

THỬ NGHIỆM PHẦN MỀM SPSS10.0 TRONG DỰ BÁO NĂNG SUẤT MỘT SỐ LOẠI CÂY TRỒNG

KS. Lương Đình Tuyển, TS. Bùi Việt Nữ, KS. Chiêu Kim Quỳnh
Trung tâm KTTV phía Nam

Trong chương trình hợp tác nghiên cứu Khí tượng nông nghiệp giữa Việt Nam - Trung Quốc, các chuyên gia Trung Quốc đã xây dựng và chuyển giao mô hình dự báo năng suất theo phần mềm SPSS 10.0 cho phía Việt Nam, phần mềm trên đã và đang được áp dụng tại Trung Quốc. Để có thể khẳng định được độ chính xác của mô hình dự báo năng suất trong điều kiện khí hậu nhiệt đới gió mùa của miền Nam Việt Nam, chúng tôi tiến hành thử nghiệm tính toán dự báo năng suất lúa tại tỉnh Cần Thơ và năng suất ngô tại tỉnh Đồng Nai.

1. Nội dung của phần mềm SPSS 10.0

Về bản chất, mô hình dự báo năng suất theo phần mềm SPSS 10.0 dựa trên nguyên tắc sử dụng thuật toán hồi quy bội từng bước, với quá trình sàng lọc dần và xác định mối tương quan giữa sự biến động năng suất do thời tiết tạo nên với những yếu tố chính của thời tiết.

a. Cơ sở lý luận

Như chúng ta đã biết, có nhiều yếu tố ảnh hưởng đến năng suất nhưng có thể tách chúng ra thành ba nhóm sau:

- 1) Các tác động của các thành phần không ngẫu nhiên,
- 2) Các tác động của các thành phần ngẫu nhiên,
- 3) Những tác động nhiễu ngẫu nhiên.

Phương trình tổng quát có dạng như sau:

$$Y = Y_{xt} + Y_{tt} + \Delta Y \quad (1)$$

Trong đó:

Y - năng suất cây trồng;

Y_{xt} - thành phần năng suất xu thế do nhóm các tác động không ngẫu nhiên tạo nên (là hàm của thời gian) và được tính trong trạng thái thời tiết ở mức trung bình.

$$Y_{xt} = F(t); t - \text{năm} \quad (2)$$

Y_{tt} - thành phần năng suất do tác động của khí hậu thời tiết tạo nên.

ΔY - sai số ngẫu nhiên (thường được bỏ qua trong dự báo năng suất).

Năng suất dự báo được tính theo công thức sau:

$$Y_{db} = Y_{xtdb} + Y_{ttdb} \quad (3)$$

Trong đó:

Y_{db} - năng suất dự báo,

Y_{xtdb} - năng suất xu thế dự báo,

Y_{ttdb} - năng suất thời tiết dự báo.

Sau khi tính được năng suất xu thế sẽ tính được năng suất thời tiết. Năng suất thời tiết là hiệu số giữa năng suất thực và năng suất xu thế.

$$Y_u = Y_T - Y_{xt} \quad (4)$$

Ở đây Y_T - năng suất thực.

Các giá trị Y_u sẽ được sử dụng ở các bước tiếp trong mô hình để tìm phương trình dự báo năng suất thời tiết theo phương pháp hồi quy từng bước.

Năng suất thời tiết là hàm của các yếu tố khí tượng được biểu thị bằng công thức sau:

$$Y_u = F(t, r, s, \dots \text{tuần, tháng}). \quad (5)$$

Tiến hành tính các hệ số tương quan riêng giữa Y_u với các yếu tố thời tiết, chọn những tuần, tháng có giá trị tuyệt đối của hệ số tương quan cao. Từ đó thiết lập hồi quy giữa năng suất thời tiết với tất cả các tuần, tháng đã chọn sẽ có phương trình năng suất thời tiết dự báo (Y_{udb}).

Mức độ phù hợp giữa năng suất tính theo mô hình và năng suất thực được đánh giá khi so sánh sai số dự báo (S_{db}) với sai số cho phép (S_{cf}).

Các sai số được tính như sau:

$$S_{cf} = 0,674 * \sigma \quad (6)$$

$$S_y = \pm \sigma * \sqrt{1 - r^2} \quad (7)$$

$$S_{db} = Y_t - Y_{db} \quad (8)$$

Trong đó:

S_{cf} - sai số cho phép.

σ - độ lệch chuẩn của năng suất thực, tính theo công thức:

$$\sigma = \sqrt{\frac{\sum(Y_t - Y_{tb})^2}{(n-1)}} \quad (9)$$

r - hệ số tương quan,

S_y - sai số chuẩn,

S_{db} - sai số dự báo,

Y_t - năng suất thực,

Y_{tb} - năng suất trung bình của dãy số liệu,

Y_{db} - năng suất dự báo (tính toán),

n - số số hạng của chuỗi số liệu.

Nếu $|Y_t - Y_{db}| \leq S_{cf}$ thì kết quả dự báo coi là phù hợp và nếu ngược lại là chưa phù hợp.

b. Các bước tính toán

- Nhập số liệu vào môi trường SPSS,

- Kiểm tra hồi qui bội giữa năng suất với các yếu tố khí tượng bằng các đồ thị phân dư,

- Các tính toán thống kê được thực hiện nhờ các menu trong mô hình theo các ứng dụng khác nhau.

Phương pháp lựa chọn trong phần mềm chuyển giao là sử dụng phương pháp loại trừ dần (Backward). Phương pháp loại trừ dần khởi đầu với tất cả các biến đều ở trong phương trình và sau đó tuân tự loại trừ chúng bằng tiêu chuẩn loại trừ. Có hai tiêu chuẩn loại trừ (tiêu chuẩn ra) trong SPSS. Tiêu chuẩn thứ nhất là giá trị F tối thiểu biến phải đạt được để ở lại trong phương trình gọi là F ra, ký hiệu trong chương trình là FOUT. Các biến có giá trị F nhỏ hơn F ra sẽ bị loại ra khỏi phương trình. Tiêu chuẩn thứ hai là xác suất tối đa của F ra, ký hiệu trong chương trình là POUT, mà một biến phải nhỏ hơn, nếu vượt quá xác suất tối đa cho phép này thì biến sẽ bị loại. Các giá trị mặc định của FOUT là 2.71 và giá trị mặc định của POUT là 0.10 [8].

Để xây dựng mô hình thể hiện sự tương quan giữa một biến phụ thuộc (năng suất) vào nhiều biến độc lập (các yếu tố khí tượng), tính ma trận tương quan của từng biến, sau đó tập trung chú ý vào hệ số tương quan cao của các biến độc lập (vì trị tuyệt đối của hệ số tương quan càng lớn thì liên hệ tuyến tính càng tốt). Từ những tuân, tháng của các yếu tố đã chọn xây dựng phương trình năng suất thời tiết theo phương pháp loại trừ dần.

2. Tính toán dự báo năng suất

a. Dự báo năng suất lúa tại tỉnh Cần Thơ

1) Số liệu

Số liệu năng suất lúa trung bình các vụ trong năm từ 1985 ÷ 2000.

Số liệu khí tượng của Trạm khí tượng Cần Thơ từ 1985 ÷ 2000.

2) Kết quả tính toán

Việc tính toán được áp dụng cho 2 vụ đông xuân và hè thu là các vụ chính có dãy số liệu dài. Theo nội dung các bước tính toán trên, kết quả tính được như sau:

a) Vụ đông xuân

Hàng năm, vụ đông xuân thường được bắt đầu gieo trồng từ tháng XI. Để dự báo năng suất vụ đông xuân, tiến hành tìm hiểu mối tương quan giữa năng suất thời tiết và chuỗi số liệu thời tiết từ tháng I đến tháng X hàng năm.

Thực hiện tính năng suất xu thế:

Theo phương trình (2), phương trình năng suất xu thế có dạng như sau:

$$Y_{xt} = - 2036,479 + 1,047 * \text{năm}. \quad R = 0,892 \quad (10)$$

Sau khi xác định năng suất xu thế (Y_{xt}), theo công thức (4) tiếp tục tính năng suất thời tiết (Y_{tt}).

Từ năng suất thời tiết lập ma trận tương quan với các yếu tố thời tiết đã chọn được các hệ số tương quan riêng cao giữa năng suất thời tiết với các yếu tố trên.

Tiếp tục tìm tương quan giữa Y_{tt} với các yếu tố của các tuân theo phương pháp loại trừ dần của mô hình trên, phương trình có dạng như (11):

$$Y_{\text{tđb}} = -43,337 + 0,940 * T_{X_{2.II}} + 0,08648 * S_{1.V} + 0,07595 * S_{2.VIII} + 0,02381 * R_{3.IX} \quad (11)$$

Ở đây hệ số tương quan chung $R = 0,829$.

Các hệ số tương quan riêng $r_{T_{X_{2.II}}} = 0,461$; $r_{S_{1.V}} = 0,437$; $r_{S_{2.VIII}} = 0,396$; $r_{R_{3.IX}} = 0,442$.

$T_{X_{2.II}}$ - nhiệt độ tối cao tuần 2 tháng II.

$S_{1.V}$ - tổng số giờ nắng của tuần 1 tháng V.

$S_{2.VIII}$ - tổng số giờ nắng của tuần 2 tháng VIII.

$R_{3.IX}$ - tổng lượng mưa tuần 3 tháng IX

Tính được $Y_{\text{tđb}}$ theo (11) và Y_{xt} theo (10) cho từng năm và tính được năng suất dự báo.

Để kiểm tra mức chính xác của phương pháp như trên đã nêu, dựa vào các công thức 6, 7, 8 và 9 tính được:

$$\sigma = 2,51; S_{cf} = 1,69 \text{ tạ/ha}; S_y = \pm 1,28 \text{ tạ/ha.}$$

Kết quả có 14/16 lần dự báo có sai số nhỏ hơn sai số cho phép ($< 1,69 \text{ tạ/ha}$) đạt 87,5%. Như vậy có thể dùng công thức trên trong dự báo tác nghiệp.

b) Vụ hè thu

Hàng năm vụ hè thu thường được bắt đầu gieo trồng từ tháng V. Để dự báo năng suất vụ hè thu, tiến hành tìm hiểu mối tương quan giữa năng suất thời tiết và chuỗi số liệu khí tượng từ tháng V năm trước đến hết tháng IV năm sau.

Tương tự như tính cho vụ đông xuân, kết quả như sau:

Năng suất xu thế:

$$Y_{xt} = -709,267 + 0,374 * \text{năm.} \quad R = 0,754 \quad (12)$$

Phương trình tính năng suất thời tiết dự báo:

$$Y_{\text{tđb}} = 69,499 - 0,0634 * S_{2.IV} - 1,103 * T_{Tb_{3.VI}} - 1,048 * T_{X_{2.II}} \quad (13)$$

Hệ số tương quan chung $R = 0,878$.

Các hệ số tương quan riêng $r_{S_{2.IV}} = -0,484$; $r_{T_{Tb_{3.VI}}} = -0,576$; $r_{T_{X_{2.II}}} = -0,587$.

$S_{2.IV}$ - tổng số giờ nắng tuần 2 tháng IV.

$T_{Tb_{3.VI}}$ - nhiệt độ trung bình tuần 3 tháng VI.

$T_{X_{2.II}}$ - nhiệt độ cao nhất tuần 2 tháng II.

Tính được: $\sigma = 1,60$ $S_{cf} = 1,08$

Kết quả dự báo có 12/16 lần sai số dưới mức cho phép S_{cf} (mức chính xác đạt 75%). Như vậy có thể dùng công thức trên để dự báo.

b. Dự báo năng suất cho cây ngô lai

1) Đặc điểm chuỗi số liệu nghiên cứu

Số liệu khí tượng thu thập theo tuần gồm 6 yếu tố từ 1985 đến năm 2001.

Có những biến động mạnh về năng suất: trước năm 1995 chủ yếu gieo trồng giống ngô địa phương năng suất thấp; sau 1995 gieo trồng chủ yếu các giống ngô lai. Xét ưu thế sử dụng giống trong thực tế, nên mô hình dự báo năng suất chỉ sử dụng các số liệu từ năm 1995 đến năm 2001 (là những năm gieo trồng 100% giống lai).

2) Kết quả tính toán

a) Vụ đông xuân

Hàng năm, vụ đông xuân thường được gieo vào trung tuần tháng XI và kết thúc giữa tháng III năm sau. Như vậy, số liệu khí tượng sử dụng trong mô hình dự báo năng suất ngô bao gồm tất cả các tháng từ đầu năm cho đến tháng X.

Tương tự như tính toán cho cây lúa:

$$Y_{xt} = 1911,513 - 0,937* \text{năm}. \quad R = 0,521 \quad (14)$$

Phương trình tính năng suất thời tiết dự báo:

$$Y_{udb} = -98,684 - 0,141*S_{1.IX} + 3,907*T_{2.III} \quad (15)$$

$$R = 0,954$$

Khi xác định được Y_{udb} và Y_{xt} cho từng năm sẽ tính được năng suất dự báo.

$$S_{cf} = 2,29$$

Các sai số dự báo tính được đều có giá trị nhỏ hơn 2,29, như vậy theo lý thuyết phương trình dự báo năng suất trong vụ đông xuân cho cây ngô là có thể áp dụng được.

b) Vụ hè thu

Hàng năm, vụ hè thu thường được bắt đầu gieo trồng từ tháng V và kết thúc vào tháng VIII. Để dự báo năng suất vụ hè thu, tiến hành tìm hiểu mối tương quan giữa năng suất thời tiết và chuỗi số liệu khí tượng từ tháng V năm trước đến hết tháng IV năm sau.

Tương tự như tính toán cho vụ đông xuân, tính được:

$$Y_{xt} = 1975,307 - 0,969* \text{năm}. \quad (16)$$

$$R = 0,317$$

Phương trình tính năng suất thời tiết dự báo:

$$Y_{udb} = 131,966 - 4,886* T_{2.V} \quad (17)$$

$$R = 0,897$$

Tính được Y_{udb} theo (17) và Y_{xt} theo (16) cho từng năm, sẽ tính được năng suất dự báo.

$$\text{Tính } S_{cf} = 3,8$$

Kết quả dự báo có 5/6 trường hợp nhỏ hơn sai số cho phép, vì vậy phương trình dự báo năng suất trên đáp ứng 83%.

c) Vụ mùa

Vụ mùa thường bắt đầu gieo từ tháng VIII và kết thúc vào tháng XI. Để dự báo năng suất vụ mùa, tiến hành tìm hiểu mối tương quan giữa năng suất thời tiết và chuỗi số liệu khí tượng từ tháng VIII năm trước đến hết tháng VII năm sau.

Xác định năng suất xu thế:

$$Y_{xt} = 705,620 - 0,334* \text{năm}. \quad (18)$$

$$R = 0,444$$

Phương trình tính năng suất thời tiết dự báo:

$$Y_{\text{udb}} = 61,785 - 0,661 * U_{1,\text{VIII}} - 0,469 * N_{r_{3,\text{VIII}}} \quad (19)$$
$$R = 0,963$$

Tính năng suất dự báo theo công thức (18 & 19).

$$\text{Tính: } S_{\text{cf}} = 0,95$$

Tất cả các trường hợp đều có sai số dự báo nhỏ hơn sai số cho phép, vì vậy phương trình dự báo năng suất trên là phù hợp.

3. Kết luận và đề nghị

1. Mô hình dự báo năng suất do chuyên gia Trung Quốc chuyển giao có thể áp dụng tốt đối với chuỗi số liệu không có những đột biến lớn về năng suất.
2. Các kết quả tính toán đối với cây ngô mặc dù khá phù hợp với năng suất thực tế song do chuỗi số liệu quá ngắn nên cần tiếp tục nghiên cứu.
3. Áp dụng mô hình dự báo năng suất lúa cho tỉnh Cần Thơ.
4. Nên tiếp tục thử nghiệm mô hình dự báo năng suất cho các cây trồng khác và các tỉnh khác.

Tài liệu tham khảo

1. Cần Thơ 25 năm xây dựng và phát triển 1975-2000. NXB Cần Thơ.
2. Đặc điểm khí hậu tỉnh Hậu Giang. UBKHKT tỉnh Hậu Giang, 1980.
3. Đặc điểm thủy văn tỉnh Hậu Giang. Đài KTTV Hậu Giang, 1986.
4. Nghiên cứu khí hậu nông nghiệp nhiệt đới ẩm Đông Nam Á. NXB Nông nghiệp. Hà Nội, 1986.
5. Ứng dụng SPSS for Windows để xử lý và phân tích dữ liệu nghiên cứu. NXB KHKT, 1997.