Postępy Nauk Rolniczych nr 1/2005

Metody statystyczne analizy składowych plonu

Marcin Kozak, Wiesław Mądry
Katedra Biometrii, Szkoła Główna Gospodarstwa Wiejskiego
ul. Nowoursynowska 159, 02-776 Warszawa
e-mail: marcin.kozak@omega.sggw.waw.pl, w.madry@agrobiol.sggw.waw.pl

Słowa kluczowe: składowe plonu, rozwój sekwencyjny, rozwój niesekwencyjny, metody statystyczne

Wstęp

Analiza zależności między plonem roślin a jego składowymi, zwana zwykle analizą składowych plonu (ang. *yield component analysis*), interesuje badaczy w wielu dyscyplinach nauk rolniczych, zwłaszcza w agronomii i ogrodnictwie [7, 9, 27, 21, 23, 30, 34, 39]. W badaniach stosowanych dostarcza ona podstawowych przesłanek do identyfikacji pośrednich kryteriów efektywnej selekcji genotypów o wysokim potencjale plonowania [1, 2, 7, 12, 33, 37, 44, 48] oraz do efektywnego wykorzystania zasobów przyrodniczych w rolnictwie poprzez trafną rejonizację odmian [9, 14, 17, 26, 41] oraz ich uprawę [9, 30, 34, 39, 50].

Fraser i Eaton [8] omówili w swej pracy metody statystyczne stosowane w analizie składowych plonu. Z ich przeglądu wynika, że do roku 1983 były to raczej ogólne metody statystyczne niż takie, które uwzględniałyby statystyczny charakter omawianego zagadnienia; większość omawianych w pracy [8] metod nie uwzględniała multiplikatywnej zależności plonu od składowych, jak również kolejności rozwoju tych cech.

Celem pracy jest przedstawienie metod statystycznych proponowanych w literaturze do zastosowania w analizie multiplikatywnych składowych plonu. Nazwa tych składowych pochodzi stąd, że plon jest wyrażony jako ich iloczyn, co przedstawia model o postaci [4, 21, 23, 31]:

$$Y = \prod_{i=1}^{k} X_i = X_1 \dots X_k \tag{1}$$

gdzie Y jest plonem, $X_i, i = 1, \dots, k$ zaś są jego składowymi, zwanymi multiplikatywnymi składowymi plonu [4, 21, 31].

W pracy zakładamy, że jedne składowe plonu rozwijają się równocześnie i współzależnie w trakcie ontogenezy [15, 21, 31], inne zaś w określonej kolejności (sekwencji) w trakcie kolejnych etapów ontogenezy [1, 9, 21, 23]. Sekwencyjny rozwój składowych plonu implikuje postać przyczynowo-skutkową związków między składowymi. Założenie to jest potrzebne, gdyż metody analizy multiplikatywnych składowych plonu dzielą się na metody niesekwencyjnej analizy składowych plonu (są one dostosowane do pierwszej grupy składowych), oraz metody sekwencyjnej analizy składowych plonu (dostosowane dla drugiej grupy składowych).

W niniejszej pracy przedstawimy krótką charakterystykę ważniejszych metod, a przede wszystkim komentarz odnośnie ich formalnej i merytorycznej poprawności oraz przydatności praktycznej. Czytelnik zainteresowany szczegółami danej metody może sięgnąć do literatury cytowanej w niniejszej pracy.

Metody niesekwencyjnej analizy składowych plonu

Metoda Hardwicka i Andrewsa

Jedną z metod analizy składowych plonu rozwijających się niesekwencyjnie zaprezentowali Hardwick i Andrews [15]. Zastosowali oni następującą transformację logarytmiczną plonu Y oraz jego składowych X_i, $i = 1, \ldots, k$:

$$\ln Y = \sum_{i=1}^{k} \ln X_i \tag{2}$$

Biorąc pod uwagę (2), autorzy dokonali podziału wariancji transformowanego plonu na składniki wariancyjno-kowariancyjne cech składowych. Na tej podstawie autorzy zaproponowali wskaźnik W następującej postaci:

$$W = \sigma^2_{\ln Y}\left(\sum_{i=1}^{k}\sigma_i\right)^{-2} = \left(\sum_{i=1}^{k}\sigma_i^2 + Q\right)\left(\sum_{i=1}^{k}\sigma_i^2 + Z\right)^{-1} \tag{3}$$

gdzie $\sigma^2_{\ln Y}$, σ_i^2, $i = 1, \ldots, k$, jest wariancją logarytmowanego plonu i i-tej logarytmowanej składowej, $Z = \sum_{i=1}^{k}\sum_{i'=1, i'\neq i}^{k}\sigma_i \cdot \sigma_{i'}$, zaś $Q = \sum_{i=1}^{k}\sum_{i'=1, i'\neq i}^{k}\mathrm{Cov}(\ln X_i, \ln X_{i'})$.

Wskaźnik (3) jest tak zbudowany, że jego wartość bliska zeru oznacza zachowanie kompensacyjne (antagonistyczne, tłumiące) składowych plonu. Polega ono na tym, że zwiększanie wartości jednych składowych plonu jest kompensowane przez zmniejszanie wartości innych składowych. Jeżeli stopień wzajemnej kompensacji (ujemnych korelacji) składowych plonu jest duży, to zmienność plonu, określona za pomocą wariancji logarytmu plonu, $\sigma^2_{\ln Y}$, jest odpowiednio mała. O przeciwnym sposobie zachowania składowych plonu świadczą wartości wskaźnika W bliskie jedności.

Wtedy wszystkie składowe plonu są wzajemnie skorelowane dodatnio. Wartość wskaźnika W zależy od liczby składowych plonu, a więc nie jest to miara porównywalna dla różnych badań. Zatem autorzy wprowadzili względny wskaźnik W, który jest niezależny od liczby zmiennych w modelu.

Jak się okazuje, metoda zaproponowana przez Hardwicka i Andrewsa [8] nie jest pozbawiona wad. W swej pracy Piepho [29] przeprowadził konstruktywną krytykę tej metody. Zauważył on, że wskaźnik W wcale nie identyfikuje właściwości składowych sugerowanych przez autorów pomysłu. Piepho podał inną postać wskaźnika dla każdej, i-tej składowej plonu. Gdy jego wartość równa jest 1/2, wpływ i-tej logarytmowanej składowej na plon jest w pełni kompensowany przez inne składowe, co sprawia, że nie uczestniczy ona w formowaniu zmienności plonu. Wartość wskaźnika większa od 1/2 oznacza, że wzrost i-tej logarytmowanej składowej powoduje (sumarycznie) wzrost pozostałych logarytmowanych składowych (niestety nie wiadomo których). Im wartość wskaźnika jest bliższa jedności, tym relatywny stopień tych współzależności jest większy. Gdy wartość wskaźnika zaproponowanego przez Piepho jest mniejsza od 1/2, mamy do czynienia z odwrotną sytuacją, gdyż przy zwiększaniu wartości i-tej logarytmowanej składowej następuje zmniejszanie pozostałych logarytmowanych składowych.

Warto zwrócić uwagę na pewien atrybut opisywanej metody, niezwykle istotny w analizie składowych plonu. Otóż metoda ta wykorzystuje transformację logarytmiczną; dzięki temu można przeprowadzić analizę składowych plonu według modelu addytywnego, a nie multiplikatywnego. Dzięki transformacji logarytmicznej w analizie nie pojawia się błąd niedopasowania modelu (gdyż suma logarytmów składowych równa jest logarytmowanemu plonowi). Niestety, zastosowanie transformacji logarytmicznej powoduje problemy interpretacyjne, wynikające ze zmiany skali zmiennych [40]. Okazuje się, że zazwyczaj wartości współczynników korelacji prostej i parametrów funkcji regresji wielokrotnej w skali oryginalnej i logarytmicznej różnią się od siebie, nawet znacznie [10, 40]. Z tego powodu obraz zależności plonu od jego składowych przedstawiony w skali logarytmicznej jest zniekształcony w stosunku do naturalnej skali badanych cech.

Metoda Piepho

W cytowanej pracy Piepho [29] znajdują się jednak nie tylko zmodyfikowane formuły na wskaźniki proponowane przez Hardwicka i Andrewsa [15], które według Piepho nie są wystarczające w analizie składowych plonu. Autor ten zaproponował własną metodę analizy składowych plonu, będącą modyfikacją metody Hardwicka i Andrewsa.

Rozpatrzmy transformowany model plonu względem składowych (1). Prawdziwa jest relacja [29]:

$$\sigma^2_{\ln Y} = \sum_{i=1}^{k} \mathrm{Cov}(\ln Y, \ln X_1) = \sum_{i=1}^{k} C_i \tag{4}$$

gdzie $\sigma^2_{\ln Y}$ jest wariancją logarytmowanego plonu, $C_i = \mathrm{Cov}(\ln Y, \ln X_i)$ zaś jest kowariancją między plonem a i-tą składową w skali logarytmicznej.

Kowariancje C_i mogą być dodatnie lub ujemne. Stanowią one miarę wkładu i-tej składowej plonu w zmienność plonu (w skali logarytmicznej) – można je więc interpretować jako miary wpływu i-tej składowej na plon. Piepho rozpatruje względne postaci kowariancji c_i,

$$c_i = \frac{C_i}{\sigma^2_{\ln Y}} \tag{5}$$

które są porównywalną miarą nie tylko kierunku, ale też wielkości liniowego wpływu (determinacji) i-tej składowej na plon. Jeżeli kowariancje są dodatnie, wówczas kowariancje względne określają strukturę zmienności (wariancji) plonu względem determinacji przez i-te składowe. Jednak, w przypadku występowania ujemnych wartości C_i w równaniu (5), kowariancje względne c_i nie mogą określać wspomnianej struktury zmienności plonu.

Reasumując, należy stwierdzić, że w metodzie Piepho istotne są dwie kwestie metodyczne. Po pierwsze, u podstaw tej metody jest przekształcenie logarytmiczne zmiennych (o którym dyskutowaliśmy wcześniej) oraz, po drugie, proponowane przez Piepho wskaźniki nie są porównywalne. Każda składowa występuje w innej skali, wobec czego ich wariancje nie są porównywalną miarą zmienności tych cech; tym samym wskaźniki znaczenia składowych w kształtowaniu zmienności plonu, proponowane przez Piepho, nie są porównywalne między sobą, co sprawia, że ich użyteczność w prawidłowej analizie składowych plonu jest dyskusyjna.

Analiza regresji wielokrotnej i analiza ścieżek

Powszechnie stosowaną metodą do analizy składowych plonu jest analiza regresji wielokrotnej, oparta na liniowym modelu regresji plonu od jego składowych. W tym modelu składowe plonu (zmienne przyczynowe) traktowane są jako zmienne losowe. Metoda ta jest stosowana zarówno w postaci klasycznej, jak i w różnych postaciach zmodyfikowanych i dostosowanych do potrzeb rozpatrywanego zjawiska. Regresję prostą, opisującą zależność plonu od poszczególnych składowych, stosowali np. Evans [6] w badaniach nad pszenicą czy Mason i Rath [27] dla truskawki; regresję wielokrotną w swej klasycznej postaci stosowali np. Scott i in. [38] dla pszenicy oraz Worley i in. [45] dla bawełny kosmatej.

Jednak wydaje się, że klasyczna analiza regresji nie jest w pełni efektywną metodą analizy składowych plonu [21]. Najczęściej stosowaną modyfikacją metody liniowej regresji wielokrotnej z losowymi zmiennymi przyczynowymi do analizy składowych

plonu jest analiza ścieżek (ang. *path analysis*) opracowana przez Wrighta [46, 47]. Jej celem jest przede wszystkim ułatwienie i wzbogacenie charakterystyki wpływu skorelowanych zmiennych przyczynowych na zmienną skutkową. Analiza ścieżek daje możliwość przeprowadzenia podziału współczynnika korelacji prostej między każdą ze składowych a plonem na efekt bezpośredni (współczynnik ścieżki) danej składowej na plon oraz efekty pośrednie tej składowej na plon poprzez pozostałe składowe. Współczynniki ścieżek są niemianowane, co sprawia, że są porównywalną miarą wielkości wpływu składowych na plon. Należy zwrócić uwagę, że w przypadku sekwencyjnego rozwoju składowych, w analizie ścieżek wyznaczane są tzw. pozorne efekty pośrednie [19], nie mające odniesienia do rzeczywistości. Są to efekty pośrednie jednych składowych plonu poprzez inne, rozwijające się wcześniej w ontogenezie. Efekt taki nie może być traktowany jako rzeczywiście wywierany, ponieważ składowe, które rozwijają się później w ontogenezie nie mogą faktycznie modyfikować zarówno ilościowego kształtowania, jak i wpływu na plon tych składowych, które już się wcześniej, przed nimi, ukształtowały. Podział współczynnika korelacji, w przypadku wystąpienia efektów pozornych, prowadzi do fałszywych wniosków. Natomiast w przypadku niesekwencyjnego rozwoju składowych plonu wyodrębnione ich efekty pośrednie na plon występują w rzeczywistości, a nie tylko pozornie.

Charakterystykę matematyczną i interpretacyjną analizy ścieżek można znaleźć np. w pracach [19, 25, 26]. Ponadto metoda ta w kontekście zastosowania do analizy składowych plonu została szeroko omówiona przez Kozaka [21].

Analiza ścieżek jest najczęściej stosowaną metodą analizy składowych plonu, zarówno w literaturze polskiej, jak i zagranicznej, dawniej i wciąż. Ilustrują to prace Mądrego i in. [29], Rozbickiego [34] oraz Rozbickiego i Mądrego [35, 36] dla pszenżyta ozimego, Samonte i in. [37] dla ryżu, Zająca i in. [50] dla pszenicy jarej, Kumar i Kumar [24] dla kukurydzy, Xing i in. [48] dla transgenicznej bawełny (zawierającej gen *Bt – Bacillus thuryngensis*), Jag i in. [17] dla pszenicy, i wielu innych.

Metoda Hühna

Hühn [16] podał metodę analizy uwarunkowania zmienności plonu przez dwie składowe, tj. liczbę roślin na jednostce powierzchni oraz średni plon z rośliny. Autor ten podał następujący analityczny wzór na współczynnik zmienności plonu:

$$v = \frac{\sqrt{v_1^2 + v_2^2 + 2\rho_{12} \cdot v_1 \cdot v_2 + \left(1 + \rho_{12}^2\right) \cdot v_1^2 \cdot v_2^2}}{1 + \rho_{12} \cdot v_1 \cdot v_2} \tag{6}$$

gdzie v jest współczynnikiem zmienności dla plonu Y, $v = \frac{\sigma_Y}{\bar{Y}}$, v_1 i v_2 zaś są współczynnikami zmienności składowych X_1 i X_2.

Współczynnik zmienności plonu jest uzależniony od trzech parametrów – współczynników zmienności składowych (v_1 i v_2) oraz od współczynnika korelacji pomiędzy X_1 i X_2 (ρ_{12}). Co prawda Hühn wyznaczył wzory jedynie dla klasycznego przykładu dwóch składowych, tj. liczby roślin na jednostce powierzchni oraz plonu z rośliny, to podaje on, że wzory dla trzech i większej liczby składowych mogą zostać łatwo wyprowadzone. Wzór na współczynnik zmienności plonu podany przez Hühna ma skomplikowaną postać. Z tego powodu wydaje się, że metoda ta nie umożliwia zrozumiałej i czytelnej interpretacji wyników. W przypadku braku korelacji między składowymi wzór znacznie się upraszcza, plon zaś jest determinowany przez poszczególne składowe w stopniu określonym przez zmienność tychże. Za pomocą tej metody możemy stwierdzić, która z cech ma większy, a która mniejszy wpływ na zmienność plonu, nie jesteśmy jednak w stanie podać udziału składowych w zmienności plonu.

Metody sekwencyjnej analizy składowych plonu

Metody analizy składowych plonu omówione w poprzednim rozdziale oparte są na modelowaniu statystycznym, w którym pomija się kolejność rozwoju cech składowych. Na potrzebę uwzględnienia sekwencyjnego rozwoju składowych plonu w podejściach do modelowania plonu pierwsi zwrócili uwagę Grafius [14], Rasmusson i Cannell [33] oraz Thomas i in. [42]. Następnie podkreślali ją również Tai [41], Scott i in. [38], Eaton i Kyte [4], Jolliffe i in. [18]. Jak wspomnieliśmy we wstępie, nieodzowne wydaje się rozróżnienie tych dwóch rodzajów składowych plonu i odpowiednich podejść do ich analizy.

Metoda Thomasa i in.

Pierwsza metoda analizy składowych plonu, uwzględniająca ich sekwencyjny rozwój w procesie ontogenezy, była oparta na transformacji zbioru skorelowanych cech do zbioru nieskorelowanych cech, podanej przez Rao [32]. Metoda ta została dostosowana do analizy składowych plonu przez Thomasa i in. [42]. Jak zauważyli Fraser i Eaton [7], transformacja ta nie usuwa wszystkich korelacji pomiędzy zmiennymi. Faktycznie znika jedynie korelacja pomiędzy pierwszą i drugą zmienną transformowaną. Dlatego metoda ta nie była często stosowana w praktyce, z wyjątkiem prac Thomasa i in. [42] oraz Thurlinga [43] dla rzepiku oraz rzepaku.

Metoda Tai'a – analiza ścieżek złożonych

Tai [41] zaproponował zastosowanie analizy ścieżek złożonych dla składowych plonu w zagadnieniach związanych z interakcją genotypowo-środowiskową, z uwzględnieniem sekwencyjnego rozwoju tych cech. Metoda ta została zastosowana np. przez Lyncha i Tai'a [26] dla ziemniaka. Jednak w takim podejściu, tak jak

w przypadku klasycznej analizy ścieżek stosowanej w analizie sekwencyjnej, wyznacza się pozorne efekty pośrednie [21], przez co, pomimo prawidłowego założenia o sekwencyjnym rozwoju składowych, interpretacja o wpływie składowych na plon nie jest w pełni zgodna z rzeczywistością [21].

Sekwencyjna analiza składowych plonu (SYCA)

Kolejną metodą statystyczną analizy składowych plonu, która uwzględnia kolejność ich rozwoju, jest sekwencyjna analiza składowych plonu (SYCA, ang. *sequential yield components analysis*). Opracowali ją i zastosowali Eaton i Kyte [4]. Tak jak w kilku omówionych już metodach, w metodzie SYCA przeprowadza się transformację logarytmiczną (2) funkcji (1).

Kluczowym zabiegiem, wykonywanym w tej metodzie, jest ortogonalizacja logarytmowanych obserwacji składowych plonu, prowadzona zgodnie z kolejnością ich rozwoju. Metoda SYCA pozwala określić, przy formalnym zachowaniu wzajemnej niezależności (ortogonalności) składowych plonu, względną wielkość i kierunek ich liniowego wpływu na plon z uwzględnieniem kolejności kształtowania się cech w trakcie ontogenezy.

Metoda SYCA była stosowana np. przez Eatona i Kyte [4] dla truskawki oraz Yasa i Eatona [49] dla żurawiny. Eaton i in. [3] opracowali metodę dwukierunkowego podziału zmienności plonu (ang. *two-dimensional partitioning of yield variation*, w skrócie TDP), która jest wzbogaceniem metody SYCA o analizy wariancji dla ortogonalnych logarytmowanych składowych plonu oraz logarytmowanego plonu. Metodę TDP zastosowali np. Eaton i in. [3] dla ogórka, McArthur i Eaton [30] dla truskawki, Freeman i in. [8] dla maliny, Gołaszewski [11] oraz Spaner i in. [39] dla kukurydzy, Gołaszewski i in. [12] dla grochu pastewnego i bobu, Kozak [20] dla pszenżyta ozimego.

Jednak wadą metody SYCA, niestety znaczącą, jest wprowadzenie koniecznej transformacji logarytmicznej zmiennych. Problem ten został omówiony wcześniej, podczas prezentacji metody Hardwicka i Andrewsa [15], wobec czego nie będziemy się na nim koncentrować.

Metoda Sparnaaija i Bosa

Sparnaaij i Bos [40] zaproponowali metodę, która nie wymaga stosowania transformacji logarytmicznej. Autorzy ci uwzględnili następujący niemalejący ciąg współczynników determinacji dla liniowych funkcji regresji prostych plonu względem iloczynów kolejnych składowych:

$$r_1^2(Y|X_1 = V_1),\ldots,r_i^2(Y|X_1\ldots X_i = V_i),\ldots,r_1^2(Y|X_1X_2\ldots X_k = V_k = Y) = 1 \quad (7)$$

Cechy $V_1, \ldots, V_k$ w (7) nazwane zostały przez autorów pierwotnymi (głównymi) cechami rozwojowymi (ang. *primary characters*, patrz [21, 23, 40]).

Analizując współczynniki determinacji (7) można zauważyć, że określają one uwarunkowanie plonu przez proces wykształcenia kolejnych składowych. Wobec tego, wzrost wartości współczynnika determinacji plonu (7) w danym, i-tym etapie kształtowania plonu, czyli na skutek wprowadzenia kolejnej w ontogenezie i-tej składowej, Sparnaaij i Bos nazwali determinacją uzupełniającą (ang. *complementary determination*) i oznaczyli go jako $cd(X_i, Y)$. Współczynnik ten informuje o wzroście determinacji cechy złożonej (tutaj plonu) w rezultacie ontogenezy, w okresie od ukształtowania się $(i - 1)$-ej w ontogenezie składowej do ukształtowania się i-tej składowej, czyli po wprowadzeniu tej ostatniej do modelu.

Rozpatrzmy na przykład klasyczne składowe plonu ziarna zbóż na jednostce powierzchni, tj. liczbę kłosów na jednostce powierzchni (X_1), średnią liczbę ziaren w kłosie (X_2), oraz średnią masę ziarniaka (X_3). Głównymi cechami rozwojowymi w tym przypadku będą $V_1 = X_1$, $V_2 = X_1X_2$, czyli liczba ziaren na jednostce powierzchni, oraz $Y = V_3$, czyli plon ziarna na jednostce powierzchni. W takim wypadku pierwszy współczynnik determinacji uzupełniającej, $cd(X_1, Y)$, informowałby o liniowej determinacji plonu Y przez pierwszą składową X_1, czyli liczbę kłosów na jednostce powierzchni. Drugi współczynnik determinacji uzupełniającej, $cd(X_2, Y)$, informuje o wzroście determinacji plonu przez ukształtowaną jego drugą składową, tj. średnią liczbę ziaren w kłosie X_2. Trzeci współczynnik determinacji uzupełniającej, $cd(X_3, Y)$, mówi o wzroście determinacji plonu Y przez wykształconą średnią masę ziarna, jako trzecią składową plonu ukształtowaną w trakcie ostatniego etapu ontogenezy.

Nowatorskie podejście Sparnaaija i Bosa wytyczyło nowy kierunek metodyki statystycznej w analizie składowych plonu, opartej na pierwotnych cechach rozwojowych, stanowiących alternatywę dla składowych plonu. Metoda ma prostą i bezpośrednią interpretację. Należy wszakże zwrócić uwagę na fakt, że interpretacja wyników uzyskanych za pomocą tej metody odnosi się do pierwotnych cech rozwojowych, a nie do składowych; w związku z tym, choć jest ona niewątpliwie poprawna z biologicznego punktu widzenia, to jednak nie ukazuje jasno faktycznego wpływu indywidualnych składowych na plon.

Opisywana metoda została zastosowana przez jej autorów do analizy uwarunkowania wrażliwości chryzantemy (*Dendranthema grandiflora*) na szkodnika liści *Liriomyza trifolii* [2, 40]. W tym zagadnieniu wrażliwość roślin na szkodnika jest także traktowana jako cecha złożona, podobnie jak plon.

Niestety, metoda Sparnaaija i Bosa [40] nie jest pozbawiona wad. Kolejny, i-ty współczynnik determinacji uzupełniającej określony jest przez różnicę między współczynnikiem determinacji dla regresji prostej plonu od i-tej pierwotnej cechy rozwojowej, a współczynnikiem determinacji dla regresji prostej plonu od $(i - 1)$-ej pierwotnej cechy rozwojowej. Taka ocena współczynnika determinacji uzupełniającej powoduje niedoszacowanie jego wartości dla $i > 1$, z powodu niedoszacowania

współczynnika determinacji plonu po zakończeniu i-tego etapu ontogenetycznego, czyli wykształcenia *i*-tej cechy pierwotnej. Rozpatrzymy to dla $i = 2$. Niedoszacowany jest współczynnik determinacji plonu, gdyż uwzględniamy w nim jedynie drugą cechę pierwotną V_2, nie biorąc pod uwagę, że na plon może wpływać, w pewnym stopniu niezależnie od V_2, również pierwsza pierwotna cecha rozwojowa V_1. Należałoby więc rozpatrywać model liniowy plonu względem V_1 oraz V_2, nie zaś tylko względem V_2. Z tego względu niedoszacowany jest również współczynnik $cd(X_2, Y)$, przez co wpływ drugiej, w ogólnym przypadku *i*-tej, pierwotnej cechy rozwojowej na plon jest obciążony.

Sekwencyjna analiza plonu względem jego składowych

Sparnaaij i Bos [40] przeprowadzili porównanie wnioskowania za pomocą dwóch sposobów zastosowania metody SYCA. W pierwszym, klasycznym podejściu zastosowano skalę logarytmiczną zmiennych, w drugim zaś oryginalną skalę zmiennych. Uzyskane wyniki i wnioski znacznie się różniły. Gołaszewski [11] uznał, że ominięcie transformacji logarytmicznej w metodzie SYCA może być korzystne dla wiarygodności wnioskowania, więc zaproponował zastosowanie tej metody dla zmiennych w skali oryginalnej.

Rozpatrzmy zalety i wady takiego podejścia. Jego zaleta polega na tym, że analizę statystyczną wykonujemy w naturalnej i bardziej zrozumiałej oryginalnej skali plonu i jego składowych, zamiast przeprowadzać ją w skali logarytmicznej plonu i jego składowych. Wadą tego podejścia jest zmniejszenie determinacji plonu przez jego składowe z powodu przybliżenia modelu multiplikatywnego (1) za pomocą modelu liniowego. Wada ta może być czasami (chociaż z naszych badań wynika, że raczej rzadko) poważniejsza, jeżeli determinacja plonu przez jego składowe byłaby znacząco mniejsza od 100%, mimo że rzeczywista determinacja plonu przez jego składowe w modelu (1) jest równa 100% [21]. Kozak [22] pokazał, że determinacja modelu liniowego, przybliżająca multiplikatywną zależność plonu od składowych, może być nawet mniejsza od 80%. W związku z tym, wartość naukowa i przydatność praktyczna tej metody analizy składowych plonu również jest w pewnym stopniu kontrowersyjna.

Sekwencyjna analiza plonu względem pierwotnych cech rozwojowych

Wydaje się, że dwie z omówionych metod, tj. metoda Sparnaaija i Bosa oraz sekwencyjna analiza plonu względem jego składowych, dość dobrze przybliżają biologiczny obraz zależności plonu od jego składowych rozwijających się sekwencyjnie. Kozak [21] zaproponował metodę, która stanowi pewną syntezę obu tych metod. Ze względu na jej istotę została ona przez autora nazwana sekwencyjną analizą plonu względem pierwotnych cech rozwojowych. W najprostszym ujęciu, polega ona na pewnej modyfikacji metody Sparnaaija i Bosa [40], niwelującej wspomniane

wcześniej wady tej metody. Można to uczynić, stosując opisaną sekwencyjną analizę plonu dla pierwotnych cech rozwojowych w oryginalnej skali zmiennych (w przeciwieństwie do poprzedniej metody, w której stosowano sekwencyjną analizę plonu dla składowych w oryginalnej skali). Estymację efektu każdej, poczynając od drugiej, składowej plonu przeprowadza się tak, jak w poprzedniej metodzie [40].

Omawiana metoda jak dotychczas wydaje się najlepszą metodą analizy składowych plonu rozwijających się sekwencyjnie, gdyż – po pierwsze – uwzględnia matematyczną postać zależności plonu od jego składowych (przez co estymacja efektów tych składowych nie jest obciążona błędem niedopasowania modelu) oraz – po drugie – analiza składowych plonu prowadzona jest w oryginalnej skali zmiennych (przez co estymacja efektów tych składowych nie jest obciążona błędem wynikającym z transformacji zmiennych).

Podsumowanie

W pracy przedstawiliśmy różne podejścia statystyczne do analizy składowych plonu. Procedury metodyczne analizy składowych plonu podlegały i wciąż podlegają ewolucji, począwszy od pierwszej pracy Engledowa i Wadhama [5]. Ocena porównawcza metod przedstawionych w niniejszej pracy skłania nas do stwierdzenia, że wybór metody analizy składowych plonu ma znaczący wpływ na jej wyniki, a przez to na oszacowanie udziału i interpretację znaczenia składowych w kształtowaniu plonu.

W praktycznym wyborze metody analizy składowych plonu kluczowe są dwie decyzje. Najpierw badacz powinien uznać, czy składowe plonu w rozpatrywanych badaniach rozwijają się sekwencyjnie, czy niesekwencyjnie. Decyzja ta implikuje wybór grupy metod, spośród których pochodzić będzie metoda zastosowana do rozpatrywanego zagadnienia. Zastosowanie metody niesekwencyjnej dla składowych sekwencyjnych, lub odwrotnie, prowadzi do uzyskania wyników, które fałszują prawidłowości i charakter uwarunkowania plonu przez jego składowe, zgodnie z ich sukcesywnym rozwojem i ilościowym kształtowaniem w trakcie ontogenezy [1, 4, 7, 9, 12, 21, 27]. Drugą ważną decyzją jest wybór właściwej metody analizy plonu z odpowiadającej grupy metod, tj. metod niesekwencyjnych lub sekwencyjnych. I tak, spośród metod niesekwencyjnych wydaje się, że jak na razie najlepszą jest analiza ścieżek, natomiast spośród metod sekwencyjnych – sekwencyjna analiza plonu względem pierwotnych cech rozwojowych. Jeżeli przybliżenie zależności (1) za pomocą modelu liniowego jest wystarczająco dobre (tj. współczynnik determinacji jest bliski 100%), za zasadne można uznać również zastosowanie sekwencyjnej analizy plonu względem jego składowych. Wydaje się, że stosowanie innych metod sekwencyjnych jest nieprawidłowe, gdyż obecny stan wiedzy dotyczącej aplikacji metod analizy składowych plonu pozwala sądzić, że prowadzą one do wniosków zniekształconych (np. ze względu na zastosowanie transformacji logarytmicznej).

Literatura

[1] Board J.E., Kang M.S., Bodrero M.L. 2003. Yield components as indirect selection criteria for late-planted soybean cultivars. *Agron. J.* 95: 420–429.

[2] Bos I., Sparnaaij L.D. 1993. Component analysis of complex characters in plant breeding. II. The pursuit of heterosis. *Euphytica* 70: 237–245.

[3] Eaton G.W., Bowen P.A., Jolliffe P.A. 1986. Two-dimensional partitioning of yield variation. *HortScience* 21: 1052–1053.

[4] Eaton G.W., Kyte T.R. 1978. Yield component analysis in strawberry. *J. Am. Soc. Horticult. Sci.* 103: 578–583.

[5] Engledow F.W., Wadham S.M. 1923. Investigation of yield in the cereals. *J. Agric. Sci.* UK 95: 111–115.

[6] Evans L.T. 1978. The influence of irradiance before and after anthesis on grain yield and its components in microcrops of wheat grown in a constant day length and temperature regime. *Field Crops Res.* 1: 5–19.

[7] Fraser J., Eaton G.W. 1983. Applications of yield component analysis to crop research. *Field Crop Abstracts* 36: 787–796.

[8] Freeman J.A., Eaton G.W., Baumann T.E., Daubeny H.A., Dale A. 1989. Primocane removal enhances yield components of raspberries. *J. Am. Soc. Horticult. Sci.* 114: 6–9.

[9] Garcia del Moral L.F., Rharrabti Y., Villegas D., Royo C. 2003. Evaluation of grain yield and its components in durum wheat under Mediterranean conditions: An ontogenetic approach. *Agron. J.* 95: 266–274.

[10] Gierdziewicz M., Zając T. 1999. Porównanie metod szacowania siły związku między plonem ziarna pszenżyta ozimego a jego elementami strukturalnymi. Mat. Konf. *Pam. Puł.* 114: 111–118.

[11] Gołaszewski J. 1996. A method of yield component analysis. *Listy Biom.* 33: 79–88.

[12] Gołaszewski J., Idźkowska M., Milewska J. 1998. The TDP method of seed yield component analysis in grain legume breeding. *J. Appl. Genet.* 39: 299–308.

[13] Gołaszewski J., Idźkowska M., Milewska J., Koczowska I., Kosiorek A. 2000. Analiza komponentów plonowania roślin uprawnych. *Frag. Agron.* XVII, 4(68): 15–27.

[14] Grafius J.E. 1969. Stress: A necessary ingredient of genotype by environment interactions. Internat. Barley Genet. II, Wash. State Univ. Press: 346–355.

[15] Hardwick, R.C., Andrews, D.J. 1980. Genotypic and environmental variation in crop yield. a method of estimating the interdependence of the components of yield. *Euphytica* 26: 177–188.

[16] Hühn M. 1987. Stability analysis of winter-rape (Brassica napus L.) by using plant density and mean yield per plant. *J. Agron. & Crop Sci.* 159: 73–81.

[17] Jag S., Hariprasad A.S., Lakshmi K., Mani V.P., Chauhan V.S., Shoran J., Kant L. 2000. Association and contribution of yield attributes to seed yield in wheat under varying environments in north western hills. *Ann. Agric. Res.*, 21: 274–278.

[18] Jolliffe P.A., Eaton G.W., Lovett Doust J. 1982. Sequential analysis of plant growth. *New Phytol.* 92: 287–296.

[19] Kang K.M., Seneta E. 1980. Path analysis: An exposition. *Dev. Stat.*, red. Krishnaiah P.R. 3: 217–245.

[20] Kozak M. 2001. Use of two-dimensional partitioning of yield variation for winter triticale. 3rd Intern. Conf. PhD Stud., Univ. Miskolc: 183–189.

[21] Kozak M. 2002. Statystyczna analiza uwarunkowania zmienności plonu przez jego składowe. Praca doktorska, SGGW Warszawa.

[22] Kozak M. 2004. Przydatność modelu liniowego w analizie składowych plonu. *Coll. Biom.* 34: 117–126.

[23] Kozak M., Mądry W. 2004. Statystyczna analiza multiplikatywnych składowych plonu roślin – podstawy modelowania. *Post. Nauk Rol.* 5: 13–25.

[24] Kumar M.V.N., Kumar S.S. 2000. Studies on character association and path coefficients for grain yield and oil content in maize (*Zea mays* L.). *Ann. Agric. Res.* 21: 73–78.

[25] Li C.C. 1975. Path analysis – a primer. Boxwood Press, Pacific Grove, California.

[26] Lynch D.R., Tai G.C.C. 1989. Yield and yield component response of eight potato genotypes to water stress. *Crop Sci.* 29: 1207–1211.

[27] Mason D.T., Rath N. 1980. The relative importance of some yield components in East of Scotland strawberry plantations. *Ann. Appl. Biol.* 95: 399–405.

[28] Mądry W., Kozak M. 2000. Analiza ścieżek i sekwencyjna analiza plonu w badaniach zależności plonu od cech łanu. Cz. I. Opis metod. *Rocz. Nauk Rol.*, Seria A, 115: 143–157.

[29] Mądry W., Pietrzykowski R., Rozbicki J. 1995. Analiza współczynników ścieżek dla cech rozwijających się w trakcie ontogenezy oraz plonu ziarna pszenżyta ozimego. *Rocz. Nauk Rol.* Seria A, 111: 9–21.

[30] McArthur D.A.J, Eaton G.W. 1988. Strawberry yield response to fertilizer, paclobutrazol and chlormequat. *Scientia Horticult.* 34: 33–45.

[31] Piepho H.P. 1995. A Simple procedure for yield component analysis. *Euphytica* 84: 43–48.

[32] Rao C.R. 1952. Advanced statistical methods in biometric research. John Wiley & Sons, New York.

[33] Rasmusson D.C., Cannell R.Q. 1970. Selection for grain yield and components of yield in barley. *Crop Sci.* 10: 51–54.

[34] Rozbicki J. 1997. Agrotechniczne uwarunkowania wzrostu, rozwoju i plonowania pszenżyta ozimego. Rozpr. hab., Fundacja Rozwój SGGW Warszawa.

[35] Rozbicki J., Mądry W. 1998. Uwarunkowanie plonu ziarna pszenżyta ozimego przez jego składowe i wybrane cechy botaniczno-rolnicze łanu w zmiennych warunkach uprawowych i pogodowych. *Biul. Inst. Hod. Aklim. Roślin* 205/206: 195–204.

[36] Rozbicki J., Mądry W. 1999. Zależność plonu ziarna pszenżyta ozimego od jego składowych na plantacjach produkcyjnych. *Rocz. Nauk Rol.*, Seria A, 114: 109–115.

[37] Samonte S.O., Wilson L.T., McClung A.M. 1998. Path analysis of yield-related traits of fifteen diverse rice genotypes. *Crop Sci.* 38: 1130–1136.

[38] Scott W.R., Dougherty C.T., Langer R.H.M. 1977. Development and yield components of hihg-yielding wheat crops. *New Zeal. J. Agric. Res.* 20: 205–212.

[39] Spaner D., Mather D.E., Brathwaite R.A.I. 1996. Yield and quality attributes of rainfed green corn in Trinidad. *HortTechnology* 6: 131–134.

[40] Sparnaaij L.D., Bos I. 1993. Component analysis of complex characters in plant breeding. I. Proposed method for quantifying the relative contribution of individual components to variation of the complex character. *Euphytica* 70: 225–235.

[41] Tai G.C.C. 1975. Analysis of genotype-environment interactions based on the method of path coefficient analysis. *Can. J. Genet. and Cytology* 17: 141–149.

[42] Thomas R.L., Grafius J.E., Hahn S. K. 1971. Transformation of sequential quantitative characters. *Heredity* 26: 189–193.

[43] Thurling N. 1974. Morphophysiological determinants of yield in rapeseed (*Brassica campestris* and *Brassica napus*). II Yield components. *Aust. J. Agric. Res.* 25: 711–721.

[44] Toker C., Cagirgan M.I. 2004. The use of phenotypic correlations and factor analysis in determining characters for grain yield selection in chickpea (*Cicer arietinum* L.). *Hereditas* 140: 226–228.

[45] Worley S.:, Culp T.W., Harrell D.C. 1974. The relative contributions of yield components to lint yield of upland cotton, *Gossypium hirsutum* L. *Euphytica* 23: 399–403.

[46] Wright S. 1921. Correlation and causation. *J. Agric. Res.* 20: 557–585.

[47] Wright S. 1934. The method of path coefficients. *Ann. Math. Stat.* 5: 161–215.

[48] Xing C.Z., Jing S.R., Guo L.P., Yuan Y.L., Wang H.L. 2000. Study on heterosis and combining ability of transgenic Bt (*Bacillus thuringiensis*) cotton. *Acta Gossypii Sinica* 12: 6–11.

[49] Yas A.M., Eaton G.W. 1982. Effect of cotton grass on the yield components of cranberry. *Scientia Horticult.* 18: 125–129.

[50] Zając T., Szafrański W., Oleksy A., Witkowicz R. 1999. Indywidualny wkład komponentów struktury plonu ziarna z jednostki powierzchni i z kłosa pszenicy jarej uprawianej po różnych przedplonach. *Frag. Agron.* XV, 4(60): 76–86.

Statistical methods of yield component analysis

Keywords: yield components, sequential development, non-sequential development, statistical methods

Summary

Paper reviewed the methods of yield component analysis. Mostly the methods taking into account a form of relationship between yield and its components were considered; the elaboration presented interpretation possibilities of the methods and their statistical appropriateness (without considering complex mathematical formulas). Some of presented methods are just of historical significance, while the others being quite new and not well known. Moreover, sequential and non-sequential cases of the components development were treated independently. Besides the discussion on the methods, some recommendations regarding the choice of appropriate method to yield component analysis were given. Finally it was concluded that there are no methods both convenient in interpretation and appropriate from the statistical point of view.