

Autoreferat

dr inż. Marcin Jacek Studnicki

Katedra Doświadczalnictwa i Bioinformatyki

Wydział Rolnictwa i Biologii

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Warszawa, 2018

1. Imię i Nazwisko

Marcin Jacek Studnicki

2. Posiadane dyplomy

2012 – doktor nauk rolniczych, dyscyplina agronomia, Wydział Rolnictwa i Biologii, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, tytuł rozprawy: „Efektywność metod pobierania próby w tworzeniu kolekcji podstawowej roślinnych zasobów genowych”, promotor: prof. dr hab. Wiesław Mądry

2008 – magister rolnictwa, specjalność „Agronomia i agrobiznes”, Wydział Rolnictwa i Biologii, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, tytuł pracy: „Wielocechowa ocena różnorodności fenotypowej obiektów w kolekcji zasobów genowych pszenicy jarej”, promotor: prof. dr hab. Wiesław Mądry

2007 – inżynier rolnictwa, specjalność „Informatyka w rolnictwie”, Wydział Rolnictwa i Biologii, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, tytuł pracy: „Metodyka gromadzenia, oceny, przechowywania i wykorzystania roślinnych zasobów genowych w Polsce”, promotor: prof. dr hab. Wiesław Mądry

3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

2012- obecnie – adiunkt, Katedra Doświadczalnictwa i Bioinformatyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

2011-2012 – asystent, Katedra Doświadczalnictwa i Bioinformatyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311):

a) Tytuł osiągnięcia naukowego

Ocena adaptacji odmian roślin uprawnych z zastosowaniem liniowych modeli mieszanych

b) (autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa, recenzenci wydawniczy),

- 1) Studnicki M., Paderewski J., Piepho H.P., Wójcik-Gront E. 2017. Prediction accuracy and consistency in cultivar ranking for factor-analytic linear mixed models for winter wheat multi-environmental trials. *Crop Science* 57:2506-2516. [IF₂₀₁₆=1,629; MNiSW₂₀₁₆=30 pkt; mój udział procentowy szacuję na 70%]
- 2) Derejko A., Studnicki M., Mądry W., Gacek E. 2016. A comparison of winter wheat cultivar rankings in groups of Polish locations. *Cereal Research Communications* 44:628–638. [IF₂₀₁₆=0,496; MNiSW₂₀₁₆=15 pkt; mój udział procentowy szacuję na 60%]
- 3) Studnicki M., Wijata M., Sobczyński G., Samborski S., Gozdowski D., Rozbicki J. 2016. Effect of genotype, environment and crop management on yield and quality traits in spring wheat. *Journal of Cereal Science* 72:30-37. [IF₂₀₁₆=2,223; MNiSW₂₀₁₆=35 pkt; mój udział procentowy szacuję na 50%]
- 4) Studnicki M., Mądry W., Noras K., Wójcik-Gront E., Gacek E. 2016. Yield response of winter wheat cultivars to environments modeled by different variance-covariance structures in linear mixed models. *Spanish Journal of Agricultural Research* 14: e0703. [IF₂₀₁₆=0,687; MNiSW₂₀₁₆=25 pkt; mój udział procentowy szacuję na 80%]
- 5) Studnicki M., Mądry W., Derejko A., Noras K., Wójcik-Gront E. 2015. Four-way data analysis within the linear mixed modelling framework. *Scientia Agricola* 72: 411-419. [IF₂₀₁₅=0,954; MNiSW₂₀₁₅=30 pkt; mój udział procentowy szacuję na 80%].

Łączny IF ww. publikacji zgodnie z rokiem wydania lub dla prac opublikowanych w 2017 roku zgodny z IF₂₀₁₆ wynosi 5,989. Łączna liczba punktów wg. MNiSW zgodnych z rokiem wydania lub dla prac opublikowanych w 2017 roku zgodny z listą z roku 2016 wynosiła 135 punktów.

- a) omówienie celu naukowego/artystycznego ww. pracy/prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.**

Wstęp

W celu przeprowadzenia efektywnej selekcji genotypów przez hodowcę, a także później rekomendacji odmian do uprawy w gospodarstwach należy ocenić reakcję plonu oraz innych ważnych użytkowo cech (np. cech jakości ziarna) odmian roślin uprawnych w zróżnicowanych warunkach środowiska (Liu et al. 2013). To zadanie realizowane jest poprzez właściwą ocenę występowania i znaczenia interakcji pomiędzy genotypem (G) a środowiskiem (E), zwanej interakcją genotypowo-środowiskową GE. Ta interakcja jest definiowana jako zróżnicowana reakcja plonu jak i innych ważnych cech odmian w poszczególnych środowiskach i jest ona powszechnym zjawiskiem w genetyce oraz hodowli roślin uprawnych. Ocena interakcji GE stanowi podstawę do badania adaptacyjności odmian w zróżnicowanych warunkach środowiska. Badanie adaptacji odmian na warunki środowiska polega na badaniu przystosowania i produktywności odmian do rozpatrywanych środowisk (Mądry i Iwańska 2011). Z punktu widzenia hodowcy ważne jest aby odmiana charakteryzowała się relatywnie wysokim plonem we wszystkich lub w możliwie maksymalnej liczbie rozpatrywanych środowisk. Hodowca zainteresowany jest uzyskaniem odmiany o szerokiej adaptacji, która zwiększa jego sukces komercyjny. Odmiana charakteryzująca się szeroką adaptacją może być z powodzeniem stosowana w dużym zakresie zmienności środowiskowej (warunków glebowych i klimatycznych), charakteryzując się również dużą stabilnością. Poza odmianami o szerokiej adaptacji wyróżniamy także odmiany charakteryzujące się wąską adaptacją, czyli takie które są relatywnie wysoko produktywne (plenne) tylko w pewnej grupie środowisk. Wąska adaptacja odmian w ostatnim czasie staje się coraz bardziej

interesująca dla badaczy i hodowców. Pozwala ona osiągnąć sukces odmiany w specyficznych warunkach środowiska np. w uprawach ekologicznych.

Podstawowym narzędziem wykorzystywanym do badania i oceny interakcji genotypowo-środowiskowa, a co za tym idzie oceny stabilności i adaptacyjności odmian są wielośrodowiskowe serie doświadczeń polowych (ang. multi-environmental trials METs) (Annicchiarico et al. 2010). Wielośrodowiskowe serie doświadczeń polegają na powtarzaniu jednoczynnikowych (np. doświadczenia odmianowe), a coraz częściej wieloczynnikowych (np. doświadczenia odmianowo – agrotechniczne) doświadczeń w różnych środowiskach. To znaczy kilku miejscowościach (lokalizacjach) i/lub wielu latach (sezonach wegetacyjnych). Kombinacja lat z miejscowościami określana jest często jako pojedyncze środowisko. Wielośrodowiskowe serie doświadczeń bywają często rozbudowane, oceniana jest w nich duża liczba odmian w dużej liczbie miejscowości i przez wiele lat (Curti i in. 2014). Obserwacje plonu lub innych cech w rozbudowanych wielośrodowiskowych seriach doświadczeń bardzo często tworzą niekompletną wielokierunkową klasyfikację danych. Jednym z powodów występowania niekompletności w zbiorze danych jest zmienna (różna) liczba oraz skład badanych odmian w poszczególnych latach, a także bardzo często zróżnicowana liczba i skład odmian w miejscowościach. W przypadku niekompletnych zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń dokładna ocena efektów głównych i interakcji GE jest wyzwaniem.

Jedną z powszechnie wykorzystywanych metod analizy zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń jest klasyczna analiza wariancji ANOVA. Jednak ANOVA ma bardzo restrykcyjne założenia odnośnie wariancji i kowariancji efektów oraz błędu losowego (Crossa i in. 2010). Zwłaszcza dla wielośrodowiskowych serii doświadczeń są one nierealistyczne i ograniczają wiarygodność wnioskowania (So i Edwards 2009). Dodatkowo klasyczna ANOVA nie jest efektywnym narzędziem do analizy danych o znacznym stopniu niekompletności pochodzących z wielośrodowiskowych serii doświadczeń. Często stosowanym rozwiązaniem problemu z brakującymi obserwacjami jest ocena ograniczonej liczby odmian wybranych z całego zbioru danych pochodzących z wielośrodowiskowej serii doświadczeń tak aby stanowiły one kompletną klasyfikację (Mądry i in. 2017). Takie podejście umożliwia analizę i ocenę odmian z wykorzystaniem analizy wariancji ANOVA, jednak jednocześnie tracimy wiele ważnych i użytecznych informacji. W tym

wypadku wyniki z wielośrodowiskowej serii doświadczeń nie są w pełni i efektywnie wykorzystane, co uniemożliwia właściwą ocenę adaptacji odmian roślin uprawnych na warunki środowiska i wiarygodną ich rekomendację.

Bardziej elastyczne i realistyczne założenia odnośnie wariancji i kowariancji efektów losowych, w tym efektu interakcji GE, są osiągalne dzięki zastosowaniu w analizie danych pochodzących z wielośrodowiskowej serii doświadczeń podejścia opartego na liniowych modelach mieszanych (ang. Linear Mixed Model, LMM). Tradycyjnie liniowe modele mieszane mogą być zaprezentowane w postaci macierzowej za pomocą następującej formuły:

$$Y = Xb + Zu + e$$

gdzie: Y jest wektorem obserwacji (np. plonu), X jest macierzą układu dla efektów stałych, b jest wektorem efektów stałych, Z jest macierzą układu dla efektów losowych, u jest wektorem efektów losowych, e jest wektorem reszt (błędów losowych).

Zakłada się rozkład normalny dla efektów losowych u i e o postaci:

$$u \sim N(0, G)$$

$$e \sim N(0, R)$$

gdzie: G i R są odpowiednio strukturami (macierzami) wariancji i kowariancji dla efektów losowych i reszt (błędów losowych).

W liniowych modelach mieszanych stałe efekty b oszacowane są z zastosowaniem najlepszych nieobciążanych estymatorów (ang. best linear unbiased estimator BLUE), natomiast do estymacji efektów losowych u stosowane są najlepsze nieobciążone predyktory (ang. best linear unbiased prediction BLUP). Estymatory typu BLUP w porównaniu do estymatorów BLUE charakteryzują się mniejszą wariancją błędu oszacowania (Robinson 1991). Do estymacji poszczególnych parametrów w liniowych

modelach mieszanych jedną z wykorzystywanych metod jest metoda największej wiarygodności z restrykcją REML (ang. restricted maximum likelihood), która została opracowana przez Petersona i Thompsona (1971). Ponadto metoda REML dobrze radzi sobie z niekompletnymi zbiorami danych (Searl i in. 1992). Właśnie z tych dwóch powodów liniowe modele mieszane wraz z metodą REML są coraz częściej stosowane do analizy danych pochodzących z wielośrodowiskowych serii doświadczeń.

Zastosowanie liniowych modeli wraz z metodą REML umożliwia wykorzystanie bardziej realistycznych struktur wariancji i kowariancji dla efektów losowych (macierz G) i dla błędów losowych (macierz R). Z punktu widzenia hodowli i rekomendacji odmian ważne jest możliwie wiarygodne opisanie reakcji odmian na środowiska poprzez zastosowanie odpowiedniego typu struktury dla macierzy G w założonym liniowym modelu mieszanym. Najprostsza struktura wariancji i kowariancji zawiera tylko same wariancje dla poszczególnych efektów, zwane komponentami wariacyjnymi, natomiast kowariancje pomiędzy efektami są zerowe. Struktura ta zwana jest strukturą komponentów wariacyjnych (ang. variance components matrix VC) jest najczęściej (domyślnie) stosowaną w liniowych modelach mieszanych, również jest ona podstawową strukturą w trakcie klasycznej mieszanej analizy wariancji (ang. mixed ANOVA). Jedną z prostszych, ale już elastycznych struktur wariancji i kowariancji możliwą do oszacowania w liniowych modelach mieszanych z zastosowaniem metody REML jest macierz typu „compound symmetry” – CS. Ta struktura zakłada jednakowe wariancje i jednakowe kowariancje (Hu i Spilke 2011). W odróżnieniu od klasycznej analizy wariancji zastosowanie tej struktury pozwala na uwzględnienie kowariancji pomiędzy efektami. Najbardziej realistyczną i elastyczną z możliwych do zastosowania struktur wariancji i kowariancji jest macierz typu niestrukturalnego – UN (ang. unstructured). Pozwala ona na oddzielne oszacowanie wariancji i kowariancji dla poszczególnych elementów macierzy (Meyer 2009), zakłada ona heterogeniczność wariancji i kowariancji. Ze względu na dużą liczbę parametrów do oszacowania w tego typu struktury, jest ona kłopotliwa w praktycznym zastosowaniu. Struktura UN stwarza duże trudności numeryczne i obliczeniowe (Kelly i in. 2007). Dlatego też coraz częściej w analizie danych z wielośrodowiskowych serii doświadczeń wykorzystuje się strukturę wariancji i kowariancji typu „factor analytic” – FA (Stefanova i Buirchell 2010; Bergueno i in. 2011). Struktura FA jest aproksymacją macierzy niestrukturalnej, z wykorzystaniem analizy czynnikowej z dekompozycją Choleskiego

(Meyer 2009; So i Edwards 2009). Zaletą tej struktury jest to, że jej elastyczność jest porównywalna z macierzą niestrukturalną, ale przy mniejszej liczbie parametrów, które należy oszacować (Meyer i Kirkpatrick 2008; Kelly i in. 2007).

W trakcie analizy zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń z wykorzystaniem liniowych modeli mieszanych wykorzystuje się jednoetapowe lub wieloetapowe podejście (Piepho i in. 2012). W jednoetapowym podejściu do analizy stosuje się jeden model w którym uwzględnia się układ doświadczalny, wykorzystując bezpośrednio obserwację z pojedynczych powtórzeń. Ze względu na trudności i ograniczenia obliczeniowe to podejście może być trudne do rutynowego stosowania. W praktyce częściej stosowane jest podejście wieloetapowe. W trakcie wieloetapowego podejścia stosuje się najczęściej dwa etapy (dwuetapowa analiza). W pierwszym etapie przeprowadza się analizy dla każdego pojedynczego doświadczenia przeprowadzonego w ramach wielośrodowiskowej serii doświadczeń oddzielnie. Stosując w tym celu pojedyncze obserwacje z powtórzeń i uwzględniając w modelu układ doświadczalny. Na tym etapie wyznaczane są średnie poprawione (ang. adjusted mean) dla badanych odmian w poszczególnych środowiskach. W drugim etapie średnie poprawione pochodzące z pierwszego etapu wykorzystywane są do łącznej analizy (ang. combined analysis). Jednym z głównych problemów w zastosowaniach tego podejścia jest to, że każde z pojedynczych doświadczeń charakteryzuje się zróżnicowaną wariancją błędów, więc i średnie poprawione są oszacowywane z różną dokładnością. Rozwiązaniem tego problemu może być ważenie średnich poprawionych wariancją błędów z każdego pojedynczego doświadczenia oddzielnie, a później wykorzystanie ich w drugim etapie analizy. Stosowane są w tym celu różne podejścia. W badaniach efektywności różnych metod ważenia średnich poprawionych opartych na rzeczywistych przykładach i zbiorach danych wielośrodowiskowych serii doświadczeń okazuje się, że zastosowane metody ważenia średnich poprawionych nie poprawiają znacząco efektywności liniowych modeli mieszanych (Mohring i Piepho 2009). Co sugeruje, że nieważenie średnich poprawionych także pozwala na uzyskanie wiarygodnych wyników analiz przeprowadzonych w drugim etapie, a także nie wpływa znacząco na poprawność i wiarygodność wniosków.

Ważnym zagadnieniem związanym z zastosowaniem liniowych modeli mieszanych w analizie wielośrodowiskowych serii doświadczeń jest określenie rodzaju poszczególnych efektów (stały lub losowy), zwłaszcza dla efektu odmian (genotypów)

i środowisk (miejscowości, lat lub kombinacji miejscowości i lat). Tę decyzję można podjąć w oparciu o uzasadnienie merytoryczne traktowania jednych efektów jako stałe, a innych za losowe. Uznanie pewnych efektów za stałe, a innych za losowe ma również wpływ na właściwości i kształt struktury wariancji i kowariancji w liniowym modelu mieszanym. Jeżeli odmianę uznamy jako efekty stałe, a środowisko jako losowy wtedy dla elastycznych macierzy typu niestrukturalnego UN oszacowywane są wariancje zmiennej (np. plonu) w miejscowościach oddzielne dla każdej odmiany, a kowariancje są wyznaczone pomiędzy parami odmian. Natomiast, jeżeli uznamy odmianę za losowy efekt, a środowisko jako stałe wtedy w elastycznych strukturach wariancji i kowariancji oszacujemy wariancję odmian oddzielnie dla każdego środowiska, a kowariancje są wyznaczone dla par środowisk. Obydwa powyższe podejścia są bardzo często stosowane w literaturze, często nawet zamiennie. Można uznać, że wyłoniły się dwie szkoły. Pierwsza z nich, którą można nazwać „Brytyjsko-Australijską” preferuje losowe efekty odmian i stałe efekty środowiska (Smith i in. 2005), natomiast druga „Niemiecka” zakłada że odmiany są efektem stałym a miejscowości losowym (Hu i Spilke 2011, Laidig i in. 2014).

Przykładem wielośrodowiskowej serii doświadczeń w Polsce jest Porejestrowe Doświadczalnictwo Odmianowe (PDO). Jest to system doświadczeń odmianowych i odmianowo-agrotechnicznych (w zależności od gatunku rośliny uprawnej), który został założony w 1998 roku w celu rekomendacji odmian do uprawy na terenie Polski oraz w jej poszczególnych regionach. W doświadczeniach w systemie PDO biorą udział odmiany oficjalnie zarejestrowane w Polsce (Krajowy Rejestr- KR), jak i te pochodzące z Katalogu Wspólnotowego (Common Catalogue of Varieties of Agricultural Plant Species of European Union – CCV). Liczba badanych odmian oraz miejscowości, w których przeprowadza się doświadczenia zależy od znaczenia poszczególnych gatunków roślin. Najliczniejsza liczba doświadczeń w Porejestrowym Doświadczalnictwie Odmianowym przeprowadzana jest dla pszenicy ozimej. Gatunek ten jest także reprezentowany przez największą liczbę uczestniczących w tej serii doświadczeń odmian. Rekomendacja odmian przeprowadzana jest na poziomie poszczególnych województw przez specjalnie utworzone Wojewódzkie Zespoły Porejestrowego Doświadczalnictwa Odmianowego. W ramach prac zespołów powstają Listy Odmian Zalecanych (LZO) w województwie dla poszczególnych gatunków roślin uprawnych. We wszystkich pięciu publikacjach, stanowiących prezentowane osiągnięcie

naukowe, zostały wykorzystane dane pochodzące z doświadczeń dla zbóż (pszenica ozima i jara) wykonywanych w ramach Porejestrowego Doświadczalnictwa Odmianowego. Pojedyncze doświadczenia dla rozpatrywanych gatunków są to doświadczenia dwuczynnikowe (odmiana, poziom agrotechniki), przeprowadzone w układzie split-plot w dwóch powtórzeniach.

Celem publikacji stanowiących niniejsze osiągnięcie naukowe była ocena liniowych modeli mieszanych z różnymi typami struktur wariancji i kowariancji oraz poznanie właściwości struktur wariancji i kowariancji w zróżnicowanych zastosowaniach (Publikacja nr 1 i Publikacja nr 4). Ponadto, celem było opracowanie liniowych modeli mieszanych do oceny adaptacji odmian w klasyfikacji trój lub czterokierunkowej oraz przedstawienie możliwości wnioskowania z przedstawionych modeli i ich wykorzystanie do merytorycznej oceny odmian roślin uprawnych (Publikacja nr 2, Publikacja nr 3 i Publikacja nr 5).

Ocena liniowych modeli mieszanych z różnymi typami struktur wariancji i kowariancji

W przypadku analizy danych pochodzących z wielośrodowiskowych serii doświadczeń z wykorzystaniem liniowych modeli mieszanych, rekomendowane jest w pierwszej kolejności porównanie modeli z różnymi strukturami wariancji i kowariancji (So i Edwards 2009; Bergueno i in. 2011) praktycznie dla każdego zbioru danych pochodzących z wielośrodowiskowych serii doświadczeń oddzielnie. W drugiej kolejności wybrana najlepsza struktura wariancji i kowariancji zostaje zastosowana do poprawnej analizy danych i oceny odmian roślin uprawnych. Ocena liniowych modeli mieszanych z różnymi strukturami wariancji i kowariancji jest bardzo skomplikowana oraz nie zawsze jest możliwa do przeprowadzenia w ramach rutynowych działań hodowców czy innych badaczy odmian. Dlatego cenne są badania ogólnych właściwości tego typu modeli, a zwłaszcza struktur wariancji i kowariancji. Porównanie liniowych modeli mieszanych obejmuje ocenę dopasowania modelu (ang. goodness of fit) oraz dokładności oszacowania efektów (ang. prediction accuracy). Takie badanie i ocena modeli z różnymi strukturami przeprowadza się z wykorzystaniem różnych podejść i metod.

Ocenę dopasowania modeli z różnymi strukturami wariancji i kowariancji stosuje się najczęściej z wykorzystaniem kryteriów informacyjnych (Hu i Spilke 2011) między innymi z kryterium informacyjnym Akaike AIC (ang. Akaike information criterion) lub Bayesowskim kryterium informacyjnym BIC (ang. Bayesian information criterion). Obydwa te kryteria oparte są na wartości logarytmu funkcji wiarygodności (ang. log-likelihood) i odpowiedniej kary za liczbę parametrów w modelu. Podejście to było dość powszechnie stosowane w ocenie różnych typów struktur wariancji i kowariancji w liniowych modelach mieszanych dla danych pochodzących z wielośrodowiskowych serii doświadczeń. Jednakże te badania były oparte na zbiorach danych które charakteryzowały się małym stopniem niekompletności i pochodziły z doświadczeń stosunkowo niewielkich, w których badano stosunkowo małą liczbę odmian w ograniczonej liczbie miejscowości. To znacząco ogranicza możliwości poznania właściwości struktur wariancji i kowariancji w bardzo rozbudowanych oraz niekompletnych wielośrodowiskowych seriach doświadczeń.

Ocenę dopasowania modeli z różnymi macierzami kowariancji dla rozbudowanego zbioru danych z znacznym stopniem niekompletności przeprowadzono w Publikacji nr 4. Celem niniejszej pracy było porównanie 3 typów struktur wariancji i kowariancji – struktury CS, UN i FA ze składowymi od 1 do 17. Ocena została wykonana z wykorzystaniem obserwacji dla plonu 63 odmian pszenicy ozimej w 18 miejscowościach przeprowadzonych w 3 sezonach wegetacyjnych (2008/2009-2010/2011). Dane pochodziły z Porejstrowego Doświadczalnictwa Odmianowego. Ze względu na zmienną liczbę badanych odmian w poszczególnych sezonach wegetacyjnych zbiór danych stanowił 68,4% obserwacji w stosunku do kompletnej klasyfikacji. Analizę danych przeprowadzono wykorzystując dwuetapowe podejście (Möhring i Piepho 2009; Piepho i in. 2012). W pierwszym etapie w każdym pojedynczym doświadczeniu z serii przeprowadzono analizę w modelu zgodnym z układem doświadczalnym i wyznaczono średnie poprawione. Zostały one wykorzystane w drugim etapie, którego celem była łączna analiza oparta na trójkierunkowym liniowym modelu mieszanym, w którym dla losowego efektu odmian w poszczególnych miejscowościach założono badane typy struktur macierzy wariancji i kowariancji. Model ten posłużył do oszacowania średnich poprawionych dla kombinacji odmian i miejscowości, według algorytmu zaproponowanego przez Welham i in. 2004, opartego na oszacowaniu odpowiednich

losowych BLUP i stałych BLUE efektów w zaproponowanym liniowym modelu mieszanym.

Do oceny dopasowania zaproponowanego modelu z różnymi strukturami wariancji i kowariancji w pracy zastosowano kryterium informacyjne Akaikego. Dodatkowo zaproponowano ocenę rozpatrywanych struktur opartą na ocenie zróżnicowania średnich poprawionych dla kombinacji odmiana x miejscowość oraz przedstawienie ich za pomocą wykresu reakcji odmian na środowisko. Średnie poprawione dla plonu uzyskane z modelu z macierzą typu CS, charakteryzują się małym zróżnicowaniem zakresu plonu odmian w poszczególnych miejscowościach. Zakres plonu w rozpatrywanych miejscowościach dla tej struktury kształtował się od 0,66 do 0,98 t*ha⁻¹. Przy zastosowaniu tej struktury moglibyśmy dojść do błędnego wniosku, że miejscowości charakteryzują się podobną lub nawet jednakową siłą dyskryminacyjną plon odmiany. W rzeczywistości tak nie jest, miejscowości charakteryzują się silnym zróżnicowaniem zakresu plonu w poszczególnych lokalizacjach. Ponadto model ze strukturą typu CS charakteryzował się niekorzystnymi wartościami kryterium informacyjnego Akaikego (wysokimi wartościami) oraz wartością bezwzględną dla logarytmu funkcji wiarygodności. Natomiast, średnie poprawione dla odmian w poszczególnych miejscowościach otrzymane z modelu z strukturą typu UN charakteryzowały się znacznie większą zmiennością zakresu plonu odmian w poszczególnych miejscowościach. Zakres wartości średnich plonu odmian w poszczególnych miejscowościach kształtował się od 0,69 do 2,44 t*ha⁻¹. Takie zróżnicowanie zakresu zmienności plonu w poszczególnych miejscowościach jest zgodne z rzeczywistością, gdzie poszczególne miejscowości (środowiska) charakteryzują się nie jednakową mocą dyskryminującą odmiany. Jednak ze względu na swoje wady tego typu struktura nie jest zalecana do stosowania w trakcie oceny odmian. Podstawowym problemem w jej zastosowaniu jest duża złożoność obliczeniowa, co utrudnia jej rutynowe zastosowanie w ocenie adaptacji odmian na warunki środowiska. Dlatego dodatkowo badano modele z strukturą typu FA z różną liczbą składowych. Dotychczas w literaturze był prezentowany pogląd, że użycie 2 pierwszych składowych jest wystarczające i optymalne (Stefanova i Buirchell 2010; Bergueno i in. 2011). Jednakże, te badania jak wspomniano wcześniej były przeprowadzone dla bardzo małych zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń. Jak wynika z badań przedstawionych w Publikacji nr 4 model z strukturą FA z dwiema pierwszymi

składowymi – FA(2) pomimo posiadania najniższych wartości dla kryterium informacyjnego Akaikego nie najlepiej odzwierciedlały zmienność odmian w miejscowościach. Analizując średnie poprawione otrzymane dla modelu ze strukturą FA(2) obserwujemy bardzo małą zmienność plonu odmian pszenicy ozimej w niektórych miejscowościach. Zakres średnich poprawionych plonu odmian w tych miejscowościach był często mniejszy niż $0,4 \text{ t} \cdot \text{ha}^{-1}$. Podobnie jak dla struktury CS mogło by prowadzić do błędnych wniosków. W tym przypadku błędnie sugeruje, że niektóre miejscowości nie różnicują plonu badanych odmian – odmiany w nich praktycznie się nie różnią. Okazuje się że taka liczba składowych dla modelu z strukturą FA jest jeszcze nie wystraszająca, aby w pełni opisać zmienność plonu odmian w badanych miejscowościach. Opierając się na badaniu średnich poprawionych oraz wykresie reakcji (wizualizacji) odmian na warunki środowiska, najmniejsza optymalna liczba składowych wynosi 4. W związku z tym za najlepszy model można byłoby uznać model ze strukturą wariancji i kowariancji FA(4). Wykres reakcji odmian na środowiska jest w tym przypadku najbardziej podobny do wykresu reakcji otrzymanej za pomocą modelu z macierzą UN. Zwiększenie liczby składowych nie wpłynęło znacząco na zmianę opisu reakcji odmian na miejscowości. Zaobserwowano duży stopień zgodności rankingu ze strukturą US (oceniona za pomocą korelacji rang Spearmana) efektów głównych dla odmian oraz efektu odmian w miejscowościach dla modeli z macierzą FA(4) oraz strukturami FA z większą od 4 liczbą składowych. Natomiast, słaby stopień zgodności rankingu tych efektów obserwujemy dla modeli z strukturą wariancji i kowariancji CS oraz FA z mniejszą niż 4 liczbą składowych - FA(1), FA(2), FA(3). Wyniki przedstawione w Publikacji nr 4 wskazują że mniejsza liczba składowych niż optymalna dla struktury wariancji i kowariancji typu FA może prowadzić do błędnych wniosków dotyczących oceny adaptacyjności odmian i utrudnia przez to ich wiarygodną rekomendację. Zastosowanie kryterium informacyjnego nie zawsze pozwoli na wybór najlepszej struktury wariancji i kowariancji w liniowych modelach mieszanych. Zastosowanie oceny średnich poprawionych oraz wykresów reakcji odmian na środowisko może być prostym i efektywnym narzędziem w określeniu optymalnej liczby składowych dla struktury typu FA stosowanego w liniowych modelach mieszanych wykorzystywanych do analizy danych pochodzących z wielośrodowiskowych serii doświadczeń. Ocena struktur wariancji i kowariancji zwłaszcza struktury typu FA przeprowadzona w Publikacji nr 4 pozwoliła na poznanie ich właściwości dla stosunkowo dużych wielośrodowiskowych

serii doświadczeń, które dodatkowo charakteryzują się znacznym stopniem niekompletności danych.

Drugi sposób oceny liniowych modeli mieszanych z różnymi strukturami wariancji i kowariancji polega na ocenie dokładności oszacowania efektów (ang. prediction accuracy) w modelu. Przeprowadza się ją z wykorzystaniem różnych metod, ale powszechnie do oceny modeli (nie tylko w przypadku liniowych modeli mieszanych) stosuje się metody walidacji krzyżowej (ang. cross validation). W tym podejściu zbiór danych dzieli się na dwa zbiory – uczący i walidacyjny (sprawdzający). Na podstawie zbioru uczącego oszacowuje się parametry w modelu (przeprowadza się estymację modelu), a zbioru walidacyjnego sprawdza się precyzję oszacowania z wykorzystaniem różnych statystyk. W Publikacji nr 1 przeprowadzono badanie precyzji oszacowania liniowych modeli mieszanych z różnymi strukturami wariancji i kowariancji, oceniając dokładność oszacowania efektów w modelu z zastosowaniem walidacji krzyżowej.

Celem Publikacji nr 1 była ocena precyzji oszacowania dla dwóch jednoetapowych liniowych modeli mieszanych z różnymi strukturami wariancji i kowariancji. Służyły one do wyznaczenia średnich poprawionych dla kombinacji odmiany x miejscowości (pierwszy model) i odmian x regiony (drugi model) oraz badanie zgodności rankingu odmian w tych modelach. Powszechnym podejściem do oceny interakcji genotypowo - środowiskowej jest rozpatrywanie środowisk jako pojedynczych miejscowości lub kombinacji lat z miejscowościami. Jednak z punktu widzenia rekomendacji bardziej właściwa jest ocena odmian roślin uprawnych w regionach ich docelowej uprawy (ang. target region). Rolnik który jest odbiorcą rekomendacji odmian prędzej identyfikuje się z regionem, a nie konkretną miejscowością w której przeprowadzono doświadczenia. Dlatego też Polski system rekomendacji odmian roślin uprawnych oparty jest na podziale administracyjnym i rekomendację przeprowadza się w poszczególnych województwach.

Dotychczas prace nad oceną efektywności i precyzji oszacowania w liniowych modelach mieszanych były skupione na modelach rozpatrujących oddzielnie miejscowości, w których przeprowadzono doświadczenia polowe. W Publikacji nr 1 uzupełniono te badania o model z różnymi strukturami wariancji i kowariancji uwzględniający regiony, w których zagnieżdżone są pojedyncze miejscowości. W pracy zaproponowano dwa liniowe modele mieszane. Pierwszy z nich uwzględniający same

miejsowości w których przeprowadzano doświadczenia polowe, posłużył do oszacowania średnich poprawionych w klasyfikacji odmiana x miejscowość. Drugi zaprezentowany model uwzględnia pogrupowanie miejscowości w regiony i posłużył do oszacowania średnich poprawionych w kombinacji odmiana x region. Obydwa modele zastosowano z dwoma typami struktur wariancji i kowariancji – macierzami typu CS i FA. Struktura FA była rozpatrywana z różną liczbą składowych od jednej składowej FA(1) do pięciu składowych FA(5). Zaproponowane modele z różnymi typami struktur oceniono z wykorzystaniem walidacji krzyżowej. Tradycyjnie podział na zbiór uczący i zbiór walidujący dla obserwacji pochodzących z wielośrodowiskowych serii doświadczeń przeprowadzony jest w oparciu na powtórzeniach (blokach). Losowo całe powtórzenia (bloki) przydzielane są do zbioru uczącego albo do zbioru walidującego (Piepho 1998). Niestety, w przypadku doświadczeń przeprowadzonych w serii Porejstrowego Doświadczalnictwa Odmianowego nie można zastosować tego podejścia. Wynika to z faktu, że dla zbóż standardowo doświadczenia polowe przeprowadzana są w dwóch powtórzeniach (blokach). W pracy zaproponowano inny sposób, czyli podział obserwacji na zbiór uczący i walidujący oraz odpowiednią modyfikację parametru średniego kwadratu błędu oszacowania MSEP (ang. mean squared error of prediction), który posłuży do oceny różnych struktur wariancji i kowariancji dla zaproponowanych modeli. Ocenę precyzji oszacowania dla zaproponowanych modeli przeprowadzono z wykorzystaniem zbioru danych pochodzącego z Porejstrowego Doświadczalnictwa Odmianowego. Obserwację dla plonu 47 odmian pszenicy jarej pochodziły z 42 miejscowości pogrupowanych w 6 regionów prezentujących strefy agro-ekologiczne Polski. Zbiór danych pochodził z 3 sezonów wegetacyjnych (2010/2011 – 2012/2013). Kompletność tego zbioru wynosiła 43% obserwacji w stosunku do pełnej klasyfikacji trójkierunkowej - odmian, miejscowości i sezonów wegetacyjnych.

Dla zaproponowanego jednoetapowego liniowego modelu mieszanego wykorzystywanego do oszacowania średnich poprawionych dla kombinacji odmiana x miejscowość struktura typu FA z dwiema pierwszymi składowymi FA(2) posiadała najniższe średnie wartości parametru MSEP. Spośród rozpatrywanych struktur wariancji i kowariancji dla tego modelu, FA charakteryzowała się największą predykcją oszacowania. Ponadto, struktura FA(2) odznaczała się małą zmiennością wartości parametru MSEP, posiadała najniższe wartości dla odchylenia standardowego

i współczynnika zmienności. Zupełnie odmienne właściwości zaobserwowano dla modelu, w którym uwzględniono podział miejscowości na regiony i zastosowanego do oszacowania średnich poprawionych dla kombinacji odmiana x region. Dla tego modelu wartości średnie MSEP pomiędzy rozpatrywanymi strukturami wariancji i kowariancji były mało zróżnicowane. Jednakże, najniższe wartości zaobserwowano dla najprostszej struktury typu CS.

Dobór właściwej struktury wariancji i kowariancji dla liniowych modeli mieszanych zależy od celu przeprowadzanej analizy zbiorów danych pochodzących z wielośrodowiskowej serii doświadczeń. Jeżeli rozpatrujemy miejscowości pogrupowane w regiony, a ocenę adaptacji i stabilności odmian chcemy przeprowadzić w regionach w takim wypadku nie ma potrzeby stosowania bardzo elastycznych struktur. Jeżeli regiony są dobrze wyznaczone to pomiędzy nimi nie powinny występować silne korelacje. Czyli możliwie prosta struktura uwzględniająca homogeniczne wariancje i kowariancji dla plonu odmian pomiędzy regionami w wystarczający sposób opisuje zmienność plonu pomiędzy odmianami. Natomiast, jeżeli ocenę adaptacji i stabilności odmian przeprowadza się w odniesieniu do miejscowości należy zastosować elastyczną strukturę wariancji i kowariancji, która zakłada, że są one heterogeniczne. Pomędzy pojedynczymi miejscowościami naturalnie będą występować pomiędzy nimi zróżnicowane korelacje.

Obydwie przedstawione powyżej prace poszerzają wiedzę na temat właściwości różnych typów struktur wariancji i kowariancji dla liniowych modeli mieszanych wykorzystywanych do analizy zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń w tym i struktury typu FA. A zwłaszcza w zakresie ich zastosowań dla bardzo rozbudowanych zbiorów danych charakteryzujących się dużym stopniem niekompletności oraz ich wykorzystania w trakcie oceny adaptacyjności odmian w regionach (grupach miejscowości).

Opracowanie modeli wykorzystanych do oceny adaptacyjności odmian

Dotychczas powszechnie liniowe modele mieszane wykorzystywano do analizy zbiorów danych dla wielośrodowiskowych serii doświadczeń co najwyżej dla klasyfikacji dwukierunkowej odmiany \times środowiska. Bardzo często jako pojedyncze środowisko rozpatrywano kombinacje roku (sezonu wegetacyjnego) i miejscowości w której przeprowadzono doświadczenie. Taki zabieg pozwala na uproszczenie modelu, ale niestety ogranicza zakres i możliwości wnioskowania o adaptacji odmian roślin uprawnych na warunki środowiska

Celem niektórych wielośrodowiskowych serii doświadczeń jest ocena odmian na warunki w środowiskach przy różnych poziomach dodatkowego czynnika. Takim dodatkowym czynnikiem może być zastosowana agrotechnika, zastosowanie stresów abiotycznych (np. stresu suszy) i biotycznych (np. szkodniki, choroby). Przykładem takiego działania jest zastosowanie w doświadczeniach w ramach Porejstrowego Doświadczalnictwa Odmianowego dwóch zróżnicowanych poziomów agrotechniki dla niektórych gatunków roślin uprawnych między innymi zbóż. Brakuje jednak w literaturze propozycji liniowych modeli mieszanych dla wielośrodowiskowych serii doświadczeń uwzględniających dodatkowy czynnik. Często analizę z wykorzystaniem liniowych modeli mieszanych dla zbiorów danych z uzupełniającym czynnikiem przeprowadza się oddzielnie dla każdego z poziomów tego czynnika (Laidig i in. 2014). Nie uwzględniając efektu tego czynnika w modelu. Oczywiście ma to wpływ na ograniczenie zakresu możliwości wnioskowania z tego typu liniowego modelu mieszane, jednak upraszcza to model. Wielu badaczy i hodowców oczekuje aby tego typu dodatkowy czynnik był uwzględniony w modelu wykorzystywanym do analizy danych pochodzących z wielośrodowiskowych serii doświadczeń odmianowych. Wówczas można byłoby wiarygodnie wnioskować o wpływie tego czynnika i jego interakcji z odmianami czy środowiskiem. A także wpływie tego dodatkowego czynnika na zróżnicowanie reakcji odmian na warunki środowiskowe.

W Publikacji nr 6 zaproponowano dwuetapowe podejście do analizy danych dla czterokierunkowej klasyfikacji odmiany \times sezon wegetacyjny (rok) \times miejscowość \times poziom agrotechniki pochodzących z wielośrodowiskowej serii doświadczeń. W pracy zaproponowano dwa liniowe modele mieszane w drugim etapie analizy, wykorzystywane

do analizy danych pochodzących z Porejestrowego Doświadczalnictwa Odmianowego oraz zaprezentowano możliwości i sposób wnioskowania z zastosowanego podejścia do analizy statystycznej. Wykorzystano w tym celu zbiór danych obserwacji plonu dla 51 odmian pszenicy ozimej ocenionej w 18 miejscowościach w ciągu 7 sezonów wegetacyjnych (od 2004/2005 do 2010/2011) na dwóch poziomach agrotechniki (umiarkowanym i intensywnym). Do analizy danych sklasyfikowanych czterokierunkowo zaproponowano dwuetapowe podejście. W pierwszym kroku analizy przeprowadzono analizę pojedynczych doświadczeń według modelu dla układu split-block, w którym wyznaczono średnie poprawione. W kolejnym kroku analizy zastosowano łączną analizę z wykorzystaniem liniowych modeli mieszanych. Na tym etapie analizy zaproponowano dwa liniowe modele mieszane. Pierwszy model posłużył do oceny istotności poszczególnych efektów głównych i interakcyjnych. Wielu badaczy i hodowców w pierwszych krokach swoich działań w trakcie oceny odmian interesuje się, czy poszczególne efekty (główne oraz interakcyjne) miały istotny wpływ na kształtowanie się wartości rozpatrywanej zmiennej (np. plonu). Model stanowił typową klasyfikację krzyżową dla 4 czynników (sezon wegetacyjny, miejscowość, odmiana, poziom agrotechniki) z założeniem standardowej struktury wariancji i kowariancji typu VC. Dla efektów stałych do oceny ich istotności zastosowano test F Walda, natomiast dla efektów losowych test oparty na ilorazie wiarygodności (ang. likelihood-ratio test). Użycie tego liniowego modelu mieszanego pozwoliło na ocenę znaczenia i stopnia wpływu na plon zarówno stałych jak i losowych efektów w modelu. Drugi model miał na celu oszacowanie średnich poprawionych dla kombinacji odmiana x miejscowość x poziom agrotechniki poprzez wszystkie sezony wegetacyjne. W tym modelu zaproponowano efekt zagnieżdżony odmian w miejscowościach którego wariancje i kowariancje były modelowane z zastosowaniem struktury typu FA z 6 pierwszymi składowymi. Dzięki temu można było uwzględnić zróżnicowane wariancje plonu odmian w miejscowościach oraz kowariancje plonu odmian w wszystkich parach środowisk. Średnie poprawione w tym modelu otrzymano z wykorzystaniem procedury opisanej przez Welham i in. 2004, tak otrzymane średnie poprawione pozwoliły na merytoryczną ocenę adaptacji ocenian pszenicy ozimej przy dwóch poziomach agrotechniki. Zastosowanie modelu z zagnieżdżonym losowym efektem odmian i zamodelowaniem ich z wykorzystaniem struktury typu FA w poszczególnych miejscowościach pozwolił na wykazanie odmian charakteryzujących się szerokim stopniem adaptacji do warunków środowiska dla obydwu poziomów agrotechniki. Ten cel został osiągnięty poprzez

zastosowanie analizy skupień dla średnich poprawionych uzyskanych w liniowym modelu mieszanym z strukturą wariancji i kowariancji typu FA do grupowania odmian o podobnych uwarunkowaniu adaptacji (ang. adaptive response pattern) na warunki środowiska.

Obydwa zaproponowane w Publikacji nr 6 liniowe modele mieszane w drugim etapie analizy statystycznej pozwoliły na kompleksową ocenę adaptacyjności odmian na warunki środowiskowe przy dwóch różnych poziomach agrotechniki dla niekompletnego zbioru danych sklasyfikowanego czterokierunkowo. Umożliwiły one ocenę zmienności plonu badanych 51 odmian pszenicy ozimej, wykazanie odmian charakteryzujących się szerokim stopniem adaptacji do warunków klimatycznych Polski przy dwóch poziomach agrotechniki. Dotychczas analiza danych dla tak obszernego zbioru zawierającego dużą liczbę odmian ocenianych w możliwie szerokim zakresie środowisk (znaczna liczba miejscowości i sezonów wegetacyjnych) dla obydwu poziomów agrotechniki nie była przeprowadzana. Natomiast, wnioski merytoryczne z przeprowadzonej oceny odmian pszenicy ozimej pozwalają na efektywną ich rekomendację do uprawy.

Elastyczne modelowanie wariancji odmian w miejscowościach oraz kowariancji dla par miejscowości zaprezentowane w Publikacji nr 6, może nie być wystraszające do wyczerpującej oceny odmian przy różnych poziomach agrotechniki. Wynika to z faktu, że, adaptacyjność odmian na warunki środowiska jest zależna od poziomu agrotechniki na jakim były one uprawiane. To jest powód aby uwzględnić w trakcie analiz z wykorzystaniem liniowych modeli mieszanych, fakt, że zmienność odmian w poszczególnych miejscowościach, a także korelacje plonu odmian pomiędzy miejscowościami mogą być różne przy zastosowanych poziomach agrotechniki. Takie rozwiązanie uwzględniono w liniowym modelu mieszanym zaproponowanym w Publikacji nr 4 i zastosowanym do analizy z wielośrodowiskowej serii doświadczeń. W tej pracy, podobnie jak w Publikacji nr 6, do analizy zbioru danych pochodzących z Porejestrowego Doświadczalnictwa Odmianowego zastosowano dwuetapowe podejście. W drugim etapie analizy zaproponowano dwa modele. Pierwszy model posłużył do oszacowania komponentów wariacyjnych, które użyto do oceny istotności i siły (stopnia) wpływu poszczególnych efektów głównych i interakcyjnych. Drugi model posłużył do oszacowania średnich poprawionych dla cechy, które zostały wykorzystane do oceny związków pomiędzy rozpatrywanymi zmiennymi (zastosowano współczynnik korelacji Pearsona i analizę składowych głównych PCA). Ten model zawierał efekt odmian

zagnieżdżony w miejscowościach oraz w poziomach agrotechniki, dla którego zastosowano strukturę wariancji i kowariancji typu FA z trzema pierwszymi składowymi. Zastosowanie podwójnie zagnieżdżonego efektu odmian (w miejscowościach i poziomach agrotechniki) oraz elastycznej struktury wariancji i kowariancji zezwoliło na uwzględnienie zróżnicowanych wariancji odmian w miejscowości i w dwóch poziomach agrotechniki. Pozwoliło to na jeszcze bliższe rzeczywistości odwzorowanie reakcji odmian roślin uprawnych na warunki w miejscowościach i ich reakcji na zastosowane poziomy agrotechniki.

Dzięki, powyżej zaproponowanej metodyce w Publikacji nr 4 przeprowadzono analizę zbioru danych dla 15 odmian pszenicy jarej, którą oceniano w 7 miejscowościach przez 2 sezony wegetacyjne (2013/2014 i 2014/2015). Zbiór danych obejmował obserwację dla plonu, składowych plonu (liczba kłosów, masa 1000 ziaren oraz liczba ziaren w kłosie) oraz cech jakości ziarna (zawartość białka, zawartość skrobi, zawartość glutenu, wartość testu sedymentacji oraz gęstość ziarna). Powszechnie liniowe modele mieszane stosowane były do tej pory dla plonu roślin, natomiast brak jest prac w których wykorzystano by te podejścia dla innych cech obserwowanych takich: jak cechy składowe plonu czy jego jakości. Dodatkowo w Publikacji nr 4 miejscowości, w których przeprowadzono doświadczenia podzielono na dwie grupy środowisk. Pierwsza, obejmowała miejscowości, które charakteryzowały się korzystnymi warunkami glebowymi do uprawy pszenicy. Natomiast, druga grupa miejscowości charakteryzowała się niekorzystnymi warunkami glebowymi. Analizę danych z zaproponowanymi liniowymi modelami mieszanymi przeprowadzono oddzielnie dla dwóch grup środowisk. Pozwoliło to na ocenę zróżnicowania wpływu oraz znaczenia odmian i poziomów agrotechniki w kształtowaniu się cech jakości ziarna pszenicy jarej. Okazuje się, że w grupie środowisk o niekorzystnych warunkach do uprawy pszenicy znaczenie odmian i poziomu agrotechniki w kształtowaniu zmienności plonu, jego składowych oraz cech jakości ziarna jest mniejszy niż w grupie środowisk o korzystnych warunkach glebowych do uprawy.

Jak już wspomniano w trakcie prezentacji Publikacji nr 1 rekomendacja odmian powinna być przeprowadzona dla regionu docelowego uprawy. W systemie Porejstrowego Doświadczalnictwa Odmianowego rekomendację przeprowadza się w obrębie poszczególnych województw. Jednak w niektórych województwach warunki klimatyczne i glebowe są na tyle zbliżone, że te same odmiany są w nich rekomendowane.

Województwa, w których ranking plonu odmian jest zbliżony mogą tworzyć pojedyncze mega-środowisko (ang. megaenvironment). W celu poprawy efektywności i skuteczności rekomendacji odmian do uprawy uzasadnione byłoby tworzenie tego typu mega-środowisk (grupy województw). Ocenę zgodności rankingu odmian w poszczególnych województwach oraz próba wyznaczenia mega-środowisk było celem Publikacji nr 3. W tej pracy wykorzystano zbiór danych plonu dla 50 odmian pszenicy ozimej ocenionych w 49 miejscowościach przez 3 sezony wegetacyjne (2010/2011-2012/2013). Miejscowości w których przeprowadzono doświadczenia polowe zostały uporządkowane w 16 województwach. Liczba miejscowości w poszczególnych województwach kształtowała się od dwóch do pięciu. W pracy wykorzystano tylko obserwację dla plonu pochodzące z intensywnego poziomu agrotechniki (A2), który umożliwia w pełni ujawnić potencjał plonowania rozpatrywanych odmian pszenicy ozimej. Do realizacji tego celu zastosowano dwuetapowe podejście do analizy zbiorów danych pochodzących z wielośrodowiskowej serii doświadczeń. Pierwszy etap analizy był podobny do tych, które były prezentowane w powyżej opisanych pracach. W drugim etapie zastosowano dwa liniowe modele mieszane, które były stosowane w różnych celach. Pierwszy model w którym rozpatrywano efekt główny regionów (województw), miejscowości zagnieżdżonych w regionach oraz pozostałe efekty główne (odmiana oraz rok) i interakcyjne w klasyfikacji krzyżowej. Model ten posłużył do oceny istotności i stopnia wpływu na plon rozpatrywanych efektów. Drugi model zawierał efekt zagnieżdżenia odmian w regionach (województwach), w których zagnieżdżone były miejscowości. Dla tego efektu zastosowano strukturę wariancji i kowariancji typu FA z dwoma pierwszymi składowymi. Na podstawie tego modelu wyznaczono średnie poprawione dla kombinacji odmiana x region które posłużyły do badania zgodności rankingu odmian pomiędzy regionami. Zgodność rankingu w regionach przeprowadzono z wykorzystaniem współczynników korelacji rang Spearmana, a podobieństwo pomiędzy regionami oceniono z wykorzystaniem metody Warda analizy skupień.

Zastosowane podejścia do analizy zaproponowane w Publikacji nr 3 z zastosowaniem liniowych modeli mieszanych pozwoliło zaobserwować istnienie grup województw, które charakteryzują się zgodnym (podobnym) rankingiem plonu odmian pszenicy ozimej. Dzięki użyciu analizy skupień zostały one połączone w mega-środowiska. Niestety, grupy te nie tworzą pewnego rodzaju region geograficzny, ponieważ nie łączą sąsiednie województwa w mega-środowiska. Jednak utworzone grupy

województw charakteryzują się zbliżonymi warunkami klimatycznymi oraz podobną charakterystyką warunków glebowych. Zaobserwowano również, że niektóre województwa miały odmienny ranking odmian od pozostałych województw czy ich grup. Przykładem jest województwo Opolskie, silne ujemne korelacje rang Spearmana. Zaproponowana metodyka oparta na liniowych modelach mieszanych może stanowić dobre narzędzie do wyznaczania mega-środowisk opartych na zbiorach danych z wielośrodowiskowych serii doświadczeń czy poszukiwanie zgodności rankingu plonu odmian. Utworzone w ten sposób mega-środowiska pozwolą na zwiększenie efektywności rekomendacji odmian do uprawy na terenie Polski.

Podsumowując, ocena właściwości różnego typu struktur wariancji i kowariancji, opracowane liniowe modele mieszane oraz przedstawione sposoby wnioskowania zaprezentowane w pracach stanowiących niniejsze osiągnięcie naukowe mogą zostać wykorzystane do poprawy efektywności oceny interakcji GE. Ponadto służyć mogą do usprawnienia badania adaptacji odmian i skuteczności rekomendacji odmian ze zbiorów danych pochodzących z wielośrodowiskowych serii doświadczeń. Zaprezentowana metodyka statystyczna przyczynia się także do prób wyjaśnienia powodów występowania luki (ang. yield gaps) pomiędzy plonem potencjalnym, a tym otrzymywanym przez rolników (van Ittersum i in. 2013). Ocena adaptacyjności odmian na warunki środowiskowe oraz poznanie przyczyn luki w plonie przyczyni się do intensywniejszego postępu hodowlanego i genetycznego dla roślin uprawnych, a w konsekwencji do wzrostu plonu z jednostki powierzchni. Wzrost plonu ważnych użytkowo roślin uprawnych jest ważnym problemem ze względu na spodziewany wzrost populacji i zużycia biopaliw na świecie. Ray i in. 2013 uważają że produkcja rolnicza a zwłaszcza roślinna powinna zostać podwojona do 2050 roku.

Spis publikacji

- Annicchiarico P., Chiapparino E., Perenzin M. 2010. Response of common wheat varieties to organic and conventional production systems across Italian locations, and implications for selection. *Field Crops Research* 116: 230–238.
- Burgueño J., Crossa J., Cotes J.M., Vicente F.S., Das B. 2011. Prediction assessment of linear mixed models for multienvironment trials. *Crop Science* 51: 944-954.

- Crossa J., Vargas M., Joshi A.K. 2010. Linear, bilinear, and linear-bilinear fixed and mixed models for analyzing genotype x environment interaction in plant breeding and agronomy. *Canadian Journal of Plant Science* 90: 561-574.
- Curti R.N., de la Vega A.J., Andrade A.J., Bramardi S.J., Bertero H.D. 2014. Multi-environmental evaluation for grain yield and its physiological determinants of quinoa genotypes across Northwest Argentina. *Field Crops Research* 166: 46-57.
- Hu X., Spilke J. 2011. Variance-covariance structure and its influence on variety assessment in regional crop trials. *Field Crops Research* 120: 1-8.
- Kelly A.M., Smith A.B., Eccleston J.A., Cullis B.R., 2007. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. *Crop Science* 47: 1063-1070.
- Laidig F., Piepho H.P., Drobek T., Meyer U. 2014. Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics* 127: 2599-2617.
- Liu S.M., Constable G.A., Reid P.E., Stiller W.N., Cullis B.R. 2013. The interaction between breeding and crop management in improved cotton yield. *Field Crop Research* 148: 49-60
- Mađry W., Derejko A., Studnicki M., Paderewski J., Gacek E. 2017. Response of winter wheat cultivars to crop management and environment in post-registration trials. *Czech Journal of Genetics and Plant Breeding* 53: 47-52.
- Meyer K. 2009. Factor-analytic models for genotype \times environment type problems and structured covariance matrices. *Genetics Selection Evolution* 41: 21.
- Meyer K., Kirkpatrick M. 2008. Perils of parsimony: Properties of reduced-rank estimates of genetic covariance matrices. *Genetics* 180:1153-1160.
- Mohring J., Piepho H.P. 2009. Comparison of weighting in two-stage analyses of series of experiments. *Crop Science* 49: 1977-1988.
- Patterson H.D., Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554.
- Piepho H.P., Möhring J., Schulz-Streeck T., Ogutu J.O. 2012. A stage-wise approach for analysis of multi-environment trials. *Biometrical Journal* 54: 844-860.
- Piepho H.P. 1998. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. *Theoretical and Applied Genetics* 97: 195-201.

- Piepho H.P., Möhring J., Schulz-Streeck T., Ogutu J.O. 2012. A stage-wise approach for analysis of multi-environment trials. *Biometrical Journal* 54: 844–860.
- Ray D.K., Mueller N.D., West P.C., Foley J.A. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* 8: e66428.
- Searle S.R., Casella G., McCulloch C.E. 1992. *Variance components*. Wiley, New York.
- So Y., Edwards J. 2009. A comparison of mixed-model analyses of the Iowa Crop Performance Test for corn. *Crop Science* 49: 1593–1601.
- Stefanova K.T., Buirchell B. 2010. Multiplicative mixed models for genetic gain assessment in lupin breeding. *Crop Science* 50: 880-891.
- van Ittersum M.K., Cassman K.G., Grassini P., Wolf J., Tittonell P., Hochman Z. 2013. Yield gap analysis with local to global relevance—A review. *Field Crops Research* 143: 4-17.
- Welham S.J., Cullis B.R., Gogel B.J., Gilmour A.R., Thompson R. 2004. Prediction in linear mixed models. *Australian and New Zealand Journal of Statistics* 46: 325-347.
- Smith A.B., Cullis B.R.; Thompson R. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science* 143: 1–14.
- Mądry W., Iwańska M. 2011. Przydatność metod i miar statystycznych do oceny stabilności i adaptacyjności odmian: przegląd literatury. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin* 260/261: 193-218.
- Robinson G.K. 1991. That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science* 6: 15–32.
- Yang R.C. 2010. Towards understanding and use of mixed-model analysis of agricultural experiments. *Canadian Journal of Plant Science* 90: 605-627.

5. Omówienie pozostałych osiągnięć naukowo – badawczych

Dotychczas moje zainteresowania naukowo-badawcze poza omówionym wyżej osiągnięciem naukowym koncentrowały się na następujących tematach i zagadnieniach:

- a) Ocena adaptacyjności oraz stabilności odmian roślin uprawnych z wykorzystaniem innych metod statystycznych niż opartych na liniowych modelach mieszanych

Zagadnienie to miało na celu ocenę adaptacyjności oraz stabilności rodów hodowlanych dla pszenicy ozimej i realizowano je w współpracy z polskimi spółkami hodowli roślin, między innymi z Małopolską Hodowlą Roślin oddział w Kobierzycach, Poznańską Hodowlą Roślin, Instytutem Hodowli i Aklimatyzacji Roślin PIB w Radzikowie, a także w współpracy z kolegami z Katedry Doświadczalnictwa i Bioinformatyki SGGW. W ramach tego zagadnienia oceniono również rody i odmiany roślin ogrodniczych (agrest, porzeczka czarna) w współpracy z Instytutem Ogrodnictwa w Skierniewicach. Ze względu na to że do realizacji powyżej zaprezentowanych celów zastosowano kompletne zbiory danych, pochodzących z wielośrodowiskowych serii doświadczeń dla plonu jak i cech jego jakości, można było zastosować klasyczne podejście do oceny adaptacyjności i stabilności odmian. W celu realizacji powyższego zagadnienia i tematyki zastosowano między innymi klasyczną analizę wariancji z efektami losowymi, stały model AMMI (ang. additive main-effects and multiplicative interaction), oszacowanie średnich poprawionych z wykorzystaniem metody najmniejszych kwadratów (LSMeans), a także ocena plonu nominalnego. Metody pozwoliły wykazać rody które charakteryzują się szeroką adaptacją na warunki środowiska zarówno dla plonu jak i cech jego jakości, przez co zwiększają szansę na komercyjny sukces nowo wytworzonej odmiany. Badania przeprowadzone w ramach tego zagadnienia zostały opublikowane w czasopismach z listy A MNISW (Załącznik nr 4 II.A.3, II.A.15) oraz w czasopismach z listy B MNISW (Załącznik nr 4 II.D.1, II.D.3, II.D.6, II.D.9, II.D.13, II.D.14).

- b) Ocena zmienności wielocechowej oraz tworzenie kolekcji podstawowych dla kolekcji zasobów genowych

Zagadnienie to jest kontynuacją i rozwinięciem prac podjętych w trakcie przygotowywania pracy doktorskiej. W ramach tych prac przeprowadzono ocenę zmienności cech fenotypowych obiektów (ekotypów, linii, rodów, odmian) zgromadzonych w Polskich kolekcjach zasobów genowych roślin uprawnych (agrestu, kupkówki pospolitej, pszenicy ozimej i jarej, truskawki). W większości przypadków ocena zgromadzonych w kolekcjach zasobów genowych przeprowadzona jest tylko raz przed włączeniem obiektu do kolekcji. Z tego względu zbiory danych pochodzące z takich ocen obiektów, stanowią klasyfikację niekompletną. Dlatego też do analizy statystycznej danych wykorzystano liniowe modele mieszane z metodą REML do

oszacowania parametrów. Oszacowane w ten sposób średnie poprawione dla obiektów zostały wykorzystane do ich oceny z zastosowaniem metod wielocechowych, takich jak analiza skupień, analiza czynnikowa, analiza zmiennych kanonicznych czy analiza składowych głównych. Tak kompleksowa ocena zmienności w kolekcjach zasobów genowych pozwala na efektywniejsze ich wykorzystanie w hodowli, czy przy innego rodzaju pracach badawczych. W celu jeszcze bardziej efektywnego wykorzystanie zgromadzonych obiektów w kolekcjach roślin wyznacza się kolekcje podstawowe, które w możliwie małej liczbie obiektów zawierają pełny zakres zmienności obserwowanej w kolekcjach zasobów genowych. W ramach prac nad kolekcjami podstawowymi zasobów genowych zaproponowana i oceniono przydatność kilku metod selekcji obiektów do takich kolekcji. Zaproponowane metody były oparte na metodzie warstwowego pobierania prób zgodnej z metodyką pobierania próby reprezentatywnej (metoda reprezentatywna).

W ramach ty zagadnień powstało prace w czasopismach z listy A MNISW (Załącznik nr 4 II.A.4, II.A.14, II.A.16) i listy B MNISW (Załącznik nr 4 II.D.4, II.D.5, II.D.8, II.D.10, II.D.11, II.D.12, II.D.13, II.D.14). Zagadnienie to realizowano w współpracy z Instytutem Aklimatyzacji i Hodowli Roślin PIB w Radzikowie oraz Instytutem Ogrodnictwa w Skierniewicach.

W ostatnim czasie w ramach współpracy z Katedrą Hodowli Lasu Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie przeprowadzono ocenę wielocechowej zmienności rodów i populacji matecznych dla kolekcji zasobów genowych buka zwyczajnego i sosny zwyczajnej. Zaadaptowano w tym celu metodykę powszechnie stosowaną w oceny obiektów w kolekcjach roślin uprawnych w oparciu na uogólnionych modelach liniowych z rozkładem wielomianowym oraz metodami wielocechowymi (analizą skupień i analizą składowych głównych). W ramach tej współpracy i realizacji tego tematu badawczego powstały dwie publikacje w czasopismach z listy A MNISW (Załącznik nr 4 II.A.5, II.A.8)

- c) Badanie wpływu czynników genetycznych i agrotechnicznych w kształtowaniu się plonu, jego składowych oraz cech jakości ziarna

Zagadnienie to jest wynikiem realizacji wspólnie z Katedrą Agronomi Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie dwóch grantów. W ramach wspomnianych grantów oceniono wpływ efektów genotypowych (odmian),

środowiskowych (gleby, przebiegu pogody) oraz intensywności agrotechniki na plon oraz jego uwarunkowanie składowymi plonu i innymi cechami plonotwórczymi dla odmian pszenicy ozimej i jarej. Badania pozwoliły określić sposób kształtowania plonu nowoczesnych odmian (zarejestrowanych w ostatnim dziesięcioleciu w kraju i w Europie) pszenicy ozimej i jarej przez trzy składowe – liczba kłosów na jednostce powierzchni, liczba ziaren w kłosie i masa tysiąca ziaren. Wyznaczono grupy odmian charakteryzujące się różnym znaczeniem w uwarunkowaniu plonu poszczególnych składowych, co pozwala na dopasowanie sposobu uprawy aby możliwie zwiększyć plon i zysk z produkcji danej odmiany.

Przeprowadzono także badanie wpływu warunków glebowych i klimatycznych na znaczenie efektów genotypowych i poziomów agrotechniki w zmienności plonu i cech jakości. Zmienność plonu i inne cechy jakości w środowiskach o korzystnych warunkach do uprawy pszenicy była w znacznie większym stopniu kształtowana przez genotyp i agrotechnikę niż w środowiskach o niekorzystnych do uprawy pszenicy warunków glebowych i klimatycznych. W ramach tego zagadnienia przeprowadzono dwuczynnikowe (odmianowo-agrotechniczne) doświadczenia polowe w dwóch sezonach wegetacyjnych (pszenica ozima – 2008/2009, 2009/2010; pszenica jara – 2013,2014) w stacjach doświadczalnych Centralnego Ośrodka Badania Odmian Roślin Uprawnych. W trakcie analizy statystycznej zastosowano między innymi sekwencyjną analizę ścieżek opartą na fazach ontogenetycznych pszenicy, metod wielocechowych (analizy składowych głównych, analizy skupień metodą Warda) oraz liniowych modeli mieszanych z metodą REML do oszacowania parametrów w modelu i estymacji komponentów wariacyjnych. Mój udział w realizacji przestawianego zagadnienia i związanych z nim grantów naukowych polegał nie tylko na opracowania i przeprowadzeniu analizy statystycznej ale także udział w przeprowadzeniu doświadczeń i obserwacji polowych.

Wyniki badań zostały opublikowane w pracach w czasopismach z listy A MNiSW (Załącznik nr 4 II.A.11, II.A.12, II.A.15, II.A.18) oraz w czasopismach z listy B MNiSW (Załącznik nr 4 II.D.3, II.D.6).

d) Badania zachowania się owadów z wykorzystaniem uogólnionych modeli liniowych

Temat ten jest realizowany w współpracy z Samodzielnym Zakładem Entomologii Stosowanej SGGW i Zakładem Etologii Instytutu Biologii Doświadczalnej

im. M. Nenckiego PAN w Warszawie. Pierwsze z zadań badawczych obejmowała ocenę wpływu obecności innych osobników tego samego gatunku (konkurencja) na zachowania agresywne i rozrodcze (np. sposób składowania jaj). A drugie z zagadnień polegało na określeniu wpływu wieku samic oraz otoczenia na preferencję pokarmową osobników. W trakcie doświadczeń wykorzystywanych w tym celu obserwowano zmienne liczebnościowe (ang. count data), które nie posiadają rozkładu normalnego. A także inne rodzaju zmienne które tego rozkładu nie posiadają, na przykład czas trwania pewnego typu zachowania. Z tego względu klasyczne metody statystyczne nie mogą być w tym przypadku właściwie stosowane. Dla tego typu obserwacji i zbiorów danych coraz częściej stosuje się uogólnione modele liniowe (ang. generalized linear model, GLM) i uogólnione modele mieszane (ang. generalized linear mixed model, GLMM). Mój udział w powyższym temacie badawczym polegał na opracowaniu modeli typu GLM i GLMM zastosowanych do analizy zbiorów danych pochodzących z doświadczeń i obserwacji służących do realizacji tego zagadnienia. Także, moje zadanie polegało na doborze rozkładów prawdopodobieństwa dla poszczególnych zmiennych. Zwłaszcza dla zmiennych liczebnościowych gdzie ważnym zagadnieniem była ocena, badanie i eliminacja zjawiska over-dispersion, czyli zjawiska przeszacowania wariancji w proponowanym uogólnionym modelu liniowym. W ramach tego zagadnienia powstały publikacje w czasopiśmie z listy A MNISW (Załącznik nr 4 II.A.2 II.A.9, II.A.13, II.A.17)

e) Badania epidemiologiczne dotyczące śmiertelności z powodu choroby Alzheimera

Celem tych rozważań było wykazanie zależności pomiędzy dochodem a śmiertelnością na chorobę Alzheimera w latach od 1929 do 2005 w populacji mieszkańców Stanów Zjednoczonych Ameryki USA. Z wykorzystaniem metod statystycznych wykorzystywanych do badania szeregów czasowych wykazano istnienie takiej zależności, która charakteryzuje się pewnego rodzaju cyklami silniejszej zależności pomiędzy dochodem a śmiertelnością. Takich cykli w rozpatrywanym okresie było 4 i trwały około 20 lat. Także zaobserwowano tendencję w czasie zwiększenia się siły związku pomiędzy dochodem a śmiertelnością. Dochód mieszkańców jest silnie związany z stylem ich życia, a zwłaszcza z dietą i sposobem spędzania wolnego czasu. Na podstawie związku między dochodem a śmiertelnością opracowano również model który pozwolił opracować dietę, która minimalizuje śmiertelność powodowaną

rozpatrywanym schorzeniem. Zastosowano w tym celu liniowy model regresji, który oszacowano na podstawie historycznych danych na temat spożycia makro-składników (węglowodany, białka, tłuszcze) w diecie. Tak utworzony model pozwolił na dobór optymalnych proporcji oraz ilości spożycia makro-składników, które minimalizują śmiertelność z powodu choroby Alzheimera dla każdego okresu życia człowieka. Zagadnienie to realizowano w ramach współpracy z Instytutem Biologii Doświadczalnej im. M. Nenckiego PAN w Warszawie.

W ramach realizacji tego zagadnienia powstały dwie publikacje znajdujące się na liście A MNiSW (Załącznik nr 4 II.A.6, II.A.10). Pracę nad tym zagadnieniem są dalej kontynuowane, opracowywany jest model który optymalizuje spożycie różnych typów tłuszcz w diecie.

f) Podsumowanie dotychczasowego dorobku naukowego

Jestem współautorem 36 publikacji w czasopismach recenzowanych, z czego 23 prace zostały opublikowane w czasopismach posiadających współczynnik Impact Factor. Jestem także współautorem jednej monografii w języku polskim oraz 27 streszczeń opublikowanych w materiałach konferencyjnych. Zestawienie ilościowe prac w poszczególnych kategoriach przedstawiono w Tabeli 1.

Suma punktów dla wszystkich prac zgodnie z listą Ministerstwa Nauki i Szkolnictwa Wyższego zgodnej z rokiem wydania pracy wynosi 668 pkt, łączna wartość współczynnika Impact Factor zgodny z rokiem wydania pracy wynosił 31,210. Suma punktu dla prac nie objętych zaprezentowanym osiągnięciem naukowym wynosi 533 a łączna wartość współczynnika Impact Factor wynosi 25,221.

Liczba cytowań prac, których jestem współautorem według bazy Web of Science Core Collection wynosi 39 (33 bez autocytowań) natomiast wartość współczynnika Indeks Hirscha wynosi 3. Liczba cytowani według bazy Google Scholar wynosi 111, a wartość współczynnika Indeks Hirscha wynosi 6.

Tabela 1. Zestawienie ilościowe dotychczas opublikowanych prac.

	Przed uzyskaniem stopnia doktora	Po uzyskaniu stopnia doktora	Razem
Publikacje recenzowane w czasopismach posiadających współczynnik Impact Factor		23	23
- w tym w języku polskim		2	2
- w tym w języku angielskim		21	21
Publikacje recenzowane w czasopismach nie posiadających współczynnik Impact Factor	11	2	13
- w tym w języku polskim	7	2	9
- w tym w języku angielskim	4		4
Monografie	1		1
- w tym w języku polskim	1		1
- w tym w języku angielskim			
Streszczenia w materiałach konferencyjnych	7	20	27
- w tym w języku polskim	4	12	16
- w tym w języku angielskim	3	8	9

Warszawa, 24 stycznia 2018

