



DỰ BÁO DIỆN TÍCH, NĂNG SUẤT VÀ SẢN LƯỢNG LÚA CỦA VIỆT NAM: ÁP DỤNG MÔ HÌNH ARIMA

Hồ Trọng Phúc*, Phạm Xuân Hùng

Trường Đại học Kinh tế, Đại học Huế, 99 Hồ Đắc Di, Tp. Huế, Việt Nam

* Tác giả liên hệ: Hồ Trọng Phúc <htphuc@hueuni.edu.vn>

(Ngày nhận bài: 13-4-2023; Ngày chấp nhận đăng: 22-6-2023)

Tóm tắt. Bài báo này phân tích và dự báo diện tích, năng suất và sản lượng lúa của Việt Nam đến năm 2030 sử dụng mô hình Box-Jenkins ARIMA. Nghiên cứu sử dụng chuỗi dữ liệu thời gian giai đoạn 1990-2021 được tổng hợp từ Tổng Cục Thống kê của Việt Nam. Kết quả thống kê cho thấy năm 2021, tổng diện tích canh tác lúa đạt 7,24 triệu ha, với năng suất bình quân đạt 6,06 tấn/ha và tổng sản lượng lúa đạt 43,85 triệu tấn. Qua giai đoạn 1990-2021, năng suất và sản lượng lúa tăng, với mức tăng trưởng bình quân hàng năm lần lượt là 2,10% và 2,70%; diện tích có tăng, nhưng tăng với tốc độ chậm hơn (0,58%) và có xu hướng giảm từ năm 2013 trở lại đây. Kết quả dự báo cho thấy rằng đến năm 2030, diện tích lúa tiếp tục giảm khoảng 0,8 triệu ha, xuống còn 6,42 (4,17; 8,67) triệu ha. Trong khi đó, năng suất và sản lượng lúa có xu hướng tăng, lần lượt đạt 6,90 (6,26; 7,53) tấn/ha và 46,60 (36,02; 57,19) triệu tấn. Nghiên cứu đề xuất rằng trong thập kỷ tới, các chính sách về sản xuất lúa cần tập trung thúc đẩy tăng năng suất và lợi nhuận thay vì mở rộng diện tích canh tác. Cải thiện hiệu quả sản xuất và áp dụng công nghệ tiên tiến là những định hướng phù hợp.

Từ khóa: mô hình ARIMA, dự báo, diện tích lúa, năng suất lúa, sản lượng lúa, Việt Nam

Forecasting rice area, yield and production of Vietnam: ARIMA model approach

Ho Trong Phuc*, Pham Xuan Hung

University of Economics, Hue University, 99 Ho Dac Di St., Hue, Vietnam

* Correspondence to Ho Trong Phuc <htphuc@hueuni.edu.vn>

(Received: April 13, 2023; Accepted: June 22, 2023)

Abstract. This paper analyzes and forecasts rice area, yield, and production of Vietnam by 2030 using Box-Jenkins ARIMA model. The study uses time series data over the 1990-2021 period compiled from the General Statistics Office of Vietnam. The statistical summary shows that in 2021, the total rice-cultivated area is 7.24 million ha, with an average yield of 6.06 tonnes/ha and a total rice production of 43.85 million tonnes. Over the 1990-2021 period, rice yield and production slightly increase, with an average annual

increase of 2.10% and 2.70%, respectively; the rice-cultivated area also increases, but at a lower rate (0.58%) and it tends to decrease from 2013 onwards. The forecasted results show that by 2030, the rice-cultivated area will continue to decrease by approximately 0.8 million ha, to 6.42 (4.17; 8.67) million ha. While, rice yield and production will increase and reach 6.90 (6.26; 7.53) tonnes/ha and 46.60 (36.02; 57.19) million tonnes, respectively. This study suggests that in the next decade, policies related to rice production should focus on promoting the improvements in rice yield and profits instead of expanding the rice-cultivated area. The improvement of production efficiency and the adoption of advanced technologies would be appropriate directions.

Keywords: ARIMA model, forecasting, rice area, rice yield, rice production, Vietnam

1 Đặt vấn đề

Canh tác lúa đóng vai trò quan trọng trong vấn đề đảm bảo an ninh lương thực và phát triển kinh tế, đặc biệt là các quốc gia chậm phát triển và đang phát triển phụ thuộc vào nông nghiệp [1]. Phúc và cs. [1] đã chỉ ra rằng có sự khác biệt lớn về năng suất lúa giữa các quốc gia trên thế giới. Năng suất lúa của Việt Nam năm 2018 đạt 5,82 tấn/ha, mặc dù cao hơn mức bình quân chung thế giới (4,68 tấn/ha) nhưng vẫn còn thấp so với các quốc gia trong khu vực và trên thế giới như Trung Quốc, Nhật Bản, Brazil, Hy Lạp, Thổ Nhĩ Kỳ, ... [1–3]. Điều này cho thấy Việt Nam vẫn còn tiềm năng để tăng sản lượng lúa từ việc tăng năng suất lúa.

Việc tăng năng suất lúa không chỉ giúp quốc gia sử dụng nguồn lực có hiệu quả hơn mà còn góp phần bù đắp sản lượng do sụt giảm diện tích lúa do nhu cầu chuyển đổi đất trồng lúa cho quá trình công nghiệp hóa và đô thị hóa, tác động tiêu cực của biến đổi khí hậu, và sự suy giảm chất các nguồn lực sản xuất. Số liệu thống kê cho thấy diện tích canh tác lúa có xu hướng giảm trong những năm gần đây (Hình 1). Việc duy trì mức sản lượng lúa không chỉ đảm bảo vấn đề an ninh lương thực trong nước mà còn đảm bảo vị thế quốc gia xuất khẩu gạo hàng đầu trên thị trường lúa gạo quốc tế. Số liệu thống kê cho thấy, năm 2021, sản lượng gạo xuất khẩu đạt 6,24 triệu tấn, đóng góp 3,29 tỷ USD cho nền kinh tế [4]. Do đó, việc tăng năng suất và sản lượng lúa của Việt Nam sẽ có những đóng góp lớn trong chiến lược giải quyết vấn đề thiếu hụt lương thực toàn cầu hiện nay [1].

Đã có nhiều nghiên cứu về hiệu quả sản xuất lúa để tìm kiếm giải pháp nâng cao hiệu quả sử dụng các nguồn lực trong sản xuất và nâng cao năng suất lúa ở Việt Nam [5–12]. Tuy nhiên, hiện chưa có nghiên cứu nào tiến hành đánh giá lại kết quả đạt được của sản xuất lúa gạo Việt Nam trong ba thập kỷ qua và dự báo sự biến động của diện tích, năng suất và sản lượng lúa cho

thập kỷ tới để làm cơ sở cho việc xây dựng chiến lược phát triển ngành lúa gạo trong giai đoạn tới. Do đó, nghiên cứu này tiến hành phân tích sự biến động của diện tích, năng suất và sản lượng lúa ở Việt Nam giai đoạn 1990–2021 và áp dụng mô hình dự báo ARIMA (Autoregressive Integrated Moving Average) ước tính sự biến động của diện tích, năng suất và sản lượng lúa của Việt Nam từ năm 2022 đến năm 2030. Kết quả nghiên cứu sẽ cung cấp những thông tin hữu ích cho các nhà hoạch định chính sách trong việc xây dựng chiến lược, lập quy hoạch và kế hoạch phù hợp cho phát triển ngành lúa gạo Việt Nam và các ngành liên quan trong thập kỷ tới. Bên cạnh đó, nghiên cứu còn là tài liệu tham khảo hữu ích cho các nhà nghiên cứu trong lĩnh vực sản xuất lúa gạo, đặc biệt là lĩnh vực dự báo ít nhận được sự quan tâm của các nhà nghiên cứu.

2 Tổng quan tài liệu

Dự báo kinh tế đóng vai trò quan trọng trong quá trình kế hoạch hóa nền kinh tế. Dự báo là bước tiền đề nhằm cung cấp thông tin để xác định mục tiêu và chiến lược phát triển, sau đó tiến hành lập quy hoạch, kế hoạch, và các chương trình dự án thực hiện để đạt mục tiêu đề ra [13, 14]. Phương pháp dự báo trong kinh tế khá đa dạng, tùy thuộc vào đối tượng và đặc điểm của dữ liệu. Quy trình thực hiện một nghiên cứu dự báo và các phương pháp dự báo có thể áp dụng được mô tả chi tiết trong nghiên cứu của Hyndman và Athanasopoulos [13] và Petropoulos và cs. [14].

Để dự báo cho mô hình chuỗi dữ liệu thời gian, mô hình ARIMA thường được sử dụng trong các nghiên cứu thực nghiệm [14]. Ưu điểm chính của mô hình ARIMA là sự linh hoạt và có thể xử lý nhiều loại dữ liệu chuỗi thời gian. Các mô hình ARIMA có thể cung cấp kết quả dự báo và khoảng dự báo đáng tin cậy vì chúng được ước lượng dựa trên phương pháp và lý thuyết thống kê. Ngoài ra, mô hình ARIMA dễ thực hiện và dễ giải thích kết quả vì mô hình chỉ yêu cầu ít tham số và giả thuyết. Một số nghiên cứu thực nghiệm đã áp dụng mô hình ARIMA để dự báo diện tích, năng suất và sản lượng lúa ở Việt Nam và một số quốc gia trên thế giới. Ví dụ như nghiên cứu của Khám [15] đã áp dụng mô hình ARIMA để dự báo năng suất lúa ở Hà Nội sử dụng chuỗi dữ liệu thời gian giai đoạn 1960–1998. Rahman [16] dự báo sản lượng lúa ở Bangladesh từ năm 2008 tới năm 2013 sử dụng mô hình ARIMA. Tài [17] sử dụng mô hình ARIMA và một số mô hình khác để dự báo sản lượng lúa của Việt Nam sử dụng chuỗi dữ liệu thời gian từ năm 1990 đến năm 2010. Biswas và Bhattacharyya [18] áp dụng mô hình ARIMA để dự báo diện tích và sản lượng lúa ở vùng Tây Bengal sử dụng chuỗi dữ liệu giai đoạn 1947/48–2007/08. Hamjah [19] sử dụng mô hình ARIMA để dự báo sản lượng lúa theo mùa vụ ở

Bangladesh sử dụng bộ dữ liệu thời gian từ năm 1972 tới năm 2006. Sahu và cs. [20] sử dụng mô hình ARIMA dự báo diện tích, năng suất, sản lượng và giống lúa của một số quốc gia và toàn thế giới. Hemavathi và Prabakaran [21] sử dụng mô hình ARIMA để dự báo diện tích, sản lượng và năng suất lúa ở huyện Thanjavur, Ấn Độ dựa trên bộ dữ liệu giai đoạn 1990/91–2014/15. Nghiên cứu của Hiền và cs. [22] ứng dụng mô hình ARIMA để dự báo sản lượng lúa của tỉnh Thừa Thiên Huế từ năm 2019 đến năm 2025 dựa trên bộ dữ liệu giai đoạn 1995–2018. Noorunnahar và cs. [23] sử dụng phương pháp ARIMA và XGBoost để dự báo sản lượng lúa ở Bangladesh dựa trên bộ dữ liệu giai đoạn 1961–2020.

Tổng quan tài liệu cho thấy các nghiên cứu về dự báo diện tích, năng suất và sản lượng lúa còn hạn chế, đặc biệt ở Việt Nam gần đây chưa có nghiên cứu nào tiến hành dự báo diện tích, năng suất và sản lượng lúa trên phạm vi cả nước cho các thập kỷ tới.

3 Phương pháp nghiên cứu

3.1 Dữ liệu

Nghiên cứu này sử dụng chuỗi dữ liệu thời gian về diện tích, năng suất và sản lượng lúa cả nước giai đoạn 1990–2021, được tổng hợp từ Tổng Cục Thống Kê của Việt Nam [2]. Ngoài ra, nghiên cứu cũng sử dụng dữ liệu được thu thập từ Tổ chức Nông nghiệp và Lương Thực của Liên Hợp Quốc (FAO). Thống kê mô tả bộ dữ liệu được trình bày chi tiết ở mục 4.1.

3.2 Phương pháp thống kê mô tả

Để đánh giá tăng trưởng về diện tích, năng suất và sản lượng lúa hàng năm và qua các giai đoạn, chúng tôi sử dụng các công thức tính giá trị tăng trưởng tuyệt đối, tốc độ tăng trưởng hàng năm, tốc độ tăng trưởng giai đoạn và tốc độ tăng trưởng bình quân giai đoạn [24]. Các công thức tính cụ thể như sau:

Công thức tính giá trị tăng trưởng tuyệt đối (ΔY_t)

$$\Delta Y_t = Y_t - Y_{t-1} \tag{1}$$

Công thức tính tốc độ tăng trưởng hàng năm (g_t)

$$g_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \times 100\% \tag{2}$$

Công thức tính tốc độ tăng trưởng giai đoạn (g_p)

$$g_p = \frac{Y_t - Y_0}{Y_0} \times 100\% \quad (3)$$

Công thức tính tốc độ tăng trưởng bình quân giai đoạn (\tilde{g}_p)

$$\tilde{g}_p = \left[\left(\frac{Y_t}{Y_0} \right)^{\frac{1}{n}} - 1 \right] \times 100\% \quad (4)$$

trong đó, Y_0 là giá trị năm gốc, Y_t là giá trị năm t , Y_{t-1} là giá trị năm $t - 1$, và n là số năm của giai đoạn.

3.3 Phương pháp dự báo với mô hình ARIMA

Để dự báo diện tích, năng suất và sản lượng lúa đến năm 2030, nghiên cứu này sử dụng mô hình Box-Jenkins ARIMA, được phát triển bởi Box and Jenkins [25]. Mô hình này được sử dụng rộng rãi trong việc dự báo chuỗi dữ liệu thời gian [13, 14]. Mục đích của mô hình ARIMA là để mô tả sự tự tương quan trong dữ liệu [13]. Mô hình ARIMA được tích hợp từ mô hình tự tương quan (AR) và mô hình trung bình trượt (MA). Phương trình tổng quát của mô hình ARIMA (p,d,q) được thể hiện như sau [22]:

$$\phi_p(B)(1-B)^d y_t = \delta + \theta_q(B)u_t \quad (5)$$

trong đó, $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ là quá trình tự hồi quy bậc p ;

$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ là quá trình trung bình trượt bậc q ;

$(1-B)^d y_t$ là sai phân bậc thứ d của chuỗi quan sát y_t ;

B là toán tử lùi: $B y_t = y_{t-1}$, $B^m y_t = y_{t-m}$, $y_t - y_{t-1} = (1-B)y_t$;

u_t là nhiễu trắng.

Quy trình ước lượng mô hình ARIMA gồm các bước sau:

Bước 1: Kiểm tra dữ liệu

Trước khi tiến hành các bước của mô hình dự báo, chúng ta tiến hành phân tích mô tả chuỗi dữ liệu để kiểm tra xem chuỗi dữ liệu có tính dừng hay không. Nếu dữ liệu không có tính dừng, tiến hành lấy sai phân bậc 1 và kiểm tra cho đến khi nào chuỗi dữ liệu có tính dừng.

Bước 2: Xác định mô hình ARIMA tối ưu

Để ước lượng mô hình ARIMA, trước tiên cần phải xác định giá trị p , d và q cho mô hình (tức là nhận dạng mô hình ARIMA (p,d,q) phù hợp). Trong đó, d là bậc sai phân của chuỗi dữ liệu thời gian được khảo sát, p là bậc tự hồi quy và q là bậc trung bình trượt [13]. Do dữ liệu có đặc tính như hàm tự tương quan thực nghiệm, nên hàm tự tương quan từng phần thực nghiệm có thể được sử dụng để xác định giá trị của p và q , trong trường hợp giá trị p và q thấp [14]. Ngoài ra, có thể sử dụng tiêu chuẩn thông tin AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion) để xác định giá trị p và q . Gần đây việc xác định giá trị p , d và q tối ưu cho mô hình ARIMA của chuỗi dữ liệu cần phân tích được thực hiện dễ dàng nhờ sử dụng hàm *auto.arima()* của *forecast* package [13] trong phần mềm R [13, 14]. Trong nghiên cứu này chúng tôi sử dụng hàm *auto.arima()* để tìm kiếm mô hình ARIMA (p,d,q) phù hợp cho các chuỗi dữ liệu.

Bước 3: Ước lượng và kiểm tra mô hình

Sau khi xác định được giá trị p , d và q , chúng ta tiến hành ước lượng mô hình. Hệ số của mô hình được ước lượng bằng phương pháp maximum likelihood estimation (MLE). Sau đó, chúng ta tiến hành kiểm tra lại kết quả ước lượng của mô hình để đảm bảo sự phù hợp của mô hình dự báo và đảm bảo độ tin cậy của kết quả dự báo. Việc kiểm tra tính phù hợp của mô hình thông qua việc kiểm tra phần dư của mô hình được chọn bằng cách hiện thị biểu đồ ACF của phần dư và kiểm tra portmanteau test của phần dư. Nếu chúng không giống như sai số ngẫu nhiên trắng (White noise), thì thử với mô hình điều chỉnh. Nếu phần dư giống sai số ngẫu nhiên trắng thì tiến hành bước tiếp theo đó là dự báo [13].

Bước 4: Dự báo

Sau khi kiểm tra tính phù hợp của mô hình được lựa chọn, dựa trên phương trình của mô hình ARIMA (p, d, q) phù hợp, chúng ta tiến hành dự báo và xác định khoảng tin cậy của giá trị dự báo cho các chuỗi dữ liệu nghiên cứu.

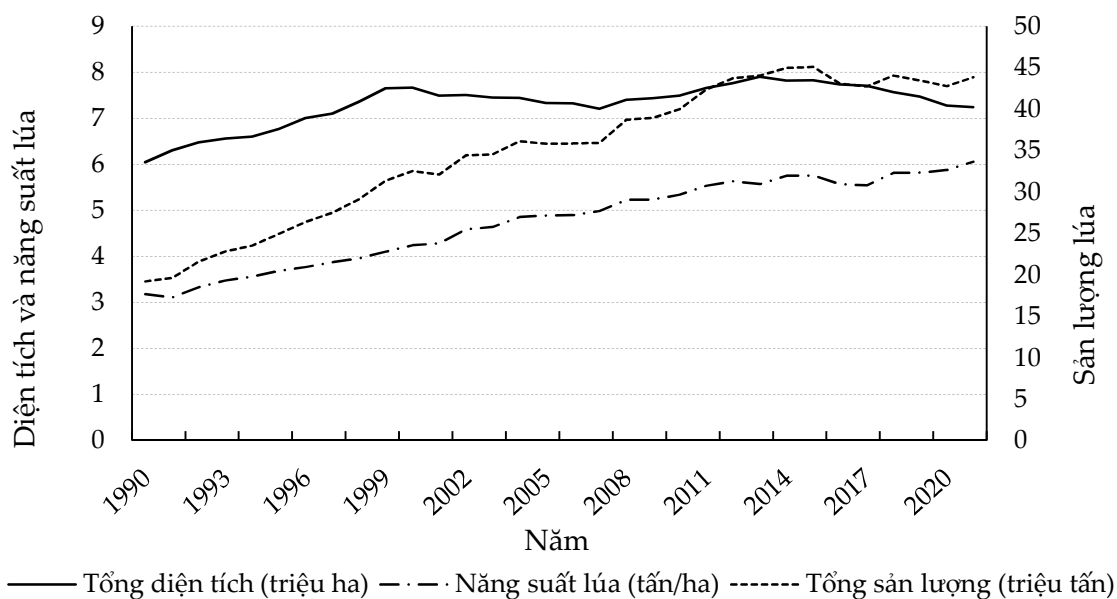
4 Kết quả nghiên cứu

4.1 Phân tích diện tích, năng suất và sản lượng lúa giai đoạn 1990–2021

Tổng diện tích, năng suất và sản lượng lúa cả năm của Việt Nam giai đoạn 1990–2021 được thể hiện ở Bảng 1. Đánh giá chung qua giai đoạn, cả diện tích, năng suất và sản lượng lúa đều

tăng. Tuy nhiên, tổng diện tích và sản lượng lúa có xu hướng giảm trong những năm gần đây (được thể hiện rõ ở Hình 1). Theo GSO [2] tổng diện tích lúa Việt Nam tăng từ 6,042 triệu ha năm 1990 lên 7,239 triệu ha năm 2021, tăng 19,79% (tương ứng với mức tăng bình quân giai đoạn 0,58%/năm). Tuy nhiên, từ năm 2013 trở lại đây, diện tích lúa có xu hướng giảm. Việc sụt giảm về diện tích, ít nhiều có ảnh hưởng tới tổng sản lượng lúa. Cụ thể, từ năm 1990 đến năm 2015, tổng sản lượng lúa tăng 19,23 triệu tấn lên 45,09 triệu tấn; tuy nhiên sau đó tổng sản lượng giảm sút đáng kể, xuống còn 43,85 triệu tấn năm 2021. Trái ngược với sự sụt giảm của diện tích và sản lượng, năng suất lúa đạt mức tăng liên tục trong 3 thập kỷ qua, ngoại trừ sự giảm nhẹ trong hai năm 2016 và 2017. Cụ thể, năng suất tăng gần gấp đôi từ 3,18 tấn/ha năm 1990 lên 6,06 tấn/ha năm 2021, tương ứng với mức tăng trưởng 90,41% (với mức tăng trưởng bình quân giai đoạn 2,1%/năm).

Bảng 1 cũng thể hiện diện tích, năng suất và sản lượng lúa phân theo mùa vụ giai đoạn 1990–2021. Có thể thấy rằng, diện tích canh tác lúa vụ Đông Xuân (ĐX) và Hè Thu – Thu Đông (HT-TĐ) có chiều hướng tăng, trong khi đó diện tích lúa Mùa giảm mạnh (cũng được thể hiện rõ ở Hình 2). Theo GSO [2], diện tích canh tác lúa vụ Đông Xuân chiếm diện tích cao nhất và tăng



Hình 1. Diện tích, năng suất và sản lượng lúa cả năm giai đoạn 1990–2021

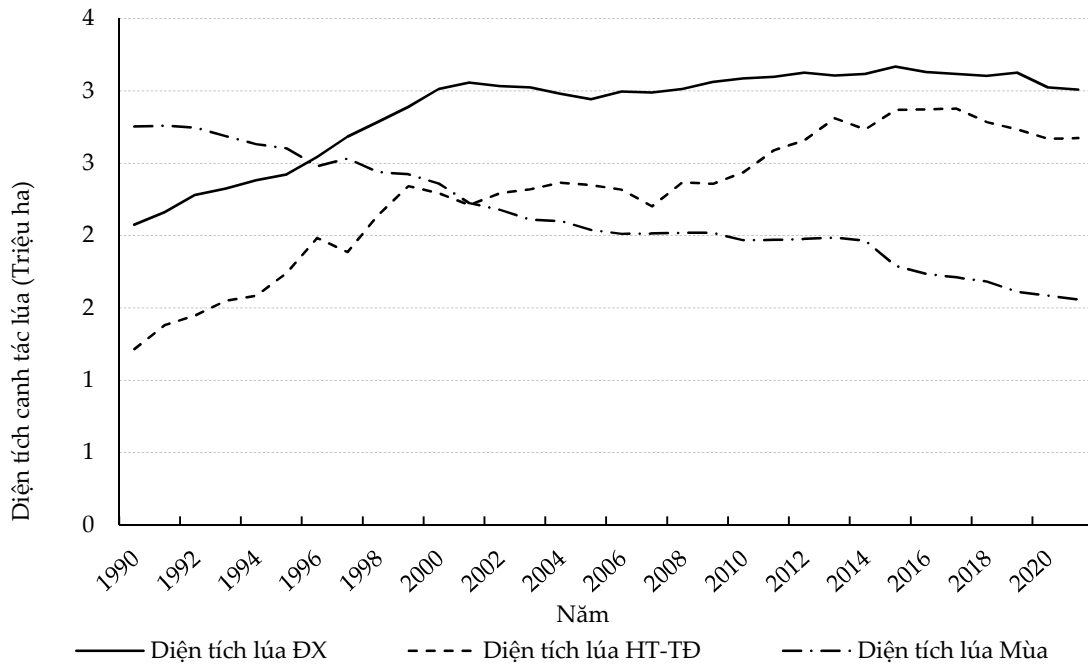
Nguồn: GSO [2]

Bảng 1. Diện tích, năng suất và sản lượng lúa của Việt Nam giai đoạn 1990–2021

| Năm | 1990 | 2000 | 2010 | 2021 | So sánh giai đoạn 2021/1990 | | |
|------------------------------|--------|--------|--------|--------|-------------------------------|-------------------------|-----------------------------|
| | | | | | Tăng trưởng giá trị tuyệt đối | % tăng trưởng giai đoạn | % tăng trưởng bình quân năm |
| Diện tích (Triệu ha) | | | | | | | |
| Tổng diện tích cả năm | 6,043 | 7,666 | 7,489 | 7,239 | 1,196 | 19,793 | 0,584 |
| Diện tích lúa ĐX | 2,074 | 3,013 | 3,086 | 3,007 | 0,933 | 45,006 | 1,206 |
| Diện tích lúa HT-TĐ | 1,216 | 2,293 | 2,436 | 2,673 | 1,458 | 119,914 | 2,575 |
| Diện tích lúa Mùa | 2,754 | 2,360 | 1,968 | 1,559 | -1,195 | -43,399 | -1,819 |
| Năng suất (Tấn/ha) | | | | | | | |
| Năng suất cả năm | 3,181 | 4,243 | 5,342 | 6,058 | 2,876 | 90,412 | 2,099 |
| Năng suất vụ ĐX | 3,793 | 5,168 | 6,227 | 6,861 | 3,067 | 80,866 | 1,930 |
| Năng suất vụ HT-TĐ | 3,365 | 3,762 | 4,797 | 5,672 | 2,307 | 68,566 | 1,699 |
| Năng suất vụ Mùa | 2,640 | 3,531 | 4,627 | 5,169 | 2,529 | 95,800 | 2,191 |
| Sản lượng (Triệu tấn) | | | | | | | |
| Tổng sản lượng cả năm | 19,225 | 32,530 | 40,006 | 43,853 | 24,627 | 128,101 | 2,696 |
| Sản lượng lúa ĐX | 7,866 | 15,571 | 19,217 | 20,629 | 12,763 | 162,266 | 3,159 |
| Sản lượng lúa HT-TĐ | 4,091 | 8,625 | 11,686 | 15,163 | 11,073 | 270,700 | 4,317 |
| Sản lượng lúa Mùa | 7,269 | 8,333 | 9,103 | 8,056 | 0,787 | 10,824 | 0,332 |

Nguồn: GSO [2]

từ 2,07 triệu ha năm 1990 lên 3,01 triệu ha năm 2021, mức tăng trưởng bình quân giai đoạn đạt 1,21%/năm. Diện tích lúa vụ Hè Thu – Thu Đông tăng từ 1,22 triệu ha năm 1990 lên 2,67 triệu ha năm 2021, mức tăng trưởng bình quân giai đoạn đạt 2,57%/năm. Ngược lại, diện tích lúa vụ Mùa có xu hướng giảm mạnh, từ 2,75 triệu ha năm 1990 xuống còn 1,56 triệu ha năm 2021, tương ứng với mức giảm bình quân giai đoạn là 1,82%/năm.

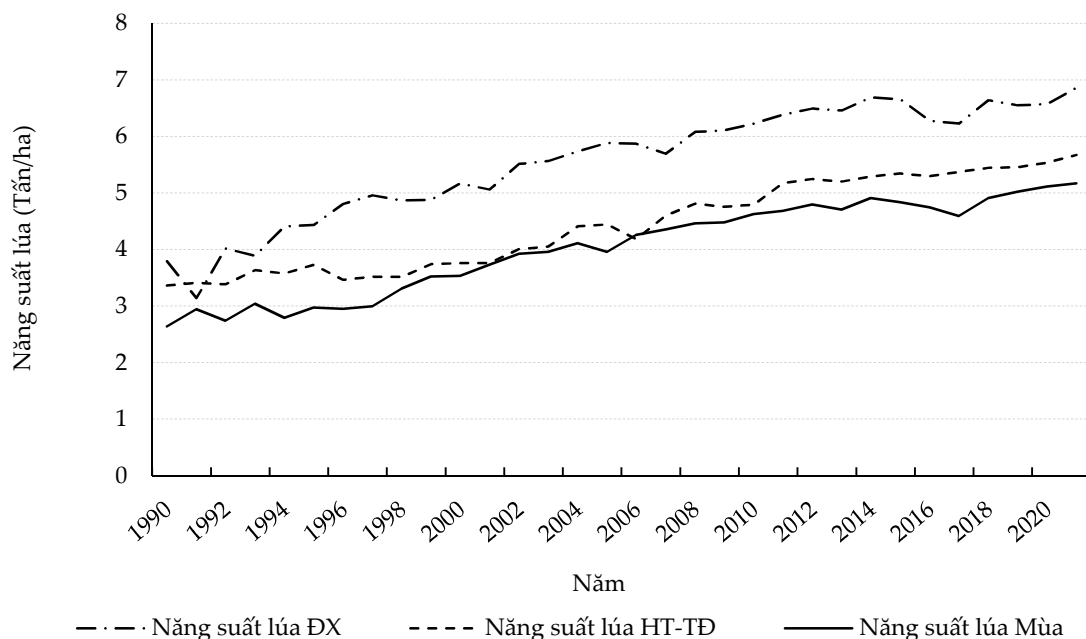


Hình 2. Diện tích canh tác lúa phân theo mùa vụ giai đoạn 1990–2021

Nguồn: GSO [2]

Hình 3 cho thấy rằng năng suất lúa tăng ở các mùa vụ đều tăng qua giai đoạn 1990–2021. Trong đó, năng suất lúa vụ Đông Xuân cao nhất, theo sau là vụ Hè Thu – Thu Đông. Năng suất lúa Mùa thấp nhất. Theo số liệu thống kê của GSO [2], năng suất lúa ở vụ Đông Xuân tăng từ 3,79 tấn/ha năm 1990 lên 6,86 tấn/ha năm 2021, tương đương mức tăng bình quân giai đoạn 1,93%/năm. Năng suất lúa ở vụ Hè Thu – Thu Đông tăng từ 3,36 tấn/ha năm 1990 lên 5,67 tấn/ha năm 2021, tương đương mức tăng bình quân giai đoạn 1,70%/năm. Mặc dù có năng suất thấp hơn cả, nhưng năng suất lúa vụ Mùa có bước tăng cao hơn so với lúa vụ Đông Xuân và vụ Hè Thu – Thu Đông, với mức tăng bình quân giai đoạn là 2,19%/năm, từ 2,64 tấn/ha năm 1990 lên 5,17 tấn/ha năm 2021 (Bảng 1).

Sản lượng lúa vụ Đông Xuân và vụ Hè Thu – Thu Đông có sự tăng trưởng đáng kể qua giai đoạn 1990–2021 do sự tăng về quy mô diện tích và năng suất lúa. Cụ thể, sản lượng lúa vụ Đông Xuân tăng từ 7,87 triệu tấn năm 1990 lên 20,63 triệu tấn năm 2021, tương ứng với mức tăng bình quân giai đoạn là 3,16%/năm. Tương tự, sản lượng lúa vụ Hè Thu – Thu Đông tăng từ



Hình 3. Năng suất lúa phân theo mùa vụ giai đoạn 1990–2021

Nguồn: GSO [2]

4,09 triệu tấn năm 1990 lên 15,16 triệu tấn năm 2021, tương ứng với mức tăng trưởng bình quân giai đoạn là 4,32%/năm. Sản lượng lúa vụ Mùa mặc dù có sự tăng qua giai đoạn, nhưng mức tăng không đáng kể từ 7,23 triệu tấn năm 1990 lên 8,06 triệu tấn năm 2021 (Bảng 1), nguyên nhân là do sự sụt giảm mạnh về diện tích (Hình 2).

4.2 Dự báo diện tích, năng suất và sản lượng lúa đến năm 2030

Bước 1: Kiểm tra tính dừng của chuỗi dữ liệu

Thống kê mô tả chuỗi dữ liệu diện tích, năng suất và sản lượng lúa của giai đoạn 1990–2021 được thể hiện ở Bảng 2. Chuỗi dữ liệu sử dụng trong mô hình ARIMA được giả định là chuỗi dừng, vì vậy chúng ta cần phải xem xét chuỗi dữ liệu về diện tích, năng suất và sản lượng lúa giai đoạn 1990–2021 có dừng hay không. Để kiểm tra tính dừng của chuỗi dữ liệu, hai kiểm định phổ biến được dùng là: Augmented Dickey–Fuller (ADF) và Perron–Phillips (PP), hay còn được gọi là kiểm định đơn vị (unit root test). Kết quả kiểm định ADF và PP của chuỗi dữ liệu được thể hiện ở Bảng 3.

Bảng 2. Thống kê mô tả diện tích, năng suất và sản lượng lúa của Việt Nam giai đoạn 1990–2021

| Biến | Diện tích (triệu ha) | Năng suất (tấn/ha) | Sản lượng (triệu tấn) |
|--------------------|---------------------------------|-------------------------------|----------------------------------|
| Số quan sát | 32 | 32 | 32 |
| Giá trị trung bình | 7,299 | 4,757 | 35,050 |
| Độ lệch chuẩn | 0,471 | 0,919 | 8,285 |
| Giá trị nhỏ nhất | 6,043 | 3,113 | 19,225 |
| Giá trị lớn nhất | 7,903 | 6,058 | 45,091 |
| Trung vị | 7,441 | 4,892 | 35,897 |
| Skewness | -1,102 | -0,326 | -0,482 |
| Kurtosis | 3,398 | 1,739 | 1,962 |

Nguồn: GSO [2]

Kết quả kiểm định ADF và PP (Bảng 3) cho thấy chỉ có chuỗi diện tích có giá trị p -value có mức ý nghĩa bé hơn 5%, nên ta bác bỏ giả thiết H_0 : Chuỗi dữ liệu có nghiệm đơn vị, tức là chuỗi không có tính dừng, và chấp nhận giả thiết H_1 : Chuỗi dữ liệu không có nghiệm đơn vị, tức là chuỗi có tính dừng. Trong khi đó, chuỗi năng suất và sản lượng lúa có kết quả kiểm định ADF và PP với giá trị p -value lớn hơn 5%, tức là không thể bác bỏ giả thiết H_0 : Chuỗi dữ liệu có nghiệm đơn vị, tức là chuỗi không có tính dừng. Do đó, chúng ta cần khắc phục tính dừng của chuỗi dữ liệu năng suất và sản lượng.

Bảng 3. Kết quả kiểm định ADF và PP đối với chuỗi dữ liệu

| Chuỗi dữ liệu | Kiểm định | Giá trị t | p-value |
|----------------------|------------------|------------------|-----------------------------|
| Diện tích | ADF | -3,632 | 0,005 |
| | PP | -3,087 | 0,028 |
| Năng suất | ADF | -1,095 | 0,717 |
| | PP | -1,281 | 0,638 |
| Sản lượng | ADF | -2,359 | 0,154 |
| | PP | -2,477 | 0,121 |

Nguồn: Tính toán của tác giả

Khắc phục tính dừng của chuỗi dữ liệu năng suất và sản lượng lúa bằng cách lấy sai phân cấp 1 và tiến hành kiểm tra tính dừng của hai chuỗi này bằng hai kiểm định ADF và PP. Kết quả kiểm định ADF và PP được trình bày ở Bảng 4 cho thấy rằng sai phân cấp 1 của năng suất và sản lượng lúa có tính dừng (các giá trị p -values của kiểm định ADF và PP đều nhỏ hơn 1%).

Bước 2: Xác định mô hình ARIMA(p,d,q) tối ưu

Các giá trị p, q được xác định dựa vào biểu đồ hệ số tự tương quan (ACF) và bảng hệ số tự tương quan từng phần (PACF) [13, 22]. Trong nghiên cứu này, chúng tôi sử dụng hàm *auto.arima()* của *forecast* package trong phần mềm R [13] để tìm kiếm mô giá trị tối ưu. Kết quả kiểm tra cho thấy mô hình ARIMA(0,2,1) là thích hợp để dự báo cho diện tích lúa và sản lượng lúa. Trong khi đó, mô hình ARIMA(0,1,0) được lựa chọn để dự báo năng suất lúa.

Bước 3: Ước lượng và kiểm tra mô hình

Chúng tôi ước lượng mô hình ARIMA(0,2,1) cho chuỗi diện tích và sản lượng lúa, và sử dụng mô hình ARIMA(0,1,0) để ước lượng cho chuỗi năng suất lúa. Sau đó tiến hành kiểm tra mức độ phù hợp của mô hình với chuỗi dữ liệu nghiên cứu bằng cách phân tích phần dư của các mô hình đã ước lượng sử dụng biểu đồ ACF, histogram, normal Q-Q plot, và kiểm định portmanteau test. Kết quả ước lượng các mô hình ARIMA tối ưu cho diện tích, năng suất và sản lượng lúa được trình bày ở Bảng 5.

Bảng 4. Kết quả kiểm định ADF và PP đối với chuỗi dữ liệu sai phân cấp 1

| Chuỗi dữ liệu | Kiểm định | Giá trị t | p -value |
|---------------|-----------|-----------|------------|
| Năng suất | ADF | -6,448 | 0,000 |
| | PP | -6,471 | 0,000 |
| Sản lượng | ADF | -4,877 | 0,000 |
| | PP | -4,920 | 0,000 |

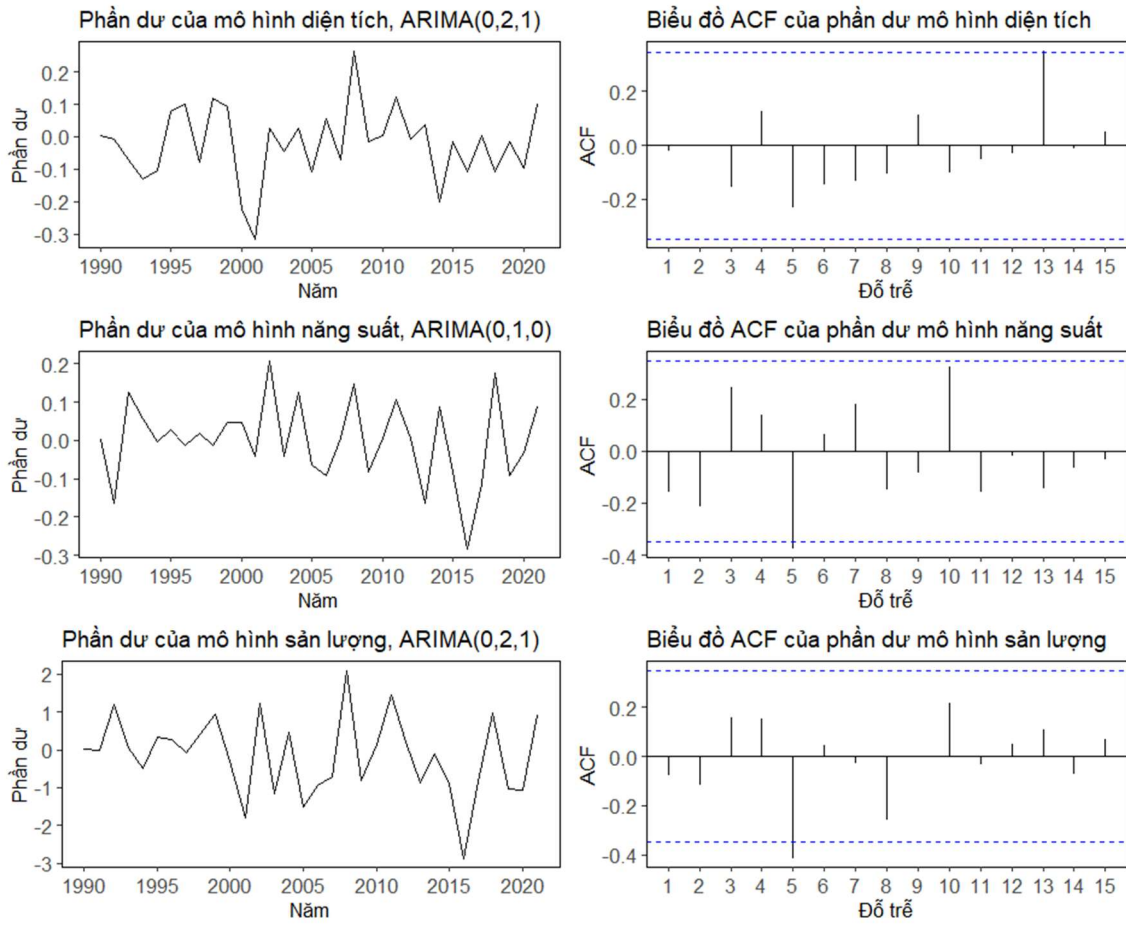
Nguồn: Tính toán của tác giả

Bảng 5. Kết quả ước lượng của các mô hình ARIMA tối ưu

| Chuỗi dữ liệu | Mô hình ARIMA tối ưu | Hệ số ước lượng và sai số chuẩn (S.E.) | AIC | BIC | Sigma ² | Log-likelihood |
|---------------|----------------------|--|--------|--------|--------------------|----------------|
| Diện tích | (0,2,1) | MA(1): -0,516 S.E.: 0,183 | -38,76 | -35,96 | 0,01 | 21,38 |
| Năng suất | (0,1,0) | Drift: 0,093 S.E.: 0,019 | -47,06 | -44,19 | 0,01 | 25,53 |
| Sản lượng | (0,2,1) | MA(1): -0,851 S.E.: 0,104 | 94,85 | 97,66 | 1,20 | -45,43 |

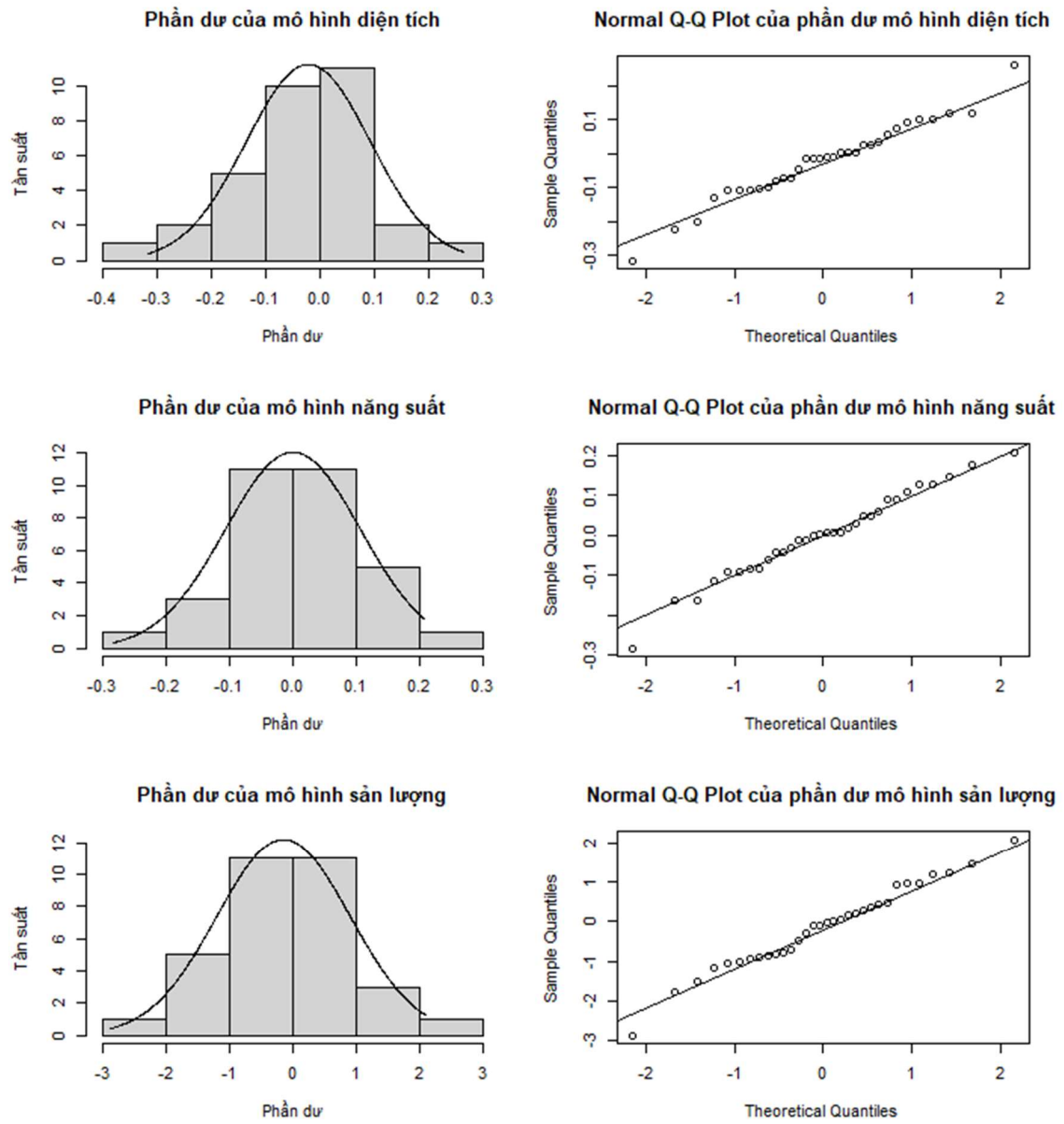
Nguồn: Tính toán của tác giả

Biểu đồ phân phối phần dư và biểu đồ ACF của phần dư của các mô hình ARIMA được lựa chọn cho chuỗi diện tích, năng suất và sản lượng lúa được trình bày ở Hình 4. Kết quả cho thấy rằng tất cả các tự tương quan đều nằm trong giới hạn, cho thấy phần dư hoạt động giống như sai số ngẫu nhiên trắng (White noise). Biểu đồ histogram và biểu đồ xác suất chuẩn (Normal Q-Q plot) ở Hình 5 cho thấy rằng phần dư của các mô hình được lựa chọn có phân phối chuẩn (normal distribution). Kết quả kiểm định portmanteau test có giá trị p (p -value) lớn, cũng xác nhận rằng phần dư của các mô hình được chọn là nhiễu trắng. Điều này khẳng định rằng kết quả dự báo diện tích, năng suất và sản lượng lúa dựa trên các mô hình được lựa chọn là phù hợp và đáng tin cậy.



Hình 4. Biểu đồ phân phối và ACF của phần dư của các mô hình được lựa chọn

Nguồn: Tính toán của tác giả



Hình 5. Biểu đồ histogram và Normal Q-Q plot của phần dư của các mô hình được lựa chọn

Nguồn: Tính toán của tác giả

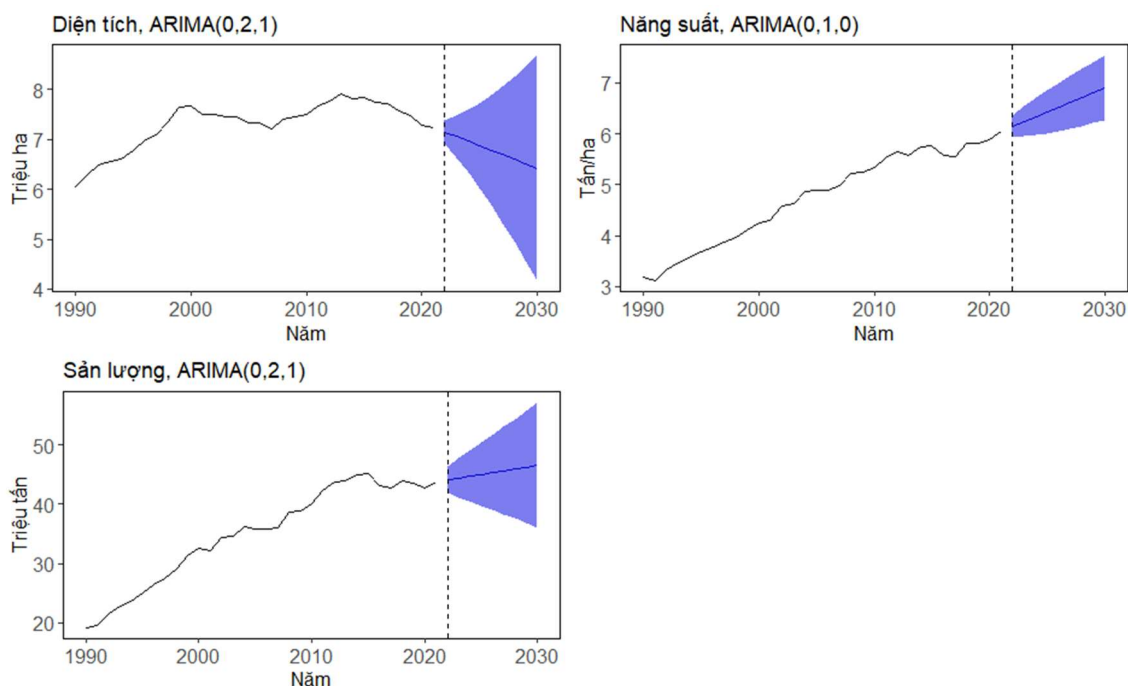
Bước 4: Dự báo

Giá trị dự báo và giá trị khoảng tin cậy 95% của diện tích, năng suất và sản lượng lúa cho giai đoạn 2022–2030 được trình bày ở Bảng 6 và Hình 6. Kết quả cho thấy rằng diện tích được dự báo có xu hướng giảm trong giai đoạn tới. Cụ thể, đến năm 2030 diện tích lúa giảm xuống còn 6,42 (4,17; 8,67) triệu ha, so với năm 2021 (7,24 triệu ha) diện tích lúa sẽ giảm đi khoảng 0,8 triệu ha, tương đương mức giảm bình quân khoảng 0,1 triệu ha/năm. Trong khi đó, năng suất và sản lượng lúa được dự báo có xu hướng tăng cho thập kỷ tới. Năng suất lúa sẽ tăng từ 6,06 tấn/ha năm 2021 lên đạt 6,90 (6,26; 7,53) tấn/ha năm 2030. Nhờ sự tăng về năng suất nên tổng sản lượng lúa sẽ tăng từ 43,85 triệu tấn năm 2021 lên đạt khoảng 46,60 (36,02; 57,19) triệu tấn năm 2030. Nghiên cứu cũng tiến hành mở rộng phạm vi thời gian dự báo đến năm 2050, kết quả cho thấy rằng diện tích tiếp tục giảm xuống còn khoảng 4,59 triệu ha; trong khi đó năng suất và sản lượng dự kiến đạt khoảng 8,75 tấn/ha và 52,72 triệu tấn.

Bảng 6. Kết quả dự báo diện tích, năng suất và sản lượng lúa của Việt Nam giai đoạn 2022–2030

| Năm | Diện tích | | | Năng suất | | | Sản lượng | | |
|------|----------------|--------------------|----------|----------------|--------------------|----------|----------------|--------------------|----------|
| | Giá trị dự báo | Khoảng tin cậy 95% | | Giá trị dự báo | Khoảng tin cậy 95% | | Giá trị dự báo | Khoảng tin cậy 95% | |
| | | Cận dưới | Cận trên | | Cận dưới | Cận trên | | Cận dưới | Cận trên |
| 2022 | 7,149 | 6,913 | 7,384 | 6,153 | 5,941 | 6,364 | 44,156 | 42,010 | 46,302 |
| 2023 | 7,057 | 6,636 | 7,479 | 6,246 | 5,947 | 6,545 | 44,462 | 41,192 | 47,732 |
| 2024 | 6,966 | 6,340 | 7,592 | 6,339 | 5,972 | 6,705 | 44,768 | 40,471 | 49,065 |
| 2025 | 6,875 | 6,023 | 7,726 | 6,432 | 6,008 | 6,855 | 45,074 | 39,770 | 50,377 |
| 2026 | 6,784 | 5,687 | 7,880 | 6,525 | 6,051 | 6,998 | 45,380 | 39,064 | 51,695 |
| 2027 | 6,692 | 5,332 | 8,053 | 6,617 | 6,099 | 7,136 | 45,686 | 38,341 | 53,030 |
| 2028 | 6,601 | 4,959 | 8,243 | 6,710 | 6,150 | 7,270 | 45,992 | 37,595 | 54,389 |
| 2029 | 6,510 | 4,570 | 8,450 | 6,803 | 6,205 | 7,402 | 46,298 | 36,822 | 55,773 |
| 2030 | 6,418 | 4,165 | 8,672 | 6,896 | 6,261 | 7,531 | 46,604 | 36,021 | 57,186 |

Nguồn: Tính toán của tác giả



Hình 6. Biểu đồ dự báo diện tích, năng suất và sản lượng lúa đến năm 2030

Nguồn: Tính toán của tác giả

Kết quả dự báo về diện tích, năng suất và sản lượng lúa cho thấy rằng, trong thập kỷ tới để gia tăng sản lượng để đáp ứng nhu cầu an ninh lương thực và duy trì vị thế xuất khẩu gạo đòi hỏi các chính sách cần phải tập trung vào việc tăng năng suất và giá trị sản phẩm đầu ra (tăng lợi nhuận). Những nghiên cứu về hiệu quả kỹ thuật canh tác lúa gần đây [1, 11, 12] cho thấy rằng vẫn còn tiềm năng để tăng năng suất lúa thông qua việc sử dụng và phân bổ các nguồn lực sản xuất một cách hiệu quả và ứng dụng giống lúa chất lượng cao để gia tăng lợi nhuận cho người trồng lúa và đáp ứng yêu cầu chất lượng của gạo xuất khẩu.

5 Kết luận

Nghiên cứu này phân tích thực trạng canh tác lúa ở Việt Nam giai đoạn 1990–2021 và tiến hành dự báo biến động diện tích, năng suất và sản lượng lúa đến năm 2030 sử dụng mô hình

Box–Jenkins ARIMA. Nghiên cứu sử dụng chuỗi dữ liệu thời gian giai đoạn 1990–2021 được tổng hợp từ Tổng Cục Thống Kê của Việt Nam.

Kết quả thống kê cho thấy năm 2021, tổng diện tích canh tác lúa đạt 7,24 triệu ha, với năng suất bình quân đạt 6,06 tấn/ha và tổng sản lượng lúa đạt 43,85 triệu tấn. Qua giai đoạn 1990–2021, năng suất và sản lượng lúa tăng, với mức tăng bình quân hàng năm tương ứng lần lượt là 2,10% và 2,70%; diện tích có tăng, nhưng tăng với tốc độ chậm hơn (0,58%) và có xu hướng giảm từ năm 2013 trở lại đây. Kết quả ước lượng cho thấy, mô hình ARIMA tối ưu để dự báo diện tích, năng suất và sản lượng lúa lần lượt là ARIMA(0,2,1), ARIMA(0,1,0) và ARIMA(0,2,1). Kết quả dự báo chỉ ra rằng đến năm 2030, diện tích lúa giảm khoảng 0,8 triệu ha, còn 6,42 (4,17; 8,67) triệu ha. Trong khi đó, năng suất và sản lượng lúa có xu hướng tăng và lần lượt đạt 6,90 (6,26; 7,53) tấn/ha và 46,60 (36,02; 57,19) triệu tấn. Nghiên cứu đề xuất rằng trong thập kỷ tới, các chính sách về sản xuất lúa cần tập trung thúc đẩy việc tăng năng suất và lợi nhuận thay vì mở rộng diện tích canh tác lúa. Cải thiện hiệu quả sản xuất và áp dụng công nghệ tiên tiến là những định hướng phù hợp.

Lời cảm ơn

Nghiên cứu này là một hợp phần của đề tài Đại học Huế, với mã số DHH2023-06-136.

Tài liệu tham khảo

1. Phuc, H. T., Burton, M., Ma, C., and Hailu, A. (2022), Quantifying heterogeneity, heteroscedasticity and publication bias effects on technical efficiency estimates of rice farming: A meta-regression analysis, *Journal of Agricultural Economics*, 73(2), 580–597.
2. GSO (2023), Diện tích, sản lượng và năng suất lúa cả năm của Việt Nam. Available: <https://www.gso.gov.vn/px-web-2/?pxid=V0612&theme=N%C3%B4ng%20%C3%A2m%20nghi%1%BB%87p%20v%C3%A0%20th%E1%BB%A7y%20s%E1%BA%A3n>.
3. FAOSTAT (2023), Area harvested, yield, and production quantity of global rice, paddy. Available: <http://www.fao.org/faostat/en/#data/QC>.
4. GSO (2023), Một số mặt hàng xuất khẩu chủ yếu. Available: <https://www.gso.gov.vn/px-web-2/?pxid=V0812&theme=Th%C6%B0%C6%A1ng%20m%E1%BA%A1i%20gi%C3%A1%20c%E1%BA%A3>.

5. Hien, N. T. M., Kawaguchi, T., and Suzuki, N. (2003), A study on technical efficiency of rice production in the Mekong Delta-Vietnam by stochastic frontier analysis, *J. Fac. Agr., Kyushu Univ.*, 48(1/2), 325–57.
6. Huy, H. T. (2009), Technical efficiency of rice-producing households in the Mekong Delta of Vietnam, *Asian Journal of Agriculture and Development*, 6(2), 35–50.
7. Khai, H. V. and Yabe, M. (2011), Technical efficiency analysis of rice production in Vietnam, *Journal of ISSAAS*, 17(1), 135–146.
8. Linh, V. H. (2012), Efficiency of rice farming households in Vietnam, *International Journal of Development Issues*, 11(1), 60–73.
9. Tung, D. T. (2013), Changes in the technical and scale efficiency of rice production activities in the Mekong Delta, Vietnam, *Agricultural and Food Economics*, 1(1), 1–11.
10. Linh, T. T., Nanseki, T., and Chomei, Y. (2015), Productive efficiency of crop farms in Viet Nam: A DEA with a smooth bootstrap application, *Journal of Agricultural Science*, 7(5), 37.
11. Phuc, H. T. and Napasintuwong, O. (2015), Profit inefficiency among hybrid rice farmers in Central Vietnam, *Agriculture and Agricultural Science Procedia*, 5, 89–95.
12. Phuc, H. T. (2021), *Profit Efficiency and Rice Variety Choice in Rice Farming in Vietnam: A Stochastic Frontier Analysis Approach*, Doctoral Thesis, UWA School of Agriculture and Environment, The University of Western Australia, Australia.
13. Hyndman, R. J. and Athanasopoulos, G. (2018), *Forecasting: Principles and Practice*, 2nd ed., Monash University, Australia.
14. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., Bergmeir, C., Bessa, R. J., Bijak, J. and Boylan, J. E. (2022), *Forecasting: Theory and practice*, *International Journal of Forecasting*.
15. Dương Văn Khảm (2000), Áp dụng phương pháp tích phân hồi quy bội và mô hình ARIMA trong việc dự báo năng suất lúa ở Hà Nội, *Tạp chí Khí tượng Thủy Văn*, 1–10.
16. Rahman, N. (2010), Forecasting of boro rice production in Bangladesh: An ARIMA approach, *Journal of the Bangladesh Agricultural University*, 8(1), 103–112.
17. Võ Văn Tài (2012), Dự báo sản lượng lúa Việt Nam bằng các mô hình toán học, *Tạp chí Khoa học Trường Đại học Cần Thơ*, 23b, 125–134.
18. Biswas, R. and Bhattacharyya, B. (2013), ARIMA modeling to forecast area and production of rice in West Bengal, *Journal of crop and weed*, 9(2), 26–31.

19. Hamjah, M. A. (2014), Rice production forecasting in Bangladesh: An application of Box-Jenkins ARIMA model, *Mathematical theory and modeling*, 4(4), 1–11.
20. Sahu, P., Mishra, P., Dhekale, B., Vishwajith, K., and Padmanaban, K. (2015), Modelling and forecasting of area, production, yield and total seeds of rice and wheat in SAARC countries and the world towards food security, *American Journal of Applied Mathematics and Statistics*, 3(1), 34–48.
21. Hemavathi, M. and Prabakaran, K. (2018), ARIMA model for forecasting of area, production and productivity of rice and its growth status in Thanjavur District of Tamil Nadu, India, *Int. J. Curr. Microbiol. App. Sci*, 7(2), 149–156.
22. Phạm Thị Thảo Hiền, Nguyễn Ngọc Ánh, Tôn Nữ Tuyết Trinh và Nguyễn Đức Hồng (2020), Ứng dụng mô hình ARIMA dự báo sản lượng lúa Tỉnh Thừa Thiên Huế đến năm 2025, *Tạp chí Khoa học và công nghệ nông nghiệp Trường Đại học Nông Lâm Huế*, 4(2), 1915–1921.
23. Noorunnahar, M., Chowdhury, A. H., and Mila, F. A. (2023), A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh, *PloS one*, 18(3), e0283452.
24. Đinh Phi Hồ và Nguyễn Văn Phương (2015), *Kinh tế phát triển: Cơ bản và nâng cao*, Nxb. Kinh tế TP. Hồ Chí Minh.
25. Box, G. E. P. and Jenkins, G. M. (1970), *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.