978	3900	3542	3394	3978.2
2002	3870	3900	4118	4134	3870.2
2003	3727	3700	3739	3696	3727.9

B. Units The error calculation techniques like MSE, RMSE and Average Error has been used in table 4.

Soft computing technique	MSE	RMSE	Average Error
Fuzzy Logic	3080.569091	55.50287	0.013321
Fourier Equation	90642.55	301.069	0.068291
Artificial Neural Network	99443.18	315.34	0.067069
Genetic Algorithm	1.4554545	1.20639	0.000251

According to the table 3 and 4, Accuracy of the prediction and error calculation is more optimized using Genetic Algorithm. In Graph 1, Actual value is indicated by orange colour and Genetic Algorithm (GA) prediction data is shown by green colour. It is clearly showing that the predicted data is overlapped with the actual data. If prediction accuracy is more accurate then we can easily forecast the data set.



Graph 1. Accuracy of prediction data using soft computing Vs Actual data.

V. CONCLUSION

In this study, the time series data of rice production from 1981 to 2003 is used for prediction. Soft computing techniques like Fuzzy logic, Statistical Equations (Linear, Exponential, Gaussian, Polynomial, Fourier series and Sum of Sine), Artificial Neural Network (ANN) and Genetic algorithm (GA) have been used for forecasting data. In statistical equation, Fourier series is selected based on error analysis. The error is further reduced by Artificial Neural Network (ANN) based on properly optimized constant value of Fourier series. Finally, Genetic Algorithm (GA) has been applied after using Neural Network to get more accurate data prediction.

It has been noticed that the results of the proposed hybrid model become better in performance.

REFERENCES

1. "Fuzzy time series and its models, Fuzzy Sets Systems",54: 269-77, Song, Q. and Chissom, B.S. 1993.
2. "Fuzzy set. Information and Control", 8: 338–353, Zadeh, LA. (1965).

3. "The concept of a linguistic variable and its application to approximate reasoning", Part 1–3, Information. Science. 8 199 249, Zadeh LA (1975).
4. "Forecasting enrolments with fuzzy time series", Part I, Fuzzy Sets Systems, 54: 19, Song, Q. and Chissom, B.S. 1993.
5. "Forecasting enrolments with fuzzy time series", Part II, Fuzzy Sets Systems, 64: 1-8, Song, Q. and Chissom, B.S. 1994.
6. "A fuzzy time series prediction method based on consecutive values", IEEE International Fuzzy Systems Conference, August 22-25, (1999)703-707, I. Kim and S.R. Lee.
7. Data Source Url-"https://www.researchgate.net/publication/320948199_A_Computational_Method_for_Rice_ProductionForecasting Based on High-Order Fuzzy Time Series".
8. M. Adya and F. Collopy, How Effective are Neural Networks at Forecasting and Prediction, Journal of Forecasting, vol. 17, no. 5–6, 1998, pp. 481–495.
9. "Acceleration of Back Propagation through Initial Weight Pre-training with Delta Rule", in: Proceedings of the IEEE International Conference on Neural Networks, vol. 1, pp. 580–585, San Fransisco, CA, 1993.
10. "Analysis and correlation methods for spatial data", ASAE Paper No. 95-1335. ASAE, St. Joseph, Mich, Drummond, S.T., Sudduth, K.A., Birrell, S.J., 1995.
11. "Analysis of spatial factors influencing crop yield", In: Proc. of Int. Conf. on Precision Agri. pp. 129–140, Sudduth, K., Fraisse, C., Drummond, S., Kitchen, N., 1996.
12. "Use of artificial neural networks for predicting rice crop damage by greater flamingos in the Camargue", France. Ecol. Mod. 120, 349–358, Tourenq, C., Stephane, A., Laurent, D., Lek, S., 1999.
13. "Prediction of wheat yield using artificial neural networks", Ín: 25th Conf on Agricultural and Forest Meteorology, Safa, B., Khalili, A., Teshnehlab, A.M., Liaghat, A.M., 2002.
14. "Prediction of Tangerine Yield Using Artificial Neural Network". Chiang Mai University, Chiang Mai 50200, Boonprasom, P., Bumroongitt, G., 2002.
15. "Electricity estimation using genetic algorithm approach: a case study of Turkey", Energy (30), ppl003 1012,2003, H. Ozturk, H. Ceylan, O. E. Canyurt, A. Hepbasli,
16. "Energy demand estimation based on two-different genetic algorithm approaches", Energy Source, Vol.26(14),1313-1320,2004, O. E. Canyurt, A. Hepbasli, H. Ozturk.
17. "Genetic algorithm approach to estimate transport energy demand in Turkey", Energy policy, pp 89-98,2004, S. Haldenbilen, H. Ceylan.
18. Genetic Algorithm Tool Box for used with Matlab, Automatic Control and System Engineering, University of Sheffield.
19. "Genetic algorithms in search, optimization, and machine learning", Addison-Wesley Publishing Company, Reading, Goldberg DE (1989).

AUTHORS PROFILE

Surjeet Kumar University Research Scholar, University of Kalyani, West Bengal, India.

Prof. Manas Kumar Sanyal, Former Dean and Chairman of Departmental Research Committee, University of Kalyani, West Bengal, India.