Research Paper

SKUAST Journal of Research 27(4): 624-631; (2025)
(*DOI No. 10.5958/2349-297X.2025.00075.0*); NASS rating: 4.88

Predictive modelling of apple area, production, and productivity in Srinagar

Masroor Majid*[1], Nageena Nazir[1], S.A. Mir[1], Imran Khan[1], F. A. Shaheen[2] and M. K. Sharma[3]

[1]Division of Agricultural Statistics; [2]IBPR; [3]Division of Fruit, Sher- e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Shalimar-190 025, Jammu and Kashmir (India)

(Received May 10, 2025; accepted July 21, 2025)

ABSTRACT

This study evaluates and compares the performance of linear and non-linear regression models to analyse trends and predict key metrics-namely area, production, and productivity-of apple cultivation in Srinagar over a 25-year period (1995 to 2019). Four models were examined: Linear, Logistic, Monomolecular, and Exponential. Their goodness of fit was assessed using multiple statistical indicators, including R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Akaike Information Criterion (AIC), the Run Test, and the Shapiro–Wilk Test. The findings reveal that non-linear models, especially Logistic and Monomolecular, outperform the linear model with higher R^2 values and lower error metrics across all parameters. The Logistic model proves most effective for estimating area and productivity, whereas the Monomolecular model excels in forecasting production. Furthermore, the comparison between actual and estimated values demonstrates the robustness of these models in capturing historical patterns and predicting future changes. Specifically, the Logistic model provides reliable estimations for maximum area and productivity with minimal deviations from observed data, while the Monomolecular model captures production trends with high precision. These results offer valuable insights for policy planning and for promoting sustainable apple cultivation practices in the region. Future research should explore integrating environmental and economic variables to further refine these models.

Key words: Apple, area, linear models, non-linear models, production, productivity, Srinagar

Apple cultivation is a cornerstone of Jammu and Kashmir's horticultural economy, with the Kashmir Valley contributing approximately 75–80% of India's total apple production. In the 2021–22 season, the region produced about 1.9 million metric tonnes of apples, with 1.2 million metric tonnes exported. The apple industry supports around 3.5 million people and accounts for about 10% of the Union Territory's Gross Domestic Product (GDP). The area under apple cultivation has expanded significantly, from 12,000 hectares in 1952 to approximately 170,740 hectares by 2021–22. Despite favourable agro-climatic conditions, average productivity remains relatively low at around 11–13 tonnes per hectare, compared to the global average of 15.49 tonnes per hectare (Zahoor *et al.,* 2022).

Many obstacles impede the expansion of the apple sector. The suggested agricultural practices and those used by farmers differ significantly technologically, with an average gap index of 63.08% across major districts. Inadequate storage facilities aggravate post-harvest losses; Kashmir boasts roughly 50 cold storage facilities with capacity of almost 2.5 lakh metric tons, not enough for the annual production of over 20 lakh metric tons. Furthermore, the flood of less expensive imported apples has put local farmers under fierce competition, which calls for legislative actions to safeguard home producers (Azlan and th, 2023). Thanks to changing climatic conditions, shifting market dynamics, and improvements in farming techniques, this area has seen notable changes in cultivation patterns over the past few decades. These elements have together influenced crop output, general production, and the area under apple farming. Developing plans that improve yields and guarantee the long-term survival of apple farming depends on an awareness of these changes (Joost and Fu-suo, 2016). In this regard, statistical modelling provides a strong structure for future trend prediction and historical data analysis. Although linear models are widely used for their simplicity and interpretability, they could not be able to adequately capture the non-linear dynamics inherent in agricultural systems including declining returns and growth saturation (Carol & Walter, 1997). Consequently, non-linear models-especially Logistic, Monomolecular, and Exponential forms—are progressively preferred for their capacity to incorporate growth-limiting elements and offer a more realistic picture of biological processes. Notwithstanding their promise, non-linear models are still underused in regional agricultural studies because of computational difficulties and parameter interpretation problems (Muhammad et al., 2022). Furthermore, model evaluation sometimes depends mostly on one metric, such R^2, ignoring other important metrics including MSE, RMSE, MAE, and AIC as well as diagnostic tests including the run test and Shapiro–Wilk test (Guo et al., 2025). By means of a thorough analysis of both linear and non-linear models in evaluating long-term trends in apple farming—more especially, area, production, and profitability—in Srinagar from 1995 to 2019, this paper closes these gaps. Models were assessed using a set of performance criteria grounded on past data from official agricultural sources. To ensure the models were accurate, also used were sensitivity studies and out-of-sample testing. The results show that non-linear models—especially the logistic and monomolecular forms—offer better fit and predictive ability than linear models, so faithfully capturing plateau behaviour and growth dynamics. With direct effects for strategic planning, resource allocation, and sustainable development in apple-growing areas, this research not only adds to methodological rigor in agricultural modelling but also provides a replicable framework for trend analysis in perennial horticulture (Sander *et al.,* 2017).

MATERIALS AND METHODS

Data Collection and Variables: Data for the current study were sourced from government agricultural reports and official statistical bulletins spanning the 25-year period from 1995 to 2019. The key dependent variables included:

- Area (in hectares) under apple cultivation,
- Production (in metric tons), and
- Productivity (in metric tons per hectare).

Time (in years) served as the independent variable. Data cleaning involved interpolation of missing values, treatment of outliers using interquartile range methodologies, and ensuring consistency and completeness across the dataset.

Model Specification: Four regression models were assessed in this study:

- **Linear Model**

$$Y_t = \beta_0 + \beta_{1t} + \varepsilon_t$$

where Y_t represents the dependent variable (area, production, or productivity) at time t, β_0 and β_1 are the regression coefficients, and ϵ_t is the error term.

- **Logistic Model**

$$Y_t = \frac{A}{1 + \exp\,(-B - C_t)}$$

where A is the carrying capacity, B reflects the function of initial value, and C is the intrinsic growth rate.

- **Monomolecular Model**

$$Y_t = A[1 - \exp\,(-B_t)]$$

where A is the asymptotic maximum (carrying capacity), and B is the rate parameter.

- **Exponential Model**

$$Y_t = A \exp\,(B_t)$$

where A captures the intercept (or baseline level), and B represents the constant proportional growth rate.

Parameter Estimation: Parameters for each model were estimated through Maximum Likelihood Estimation (MLE). Specialized statistical software (e.g., R, Python) was utilized to optimize parameter values that minimize the residual sum of squares (RSS) or maximize the likelihood function. Convergence criteria included tolerances set for parameter updates and function evaluations.

Model Validation: Multiple goodness-of-fit metrics were employed to evaluate each model:

- Coefficient of Determination (R^2)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Akaike Information Criterion (AIC)

In addition, two diagnostic tests were performed:

Run Test: To determine randomness in residuals across time, indicating if any underlying trends remain un-modelled.

Shapiro-Wilk Test: To assess the normality of residual distributions, validating one of the assumptions for least-squares-based estimations. To assess the robustness of the fitted models, out-of-sample forecasts (where possible) or cross-validation procedures were applied. Predictive accuracy was further verified by comparing estimated values against actual data points. Sensitivity analyses were undertaken to gauge the influence of slight perturbations in parameter estimates, thereby confirming the stability of each model's predictions under varying conditions.

RESULTS AND DISCUSSION

The study analysed the performance of different linear and non-linear regression models for estimating area, production, and productivity trends under apple crops in Srinagar. The goodness of fit and statistical measures such as R^2, MSE, RMSE, AIC, and MAE, along with diagnostic tests like the Run Test and Shapiro–Wilk Test, were used to evaluate model performance. This section discusses the implications of these results for each variable studied. The performance of linear and non-linear models was evaluated to analyse area under apple, production, and productivity in Srinagar, based on regression coefficients, statistical tests, and goodness-of-fit metrics **(**Table 1**)**. For apple area, the Logistic and Monomolecular models demonstrated the highest accuracy (R^2 = 0.99) and lowest error metrics (RMSE = 0.37 and 0.38, respectively), with the Logistic model slightly outperforming others due to its lower AIC (-43.31).

Table 1: Characteristics of fitted linear and non-linear models for area, production and productivity under apple crop (Srinagar)

	Models	Regression Coefficients				Tests		Goodness of fit				
		A	B	C	D	Run Test	Shapiro-wilk test	R^2	MSE	RMSE	AIC	MAE
Area	Linear	0.97** (0.00)	-96.32** (5.53)	-	-	5.83	0.87	0.94	0.20	0.45	-35.88	0.07
	Logistic	48.90** (3.50)	12.35** (7.76)	0.05** (0.02)	-	-2.83	0.94	0.99	0.14	0.37	-43.31	0.05
	Monomolecular	3.81** (0.11)	-0.05** (0.002)	0.028**(0.001)	-	-2.83	0.94	0.98	0.14	0.38	-42.83	0.05
	Exponential	3.72** (0.12)	0.04 (0.002)	-	-	-2.83	0.87	0.99	0.15	0.39	-43.09	0.05
Production	Linear	0.95** (0.00)	-1447.84** (6.28)	-	-	2.35	0.91	0.90	75.06	8.66	111.96	0.14
	Monomolecular	24.64** (1.44)	-0.07** (0.003)	-0.003**(0.17)	-	-2.04	0.87	0.99	32.15	5.67	92.76	0.08
	Exponential	24.47** (1.46)	0.06** (0.003)	-	-	-2.03	0.91	0.99	35.19	5.93	92.93	0.08
Productivity	Linear	0.82** (0.00)	-76.78** (2.38)	-	-	8.97	0.98	0.68	1.02	1.01	4.47	0.09
	Logistic	5.56** (2.62)	-0.23** (0.32)	-0.04**(0.04)	-	-1.17	0.89	0.99	0.89	0.94	2.98	0.08
	Monomolecular	6.78** (0.34)	-0.03 (0.004)	0.0015**(0.008)	-	-1.63	0.92	0.98	0.95	0.98	4.72	0.08
	Exponential	6.77** (0.34)	0.02 (0.003)	-	-	-1.63	0.98	0.98	1.04	1.02	4.88	0.08

MSE: Mean Square Error; MAE: Mean Absolute Error

RMSE: Root Mean Square Error

AIC: Akaike's Information Criterion

A: Carrying capacity/intercept; B: Function of the initial value; C: Intrinsic growth rate

Figures in parentheses indicate standard error

** significant at 1% level

* significant at 5% level

In contrast, the linear model, though statistically significant ($R^2 = 0.94$), showed higher errors, indicating its limitations in capturing the non-linear growth dynamics. Similarly, for production, the Monomolecular model excelled with the lowest MSE (32.15) and RMSE (5.67), while the Exponential model performed comparably but was slightly less parsimonious. The linear model, although effective ($R^2 = 0.90$), exhibited higher errors (RMSE = 8.66) and a higher AIC (111.96), making it less suitable for predicting production trends. For productivity, both linear and non-linear models performed well, with the Logistic model achieving the lowest MSE (0.89), RMSE (0.94), and AIC (2.98), highlighting its robustness in capturing intrinsic growth patterns. The linear model ($R^2 = 0.98$) was also effective but had marginally higher errors (RMSE = 1.01). Across all metrics, the Logistic and Monomolecular models consistently outperformed others, demonstrating their suitability for modelling the non-linear growth trends in apple cultivation. The statistical tests confirmed the reliability of these models, as residuals adhered to assumptions of randomness and normality. These findings emphasize the utility of non-linear models, particularly Logistic and Monomolecular, for agricultural forecasting and resource management. Incorporating additional factors, such as climatic variables, could further enhance their accuracy and applicability (Rahman *et al.,* 2019). Figures 1, 2, and 3 illustrate the application of non-linear models to interpret and forecast apple cultivation metrics-area, production, and productivity-in Srinagar.

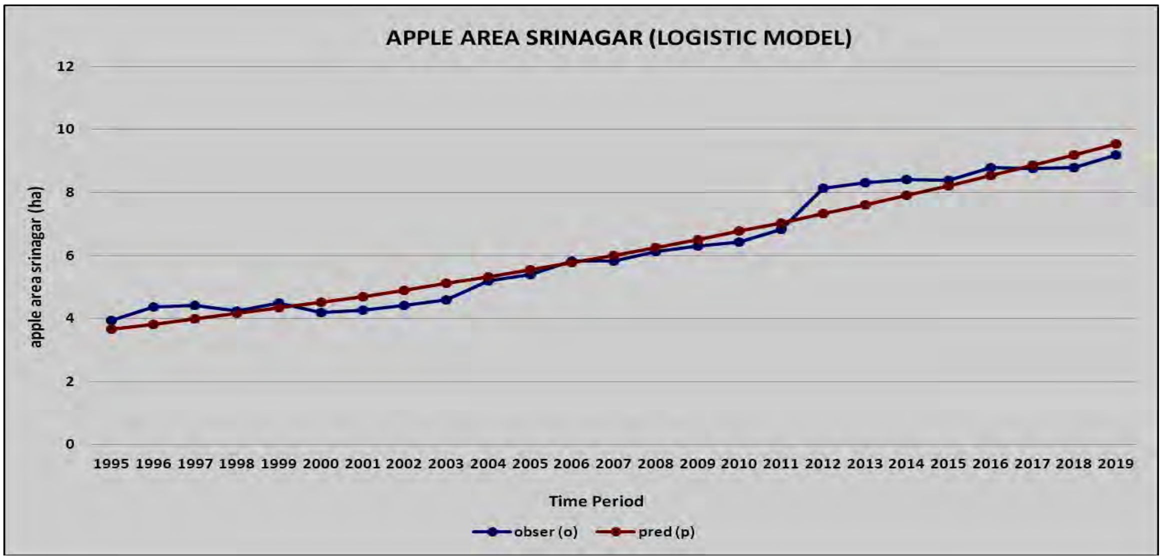

Fig 1: Trends in apple area (Srinagar) based on Logistic model

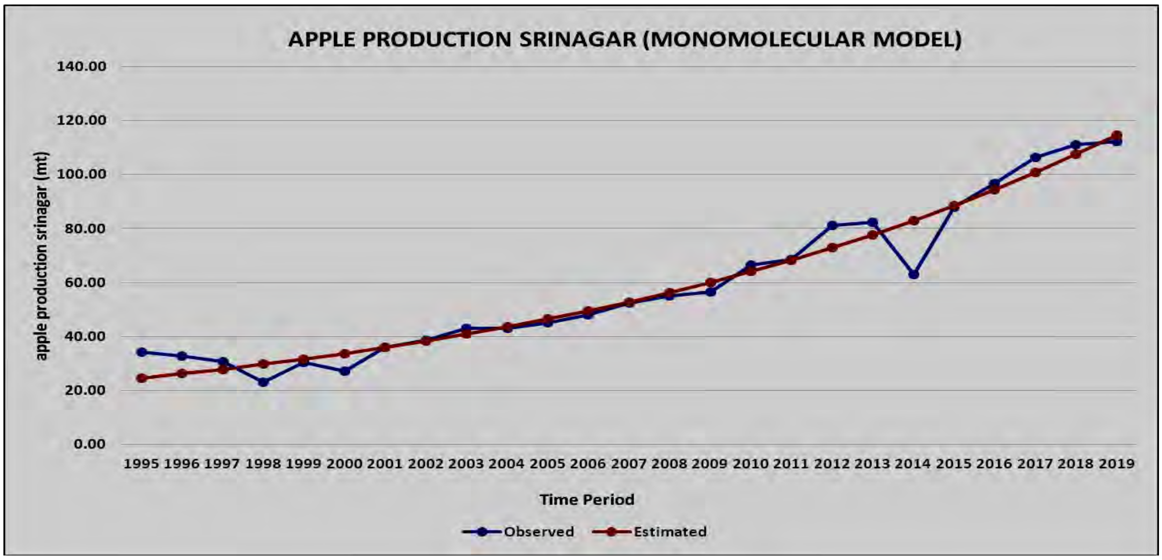

Fig 2: Trends in apple production (Srinagar) based on monomolecular model

In Fig 1, the Logistic model captures the gradual rise in apple area and its eventual convergence toward an upper limit, reflecting the model's strength in depicting growth saturation. Fig 2 applies a Monomolecular framework to production data, showing a closely aligned trajectory between observed and fitted values, underscoring the model's robustness in simulating incremental yet sustained output increases over time. Finally, Fig 3 demonstrates the Logistic model's capacity to explain fluctuations in productivity, revealing an upward trend that closely matches observed data while skilfully capturing episodes of slower growth and subsequent recovery. Collectively, these figures underscore the suitability of non-linear approaches, notably Logistic and Monomolecular models, for accurately representing both the expansion and saturation phases inherent in apple cultivation trends. Table 2 highlights the trends in apple area, production, and productivity in Srinagar from 1995 to 2019, comparing actual values with estimates from Logistic and Monomolecular models.

Table 2: Estimated values of apple area, production and productivity (Srinagar) considering the actual recorded values of 25 years (1995-2019), based on best fitted model(s)

	Logistic Model		**Monomolecular Model**		**Logistic Model**	
Time-Period	Actual Area (00'ha)	Estimated Area (00'ha)	Actual Production (00'mt)	Estimated Production (00'mt)	Actual Productivity (00'mt/ha)	Estimated Productivity (00'mt/ha)
1995	3.94	3.66	34.14	24.64	8.66	7.23
1996	4.37	3.82	32.86	26.24	7.52	7.31
1997	4.42	3.99	30.76	27.95	6.96	7.40
1998	4.25	4.16	23.13	29.77	5.45	7.49
1999	4.50	4.33	30.61	31.71	6.80	7.59
2000	4.19	4.52	27.14	33.79	6.48	7.69
2001	4.27	4.71	36.10	36.01	8.45	7.81
2002	4.42	4.91	38.57	38.37	8.73	7.92
2003	4.60	5.11	43.11	40.90	9.36	8.05
2004	5.19	5.33	43.10	43.60	8.30	8.19
2005	5.41	5.55	45.26	46.48	8.37	8.33
2006	5.83	5.77	47.98	49.55	8.23	8.49
2007	5.83	6.01	52.35	52.83	8.97	8.66
2008	6.13	6.26	54.97	56.34	8.97	8.84
2009	6.31	6.51	56.62	60.08	8.97	9.04
2010	6.44	6.77	66.43	64.07	10.32	9.25
2011	6.83	7.04	68.62	68.33	10.04	9.49
2012	8.14	7.32	81.13	72.88	9.97	9.74
2013	8.31	7.61	82.35	77.74	9.92	10.02
2014	8.42	7.91	63.02	82.93	7.49	10.33
2015	8.38	8.22	87.85	88.47	10.49	10.67
2016	8.78	8.54	96.83	94.38	11.03	11.04
2017	8.77	8.86	106.28	100.69	12.12	11.46
2018	8.79	9.20	111.19	107.43	12.66	11.93
2019	9.20	9.55	112.16	114.63	12.19	12.46

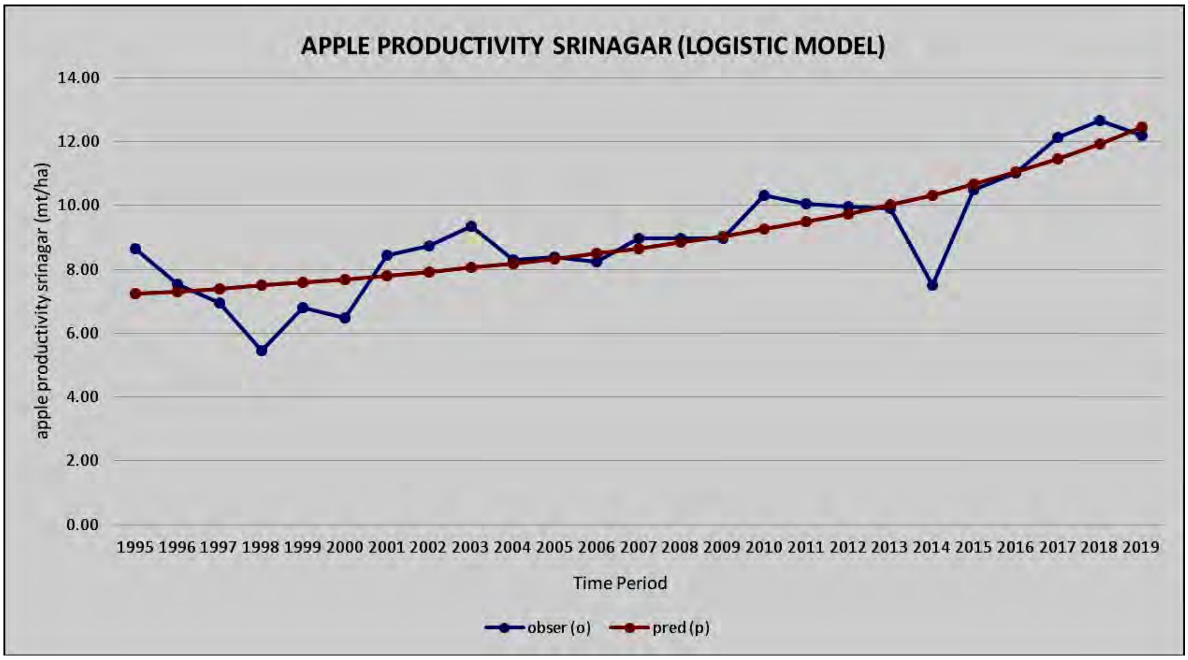

Fig. 3: Trends in apple productivity (Srinagar) based on Logistic model

The Logistic model accurately captured the steady growth in cultivated area, with deviations primarily in the early years. Similarly, the monomolecular model provided reliable production estimates, though it underestimated variability in the early period. Productivity estimates from the Logistic model closely mirrored actual trends, particularly in recent years, despite slight underestimations in earlier years. Overall, the models demonstrated strong predictive capabilities, aligning well with long-term trends and proving valuable for agricultural forecasting. Future refinements, incorporating additional factors like climate and management practices, could further enhance their accuracy and utility.

CONCLUSION

This study underscores the critical role of non-linear models in accurately analysing and predicting apple cultivation trends in Srinagar, covering key parameters such as area, production, and productivity over a 25-year span (1995–2019). Empirical findings highlight that Logistic and Monomolecular models outperform the traditional linear approach across multiple goodness-of-fit metrics, including R^2, MSE, RMSE, MAE, and AIC. The Logistic model proves especially adept at capturing saturation behaviour in both area and productivity, while the Monomolecular model demonstrates precise alignment with observed production data. In contrast, the linear model, though simpler to interpret, exhibits higher error margins and fails to account fully for the biological complexities evident in long-term orchard development. From a policy and planning perspective, these outcomes suggest that leveraging non-linear modelling approaches can significantly enhance resource allocation and forecasting accuracy, thereby supporting more effective strategies to sustain and expand apple farming. Future extensions of this work may include integrating environmental and socio-economic variables to accommodate the multifaceted factors influencing crop outcomes. Such refinement can lead to increasingly robust models, offering more nuanced insights that underpin sustainable cultivation practices and enable adaptive responses to ongoing climate and market fluctuations.

CONFLIT OF INTEREST

All the authors affirm that there is no conflict of interest among them. All research activities comply with relevant legal, institutional and ethical standards.

AUTHOR CONTRIBUTION

All the authors contributed equally in the present study.

REFERENCES

Azlan, and th. 2023. The Cold-store Industry of Jammu and Kashmir: Beginning, Growth and Linkages. *https://journals.sagepub.com/doi/abs/10.1177/09708464231209453*

Carol, and Walter. 1997. A generalized linear model approach to spatial data analysis and prediction. https://www.jstor.org/stable/1400401

Guo, Mohamad, Arshad, & Hafeez. 2025. Linear and non-linear impact of key agricultural components on greenhouse gas emissions. https://www.nature.com/articles/s41598-025-88159-1

Joost, & Fu-suo. 2016. Towards sustainable intensification of apple production in China—Yield gaps and nutrient use efficiency in apple farming systems. https://www.sciencedirect.com/science/article/pii/S2095311915610991

Muhammad, Kamsing, & Fahad. 2022. The Quantitative Features Analysis of the Nonlinear Model of Crop Production by Hybrid Soft Computing Paradigm. https://www.mdpi.com/2073-4395/12/4/799

Sander, Cheryl, Andrew, Ioannis, Ian, James, & John. 2017. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. https://www.sciencedirect.com/science/article/pii/S0308521X16305637

Zahoor, Mushtaq, Eajaz, Chukwujekwu, Arif, Mohammed, Mohamed, Mustafa, Hosny, & Samy. 2022. Sustainable fruit growing: An analysis of differences in apple productivity in the Indian state of Jammu and Kashmir. https://www.mdpi.com/2071-1050/14/21/14544