DOI: 10.37317/biul-2008-0011

ANNA RAJFURA
WIESŁAW MĄDRY
Katedra Doświadczalnictwa i Bioinformatyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Wydzielanie grup miejscowości na podstawie serii doświadczeń wielokrotnych ze zmiennym składem odmian w latach przy użyciu pakietu SEQRET
Część I. Teoretyczne podstawy analizy retrospektywnej

The clustering of locations based on multi-environment trials with different cultivars across years using the SEQRET package
Part I. Theoretical development of retrospective analysis

W pracy przedstawiono teoretyczne podstawy analizy retrospektywnej na niekompletnych danych historycznych pochodzących z wieloletnich serii doświadczeń w wielu miejscowościach. Analiza ta stanowi integralną część metody pattern analysis i stosuje się do wydzielania grup miejscowości, w których odmiany są podobnie zróżnicowane pod względem plonu. Praca prezentuje aspekt wyznaczania odległości pomiędzy miejscowościami uśrednionych poprzez lata. Opisana metodyka została wykorzystana w procedurach pakietu SEQRET, co zaprezentowano na przykładzie w Części II.

Słowa kluczowe: retrospektywna sekwencyjna analiza skupień, niekompletne historyczne bazy danych, interakcja genotypowo-środowiskowa

This work presents the theoretical development of retrospective analysis methods, which are appropriate to study unbalanced historical data sets from multi-environmental series of experiments carried out for many years. These methods are the first part of pattern analysis methods and they are used for clustering locations in the way by which they discriminate among genotypes. The paper describes an idea of calculating proximities between locations and averaging proximity matrices over years. These procedures were used in the SEQRET package, which has been exemplified in Part II of the work.

Key words: retrospective sequential pattern analysis, unbalanced historical databases, genotype × environment interaction

WSTĘP

Do celów badania odmian w programie hodowlano-odmianowym przeprowadza się serie doświadczeń w wielu miejscowościach, powtórzonych przez wiele lat. W takich seriach doświadczeń te same odmiany są badane w wielu latach. Jednak w seriach doświadczeń przedrejestrowych w poszczególnych latach zmienia się skład odmian (jedynie odmiany wzorcowe powtarzane są w wielu latach). Ponadto zbiór miejscowości nie zawsze jest taki sam w każdym roku. Dane pochodzące z serii doświadczeń przedrejestrowych zebrane z wielu lat tworzą historyczną bazę danych, która stanowi trójkierunkową klasyfikację genotypy × miejscowości × lata (ang. GLY — genotype, location, year) o bardzo dużej niekompletności. Mimo tego, wymieniony rodzaj danych jest przydatny także do badaniu podobieństw między miejscowościami, w których przeprowadzane były doświadczenia. Podobieństwa między miejscowościami badane są pod względem rodzaju zróżnicowania odmian dla ważnej cechy (głównie plonu). Ten sposób postępowania umożliwia grupowanie miejscowości, w których odmiany są zróżnicowane podobnie. Takie grupowanie miejscowości jest bardzo ważnym aspektem badań metodyczno-hodowlanych nad doskonaleniem sieci stacji testujących odmiany w procesie ich hodowli i rejestracji (Lawrence i DeLacy, 1993; DeLacy i in., 1994; Gauch i Nobel, 1997; Zhang i in., 2006). Do wykonania analiz tego rodzaju można zastosować metody pattern analysis, czyli łączne zastosowanie metod analizy skupień oraz analizy składowych głównych (Williams, 1976), dzięki czemu możliwe jest grupowanie badanych miejscowości oraz ocena podobieństw pomiędzy nimi ze względu na sposób różnicowania odmian pod kątem obserwowanej cechy (np. plonowania).

Braki danych w klasyfikacji genotypy × miejscowości × lata mogą znacznie utrudnić ich analizę i wnioskowanie w kontekście podobieństwa miejscowości. W szczególności wymienione wcześniej metody analizy skupień wymagają wyznaczenia odległości pomiędzy miejscowościami, podczas, gdy przy braku wyników ze wszystkich miejscowości klasyczne podejście analityczne nie może być zastosowane. Kolejną cechą danych historycznych jest występowanie interakcji genotypowo-środowiskowej w poszczególnych latach, którą powinna uwzględniać metoda grupowania miejscowości.

DeLacy i wsp. (1996) zaproponowali metodę odpowiednią do analizowania niekompletnych danych historycznych pod kątem oceny odmiennego zróżnicowania badanej cechy odmian w rozpatrywanych miejscowościach i grupowania podobnych miejscowości. Opiera się ona na zastosowaniu pewnego sposobu uśredniania macierzy odległości między miejscowościami poprzez lata, a dodatkową jej zaletą jest możliwość przeprowadzenia grupowania miejscowości na podstawie ocen efektów interakcji genotypowo-środowiskowej, a nie na wartościach cechy obserwowanej w doświadczeniu. Metodykę zaproponowaną w pracy DeLacy i wsp. (1996) zastosowano w pakiecie komputerowym SEQRET (nazwa od ang. SEQuential RETrospective), którego autorami są DeLacy i wsp. (1998).

Celem pierwszej części pracy jest przedstawienie podstaw teoretycznych metody statystycznej do opracowania danych w klasyfikacji trójkierunkowej genotypy × miejscowości × lata, które są w znacznym stopniu niekompletne. Metody te służą

głównie uzyskaniu kompletnej macierzy odległości pomiędzy miejscowościami. Wyniki analiz danych rzeczywistych z serii doświadczeń przedrejestrowych z pszenicą ozimą przedstawiono w drugiej części pracy.

IDEA METODY

W wyniku przeprowadzenia w programie hodowlanym serii doświadczeń wielokrotnych w kolejnych latach, ale ze zmiennym składem odmian w poszczególnych latach, uzyskuje się dane obserwowanej cechy (głównie plonu), które tworzą trójkierunkową klasyfikację danych genotypy × miejscowości × lata (ang. GLY — genotype, location, year) o bardzo dużej niekompletności. Na podstawie danych dla każdego roku wyliczane są odległości między miejscowościami, które tworzą macierz odległości. Jeśli w pewnym roku brakuje danych z miejscowości, to niemożliwe jest wyznaczenie odległości i na skutek tego macierz odległości dla tego roku ma pusty wiersz i kolumnę. Proponowana metoda rozwiązuje ten problem przez zastosowanie uśredniania odległości między miejscowościami poprzez lata, a następnie wyeliminowanie tych miejscowości, które generują występowanie pustych komórek. Sposób ten daje w wyniku zrównoważoną macierz odległości (bez pustych komórek), którą można następnie analizować standardowymi metodami analizy skupień. Narzędziem do identyfikacji braków danych są opisane dalej macierze incydencji.

Macierze incydencji

Niech y_{ijk} będzie oceną wartości cechy obserwowanej w doświadczeniu w i-tej miejscowości, j-tym roku, dla k-tego genotypu, $n_{g(j)}$ liczbą różnych genotypów uprawianych w miejscowościach w j-tym roku, a n_g maksymalną liczbą genotypów spośród wszystkich lat. Miejscowości w liczbie n_l oraz lata w liczbie n_y identyfikowane są jednoznacznie w całym zbiorze danych, natomiast genotypy identyfikowane są jednoznacznie tylko w ramach roku. Zakłada się przy tym, że genotypy w każdym roku są reprezentacją zasobów genowych w badaniu. Przyjmijmy, że wartości cechy obserwowanej w doświadczeniu (np. plonowania) tworzą macierz o n_l wierszach i n_g kolumnach dla każdego z n_y lat. W macierzach tych występują puste wiersze dla tych miejscowości, z których brak wartości obserwowanej cechy w danym roku i również dla ostatnich n_g - $n_{g(j)}$ kolumn, gdy liczba genotypów uprawianych w danym roku jest mniejsza od maksymalnej n_g. W przedstawianej metodzie definiuje się macierze incydencji $\mathbf{\Delta}$, $\mathbf{\Gamma}$, $\mathbf{H}$, $\mathbf{Z}$, $\mathbf{K}$ (DeLacy i in., 1990), których elementy odnoszą się do istnienia lub nieistnienia wyników lub par wyników z doświadczeń.

Pierwsza z macierzy incydencji $\mathbf{\Delta} = (\delta_{ijk})$ opisuje istnienie wyników ze względu na genotypy w następujący sposób:

$$\delta_{ijk} = \begin{cases} 1, \textit{ gdy wynik } y_{ijk} \textit{ istnieje}, \\ 0, \textit{ w przeciwnym przypadku}. \end{cases} \quad (1)$$

Przy oznaczeniach: $\delta_{ij\bullet} = \sum_{k=1}^{n_g} \delta_{ijk}$, $\delta_{i\bullet\bullet} = \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk}$, $\delta_{\bullet\bullet\bullet} = \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk}$, (2)

$\delta_{ij\bullet}$ jest liczbą wyników dla genotypów uprawianych w *i*-tej miejscowości w *j*-tym roku, $\delta_{i\bullet\bullet}$ jest liczbą wyników dla genotypów uprawianych w *i*-tej miejscowości w ciągu wszystkich lat, $\delta_{\bullet\bullet\bullet}$ jest liczbą wyników dla genotypów uprawianych we wszystkich miejscowościach w ciągu wszystkich lat. Zauważmy, że element $\delta_{ij\bullet}$ przyjmuje wartość zero, tylko wtedy, gdy brak wyników z *i*-tej miejscowości dla *j*-tego roku.

Druga macierz incydencji $\mathbf{\Gamma} = (\gamma_{ij})$ opisuje istnienie wyników ze względu na miejscowości:

$$\gamma_{ij} = \begin{cases} 1, \textit{ gdy } \delta_{ij\bullet} > 0, \\ 0, \textit{ w przeciwnym przypadku.} \end{cases} \tag{3}$$

Przy oznaczeniach: $\gamma_{i\bullet} = \sum_{j=1}^{n_y} \gamma_{ij}$, $\gamma_{\bullet j} = \sum_{i=1}^{n_l} \gamma_{ij}$, $\gamma_{\bullet\bullet} = \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \gamma_{ij}$, (4)

$\gamma_{i\bullet}$ jest liczbą lat, w których pojawiła się *i*-ta miejscowość, $\gamma_{\bullet j}$ jest liczbą miejscowości w których przeprowadzono doświadczenia w *j*-tym roku, $\gamma_{\bullet\bullet}$ jest liczbą wyników uzyskanych we wszystkich kombinacjach miejscowość x rok (we wszystkich seriach doświadczeń wielokrotnych w przeciągu wszystkich lat).

Kolejne trzy macierze incydencji opisują istnienie par wyników (np. dla tego samego genotypu z dwóch różnych miejscowości w tym samym roku). Dla dwóch wyników y_{ijk}, $y_{i'jk}$ definiujemy:

$$\eta_{ii'jk} = \delta_{ijk} \cdot \delta_{i'jk} = \begin{cases} 1, \textit{ gdy oba wyniki } y_{ijk} \textit{ i } y_{i'jk} \textit{ istnieją}, \\ 0, \textit{ w przeciwnym przypadku.} \end{cases} \tag{5}$$

Przy oznaczeniach: $\eta_{ii'j\bullet} = \sum_{k=1}^{n_g} \eta_{ii'jk}$, $\eta_{ii'\bullet\bullet} = \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \eta_{ii'jk}$, $\eta_{i\bullet\bullet\bullet} = \sum_{i'\neq i} \eta_{ii'\bullet\bullet}$, (6)

$\eta_{ii'j\bullet}$ jest liczbą par wyników dla tych samych genotypów uprawianych w miejscowościach i, i' w j-tym roku (czyli liczbą genotypów wspólnych dla miejscowości i, i'), $\eta_{ii'\bullet\bullet}$ liczbą par wyników dla genotypów wspólnych dla miejscowości i, i' zsumowaną poprzez wszystkie lata; a $\eta_{i\bullet\bullet\bullet}$ liczbą par wyników dla genotypów wspólnych dla miejscowości i-tej i każdej innej niż i-ta zsumowaną poprzez wszystkie lata. Oznaczając $\mathbf{H} = (\eta_{ii'\bullet\bullet})$ otrzymujemy symetryczną macierz incydencji wymiaru $n_l \times n_l$.

Dalej definiujemy:

$$\zeta_{ii'j} = \begin{cases} 1, & gdy \quad \eta_{ii'j\bullet} > 0, \\ 0, & w\ przeciwnym\ przypadku. \end{cases} \tag{7}$$

Przy oznaczeniach $\zeta_{ii'\bullet} = \sum_{j=1}^{n_y} \zeta_{ii'j} \quad \zeta_{i\bullet\bullet} = \sum_{i' \neq i} \zeta_{ii'\bullet}$. (8)

element $\zeta_{ii'\bullet}$ jest liczbą lat, w których istnieją wspólne genotypy dla miejscowościami *i*, *i'*, a $\zeta_{i\bullet\bullet}$ jest liczbą par wyników dla miejscowości *i*-tej oraz innej niż *i*-ta zsumowaną poprzez wszystkie lata. Oznaczając $\mathbf{Z} = \left(\zeta_{ii'\bullet}\right)$ otrzymujemy symetryczną macierz incydencji wymiaru n_l x n_l. Jeśli pewne dwie miejscowości nigdy nie wystąpiły razem w tym samym roku, element $\zeta_{ii'\bullet}$ będzie miał wartość zero. Pomiędzy takimi miejscowościami nie można wyznaczyć odległości.

Następnie definiujemy:

$$\kappa_{ii'} = \begin{cases} 1, & gdy \quad \zeta_{ii'\bullet} > 0, \\ 0, & w\ przeciwnym\ przypadku. \end{cases} \tag{9}$$

Zatem $\kappa_{ii'}$ wyraża, czy para wyników dla miejscowości *i, i'* wystąpiła jednocześnie przynajmniej w jednym roku i wówczas

$$\kappa_{i\bullet} = \sum_{i \neq i'} \kappa_{ii'} \tag{10}$$

jest liczbą miejscowości, z którymi *i*-ta miejscowość była porównywana przynajmniej w jednym roku. Oznaczając $\mathbf{K} = \left(\kappa_{ii'}\right)$ otrzymujemy kwadratową symetryczną macierz incydencji wymiaru $n_l \times n_l$.

Macierze **H**, **Z** i **K** mają element zerowy poza przekątną, kiedy brak wspólnych genotypów dla dwóch miejscowości i w konsekwencji macierz odległości między miejscowościami ma odpowiedni element pusty.

Model

Przy $\delta_{ij\bullet} > 0$, przyjęty model opisany został następującym równaniem:

$$y_{ijk} = m_{ij} + \left(g \mid y \mid l\right)_{ijk} \tag{11}$$

gdzie *i* jest liczbą miejscowości, *i*=1, ..., n_l; *j* jest liczbą lat, w których wystąpiła *i*-ta miejscowość, *j*=1, ..., $\gamma_{i\bullet}$; *k* jest liczbą genotypów uprawianych w kombinacji *i*-ta miejscowość x *j*-ty rok, *k*=1, ..., $\delta_{ij\bullet}$, m_{ij} jest średnią dla kombinacji *i*-ta miejscowość x *j*-ty rok oraz $\left(g \mid y \mid l\right)_{ijk}$ jest efektem *k*-tego genotypu w kombinacji *i*-ta miejscowość x *j*-ty rok. Dalej oceny wielkości $\left(g \mid y \mid l\right)_{ijk}$ oznaczane będą przez x_{ijk}.

Transformacje danych

Przed wykonaniem łącznej analizy w latach zalecane jest transformowanie danych. Analizy oparte na danych nietransformowanych wykorzystują całą informację o zmienności w układzie danych, podczas gdy transformacje pozwalają usuwać niektóre składniki zmienności uwypuklając różne aspekty wzorca tkwiące w danych. Oczywiście wybór transformacji powinien być uzależniony od celu badań. (DeLacy i in., 1990). W metodach *pattern analysis* stosowanych do badania zróżnicowania między miejscowościami zalecane jest stosowanie centrowania (Lawrence i DeLacy (1993)) lub standaryzowania (DeLacy i in. (1994) i Mirzawan i in. (1994)) danych względem miejscowości.

Centrowanie danych względem miejscowości polega na odjęciu od każdego wyniku (np. plonu) średniej wyników poprzez miejscowości w danym roku (przez co z wartości cechy wydobyty zostaje efekt interakcji). Natomiast standaryzowanie obejmuje ponadto podzielenie otrzymanych wyników przez odchylenie standardowe dla miejscowości. Po tej transformacji średnie dla miejscowości wynoszą zero, a wariancje jeden dla każdego roku.

Rezultat standaryzacji danych w przedstawianej metodzie opisuje wzór:

$$w_{ijk} = \frac{x_{ijk}}{v_{ij}}, \text{ gdzie } v_{ij}^2 = \frac{1}{\delta_{ij\bullet} - 1} \sum_{k=1}^{n_g} \delta_{ijk} x_{ijk}^2 , \tag{12}$$

gdzie x_{ijk} są ocenami metody najmniejszych kwadratów, a v_{ij}^2 jest fenotypową wariancją dla miejscowości w danym roku.

Własności transformacji stosowanych w pakiecie SEQRET zostały opisane w pracy Fox i Rosielle (1982).

Odległości między miejscowościami w jednym roku oraz w wielu latach

Zalecaną miarą do wydzielania grup miejscowości o podobnym różnicującym wpływie na genotypy jest kwadrat odległości euklidesowej między miejscowościami. Niech $D_{ii'j}$ oznacza kwadrat odległości euklidesowej między miejscowościami *i, i'* w *j*-tym roku. Pamiętając, że $\eta_{ii'j\bullet}$ wyraża liczbę par wyników dla miejscowości *i, i'* w *j*-tym roku, $D_{ii'j}$ można wyrazić wzorem:

$$D_{ii'j} = \frac{1}{\eta_{ii'j\bullet}} \sum_{k=1}^{n_g} \eta_{ii'jk} \left(w_{ijk} - w_{i'jk} \right)^2 , \tag{13}$$

przy założeniu, że para wyników istnieje w *j*-tym roku dla miejscowości *i, i'*. Jeśli miejscowości *i, i'* nie występują jednocześnie w *j*-tym roku, to $\eta_{ii'j\bullet}$ jest zerem i nie można wyliczyć odległości między tymi miejscowościami w danym roku (wówczas macierz odległości dla tego roku nie zawiera wyniku dla miejscowości *i, i'*).

Dalej niech $D_{ii'}$ oznacza kwadrat odległości euklidesowej między miejscowościami *i, i'* poprzez wszystkie lata. Pamiętając, że $\eta_{ii'\bullet\bullet}$ wyraża liczbę par wyników dla miejscowości *i, i'* w przeciągu wszystkich lat, $D_{ii'}$ można wyrazić wzorem:

$$D_{ii'} = \frac{1}{\eta_{ii'\bullet\bullet}} \sum_{j=1}^{n_y} \eta_{ii'j\bullet} D_{ii'j} \,. \tag{14}$$

Jeśli dwie miejscowości nigdy nie wystąpiły jednocześnie w tym samym roku, $\eta_{ii'\bullet\bullet}$ jest zerem i wtedy kwadrat odległości euklidesowej między dwiema miejscowościami nie jest określony. Wówczas macierz odległości wyznaczanych poprzez lata nie zawiera wyniku dla miejscowości *i, i'* (ma pustą komórkę).

Eliminowanie pustych komórek w macierzy odległości

Macierz odległości pomiędzy miejscowościami poprzez lata **D** ma wymiar $n_l \times n_l$. Jeśli nie występują w niej puste komórki, grupowanie miejscowości można przeprowadzić klasycznymi metodami analizy skupień, w przeciwnym przypadku prezentowana metoda przeprowadza eliminację miejscowości, które generują te braki według opisanej niżej reguły.

Wśród miejscowości z najmniejszymi wartościami $\kappa_{i\bullet}$ należy znaleźć te z najmniejszymi $\eta_{i\bullet\bullet\bullet}$. Jeśli $\kappa_{i\bullet}$ jest mniejsze niż $n=n_l$-1, usunąć *i*-ty wiersz i *i*-tą kolumnę z macierzy **D**, **H**, **Z** i **K**, co zmniejszy stopień tych macierzy o 1. Od nowa wyliczyć $\kappa_{i\bullet}$ oraz $\eta_{i\bullet\bullet\bullet}$. Powtarzać, aż najmniejsze $\kappa_{i\bullet}=n$, tj. gdy wszystkie $\kappa_{i\bullet}$ są równe i nie ma pustych komórek w macierzach **D**, **H**, **Z** i **K** (takie macierze nazwano zredukowanymi). Przy równych liczebnościach (niektóre $\eta_{i\bullet\bullet\bullet}$ równe i $\kappa_{i\bullet}$ mniejsze niż *n*) eliminować należy pierwszą lub ostatnią (najmniejsze lub największe *i*) wiersz oraz kolumnę.

Skutkiem zastosowania opisanej reguły jest pozostawienie w cyklu eliminacji miejscowości z największą liczbą porównań dla ustalonej liczby innych miejscowości, do których były porównywane w przynajmniej jednym roku.

Dołączenie miejscowości wyeliminowanych z powodu braku porównań

Uzyskanie zredukowanej macierzy odległości może spowodować wyeliminowanie znacznej liczby miejscowości. Z tego względu po wydzieleniu grup miejscowości pozostawionych, można przyporządkować miejscowość wyeliminowaną do jednej z istniejących grup. Opisywana metoda umożliwia dołączenie wyeliminowanej miejscowości *i'* do grupy *I* o najbliższym centroidzie, a kwadrat odległości euklidesowej między obiektami określa wzór:

$$d^2_{i'I} = \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk} (x_{i'jk} - \bar{x}_{Ijk})^2 \text{ , gdzie } \bar{x}_{Ijk} = \sum_{i \in I} \delta_{ijk} x_{ijk} \,. \tag{15}$$

Ocena dopasowania modeli

Niech wyrażenie $\hat{x}_{ijk}$ oznacza przewidywaną wartość *k*-tej odmiany w *i*-tej miejscowości w *j*-tym roku. Kiedy grupy miejscowości zostaną już wydzielone, można przyjąć za $\hat{x}_{ijk}$ wartość średnią obliczoną dla *I*-tej grupy według wzoru:

$$\hat{x}_{ijk} = \overline{x}_{Ijk} = \sum_{i \in I}^{n_l} \delta_{ijk} x_{ijk} \ . \tag{16}$$

Dalej niech *SS(T)* oznacza całkowitą sumę kwadratów:

$$SS(T) = \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk} (x_{ijk} - \overline{x}_{ij\bullet})^2 = \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk} x_{ijk}^{\ 2} \ , \tag{17}$$

gdyż x_{ijk} sumują się do zera dla każdej kombinacji miejscowość x rok. Wyrażenie *SS(T)* można zapisać w postaci sumy składników:

$$SS(T) = \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk} \hat{x}_{ijk}^{\ 2} + \sum_{i=1}^{n_l} \sum_{j=1}^{n_y} \sum_{k=1}^{n_g} \delta_{ijk} (x_{ijk} - \hat{x}_{ijk})^2 = SS(Exp) + SS(Res) \tag{18}$$

(nazwy od ang. explained and residual), a współczynnik determinacji wyznaczony ze wzoru:

$$R^2 = \frac{SS(Exp)}{SS(T)} \tag{19}$$

przyjąć jako całkowitą (tzn. wyznaczoną poprzez lata) miarę efektywności modelu.

Sumy kwadratów odchyleń można też rozdzielić na lata:

$$SS(T)_j = \sum_{i=1}^{n_l} \sum_{k=1}^{n_g} \delta_{ijk} \hat{x}_{ijk}^{\ 2} + \sum_{i=1}^{n_l} \sum_{k=1}^{n_g} \delta_{ijk} (x_{ijk} - \hat{x}_{ijk})^2 = SS(Exp)_j + SS(Res)_j \tag{20}$$

a współczynnik determinacji wyznaczony ze wzoru:

$$R_j^2 = \frac{SS(Exp)_j}{SS(T)_j} . \tag{21}$$

przyjąć jako miarę efektywności modelu w *j*-tym roku.

PODSUMOWANIE

Opisana metoda wykorzystująca uśrednianie macierzy odległości poprzez lata (stąd określenie „analiza retrospektywna") pozwala przeprowadzić analizę skupień dla danych w znacznym stopniu niekompletnych. Dodatkową zaletą jest możliwość przydziału wyeliminowanych miejscowości do wydzielonych grup, dzięki czemu uzyskuje się informacje o tych miejscowościach, które z przyczyn matematycznych, a nie przyrod-

niczych zostały wyeliminowane ze standardowego postępowania. Metody grupowania obiektów stanowią część postępowania określanego mianem pattern analysis. Opis danych uzupełniają metody ordynacyjne analizujące zależności między badanymi obiektami w wybranym ciągu lat (stąd określenie „analiza sekwencyjna"), pominięte w tej pracy, chociaż zawarte w pakiecie SEQRET.

LITERATURA

DeLacy I. H., Basford K. E., Cooper M., Fox P. N. 1996. Retrospective analysis of historical data sets from multi-environment trials-Theoretical development. In: Cooper M., Hammer G.L. (eds) Plant Adaptation and Crop Improvement. CAB International: 243 — 267.

DeLacy I. H., Cooper M., Lawrence P. K. 1990. Pattern analysis over years of regional variety trials: relationship among sites. In: Kang M.S. (ed.) Genotype-by-Environment Interaction and Plant Breeding. Louisiana State University, Baton Rouge, Louisiana (12-13 February): 189 — 213.

DeLacy I.H., Fox P.N., Corbett J.D., Crossa J., Rajaram S., Fisher R.A. and van Ginkel M. 1994. Long-term association of locations for testing spring bread wheat. Euphytica 72: 95 — 106.

DeLacy I. H., Cooper M., Basford K. E. 1996. Relationships among analytical methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection. In: Kang, M. S., Gauch, H.G. (eds) Genotype-by-Environment Interaction: New Perspectives. CRC Press, Boca Raton, Florida: 51 — 84.

DeLacy I. H., Basford K. E., Cooper M., Fox P. N. 1998. The SEQRET Package: Computer Programs for Retrospective Pattern Analysis, Version 1.1. The University of Queensland, Brisbane 4072, Australia.

Fox P. N., Rosielle A. A. 1982. Reducing the influence of environmental main-effects on pattern analysis of plant breeding environments. Euphytica 31: 645 — 656.

Gauch H. G., Zobel R. W. 1997. Identifying mega-environments and targeting genotypes. Crop Sci. 37: 311 — 326.

Lawrence P. K. and DeLacy I. H. 1993. Classification of locations in regional cotton variety trials where trial entries change over years. Field Crops Research 34: 195 — 207.

Mirzawan P. D. N., Cooper M., DeLacy I. H., Hogarth D. M. 1994. Retrospective analysis of the relationships among the test environments of the Southern Queensland sugarcane breeding program. Theoretical and Applied Genetics 88: 707 — 716.

Williams W. T. 1976. Pattern Analysis in Agricultural Science. Elsevier Scientific Publishing Company, Amsterdam.

Zhang Y., He Z., Zhang A., van Ginkel M., Ye G. 2006. Pattern analysis on grain yield of Chinese and CIMMYT spring wheat cultivars grown in China and CIMMYT. Euphytica 147: 409 — 420.