J. Bio-Sci. 31(2): 25-34, 2023 **ISSN 1023-8654**
http://www.banglajol.info/index.php/JBS/index
DOI:https://doi.org/10.3329/jbs.v31i2.74143

EVALUATION OF THE CURRENT SITUATION OF TEA PRODUCTION AND CONSUMPTION IN BANGLADESH THROUGH DIFFERENT STATISTICAL MODELS

Md. Abdus Salam[1], Parvej Hasan Jon[1], Iftekhar Ahmad[1*], Mahbub Alam[1], Mohammed Taj Uddin[2], Hafijur Rahman[1], Md Ismail Haque[1], Md. Zahidul Islam[1] and Mitu Samadder[1]

[1]Department of Food Engineering and Tea Technology, Shahjalal University of Science and Technology, Sylhet, Bangladesh
[2] Department of Statistics, Shahjalal University of Science and Technology, Sylhet, Bangladesh

Abstract

Tea is considered a valuable non-alcoholic beverage worldwide and is gaining popularity as a healthy drink due to its multifarious medicinal properties. The tea industry is a pivotal economic driver of Bangladesh with rising production and consumption. Several statistical models were used to anticipate the best-fitted model and pattern of production and consumption until 2025. Data collected from numerous authentic sources and analyzed. Rational Quadratic Gaussian Process Regression (GPR) and Quadratic Support Vector Machines (SVM) models were chosen for tea production and consumption respectively based on RMSE and R-Square value. In 2022, this study predicts tea production to be 93.83 millon kg and consumption to be 98.48 million kg while intersecting each other. Our study suggests an existing gap in the production and consumption trend and this issue needs to be addressed imperatively.

Key words: GPR, Forecasting, Residuals analysis, Trend Analysis, Tea consumption and production.

Introduction

One of Bangladesh's primary exports is tea, which generates significant foreign exchange profits. Due to the rising local tea consumption and the highly competitive worldwide market, Bangladesh is facing significant difficulties. Bangladesh currently produces 1% of the world's tea. Because of increased domestic tea consumption, the country is gradually losing the global export market. Bangladesh must enhance tea production by cultivating more areas for tea production in order to expand its export market (Saha et al. 2021). In addition, they have recently engaged in small-scale tea farming to enhance the amount of tea produced in their nations (Arefin and Hossain 2022; Hossain et al. 2012). Growing domestic demand and a lack of the implementation of contemporary methods may contribute to lower production (Mamun and Ahmed 2011). In the dry season or during extended droughts, there is also a scarcity of long- term water supply for irrigation (Gilgen et al. 2001). In three regions of the Panchagar district, according to a recent feasibility assessment conducted by the Bangladesh Tea Board, there may be roughly 6000 acres of land available (Sultana et al. 2014). With the colonization of the Chittagong hill tracts, an additional 7500 hectares of suitable land will become accessible (Ahmad and Hossain 2013). It follows that the nation's tea production must rise to keep up with the rising per capita consumption, if not to maintain a constant export volume (Sultana et al. 2014).

Considering the export benefit in the context of both local and international markets, it may be necessary to conduct statistical research on the developing trend for production and consumption in the competitive market. Several trend models have been published in other research. Used an exponential growth model to track the expansion of tea acreage, output, and yield in Bangladesh (Saha et al. 2021). Although the majority of the research produced positive results, they did not conduct a statistical trend analysis to compare production and consumption. To the best of our knowledge, there exists no study in the literature that

*Author for correspondence: iftekhar-ttc@sust.edu

compares trend models on tea production and consumption in Bangladesh and evaluates the best trend model between them. It may help to promote producers while also influencing policymakers in their decision-making and improvement. There is an urgent need for research to update and improve models based on current data for reliable forecasting. With this context in mind, the study's objectives were to find trends and appropriate time series models to anticipate tea production and consumption in Bangladesh.

Materials and Methods

Dataset

A statistical bulletin produced by the Bangladesh Tea Board included time series data on yearly black tea production in million kg from 2000 to 2022. Also, tea area was collected from Bangladesh Tea Research Institute (BTRI), and rainfall data was collected from Bangladesh Meteorology Department (BMD) Bangladesh.

Methods

This study applied Gaussian process regression (GPR): Rational Quadratic GPR, Squared Exponential GPR, Matern 5/2 GPR, Exponential GPR, and Quadratic Support Vector Machine (Quad-SVM) to analyze trends and predict Bangladesh tea production and consumption. The trends models are below:

- **Rational Quadratic GPR**

  $k\ (xi, xj|\theta)\frac{r^2}{2\alpha\sigma^2{}_l})$

- **Squared Exponential GPR**

  $k\ (xi, xj|\theta)$ =exp$\frac{-(x_i-x_j)^T(x_i-x_j)}{2\sigma_l{}^2})$

- **Matern 5/2 GPR**

  $k\ (xi, xj|\theta)\frac{\sqrt{3r}}{\sigma_l})exp(-\frac{\sqrt{3r}}{\sigma_l})$

- **Exponential GPR**

  $k\ (xi, xj|\theta)$ =$\sigma_f{}^2$exp $\frac{r}{\sigma_l}$), where: $\sqrt{(x_i-x_j)^T(x_i-x_j)}$

- **Quadratic Support Vector Machine (Quad-SVM)**

  K(x,y) = $(x\cdot y+c)^2$, where c is a constant parameter.

Here x_i is the input vector of the predictors, the maximum posterior estimate corresponds to θ, and the standard deviation of the signal is denoted as σ_f. Alpha represents the non-negative parameter governing covariance (Lin et al. 1979, Rodgers and Nicewander 1988, Kuh and Welsch 2005, Durgesh and Lekha 2010, Weerapura and Abeynayake 2013, Zhang et al. 2018, Sathyadeepa and Sivakumar 2021).

Evaluation of models performance

The best-fitted model was selected by analyzing R-square and Root Mean Square Error (RMSE) values. Higher R-square values and the lower RMSE specify the best-fitted model to forecast tea production and consumption. Therefore, different graphical approaches were used in this work to test those two critical assumptions. A predicted plot against actual values is useful and widely practiced to check the normality of error. The most useful residual plot is a graph that shows the difference between the predicted values and the actual values. It is used to check if a linear model is suitable for the data and if the residuals have constant variance (Zhang et al. 2018).

Result and Discussion

Once the regression model has been trained in the response plot, actual vs. prediction plot and residual plot are generated that is shown as a graphical expression against crop year, rain (mm), temperature (^{o}C), and actual production (Mkg) to present tea production. On the other hand, tea consumption has been trained against a population that shows the correlation between them. Through holdout or cross-validation, the network is trained with 80% of the data points and tested with the remaining 20%. So, to speak, these projections are predicted on the holdout observations result. The studied results were divided into tea production and consumption.

Tea Production Modelling

In this section, we have used four effective models namely Rational quadratic, Exponential, Squared Exponential, and Matern 5/2 GPR for tea production in Bangladesh. The depicted response plot in Fig. 1 vividly showcases the accuracy of our predictive model, as evidenced by the close correspondence of the predicted values to the actual values, signifying negligible prediction errors.

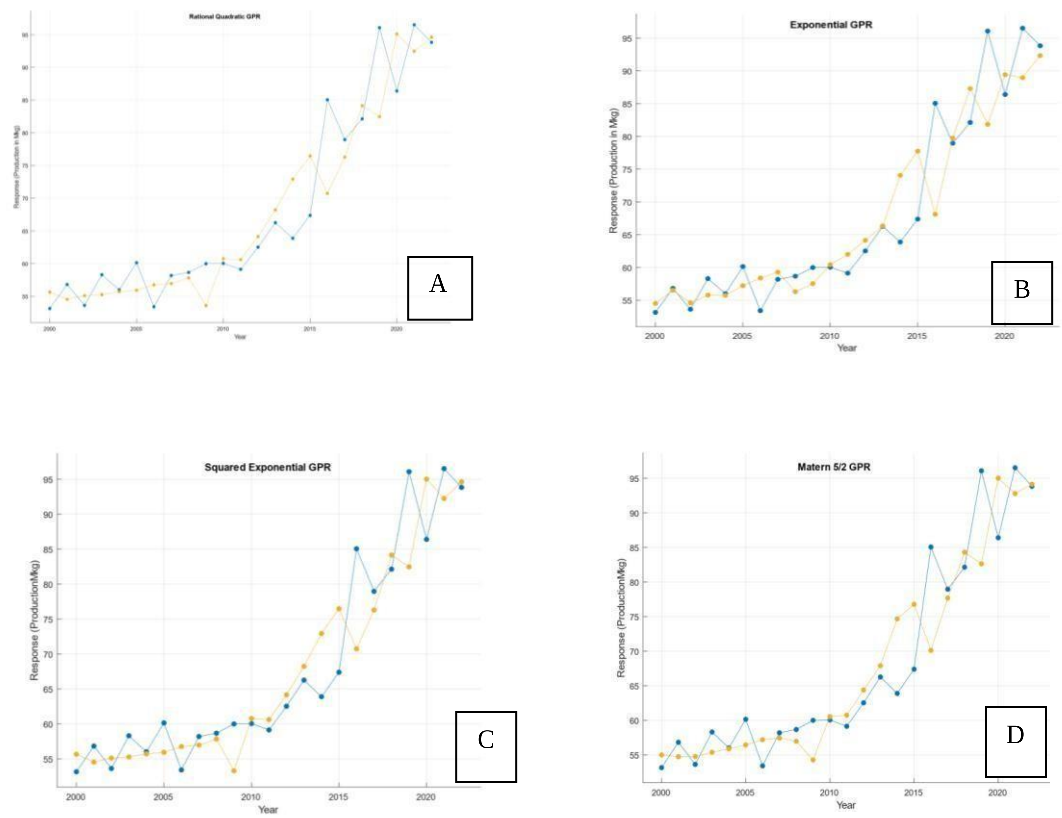


Fig. 1: The response plot of (a) Rational quadratic GPR, (b) Exponential GPR, (c) Squared exponential GPR, (d) Matern 5/2 GPR.

This alignment underscores the model's capacity to consistently and dependably approximate real-world data, instilling confidence in its reliability. Furthermore, it implies that the model's predictions adhere to a coherent and consistent pattern closely resembling the true data, negating the suggestion of randomness or sporadic behavior. In essence, this plot serves as compelling evidence for the model's proficiency in generating precise predictions that closely mirror actual values, establishing its credibility and trustworthiness in facilitating data-informed decision-making across diverse domains, which is compatible with other previous studie (Weerapura and Abeynayake 2013).

Fig. 2 provides a visual representation of how our predictive model's projected responses compare to the actual observed responses.

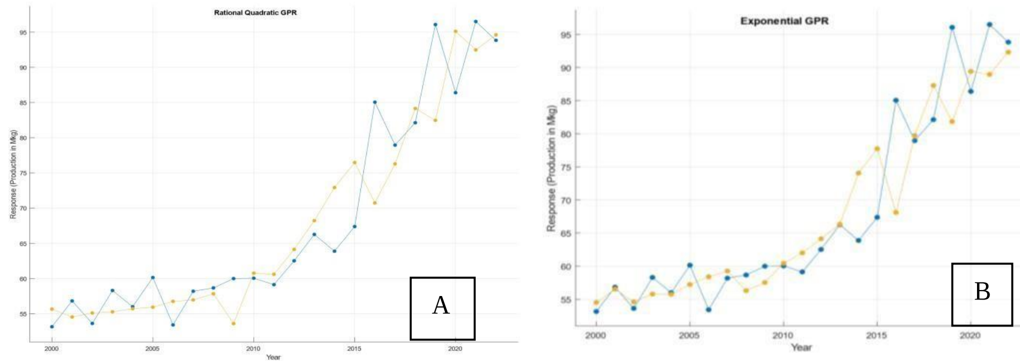
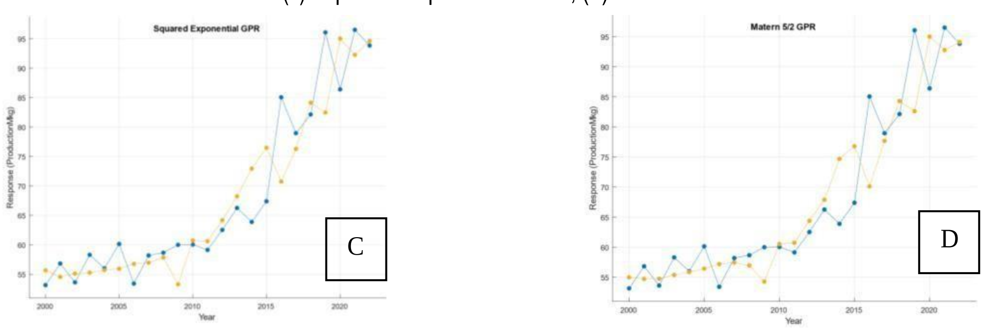


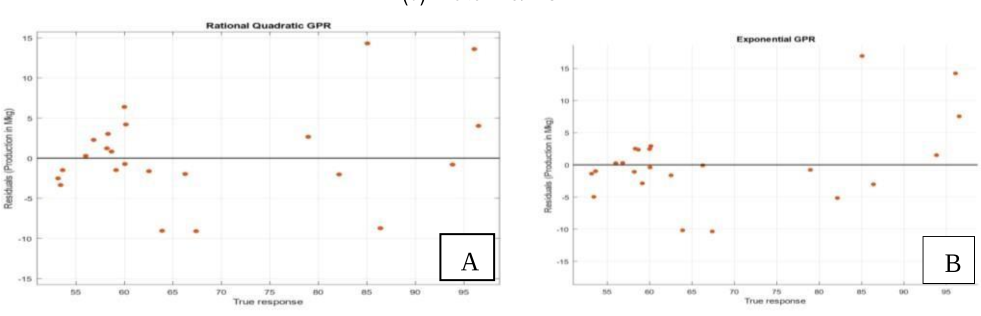
Fig. 2: The Predicted vs. Actual response plot (a) Rational Quadratic GPR, (b) Exponential GPR, (c) Squared Exponential GPR, (d) Matern 5/2 GPR.



The alignment of all data points along a diagonal line signifies the ideal scenario where the predicted responses match the actual responses. In this case, there exists only a marginal prediction error. The vertical separation of the points from the diagonal line signifies the error margin. This model suggests that though the model fits well, there remain few deviations from the prediction line.

After the model has been trained, this study used the residuals plot to further assess its performance. The difference between the predicted and actual observations is shown in the residual graphic. This scatter plot shows the error which is useful to select fitted model forecasting. The residuals were not symmetrically distributed around 0, clearly a nonlinear pattern, which is supported by another report (Zhang et al. 2018). The residuals plot of Rational Quadratic GPR in Fig. 3 shows the most significant relationship with 0 and less error than other models.

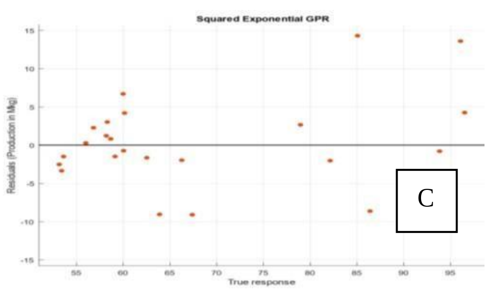
Fig. 3: The Residuals plot (a) Rational quadratic GPR, (b) Exponential GPR, (c) Squared Exponential GPR, (d) Matern 5/2 GPR.

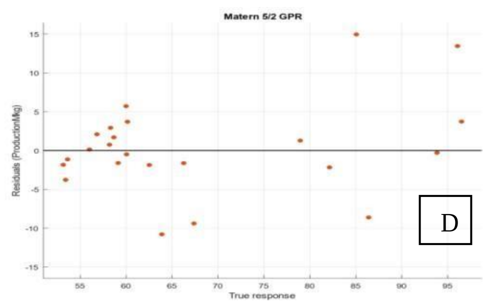


Model parameters are also useful for selecting the best fitted model for tea production which is presented in Table 1:

Table 1. Model summary of tea production.

Model type	RMSE	MSE	R-Squared	MAE
Rational Quadratic GPR	5.7595	33.1729	0.8527	4.1599
Squared Exponential GPR	5.7760	33.3632	0.8519	4.1802
Exponential GPR	5.8765	34.5340	0.8467	4.0881
Matern 5/2 GPR	6.1229	37.4904	0.8336	4.0833





After analysis, this study shows that the RMSE value is the lowest and the R-Square value the highest close to 1 of the Rational Quadratic GPR model to compare the other three models which is shown in Table 1. The forecasted result for all models is shown below in Table 2.

Table 2. Forecasted result of tea production in million kg.

Year	Actual	Matern 5/2 GPR	Exponential GPR	Squared exponential GPR	Rational quadratic GPR
2018	82.13	84.29	87.29	84.15	84.15
2019	96.07	82.61	81.83	82.46	82.46
2020	86.39	95.00	83.42	95.01	95.11
2021	96.51	92.77	88.96	92.24	92.46
2022	93.83	94.12	92.32	94.61	94.61
2023	-	95.41	92.97	99.18	95.02
2024	-	97.22	92.48	94.05	96.24
2025	-	94.96	91.31	92.59	91.27

Tea consumption modelling

In this section, we have used four effective models namely rational quadratic, exponential, squared exponential, and matern 5/2 GPR for tea production in Bangladesh.

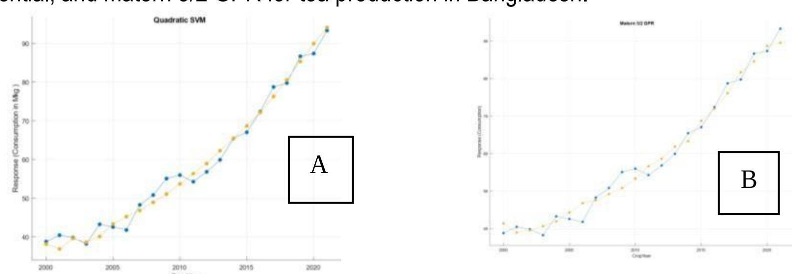


Fig. 4: The Response plot, (a) Quadratic SVM, and (b) Matern 5/2 GPR.

The response plot, which is displayed in Fig. 4 anticipates response against consumption year and population and can be used to present the results. This study shows that the prediction line and actual consumption line follow a similar course, and both are on an upward trend, probably due to population growth.

Fig. 5 shows model's predictions closely match with real data lines. This means that the model is reliable at making accurate predictions that are similar to the true values.

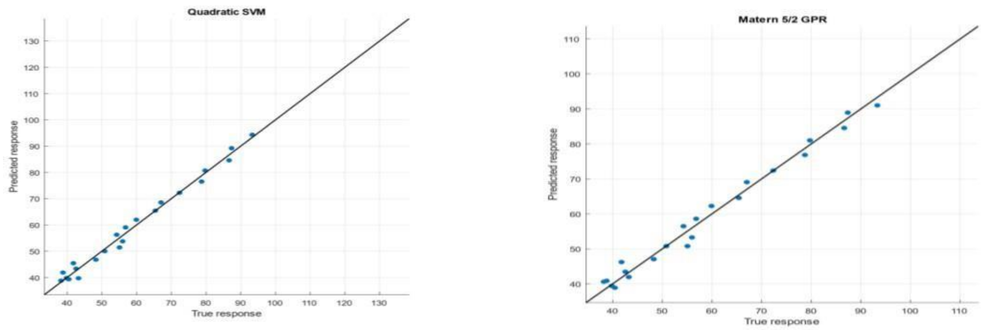


Fig. 5: The Predicted vs. Actual response plot, (a) Quadratic SVM, and (b) Matern 5/2 GPR.

The scattered points around the straight line represent the variance between actual data points and the predicted linear relationship, with the straight line itself indicating both the degree of correlation and the extent of deviation from the ideal model for a specific input. Both model shows a high degree of correlation and minimal deviation from the response.

The residuals plot was used to find the difference between the predicted and actual reactions is shown in Fig. 6.

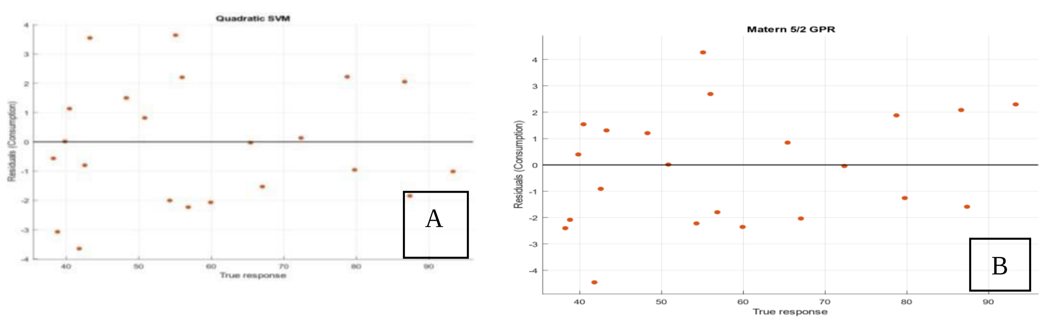


Fig. 6: The Residual plot of (a) Quadratic SVM, (b) Matern 5/2 GPR.

The residuals are not symmetrically distributed around 0, clearly a nonlinear pattern, which is compatible with another report (Zhang et al. 2018, Sathyadeepa and Sivakumar 2021).

Table 3. Model statistics of tea consumption.

Model type	RMSE	MSE	R-Squared	MAE
Quadratic SVM	2.0817	4.3336	0.9868	1.7380
Matern 5/2 GPR	2.3706	5.6198	0.9829	2.0949

Once the network has undergone thorough training, it is essential to assess the performance of each feature subset for every model employed in this research study. Table 3 provides a comprehensive comparison. After model analysis, this study shows that the RMSE value is the lowest and R Square value the highest close to 1 of Quadratic SVM than the Matern 5/2 model. The forecasted result is shown below in Table 4.

Table 4. Forecasted value of consumption.

Crop Year	Actual	Quadratic SVM	Matern 5/2 GPR
2018	79.74	80.69	80.99
2019	86.64	84.58	84.55
2020	87.35	89.19	88.93
2021	93.30	94.31	91.00
2022	-	97.48	94.92
2023	-	104.06	99.53
2024	-	108.88	96.70
2025	-	113.96	105.30

According to research, there is a noticeable and steady upward trend in tea consumption, with the projected figures indicating that by the year 2025, global tea consumption will reach a substantial 105.307 million kilograms. However, it's noteworthy that the Matern 5/2 model, employed in this study, did not demonstrate a more favorable trend concerning population growth. This suggests that population growth alone may not be the sole driving force behind the increase in tea consumption. Other factors, such as changing consumer preferences, marketing strategies, and health trends, may also contribute to this rise in tea consumption.

This study's choice of predictive models, specifically the Rational Quadratic GPR for tea production and the Quadratic Support Vector Machine for consumption forecasting, underscores the importance of using advanced machine learning techniques to make accurate predictions in the tea. The evaluation criteria, which include RMSE and R-squared values, confirm the advantageous eminence of these models in making these predictive decisions.

The visual representation of the results in Fig. 7 highlights the likelihood of future fluctuations in both tea production and consumption. This insight is crucial for decision-makers within the tea industry, as it allows them to anticipate and prepare for potential changes in market production and consumption. It also emphasizes the value of data-driven decision-making in maintaining a competitive edge in the tea market.

In conclusion, the findings of our study provide valuable insight into our understanding of tea consumption and production trends. As the world's taste for tea continues to evolve, these insights will be instrumental in guiding industry stakeholders toward informed and strategic choices.

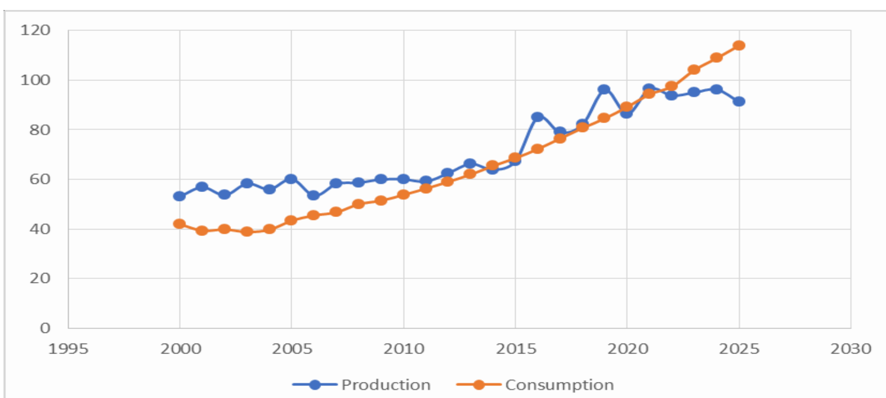


Fig. 7: The relation between forecasted production and consumption.

The graph in Fig. 7 demonstrates the link between the projected value of tea production and consumption in Bangladesh. The intersection of production and consumption in 2022 highlights a significant challenge facing the Bangladesh tea industry, namely the production-consumption gap. The high correlation coefficient of 0.94 between consumption and production indicates that there is a strong positive relationship between these two variables. The tea consumption rate is 4.06% against the tea production rate of 1.03%. To address the steadily increasing consumption rate for tea in Bangladesh, it is crucial to focus on developing effective strategic measures urgently.

Conclusion

This study highlights the effectiveness of the Rational Quadratic GPR and Quadratic SVM models in predicting tea production and consumption in Bangladesh. These models demonstrated statistical robustness and generated accurate forecasts for both tea production and consumption. In the year 2022, tea production reached 93.83 Mkg, while consumption concurrently stood at 97.48 Mkg, resulting in a gap of 3.65 Mkg. Looking ahead, the forecast indicates that production will increase in the following year, but consumption is anticipated to surpass production. To bridge this gap, potential strategies could involve importing tea or enhancing productivity within the domestic tea industry.

Acknowledgement

Sincere thanks to the supervisor Prof. Dr. Iftekhar Ahmad and Co-supervisor Mahabub Alam, Department of Food Engineering and Tea Technology, Shahjalal University of Science and Technology for their overall support.

Conflict of interest

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

References

Ahmad I and Mohammad AH (2013). Present status of Bangladesh tea sector in respect of world tea: an overview, Journal of Applied Science and Technology 9(1):116-28.

Arefin, Md R and Md IH (2022). Present Status and Future Prospects of Tea production and Research on Varietal Improvement in Bangladesh, Turkish Journal of Agriculture-Food Science and Technology 10(12): 2324-33.

Belsley, David A, Edwin K and Roy EW (2005). Regression diagnostics: Identifying influential data and sources of collinearity 2nd Ed., John Wiley and Sons, Inc., pp. 1-321.

Keya D, N Sultana, Provash KK and MN Hasan (2022). A comparison of trend models for predicting tea production in Bangladesh, Nepalese Journal of Statistics 6(1): 51-62.

Durgesh, Srivastava and Lekha B (2010). Data classification using support vector machine, Journal of Theoretical and Applied Information Technology 12(1): 1-7.

Gilgen DD, Nick MT and Lyliane R (2001). Intestinal helminth infections, anaemia and labour productivity of female tea pluckers in Bangladesh. Tropical Medicine and International Health 6(6): 449-57.

Hossain MI, M Ahmed, Mamun MSA and Golam S (2012). Yield and quality performances of four tea test clones, Journal of the Bangladesh Agricultural University 10(1):33-38.

Lee R, Joseph and WA Nicewander (1988). Thirteen ways to look at the correlation coefficient, The American Statistician 42(1): 59-66.

Lin JC, Joseph K, Martin K and Paul BW (1979). Microwave apexcardiography, IEEE Transactions on Microwave Theory and Techniques 27(6): 618-20.

Mamun MSA and M Ahmed (2011). Integrated pest management in tea: prospects and future strategies in Bangladesh, The Journal of Plant Protection Sciences 3(2): 1-13.

Saha JK, KMM Adnan, Swati AS and Shefali B (2021). Analysis of growth trends in area, production and yield of tea in Bangladesh, Journal of Agriculture and Food Research 4: 100136.

Sathyadeepa S and T Sivakumar (2021). Optimal design of quadratic support vector machine based end stage renal disease predictor model, In: Journal of Physics: Conference Series, January 2020. IOP Publishing.

Sultana J, MNA Siddique, Md Kamaruzzaman and MA Halim (2014). Conventional to ecological: Tea plantation soil management in Panchagarh District of Bangladesh, Journal of Science, Technology & Environment Informatics 1(1): 27–37.

Weerapura WHEBP and Abeynayake NR (2013). Forecasting of tea production using time series models. Proceedings of 12th Agricultural Research Symposium (2013): 351-355.

Zhang N, Jiang X, Jing Z and Keenan L (2018). Gaussian process regression method for classification for high-dimensional data with limited samples. 8th Eighth International Conference on Information Science and Technology, June 30-6 July 2018, IEEE, Granda, Cordoba.

(Manuscript received on 14th July 2023 and revised on 8th October 2023)