

ỨNG DỤNG MÔ HÌNH ARIMA DỰ BÁO SẢN LƯỢNG LÚA TỈNH THỪA THIÊN HUẾ ĐẾN NĂM 2025

Phạm Thị Thảo Hiền*, Nguyễn Ngọc Anh, Tôn Nữ Tuyết Trinh, Nguyễn Đức Hồng

Trường Đại học Nông Lâm, Đại học Huế

*Tác giả liên hệ: phamthithaohien@huaf.edu.vn

Nhận bài: 18/11/2019 Hoàn thành phản biện: 03/03/2020 Chấp nhận bài: 19/06/2020

TÓM TẮT

Nghiên cứu sử dụng số liệu thống kê về năng suất lúa từ 1995 – 2018 và sử dụng phần mềm Excel 2010 để nhập số liệu và Eview 10.0, SPSS 16.0 để chạy mô hình nhằm xây dựng mô hình dự báo ARIMA thích hợp cho dự báo sản lượng lúa ở Thừa Thiên Huế. Sử dụng tiêu chuẩn thông tin BIC (Bayesian Information Criterion) để tìm ra mô hình phù hợp, kết quả cho thấy: mô hình ARIMA (1,1,0) là thích hợp để giải thích sự biến động sản lượng lúa tỉnh Thừa Thiên Huế trong giai đoạn trên và cũng đưa ra dự báo về sản lượng lúa tỉnh Thừa Thiên Huế năm 2019 đến 2025.

Từ khóa: ARIMA, Dự báo, Sản lượng lúa

APPLICATION OF ARIMA MODEL TO FORECAST THUA THIEN HUE'S PADDY OUTPUT

Pham Thi Thao Hien, Nguyen Ngoc Anh, Ton Nu Tuyet Trinh, Nguyen Duc Hong

University of Agriculture and Forestry, Hue University

ABSTRACT

In this paper, the statistics on paddy output from 1995 to 2018 were used to build an ARIMA model aims to forecast Thua Thien Hue province's paddy output in the coming years. Using Bayesian Information Criterion (BIC) as the goodness of fit measure, the results showed that ARIMA (1, 1, 0) was the best model to explain the fluctuation of paddy output in above period and possible to make good prediction on Thua Thien Hue province's paddy output from 2019 to 2025.

Keywords: ARIMA, Forecast, Paddy output

1. MỞ ĐẦU

Lúa là cây lương thực chính được canh tác không chỉ ở các nước châu Á mà còn được trồng nhiều nơi trên toàn thế giới. Ngày nay, với sự gia tăng nhanh về dân số làm cho tình hình an ninh lương thực bị ảnh hưởng. Để đảm bảo nhu cầu về lương thực cho con người, nâng cao năng suất cây trồng, đặc biệt là năng suất cây lúa được coi như là biện pháp chính để giải quyết vấn đề trên.

Dự báo sản lượng lúa là một trong những việc cần thiết cho việc phát triển ngành kinh tế nông nghiệp của nước ta nói chung và Thừa Thiên Huế nói riêng. Dự

đoán được sản lượng lúa tăng, hoặc giảm sẽ làm cơ sở hoạch định các chính sách trong nông nghiệp và các chính sách xã hội khác. Thừa Thiên Huế có tổng diện tích đất nông nghiệp gần 412.000 ha, trong đó có gần 70.000 ha dành cho trồng lúa và hoa màu. Năng suất lúa bình quân đạt khoảng 55 tạ/ha. Tuy nhiên, năng suất lúa có sự khác biệt lớn giữa các vùng do yếu tố giống và kỹ thuật canh tác. Với sự tác động của biến đổi khí hậu đã không những làm giảm năng suất cây trồng mà còn giảm diện tích lúa, nhất là ở các vùng ven biển. Việc dự báo sản lượng lúa sẽ gặp nhiều khó khăn và không chính xác.

Hiện nay, có nhiều nghiên cứu thực nghiệm về dự báo sản lượng lúa được tiến hành nhiều nơi trên thế giới (Box và Jenkins, 1970; Sakamoto và cs., 2015; Ramakrishna và Vijaya, 2017). Việc ứng dụng công nghệ GIS để dự báo năng suất và sản lượng lúa ở Đồng bằng Sông Hồng đã được tiến hành (Dương Văn Khâm, 2006; Nguyễn Thị Hà, 2008). Tuy nhiên với công nghệ này cần nhiều số liệu về khí tượng, số liệu về năng suất, sản lượng cây trồng qua các năm. Võ Văn Tài (2012) đã sử dụng các mô hình toán học khác nhau của hồi quy và chuỗi thời gian ARIMA để dự báo sản lượng lúa của Việt Nam. Mô hình ARIMA đưa ra các kết quả dự báo khá phù hợp với sản xuất lúa. Tuy nhiên, với từng vùng khác nhau, biến động năng suất lúa là hoàn toàn khác nhau. Xuất phát từ thực tế đó, chúng tôi sử dụng các phương pháp Box-Jenkins để xây dựng mô hình ARIMA (Autoregressive Integrated Moving Average) để dự báo sự biến động sản lượng lúa tỉnh Thừa Thiên Huế. Từ đó làm cơ sở để xây dựng các chính sách nông nghiệp tại địa phương.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thu thập và xử lý số liệu

Số liệu phục vụ cho nghiên cứu được tổng hợp từ các báo cáo tại trang thông tin điện tử của Tổng cục Thống kê Việt Nam (GSO). Cụ thể, nghiên cứu đã thu thập số liệu sản lượng lúa tỉnh Thừa Thiên Huế từ năm 1995 đến năm 2018.

Sau đó, sử dụng phần mềm EXCEL 2013 để nhập số liệu và Eview 10.0, SPSS 16.0 để chạy mô hình.

2.1 Các bước thực hiện

Box và Jenkins (1970) lần đầu tiên giới thiệu mô hình ARIMA (autoregressive integrated moving average) trong phân tích chuỗi thời gian, được hiểu là phương pháp Box-Jenkins. Mô hình tự tương quan tích

hợp với trung bình trượt ARIMA (p,d,q) có phương trình tổng quát như sau:

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

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 ;

Phương pháp Box-Jenkins gồm bốn bước:

Bước 1: Nhận dạng mô hình

Nhận dạng mô hình ARIMA (p,d,q) thích hợp là việc tìm các giá trị thích hợp của p, d và q (với d là bậc sai phân của chuỗi dữ liệu thời gian được khảo sát, p là bậc tự hồi quy và q bậc trung bình trượt). Các giá trị này được xác định dựa vào biểu đồ tự tương quan (ACF) và biểu đồ tự tương quan riêng phần (PACF). Trong đó, việc lựa chọn mô hình AR(p) phụ thuộc vào biểu đồ PACF nếu nó có giá trị cao tại các độ trễ 1, 2, ..., p và giảm đột ngột sau đó, đồng thời dạng hàm ACF tắt lịm dần. Tương tự, việc lựa chọn mô hình MA(q) dựa vào biểu đồ ACF nếu nó có giá trị cao tại các độ trễ 1, 2, ..., q và giảm mạnh sau q, đồng thời dạng hàm PACF tắt lịm dần.

Bước 2: Ước lượng các thông số của mô hình ARIMA (p, d, q)

Tiến hành ước lượng các tham số cho các mô hình có khả năng phù hợp đã được nhận dạng. Ở đây, mô hình có khả năng phù hợp đã được nhận dạng. Mô hình có hệ số xác định R^2 , tiêu chuẩn thông tin BIC (Bayesian Information Criterion) và sai số bình phương trung bình RMSE

(Root mean square error) nhỏ nhất được coi là mô hình phù hợp nhất. Bên cạnh đó để đánh giá độ tin cậy của mô hình dự báo, nghiên cứu sử dụng chỉ số đánh giá độ chính xác của mô hình dự báo MAPE (Mean Absolute Percent Error). MAPE lớn hơn hoặc bằng 50% thì dự báo không chính xác, 20% - 50% là hợp lệ, 10%-20% là dự báo tốt, dưới 10% là dự báo hoàn hảo.

Bước 3: Kiểm tra mô hình

Mô hình ước lượng sau đó phải được kiểm tra lại để đảm bảo tính đại diện cho chuỗi dữ liệu quan sát. Việc này sẽ được thực hiện trên dãy giá trị sai số của mô hình nhằm xác định xem chúng có phải

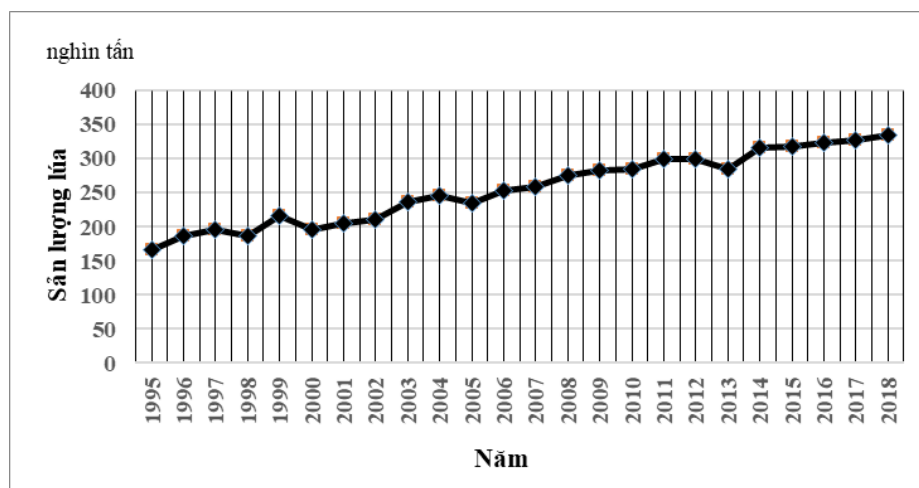
là sai số ngẫu nhiên trắng (white noise) hay không. Ở đây, biểu đồ ACF của phần dư sẽ cho phép kiểm tra tiêu chuẩn này. Ngoài ra, kiểm định Breusch-Godfrey (BG) và ARCH cũng được thực hiện trên phần dư nhằm kiểm tra về hiện tượng tự tương quan và phương sai số thay đổi.

Bước 4: Dự báo

Dựa trên phương trình của mô hình ARIMA, ta tiến hành xác định khoảng tin cậy của dự báo.

3. KẾT QUẢ NGHIÊN CỨU

3.1. Tổng quan về dữ liệu nghiên cứu



Hình 1. Sản lượng lúa tỉnh Thừa Thiên Huế từ năm 2005 đến năm 2018

Biểu đồ trên cho thấy một sự tăng trưởng liên tục sản lượng lúa qua các năm.

Bảng 1. Thống kê mô tả về chuỗi sản lượng lúa tỉnh Thừa Thiên Huế

	Sản lượng lúa
Observations	24
Mean	255,9708333
Median	256,1
Maximum	334,4
Minimum	166,6
Std.Dev	51,83534533
Skewness	-0,071807645
Kurtosis	-1,310794452
Shapiro – Wilk	0,943
P-valued	0,189

Chuỗi dữ liệu sản lượng lúa có phân phối chuẩn theo kiểm định Shapiro – Wilk ở mức ý nghĩa 5% (vì p-value > 0,05). Ngoài ra, từ các giá trị Skewness, Kurtosis cũng chứng tỏ rằng chuỗi biến động không nhiễu và có xu hướng lệch trái.

3.2. Mô hình dự báo sản lượng lúa

3.2.1. Kiểm tra tính dừng của chuỗi dữ liệu sản lượng lúa

Bảng 2. Kết quả kiểm tra ADF và PP đối với chuỗi dữ liệu sản lượng lúa

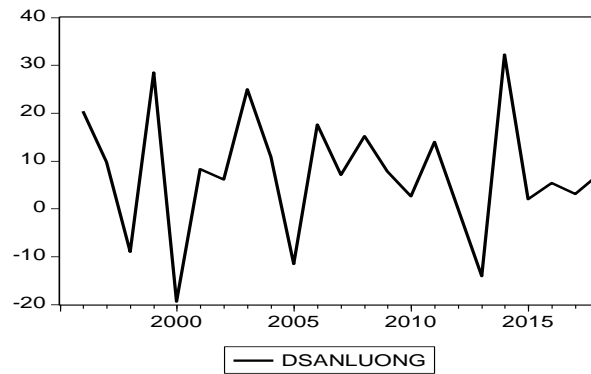
Kiểm định	Giá trị t	P-value
ADF	-0,916870	0,7641
PP	-0,885750	0,7742

So sánh p-value của kiểm định ADF và PP với 0,05 ta chấp nhận giả thiết: H_0 : chuỗi dữ liệu có nghiệm đơn vị, tức là chuỗi không dừng.

3.2.2. Khắc phục tính dừng của chuỗi số liệu

Chuỗi dữ liệu sử dụng trong ARIMA được giả định là chuỗi dừng, vì vậy ta cần phải xem xét chuỗi dữ liệu nghiên cứu có dừng hay chưa. Ta tiến hành kiểm tra tính dừng thông qua 2 kiểm định phổ biến: Augmented Dickey-Fuller (ADF) và Perron-Phillips (PP) được gọi là kiểm định nghiệm đơn vị (unit root test).

Lấy sai phân cấp 1 của chuỗi số liệu sản lượng ta được chuỗi mới dlsanluong và tiến hành kiểm tra tính dừng của chuỗi này. Chuỗi số liệu sau khi lấy sai phân có đồ thị như sau:



Hình 2. Chuỗi dừng sau khi lấy sai phân bậc 1

Sau đó ta tiến hành kiểm tra tính dừng thông qua 2 kiểm định phổ biến:

Augmented Dickey-Fuller (ADF) và Perron-Phillips (PP) cho chuỗi số dlsanluong

Bảng 3. Kết quả kiểm tra ADF và PP đối với chuỗi số dlsanluong

Kiểm định	Giá trị t	P-value
ADF	-7,679048	0,0000
PP	-8,331381	0,0000

Từ Bảng 3 ta thấy chuỗi sai phân bậc 1 của sản lượng lúa tỉnh Thừa Thiên Huế là chuỗi dừng.

Để xây dựng mô hình ARIMA, chúng tôi sử dụng chuỗi dữ liệu 24 quan sát từ 1995 đến năm 2018.

3.2.3. Xây dựng mô hình ARIMA cho sản lượng lúa tỉnh Thừa Thiên Huế

Bước 1: Nhận dạng (xác định các giá trị p, d, q)

Chuỗi dữ liệu dừng ở sai phân bậc 1, ta có $d = 1$.

Các giá trị p, q được xác định dựa vào bảng hệ số tự tương quan (ACF) và bảng hệ số tự tương quan riêng phần (PACF).

Bảng 4. Các hệ số tự tương quan ACF và hệ số tự tương quan riêng PACF của sai phân bậc 1 của chuỗi sản lượng lúa tỉnh Thừa Thiên Huế

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0,469	-0,469	5,7551	0,016
		2	-0,028	-0,318	5,7763	0,056
		3	0,014	-0,217	5,7823	0,123
		4	-0,065	-0,254	5,9118	0,206
		5	0,137	-0,054	6,5107	0,260
		6	-0,200	-0,260	7,8582	0,249
		7	0,163	-0,098	8,8155	0,266
		8	0,096	0,146	9,1710	0,328
		9	-0,227	-0,065	11,281	0,257
		10	-0,014	-0,244	11,290	0,335
		11	0,044	-0,217	11,383	0,412
		12	0,119	-0,071	12,124	0,436

Bước 2: Ước lượng mô hình

Từ Bảng 4 ta sử dụng các hệ số tự tương quan ACF để chọn bậc q cho MA, hệ số tự tương quan riêng PACF để chọn bậc p cho AR. Ta sẽ xét các mô hình sau: ARIMA (1, 1, 1), ARIMA (1, 1, 0), ARIMA (2, 1, 0), ARIMA (2, 1, 1), ARIMA (3, 1, 0), ARIMA (3, 1, 1), ARIMA (4, 1, 0), ARIMA (4, 1, 1).

Các mô hình đã nhận dạng được kiểm tra lại tính phù hợp dựa trên các thông số kiểm định: hệ số xác định R^2 , tiêu chuẩn thông tin BIC (Bayesian Information Criterion), sai số bình phương trung bình RMSE và chỉ số đánh giá độ chính xác của mô hình dự báo MAPE (Mean Absolute Percent Error) .

Bảng 5. Kết quả thống kê một số tiêu chuẩn của các mô hình ARIMA thử nghiệm

Mô hình	R^2	BIC	RMSE	MAPE
ARIMA (1,1,1)	0,960	5,277	10,651	3,370*
ARIMA (1,1,0)	0,946	5,196*	11,726	3,902
ARIMA (2,1,0)	0,952	5,440	11,294	3,674
ARIMA (2,1,1)	0,961*	5,242	10,467*	3,450
ARIMA (3,1,0)	0,954	5,400	11,328	3,577
ARIMA (3,1,1)	0,961*	5,432	10,754	3,463
ARIMA (4,1,0)	0,959	5,485	11,042	3,458
ARIMA (4,1,1)	0,961*	5,623	11,050	3,507

*giá trị tốt nhất dựa theo tiêu chuẩn lựa chọn

Từ Bảng 5 ta thấy mô hình ARIMA (2, 1, 1) là mô hình thỏa mãn nhiều nhất các tiêu chuẩn sử dụng, nhưng sau khi tiến hành xây dựng các mô hình trên thì thu được kết quả: mô hình ARIMA (1,1,0) có các ước lượng của tham số có ý nghĩa thống kê (với p-value < 0,05) và tiêu chuẩn BIC nhỏ nhất. Do đó mô hình ARIMA (1,

1, 0) là mô hình được sử dụng cho việc ước lượng tiếp theo.

Bước 3: Kiểm tra mô hình

Mô hình sau đó được kiểm tra mức độ phù hợp với chuỗi dữ liệu nghiên cứu bằng cách phân tích phần dư.

Bảng 6. Các hệ số tự tương quan ACF và hệ số tự tương quan riêng PACF của bình phương phần dư mô hình ARIMA(1, 1, 0)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0,155	-0,155	0,6288	0,428
		2	-0,315	-0,347	3,3434	0,188
		3	-0,042	-0,188	3,3942	0,335
		4	-0,001	-0,198	3,3943	0,494
		5	0,062	-0,081	3,5183	0,621
		6	-0,133	-0,264	4,1210	0,660
		7	0,187	0,088	5,3740	0,614
		8	0,108	0,073	5,8189	0,668
		9	-0,300	-0,213	9,5212	0,391
		10	-0,135	-0,254	10,326	0,412
		11	0,132	-0,149	11,162	0,430
		12	0,260	0,069	14,703	0,258

ACF của phần dư trong Bảng 6 cho thấy sai số là ngẫu nhiên trắng. Hơn nữa, kết quả kiểm định Breusch-Godfrey cũng cho thấy không tồn tại hiện tượng tự tương

quan, đồng thời kiểm định ARCH cũng chỉ ra rằng không có hiện tượng phương sai sai số thay đổi.

Bảng 7. Kết quả kiểm định Breusch-Godfrey và kiểm định ARCH

Kiểm định	Thông kê F	P-value
Breusch-Godfrey	3,513267	0,076340
ARCH	0,372506	0,548873

Kết quả kiểm tra cho thấy mô hình ARIMA (1, 1, 0) là thích hợp và có thể dự báo.

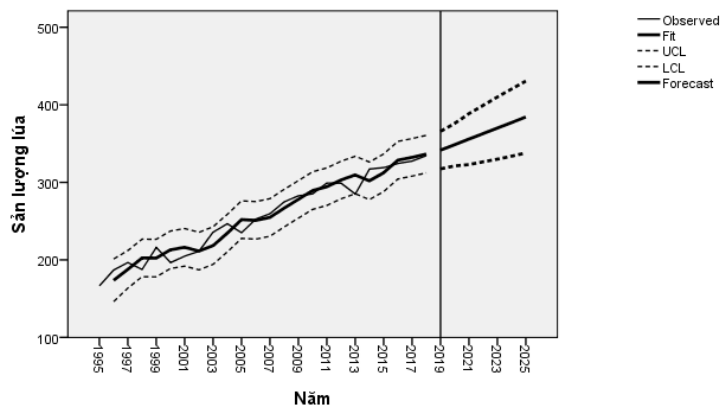
Những dự báo về sản lượng lúa tỉnh Thừa Thiên Huế dựa trên mô hình ARIMA được trình bày dưới bảng sau.

Bước 4: Dự báo

Bảng 8. Kết quả dự báo sản lượng lúa tỉnh Thừa Thiên Huế từ năm 2019 đến năm 2025 với độ tin cậy 95%

Năm	2019	2020	2021	2022	2023	2024	2025
Giá trị dự báo	341,6	348,7	355,8	362,9	370,0	377,1	384,2
Giới hạn cận trên	365,8	376,1	388,7	399,3	410,1	420,4	430,6
Giới hạn cận dưới	317,3	321,3	322,9	326,5	329,9	333,8	337,9

ĐV: nghìn tấn



Hình 3. Biểu đồ dự báo sản lượng lúa phạm vi trong và ngoài mẫu nghiên cứu

Từ Bảng 8 ta nhận thấy sản lượng lúa tỉnh Thừa Thiên Huế dự báo có xu hướng tăng phù hợp với tình hình tỉnh Thừa Thiên Huế đang tái cơ cấu nông nghiệp theo hướng nâng cao giá trị gia tăng, phát triển bền vững ứng dụng công nghệ cao.

4. KẾT LUẬN

Ứng dụng mô hình arima dự báo sản lượng lúa tỉnh Thừa Thiên Huế đến năm 2025 cho thấy mô hình ARIMA (2, 1, 1) là mô hình thỏa mãn nhiều nhất các tiêu chuẩn sử dụng, nhưng sau khi tiến hành xây dựng các mô hình trên thì thu được kết quả mô hình ARIMA (1, 1, 0) có các ước lượng của tham số có ý nghĩa thống kê (với p -value < 0,05) và tiêu chuẩn BIC nhỏ nhất. Mô hình ARIMA (1, 1, 0) là mô hình phù hợp nhất trong các mô hình thử nghiệm và có thể dùng để dự báo sản lượng lúa ở tỉnh Thừa Thiên Huế. Kết quả từ mô hình ARIMA (1, 1, 0) có thể cung cấp các thông tin để làm chính sách trong việc tìm kiếm những giải pháp thích hợp để làm tăng năng suất lúa và định hướng sản xuất lúa trong tương lai gần.

TÀI LIỆU THAM KHẢO

1. Tài liệu tiếng Việt

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.

Nguyễn Thị Hà. (2008). *Nghiên cứu dự báo năng suất ngô, đậu tương, lạc và xây dựng quy trình giám sát khí tượng nông nghiệp cho 4 cây trồng chính (lúa, ngô, lạc, đậu tương) bằng thông tin mặt đất ở Việt Nam*, Báo cáo tổng kết đề tài nghiên cứu khoa học và công nghệ cấp bộ, Viện Khoa học Khí tượng Thủy văn và Môi trường, Hà Nội.

Dương Văn Khâm. (2006). *Nghiên cứu áp dụng công nghệ viễn thám (RS) và hệ thống thông tin địa lý (GIS) trong khí tượng thủy văn*. Báo cáo tổng kết đề tài nghiên cứu khoa học và công nghệ cấp bộ, Viện Khoa học Khí tượng Thủy văn và Môi trường, Hà Nội.

2. Tài liệu tiếng nước ngoài

Box, G. E. P., & Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Ramakrishna, G., Vijaya, K. R. (2017). ARIMA model for forecasting of rice production in India by Sas. *International Journal of Applied Mathematics & Statistical Sciences (IJAMSS)*, 6(4), 67-72.

Sakamoto, T., Masayuki Yokozawa, Hitoshi Toritani, Michio Shibayama, Naoki Ishitsuka, Hiroyuki Ohno (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of Environment*, 96, 366-374.