



# Forecasting Production of Potato for a Sustainable Future: Global Market Analysis

Pradeep Mishra<sup>1</sup>  · Amel Ali Alhussan<sup>2</sup> · Doaa Sami Khafaga<sup>2</sup> · Priyanka Lal<sup>3</sup> · Soumik Ray<sup>4</sup> · Mostafa Abotaleb<sup>5</sup> · Khder Alakkari<sup>6</sup> · Marwa M. Eid<sup>7</sup> · El-Sayed M. El-kenawy<sup>8</sup>



Received: 17 February 2024 / Accepted: 6 March 2024

© The Author(s), under exclusive licence to European Association for Potato Research 2024

## Abstract

This study goes into the essential challenge of estimating potato output in order to ensure sustainable agricultural practices while also providing vital insights into global market patterns. The potato production data series compares the accuracy of two popular forecasting models, ARIMA (AutoRegressive Integrated Moving Average) and ETS (Error-Trend-Seasonality), in predicting potato production. The study assesses the efficacy of these models with a particular focus on their relevance to the agricultural markets of India, China, and the USA, three major potato-producing countries. This research builds ARIMA and ETS models and thoroughly assesses their forecasting performance using historical production data series from these important nations. The results show that the ETS model, especially when considering the chosen countries, consistently performs better in predicting potato production for the testing data set than the ARIMA model. According to the models, China and India will keep contributing more to the potato market, solidifying their positions as key players. It is anticipated that the US economy will plateau and stabilize. For the anticipated year 2027, the expected potato output for China, India, and the USA is 100,417, 61,882, and 18,229 thousand tonnes, respectively. Nonetheless, the increasing diversity of confidence intervals in extended forecasts illustrates the intricacy of agricultural productivity and the numerous factors that could impact outcomes. We believe that this research significantly advances sustainable farming methods by offering a thorough analysis of worldwide potato production projections. It also improves our comprehension of the dynamics of the potato market, providing insightful information that can guide decision-making at different levels. In the conclusions, we stated that the studies not only have consequences for the potato sector, but they also highlight how crucial it is to use cutting-edge forecasting methods in order to promote sustainable food production and guarantee future food security.

**Keywords** ETS · Forecasting · Production · Potato · Sustainable agriculture

Extended author information available on the last page of the article

## Introduction

Potatoes are widely recognized as a comprehensive and essential dietary staple, offering not only crucial vitamins but also playing a pivotal role in global food security. Their extensive cultivation and consumption make them a significant dietary component worldwide. Given the challenges of a growing population, diminishing land availability, and evolving dietary trends, understanding the forecasting of potato production becomes imperative. This understanding is vital for ensuring the resilience and sustainability of food supply chains to meet changing demands. Originating from South America, the potato (*Solanum tuberosum* L.) is now cultivated in over 100 countries globally. The total global production of potatoes stands at 376 million tonnes, with China being the leading producer, contributing 94 million tonnes. The combination of China and India accounts for approximately one-third of the global potato production ([www.fao.org](http://www.fao.org)). The top three potato-producing nations, namely China, India, and the United States of America (USA), have a substantial impact on the global food supply chain, with fluctuations in their production and prices influencing the entire system (Wang et al. 2023). Hossain and Abdulla (2016) utilized potato production data over the period 1971 to 2013 and forecasted potato production in Bangladesh and found ARIMA (0,2,1) as the best fitted model.

The projected potato market size is estimated to be USD 115.74 billion in 2024, with expectations to reach USD 137.46 billion by 2029, showcasing a Compound Annual Growth Rate (CAGR) of 3.5% during the forecast period (2024–2029) (source: <https://www.mordorintelligence.com/industry-reports/potato-market/market-size>). Due to its inherent resistance to pests, diseases, and various climatic conditions, potato has become a dominant crop in developing countries, surpassing other food crops for enhanced food security (Zaheer and Akhtar 2016). The extended shelf life of potatoes has also increased their popularity in traditionally non-potato-consuming regions. Consequently, the rise in potato demand is intricately linked to increased potato production. However, while the potato market is expected to grow, questions arise regarding the sustainability of production growth at the same rate.

Recognizing this, there is a crucial need to comprehend the forecasting patterns of potato production. This study aims to offer insights into the global potato production market, utilizing forecasting techniques to predict trends and challenges. By thoroughly examining the production patterns, the research employs various forecasting techniques, with a particular focus on training data and utilizing ARIMA and ETS forecasting models. The objective is to analyze the future trajectory of potato production in key global players such as China, India, and the United States of America (Mishra et al. 2021a, 2023b, 2023c). The forecasted production analysis will delve into regional variations, identifying potential growth areas or vulnerabilities (Sahu et al. 2024).

Moreover, the study will investigate the influence of economic factors, trade dynamics, and emerging technologies on the potato production landscape (Mishra et al. 2023c). By offering a forward-looking perspective, this research aims to

be a valuable resource for policymakers. The goal is to assist them in navigating the intricacies of the global potato market and contribute to the development of strategies that foster a resilient and sustainable future for potato production on a global scale. Forecasts represent different levels of potato production, demand, and consumption trade in countries (Mishra et al. 2023b). Numerous studies have been conducted using different forecasting models to predict potato production in various regions (Singh and Deo 2015; Novkovic et al. 2015; Rahman et al. 2022). India is predicted to produce around 46,712 thousand metric tonnes (Yadav et al. 2024). This study adds to the existing body of knowledge by concentrating on key global players and employing advanced forecasting models for a comprehensive understanding of the future of potato production.

The current study aims to answer the following researchable questions:

- a. What are the global trends in the potato production in top producing countries?
- b. What is the estimated demand using different forecasting models?
- c. Which forecasting model gives best results?
- d. What specific variables have the most significant impact on refining forecasting models for potato production?
- e. What are the possible impacts of trends in production on food security?
- f. What should be strategies for economic impact, and marketing channels for potato industry?
- g. How can the integration of technology contribute to sustainable increases in potato production?
- h. How can forecasting models be adapted to better account for unforeseen events that may disrupt historical production patterns?

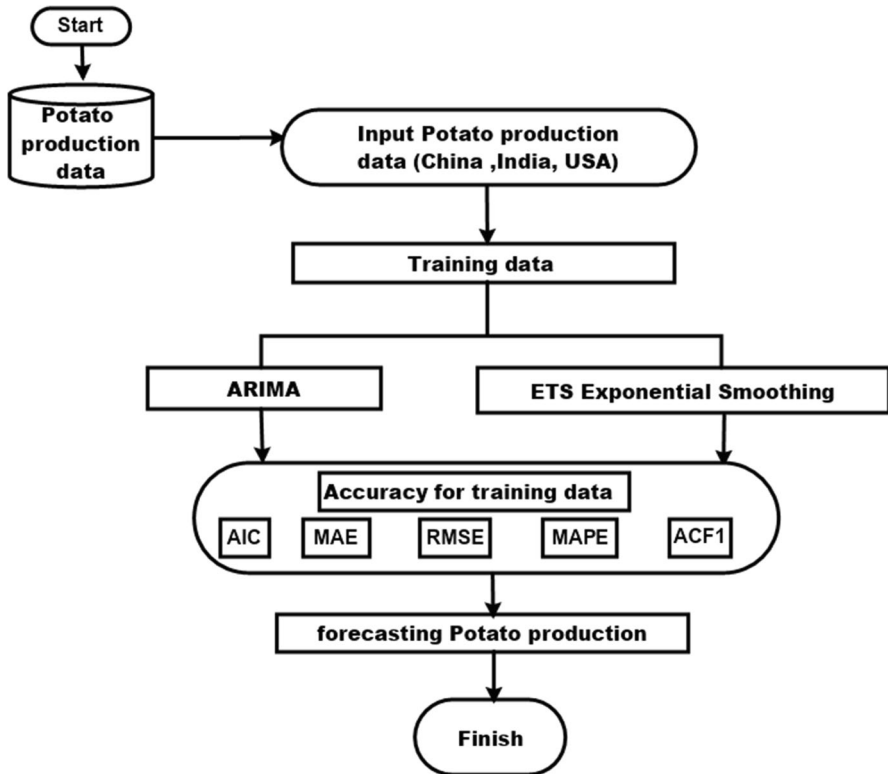
Thus, the study also highlights the limitations of the forecasting models, emphasizing the need for caution in interpreting results due to unforeseen events like natural disasters, pandemics, and geopolitical upheavals that can disrupt trends and invalidate predictions.

## Material and Methods

The data for this study was acquired from the Food and Agriculture Organisation (FAO) of the United Nations website <http://www.fao.org/faostat/en/#data/QC>, which offers yearly information on potato production for India, China, and the USA. The dataset encompasses the timeframe spanning from 1961 to 2022 and is quantified in units of one thousand metric tonnes. In present investigation, schema presented in Fig. 1.

The scheme appears to be a flowchart for forecasting potato production using time series analysis. Here is a description of the flow:

- 1- Start: The initial step of the process.



**Fig. 1** Schema representation of forecasting of potato production in China, India, and the USA

- 2- Potato production data: This seems to be the first stage where the relevant data for potato production is identified as the input for the process.
- 3- Input Potato production data (China, India, USA): This step specifies that the potato production data will be coming from three countries: China, India, and the USA.
- 4- Training data: The gathered data is then designated as training data, which will be used to fit the forecasting models.
- 5- The training data splits into two paths, indicating that two different forecasting methods will be used:

→ ARIMA: Autoregressive Integrated Moving Average, a popular statistical method for time series forecasting.

→ ETS Exponential Smoothing: Exponential Smoothing with Error, Trend, and Seasonal components, another common forecasting method.

- 6- Accuracy for training data: Both methods are evaluated based on several accuracy metrics:
  - AIC: Akaike Information Criterion, a measure of the relative quality of statistical models for a given set of data.
  - MAE: Mean Absolute Error, a measure of errors between paired observations expressing the same phenomenon.
  - RMSE: Root Mean Square Error, a measure of the differences between values predicted by a model and the values observed.
  - MAPE: Mean Absolute Percentage Error, a measure of prediction accuracy in a forecasting model.
  - ACF1: Autocorrelation Function at lag 1, a measure of the correlation between observations of a time series that are separated by one time period.
- 7- Forecasting Potato production: Based on the evaluations, the more accurate forecasting model is likely selected to predict future potato production.
- 8- Finish: The end of the process.

The flowchart outlines a structured approach to forecasting, beginning with data collection and ending with the selection of a forecasting model based on various accuracy metrics.

## Results and Discussion

The plot appears to show the trend of potato production in India over several decades, starting from around 1960 up to the early 2022s. The trend is represented in Fig. 2. The production values are plotted on the y-axis, while the x-axis represents the years. The y-axis is labeled in units of thousand tonnes, indicating the scale of production measured. From the visual, we can observe a general upward trend in potato production over the years. The growth appears to be relatively steady without drastic fluctuations, suggesting a consistent increase in production capacity or yield over time. There are a few notable periods where the slope of the line becomes steeper, indicating more rapid growth during those times. Overall, the plot suggests a positive trend in the agriculture of potatoes in India, reflecting improvements in production methods, expansion of cultivation areas, or possibly increasing demand driving the growth in production.

The graph represented in Fig. 3 is a time series plot that illustrates the progression of potato production in China from the 1960s through to the early 2022s. The x-axis denotes time, segmented by years, and the y-axis signifies the volume of production in thousand tonnes. A cursory examination reveals a marked increase in potato production over the span of six decades. The trajectory of growth is not linear; rather, there are periods of more pronounced growth interspersed with

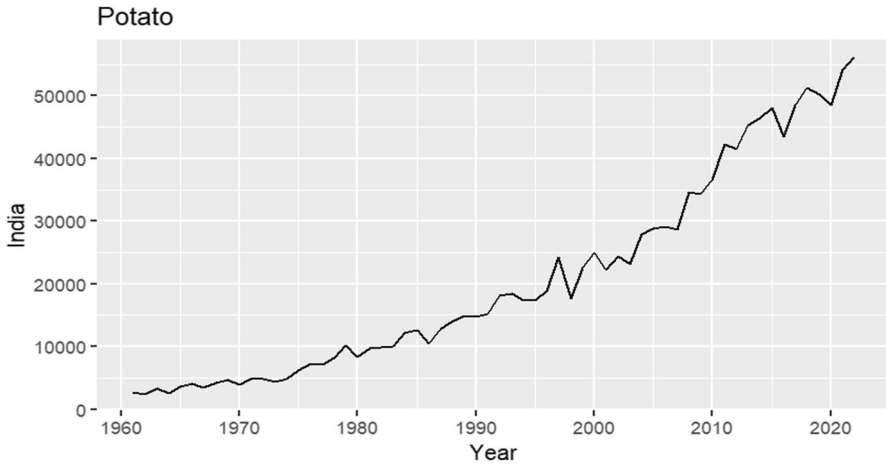


Fig. 2 Trend of potato production in India

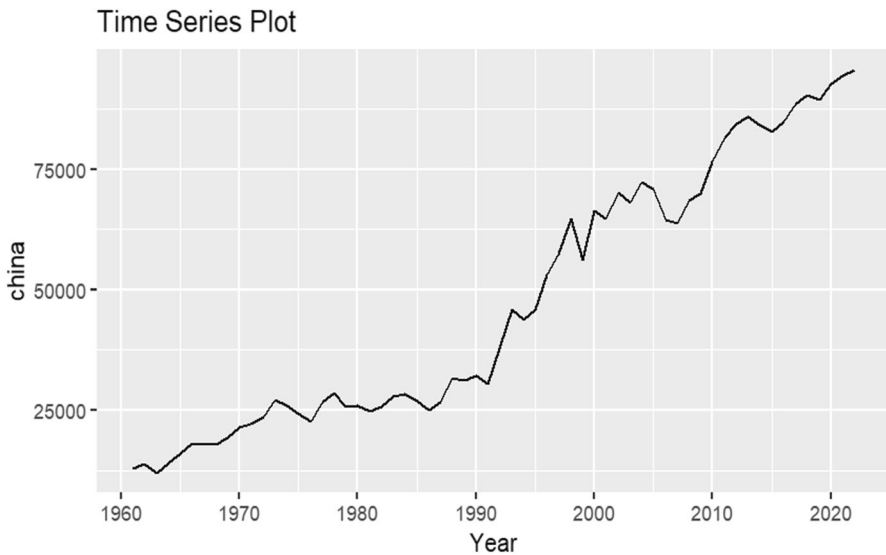


Fig. 3 Trend of potato production in China

phases of gradual increase. This could indicate varying factors at play over the years, such as advancements in agricultural technology, policy changes, shifts in domestic demand, or even international market influences. One notable observation is that the production appears to remain relatively stable or to grow only slightly during the initial two decades, followed by a more steep and consistent ascent from the 1980s onwards. The most significant growth occurs after the year 2000, where the plot shows a steep climb, suggesting a substantial expansion in

potato production capacities or efficiencies during this period. This plot serves as a strong indicator of China's increasing role in potato production globally, and it could be reflective of broader agricultural trends and economic policies implemented within the country during the late twentieth and early twenty-first centuries.

The provided plot depicts the trend of potato production in the United States of America from the 1960s to the early 2020s. The y-axis is labeled in thousand tonnes, which is the measure of production quantity, and the x-axis is marked with years, indicating the timeline.

From the graph represented in Fig. 4, we can discern that the production of potatoes in the USA has experienced more variability compared to the consistent growth seen in the previous graphs for India and China. Starting from the 1960s, there is a general upward trend in production until around the mid-1990s, where it peaks. Following this peak, there is a noticeable fluctuation in production levels, with several ups and downs. These fluctuations could be due to a multitude of factors, such as changes in consumer demand, variations in annual yields caused by climatic conditions, shifts in agricultural policy, or economic factors affecting the agriculture sector. The graph also shows that despite the variability, production has largely remained within a certain range, suggesting that while year-to-year production may change, overall capacity has not drastically increased or decreased. The period from the late 2000s to the early 2020s is characterized by a slight decline in production, with a few years showing more significant dips. This may reflect market saturation, changes in dietary habits, competition from other crops or international producers, or possibly environmental and sustainability considerations impacting production practices in the USA.

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots for three different countries: China, India, and the United States of America are represented in Fig. 5. These plots are commonly used in time series

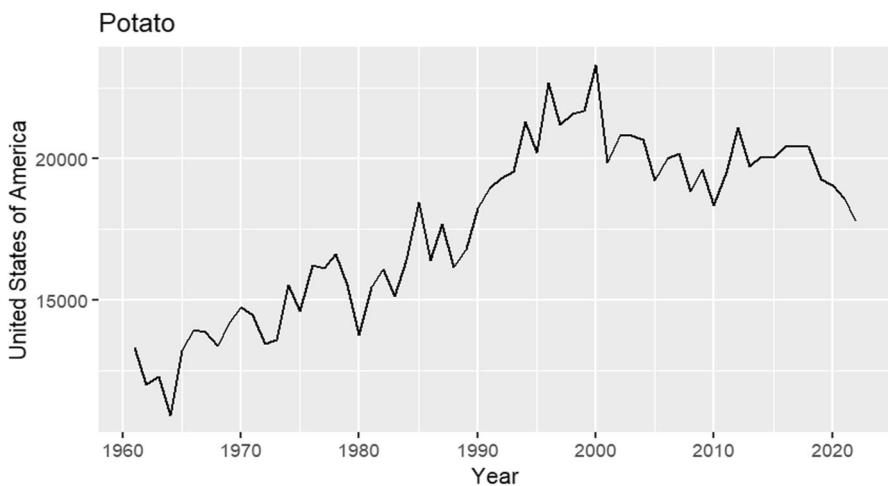
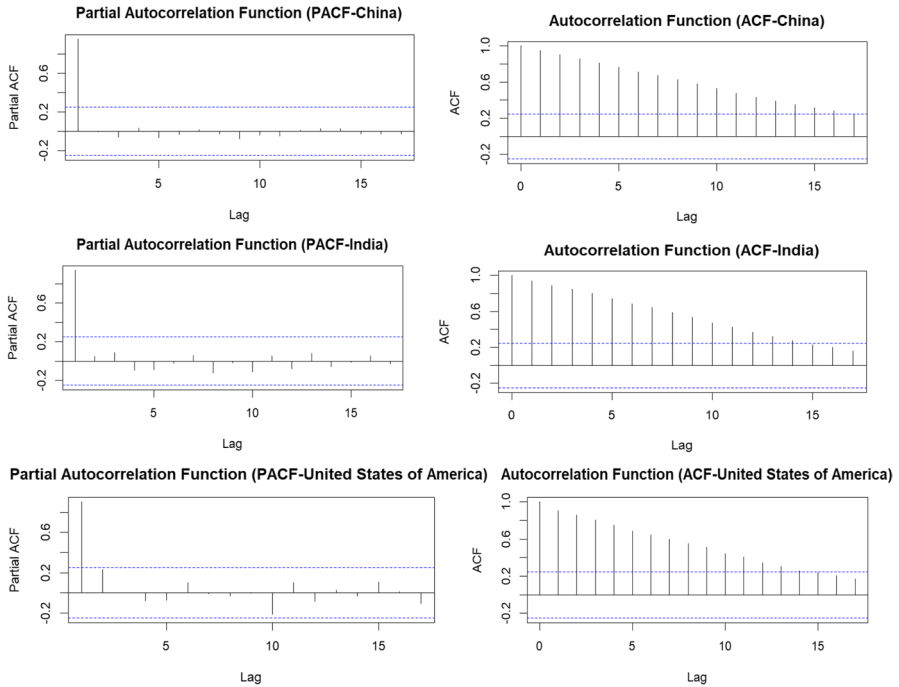


Fig. 4 Trend of potato production in USA



**Fig. 5** ACF and PACF plots for China, India, and the USA

analysis to determine the correlation of a signal with a delayed copy of itself at different points in time (lags) (Ray et al. 2023).

- 1- ACF for China: The ACF plot for China shows strong autocorrelations that slowly taper off as the lag increases. This suggests a gradual decline in the correlation of the production data with its past values over time. The fact that several lags have autocorrelation values above the significance level (outside the blue dotted lines) indicates that the data points are not independent of each other.
- 2- PACF for China: The PACF plot for China shows a sharp cut-off after the first lag, with all subsequent lags falling within the significance level (blue dotted lines). This indicates that the autocorrelation for any lag beyond the first is essentially explained by the lag-1 autocorrelation, pointing to a possible AR(1) process, where AR stands for “autoregressive.”
- 3- ACF for India: Similar to China’s ACF plot, India’s plot also shows strong autocorrelations at the initial lags that decrease over time. The autocorrelation is significant for several lags, which again suggests that the values are not independent and have a strong relationship with their past values.
- 4- PACF for India: The PACF plot for India demonstrates that the partial autocorrelations are within the significance level quite quickly after the first few lags, suggesting that only the immediate past values have a strong influence on the current value, with a potential AR process at play.



- 5- ACF for the United States of America: The ACF plot indicates significant auto-correlation at the first couple of lags, with a notable decrease thereafter, and remaining within the confidence bounds. This points to a less strong dependency on past values as compared to China or India.
- 6- PACF for the United States of America: The PACF plot shows that the partial autocorrelations are mostly within the significance level from the start, suggesting that each value in the series is largely independent when the effects of the previous values are accounted for.

In summary, the ACF and PACF plots suggest that the time series data for each country exhibit different characteristics and dependencies on past values. China and India show stronger temporal dependencies, which could be modeled by an AR process. In contrast, the USA shows weaker temporal dependencies, suggesting a more complex or different type of process could be at play. These analyses are crucial for building accurate time series forecasting models.

Table 1 summarizes the results of fitting ARIMA (Auto Regressive Integrated Moving Average) models to the potato production time series data for three countries: China, India, and the USA. ARIMA models are a popular choice for time series forecasting because they account for different types of patterns in the data, such as trends and seasonality. Here is a detailed description of the results for each country:

• **China**

- **Model:** The ARIMA model used for China is (0,1,0) with a drift term.
- **Parameters:** This model does not use any autoregressive (AR) or moving average (MA) terms, as indicated by the 0 s. The drift term is 1356.1, suggesting a linear trend over time.
- **AIC:** The Akaike Information Criterion (AIC) for this model is 1166.2, which is a measure of the relative quality of the statistical model for a given set of data. Lower AIC values generally indicate a better model.
- **Training:** The Root Mean Square Error (RMSE) on the training data is 3300.3. RMSE is a measure of the differences between values predicted by the model and the values actually observed. The lower the RMSE, the better the model's fit.

**Table 1** The results of estimating ARIMA models for three countries

Countries	Model	Parameters ARIMA			AIC	Training		Testing
		drift	AR	MA		RMSE	MAPE	
China	(0,1,0)	1356.1	-	-	1166.2	3300.3	6.16	7.21
India	(4,2,0)	-	-0.450	-	1089.0	1859.7	8.41	10.21
USA	(0,1,1)	-	-	-0.396	1031.4	1089.2	5.02	7.36

- **Testing:** The Mean Absolute Percentage Error (MAPE) on the testing data is 6.16%, indicating that, on average, the model's forecasts are off by this percentage.
- **India**
  - **Model:** The ARIMA model for India is (4,2,0), indicating that the data is differenced twice ( $I=2$ ) to make it stationary and that it includes four autoregressive terms ( $AR=4$ ).
  - **Parameters:** The AR coefficient of  $-0.450$  suggests that there is a negative correlation with the fourth lagged value.
  - **AIC:** An AIC of 1088.97 implies that, relative to the other models, this one has a better fit to the data as it is the lowest AIC presented.
  - **Training:** The RMSE is 1859.7, which is lower than China's RMSE, suggesting a better fit to the training data.
  - **Testing:** The MAPE is higher for both the training (8.41%) and testing (10.21%) datasets compared to China, which means the predictions for India are less accurate.
- **USA**
  - **Model:** The ARIMA model used is (0,1,1), meaning that the data is differenced once to make it stationary and includes one moving average term ( $MA=1$ ).
  - **Parameters:** The MA coefficient of  $-0.396$  indicates the influence of the first lagged forecast error.
  - **AIC:** The AIC is 1031.4, suggesting it is a better model than China's but not as good as India's based on the AIC values alone.
  - **Training:** The RMSE is 1089.2, which is significantly lower than both China and India, indicating a very good fit to the training data.
  - **Testing:** The MAPE is 5.02% for the training set and 7.36% for the testing set, showing that the USA's model has the lowest prediction error among the three on the training set and is comparable on the testing set.

In conclusion, each country's ARIMA model has different characteristics based on the data's behavior. India's model is the most complex, with four AR terms and two levels of differencing. The USA's model has the lowest RMSE for the training data, suggesting a good fit, but its testing MAPE is comparable to China's. China's model is the simplest and has a drift term, which implies a trend but not necessarily the best fit or forecasting accuracy as indicated by the higher MAPE compared to the USA.

Table 2 provides the results from fitting Exponential Smoothing State Space Model (ETS) to the potato production data for China, India, and the USA. ETS models are another commonly used approach for forecasting time series data, which captures various components such as error, trend, and seasonality. Here is a breakdown of the ETS models' results for each country:

**Table 2** The results of estimating ETS models for three countries

Countries	Model	Parameters ETS			Initial stats		AIC	Training		Testing MAPE
		$\alpha$	$\beta$	$\phi$	$l$	$b$		RMSE	MAPE	
China	(M,A,N)	0.890	0.0001	-	11375.2	963.1	1263.8	3293.2	6.12	6.91
India	(M,A,N)	0.347	0.068	-	2347.1	232.8	1178.1	1985.8	8.58	9.61
USA	(A,N,N)	0.598	-	-	12813.7	-	1129.0	1089.2	5.07	5.99

• **China**

- **Model:** The ETS model is multiplicative error, additive trend, and no seasonality (M,A,N).
- **Parameters:** The smoothing parameter for the level ( $\alpha$ ) is 0.890, indicating a high weight on recent observations for the level component. The smoothing parameter for the trend ( $\beta$ ) is 0.0001, which is very small, suggesting a nearly flat trend.
- **Initial stats:** The initial level ( $l$ ) is 11375.2, and the initial trend ( $b$ ) is 963.1.
- **AIC:** The Akaike Information Criterion is 1263.8, which helps to compare the quality of the model among different fits. The lower the AIC, the better the model in terms of the trade-off between goodness of fit and complexity.
- **Training:** The Root Mean Square Error (RMSE) on the training set is 3293.2. This is a measure of the average magnitude of the errors between predicted and actual values.
- **Testing:** The Mean Absolute Percentage Error (MAPE) is 6.12% for the training set and 6.91% for the testing set, showing a relatively consistent forecasting accuracy between the training and testing phases.

• **India**

- **Model:** The ETS model is also multiplicative error, additive trend, and no seasonality (M,A,N).
- **Parameters:** The smoothing parameter for the level ( $\alpha$ ) is 0.347, indicating moderate weighting of recent observations, and the smoothing parameter for the trend ( $\beta$ ) is 0.068.
- **Initial stats:** The initial level ( $l$ ) is 2347.1, and the initial trend ( $b$ ) is 232.8.
- **AIC:** The AIC for India’s model is 1178.1, suggesting a better fit compared to China’s model based on AIC values.
- **Training:** The RMSE is 1985.8, which is lower than China’s, indicating a better fit to the training data.
- **Testing:** The MAPE is 8.58% for the training set and 9.61% for the testing set, which shows a slight increase in error from training to testing.

• **USA**

- **Model:** The ETS model for the USA is additive error, no trend, and no seasonality (A,N,N), indicating a stable series without a clear trend.

- **Parameters:** The smoothing parameter for the level ( $\alpha$ ) is 0.598, and there are no parameters for trend ( $\beta$ ) or seasonal component ( $\phi$ ) as they are not used in this model.
- **Initial stats:** The initial level (l) is 12,813.7. There is no initial trend (b) value as the model does not include a trend component.
- **AIC:** The AIC is 1129.04, which is lower than China's and slightly less than India's, indicating a potentially better model fit.
- **Training:** The RMSE for the training set is 1089.2, which is the same as the RMSE given for the ARIMA model of the USA, suggesting a consistent fit.
- **Testing:** The MAPE is 5.07% for the training set and 5.99% for the testing set, showing a small increase in error but still maintaining a reasonable level of accuracy.

In conclusion, the ETS models show that each country's time series behaves differently. China's model places a high weight on recent observations for the level component with almost no trend. India's model has moderate smoothing parameters for both level and trend. The USA's model does not incorporate trend or seasonal adjustments, reflecting a stable series. The RMSE and MAPE values indicate that the ETS models fit the training data relatively well and provide a decent level of forecast accuracy in the testing phase, with the USA model showing the lowest error rates in both training and testing.

Table 3 presents a forecast for China's potato production for the years 2023 to 2027. The forecast includes point estimates as well as high and low 95% confidence intervals. Here is a detailed description of the forecast:

1. **Point forecast:** This column shows the predicted value of potato production for the given year. It is the central estimate around which the confidence intervals are constructed. For instance, the forecast predicts that China will produce approximately 96,554.59 thousand tonnes of potatoes in 2023.
2. **Hi 95:** This column represents the upper bound of the 95% confidence interval. There is a 95% chance that the actual value of potato production will be less than this. For 2023, the upper bound is around 112,044.9 thousand tonnes.
3. **Lo 95:** This column shows the lower bound of the 95% confidence interval. There is a 95% chance that the actual value will be more than this. For 2023, the lower bound is approximately 81,064.32 thousand tonnes.
4. **Year:** This column indicates the year for which the forecast is made.

**Table 3** Forecast for China

Year	Point forecast	Lo 95	Hi 95
2023	96554.59	81064.32	112044.9
2024	97520.42	76632.45	118408.4
2025	98486.25	73235.00	123737.5
2026	99452.08	70400.93	128503.2
2027	100417.91	67930.93	132904.9

The forecast shows a gradual increase in the point forecast for potato production from 2023 to 2027. The increasing trend is consistent across the 5 years, with the forecast for 2027 being roughly 4863.32 thousand tonnes higher than the forecast for 2023.

The confidence intervals widen as we move further into the future, which is common in forecasting since the level of uncertainty typically increases with the forecasting horizon. For example, the range between the high and low predictions for 2023 is about 30,980.57 thousand tonnes, whereas for 2027, it is approximately 53,973.96 thousand tonnes, indicating increased uncertainty in the later forecast.

In summary, the forecast suggests that China’s potato production is expected to increase year over year for the next 5 years, with a point forecast reaching just over 100,000 thousand tonnes by 2027. However, there is a range of uncertainty around these predictions, as reflected by the confidence intervals.

Table 4 provides a forecast for India’s potato production from 2023 to 2027, expressed in thousand tonnes. It includes high and low estimates at a 95% confidence interval, as well as the point forecasts for each year. Here is an in-depth analysis of the table:

1. **Point forecast:** This is the predicted central value of potato production for the specified year. The forecast suggests a steady increase in production over the five-year period. Starting at approximately 56,438.15 thousand tonnes in 2023, the forecasted production rises each year to reach about 61,882.58 thousand tonnes by 2027.
2. **Hi 95 (high 95% confidence interval):** This figure represents the upper end of the forecast interval, meaning there is a 95% probability that the actual production will be lower than this value. The interval expands from about 69,101.06 thousand tonnes in 2023 to approximately 81,432.79 thousand tonnes by 2027, showing increasing uncertainty in the forecast as the timeline extends.
3. **Lo 95 (low 95% confidence interval):** Conversely, this figure is the lower end of the forecast interval, with a 95% probability that the actual production will be higher than this value. The lower interval also grows wider from around 43,775.23 thousand tonnes in 2023 to about 42,332.37 thousand tonnes in 2027. Interestingly, the lower bound slightly decreases from 2025 to 2027, which might suggest that while overall production is expected to increase, there is increasing uncertainty or potential for less growth than expected in the latter years.
4. **Year:** This column simply denotes the year for which the forecast is applicable.

**Table 4** Forecast for India

Year	Point forecast	Lo 95	Hi 95
2023	56438.15	43775.23	69101.06
2024	57799.26	43788.34	71810.18
2025	59160.36	43544.21	74776.52
2026	60521.47	43054.22	77988.73
2027	61882.58	42332.37	81432.79

From the forecast data, we can infer a positive growth trend in India's potato production, with an annual increase in the central forecast. However, the widening range between the high and low estimates indicates greater uncertainty in the projections as time progresses. The forecast model seems to be quite certain about the growth in the initial years (with a tighter confidence interval) but indicates more variability and less certainty in the outer years. This could be due to various factors such as economic conditions, agricultural developments, potential policy changes, or external market forces that could affect production figures.

Table 5 displays the forecasted potato production for the United States of America for the years 2023 through 2027. The forecast is structured to provide an estimate (Point Forecast) along with a range that captures the uncertainty of the predictions (Hi 95 and Lo 95, representing the high and low ends of the 95% confidence interval, respectively). Here are the details:

1. **Point forecast:** This column represents the predicted value of potato production for the given year. Notably, the point forecast remains constant at 18,229.56 thousand tonnes for each year from 2023 to 2027. This suggests that the forecasting model predicts no change in the production level across these 5 years.
2. **Hi 95 (high 95% confidence interval):** This value provides the upper limit of the 95% confidence interval for the forecast. It indicates that there is a 95% chance that the actual production will be less than this value. The high end of the confidence interval increases slightly each year, from 20,399.69 thousand tonnes in 2023 to 21,614.24 thousand tonnes by 2027. This incremental increase suggests that while the point forecast is static, the potential for higher production levels grows with each year.
3. **Lo 95 (low 95% confidence interval):** This value is the lower limit of the 95% confidence interval. It suggests a 95% probability that the actual production will be more than this amount. Similar to the high end, the low end of the interval decreases slightly over the years, from 16,059.43 thousand tonnes in 2023 to 14,844.88 thousand tonnes by 2027. The decreasing lower bound of the confidence interval indicates growing uncertainty in the lower range of the forecast.
4. **Year:** This is the year for which the forecast is applicable.

**Table 5** Forecast for USA

Year	Point forecast	Lo 95	Hi 95
2023	18229.56	16059.43	20399.69
2024	18229.56	15700.51	20758.61
2025	18229.56	15386.54	21072.58
2026	18229.56	15103.96	21355.17
2027	18229.56	14844.88	21614.24

The forecasted values for potato production in the USA are unique in that the central point estimate remains unchanged over the five-year period, indicating a model prediction of stable production. However, the widening gap between the Hi 95 and Lo 95 confidence intervals each year reflects increasing uncertainty in the forecast as time progresses, which is a common occurrence in longer-term forecasts. The table suggests that while the expected level of potato production is to remain steady, there is less certainty about this forecast the further into the future we look. This could be due to factors such as economic conditions, market demands, climate change, and other variables that could influence agricultural output in the coming years.

The plots provided in Fig. 6 illustrate the historical data and forecasts for potato production in China, India, and the United States of America, using Exponential Smoothing State Space Model (ETS) methodology.

### 1. China

- The historical data shows a long-term increasing trend in potato production from the 1960s to the present.
- The forecast (shaded area) starts from just after 2020 and extends into the future.

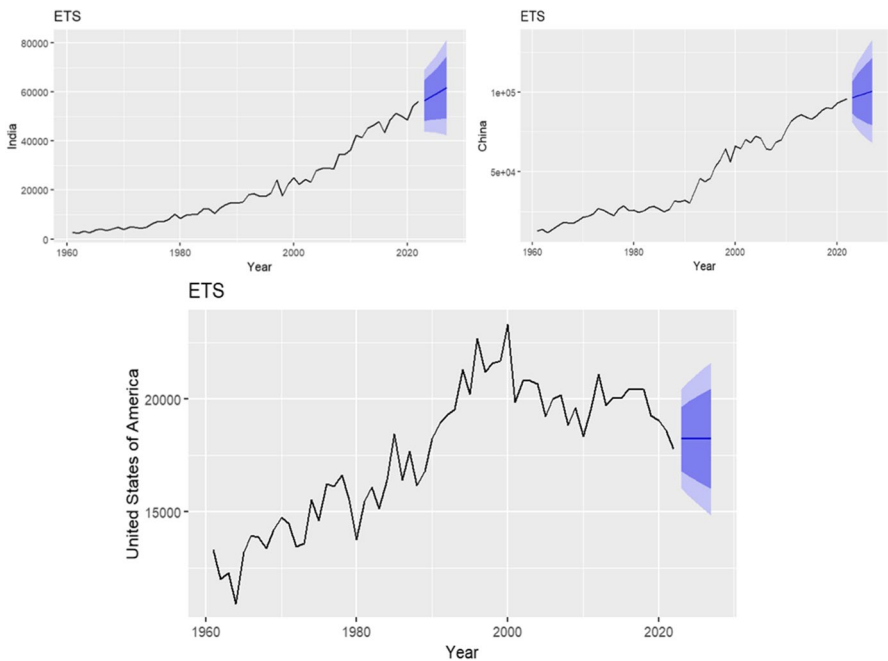


Fig. 6 Forecasts in China, India, and the USA using ETS

- The shaded forecast area has a darker middle region representing the point forecasts and a lighter region showing the 95% confidence intervals.
- The confidence intervals suggest that there is some uncertainty in the forecast, but the overall trend is expected to continue upwards.
- The increasing breadth of the confidence interval over time indicates increasing uncertainty in the longer-term predictions.

2. **India**

- Similar to the first plot, the historical trend for India also shows a steady increase in production over the decades.
- The forecast section again shows the expected continuation of this trend, with the point forecasts in the darker shade.
- As with the China plot, the confidence intervals are wider as the forecast extends further into the future, implying greater uncertainty in the long-term forecast.
- The plot indicates that the growth in potato production is expected to persist, with the exact values likely to fall within the projected confidence intervals.

3. **United States of America**

- This plot shows a more variable trend in historical potato production with several peaks and troughs.
- The forecast suggests a slight upward trend or stabilization in production levels.
- The 95% confidence intervals here also widen as the forecast moves further from the last historical data point, which is typical due to the cumulative effects of uncertainty in the forecasts.
- Unlike the China and India plots, the USA plot indicates a less clear trend, reflecting the past variability and suggesting a less certain future production level.

In all three plots, the historical data provides the basis for the ETS models to generate forecasts. The confidence intervals in the forecast periods capture the range within which the actual values are expected to fall with a 95% probability. The consistency of the historical trends in China and India is reflected in their respective forecasts, while the variability in the USA’s historical data leads to a more cautious forecast. Both model were performed significantly for training and testing data point as measured by the goodness of fit, i.e., RMSE and MAPE (Table 6). From the table,

**Table 6** Comparison of ARIMA and ETS model in training and testing set

Country	ARIMA			ETS		
	Training		Testing	Training		Testing
	RMSE	MAPE	MAPE	RMSE	MAPE	MAPE
China	3300.3	6.16	7.21	3293.2	6.12	6.91
India	1859.7	8.41	10.21	1985.8	8.58	9.61
USA	1089.2	5.02	7.36	1089.2	5.07	5.99



one can stated that the performance of both ARIMA and ETS model are at par for training set, but ETS model performed quite batter in testing set than the ARIMA model.

## Discussion

The forecast models suggest continued growth in potato production for China and India. This growth can be attributed to several factors, such as the expansion of arable land, improvements in yield due to advancements in agricultural practices and technology, and possibly increasing domestic and international demand for potatoes. The consistency in the upward trend for these countries implies that they may strengthen their position as major potato producers on the global stage.

The USA presents a different picture, with the forecasts suggesting a stable production level. The variability observed in the historical data and reflected in the forecasts could be due to market saturation, shifts in consumer preferences, or even changes in domestic agricultural policies. The stability in the forecasts might also indicate that the USA has reached an equilibrium in production capacity, or it may reflect the impact of environmental considerations and sustainable farming practices that limit further expansion. The widening confidence intervals in the forecasts for all countries highlight the inherent uncertainty in agricultural forecasting, which can be influenced by numerous unpredictable factors, including climate change, economic shifts, and policy changes. These factors can have significant effects on production capabilities and market dynamics.

Furthermore, the limitations of the forecasting models. While ARIMA and ETS models are robust and widely used, they rely on historical data and assume that past patterns will continue into the future, which may not always be the case. Unforeseen events, such as natural disasters, pandemics, and geopolitical upheavals, can disrupt trends and invalidate predictions. In conclusion, while forecasts provide valuable insights and a basis for planning, they must be interpreted with caution, and it is crucial for stakeholders to remain adaptable and responsive to actual developments in the agricultural sector. We underline the need for continuous monitoring of the market and environmental conditions to adjust strategies as required to address the challenges and opportunities that arise in the dynamic field of agriculture.

## Conclusion

The analysis of potato production in China, India, and the USA using ARIMA and ETS forecasting models has provided valuable insights into future production trends. The models indicate a continuing upward trajectory for China and India, reinforcing their roles as significant contributors to the global potato market. For the USA, the forecast suggests a period of stabilization, with production levels expected to plateau. Forecasted production of potato in the study was found to be 100,417.91, 61,882.58, and 18,229.56 thousand tonnes for China, India, and the USA, respectively. These forecasts are crucial for policymakers, farmers, and stakeholders in

the agricultural industry as they plan for future demand, supply chain logistics, and market dynamics. However, the increasing uncertainty in long-term forecasts, as evidenced by the widening confidence intervals, underscores the complex nature of agricultural production and the multitude of factors that can influence outcomes.

## Further Research

Given the limitations of the current models and the unpredictability inherent in agricultural production, further research is essential in several areas:

1. **Model refinement:** Developing more sophisticated models that can incorporate additional variables such as climate data, soil health metrics, and technological advancements in agriculture could enhance the accuracy of production forecasts.
2. **Climate change effects:** With the growing impact of climate change on agriculture, research into adaptive farming practices and crop resilience is crucial. Understanding how changing weather patterns affect potato yields will be vital.
3. **Policy impact assessment:** Examining the role of agricultural policies and subsidies on production levels can provide insights into how governmental actions influence market outcomes. This can help in formulating policies that stabilize markets and ensure food security.
4. **Technological innovations:** As new agricultural technologies emerge, studying their adoption rates and impact on production efficiency can help predict future changes in production trends.
5. **Consumer behavior:** Investigating changes in consumer preferences and their effect on demand for potatoes can help align production with market needs.
6. **Supply chain dynamics:** The interplay between production levels and supply chain efficiency warrants further study, especially in light of disruptions caused by global events.
7. **Comparative studies:** Comparing forecasts with those of other staple crops could provide a more comprehensive view of the agricultural sector's future and the potential for crop substitution effects.

In conclusion, while the current research has provided a foundation for understanding future potato production trends, ongoing investigation is needed to navigate the complexities of global food production and to support informed decision-making in a changing world.

**Acknowledgements** Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R 308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Data Availability** On reasonable request, the corresponding author will provide data supporting the study's results. The raw data cannot be made public for reasons of confidentiality and privacy. However, researchers who satisfy the requirements for access to confidential data can be given access to aggregated and anonymized data as well as the statistical analysis codes. To request access to the data, interested researchers.

## Declarations

**Ethics Approval and Consent to Participate** Not applicable.

**Conflict of Interest** The authors declare no competing interests.


## References

- Hossain MM, Abdulla F (2016) Forecasting potato production in Bangladesh by ARIMA model. *J Adv Stat* 1(4):191–198
- Mishra P, Mohammad A, Al G (2023c) Forecasting potato production in major south Asian countries: a comparative study of machine learning. *Potato Res.* <https://doi.org/10.1007/s11540-023-09683-z>
- Mishra P, Yonar A, Yonar H, Kumari B, Abotaleb M, Das SS, Patil SG (2021a) State of the art in total pulse production in major states of India using ARIMA techniques. *Current Research in Food Science.* 1(4):800–806
- Mishra P, Al Khatib AMG, Lal P, Anwar A, Nganvongpanit K, Abotaleb M, ... Punyapornwithaya V (2023b) An overview of pulses production in India: retrospect and prospects of the future food with an application of hybrid models. *Natl Acad Sci Lett* 1–8 | <https://doi.org/10.1007/s40009-023-01267-2>
- Novkovic N, Mutavdzic B, Ilin Z, Ivanisevic D (2015) Potato production forecasting. *AgroZanje - Agro-Knowledge J* 14(1/4):345–355
- Rahman MM, Islam MA, Mahboob MG, Mohammad NS, Ahmed I (2022) Forecasting of potato production in Bangladesh using ARIMA and mixed model approach. *Scholars J Agric Vet Sci.* <https://doi.org/10.36347/sjavs.2022.v09i10.001>
- Ray S, Lama A, Mishra P, Das SS, Gurung B (2023) An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. *Appl Soft Comput* 149:110939. <https://doi.org/10.1016/j.asoc.2023.110939>
- Sahu PK, Das M, Sarkar B et al (2024) Potato production in India: a critical appraisal on sustainability, forecasting, price and export behaviour. *Potato Res.* <https://doi.org/10.1007/s11540-023-09682-0>
- Singh DP, Deo S (2015) Structural time series model for forecasting potato production. *AryaBhatta J Math Inform* 7(2):329–332
- Wang Z-J, Liu H, Zeng F-K, Yang Y-C, Dan Xu, Zhao Y-C, Liu X-F, Kaur L, Liu G, Singh J (2023) Potato processing industry in China: current scenario, future trends and global impact. *Potato Res* 66(2):543–562
- Yadav S, Mohammad A, Al G (2024) Decoding potato power: a global forecast of production with machine learning and state - of - the - art techniques. *Potato Res.* <https://doi.org/10.1007/s11540-024-09705-4>
- Zaheer K, Akhtar MH (2016) Potato production, usage, and nutrition—a review. *Crit Rev Food Sci Nutr* 56(5):711–721

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Pradeep Mishra<sup>1</sup>  · Amel Ali Alhussan<sup>2</sup> · Doaa Sami Khafaga<sup>2</sup> · Priyanka Lal<sup>3</sup> · Soumik Ray<sup>4</sup> · Mostafa Abotaleb<sup>5</sup> · Khder Alakkari<sup>6</sup> · Marwa M. Eid<sup>7</sup> · El-Sayed M. El-kenawy<sup>8</sup>

✉ Pradeep Mishra  
pradeepjnkvv@gmail.com

- <sup>1</sup> College of Agriculture, Jawaharlal Nehru KrishiVishwa Vidyalaya (JNKVV), Rewa 486001, India
- <sup>2</sup> Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, 11671 Riyadh, Saudi Arabia
- <sup>3</sup> Department of Agricultural Economics and Extension, School of Agriculture, Lovely Professional University, Punjab, India
- <sup>4</sup> Centurion University of Technology and Management, Paralakhemundi, Odisha, India
- <sup>5</sup> Department of System Programming, South Ural State University, Chelyabinsk, Russia
- <sup>6</sup> Faculty of Economics, Statistics and Programming, Tishreen University, Tartous, Syria
- <sup>7</sup> Department of Artificial Intelligence, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
- <sup>8</sup> Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt