

Prof. dr hab. Henryk Bujak  
Uniwersytet Przyrodniczy we Wrocławiu  
Katedra Genetyki, Hodowli Roślin i Nasiennictwa

## RECENZJA

### **rozprawy doktorskiej mgr Kingi Noras pt. „Analiza danych niekompletnych z serii doświadczeń odmianowo-uprawowych na podstawie liniowych modeli mieszanych”**

Praca została wykonana na Wydziale Rolnictwa i Biologii Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie w Katedrze Doświadczalnictwa i Bioinformatyki pod promotorstwem prof. dr hab. Wiesława Mądrego. Promotorem pomocniczym ocenianej rozprawy doktorskiej był dr inż. Marcin Studnicki.

Rejonizacja odmian, a co za tym idzie określenie stabilności ich plonowania w różnych warunkach klimatyczno-glebowych stanowi ważny czynnik w podnoszeniu produkcji rolniczej. Odmiany wpisane do rejestrów odmian na podstawie badań prowadzonych przez jednostki rejestrowe wykazują różne reakcje na zmieniające się warunki glebowe, klimatyczne czy agrotechniczne, natomiast wielokrotnie brak jest szczegółowych informacji o tych reakcjach. Utworzony w Polsce Porejestrowe Doświadczalnictwo Odmianowe (PDO), organizuje sieć doświadczeń porejestrowych z zarówno z odmianami wpisanymi do krajowego (KR) jak i wspólnotowego katalogu odmian (CCA) dostarcza tych informacji. Istotą systemu PDO jest jego decentralizacja, to znaczy że o doborze odmian do doświadczeń, za wyjątkiem odmian wzorcowych, decydują powołane w tym celu zespoły regionalne. Powoduje to dużą zmienność i rotację badanych odmian, co w konsekwencji utrudnia analizę statystyczną otrzymanych wyników z serii kilkuletnich doświadczeń odmianowo - agrotechnicznych i poprawną rejonizację. Uzyskane układy o niekompletnych danych są trudne do analizy statystycznej, dlatego częściej dokonuje się analiz dla układów kompletnych ograniczając liczbę odmian do tych, które powtarzają się we wszystkich latach i miejscowościach lub ograniczając liczbę miejscowości, zwykle do kilku, w których wszystkie odmiany były badane przez analizowany okres. Dla takich kompletnych danych stosuje się zwykle metody analizy wariancji (ANOVA). Niewiele natomiast jest prac, w których do oceny stabilności i rekomendacji odmian zostały wykorzystane dane z układów niekompletnych.

Przedstawiona do oceny rozprawa doktorska mgr Kingi Noras wychodzi naprzeciw problemom niekompletności układów danych pochodzących z serii doświadczeń wielośrodowiskowych (powtarzanych w latach i miejscowościach) i przedstawia możliwości zastosowania dla takich danych metod statystycznych opartych na liniowych modelach mieszanych oraz analizie skupień. Zaproponowane przez Doktorantkę liniowe modele mieszane w klasyfikacji Odmiana x Agrotechnika x Miejscowość x Rok (GxMxLxY) zostały zilustrowane i zweryfikowane z wykorzystaniem danych pochodzących z doświadczeń porejestrowych z odmianami pszenicy ozimej.

Przedstawiony do oceny maszynopis rozprawy doktorskiej mgr Kingi Noras posiada 121 stron, w tym 19 tabel (tabele 13-19 w załączniku na końcu maszynopisu), 17 wykresów i 3 rysunki.

Przechodząc do oceny formalnej pracy doktorskiej można stwierdzić, że podział treści na rozdziały i podrozdziały jest powszechnie stosowany w tego typu opracowaniach i nie budzi zastrzeżeń. Rozprawa zawiera następujące rozdziały: streszczenia w języku polskim i angielskim, wstęp i cel pracy z podrozdziałami (Wprowadzenie oraz Cel i zakres pracy), przegląd literatury z podrozdziałami, materiały i metody z wydzielonymi podrozdziałami, wyniki z podrozdziałami połączone z dyskusją w każdym z nich, wnioski, spis literatury, spis wykresów i rysunków, spis tabel oraz załącznik z tabelami od 13 do 19.

Wprowadzenie do badań przedstawione we wstępie i przeglądzie literatury jest szczegółowe, wielowątkowe, a równocześnie jasno i syntetycznie informuje czytelnika o zagadnieniach, które są celem dysertacji. Doktorantka szczegółowo przedstawiła znaczenie pszenicy oraz rekomendacji odmian na podstawie doświadczeniach wielokrotnych oraz metody statystyczne stosowane do opracowania uzyskanych z tych doświadczeń danych eksperymentalnych, a zwłaszcza danych niekompletnych. Jak słusznie zauważa, niewiele jest publikacji naukowych, w których przedstawiono sprawdzone i uznane metody opracowania danych niekompletnych z doświadczeń odmianowo – agrotechnicznych, a najczęściej wykorzystywane są w tym celu modele liniowe modele mieszane z metodą REML. W literaturze można znaleźć za to wiele prac, w których opracowane zostały dane z układów kompletnych.

Doktoranta definiuje pojęcie wielośrodowiskowych serii doświadczeń oraz przedstawia kilka ich typów w zależności od liczby lat lub miejscowości, w których są prowadzone, co przekłada się z kolei na zakres wnioskowania statystycznego. Najbardziej efektywne z nich to takie, które prowadzone są w wielu latach i w wielu miejscowościach ponieważ pozwalają one na pełną ocenę reakcji adaptacyjnej odmian oraz jej powtarzalności w latach.



Doświadczenia tego typu są najczęściej stosowane w końcowych etapach hodowli roślin, badaniach rejestrowych oraz porejestrowych. Na podstawie ich wyników można ocenić stabilność plonowania badanych odmian oraz ich reakcje adaptacyjne. Dalej przedstawione zostały korzyści płynące z serii doświadczeń porejestrowych oraz trudności dotyczące prawidłowego statystycznego opracowania uzyskanych wyników. Bardzo interesującymi są podrozdziały, w których Doktorantka przedstawia możliwości zastosowania liniowych modeli mieszanych do opracowywania wyników serii doświadczeń odmianowych. Do prawidłowego statystycznego opracowania wyników z serii doświadczeń wielokrotnych niekompletnych niezwykle trudno jest dobrać właściwy model mieszany, dlatego najczęściej czyni się to na podstawie pewnych założeń o stałości i losowości poszczególnych czynników (odmian, lat, miejscowości) oraz wyborze o właściwej macierzy kowariancji. W zależności od przyjętych założeń, do opracowania tego typu danych, można zastosować wiele mniej lub bardziej skomplikowanych modeli mieszanych, które mogą dawać dość różne oceny odmian. W ocenianej rozprawie doktorskiej, mgr Knaga Noras postanowiła wykorzystać, dla niekompletnych danych plonu odmian pszenicy ozimej, liniowe modele mieszane z różnymi macierzami kowariancji efektów losowych oraz ocenić ich dopasowanie na podstawie kryteriów informacyjnych i wizualizacji estymacji efektów Odmiana - Miejscowość. Najlepiej dopasowany 4-czynnikowy model mieszany został następnie wykorzystany do oceny reakcji adaptacyjnej odmian na środowisko oraz intensywność agrotechniki. Jest to nowy bardzo interesujący sposób opracowania wyników doświadczeń wielośrodowiskowych, a zwłaszcza takich, w których brakuje obserwacji dla wszystkich odmian, z czym niejednokrotnie, jak już wcześniej podkreślałem, mają do czynienia doświadczalnicy.

W mojej ocenie, Doktorantka zademonstrowała dobrą znajomość literatury z zakresu prowadzonych badań a opracowane rozdziały świadczą o tym, że jest dobrze przygotowana do dalszych etapów swoich badań. Pozwoliło to mgr Kindze Noras na właściwe sformułowanie celu i zakresu pracy oraz prawidłowe i konsekwentne zaplanowanie i przeprowadzenie kolejnych etapów badań.

Piśmiennictwo jest bogate i obejmuje 133 pozycje literatury związane z tematyką rozprawy, zdecydowaną większość stanowią opracowania w języku angielskim, a o ich ważności i aktualności podjętego w pracy problemu świadczy najnowsza opublikowana w ostatnich pięciu latach literatura. Dodatkowo Doktorantka korzystała także z informacji zawartych na 4 stronach internetowych.

W rozdziale „Materiały i metody” Doktorantka szczegółowo charakteryzuje materiał badawczy i warunki prowadzenia doświadczeń z odmianami pszenicy ozimej, a następnie

przedstawia zastosowane metody analizy danych w niekompletnej klasyfikacji  $G \times M \times L \times Y$  uzyskane z doświadczeń polowych. Doktorantka do opracowania wykorzystwała plony ziarna odmian pszenicy ozimej z doświadczeń polowych prowadzonych w ramach systemu PDO w 18 miejscowościach przez okres 3 lat, które utworzyły niekompletny spójny czteroczynnikowy układ obejmujący 61 odmian. Do analizy statystycznej danych niekompletnych zastosowała dwustopniowe podejście. W pierwszym etapie wykonała, dla danych surowych, dwuczynnikowe analizy wariancji, a uzyskanych z pojedynczych doświadczeń poprawionych średnich dla odmian utworzyła czterokierunkową niekompletną klasyfikację krzyżową  $G \times M \times L \times Y$ . Średnie poprawione wykorzystane zostały w drugim etapie do łącznej analizy z wykorzystaniem dwóch liniowych modeli mieszanych, 4-czynnikowego standardowego liniowego modelu mieszanego oraz 4-czynnikowego liniowego modelu mieszanego z efektami odmian zagnieżdżonymi w obrębie miejscowości i sposobów agrotechniki. Nie mam zastrzeżeń do zaproponowanych metod analizy danych, a prowadzone analizy statystyczne są kolejnymi rozwinięciami wcześniejszych metod. Metodyka ta jest bardzo nowatorska i wnosi bardzo dużą wartość do ocenianej rozprawy. Kolejne etapy analizy wykonane zostały poprawnie, a uzyskane wyniki są wiarygodne i dają odpowiedzi na postawione w celach problemu badawcze. Weryfikacja założeń metodycznych analizy danych była możliwa dzięki zgromadzeniu przez Doktorantkę odpowiedniego materiału badawczego i logicznego podejścia do analizy tej olbrzymiej bazy danych.

W tabeli 4 w tym rozdziale proponuję zmienić tytuł, ponieważ dane w niej zawarte nie charakteryzują odmian (np. ich wartości jakościowej, odporności na patogeny itd.) a jedynie informują o roku wpisania do KR lub CCA oraz hodowcy. Pewne nieścisłości dotyczą wpisu odmian do KR i CCA (str. 46). W momencie wejścia Polski do Unii Europejskiej wszystkie odmiany znajdujące się w KR (Krajowym Rejestrze) zostały automatycznie wpisane do CCA (Wspólnotowego Katalogu Odmian) i od tego momentu każda nowo zarejestrowana odmiana (wpisana do KR) jest automatycznie wpisywana do CCA.

Zgodnie z metodyką badań WGO dla doświadczeń rejestrowych oraz doświadczeń PDO odmiany pszenicy badane są na dwóch poziomach agrotechniki podstawowym (A1) i wysokim (A2). Na str. 47 Autorka rozprawy pisze: „Pierwszy poziom agrotechniki A1 (zwany umiarkowanie intensywnym) charakteryzuje się ...” Proponuję stosować terminologię zgodnie z metodyką opracowaną przez COBORU. Podobną nieścisłość zauważyłem jeszcze w innych miejscach pracy.

Na str. 55 przedstawiony został 4-czynnikowy liniowy model mieszany (I), w którym efekt  $L_j$  to „stały efekt główny j-tego środowiska”. Czy nie lepiej nazwać go efektem stałym j-



tej miejscowości (ang. L – locations)? Tym bardziej, że w pierwszym zdaniu pod modelem Autorka pisze: „W tym modelu czynnikami o stałych efektach są miejscowości oraz intensywność agrotechniki, natomiast ...” Efektem środowiska mogłaby być wtedy interakcja „LM<sub>jk</sub>”. Podobna uwaga dotyczy modelu II na str. 57.

Wyniki analiz danych według zaproponowanych metod statystycznych mgr Kinga Noras podzieliła na trzy części (podrozdziały), przyjmując zasadę, że po omówieniu uzyskanych rezultatów swoich analiz dyskutuje je z wynikami innych autorów. Połączenie wyników z dyskusją moim zdaniem znacznie ułatwia śledzenie i weryfikację poprawności przeprowadzonych analiz i uzyskanych wyników w konfrontacji z danymi literaturowymi bez konieczności powrotu do nich gdyby utworzono następny rozdział „Dyskusja wyników”. W pierwszej części wyników, przeprowadzona analiza według klasycznego liniowego modelu mieszanego pozwoliła na ocenę istotności wszystkich efektów losowych i stałych wydzielonych w modelu. W przyjętym modelu wszystkie efekty stałe oraz większość losowych (za wyjątkiem (Agrotechnika x Rok i Odmiana x Agrotechnika x Rok)) były istotne. Następnie przedstawiono wyniki oceny dopasowania 4-czynnikowego modelu z zagnieżdżonymi efektami odmian na podstawie kryteriów informacyjnych (AIC oraz BIC) i logarytmu z maksymalnej wartości funkcji wiarygodności (LogL) oraz na podstawie oceny reakcji odmian na środowisko. Niestety w przypadku kryteriów informacyjnych możliwości obliczeniowe posiadanego sprzętu komputerowego nie pozwoliły na wydzielenie wszystkich 18 macierzy dla analiz FA (factor analytic), a jedynie 14, co nie umniejsza wartości uzyskanych wyników, ponieważ przyjęte kryteria AIC, jak i BIC dla modeli o większej liczbie składowych (FA(11); FA(12); FA(13); FA(14)) przyjmują najwyższe wartości. Najlepiej dopasowane wydają się być modele z najprostszymi macierzami kowariancji, jednak zastosowane kryteria nie pozwoliły na jednoznaczną ocenę zastosowanych macierzy kowariancji, a ich oceny były niekiedy sprzeczne. Doktorantka pojęła zatem próbę oceny dopasowania modeli z różnymi macierzami kowariancji na podstawie wizualizacji wyników estymacji efektów Odmiana - Miejscowość i przedstawieniu ich na wykresach reakcji plonu na warunki miejscowości. Należy podkreślić, że jest to nowatorskie podejście Doktorantki, nie stosowane wcześniej i nie opisywane w literaturze. Jednak przedstawiony na wykresie 4 obraz reakcji plonowania odmian uzyskany za pomocą modeli (II) z macierzą kowariancji CS (compound symmetry) jest, jak zauważa Autorka, całkowicie nierealistyczny. Zachodzi zatem pytanie, czy uśrednienie plonów odmian z obu poziomów agrotechniki A1 i A2 nie spowodowało spłylenia różnic reakcji odmian na warunki miejscowości (środowisk)?

Zaproponowane podejście (macierze FA) umożliwiło precyzyjną identyfikację środowisk o dużej mocy dyskryminacyjnej, czyli dzięki macierzom FA udało się zidentyfikować środowiska (miejscowości), które charakteryzują się największym zróżnicowaniem plonów badanych odmian, co jest dużym plusem tej metody. Przeprowadzone analizy pozwoliły na wybór modelu z założeniem macierzy typu UN (unstructured) jako optymalnej do zastosowania w analizie danych niekompletnych z doświadczeń odmianowych wielośrodowiskowych.

W kolejnej trzeciej części wyników analiz przedstawiono rezultaty analiz na podstawie rekomendowanego dla układów niekompletnych modelu (II) z założeniem macierzy kowariancji typu UN, które pozwoliły na ocenę reakcji odmian na zastosowane czynniki oraz podział odmian w zależności od ich reakcji na zmieniające się warunki środowiska na odmiany o wąskiej i szerokiej adaptacji. Odmiany stabilne i o szerokiej adaptacji plonowały na podobnym poziomie we wszystkich miejscowościach (powyżej średniej), natomiast wysokie plony w niektórych tylko miejscowościach świadczą o wąskiej adaptacji odmiany do środowiska i niestabilności jej plonowania. Jednak odmiany takie mogą być rekomendowane do konkretnych mikrorejonów (miejscowości) uprawy. Zastosowan w pracy analiza skupień pozwoliła na wydzielenie odmian o podobnej reakcji na warunki środowiskowe. Analiza skupień została wykonana oddzielnie dla każdego z poziomów agrotechniki. Jest to moim zdaniem poprawne podejście, ponieważ odmiany pszenicy różnią się pod względem reakcji na poziom agrotechniki i można wśród nich wyróżnić formy intensywne, które silnie reagują zwyżką plonów pod wpływem wyższego poziomu agrotechniki (nawożenie, zabiegi ochrony), które są polecane dla gospodarstw stosujących nowoczesne, techniki uprawy oraz odmiany, ekstensywne, których przyrost plonu pod wpływem warunków agrotechnicznych nie jest tak spektakularny. Odmiany drugiego typu polecane są dla gospodarstw prowadzących produkcję w sposób tradycyjny. Uzyskane wyniki są bardzo wartościowe dla praktyki doświadczalnej i rekomendacji odmian, ponieważ wykazały, że dla danych o dużym stopniu niekompletności możliwy jest dobór właściwych liniowych modeli mieszanych, które pozwalają na efektywną ocenę statystyczną i wnioskowanie o rolniczej przydatności dużej liczby odmian.

Rozprawę kończą zwięźle sformułowane wnioski, które uwypuklają otrzymane przez Doktorantkę wyniki analiz z zaproponowanymi liniowymi modelami mieszanymi. Na niektóre z nich zwróciłem uwagę, wcześniej, ponieważ mają one duże znaczenie dla praktyki hodowlanej oraz oceny odmian.



W trakcie opracowania rozprawy Doktorantka nie ustrzegła się pewnych błędów, skrótów myślowych, zwrotów żargonowych, czy nieścisłości, na które chciałbym zwrócić uwagę:

s. 13 „..., gdyż na plon i cechy jego jakości ma wpływ wiele czynników ...”; powinno być „..., gdyż na plon i jego jakość ma wpływ wiele czynników ...”;

s. 17 „... oraz miejscowości badawczych”; raczej powinno być „... oraz miejscowości w których prowadzone są badania”; podobnie na str. 24; 41;

s. 20 „... są najlepszym predyktorem...” proponuję użyć polskiego słowa zastępującego anglojęzyczne słowo „prediction” polskim „prognoza”; podobnie na str. 30,

str. 20 „Sektor hodowli roślin...”; bardziej pasuje tu; „Hodowla roślin ...”,

str. 21-22 „Odmiany o wąskiej adaptacji charakteryzują się ograniczonym rynkiem zbytu, więc głównym celem hodowcy jest wyhodowanie odmiany o możliwie szerokiej adaptacji.” ; O jaki rynek zbytu chodzi? Odmiana dobrze plonująca w danych warunkach mikroregionu (wąska adaptacja do lokalnych warunków) przyniesie rolnikowi na wet wyższe dochody niż odmiana o szerokiej adaptacji, która w tych samy warunkach może plonować niżej.

str. 31 ang. Maximum Likelihood – metoda najmniejszej, czy największej wiarygodności (prawdopodobności)?

str. 40 „... w ramach krajowej serii doświadczeń rodzaju L.”; powinno być: „... w ramach krajowej serii L doświadczeń.” ,

str. 41 „W analizowanym okresie średnio poprzez miejscowości badano 42 odmiany...”; powinno być: „W analizowanym okresie średnio w miejscowościach badano 42 odmiany...” ,

str. 52 „Przebieg pogody w okresie zimowym nie stwarzał większych zagrożeń dla zimujących upraw.”; powinno być: „Przebieg pogody w okresie zimowym nie stwarzał większych zagrożeń dla zimujących roślin.”

str. 73 powtórzenie zdania ze strony 69 w. 2-6.

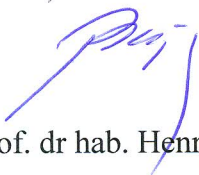
Inne uwagi zaznaczyłem i naniosłem ołówkiem w tekście maszynopisu.

Zauważone w pracy nieliczne niedociągnięcia i usterki, głównie literowe, stylistyczne, skróty myślowe oraz zwroty żargonowe nie umniejszają jej dużej wartości merytorycznej. Mogą być one z łatwością usunięte w trakcie przygotowania tych ciekawych wyników badań do druku, a wydają się być konsekwencją konieczności szybkiego przygotowania dysertacji.

## **Wniosek końcowy**

Przedstawiona do oceny rozprawa doktorska Pani Kingi Noras wnosi szereg nowych istotnych elementów do opracowywania wyników doświadczeń odmianowych prowadzonych w wielu środowiskach nie tylko z punktu widzenia naukowego, lecz także poznawczego oraz praktycznego. Nader często w badanych odmianowych otrzymuje się niekompletne dane, których opracowane statystyczne i wiarygodność wnioskowania nastęrcza wielu problemów. Uważam, zatem że temat podjęty przez Doktorantkę jest bardzo aktualny, a pytania postawione w celach badawczych sformułowano w sposób jasny i prawidłowy. Wytyczony cel został zrealizowany dzięki zastosowaniu właściwych metod analiz statystycznych, a uzyskane wyniki i wnioski znajdują uzasadnienie w materiale dokumentacyjnym przedstawionym w tabelach i na rysunkach. Rozprawa wnosi nowe elementy wiedzy w dziedzinie doświadczalnictwa i biometrii, a szczególnie w doborze właściwych liniowych modeli mieszanych do wielostronnej oceny odmian na podstawie niekompletnych danych pochodzących z doświadczeń prowadzonych w wielu miejscowościach w okresie kilku lat.

Praca mgr Kingi Noras pt. „Analiza danych niekompletnych z serii doświadczeń odmianowo-uprawowych na podstawie liniowych modeli mieszanych” spełnia wszystkie wymogi stawiane rozprawom doktorskim i stawiam wniosek do Rady Wydziału Rolnictwa i Biologii Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie o dopuszczenie Jej do dalszych etapów przewodu doktorskiego. Ze względu na nowatorstwo pracy, wagę podjętego tematu, zakres przeprowadzonych badań i analiz statystycznych oraz kompleksowość w podejściu do rozwiązywanego zagadnienia wnioskuję o wyróżnienie rozprawy nagrodą.

  
Prof. dr hab. Henryk Bujak

Wrocław, dnia 28.09.2015 r.