

PHÂN TÍCH VÀ DỰ BÁO SẢN XUẤT LÚA THEO MÙA VỤ CANH TÁC Ở VIỆT NAM ĐẾN NĂM 2030

Hồ Trọng Phúc¹, Phạm Minh Tâm², Hoàng Thị Liễu¹, Trần Minh Trí¹,
Nguyễn Thị Minh Phương¹

Ngày nhận bài: 18/08/2023

Ngày nhận bản sửa: 05/09/2023

Ngày duyệt đăng: 29/09/2023

Tóm tắt: Bài báo này tiến hành dự báo diện tích canh tác, năng suất và sản lượng lúa của Việt Nam theo mùa vụ đến năm 2030 sử dụng mô hình ARIMA. Nghiên cứu sử dụng dữ liệu chuỗi thời gian giai đoạn 1990–2021. Kết quả dự báo cho thấy đến năm 2030, diện tích canh tác lúa vụ Đông Xuân (ĐX) và vụ Mùa giảm lần lượt từ 3,01 triệu ha và 1,56 triệu ha xuống còn 2,75 triệu ha và 1,21 triệu ha. Trong khi đó, diện tích canh tác vụ Hè Thu và Thu Đông (HT–TĐ) tăng từ 2,67 triệu ha lên 3,09 triệu ha. Năng suất lúa các vụ được dự báo đều tăng lên, trong đó vụ ĐX tăng từ 6,86 tấn/ha lên 7,74 tấn/ha, vụ HT–TĐ tăng từ 5,67 tấn/ha lên 6,32 tấn/ha, và vụ Mùa tăng từ 5,17 tấn/ha lên 5,89 tấn/ha. Sản lượng lúa cũng được dự báo tăng, cụ thể vụ ĐX và vụ HT–TĐ có xu hướng tăng mạnh, lần lượt tăng từ 20,63 triệu tấn và 15,16 triệu tấn lên 24,33 triệu tấn và 18,37 triệu tấn. Trong khi đó, sản lượng vụ Mùa tăng, nhưng không đáng kể, ở mức 8,13 triệu tấn. Nghiên cứu đề xuất rằng cần có chính sách để hạn chế sự sụt giảm diện tích lúa vụ ĐX và vụ Mùa.

Từ khóa: Mô hình ARIMA; Dự báo; Sản xuất lúa; Mùa vụ canh tác; Việt Nam

1. Đặt vấn đề

Dự báo kinh tế đóng vai trò quan trọng trong quá trình kế hoạch hóa nền kinh tế. Dự báo là bước đầu tiên cung cấp thông tin để xác định mục tiêu và xây dựng chiến lược phát triển ngành (Hyndman & Athanasopoulos, 2021; Petropoulos & cs., 2022). Có nhiều phương pháp dự báo có thể áp dụng để dự báo trong kinh tế, tùy thuộc vào đối tượng dự báo và dữ liệu sẽ có các phương pháp dự báo phù hợp. Tổng quát, có năm cách tiếp cận để dự báo kinh tế dựa vào dữ liệu chuỗi thời gian, bao gồm: (1) phương pháp làm trơn hàm mũ (Exponential smoothing method), (2) mô hình hồi quy đơn phương trình (Single-equation regression models), (3) mô hình hồi quy phương trình đồng thời (Simultaneous-equation regression models), (4) mô hình tự hồi quy kết hợp trung bình trượt (AutoRegressive Integrated Moving Average – ARIMA models), và (5)

¹ Trường Đại học Kinh tế, Đại học Huế.

² Sở Kế hoạch Đầu Tư, Tỉnh An Giang, email: pminhtamts@gmail.com

mô hình tự hồi quy véctor (Vector AutoRegressive models) (Gujarati & Porter, 2022). Trong đó, mô hình dự báo ARIMA được áp dụng phổ biến trong các nghiên cứu thực nghiệm (Petropoulos & cs., 2022). Bởi vì mô hình ARIMA có ưu điểm là cho kết quả dự báo ngắn hạn đáng tin cậy so với các phương pháp dự báo khác (Trang, 2017). Các bước để thực hiện một nghiên cứu dự báo được trình bày chi tiết trong nghiên cứu của Hyndman & Athanasopoulos (2021) và Petropoulos & cs. (2022).

Mô hình ARIMA được áp dụng rộng rãi trong các nghiên cứu thực nghiệm về dự báo kinh tế ở Việt Nam và các quốc gia trên thế giới (Trang, 2017). Trong nông nghiệp, một số bài báo đã sử dụng mô hình ARIMA để dự báo sản xuất lúa ở Việt Nam và một số quốc gia trên thế giới (Khảm, 2000; Rahman, 2010; Tài, 2012; Hamjah, 2014; Sahu & cs., 2015; Hiền & cs., 2020; Phúc & Hùng, 2023). Tuy nhiên, chưa có nghiên cứu nào dự báo diện tích, năng suất và sản lượng lúa theo mùa vụ canh tác ở Việt Nam.

Kết quả thống kê GSO (2023) cho thấy có sự biến động mạnh về diện tích, năng suất và sản lượng lúa của Việt Nam theo mùa vụ canh tác trong những năm gần đây. Sự biến động khó lường đang gây ra những khó khăn trong vấn đề hoạch định chiến lược phát triển ngành lúa gạo. Để có cơ sở xây dựng chiến lược phát triển ngành trong giai đoạn tới, bài báo này tiến hành dự báo sự biến động diện tích, năng suất và sản lượng lúa của Việt Nam theo mùa vụ đến năm 2030 áp dụng mô hình ARIMA.

Nghiên cứu này có một số đóng góp như sau: (1) Đây là nghiên cứu đầu tiên tiến hành áp dụng mô hình ARIMA để dự báo sự biến động diện tích canh tác, năng suất và sản lượng lúa của Việt Nam theo mùa vụ canh tác. (2) Nghiên cứu cung cấp thông tin hữu ích cho các nhà hoạch định chính sách xác định mục tiêu và chiến lược phát triển ngành lúa gạo theo mùa vụ trong giai đoạn tới.

2. Phương pháp nghiên cứu

2.1. Phương pháp dự báo với mô hình ARIMA

Mô hình ARIMA, giới thiệu bởi Box & Jenkins (1970), được áp dụng rộng rãi để dự báo dữ liệu chuỗi thời gian, và mục đích của mô hình là để giải thích sự tự tương quan trong chuỗi dữ liệu nghiên cứu (Hyndman & Athanasopoulos, 2021; Petropoulos & cs., 2022). Mô hình ARIMA được tích hợp (I(d)) từ mô hình tự hồi quy (AR(p)) và mô hình trung bình trượt (MA(q)), được gọi là mô hình trung bình trượt tích hợp tự hồi quy. Phương trình tổng quát của mô hình ARIMA(p,d,q) được thể hiện như sau (Hyndman & Athanasopoulos, 2021):

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta \varepsilon_{t-1} \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1)$$

trong đó, y'_t là chuỗi dữ liệu hạ bậc, các giá trị y'_{t-p} và ε_{t-q} lần lượt là giá trị trễ của y'_t và ε_t . p là bậc tự hồi quy, d là bậc hạ bậc đầu tiên, và q là bậc trung bình trượt. Diễn đạt công thức (1) dưới dạng ký hiệu dịch chuỗi lùi (backshift notation) như sau:

$$(1 - \phi_1 B - \dots - \phi_p B^p) (1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (2)$$

Trong đó,

$(1 - \phi_1 B - \dots - \phi_p B^p)$ là quá trình tự hồi quy bậc p (AR(p));
 $(1 + \theta_1 B + \dots + \theta_q B^q)$ là quá trình trung bình trượt bậc q (MA(q));
 $(1 - B)^d y_t$ là sai phân bậc thứ d của chuỗi quan sát y_t ;
 B là toán tử lùi: $B y_t = y_{t-1}$, $B^m y_t = y_{t-m}$, $y_t - y_{t-1} = (1 - B)y_t$;
 ε_t là nhiễu trắng.

Nghiên cứu này sử dụng gói *forecast* của Hyndman & Athanasopoulos (2021) để ước lượng mô hình dự báo. Trong gói *forecast* của Hyndman & Athanasopoulos (2021), phần mềm *R* sử dụng tham số hơi khác, công thức (2) được viết lại như sau:

$$(1 - \phi_1 B - \dots - \phi_p B^p) (y'_t - \mu) = (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t \quad (3)$$

trong đó, $y'_t = (1 - B)^d y_t$ và μ là giá trị trung bình của y'_t . Để chuyển đổi qua công thức (3), chúng ta đặt $c = \mu(1 - \phi_1 - \dots - \phi_p)$.

Quy trình ước lượng mô hình ARIMA bao gồm 7 bước, được trình bày trong nghiên cứu của Hyndman & Athanasopoulos (2021). Trong đó, bước 3–5 có thể được thực hiện tự động nếu sử dụng hàm *auto.arima()*. Trong bài báo này, chúng tôi áp dụng hàm *auto.arima()* để dự báo, nên các bước thực hiện được gói gọn như sau:

Bước 1: Kiểm tra tính dừng của dữ liệu

Theo Gujarati & Porter (2022), trước khi sử dụng các kiểm định chính thức, chúng ta nên phân tích biểu đồ bằng cách vẽ biểu đồ chuỗi thời gian đang nghiên cứu để kiểm tra xem liệu rằng dữ liệu chuỗi thời gian nghiên cứu có tính dừng hay không. Một kiểm định đơn giản về tính dừng của dữ liệu chuỗi thời gian có thể áp dụng là hàm tự tương quan (ACF). Một kiểm định tính dừng (hoặc không cố định) của chuỗi thời gian đã trở nên phổ biến rộng rãi trong nhiều năm qua là kiểm định đơn vị (Unit root test). Có một số kiểm định đơn vị có thể áp dụng, trong đó kiểm định Dickey-Fuller (DF), kiểm định Dickey-Fuller tăng cường (Augmented Dickey-Fuller), kiểm định đơn vị Phillips-Perron (PP) thường được sử dụng (xem chi tiết về các kiểm định trong nghiên cứu của Gujarati & Porter (2022)). Trong bài báo này, chúng tôi áp dụng kiểm định Kwiatkowski-Phillips-Schmidt-Shin (KPSS) (Kwiatkowski & cs., 1992). Trong kiểm định này, giả thuyết H_0 : là dữ liệu có tính dừng, và chúng tôi tìm kiếm bằng chứng cho thấy giả thuyết H_0 là sai. Do đó, các giá trị p nhỏ (ví dụ: nhỏ hơn 0,05) có nghĩa rằng cần phải lấy sai phân chuỗi dữ liệu. Việc kiểm tra có thể được tính toán bằng hàm *ur.kpss()* từ gói *urca* trong phần mềm *R*. Nếu giả thiết H_0 bị bác bỏ, điều này có nghĩa là chuỗi dữ liệu không có tính dừng. Chúng ta lấy sai phân chuỗi dữ liệu, và áp dụng lại kiểm định KPSS cho tới khi chuỗi dữ liệu sai phân có tính dừng (Hyndman & Athanasopoulos, 2021).

Bước 2: Xác định mô hình ARIMA phù hợp

Nhận diện mô hình ARIMA phù hợp tức là xác định các giá trị thích hợp của p , d và q . Các công cụ chính để xác định giá trị phù hợp của p và q là hàm tự tương quan

(ACF), hàm tự tương quan từng phần (PACF) và các biểu đồ tương quan thu được, đơn giản là đồ thị của ACF và PACF theo độ dài độ trễ (Gujarati & Porter, 2022). Gần đây việc xác định giá trị phù hợp của p , d và q cho mô hình ARIMA được thực hiện dễ dàng nhờ sử dụng hàm *auto.arima()* của gói *forecast*, phát triển bởi Hyndman & cs. (2020), trong phần mềm R (Hyndman & Athanasopoulos, 2021; Petropoulos & cs., 2022). Trong nghiên cứu này chúng tôi sử dụng hàm *auto.arima()* để xác định giá trị p , d và q phù hợp cho mô hình ARIMA.

Bước 3: Ước lượng và kiểm tra mô hình được chọn

Sau khi xác định được giá trị p , d và q phù hợp, chúng ta tiến hành ước lượng mô hình ARIMA(p,d,q) được chọn. Hệ số của mô hình được ước lượng bằng phương pháp maximum likelihood estimation (MLE) (Hyndman & Athanasopoulos, 2021). 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 hiển 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, thì thử lại với mô hình ARIMA có sự điều chỉnh. Nếu phần dư của mô hình giống sai số ngẫu nhiên trắng thì tiến hành bước tiếp theo đó là dự báo (Hyndman & Athanasopoulos, 2021). Bên cạnh đó, có thể sử dụng Ljung Box để kiểm tra phần dư của mô hình được chọn. Ljung Box có giá trị mức ý nghĩa 5%, nếu giá trị p -value thấp hơn 0,05 chúng ta bác bỏ H_0 , có nghĩa rằng giá trị tổng bình phương sai số (sum of squared errors – SSE) lớn (hay mô hình không tốt), nếu giá trị p -value lớn hơn 0,05 thì chúng ta chấp nhận H_0 , có nghĩa rằng giá trị SSE nhỏ (hay mô hình tốt). Chúng ta cũng có thể sử dụng đồ thị Histogram và Normal Q-Q plot để kiểm tra liệu rằng phần dư của các mô hình ARIMA được chọn có phân phối chuẩn hay không.

Bước 4: Dự báo

Dựa trên kết quả ước lượng của mô hình ARIMA(p,d,q) được chọn, chúng ta tiến hành dự báo và xác định khoảng tin cậy của các giá trị dự báo (cách thức tính giá trị dự báo có thể xem Hyndman & Athanasopoulos (2021)).

2.2. Dữ liệu

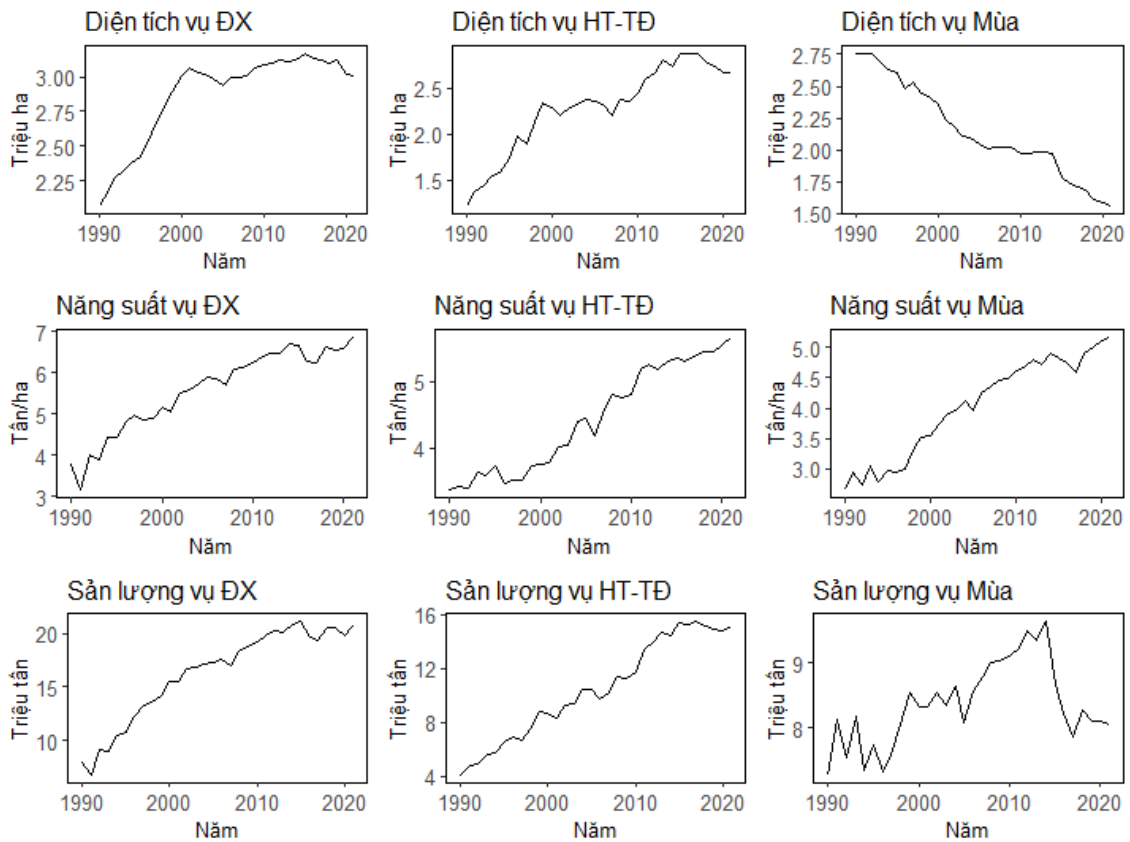
Nghiên cứu này sử dụng dữ liệu chuỗi thời gian về diện tích canh tác, năng suất và sản lượng lúa của Việt Nam theo mùa vụ giai đoạn 1990–2021, được tổng hợp từ Tổng Cục Thống Kê của Việt Nam (GSO). Kết quả thống kê mô tả về diện tích canh tác, năng suất và sản lượng lúa theo mùa vụ canh tác được thể hiện ở Bảng 1.

Bảng 1: Diện tích, năng suất và sản lượng lúa theo mùa vụ canh tác, giai đoạn 1990 – 2021

Năm	1990	1995	2000	2005	2010	2015	2021	So sánh giai đoạn 2021/1990		
								Tăng trưởng giá trị tuyệt đối	% tăng trưởng giai đoạn	% tăng trưởng bình quân năm
<i>Diện tích canh tác (triệu ha)</i>										
Vụ ĐX	2,074	2,421	3,013	2,942	3,086	3,168	3,007	0,933	45,006	1,206
Vụ HT–TĐ	1,216	1,742	2,293	2,349	2,436	2,869	2,673	1,458	119,914	2,575
Vụ Mùa	2,754	2,602	2,360	2,038	1,968	1,791	1,559	−1,195	−43,399	−1,819
<i>Năng suất (tấn/ha)</i>										
Vụ ĐX	3,793	4,434	5,168	5,891	6,227	6,658	6,861	3,067	80,866	1,930
Vụ HT–TĐ	3,365	3,731	3,762	4,442	4,797	5,347	5,672	2,307	68,566	1,699
Vụ Mùa	2,640	2,969	3,531	3,958	4,627	4,834	5,169	2,529	95,800	2,191
<i>Sản lượng (triệu tấn)</i>										
Vụ ĐX	7,866	10,737	15,571	17,332	19,217	21,092	20,629	12,763	162,266	3,159
Vụ HT–TĐ	4,091	6,501	8,625	10,436	11,686	15,341	15,163	11,073	270,700	4,317
Vụ Mùa	7,269	7,726	8,333	8,065	9,103	8,658	8,056	0,787	10,824	0,332

Nguồn: GSO (2023)

Đánh giá chung qua giai đoạn, cả diện tích, năng suất và sản lượng lúa đều tăng. Tuy nhiên, tổng diện tích canh tác và sản lượng lúa có xu hướng giảm trong những năm gần đây (Hình 1). Có thể thấy rằng diện tích canh tác lúa vụ ĐX và HT–TĐ có chiều hướng tăng, trong khi đó diện tích canh tác lúa vụ Mùa giảm mạnh. Theo GSO (2023), diện tích canh tác lúa vụ ĐX chiếm diện tích cao nhất và tăng 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 canh tác lúa vụ HT–TĐ 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 canh tác 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 1: Đồ thị diện tích, năng suất và sản lượng lúa của Việt Nam theo mùa vụ canh tác, giai đoạn 1990–2021

Nguồn: GSO (2023)

Hình 1 cho thấy rằng năng suất lúa ở các mùa vụ mặc dù có sự biến động, nhưng đều tăng qua giai đoạn 1990–2021. Trong đó, năng suất lúa vụ ĐX cao nhất, theo sau là vụ HT–TĐ, và năng suất lúa vụ Mùa thấp nhất. Theo số liệu thống kê của GSO (2023), năng suất lúa ở vụ ĐX 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 HT–TĐ 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 năng suất thấp hơn, nhưng năng suất lúa vụ Mùa có bước tăng cao hơn so với lúa vụ ĐX và vụ HT–TĐ, 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.

Sản lượng lúa vụ ĐX và HT–TĐ có sự tăng trưởng đáng kể qua giai đoạn 1990–2021 do sự tăng về quy mô diện tích canh tác và năng suất lúa. Cụ thể, sản lượng lúa vụ ĐX tăng từ 7,87 triệu tấn năm 1990 lên 20,63 triệu tấn năm 2021, tương ứng mức tăng bình quân giai đoạn là 3,16%/năm. Sản lượng lúa vụ HT–TĐ tăng từ 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 có sự biến động mạnh qua giai đoạn. Mặc dù có sự tăng trưởng, nhưng mức tăng không đáng kể, từ 7,23 triệu tấn năm 1990 lên 8,06 triệu tấn năm 2021. Nguyên nhân là do diện tích canh tác lúa Mùa giảm mạnh.

3. Kết quả nghiên cứu

3.1. Kiểm tra tính dừng của chuỗi dữ liệu

Bảng 1 và Hình 1 trình bày thống kê mô tả chuỗi dữ liệu diện tích canh tác, năng suất và sản lượng lúa của Việt Nam theo mùa vụ giai đoạn 1990–2021. Điều kiện dữ liệu chuỗi thời gian sử dụng trong mô hình ARIMA phải là chuỗi dừng, do đó, chúng ta cần phải kiểm tra chuỗi dữ liệu về diện tích canh tác, năng suất và sản lượng lúa theo mùa vụ có dừng hay không. Kết quả kiểm định KPSS cho thấy tất cả các chuỗi dữ liệu diện tích, năng suất và sản lượng lúa theo mùa vụ không có tính dừng, và phải lấy sai phân. Sau đó, chúng tôi tiến hành kiểm định KPSS đối chuỗi dữ liệu sai phân, kết quả cho thấy tất cả các chuỗi dừng ở bậc một, ngoại trừ chuỗi dữ liệu về diện tích vụ ĐX dừng ở sai phân bậc hai (giá trị d ở Bảng 2).

3.2. Xác định mô hình ARIMA(p,d,q) phù hợp nhất

Sử dụng hàm *auto.arima()*, chúng tôi xác định được giá trị p , d , và q phù hợp cho các chuỗi dữ liệu như sau: Đối với chuỗi dữ liệu về diện tích lúa, mô hình tối ưu để dự báo diện tích lúa vụ ĐX là ARIMA(0,2,1) và vụ HT–TĐ và vụ Mùa là ARIMA(0,1,0). Đối với chuỗi dữ liệu năng suất, mô hình tối ưu để dự báo năng suất vụ ĐX và vụ Mùa là ARIMA(1,1,0) và vụ HT–TĐ là ARIMA(0,1,1). Đối với chuỗi dữ liệu sản lượng lúa, mô hình tối ưu để dự báo sản lượng lúa vụ ĐX là ARIMA(1,1,0), vụ HT–TĐ là ARIMA(0,1,0), và vụ Mùa là ARIMA(1,1,1).

3.3. Ước lượng và kiểm định mô hình

– Ước lượng mô hình

Chúng tôi tiến hành ước lượng các mô hình ARIMA(p,d,q) được chọn cho các chuỗi dữ liệu ở mục 3.2. Kết quả ước lượng các mô hình được chọn được trình bày ở Bảng 2.

– Kiểm định mô hình

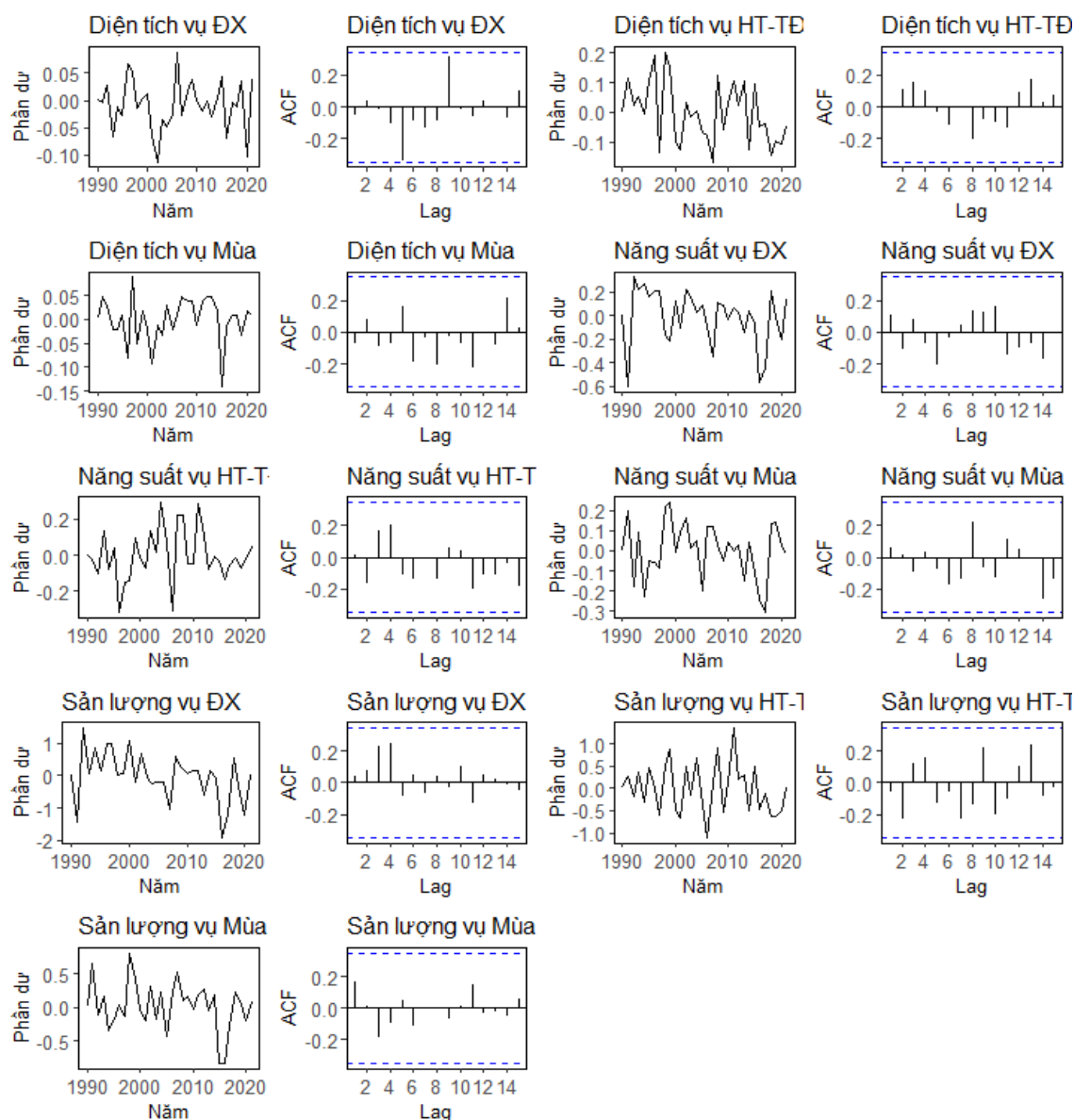
Bảng 2: Kết quả ước lượng mô hình ARIMA(p,d,q) được chọn

Chuỗi dữ liệu	Mô hình ARIMA phù hợp nhất	Hệ số ước lượng	Sai số chuẩn	AIC	AICc	BIC	Sigma2	Log-likelihood	
Diện tích canh tác									
Vụ ĐX	ARIMA(0,2,1)	MA(1):	−0,496****	0,182	−94,170	−93,730	−91,370	0,002	49,090
Vụ HT–TĐ	ARIMA(0,1,0)	Drift:	0,047****	0,018	−49,140	−48,710	−46,270	0,011	26,570
Vụ Mùa	ARIMA(0,1,0)	Drift:	−0,038****	0,008	−98,450	−98,020	−95,580	0,002	51,230
Năng suất									
Vụ ĐX	ARIMA(1,1,0) with drift	AR(1):	−0,595****	0,167	4,400	5,290	8,710	0,059	0,800
		Drift:	0,106****	0,027					
Vụ HT–TĐ	ARIMA(0,1,1) with drift	MA(1):	−0,393****	0,161	−26,250	−25,360	−21,950	0,022	16,130
		Drift:	0,075****	0,016					
Vụ Mùa	ARIMA(1,1,0) with drift	AR(1):	−0,452****	0,162	−28,610	−27,720	−24,300	0,020	17,300
		Drift:	0,079****	0,017					
Sản lượng									
Vụ ĐX	ARIMA(1,1,0) with drift	AR(1):	−0,314**	0,179	76,890	77,780	81,190	0,614	−35,440
		Drift:	0,421****	0,104					
Vụ HT–TĐ	ARIMA(0,1,0) with drift	Drift:	0,357****	0,100	55,520	55,950	58,390	0,319	−25,760
Vụ Mùa	ARIMA(1,1,1)	AR(1):	−0,917****	0,105	30,900	31,780	35,200	0,137	−12,450
		MA(1):	0,601****	0,166					

Chú thích: ***, **, * ký hiệu mức ý nghĩa thống kê tương ứng 1%, 5% và 10%. With drift là với một hằng số.

Nguồn: Tính toán của tác giả

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, kiểm định portmanteau test, và kiểm định Ljung Box.

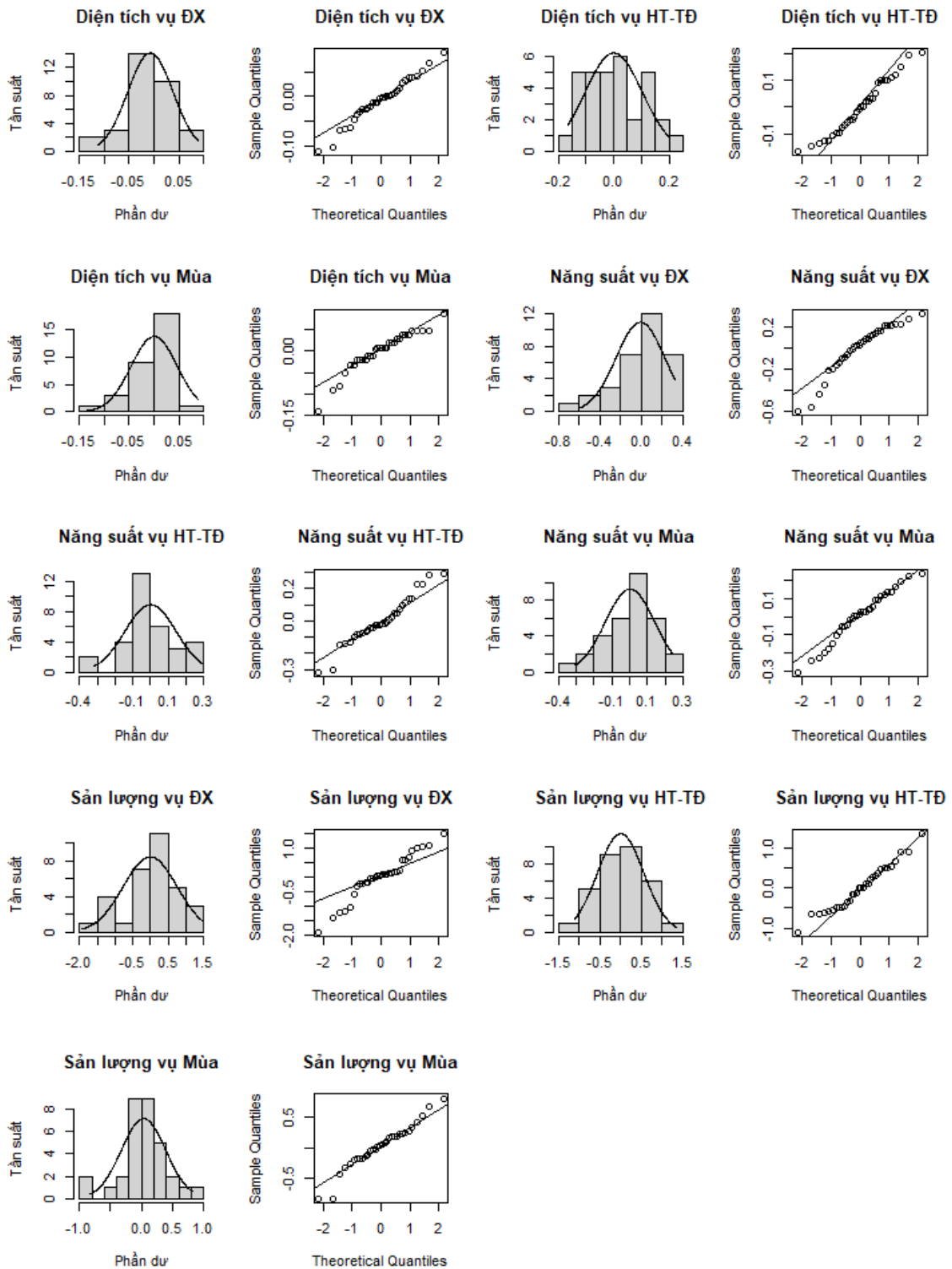


Hình 2: Đồ thị phân bố và ACF của phần dư của các mô hình ARIMA được chọn

Nguồn: Tính toán của tác giả

+ Đồ thị ACF của phần dư

Hình 2 trình bày biểu đồ ACF và biểu đồ phân phối phần dư của các mô hình ARIMA được lựa chọn. Kết quả cho thấy rằng tất cả các tự tương quan đều nằm trong giới hạn và có phân phối chuẩn, cho thấy phần dư hoạt động giống như sai số ngẫu nhiên trắng.



Hình 3: Đồ thị Histogram và Normal Q-Q plot của phần dư của các mô hình ARIMA được chọn

Nguồn: Tính toán của tác giả

Kết quả kiểm định portmanteau test có giá trị p (p -value) lớn, cũng khẳng định rằng phần dư của các mô hình được chọn là nhiễu trắng. Điều này cho phép thực hiện bước tiếp theo là dự báo diện tích canh tác, năng suất và sản lượng lúa theo mùa vụ dựa trên kết quả của những mô hình này.

+ *Kiểm định Ljung Box cho phần dư*

Bảng 3 trình bày giá trị thống kê của kiểm định Ljung Box cùng với giá trị p -value của từng mô hình dự báo tương ứng được chọn. Kết quả cho thấy rằng, giá trị p -value của tất cả các mô hình đều lớn hơn 0,05 và chúng ta chấp nhận H_0 , đồng nghĩa rằng tất cả các mô hình ARIMA được chọn là phù hợp.

Hình 3 trình bày đồ thị Histogram và Normal Q-Q plot phân phối của phần dư của các mô hình ARIMA được chọn. Kết quả cho thấy phần dư của các mô hình được chọn có phân phối chuẩn, có nghĩa rằng các mô hình được chọn là phù hợp.

Bảng 3: Kết quả kiểm định Ljung Box của các chuỗi dữ liệu

STT	Mô hình	Giá trị Q*	P-value
1	Diện tích lúa vụ ĐX ARIMA(0,2,1)	5,80	0,33
2	Diện tích lúa vụ HT-TĐ ARIMA(0,1,0)	2,40	0,88
3	Diện tích lúa vụ Mùa ARIMA(0,1,0)	3,57	0,74
4	Năng suất lúa vụ ĐX ARIMA(1,1,0) with drift	3,00	0,70
5	Năng suất lúa vụ HT-TĐ ARIMA(0,1,1) with drift	4,81	0,44
6	Năng suất lúa vụ Mùa ARIMA(1,1,0) with drift	2,00	0,85
7	Sản lượng lúa vụ ĐX ARIMA(1,1,0) with drift	5,06	0,41
8	Sản lượng lúa vụ HT-TĐ ARIMA(0,1,0) with drift	4,44	0,62
9	Sản lượng lúa vụ Mùa ARIMA(1,1,1)	3,21	0,52

Nguồn: Tính toán của tác giả

3.4. Kết quả dự báo

Bảng 4 và Hình 4 trình bày kết quả dự báo và giá trị khoảng tin cậy 95% của giá trị dự báo về diện tích canh tác, năng suất và sản lượng lúa theo mùa vụ canh tác đến năm 2030. Kết quả dự báo cho thấy rằng có sự biến động khác nhau về diện tích canh tác, năng suất và sản lượng lúa theo mùa vụ.

Đối với vụ ĐX, diện tích canh tác lúa được dự báo giảm từ 3,01 triệu ha năm 2021 xuống còn 2,75 triệu ha năm 2030, với giá trị khoảng tin cậy 95% là 1,82 – 3,67 triệu ha. Ngược lại, năng suất được dự báo tiếp tục tăng từ 6,86 tấn/ha năm 2021 lên 7,74 tấn/ha, với giá trị khoảng tin cậy là 6,79 – 8,70 tấn/ha. Mặc dù có sự sụt giảm về diện tích canh tác, nhưng do có sự gia tăng về năng suất nên sản lượng lúa được dự báo tăng từ 20,63 triệu tấn năm 2021 lên 24,33 triệu tấn, với khoảng tin cậy 95% là 20,72 – 27,95 triệu tấn.

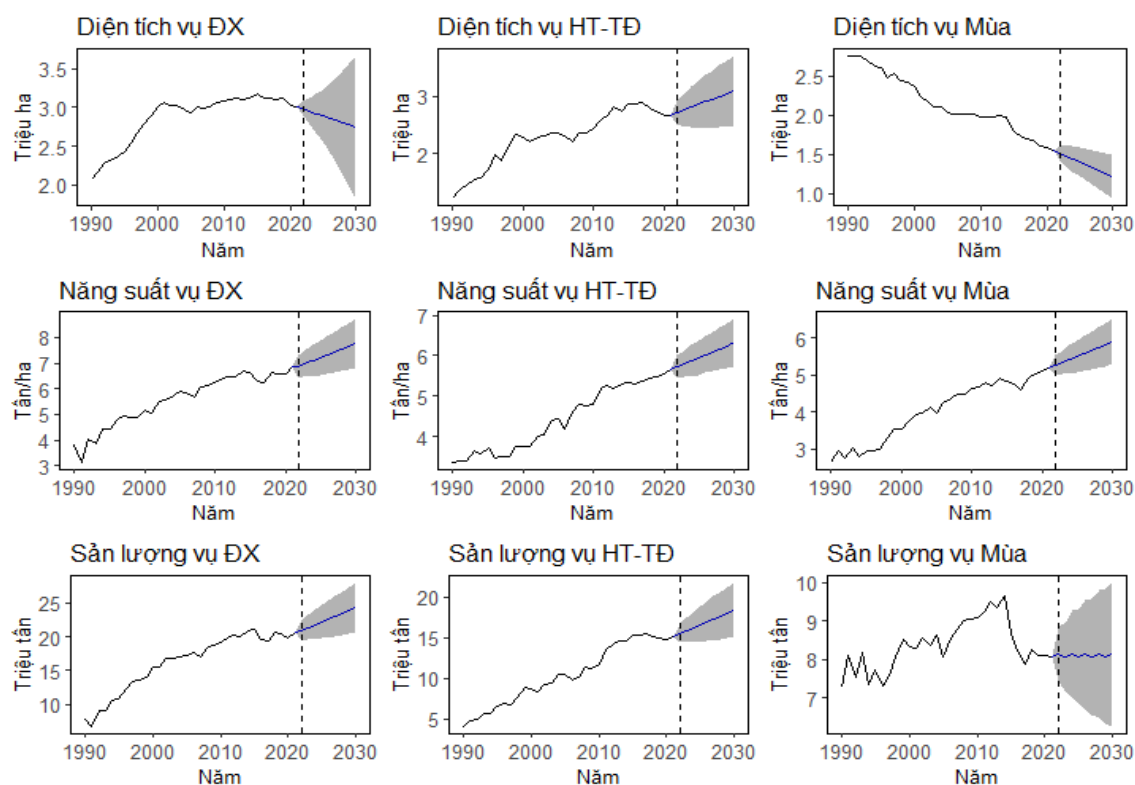
Bảng 4: Dự báo diện tích, năng suất và sản lượng lúa theo vụ đến năm 2030

Năm	Diện tích canh tác			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
Vụ ĐX									
2022	2,98	2,89	3,07	6,86	6,38	7,33	20,94	19,41	22,48
2023	2,95	2,78	3,12	7,03	6,52	7,54	21,40	19,54	23,26
2024	2,92	2,67	3,17	7,09	6,47	7,72	21,81	19,59	24,03
2025	2,89	2,55	3,24	7,22	6,55	7,90	22,23	19,73	24,73
2026	2,86	2,42	3,31	7,32	6,57	8,07	22,65	19,89	25,41
2027	2,83	2,28	3,39	7,43	6,63	8,23	23,07	20,07	26,07
2028	2,81	2,14	3,48	7,53	6,67	8,39	23,49	20,27	26,71
2029	2,78	1,98	3,57	7,64	6,73	8,55	23,91	20,49	27,34
2030	2,75	1,82	3,67	7,74	6,79	8,70	24,33	20,72	27,95
Vụ HT-TĐ									
2022	2,72	2,51	2,92	5,73	5,44	6,02	15,52	14,41	16,62
2023	2,76	2,47	3,05	5,80	5,46	6,14	15,87	14,31	17,44
2024	2,81	2,46	3,16	5,88	5,49	6,26	16,23	14,31	18,15
2025	2,86	2,45	3,27	5,95	5,53	6,37	16,59	14,37	18,80
2026	2,90	2,45	3,36	6,02	5,57	6,48	16,95	14,47	19,42
2027	2,95	2,45	3,45	6,10	5,61	6,59	17,30	14,59	20,01
2028	3,00	2,46	3,54	6,17	5,65	6,69	17,66	14,73	20,59
2029	3,04	2,47	3,62	6,25	5,70	6,80	18,02	14,89	21,15
2030	3,09	2,48	3,70	6,32	5,74	6,90	18,37	15,05	21,69
Vụ Mùa									
2022	1,52	1,43	1,61	5,26	4,98	5,54	8,15	7,42	8,87
2023	1,48	1,35	1,61	5,33	5,02	5,65	8,07	7,19	8,95
2024	1,44	1,28	1,60	5,42	5,03	5,80	8,14	7,02	9,27
2025	1,41	1,22	1,59	5,49	5,07	5,92	8,07	6,84	9,31
2026	1,37	1,16	1,57	5,57	5,11	6,04	8,14	6,72	9,55
2027	1,33	1,10	1,56	5,65	5,15	6,16	8,08	6,57	9,59
2028	1,29	1,05	1,54	5,73	5,19	6,27	8,13	6,48	9,79
2029	1,25	0,99	1,51	5,81	5,24	6,39	8,08	6,34	9,82
2030	1,21	0,94	1,49	5,89	5,29	6,50	8,13	6,26	9,99

Nguồn: Tính toán của tác giả

Đối với vụ HT–TĐ, diện tích canh tác lúa được dự báo tăng từ 2,67 triệu ha năm 2021 lên 3,09 triệu ha năm 2030, với giá trị khoảng tin cậy 95% là 2,48 – 3,70 triệu ha. Năng suất cũng được dự báo tiếp tục tăng từ 5,67 tấn/ha năm 2021 lên 6,32 tấn/ha, với giá trị khoảng tin cậy là 5,74 – 6,90 tấn/ha. Nhờ có sự gia tăng về diện tích canh tác và năng suất nên sản lượng lúa vụ HT–TĐ được dự báo tiếp tục tăng từ 15,16 triệu tấn năm 2021 lên 18,37 triệu tấn, với khoảng tin cậy 95% là 15,05 – 21,69 triệu tấn.

Đối với vụ Mùa, diện tích canh tác được dự báo giảm mạnh từ 1,56 triệu ha năm 2021 xuống còn 1,21 triệu ha năm 2030, với giá trị khoảng tin cậy 95% là 0,94 – 1,49 triệu ha. Tuy nhiên, năng suất được dự báo tiếp tục tăng từ 5,17 tấn/ha năm 2021 lên 5,89 tấn/ha, với giá trị khoảng tin cậy là 5,29 – 6,50 tấn/ha. Sản lượng lúa vụ Mùa được dự báo tăng nhẹ từ 8,06 triệu tấn năm 2021 lên 8,13 triệu tấn, với khoảng tin cậy 95% là 6,26 – 9,99 triệu tấn.



Hình 4: Đồ thị dự báo diện tích, năng suất và sản lượng lúa theo vụ đến năm 2030

Nguồn: Tính toán của tác giả

4. Kết luận và khuyến nghị

Bài báo này dự báo sự biến động của diện tích, năng suất và sản lượng lúa của Việt Nam theo mùa vụ canh tác đến năm 2030 sử dụng mô hình dự báo ARIMA. Nghiên cứu sử dụng dữ liệu chuỗi thời gian giai đoạn 1990–2021 được tổng hợp từ GSO (2023). Kết quả dự báo cho thấy rằng đến năm 2030, diện tích canh tác lúa vụ ĐX được dự báo giảm xuống còn 2,75 [1,82 – 3,67] triệu ha. Ngược lại, năng suất được dự báo tiếp tục tăng và đạt 7,74 [6,79 – 8,70] tấn/ha. Mặc dù có sự sụt giảm về diện tích

canh tác, nhưng do có sự gia tăng về năng suất nên sản lượng lúa vụ ĐX được dự báo tăng và đạt 24,33 [20,72 – 27,95] triệu tấn. Đối với vụ HT–TĐ, diện tích canh tác lúa được dự báo tăng lên 3,09 [2,48 – 3,70] triệu ha. Năng suất cũng được dự báo tiếp tục tăng lên 6,32 [5,74 – 6,90] tấn/ha. Nhờ có sự gia tăng về diện tích canh tác và năng suất nên sản lượng lúa vụ HT–TĐ được dự báo tiếp tục tăng lên 18,37 [15,05 – 21,69] triệu tấn. Đối với vụ Mùa, diện tích canh tác được dự báo giảm mạnh xuống còn 1,21 [0,94 – 1,49] triệu ha. Tuy nhiên, năng suất được dự báo tiếp tục tăng lên đạt 5,89 [5,29 – 6,50] tấn/ha. Sản lượng lúa vụ Mùa được dự báo tăng nhẹ, đạt 8,13 [6,26 – 9,99] triệu tấn.

Nghiên cứu đề xuất rằng các chính sách cần bám sát theo mùa vụ và cần tập trung duy trì ổn định diện tích canh tác lúa vụ ĐX và vụ Mùa, bên cạnh đó cần tập trung nâng cao năng suất và chất lượng sản phẩm để đảm bảo vấn đề an ninh lương thực, đảm bảo mục tiêu xuất khẩu và gia tăng giá trị lợi nhuận cho người trồng lúa.

TÀI LIỆU THAM KHẢO

- Box, G.E.P. and Jenkins, G.M. (1970). *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
- GSO (2023). Diện tích, sản lượng và năng suất lúa cả năm của Việt Nam, 1990-2021. Tổng Cục Thống Kê Việt Nam.
- Gujarati, D.N. and Porter, D.C. (2022). Chapter 22: Time Series Econometrics: Forecasting. in Gujarati, D.N. and Porter, D.C. (eds.), *Basic econometrics*. Douglas Reiner, New York, 773-837.
- Hamjah, M.A. (2014). Rice production forecasting in Bangladesh: An application of Box-Jenkins ARIMA model. *Mathematical theory and modeling* 4: 1-11.
- Hiền, P.T.T., Ánh, N.N., Trinh, T.N.T. and Hồng, N.Đ. (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: 1915-1921.
- Hyndman, R.J. and Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*, Monash University, Australia.
- Hyndman, R.J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. and Yasmeeen, F. (2020). Forecast: Forecasting functions for time series and linear models.
- Khảm, D.V. (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.
- Kwiatkowski, D., Phillips, P.C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54: 159-178.
- 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*.

Phúc, H.T. and Hùng, P.X. (2023). Dự báo diện tích, năng suất và sản lượng lúa của Việt Nam: Áp dụng mô hình ARIMA. *Tạp chí Khoa học Đại học Huế: Kinh tế và Phát triển* 132: 85–104.

Rahman, N. (2010). Forecasting of boro rice production in Bangladesh: An ARIMA approach. *Journal of the Bangladesh Agricultural University* 8: 103-112.

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: 34-48.

Tài, V.V. (2012). Dự báo sản lượng lúa Việt Nam bằng các mô hình toán học. *Tạp chí Khoa học Trường Đại học Cần Thơ* 125-134.

Trang, N.T.T. (2017). Sử dụng mô hình ARIMA và VAR dự báo lạm phát tại Việt Nam. *Kinh tế tài chính Việt Nam* 5: 42-56

ANALYZING AND FORECASTING RICE PRODUCTION BY CROPPING SEASON IN VIETNAM BY 2030

**Ho Trong Phuc, Pham Minh Tam, Hoang Thi Lieu, Tran Minh Tri,
Nguyen Thi Minh Phuong**

Abstract: This article predicts Vietnam's planted paddy area, productivity, and production by cropping season by 2030 using the ARIMA model. This study utilized time series data for the period 1990–2021. The forecasted results show that by 2030, the cultivated paddy area of the spring and winter seasons is predicted to decrease from 3.01 and 1.56 million hectares to 2.75 and 1.21 million hectares, respectively. Meanwhile, the planted paddy area of the autumn season is projected to increase from 2.67 million hectares to 3.09 million hectares. Paddy productivity in all cropping seasons is projected to increase, specifically, the spring season increases from 6.86 tons/ha to 7.74 tons/ha, the autumn season increases from 5.67 tons/ha to 6.32 tons /ha, and the winter season increases from 5.17 tons/ha to 5.89 tons/ha. Paddy production is forecasted to increase, specifically, the spring and autumn seasons sharply increase, from 20.63 and 15.16 million tons to 24.33 and 18.37 million tons, respectively. Meanwhile, the production of the winter season is forecasted to slightly increase, to 8.13 million tons. The study suggests that policies should focus on maintaining the cultivated area of the spring and winter seasons.

Keywords: ARIMA model; Forecasting; Rice production; Cropping season; Vietnam