

## CÁC THAM SỐ ỔN ĐỊNH TRONG CHỌN GIỐNG CÂY TRỒNG

### Analysis stability index in plant breeding

Nguyễn Đình Hiền<sup>1</sup>, Lê Quý Kha<sup>2</sup>

#### SUMMARY

Stability in performance is one of the most desirable properties of a genotype to be released as a variety for wide cultivation. This study was reviewed some statistical method for estimate phenotypic stability used in plant breeding from the previous papers. The Eberhart and Russel's model of stability have been described in (1) and (2) and recent developments of this model have been presented by Weikai Yan (4) using software Ggebiplot (5).

**Key words:** Stability analysis, environmental index, plant breeding.

Trong nghiên cứu giống cây trồng ngoài việc tìm các giống năng suất cao người ta còn rất chú ý đến các giống “ổn định” khi trồng trên nhiều vùng khác nhau hoặc qua nhiều vụ khác nhau. Nhưng thế nào là “ổn định” thì các nhà nông học lại không thống nhất. Có quá nhiều định nghĩa ổn định hoặc chỉ mang tính mô tả không định lượng hoặc có kèm theo các chỉ số định lượng. Trong tài liệu này chúng tôi muốn giới thiệu một số nghiên cứu về ổn định sắp xếp theo ba phần: Các tham số ổn định, phân tích chùm, phân tích kiểu gen và môi trường.

#### 1. Các tham số ổn định

Có thể sắp xếp các tham số ổn định đang xuất hiện trên các bài báo và tài liệu về ổn định hiện nay thành 4 nhóm:

**Nhóm A:** Theo dõi năng suất  $Y_{ij}$  của giống  $i$  ở địa điểm  $j$  ( $j = 1, q$ ) sau đó tính năng

Plaisted and Peterson (1959)

$$\theta_i = \frac{p}{2(p-1)(q-1)} \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2 + \frac{SS(GE)}{2(p-1)(q-1)}$$

với 
$$SS(GE) = \sum_{i=1}^p \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2$$

suất trung bình  $Y_{i*}$ , phương sai  $s_i^2$  và hệ số biến động. Hai tham số ổn định của giống  $i$  do Francis và Kannenberg (1978) đề xuất là:

$$\text{Phương sai } s_i^2 = \sum_{j=1}^q (Y_{ij} - Y_{i*})^2 / (q-1)$$

$$\text{Hệ số biến động } CV_i = \frac{s_i}{Y_{i*}} \times 100\%$$

#### Nhóm B:

Theo dõi năng suất  $Y_{ij}$  của các giống  $i$  ( $i = 1, p$ ) tại các vùng  $j$  ( $j = 1, q$ ), tính trung bình  $Y_{*j}$  của các giống  $i$  tại vùng  $j$ , trung bình  $Y_{i*}$  của giống  $i$  qua các vùng  $j$  và trung bình của tất cả các giống qua tất cả các vùng  $Y_{**}$ . Từ đó tính được các hệ số ổn định sau:

<sup>1</sup> Trường Đại học Nông nghiệp I

<sup>2</sup> Viện Nghiên cứu Ngô

*Plaisted and Peterson (1960)*

$$\theta_{(i)} = \frac{-p}{(p-1)(p-2)(q-1)} \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2 + \frac{SS(GE)}{(p-2)(q-1)}$$

*Wricke (1962)*

$$W_i^2 = \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2$$

*Shukla (1972)*

$$\sigma_i^2 = \frac{p}{(p-2)(q-1)} \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2 - \frac{SS(GE)}{(p-1)(p-2)(q-1)}$$

**Nhóm C:**

*Finlay and Wilkinson (1956)*

$$b_i = \frac{\sum_{j=1}^q (Y_{ij} - Y_{i*})(Y_{*j} - Y_{**})}{\sum_{j=1}^q (Y_{*j} - Y_{**})^2}$$

*Perkins and Jinks (1968)*

$$\beta_i = \frac{\sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})(Y_{*j} - Y_{**})}{\sum_{j=1}^q (Y_{*j} - Y_{**})^2}$$

**Nhóm D:**

*Eberhart and Russel (1966)*

$$\delta_i^2 = \frac{1}{(q-2)} \left[ \sum_{j=1}^q (Y_{ij} - Y_{i*})^2 - b_i^2 \sum_{j=1}^q (Y_{*j} - Y_{**})^2 \right]$$

*Perkins and Jinks (1968)*

$$\delta_i^2 = \frac{1}{(q-2)} \left[ \sum_{j=1}^q (Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})^2 - \beta_i^2 \sum_{j=1}^q (Y_{*j} - Y_{**})^2 \right]$$

Có thể tóm tắt lại như sau:

Gọi các hiệu số  $(Y_{ij} - Y_{i*})$  giữa năng suất  $Y_{ij}$  của giống  $i$  tại địa điểm  $j$  và năng suất trung bình của giống  $i$  qua tất cả các địa điểm  $Y_{i*}$  là các độ lệch của giống  $i$  (DG).

Gọi các hiệu số  $(Y_{ij} - Y_{i*} - Y_{*j} + Y_{**})$  là các tương tác gen\* môi trường (GE).

Nhóm A tính riêng cho từng giống và chỉ dùng tổng bình phương của các DG của giống đó.

Nhóm B tính cho từng giống nhưng dùng đến tổng bình phương của tất cả các GE của tất cả các giống.

Nhóm C tính hệ số hồi quy của giống theo địa điểm và dùng các DG hoặc các GE của giống đó cùng với các trung bình của tất cả các giống tại các địa điểm  $Y_{*j}$  và trung bình toàn bộ  $Y_{**}$ .

Nhóm D tính hệ số hồi quy của giống theo địa điểm sau đó tính tổng bình phương do hồi quy và tổng bình phương do sai số dựa trên DG hoặc trên GE của giống đó cùng với các trung bình của tất cả các giống tại các địa điểm  $Y_{*j}$  và trung bình toàn bộ  $Y_{**}$ .

Đã có những công trình nghiên cứu để chứng minh các tham số trong cùng một nhóm hoặc tương đương nhau hoặc khi dùng

để đánh giá ổn định của giống sẽ cho kết quả giống nhau.

Có thể phân khái niệm ổn định ra làm 3 loại:

**Loại I:** Một giống được coi là ổn định nếu phương sai  $s^2_i$  (tính ở nhóm A) nhỏ. Như vậy năng suất của giống này tại các địa điểm có thể coi xấp xỉ bằng nhau. Nếu các địa điểm rất khác biệt (về đất đai, nguồn nước, khí hậu,...) thì khó có thể tìm được giống ổn định theo quan niệm này. Nhưng nếu các địa điểm không quá khác biệt thì tìm được một giống ổn định, năng suất lại cao là rất quý. Đôi khi người ta gọi ổn định loại này là *ổn định theo quan điểm sinh học*. Ổn định này có quan hệ mật thiết với khái niệm nội cân bằng (Homeostasis).

Thay cho phương sai  $s^2_i$  có thể dùng hệ số biến động  $CV_i$  của nhóm A.

**Loại II:** Tính năng suất trung bình  $Y_{*j}$  của các giống tại địa điểm j. Nếu năng suất  $Y_{*j}$  thấp thì địa điểm j được gọi là không thuận lợi cho loại cây đó. Nếu năng suất  $Y_{*j}$  cao thì địa điểm j được gọi là thuận lợi. Theo dõi các năng suất trung bình  $Y_{*j}$  và năng suất  $Y_{ij}$  của giống i qua các địa điểm. Nếu  $Y_{ij}$  biến đổi "song song" với  $Y_{*j}$  thì giống i được coi là ổn định. Như vậy không giống như ổn định loại I năng suất của giống i thay đổi theo địa điểm, năng suất thấp tại địa điểm không thuận lợi và năng suất cao tại địa điểm thuận lợi. Đôi khi người ta gọi ổn định loại II này là ổn định theo quan điểm nông học.

Để đánh giá ổn định loại II có thể dùng các tham số ở nhóm B và khi sắp xếp mức độ ổn định của các giống (xếp thứ tự căn cứ vào các hệ số) thì 4 tham số ở nhóm B sẽ cho kết quả sắp xếp giống nhau. Giống ổn định là giống có tham số của nhóm B nhỏ.

Nếu dùng hiệu  $(Y_{*j} - Y_{**})$  làm *chỉ số môi trường (environment index)*  $I_j$  sau đó tính hồi quy tuyến tính của dãy  $(Y_{ij} - Y_{i**})$  theo dãy chỉ số môi trường  $I_j$  ta được các tham số ổn định là các hệ số hồi quy  $b_i$ .

Nếu tìm hồi quy của dãy  $(Y_{ij} - Y_{i**} - Y_j + Y_{**})$  theo dãy chỉ số môi trường  $I_j$  ta được các tham số ổn định là các hệ số hồi quy  $\beta_i$ . Có thể thấy  $b_i = 1 + \beta_i$

*Một giống được coi là ổn định nếu  $b_i = 1$  (hay  $\beta_i = 0$ ).* Giống ổn định và có năng suất trung bình trên tất cả các địa điểm  $Y_{i**}$  cao là giống được chú ý vì cho năng suất nhìn chung là cao trên tất cả các địa điểm. Giống không ổn định với  $b_i > 1$  (xác định qua kiểm định T) sẽ cho năng suất cao ở vùng thuận lợi nhưng không nên trồng ở vùng không thuận lợi, còn giống có  $b_i < 1$  thì thích hợp với vùng không thuận lợi nhưng lại không nên trồng ở vùng thuận lợi.

Hai mô hình ở nhóm C được nhiều nhà nông học sử dụng tuy nhiên việc dùng các hệ số hồi quy theo *chỉ số môi trường có nhược điểm là chỉ số môi trường phụ thuộc vào nhóm giống đang khảo sát*. Thí dụ khảo sát 5 giống (A, B, C, D, E) mà giống A khác xa với 4 giống (B, C, D, E) thì khi tính hồi quy sẽ kết luận A không ổn định còn B ổn định, nhưng nếu khảo sát 5 giống (A, B, F, G, H) trong đó B khác xa 4 giống (A, F, G, H) thì lại kết luận A ổn định còn B không ổn định. Như vậy nếu không tìm được một dãy "chỉ số môi trường" khách quan hơn dãy  $I_j$  thì vẫn chưa có đầy đủ cơ sở để đánh giá ổn định qua các hệ số  $b_i$  (hay  $\beta_i$ ). Nếu vẫn dùng nhóm C để đánh giá ổn định thì cần lưu ý lựa chọn các giống tham gia vào việc tính chỉ số  $I_j$  một cách thận trọng, đừng chọn các giống có những diễn biến bất thường và phải chú ý đến quan hệ giữa các giống.

**Loại III:** Sau khi tính hồi quy như đã làm với nhóm C người ta phân tích tổng biến động thành hai phần, tổng biến động do hồi quy và tổng biến động xung quanh hồi quy hay biến động do sai số.

Nếu tổng biến động do sai số nhỏ thì năng suất của giống chủ yếu biến động theo đường thẳng hồi quy còn phần bất thường do sai số gây ra không đáng kể. Nói cách khác

có thể coi như đã nắm chắc được sự thay đổi của năng suất theo địa điểm. Như vậy có thể dùng các tham số của nhóm D làm tham số ổn định. Một số người cho rằng nhóm D không thể dùng làm tham số ổn định vì tổng biến động do sai số mà lớn thì chỉ *chứng tỏ mô hình hồi quy tuyến tính không phù hợp* chứ không liên quan đến tính ổn định của giống.

## 2. Phân tích chùm (Cluster Analysis)

Như trên đã trình bày có nhiều hệ số ổn định và tất cả các hệ số đó đều dựa vào DG hoặc GE, sau đó tính phương sai hoặc hồi quy. Một số nghiên cứu về sau (Hanson (1970), Lin (1982), Lefkovitch (1985)) không đi sâu vào tính ổn định của từng giống mà sử dụng ý tưởng của phân tích chùm để lựa chọn các giống có “diễn biến giống nhau” trên các địa điểm để ghép lại thành các nhóm đồng đều hơn, sau đó tìm các nhóm có “diễn biến giống nhau” để ghép lại thành nhóm to hơn v.v... Muốn đánh giá thế nào là có “diễn biến giống nhau” phải tính các hệ số “tương tự” (Similarity index) hoặc ngược lại tính hệ số “khác biệt” (Dissimilarity index) sau đó lựa chọn thuật toán ghép nhóm. Kết quả cuối cùng là tìm ra các nhóm có năng suất diễn biến giống nhau qua địa điểm gọi là *các nhóm ổn định*. Chọn một nhóm cụ thể, căn cứ vào diễn biến của cả nhóm qua các địa điểm để chọn giống đại diện.

## 3. Chương trình phân tích kiểu gen và môi trường (GE biplot)

Các tham số ổn định ở phần 1 có thể tính bằng máy tính cầm tay hoặc viết thành chương trình máy tính như chương trình tính ổn định theo Eberhart and Russel của Nguyễn Đình Hiến, chương trình Irristat ver 4.4. Việc tính các hệ số “tương tự” hay hệ số “khác biệt” cũng không khó nhưng việc lựa chọn thuật toán để ghép thì còn những tranh luận để có được những cách ghép hợp lý và có ý nghĩa về mặt sinh học.

Gần đây có chương trình chuyên phân tích tương tác kiểu gen  $\times$  môi trường (Weikai Yang 2001) tuy không trình bày chi tiết các chỉ số định lượng nhưng lại cho các hình ảnh hai chiều rất cụ thể để đánh giá tính ổn định một cách định tính.

Có thể giới thiệu tóm tắt GE biplot như sau:

Khi nghiên cứu năng suất của n giống tại p địa điểm chúng ta ghi kết quả vào một bảng hai chiều gồm n hàng  $\times$  p cột gọi là bảng A.

Địa điểm \ Giống	D <sub>1</sub>	D <sub>2</sub>	...	D <sub>p</sub>
V <sub>1</sub>	x <sub>11</sub>	x <sub>12</sub>		x <sub>1p</sub>
V <sub>2</sub>	x <sub>21</sub>	x <sub>22</sub>		x <sub>2p</sub>
...				
V <sub>n</sub>	x <sub>n1</sub>	x <sub>n2</sub>		x <sub>np</sub>

Sau phép quy tâm, chúng ta tìm ma trận chuyển vị  $A^T$  và tính hai tích  $Ax A^T$  và  $A^T \times A$ . Tính các thành phần chính rồi tập trung sự chú ý vào hai thành phần chính đầu. Dùng hai thành phần chính đầu làm 2 trục vuông góc của một đồ thị. Hai hình chiếu trên hai thành phần chính đầu của một hàng trong bảng A (giống) giúp chúng ta vẽ một vectơ minh họa giống. Tất cả có n vectơ giống  $V_i$  ( $i = 1, n$ ).

Tương tự, hai hình chiếu trên hai thành phần chính đầu của một cột trong bảng A (địa điểm) cho ta một vectơ minh họa địa điểm  $D_j$  ( $j = 1, p$ ).

Ma trận A có thể thay gần đúng bằng tích của hai ma trận, ma trận chứa các vectơ giống  $V_i$  và ma trận chứa các vectơ địa điểm  $D_j$ . Như vậy mối quan hệ giữa 2 nhóm vectơ giống  $V_i$  và địa điểm  $D_j$  cho ta hình ảnh gần đúng về mối quan hệ kiểu gen  $\times$  môi trường trong bảng A.

Việc gần đúng tốt hay kém tùy thuộc vào tỷ lệ giữa tổng hai giá trị riêng của hai thành phần chính đầu so với tổng của tất cả các giá trị riêng. Thông thường nếu tỷ lệ này trên 70% thì những kết luận dựa vào hình ảnh của các

vectơ giống và địa điểm tuy không thật chính xác về mặt định lượng nhưng có thể giúp ích khá nhiều về mặt định tính.

Tích vô hướng của hai vectơ  $V_i \times D_j$  về mặt hình học bằng tích của hai chiều dài (độ lớn) của hai vectơ với cosin của góc giữa hai vectơ do đó:

So sánh các hình chiếu của các vectơ giống  $V_i$  trên cùng một vectơ  $D_j$  tức là so sánh các giống trong cùng một địa điểm.

So sánh các hình chiếu của các vectơ giống  $V_i$  trên một vectơ  $D_{TB}$  (minh họa cho một địa điểm trung bình, tức là địa điểm giả tạo có năng suất của mỗi giống bằng năng suất trung bình của giống đó trên mọi địa điểm) (hình 1) giúp chúng ta đánh giá chung

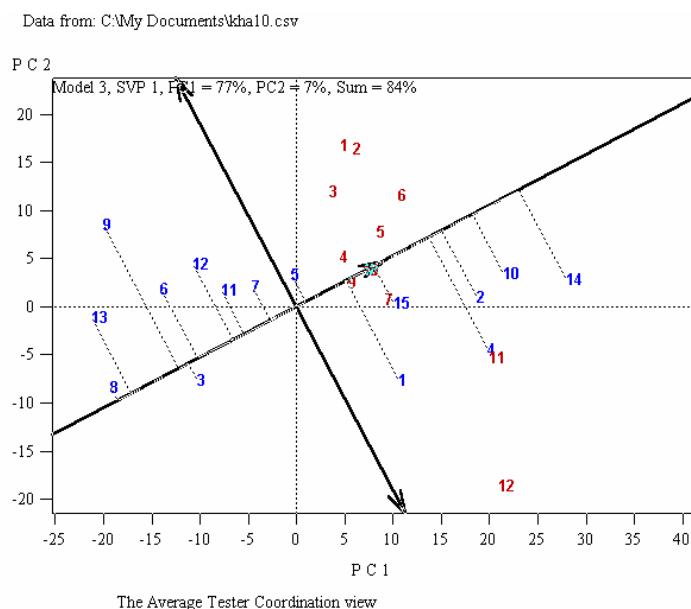
các giống trên mọi địa điểm đồng thời căn cứ vào khoảng cách của vectơ giống đối với trục đi qua địa điểm trung bình để đánh giá độ ổn định của từng giống.

So các hình chiếu của các  $D_j$  trên một vectơ  $V_i$  hoặc một nhóm  $V_i$  để đánh giá các địa điểm.

Nếu vẽ đồng thời tất cả các  $V_i$  và các  $D_j$  thì có thể tạo ra một hình ảnh giúp chúng ta nhận xét tại mỗi địa điểm nên trồng giống nào và không nên dùng giống nào.

Từ những ứng dụng đầu tiên cho bảng hai chiều kiểu gen  $\times$  môi trường đã xuất hiện nhiều nghiên cứu sâu hơn khi khảo sát đồng thời nhiều tính trạng, nghiên cứu các bảng Dialen...

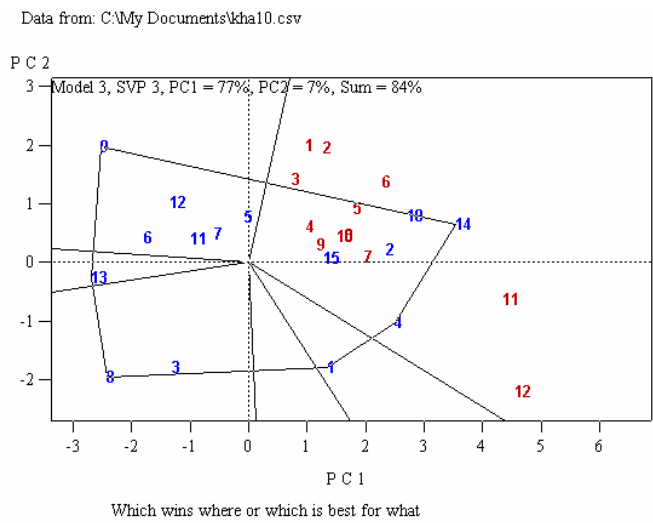
Sau đây là một vài hình ảnh của GGE biplotL



**Hình 1. Khoảng cách 15 giống đến trục đi qua địa điểm trung bình cho ta hình ảnh về độ ổn định của 15 giống (15 giống và 12 địa điểm)**

Từ hình 1 có một vài nhận xét: Căn cứ vào hình chiếu trên trục 1 (theo hướng Tây Nam - Đông Bắc) có thể thấy các giống  $V_{14}$ ,  $V_{10}$ ,  $V_2$  có năng suất trung bình cao. Các giống  $V_8$ ,  $V_{13}$ ,  $V_9$  có năng suất thấp.

Các giống  $V_{13}$ ,  $V_1$ ,  $V_4$  nằm xa trục 1 (theo hướng song song với trục 2) như vậy là giống có hệ số biến động xung quanh đường hồi quy cao tức là kém ổn định.



**Hình 2. Vẽ đồng thời 15 giống và 12 địa điểm để xem tại một địa điểm cụ thể nên trồng giống nào thì phù hợp**

Từ hình 2, nhận xét sơ bộ như sau: Các giống  $V_9$ ,  $V_{14}$ ,  $V_4$ ,  $V_1$ ,  $V_8$ ,  $V_{13}$  tạo thành một đa giác bao trùm toàn bộ các giống. Các đường thẳng góc với các cạnh chia toàn bộ mặt phẳng thành các khu vực. Giống  $V_{14}$  và  $V_4$  nằm ở khu vực tại đó có nhiều địa điểm nhưng các địa điểm  $D_{11}$ ,  $D_{12}$ ,  $D_1$ ,  $D_2$ ,  $D_6$  nằm xa gốc toạ độ nhất, như vậy có thể coi như các tích số của các vectơ  $V_{14}$ ,  $V_4$  với các vectơ địa điểm nói trên là lớn, tức là các địa điểm nói trên là thích hợp với hai giống  $V_{14}$  và  $V_4$ .

#### TÀI LIỆU THAM KHẢO

- Eberhart S. A. and Russel W. A. (1966). Stability parameters for comparing varieties. *Crop Sci* 6:36 - 40.
- Finlay K. W. and Wilinon G. N. (1963) The Analysis of adaptation in a plant breeding programme *Aust J agri Res* 14: 742 - 754.

- Hanson W.D. (1970). Genotypic Stability *Theor Appl Genet* 40: 226-231
- Perkins J.M. and Jinks J.L. (1968). Environmental and genotype- environmental components of variability *Heredity* 19: 237- 245
- Weikai Yan (2001). GGEbiplot A windows application for Graphical Analysis of multienvironmental trial data and other types of two way data. *Agronomy Journal* 93:1111-1118.
- Weikai Yan. Software Ggebiplot. (Bản Beta Ggebiplot trên mạng tại địa chỉ [w.w.w.ggebiplot.com](http://w.w.w.ggebiplot.com)).
- Lin.C.S., Binns. M. R, Lefkovitch.L. P. (1986). Stability analysis, where do we stand *Crop Science* Vol 26 896-900.

