

DARIUSZ GOZDOWSKI**WIESŁAW MĄDRY**

Katedra Doświadczalnictwa i Bioinformatyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Charakterystyka i empiryczne porównanie prostej oraz złożonej analizy ścieżek w ocenie determinacji plonu roślin przez ich cechy plonotwórcze

Część I. Prezentacja stosowanych metod

Characteristics and empirical comparison of simple and complex path analysis in grain yield determination by yield - related traits

Part I. Presentation of methods

Ocena determinacji plonu roślin uprawnych przez cechy plonotwórcze jest ważna przy określaniu, które cechy wpływają na zmienność plonu i jakie kryteria powinniśmy przyjąć w procesie hodowli wysokoplonujących genotypów. W pracy przedstawiono podstawy klasycznej analizy ścieżek (zapropionowanej przez S. Wrighta) oraz możliwości i przykłady jej zastosowania do oceny uwarunkowania zmienności plonu roślin uprawnych przez cechy plonotwórcze, które mogą bezpośrednio lub pośrednio wpływać na plon. Przedstawione zostały podejścia z wykorzystaniem ścieżek prostych oraz złożonych (wielopoziomowych), w których plon jest determinowany poprzez różne cechy plonotwórcze. Poza klasyczną analizą ścieżek omówiono w skrócie inne metody powstałe z modyfikacji tej analizy.

Słowa kluczowe: analiza ścieżek, cechy plonotwórcze, składowe plonu

Evaluation of determination of crop yield by yield-related traits is important in determining both the traits having influence on the yield variability and suitable criteria for breeding of high-yielding genotypes. This paper shows the basis of classical path analysis (established by S. Wright) as well as possibilities and examples of its using to evaluate determination of variability of yield by yield-related traits, that may have direct or indirect effects on crop yielding. The approaches including simple and complex (multilevel) path analysis, in which the yield is determined by different yield-related traits, are presented. Besides the classic path analysis, another methods being the modifications of this analysis, are shortly described.

Key words: path analysis, yield-related traits, yield components

WSTĘP

Analiza uwarunkowania zmienności plonu roślin uprawnych (plonu z rośliny lub plonu z jednostki powierzchni) przez różne cechy plonotwórcze roślin jest coraz częstszym przedmiotem badań agronomicznych (Gołaszewski, 1996; Rozbicki, 1997; Annicchiarico i Pecetti, 1998; Garcia del Moral i in., 2003, 2005; Mohammadi i in., 2003; Kozak i Mądry 2006; Kozak i in., 2007). Wśród cech plonotwórczych roślin wyróżnia się składowe plonu (ang. yield components) oraz inne morfologiczne, fenologiczne i fizjologiczne cechy plonotwórcze (ang. yield-related traits lub yield-contributing traits). Badanie uwarunkowania zmienności plonu roślin uprawnych przez cechy plonotwórcze polega na ocenie zależności między plonem, a cechami plonotwórczymi w obrębie populacji obiektów danego gatunku, które są zróżnicowane genetycznie albo w populacji obiektów jednego genotypu-odmiany, które są zróżnicowane pod względem warunków środowiskowych. Wykrycie tych plonotwórczych cech roślin, które w największym stopniu wpływają na zmienność plonu jest ważne dla określenia kryteriów selekcji wartościowych genotypów w hodowli roślin oraz dla ustalenia optymalnej technologii produkcji gatunków i odmian roślin uprawnych (Balkema-Boomstra, 1992; Rozbicki, 1997; Annicchiarico i Pecetti, 1998; Ali i in., 2003; Budzyński i in., 2003; Garcia del Moral i in., 2003, 2005; Shahinnia i in., 2005; Kozak i in., 2007).

Jedną z powszechniej stosowanych metod statystycznych do oceny zależności między plonem, a cechami plonotwórczymi jest analiza ścieżek (ang. path analysis), metoda oparta na wielokrotnej liniowej analizie regresji dla zmiennych standaryzowanych. Podstawy teoretyczne analizy ścieżek opracował Sewall Wright na początku lat 20 ubiegłego wieku, który również przedstawił możliwość jej zastosowania w badaniach rolniczych, głównie w hodowli zwierząt (Wright, 1921, 1923, 1934). Obecnie ta metoda statystyczna jest powszechnie stosowana nie tylko w różnych dziedzinach nauk rolniczych i biologicznych, ale również często znajduje zastosowanie w naukach społecznych i technicznych. Analiza ścieżek w porównaniu do zwykłej wielokrotnej liniowej analizy regresji umożliwia nie tylko porównanie stopnia bezpośredniego wpływu poszczególnych zmiennych przyczynowych na zmienną skutkową (zależną), ale również ocenę pośredniego wpływu zmiennych przyczynowych, wynikającego z ich wzajemnych współzależności. Bezpośredni wpływ danej zmiennej przyczynowej na zmienną skutkową zdefiniowany jest w analizie ścieżek jako przeciętny wpływ tej zmiennej przyczynowej, uśredniony poprzez zmienność pozostałych zmiennych przyczynowych. Zastosowanie analizy ścieżek pozwala na wielostronną, wyczerpującą ocenę związków przyczynowo-skutkowych w danym układzie relacji między rozpatrywanymi zmiennymi przyczynowymi, a zmienną skutkową.

W niniejszej pracy przedstawiono podstawy klasycznej analizy ścieżek oraz możliwości i przykłady jej zastosowania do oceny uwarunkowania zmienności plonu roślin uprawnych (przede wszystkim zbóż) przez cechy plonotwórcze, w tym także składowe plonu, które mogą bezpośrednio lub pośrednio wpływać na plon. Szczególną uwagę zwrócono na analizę ścieżek wielopoziomowych, w których zmienna zależna (plon z rośliny lub plon z jednostki powierzchni) jest determinowana bezpośrednio oraz pośrednio poprzez różne

cechy plonotwórcze. Poza klasyczną analizą ścieżek omówiono w skrócie inne metody powstałe z modyfikacji tej analizy.

PODSTAWY TEORETYCZNE ANALIZY ŚCIEŻEK

W analizie ścieżek badamy związek liniowy między zmiennymi przyczynowymi (np. składowymi plonu) a zmienną zależną (np. plonem), który można przedstawić wg następującego równania regresji wielokrotnej:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i \quad (1)$$

Po standaryzacji zmiennych według poniższych wzorów:

$$\dot{Y} = \frac{Y - \bar{Y}}{S_y} \quad (2)$$

$$\dot{X}_i = \frac{X_i - \bar{X}}{S_i} \quad (3)$$

równanie regresji wielokrotnej dla zmiennych standaryzowanych przyjmuje postać:

$$\dot{Y} = \dot{\beta}_1 \dot{X}_1 + \dot{\beta}_2 \dot{X}_2 + \dots + \dot{\beta}_n \dot{X}_n + \varepsilon_i \quad (4)$$

gdzie cząstkowe współczynniki regresji dla zmiennych standaryzowanych są nazywane współczynnikami ścieżek (ang. path coefficients) i wyrażają wpływ bezpośredni każdej z zmiennych przyczynowych na zmienną zależną (Wright, 1923). Zależność między współczynnikiem ścieżki (p_i) a odpowiadającym mu cząstkowym współczynnikiem regresji wielokrotnej jest następująca:

$$p_i = \dot{\beta}_i = \beta_i \left(\frac{S_{X_i}}{S_Y} \right) \quad (5)$$

Stosowanie analizy ścieżek wymaga spełnienia tych samych założeń, jakie są konieczne dla zastosowania regresji wielokrotnej liniowej (Oktaba, 1986).

Poza określeniem wpływu bezpośredniego poszczególnych zmiennych niezależnych na zmienną zależną, możemy określić wpływ pośredni każdej zmiennej przyczynowej na zmienną zależną poprzez pozostałe zmienne niezależne. W tym celu należy zsumować iloczyny współczynników korelacji i wartości współczynników ścieżek z pozostałymi zmiennymi niezależnymi:

$$\sum_{j=1, j \neq i}^k p_j r_{ij} \quad (6)$$

Suma efektów: bezpośredniego (wyrażonego przez współczynnik ścieżki) i efektów pośrednich wyrażonych wzorem (6) dla każdej z zmiennych przyczynowych jest równa wartości współczynnika korelacji prostej Pearsona między daną zmienną niezależną a zmienną zależną, co można wyrazić następującym równaniem

$$r_{yi} = r_{yx_i} = p_i + \sum_{j=1, j \neq i}^k p_j r_{ij} \quad (7)$$

Stopień determinacji zmiennej zależnej przez zmienne niezależne uwzględnione w modelu regresji wielokrotnej liniowej wyraża wartość współczynnika determinacji (R^2). W modelu liniowym przyjętym do analizy ścieżek współczynnik determinacji jest równy iloczynowi współczynników ścieżek i wartości odpowiednich współczynników korelacji prostej:

$$R^2 = \sum_{i=1}^k p_i r_{yi} \quad (8)$$

Przy wielu zmiennych niezależnych w modelu stosuje się również skorygowany współczynnik determinacji, uwzględniający liczbę zmiennych niezależnych, który określamy wg poniższej zależności:

$$R^2 = R^2 - (1 - R^2) \frac{(k-1)}{(n-k)} \quad (9)$$

gdzie k - liczba zmiennych niezależnych, n — liczba obserwacji

OCENA DETERMINACJI PLONU PRZEZ JEGO SKŁADOWE PRZY WYKORZYSTANIU ANALIZY ŚCIEŻEK

Plon roślin uprawnych możemy zazwyczaj przedstawić jako iloczyn zmiennych nazywanych składowymi plonu. W zależności od metody oceny tych składowych plonu (ocena bezpośrednia każdej z nich lub też pośrednia ocena jednej ze składowych), wartość ich iloczynu jest równa dokładnie wartości plonu lub też jest przybliżoną oceną plonu (Garcia del Moral i in., 1991; Dofing i Knight, 1992; Szempliński, 2003). Mimo, że zależność między składowymi plonu a plonem jest zależnością multiplikatywną, to model analizy regresji przyjęty do analizy ścieżek jest zawsze modelem liniowym. W przypadku, gdy każda ze składowych jest określana metodą bezpośrednią (iloczyn składowych jest jedynie przybliżoną oceną plonu), to możliwe jest w modelu liniowym transformowanie wartości zmiennych niezależnych (składowych plonu) poprzez ich logarytmowanie dla zwiększenia determinacji zmiennej zależnej przez zmienne przyczynowe.

Liczba składowych plonu uwzględnianych w modelu jest różna i waha się zazwyczaj od 2 do 5. W przypadku plonu zbóż najczęściej za składowe plonu przyjmuje się: liczbę kłosów na jednostce powierzchni, średnią liczbę ziarniaków w kłosie i średnią masę pojedynczego ziarniaka (Dofing i Knight, 1992; Rozbicki, 1997; Zając i in., 1997; Szafranski i Kulig, 1998; Garcia del Moral i in., 2005). Możliwe jest przyjęcie mniejszej liczby składowych (np. liczby kłosów na jednostce powierzchni i średniej masy ziarna z kłosa) lub też większej liczby składowych (np. liczba roślin na jednostce powierzchni, średnia liczba kłosów na 1 roślinie, średnia liczba kłosków w kłosie, średnia liczba ziarniaków w kłosku i średnia masa pojedynczego ziarniaka) (Sinebo, 2002; Garcia del Moral i in., 2003, 2005; Moragues i in., 2006). W przypadku innych gatunków roślin uprawnych składowe plonu są określane w podobny sposób, np. dla roślin strączkowych (np. soja) składowe mogą być następujące: liczba roślin na jednostce powierzchni, średnia liczba strąków na roślinie, średnia liczba nasion w strąku i średnia masa pojedynczego nasienia (Board i in., 1999; Ball i in., 2001).

Dane, które wykorzystywane są do oceny wpływu składowych na plon pochodzą najczęściej z ścisłych doświadczeń czynnikowych, a wszystkie dane (z wszystkich poziomów czynników) są zazwyczaj traktowane jako dane pochodzące z jednej populacji, czasami wykonuje się odrębne analizy dla poszczególnych poziomów jednego z czynników doświadczenia (Garcia del Moral i in., 2005). Pojedyncze obserwacje są średnimi z poletek, przy czym często ocena tych zmiennych jest wykonywana jedynie na niewielkiej próbie roślin reprezentującej dane poletko. Metodyka przyjęta do oceny składowych przez różnych autorów jest często inna, zazwyczaj plon ziarna określany jest na większej powierzchni, np. z całego poletka lub powierzchni 1 m², a pozostałe składowe są oceniane z mniejszej powierzchni wyznaczonej najczęściej w środkowej części poletka. Poza tym zazwyczaj tylko dwie składowe są określane bezpośrednio, a trzecia składowa plonu jest wyznaczana pośrednio na podstawie masy plonu i dwóch jego składowych. Moragues i wsp. (2006) w badaniach nad pszenicą przyjęli oznaczanie liczby kłosów i liczby ziaren w kłosie z próby roślin pobranych z 1 mb z środkowej części poletka, natomiast MTZ została określona pośrednio na podstawie masy plonu ziarna (przy wilgotności 12%) z całego poletka. W badaniach Garcia del Moral i in. (2003, 2005) dotyczących składowych plonu pszenicy wszystkie składowe plonu były określane bezpośrednio. Liczbę kłosów na jednostce powierzchni określano również na podstawie próby roślin pobranej z 1 mb, liczbę ziaren w kłosie na podstawie podpróby składającej się z 10 wybranych roślin, MTZ na podstawie trzech prób ziarna liczących 300 ziarniaków. Plon ziarna był oceniany z całego poletka i przeliczany na 12% wilgotności.

Bardzo podobną metodykę stosowano również w badaniach Garcia del Moral i in. (1991) nad jęczmieniem jarym, jedyną różnicą była metoda oceny MTZ, która oceniano na podstawie pojedynczej próby 250 ziarniaków. W badaniach Sinebo (2002) nad jęczmieniem jarym liczba ziaren w kłosie i średnia masa ziarniaka były określane na podstawie próby składającej się z 10 kłosów, natomiast liczba kłosów była określana na podstawie próby z 1 m² a plon ziarna z całego poletka.

Samonte i in. (1998, 2005) w badaniach nad ryżem, liczbę wiech na jednostce powierzchni i liczbę ziarniaków w wieszce określano na podstawie próby roślin pobranych z czterech rzędów ze środka poletka, natomiast masa 100 ziarniaków była określana na podstawie trzech powtórzeń dla każdego poletka.

Metodyka oceny składowych roślin niezbożowych, jest nieco inna, ze względu na odmienną budowę morfologiczną poszczególnych gatunków roślin uprawnych. W doświadczeniach nad soją Ball i in. (2001) liczbę strąków określano na próbie składającej się z 6 lub 12 roślin z każdego poletka, masę pojedynczego nasiona określano na próbie 100 nasion natomiast plon nasion z jednostki powierzchni na podstawie próby roślin z 12,2 mb (z rzędów wewnętrznych), co stanowiło około 25% powierzchni całego poletka. Plon nasion był przeliczany dla 13 % wilgotności nasion. Liczba roślin na jednostce powierzchni była oceniana na 4 losowo wybranych rzędach o długości 1,5 m na każdym poletku.

Nieco inną metodykę oceny składowych i plonu nasion soi zastosowali Board i in., plon nasion był określany z dużej części poletka (od 4,3 do 8,6 m²) z pominięciem rzędów zewnętrznych, a liczba strąków na roślinie, liczba nasion w strąku i masa pojedynczego nasiona na podstawie próby losowo wybranych 10 roślin.

W badaniach dotyczących składowych plonu koniczyny rocznej Iannucci i Martiniello (1998) plon nasion oceniali na podstawie próby z 1 m² przeliczając go następnie na wilgotność 15%, a składowe plonu (liczba kwiatostanów na pędzie, liczba nasion w kwiatostanie i masa 1000 nasion) były oceniane na próbie losowo wybranych 20 pędów.

Metody oceny wielkości plonu i jego składowych w doświadczeniach nad różnymi gatunkami roślin uprawnych znacznie różnią się między sobą, przy czym można stwierdzić, że plon jest określany zazwyczaj na powierzchni dość dużej, obejmującej często całe poletko. Powierzchnia pojedynczego poletka waha się najczęściej od kilku do kilkunastu m², możemy, zatem stwierdzić, że metoda oceny wielkości plonu jest w pełni reprezentatywna. Natomiast składowe plonu, są zazwyczaj oceniane na niewielkiej próbie, której reprezentatywność może budzić wątpliwości, np. ocena składowych plonu na kilku losowo wybranych roślinach jest zapewne obciążona dużym błędem doświadczalnym, ze względu na bardzo niewielką część roślin z poletka, jaką obejmuje. Poza tym wielu autorów nie precyzuje, w jaki sposób zostały pobrane podpróby roślin do oceny składowych, informacja, że „wybrano rośliny losowo” nie określa precyzyjnie stosowanej metody wyboru, co z kolei utrudnia przeprowadzenie porównywalnego doświadczenia w innych warunkach lub z innymi genotypami. Stosowana metodyka oceny poszczególnych składowych nie ma przyjętych ogólnie uznawanych standardów, co jest nie tylko problemem przy porównywaniu wyników z różnych doświadczeń, ale również porównywanie wpływu poszczególnych składowych na plon na podstawie analiz wyników uzyskanych z niektórych doświadczeń może nie być w pełni wiarygodne (Kozak i Mądry, 2006).

Po określeniu, które zmienne uznamy za składowe plonu możemy przyjąć *a priori* jakie zależności przyczynowo-skutkowe występują między badanymi zmiennymi. Jeśli uwzględnimy 3 składowe plonu zbóż to zależności między nimi można przedstawić wg poniższego schematu przedstawionego na rysunku 1.

Zazwyczaj przyjmuje się, że na schematach przyczynowo-skutkowych stosowanych dla analizy ścieżek, strzałki z jednym grotem oznaczają zależności jednokierunkowe (wyrażane współczynnikami ścieżek), a strzałki z dwoma grotami zależności dwukierunkowe (wyrażane współczynnikami korelacji).

Dla przyjętego schematu zależności możemy zatem określić wartości współczynników ścieżek przyjmując następujący model liniowy regresji wielokrotnej:

$$\dot{Y} = p_{1y} \dot{X}_1 + p_{2y} \dot{X}_2 + p_{3y} \dot{X}_3 + \varepsilon_i \quad (10)$$

Cząstkowe współczynniki regresji wielokrotnej w tym modelu nazywamy współczynnikami ścieżek (p_i) są one oceną bezpośredniego wpływu każdej ze składowych na plon ziarna, natomiast efekty pośrednie możemy określić wg poniższych równań: dla zmiennej X_1 : $r_{12} \cdot p_{y2} + r_{13} \cdot p_{y3}$

Suma efektów bezpośrednich i pośrednich dla każdej ze składowych jest równa współczynnikowi korelacji, zatem:

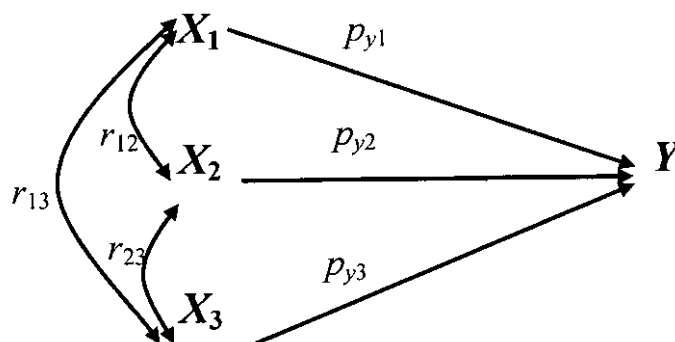
$$r_{1y} = p_{y1} + r_{12} \cdot p_{y2} + r_{13} \cdot p_{y3}$$

$$r_{2y} = p_{y2} + r_{12} \cdot p_{y1} + r_{23} \cdot p_{y3}$$

$$r_{3y} = p_{y3} + r_{13} \cdot p_{y1} + r_{23} \cdot p_{y2}$$

Udział zmienności plonu przez jego składowe wyraża współczynnik determinacji (R^2), który w omawianym modelu jest równy:

$$R^2 = r_{y1}p_{y1} + r_{y2}p_{y2} + r_{y3}p_{y3}$$



X_1 — liczba kłosów na jednostce powierzchni, number of spikes per unit area

X_2 — średnia liczba ziaren w kłosie, average number of grains per spike

X_3 — masa pojedynczego ziarniaka, weight of individual kernel

Y — plon ziarna, grain field

$p_{y,i}$ — współczynniki ścieżek, path coefficients

r_{ij} — współczynniki korelacji prostej między zmiennymi przyczynowymi; correlation coefficients between causal variables

Rys. 1. Schemat zależności przyczynowo-skutkowych między składowymi plonu (X_i) a plonem ziarna zbóż (Y)

Fig. 1. Cause and effect relationship diagram among yield components (X_i) and cereals grain yield (Y)

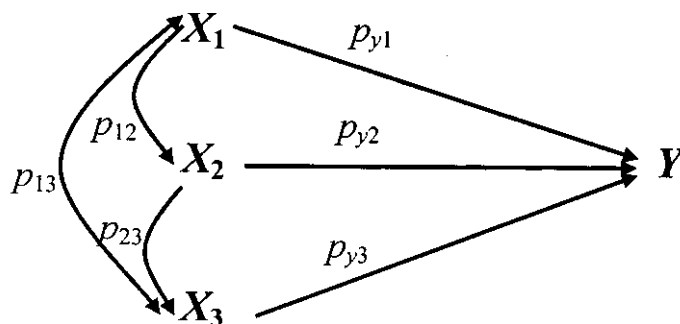
Z uwagi na to, że kształtowanie składowych plonu nie odbywa się jednocześnie, lecz każda ze składowych kształtuje się na innych etapach rozwoju rośliny, to niektórzy badacze proponują modyfikację powyższego schematu zależności przyczynowo-skutkowych. W badaniach Kleppera i wsp. (1998) nad pszenicą ozimą stwierdzono, że liczba kłosów na jednostce powierzchni kształtuje się od siewu do początku kłoszenia, liczba ziaren w kłosie od fazy strzelania w źdźbło do końca kwitnienia, a masa pojedynczego ziarniaka od fazy kłoszenia do końca wypełniania się ziarniaka, czyli fazy dojrzałości woskowej. Alternatywny model zależności przyczynowo-skutkowych dla analizy ścieżek uwzględniający sekwencyjne kształtowanie się składowych plonu zbóż proponowany przez Dofinga i Knighta (1992) oraz Garcia del Moral i wsp. (1991) oraz stosowany m.in. przez Mamana i wsp. (2004) oraz Garcia del Moral i wsp. (2003) można przedstawić według schematu przedstawionego na rysunku 2.

W tym modelu nie określamy wpływu składowych później kształtujących się na składowe tworzące we wcześniejszych etapach ontogenezy. Zależności przedstawione na rysunku 2 można opisać równaniem (10) oraz następującymi dwoma równaniami regresji:

$$X_2 = p_{12}\dot{X}_1 + \varepsilon_i$$

$$X_3 = p_{13}\dot{X}_1 + p_{23}\dot{X}_2 + \varepsilon_i$$

Badając powyższe zależności określamy wartości współczynników ścieżek (których wartości są takie same dla modelu zależności przedstawionego na rysunkach 1 i 2) wyrażające wpływ bezpośredni składowych na zmienność plonu. Ponadto badamy wpływ bezpośredni liczby kłosów na jednostce powierzchni na średnią liczbę ziaren oraz wpływ bezpośredni liczby kłosów i średniej liczby ziaren w kłosie na średnią masę pojedynczego ziarniaka.



X_1 — liczba kłosów na jednostce powierzchni, number of spikes per unit area

X_2 — średnia liczba ziaren w kłosie, average number of grains per spike

X_3 — masa pojedynczego ziarniaka, weight of individual kernel

Y — plon ziarna, grain yield

p_{ij}, p_{yi} — współczynniki ścieżek, path coefficients

Rys. 2. Schemat zależności przyczynowo-skutkowych między składowymi plonu (X_i) a plonem ziarna zbóż (Y) uwzględniający sekwencyjne kształtowanie się składowych plonu

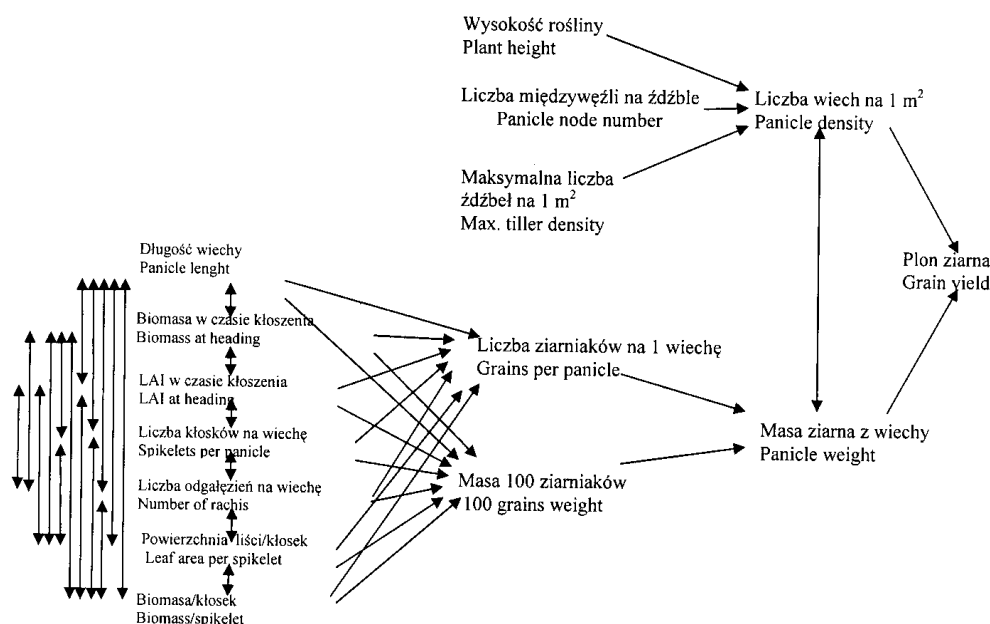
Fig. 2. Cause and effect relationship diagram among yield components (X_i) and cereals grain yield (Y) taking into account sequential yield components forming

OCENA DETERMINACJI PŁONU PRZEZ INNE CECHY PŁONOTWÓRCZE ROŚLIN

W badaniach agronomicznych analizuje się także wpływ innych cech plonotwórczych roślin na plon. Wpływ ten określany z użyciem analizy ścieżek często jest traktowany jako bezpośredni dla wszystkich badanych cech (Ball i in., 2001; Ahmed i in., 2003; Ali i in., 2003; Okuyama i in., 2005; Bhargava i in., 2007). Przyjęcie takiego modelu zależności, w którym wszystkie cechy plonotwórcze są cechami pierwszego rzędu nie uwzględnia ich rozwoju w kolejnych fazach ontogenezy rośliny. Dlatego też coraz częściej przyjmuje się złożone modele analizy ścieżek. W zależności od przyjętego modelu przyczynowo-skutkowego cechy będące zmiennymi przyczynowymi, można podzielić na cechy plonotwórcze pierwszego rzędu (ang. primary level traits), nazywane też składowymi plonu pierwszego rzędu (ang. primary yield components), cechy plonotwórcze drugiego rzędu (ang. secondary level traits) — składowe plonu drugiego rzędu (ang. secondary components) i cechy plonotwórcze trzeciego rzędu (ang. tertiary level traits) — składowe

trzeciego rzędu (ang. tertiary components) (Samonte i in., 1998; Board i in., 1999, 2003; Mohammadi i in., 2003). Pojęcie składowe plonu według niektórych autorów (Kozak i Mądry, 2006) powinno być używane wyłącznie do cech, które bezpośrednio wpływają na wielkość plonu (najczęściej iloczyn ich jest równy wartości plonu), natomiast pozostałe cechy roślin nie będące składowym, ale wpływające na plon nazywane są najczęściej cechami plonotwórczymi lub cechami powiązаныmi z plonem (ang. yield-related traits).

Cechy plonotwórcze pierwszego rzędu są to zazwyczaj cechy, które są właściwymi składowymi plonu tj. bezpośrednio wpływające na jego wielkość, cechy plonotwórcze drugiego rzędu są cechami, które bezpośrednio nie wpływają na plon, natomiast wpływają na zmienność składowych plonu. Cechy plonotwórcze trzeciego rzędu wpływają natomiast bezpośrednio na cechy plonotwórcze drugiego rzędu a wpływ ich na składowe plonu i plon jest pośredni. W zależności od przyjętego modelu przyczynowo-skutkowego możemy różne cechy zaliczyć do różnych grup cech (Sinebo, 2002). W przypadku zbóż do cech plonotwórczych drugiego lub trzeciego rzędu zalicza się najczęściej cechy morfologiczne roślin, jak również cechy charakteryzujące łan roślin jako całość np. liczba wszystkich pędów na jednostce powierzchni, udział pędów, które utrzymały się do końca okresu wegetacji, liczba kłosek w kłosie, liczba ziarniaków w kłosku, długość okresu wegetacji rośliny, długość okresu wypełniania się ziarniaków, LAI (ang. leaf area index) i biomasa w poszczególnych fazach rozwojowych rośliny itp. (Samonte i in., 1998; Garcia del Moral i in., 2005). W przypadku innych gatunków roślin uprawnych cechy te mogą być oczywiście inne (Iannucci i Martinello, 1998; Board i in., 1999) zazwyczaj są związane z wcześniejszymi etapami ontogenezy roślin (np. dla soi: liczba międzywęźli, udział międzywęźli produktywnych, wskaźniki wzrostu CGR (ang. crop growth rate) itp.). Podobnie tak, jak w przypadku określenia wpływu składowych plonu na jego zmienność przy pomocy analizy ścieżek, również przy dużej liczbie cech, które mogą być powiązane z plonem należy wstępnie (*a priori*) określić model zależności przyczynowo-skutkowych, a następnie określić z użyciem analizy ścieżek, które cechy (zmiennie przyczynowe) wpływają istotnie na cechy zależne (Samonte i in., 1998; Mohammadi i in., 2003). Dobór właściwego modelu do analizy ścieżek powinien być oparty na wiedzy dotyczącej kolejności kształtowania się każdej z cech w procesie ontogenezy roślin. Przykładowy model zastosowany przez Samonte i wsp. (1998) w ocenie determinacji plonu ryżu przez składowe i inne cechy plonotwórcze przedstawiony jest na rysunku 3. Wszystkie cechy powiązane z plonem (14 cech) zostały podzielone na 3 grupy. Pierwszą grupę stanowią 2 cechy (cechy plonotwórcze pierwszego rzędu), będące jednocześnie składowymi plonu tj. liczba wiech na 1 m² i masa ziarna z wiechy, cechy te stanowią końcowy komponent diagramu analizy ścieżek.



Rys. 3. Schemat zależności między cechami plonotwórczymi a plonem ziarna ryżu (za Samonte i in., 1998)

Fig. 3. Cause and effect relationship diagram among rice yield and yield related traits (according to Samonte *et al.* 1998)

Kolejną grupę cech (cechy drugiego rzędu) tworzą wysokość rośliny, liczba międzywęzła na źdźble i maksymalna liczba źdźbeł/m² oraz liczba ziarniaków na 1 wiechę i masa 100 ziarniaków, cechy te wpływają bezpośrednio na cechy pierwszego rzędu i tworzą 2 komponenty diagramu analizy ścieżek. Pozostałe cechy związane z plonem są cechami trzeciego rzędu i wpływają bezpośrednio na 2 spośród cech drugiego rzędu, tworząc 2 kolejne komponenty diagramu analizy ścieżek. Cały diagram przyczynowo-skutkowy przedstawiony na rysunku 3, tworzy, zatem 5 komponentów analizy ścieżek, z których każdy możemy analizować odrębnie, stosując analogiczne metody, które dotyczyły diagramu z rysunku 1. Podział dużej liczby cech na cechy kolejnych rzędów umożliwia określenie determinacji dla każdej z zmiennych zależnych w poszczególnych komponentach analizy ścieżek, przez zmienne przyczynowe. Jednocześnie eliminuje wzajemny wpływ cech silnie skorelowanych, które dzięki podziałowi na kolejne rzędy cech są w odrębnych komponentach analizy ścieżek. Takie podejście stosowane może być nie tylko do oceny determinacji plonu z jednostki powierzchni, ale również jako końcową zmienną zależną można potraktować masę ziarna z kłosa lub rośliny (Mohammadi i in., 2003). Diagram zależności przyczynowo-skutkowych przyjęty wstępnie (*a priori*) do analizy ścieżek można modyfikować w trakcie analiz wykorzystując na przykład regresję krokową, co pozwala wyeliminować cechy przyczynowe bardzo silnie skorelowane lub też

nie wykazujące istotnego związku z zmienną zależną. Włączenie do diagramu zależności przyczynowo-skutkowych wielu cech powiązanych z plonem i jednocześnie zastosowanie sekwencyjnego modelu analizy ścieżek pozwala na lepsze zrozumienie kształtowania się kolejnych cech na poszczególnych etapach ontogenezy roślin (Garcia del Moral i in., 1991, 2003). Tworzenie złożonych modeli analizy ścieżek uwzględniających kolejność kształtowania się badanych cech zakłada, że wcześniej rozwijające się cechy mogą wpływać na inne rozwijające się w późniejszych fazach ontogenezy, natomiast brak jest wpływu o odwrotnym kierunku. Takie ontogenetyczne podejście (ang. ontogenetic approach) do analizy ścieżek wydaje się bardziej właściwe, gdyż uwzględnia dobór modelu nie tylko oparty na uzyskaniu dużej determinacji zmiennych zależnych, ale przede wszystkim na powiązaniu cech z uwzględnieniem kryteriów biologicznych (Garcia del Moral i in., 1991, 200; Samonte i in., 1998).

Podobnie tak jak w przypadku oceny determinacji plonu przez jego składowe, również w przypadku modeli złożonych wykorzystujących wiele różnych cech ważna jest metodyka oceny poszczególnych cech. Do oceny wielu cech plonotwórczych wykorzystuje się te same próby roślin, które były użyte do oceny składowych plonu, tak, więc ocenę tych cech dokonuje się na określonej liczbie roślin lub też na określonej powierzchni. Z uwag na to, że wiele cech ocenianych w złożonych modelach analizy ścieżek rozwija się sekwencyjnie, to również pomiary tych cech są rozciągnięte w czasie, począwszy od wschodów roślin do osiągnięcia pełnej dojrzałości. W wspomnianym już doświadczeniu Samonte i wsp. (1998), kilka z badanych cech było określane w czasie wegetacji roślin. Biomasa roślin była określana na każdym poletku w czasie kwitnienia roślin na podstawie pięciu prostokątnych powierzchni (0,72 m × 0,30 m). Z tych samych prób roślin był oceniany w sposób destrukcyjny wskaźnik powierzchni liści łąnu LAI. Rośliny zostały wykopane a następnie wyodrębniono z nich poszczególne organy (w tym liście), a następnie powierzchnia asymilacyjna roślin została określona przy pomocy urządzenia CI-251.

W złożonych modelach analizy ścieżek wykorzystywane są również inne wskaźniki fizjologiczne informujące o dynamice wzrostu roślin. W badaniach Boarda i wsp. (1999) nad soją uwzględniono dwa wskaźniki fizjologiczne oznaczane w okresie wegetacji na podstawie akumulacji biomasy przez rośliny, były to wskaźnik przyrostu plonu (ang. crop growth rate — CGR) oraz pochłanianie światła przez rośliny (ang. light interception — LI).

Duża liczba cech może być również określana w jednym czasie, czego przykładem jest doświadczenie z kukurydzą Mohammadi i wsp. (2003), w którym 14 różnych cech (m.in.: średnica kolby, długość i szerokość ziarniaków, wysokość rośliny, liczba liści na roślinie) było oceniane w czasie zbioru. Cechy te było oceniane na pięciu wybranych z każdego poletka roślinach. Podobnie w doświadczeniu z jęczmieniem jarym, Sinebo (2002) w złożonym modelu analizy ścieżek wykorzystał przede wszystkim cechy, które mierzono na roślinach w czasie zbioru (np. plon słomy, indeks żniwny — HI, wysokość rośliny). Jediną cechą mierzoną w wcześniejszej fazie rozwojowej była wysokość pędów wegetatywnych określana w fazie strzelania w źdźbło na pięciu wybranych roślinach z każdego poletka.

Poza cechami bezpośrednio związanymi z budową morfologiczną roślin w złożonych modelach często wykorzystywane są zmienne określające czas trwania poszczególnych okresów rozwojowych roślin (Garcia del Moral i in., 1991, 2005; Sinebo, 2002; Moragues i in., 2006). W badaniach Sinebo (2002) zmiennymi przedstawionymi jako czas były: czas wypełniania ziarna jęczmienia (ang. grain filling duration — GFD), czyli różnica między czasem od siewu do pełnej dojrzałości a czasem od siewu do kłoszenia oraz indeks faz (ang. phase index), czyli iloczyn czasu wypełniania ziarna przez czasem od siewu do pełnej dojrzałości. W badaniach Garcia del Moral i wsp. (2005) zmienne powiązane z czasem rozwoju roślin to długość okresu wegetacji roślin, czas wypełniania ziarna i wskaźnik wypełniania ziarna (ang. grain filling rate), który jest wartością współczynnika regresji liniowej prostej dla suchej masy ziarna względem czasu.

Wykonywanie pomiarów wartości różnych cech morfologicznych roślin i wskaźników świadczących o szybkości wzrostu i akumulacji biomasy w czasie wegetacji często wiąże się z zwiększonym błędem doświadczalnym. Może być to spowodowane pomiarami destrukcyjnymi, które całkowicie eliminują część roślin z poletka z dalszej oceny i uniemożliwiają odniesienie kolejnych wyników do dokładnie tych samych roślin. Inną przyczyną zwiększania się błędu w kolejnych w czasie pomiarach może być nieco inna powierzchnia, na której są wykonywane poszczególne pomiary roślin, np. z uwagi na niemożliwość uwzględnienia wszystkich roślin na poletku w czasie oceny.

Mimo tych wad modele złożone analizy ścieżek, w których uwzględniana liczba cech roślin jest duża umożliwiają pełniejszą ocenę determinacji plonu z rośliny lub jednostki powierzchni, pozwalają stwierdzić, które z cech najsilniej wpływają na kształtowanie zmienności plonu (Board i in., 1999).

INNE METODY W OCENIE DETERMINACJI PŁONU

Metoda analizy ścieżek opracowana przez Wrighta (1921, 1923, 1934) jest powszechnie stosowana i akceptowana w naukach rolniczych i biologicznych. Powstają jednak modyfikacje tej metody, które są coraz częściej stosowane zamiast klasycznej analizy ścieżek (Kozak i Kang, 2006; Kozak i in., 2007). Według Shipleya (2002), stosowanie klasycznej analizy ścieżek ma wiele niedoskonałości, jedną z nich jest brak możliwości statystycznego testowania struktury modelu przyczynowo-skutkowego, istnieje jedynie możliwość testowania istotności poszczególnych pojedynczych ścieżek. Dlatego też, proponuje on rozwiniętą przez Jöreskoga (1970) i Keeslinga (1972) analizę ścieżek umożliwiającą testowanie modelu oparte na metodzie największej wiarygodności (ang. maximum likelihood) wykorzystującej modelowanie równań strukturalnych (ang. structural equation modeling — SEM). Analiza ścieżek oparta na metodzie największej wiarygodności powszechnie stosowana jest w naukach socjologicznych, psychologii oraz coraz częściej w naukach biologicznych i rolniczych, w tym w określaniu determinacji plonu przez jego składowe (Mitchell, 1992; Agrama i in., 1999). Podstawową zaletą tych metod jest wybór najlepszego modelu przyczynowo-skutkowego, co jest zwłaszcza ważne przy złożonych modelach analizy ścieżek.

Inną metodą stosowaną zamiast analizy ścieżek w ocenie determinacji plonu przez jego składowe jest analiza oparta na dwuwymiarowym podziale zmienności plonu (ang. two-dimensional partitioning of yield variation — TDP), która uwzględnia jednocześnie wpływ czynników doświadczalnych i składowych plonu na jego zmienność. Metoda ta oparta jest na analizie regresji i analizie wariancji, stosowana w ocenie determinacji plonu konieczne jest założenie sekwencyjnego kształtowanie się poszczególnych składowych plonu (Eaton i in., 1986; Gołaszewski, 1996). Zaletą tej metody jest uwzględnienie wpływu czynników doświadczenia, których wpływ w klasycznej analizie ścieżek jest pomijany, bądź wykonywane są odrębne analizy dla wybranych poziomów czynników doświadczenia. Stosowanie metody TDP w ocenie determinacji plonu przez jego składowe wymaga transformacji zmiennych (składowych i plonu) poprzez ortogonalizację Gram-Schmidta, które następnie są wykorzystywane do przeprowadzenia analizy regresji wielokrotnej, w której zmienną zależną jest plon i analiz wariancji dla każdej zmiennej. Wyniki z analizy regresji i analiz wariancji służą następnie do określenia wpływu poszczególnych składowych i jednocześnie czynników doświadczenia na zmienność plonu. Analiza ta jest więc narzędziem odpowiadającym jednocześnie na wiele pytań, co pozwoliło na dość szybkie upowszechnienie tej metody w ocenie determinacji plonu przez jego składowe (Tarimo, 1997; Spaner i in., 2001; Gołaszewski i in., 1998, 2001; Soenarto, 2001).

Kozak (2006) twierdzi, że metoda TDP, ze względu na uwzględnienie sum kwadratów z analizy wariancji nie jest jednak w pełni wiarygodną metodą oceny determinacji plonu przez poszczególne składowe, gdyż wartości sum kwadratów nie mogą być porównywalne między różnymi źródłami zmienności i różnymi analizami wariancji.

PODSUMOWANIE

Analiza ścieżek w ocenie determinacji plonu roślin uprawnych jest metodą pozwalającą stwierdzić, które z cech roślin najsilniej wpływają na zmienność plonu. Stwierdzenie, które są to cechy pozwala przyjąć bardziej precyzyjne kryteria w hodowli nowych genotypów (Balkema-Boomstra, 1992). Rozwój genetyki sprawił, że często możliwe jest już określenie, które geny odpowiadają w największym stopniu za kształtowanie się tych cech, co z kolei można wykorzystać w hodowli roślin wykorzystując metody genetyki molekularnej (Slafer, 2003). W zależności, jakie cechy uwzględnimy w ocenie determinacji plonu możemy uzyskać mniej lub bardziej precyzyjną odpowiedź, na pytanie: na co zwrócić uwagę przy hodowli nowych wysoko produktywnych genotypów w zróżnicowanych warunkach środowiskowych. Za bardzo ważny element badań nad wpływem składowych plonu i innych cech roślin na zmienność plonu należy uznać przyjęcie właściwej metodyki prowadzenia doświadczeń, zwłaszcza gdy uwzględniamy dużą liczbę cech, których ocena może być obciążona dużym błędem doświadczalnym (np. destrukcyjne pomiary w czasie okresu wegetacji roślin). Przyjęcie właściwego modelu, zwłaszcza w modelach złożonych oraz prawidłowa interpretacja wyników badań nad determinacją plonu musi opierać się na znajomości biologicznych podstaw plonowania badanego gatunku, w szczególności ważna jest sekwencja kształtowania się składowych i innych badanych cech roślin i wzajemne zależności między nimi (Acreche i Slafer, 2006).

LITERATURA

- Acreche M., Slafer G. 2006. Grain weight response to increases in number of grains in wheat in a Mediterranean area, *Field Crops Research* 98: 52–59
- Agrama H., Zakaria A., Said F., Tuinstra M. 1999. Identification of quantitative trait loci for nitrogen use efficiency in maize. *Molecular Breeding*, 5 (2): 187 — 195.
- Ahmed H., Khan, B., Khan S., Kissana N., Laghari S. 2003 . Path coefficient analysis in bread wheat, *Asian J. Plant Sci.*, 2 (6): 491 — 494.
- Ali N., Javidfar F., Elmira Y., Mirza Y. 2003. Relationship among yield components and selection criteria for yield improvement in winter rapeseed (*Brassica napus* L.). *Pak. J. Bot.*, 35 (2): 167 — 174.
- Annicchiarico P., Pecetti L. 1998. Yield vs. morphophysiological trait-based criteria for selection of durum wheat in a semi-arid Mediterranean region (northern Syria). *Field Crop Res.* 59, (3): 163 — 173.
- Balkema-Boomstra A. 1992 . The relation between grain yield and some related traits of spring barley (*Hordeum vulgare* L.) and their usefulness in a breeding program. *Euphytica* 65 (2): 99 — 106.
- Ball R., McNew R., Vories E., Keisling T., Purcell L. 2001. Path analyses of population density effects on short-season soybean yield. *Agron. J.* 93: 187 — 195.
- Bhargava A., Shukla S., Ohri D. 2007. Genetic variability and interrelationship among various morphological and quality traits in quinoa (*Chenopodium quinoa* Willd.). *Field Crops Research* 101: 104–116.
- Board J., Kang M. S., Harville B. 1999. Path analyses of the yield formation process for late-planted soybean. *Agron. J.*, 91: 128 — 135.
- Board, J., Kang, M. S., Bodrero, M. L. 2003. Yield components as indirect selection criteria for late-planted soybean cultivars. *Agron. J.* 95: 420 — 429.
- Budzyński W., Jankowski K., Szempliński W. 2003. Cultivar-related and agronomic conditions of rye yielding on good rye soil suitability complex part I. Yield and its relationship with the yield components. Vol. 6, issue 1, *Electronic Journal of Polish Agricultural Universities, ser. Agronomy*.
- Dofing M., Knight W. 1992. Alternative model for path analysis of small-grain yield. *Crop Sci.* 32: 487 — 489.
- Eaton G. W., Bowen P. A., Jolliffe P. A. 1986. Two-dimensional Partitioning of Yield Variation. *HortScience* 21: 1052 — 1053.
- Garcia del Moral L., Ramos M., Garcia del Moral B., Jimenez-Tejada P. 199 . Ontogenic approach to grain production in spring barley based on path-coefficient analysis. *Crop Sci.* 31: 1179 — 1185.
- Garcia del Moral L., Rharrabti Y., Elhani S., Martos V., Royo C. 2005. Yield formation in Mediterranean durum wheat's under two contrasting water regimes based on path-coefficient analysis. *Euphytica* 146: 203 — 212.
- Garcia del Moral L., Rharrabti Y., Villegas D., Royo C. 2003. Evaluation of grain yield components in durum wheat under Mediterranean conditions: an ontogenic approach. *Agron. J.* 95: 266 — 274.
- Gołaszewski J. 1996. A Method of Yield Component Analysis. *Biometrical Letters* 33 (2): 79 — 88.
- Gołaszewski J., Idżkowska M., Milewska, J. 1998. The TDP Method of Seed Yield Component Analysis in Grain Legume Breeding. *J. Appl. Genet.* 39 (4): 299 — 308.
- Gołaszewski J., Idżkowska M., Milewska J., Koczowska I. 2001. Yield component analysis with SYCA and TDP in Fodder Pea. *Plant Breed. Seed Sci.* 45 (2): 77 — 85.
- Iannucci A., Martiniello P. 1998. Analysis of seed yield components in four Mediterranean annual clovers. *Field Crop Res.*, 55, 235 — 243.
- Jöreskog K. 1970. A general method for analysis of covariance structures. *Biometrics* 57: 239 — 251.
- Keesling J. 1972. Maximum likelihood approaches to causal analysis. Chicago, Department of Education, University of Chicago.
- Klepper, B., Rickman, R., Waldman, S., Chevalier, P. 1998. The physiological life cycle of wheat: Its use in breeding and crop management. *Euphytica* 100: 341 — 347.

- Kozak M. 2006. Two-dimensional partitioning of yield variation: a critical note. *Plant Breed. Seed Sci.* 53: 37 — 42.
- Kozak M., Singh P. K., Verma, M. R., Hore, D. K. 2007. Causal mechanism for determination of grain yield and milling quality of lowland rice. *Field Crops Research* 102: 178 — 184.
- Kozak M., Kang M.S. 2006. Note on modern path analysis in application to crop science. *Commun. Biometry Crop Sci.* 1 (1): 32 — 34.
- Kozak, M., Mądry, W. 2006. Note on yield component analysis. *Cereal Research Communications* 34 (2–3): 933 — 940.
- Maddox D., Antonovics J. 1983. Experimental ecological genetics in *Plantago*: a structural equation approach to fitness components in *P. aristata* and *P. patagonica*. *Ecology*, 64 (5): 1092 — 1099.
- Maman N., Mason S., Lyon D., Dhunganga P. 2004. Yield components of pearl millet and grain sorghum across environments in the central great plains. *Crop Sci.* 44: 2138 — 2145.
- Mohammadi S., Prasanna B., Singh N. 2003. Sequential path model for determining interrelationships among grain yield and related characters in maize. *Crop Sci.* 43: 1690 — 1697.
- Mitchell R. J. 1992. Testing evolutionary and ecological hypotheses using path analysis and structural equation modeling, *Functional Ecology*, 6 (2): 123 — 129.
- Moragues M., Garcia del Moral L., Moralejo M., Royo C. 2006. Yield formation strategies of durum wheat landraces with distinct pattern of dispersal within the Mediterranean basin. I. Yield components. *Field Crops Res.* 95: 194 — 205.
- Oktaba W. 1986. *Metody statystyki matematycznej w doświadczalnictwie*. PWN Warszawa.
- Okuyama L., Federizzi L., Barbosa Neto J. 2005. Plant traits to complement selection based on yield components in wheat. *Ciência Rural*, 35, 5: 1010 — 1018.
- Rozbicki J. 1997: *Agrotechniczne uwarunkowania wzrostu, rozwoju i plonowania pszenżyta ozimego*. Fundacja "Rozwój SGGW".
- Samonte S., Wilson L., McClung A. 1998. Path analyses of yield and yield-related traits of fifteen diverse rice genotypes. *Crop Sci.*, 38: 1130 — 1136.
- Samonte S., Wilson L., Tabien R. 2005. Maximum node production rate and main culm node number contributions to yield and yield-related traits in rice. *Field Crops Research* 96, 313 — 319.
- Shahinnia F., Rezai M., Sayed Tabatabaei E. 2005. Variation and path analysis of important agronomic traits in two- and six-rowed recombinant inbred lines in barley (*Hordeum vulgare* L.). *Czech J. Genet. Plant Breed.* 41: 246 — 250.
- Shiple B. 2002. *Cause and correlation in biology. A user's guide to path analysis, structural equations and causal inference*. Cambridge University Press, Cambridge.
- Sinebo W. 2002. Yield relationships of barleys grown in a tropical highland environment. *Crop Sci.*, 42: 428 — 437.
- Slafer G. 2003. Genetic basis of yield as viewed from a crop physiologist's perspective. *Ann. Appl. Biol.*, 142: 117 — 128.
- Soenarto 2001. Morphological components of yield in sweet potato (*Ipomoea batatas* (L.) Lam.). *Praca doktorska*, University of British Columbia, Kanada.
- Spaner D., Todd A., McKenzie B. 2001. The effect of seeding rate and nitrogen fertilization on barley yield and yield components in a cool maritime climate. *J. Agronomy & Crop Science* 187: 105 — 110.
- Szafranski W., Kulig B. 1998. Metoda współczynników ścieżek w ocenie współzależności cech determinujących plon ziarna jęczmienia jarego. *Zesz. Nauk. AR Kraków ser. Rolnictwo*, 35: 93 — 101.
- Szempliński W. 2003. Siedliskowe i agrotechniczne uwarunkowania produkcji ziarna jęczmienia jarego na paszę w północno-wschodniej Polsce. *UWM Olsztyn*: 7 — 87.
- Tarimo A. 1997. Physiological response of groundnut to plant population density. *African Crop Science Journal* 5 (3): 267 — 272.
- Wright S. 1921. Correlation and causation. *J. Agric. Res.* 20: 557 — 585.
- Wright S. 1923. The theory of path coefficients — a reply to Niles's criticism. *Genetics* 8: 239 — 255.
- Wright S. 1934. The method of path coefficients. *Ann. Math. Statist.* 5: 161 — 215.

Zajac T., Krawontka J., Szmigiel A. 1997. Oszacowanie determinacji plonu ziarna jęczmienia jarego na podstawie predykcji regresyjnej i zmienności elementów struktury plonu. Zesz. Nauk. AR Kraków ser. Rolnictwo, 34: 141 — 151.