

**Załącznik nr 2a
do wniosku
o przeprowadzenie
postępowania habilitacyjnego**

Autoreferat

dr Jakub Paderewski

**Katedra Doświadczalnictwa i Bioinformatyki
Wydział Rolnictwa i Biologii
Szkola Główna Gospodarstwa Wiejskiego w Warszawie**

Warszawa 2018

Imię i nazwisko: Jakub Bartłomiej Paderewski

Posiadane dyplomy, stopnie naukowe:

2000. Magister matematyki, Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki. Temat pracy magisterskiej „Metoda optymalizacji genetycznej sieci neuronowych”. Promotor: dr Nguyen Hung Son.
2008. Doktorat w dyscyplinie agronomia, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, Wydział Rolnictwa i Biologii. „Przydatność Modelu AMMI do badania reakcji roślin rolniczych na warunki środowiskowe”. Promotor: prof. dr hab. Wiesław Mądry.

Informacje o poprzednim zatrudnieniu w jednostkach naukowych :

- 2001- Zatrudniony w Szkole Głównej Gospodarstwa Wiejskiego w Warszawie, Wydział Rolnictwa i Biologii, Katedra Statystyki Matematycznej i Doświadczalnictwa. Później Katedra została przemianowana na Katedrę Biometrii, obecnie, po reorganizacji nosi nazwę Katedry Doświadczalnictwa i Bioinformatyki.

Wskazane osiągnięcia wynikające z art. 16 ustawy 2

Osiągnięcie naukowe „Zastosowanie modelu AMMI do oceny reakcji odmian pszenicy ozimej na warunki środowiskowe w Polsce” składa się z cyklu pięciu publikacji naukowych. Przedstawione publikacje opisują zastosowanie modelu AMMI (ang. *additive main effects and multiplicative interactions*), wprowadzają jego modyfikacje oraz przedstawiają możliwości wnioskowania w zastosowaniu danych o plonowaniu pszenicy ozimej.

1. Yield response of winter wheat to agro-ecological conditions using Additive Main Effects and Multiplicative Interaction and Cluster Analysis, 2011; Jakub Paderewski, Hugh G. Gauch, Wiesław Mądry, Tadeusz Drzazga and Paulo C. Rodrigues. *Crop Science* (2011) 51:969–980. (IF₂₀₁₁ czasopisma 1,641, punktacja MNISW₂₀₁₁ 27)

Celem tego artykułu było przedstawienie **połączonej analizy składowych głównych modelu AMMI oraz analizy skupień dla** danych uzyskanych z pomiarów plonu ziarna **pszenicy ozimej** (*Triticum aestivum* L.) zebranych w **doświadczeniach przedrejestrach** przeprowadzonych przez wiele lat w niepełnej trzykierunkowej klasyfikacji genotyp \times lokalizacja \times rok (GLY) i wykazanie przydatności tej procedury do badania wzorców adaptacji odmian.

Rozpatrywane dane dla plonu ziarna pszenicy ozimej (*Triticum aestivum* L.) pochodziły z przedrejestrach doświadczeń odmianowych prowadzonych przez stacje hodowlane. W pierwszym etapie analizy statystycznej przeanalizowaliśmy dane osobno dla każdej kombinacji warunków środowiskowych (kombinacje lokalizacji i roku badań, GxL). Użyto programu SAS i procedury GLM, w której genotypy były traktowane jako czynnik stały, a powtórzenia i bloki jako czynniki losowe. W ten sposób obliczono najlepszy nieobciążony estymator liniowy (BLUE, *best linear unbiased estimator*) – średnie poprawione dla genotypów – oraz średni kwadrat dla błędów w każdych warunkach środowiskowych.

Tak uzyskane średnie poprawione plonu genotypów dla każdej stacji doświadczalnej i każdego roku osobno roku połączono w jedną trzy-kierunkową klasyfikację. W tej niekompletnej klasyfikacji dla kombinacji genotypy \times miejscowości \times lata (GxLxY), opisującej 21 genotypów, 15 lokalizacji i 14 lat, było 4410 podklas, z których tylko 951 miało określony średni plon (średnie poprawione plonu analizowanego modelem mieszanym analizy

wariancji w każdym roku i każdej stacji doświadczalnej osobno), co stanowiło zaledwie 21,6% wszystkich podklas.

W drugim etapie przeprowadziliśmy wieloczynnikową analizę średnich poprawionych plonu genotypów w niekompletnej, ale spójnej, trójkierunkowej klasyfikacji GxLxY według mieszanego modelu ANOVA. Genotypy i lokalizacje zostały uznane za stałe czynniki stałe, a lata jako czynnik losowy. Korzystając z metody REML oszacowaliśmy średnie poprawione dla plonu w klasyfikacji GxL (estymatory BLUE) – średnie kombinacji GxLxY poprzez lata. Dla dwukierunkowej klasyfikacji genotyp x miejscowość (kombinacje GxL z 315 podklasami) określono również liczbę lat badań dla każdej kombinacji GxL, która zmieniała się od 0 do 11, przy średniej równej 3.02. Klasyfikacja GxL została przeanalizowana za pomocą modelu AMMI przy użyciu połączonej analizy AMMI i analizy skupień.

Do grupowania genotypów o podobnych efektach interakcji została wykorzystana hierarchiczna analiza skupień metodą Warda (w której miara odmienności był kwadrat odległości euklidesowej). Dwadzieścia jeden genotypów zostało podzielone na trzy grupy. Genotypowe profile efektów interakcji GL zostały użyte do opisanie wzorców interakcji genotyp \times środowisko (GE) na plon ziarna. Aby podzielić średnie genotypów na odrębne jednorodnie podgrupy zastosowano metodę niehierarchicznej analizy skupień opracowaną przez Calińskiego i Corstena. Dzięki temu takie podgrupy genotypów byłyby homogeniczne w sensie średniej genotypowej oraz typu interakcji. W tym badaniu 11 genotypów zgrupowanych w trzech podgrupach wykazało szeroką adaptację. Spośród nich odmiana TONACJA została zidentyfikowana jako dobra odmiana kontrolna do doświadczeń przedrejestranych. Jedna odmiana miała wąską (specyficzną) adaptację, a dziewięć wykazało relatywnie słabą adaptację.

Podsumowując uzyskane wyniki badań dotyczących plonu ziarna genotypów pszenicy ozimej na zróżnicowane warunki środowiskowe w szerokim zakresie czasowo-przestrzennym zauważyliśmy, że wszystkie nowoczesne genotypy wyhodowane w Polsce w latach 2000-2004 oznaczono szeroką zdolnością adaptacyjną. Ta cecha agronomiczna nie została przypisana większości wcześniej uzyskanych genotypów, wyjątkiem były genotypy: KOBRA, KORWETA i DED315/89. Większość genotypów starszych i pośrednich wiekiem była stosunkowo słabo dostosowana do rozpatrywanych warunków środowiskowych, co zostało zademonstrowane przez ich względnie niską lub bardzo niską ocenę plonu ziarna we wszystkich stacjach doświadczalnych. Oznacza to, że współczesne polskie odmiany pszenicy ozimej są bardziej produktywnie niż stare, nie tylko na żyznych glebach (środowiska

o wysokim potencjale plonotwórczym), ale także w mniej korzystnych warunkach środowiskowych. Wyniki badania opisują postęp genetyczny w hodowli odmian pszenicy ozimej uzyskany w Polsce w trakcie okresu 23 lat (od 1982 do 2004) w średnim plonie na hektar uprawy oraz poprawę stopnia szerokiej zdolności adaptacyjnej nowych odmian, czyli stabilność produkcji w warunkach produkcyjnych.

Łączna analiza AMMI i analiza skupień była w tamtym czasie unikalna w porównaniu z innymi proponowanymi w literaturze procedurami opartymi na AMMI. To podejście do analizy AMMI adaptowanej do niepełnej trójczynnikowej klasyfikacji GxLxY dla plonu ziarna z wielośrodowiskowej serii przedrejestranych doświadczeń odmianowych przeprowadzanej przez 23 lata na pszenicy ozimej okazało się efektywne w opisie różnych wzorów reakcji genotypów na warunki agroekologiczne w stacjach doświadczalnych rozmieszczonych na większości polskich obszarów uprawy pszenicy ozimej. Opracowana metodyka statystyczna zastosowana w tej pracy była bardziej efektywna w przewidywaniu i wizualizacji różnych wzorców reakcji genotypów w porównaniu do tradycyjnej analizy AMMI, w której nie stosowane jest grupowanie genotypów. Połączona analiza AMMI i analiza skupień okazała się skuteczna do opisanego wyodrębnionych schematów reakcji plonu na warunki agroekologiczne w miejscowościach.

2. AMMI analysis of four-way Genotype × Location × Management × Year data from a wheat trial in Poland, Jakub Paderewski, Hugh G. Gauch, Wiesław Mądry and Edward Gacek. *Crop Science* (2016) 56:2157–2164. (IF₂₀₁₆ czasopisma 1,629, punktacja MNISW₂₀₁₆ 30, ujednolicona punktacja MNISW₂₀₁₃₋₂₀₁₆ 35)

Rozpatrywane (analizowane) dane dla plonu ziarna pszenicy ozimej (*Triticum aestivum* L.) pochodziły z porejestranych doświadczeń odmianowych (PDO). W każdej miejscowości i roku doświadczenia dwu-czynnikowe odmianowo-agrotechniczne były założone w doświadczalnym układzie strip-plot (pasów prostopadłych) z dwoma powtórzeniami. Ze względu na układ kombinacji doświadczalnych w serii tych doświadczeń, dane stanowiły czteroczynnikową (czterokierunkową) klasyfikację (tabelę) o postaci genotyp x miejscowość x poziom intensywności agrotechniki x rok (GxLxMxY). **Celem badań było zastosowanie dwuczynnikowego (klasycznego) modelu AMMI do analizy danych w klasyfikacji GxLxMxY i wnioskowania o adaptacji odmian do warunków środowiskowych i agrotechnicznych.** Zaproponowana metodyka statystyczna była efektywna do określenia megaśrodowisk (rejonów uprawy pszenicy ozimej w Polsce, w których odmiany wykazują

podobną interakcję genotypowo-środowiskową) oraz pozwoliła dokonać rekomendacji odmian zaadaptowanych do tych megaśrodków. Jak w każdej analizie AMMI, dodatkowym rezultatem opracowanej metody była redukcja błędu losowego (wytlumienie szumu losowego w danych), która przyczyniła się do zwiększenia precyzji ocen czynnikowych efektów głównych i interakcyjnych.

Analiza statystyczna rozpoczęła się od analiz układów strip-plot w każdej stacji doświadczalnej. Następnie połączono tak uzyskane dane w tabelę czterokierunkową genotyp x miejscowość x poziom agrotechniki x rok ($G \times L \times M \times Y$). Tabela została zawężona do podzbioru nie zawierającego braków danych (wszystkie podklasy $G \times L \times M \times Y$ miały określony plon). Tą tabelę analizowano za pomocą cztero-czynnikowego modelu mieszanego ANOVA. Określono, które interakcje są istotne. Zbiór średnich dla tej cztero-czynnikowej klasyfikacji $G \times L \times M \times Y$ przeorganizowano w dwukierunkową klasyfikację genotyp x środowisko ($G \times E$), w której 120 środowisk zostało zdefiniowanych jako kombinacje lokalizacji (L), poziomu agrotechniki (M) i roku obserwacji (Y). Ten dwukierunkowy zestaw danych analizowano modelem AMMI.

W celu diagnostyki liczby składowych głównych w modelu AMMI zastosowano procedurę sprawdzianu krzyżowego (*Cross-Validation*) z pojedynczą obserwacją w zbiorze testowym (*Leave-One-Out*). Zasada działania tej analizy jest następująca: z oryginalnego zestawu danych pojedyncza obserwacja jest usuwana (brak wartości w zbiorze uczącym) i adekwatny model, tu EM-AMMI, jest użyty do predykcji tej wartości. Powtarza się to dla każdej wartości w zbiorze danych (*exhausting version*). Średni kwadrat różnic między wartościami ukrytymi podczas estymacji parametrów modelu a wartościami przewidywanymi przez model (statystyka RMSPD, *root mean square prediction differences*) jest najmniejszy dla modelu najdokładniej przewidującego nieznane wartości.

Wnioskowanie jest trudniejsze jeśli model jest bardziej złożony. Dlatego warto również rozważyć uproszczenie modelu, jeśli wiąże się ono tylko z nieznacznym pogorszeniem właściwości predykcyjnych modelu. Model AMMI1 był w stanie zakumulować 26% interakcyjnej sumy kwadratów. Była to wartość o 80% większa od sumy kwadratów dla efektów głównych genotypu i dlatego jest ona znacząca. Suma kwadratów dla drugiej składowej głównej była znacznie niższa niż suma kwadratów dla genotypowych efektów głównych. Ponadto łączna suma kwadratów dla dalszych składowych (od drugiej wzwyż) zgadzała się z oszacowaniem błędu doświadczalnego. Stąd też wspólne rozważania dotyczące dokładności predykcyjnej i oszczędnego modelu doprowadziły do wyboru modelu AMMI1.

Pierwsza składowa główna miała ewidentną interpretację geograficzną, kontrastującą północno-wschodnią Polską (chłodniejszy klimat) i południowo-zachodnią Polską (cieplejszy klimat). Metodologiczne znaczenie tego artykułu to przedstawienie rozszerzenia analizy AMMI z analizy danych dwuczynnikowych (np. klasyfikacji GxE) do analizy danych w klasyfikacji wieloczynnikowej, takich jak rozpatrywany czteroczynnikowy zbiór danych GLMY. Uzyskane wyniki pozwoliły na konstrukcję mega-środowiskowego podziału obszaru Polski i określenia rekomendacji genotypów dla takich mega-środowisk.

3. An R function for imputation of missing cells in two-way data sets by EM-AMMI algorithm; Jakub Paderewski. *Communications in Biometry and Crop Science* (2013) 8:60-69. (SJR₂₀₁₃ czasopisma 0,237, punktacja MNISW₂₀₁₃ 10)

W doświadczalnictwie rolniczym stosowane są różne metody do potwierdzenia i interpretacji interakcji (na przykład genotypu z warunkami środowiskowymi). Część metod statystycznych do analizy danych w dwukierunkowej klasyfikacji pod kątem opisu interakcji wymaga, aby dane były kompletne, to znaczy nie mogą mieć brakujących wartości w podklasach (np. stosowane w naukach agronomicznych analizy AMMI lub GGE). W artykule zaproponowano **kod programu** w języku R **do estymacji brakujących wartości za pomocą algorytmu EM-AMMI**. Ponadto zawiera procedurę i) do sprawdzania powtarzalności tego algorytmu oraz ii) do określenia liczby składowych głównych (PCs) uzasadnionych w modelu. Pierwsza z nich może zostać użyta do oceny, czy brakujące dane zostały jednoznacznie wyznaczone, co jest szczególnie ważne w przypadku małych zbiorów danych. Druga procedura pozwala wybrać odpowiednią liczbę składowych głównych na bazie statystyki RMSPD (root mean square prediction differences, średni kwadrat z różnic między wartościami przewidywanymi przez model a wartościami obserwowanymi). Sprawdzenie liczby składowych głównych w modelu powinno być standardem w analizie danych przez modele oparte na rozkładzie według wartości osobliwych (SVD decomposition). Krótki opis procedur jest następujący:

Procedura EM.AMMI() estymuje brakujące kombinacje w dwukierunkowej tabeli zgodnie z metodą największej wiarygodności (expectation-maximization AMMI, EM-AMMI). Wyniki uzyskane procedurą są wiarygodne, gdy istnieją powiązania między plonami genotypów w różnych środowiskach i między plonami uzyskiwanymi w środowiskach przez różne odmiany. Statystyk musi podjąć decyzję o liczbie składowych głównych używanych do oszacowania brakujących wartości.

Aby sprawdzić, czy wynik procedury EM-AMMI() jest powtarzalny, procedura repeat.EM-AMMI() powtarza procedurę EM-AMMI wiele razy, za każdym razem z innym układem wartości początkowych. Wiarygodność wartości estymowanych (co implikuje powtarzalność) jest większa w przypadku większych zbiorów danych, silniejszych relacji między wierszami, silniejszych relacji między kolumnami, a także mniejszej liczbie brakujących wartości.

Optymalna liczba składowych głównych to ta, która ma najmniejszą wartość RMSPD. Do określenia tej liczby może być użyta metoda sprawdzianu krzyżowego (Cross-Validation, C-V). Popularną wersją metody C-V jest wersja z jedną wartością usuwaną ze zbioru modelującego (Leave-One-Out, L-O-O C-V). Do tego celu napisano procedurę CV.LOO(). W oryginalnym zestawie danych, przed uruchomieniem EM-AMMI, ukrywana jest jedna obserwacja, która później będzie używana jako zestaw danych walidacyjnych (wartość oceniająca predykcję). Estymację brakujących wartości wykonuje procedura EM-AMMI(), oparta na zestawie danych modelujących, który jest zbiorem danych bez tej jednej obserwacji. Taką procedurę powtarza się dla każdej obserwacji w zbiorze danych. Różnice pomiędzy ukrytą wartością a tym, co jest przypisane przez model EM-AMMI() (różnice w przewidywaniu) są podniesione do kwadratu, uśrednione i pierwiastkowane (statystyka RMSPD).

Podsumowując, autor zaprezentował opracowany przez siebie kod komputerowy, który umożliwi obliczenia analizy statystycznej EM-AMMI, która w tym czasie nie była oprogramowana. Proponowane dodatkowe analizy również zostały zamieszczone i przedstawione na ilustrujących przykładach.

4. The usefulness of EM-AMMI to study the influence of missing data pattern and application to Polish post-registration winter wheat data; Jakub Paderewski, Paulo C. Rodrigues. Australian Journal of Crop Science (2014) 8:640-645. (SJR₂₀₁₄ czasopisma 0,463, punktacja MNISW₂₀₁₄ 5)

Badanie interakcji genotypowo-środowiskowej (GEI) ma kluczowe znaczenie w doświadczalnictwie rolniczym, ponieważ zrozumienie jej umożliwia znaczną poprawę cech genotypowych roślin uprawnych. Genotypy i środowiska stanowią dwukierunkową klasyfikację czynnikową. Dane fenotypowe dla tych doświadczeń mogą być uporządkowane w dwukierunkowych tabelach z genotypami i środowiskami (przy czym środowiska mogą stanowić kombinacje lokalizacja-rok). W programach hodowli roślin niektóre genotypy

są często odrzucane, a inne dodawane w zamian z roku na rok, co powoduje brakujące wartości w takich zbiorach danych. Są to dostępne metody które radzą sobie z brakującymi wartościami w tabelach danych, ale niektóre metody nie mogą być stosowane w takich przypadkach. Jedną z najczęściej stosowanych alternatyw dla metod, które akceptują brakujące wartości, jest uzupełnienie wartości w brakujących komórkach przy użyciu metody EM-AMMI. W niniejszym opracowaniu zostało przedstawione **studium wpływu struktury brakujących wartości na efektywność algorytmu EM-AMMI**. Cztery scenariusze były rozważane: i) pierwszy, z brakującymi komórkami (podklasami kombinacji GE) występującymi całkowicie losowo; i trzy z następującymi wzorami braków danych (nielosowe), to jest ii) obserwacje występowały blokowo a bloki przy głównej diagonalu, iii) wzór diagonalny (obserwowane kombinacje ułożone były przy głównej diagonalu a braki danych w prawym górnym i lewym dolnym rogu tabeli) i iv) wzór blokowo-przekątny z odmianami kontrolnymi (występującymi we wszystkich środowiskach). Schematy odzwierciedlają najczęściej spotykane schematy brakujących danych w wieloletnich seriach doświadczeń odmianowych. Wyniki zostały porównane pod względem precyzji szacowania brakujących komórek i selekcji genotypów.

Podczas modelowania dwukierunkowych tabeli danych, przeprowadzenie diagnostyki modelu w celu wyboru najbardziej dostosowanego modelu AMMI (tj. wybranie "właściwej" liczby składników multiplikatywnych) ma kluczowe znaczenie dla analizy AMMI i dokonania właściwych zaleceń dla rolnictwa. W tym artykule badano trzy czynniki: odsetek brakujących wartości, liczbę składowych głównych w modelu AMMI i wzór brakujących komórek. Wszystkie trzy czynniki wykazały istotną interakcję między sobą w odniesieniu do statystyki RMSPD (opisującej różnicę między wartościami obserwowanymi a wartościami estymowanymi przez procedurę EM-AMMI). Wnioskiem z przeprowadzonych danych jest to, że algorytm EM-AMMI może być stosowany również w przypadkach nielosowych braków danych ale trzeba się wtedy liczyć z niewielką utratą dokładności estymacji. Zaobserwowano dużą stratę dokładności przy dopasowywaniu modelem EM-AMMI, który jest zbyt złożony, tj. zawiera zbyt wiele składowych głównych.

Podsumowując, rozpatrywano właściwości analizy statystycznej na podstawie symulacji komputerowych i w oparciu o prawdziwe dane (dane pszenicy ozimej z doświadczeń porejestrowych w Polsce). Te właściwości metody EM-AMMI nie były wcześniej badane. W pracy przedstawiono w sposób graficzny diagnozę modeli opartą na statystyce Cross-Validation i statystyce RMSPD.

5. Constrained AMMI model: application to polish winter wheat post-registration data;

Jakub Paderewski and Paulo C. Rodrigues. *Crop Science* (2018) 58:1458–1469. (IF₂₀₁₇ czasopisma 1,635, punktacja MNISW₂₀₁₇ 35)

Ograniczona analiza składowych głównych (C-PCA, constrained principal component analysis) opisuje dane z dwuwymiarowej tabeli przy założeniu liniowej zależności parametrów składowych głównych od znanych dodatkowych parametrów (tj. macierzy objaśniających). W tym badaniu użyliśmy C-PCA do uogólnienia modelu AMMI i **zapropowania ograniczonego modelu AMMI** (Constrained-AMMI). Ograniczony model AMMI był przedstawiony, gdy (i) tylko parametry składowych głównych dla środowisk miały macierz zmiennych objaśniających, (ii) tylko parametry składowych głównych dla genotypów miały macierz zmiennych objaśniających oraz (iii) oba typy parametrów miały macierze zmiennych objaśniających.

Zmiennymi objaśniającymi dla danych o plonowaniu odmian pszenicy ozimej były: mrozoodporność, wysokość, wyleganie, czas do kłoszenia, dojrzałości woskowata i podatność na choroby. W środowiskach (zdefiniowanych jako 3-czynnikowe kombinacje miejscowości x poziomu agrotechniki x rok), zmiennymi objaśniającymi były: rok, współrzędne geograficzne miejscowości, klasyfikacja gleby, odczyn gleby, wielkość opadów (skumulowane opady określono dla czterech okresów: i) zima, ii) kwiecień i maj, iii) czerwiec, iv) lipiec i sierpień) oraz poziom agrotechniki.

Opracowano model C-AMMI i porównano go ze standardowym modelem AMMI. Opis nowego modelu zawiera zarówno badania symulacyjne jak i analizy danych rzeczywistych. Procedura walidacji krzyżowej (Cross-Validation) została dostosowana do diagnozowania tej grupy modeli. Metoda L–O–O C–V (Leave–One–Out Cross-Validation) wraz ze statystyką RMSPD (Root Mean Squares of Predictive Differences) została użyta w celu selekcji zmiennych objaśniających ze zbioru objaśniającego (zarówno dla zmiennych objaśniających środowiska jak i genotypy). Weryfikacja krzyżowa była również używana do porównywania modeli C(G)–AMMI, C(E)–AMMI i AMMI, każdy z nich z dwoma składowymi głównymi.

Manuskrypt zawiera charakterystykę genotypów, środowisk w Polsce, nowe podejście do konstruowania megaśrodków (Mega–Environments). Wśród zalet modelu C-AMMI można wymienić następujące: zrozumienie interakcji GE wydaje się być bardziej interesujące i użyteczne niż precyzyjne opisanie reakcji plonu genotypów na badane środowiska, ponieważ pozwala to przewidzieć wydajność w innych środowiskach opisanych tym układem

parametrów środowiskowych. W naszej aplikacji model C(GE)–AMMI wyjaśniał nieco mniejszą zmienność (tj. sumę kwadratów) danych, ale zaobserwowano duży zysk pod względem możliwości interpretacji: parametry genotypowe i parametry środowiskowe mają znaczenie fizyczne i biologiczne, ponieważ zależą one od objaśniających je współzmiennych genotypowych i środowiskowych.

Metoda C–V i statystyka RMSPD pozwala na wybór liczby składowych głównych w modelu, co jest bardzo użyteczne przy ocenie przydatności zmiennych objaśniających. Opracowana analiza C–AMMI może zostać zastosowana do oceny innych doświadczeń wielośrodowiskowych (multi-environments trials) z innymi roślinami uprawnymi i innym opisem warunków środowiskowych. Należy jednak zachować ostrożność przy wyborze zmiennych środowiskowych lub genotypowych. Jeśli użyte zostanie zbyt wiele zmiennych, analiza C-AMMI dąży do bycia zwykłą analizą AMMI, z ocenami parametrów genotypowych i/lub środowiskowych opisanymi za pomocą regresji liniowej. Z drugiej strony brak odpowiednich zmiennych objaśniających uniemożliwia odtworzenie rzeczywistego wzorca interakcji.

Podsumowując, autorzy postawili sobie za cel opracowanie nowej analizy statystycznej. Opracowaną metodę zastosowano do opisu plonowania odmian pszenicy ozimej oraz do danych symulacyjnych. Analiza została połączona ze standardowymi typami wykresów, ale też zaprezentowano nowy typ wykresu składowych głównych.

Podsumowanie powyższych prac:

W pierwszych dwóch artykułach przedstawiono zastosowanie analizy AMMI do danych o większej liczbie czynników niż dwa. W pierwszym artykule dane uśredniono przez trzeci czynnik (lata) za pomocą estymatorów BLUE, uzyskując kompletną klasyfikację dwukierunkową. W drugim artykule analizowano dane bez brakujących wartości traktując kombinację czynników opisujących warunki środowiskowe (miejscowość, poziom intensywności agrotechniki, rok) jako pojedynczy czynnik środowiskowy.

W trzecim prezentowanym artykule napisałem kod komputerowy, który pozwala prowadzić obliczenia dla analizy statystycznej (EM-AMMI), ponieważ brakowało takiego oprogramowania. Analiza ta jest stosowana zamiast analizy AMMI gdy w zbiorze danych są brakujące informacje. Ponadto artykuł zawiera narzędzie do diagnozowania modeli metodą

Cross-Validation. Czwarty artykuł prezentuje badanie dotyczące właściwości proponowanej wcześniej procedury EM-AMMI.

Piąta praca zawiera proponowaną przeze mnie modyfikację analizy AMMI, która może zwiększyć efektywność modelu AMMI. Opracowane procedury zostały zastosowane przeze mnie do oceny plonu odmian pszenicy ozimej.

Prezentowane prace są przydatne dla osób, które chcą analizować interakcje genotypu ze środowiskiem, która występuje podczas serii doświadczeń wielo-środowiskowych. Prace skupiały się na uprawach pszenicy, ale takie analizy można zastosować dla innych gatunków roślin rolniczych.

Jestem pierwszym autorem we wszystkich prezentowanych artykułach i byłem głównym wykonawcą. Poprawki po recenzjach i odpowiedzi na recenzje były w 95% wykonane przeze mnie, również czuwałem nad ostateczną formą tekstu. Prezentowane materiały pokazują moje kompleksowe umiejętności w zakresie analizy statystycznej zbiorów danych agronomicznych.

Inne osiągnięcia naukowe i badawcze

Mój wkład koncentruje się głównie na analizie statystycznej agrotechnicznych lub ekologicznych zbiorów danych. Moją specjalnością jest adaptacja (gdy istnieje potrzeba modyfikacji) i zastosowanie metody rozkładu macierzy według wartości osobliwych (SVD, Singular Value Decomposition), która jest zasadniczym elementem modelu AMMI lub GGE. Dlatego bywałem zapraszany do współpracy w takich właśnie przypadkach.

1. Mastalerczuk G., Borawska-Jarmułowicz B., Kalaji H.M., Dąbrowski P., **Paderewski J.** 2017. Gas-exchange parameters and morphological features of festulolium (*Festulolium braunii* K. Richert A. Camus) in response to nitrogen dosage. *Photosynthetica*, 55, 1, 20-30. (IF₂₀₁₇=1,740, MNISW₂₀₁₇=25)

Opracowałem analizę cech morfologicznych festulolium w odpowiedzi na dawkowanie azotu analizowałem według analizy GGE dostosowanej do projektu 3-czynnikowego. Podobny problem został opracowany przeze mnie także w następujących artykułach:

2. Javed S., Rauf S., **Paderewski J.**, Malinowski D.P., Saleem U., Shahzad M. 2016. Evaluation of Egyptian clover (*Trifolium alexandrinum* L.) germplasm through redundancy analysis for forage yield and its components. *Crop Science*, 56 (3), 1179-1188. (IF₂₀₁₆=1,629, MNISW₂₀₁₆=30)

3. Hussain M.M., Rauf S., **Paderewski J.**, ul Haq I., Sienkiewicz-Paderewska D., Monneveux P. 2015. Multitraits evaluation of Pakistani ecotypes of berseem clover (*Trifolium alexandrinum* L.) under full-irrigation and water restriction conditions. *Journal of Applied Botany and Food Quality*, 88, 127-133. (IF₂₀₁₅=1,085, MNISW₂₀₁₅=20)

4. Hussain M.M., Kausar M., Rauf S., Farukh Zafar Khan M., **Paderewski J.**, Khan M., ul Haq I., Raza A.B.M. 2018. Selection for some functional markers for adaptability of *Helianthus argophyllus* × *Helianthus annuus* derived population under abiotic stress conditions. *Helia* 2018, 41(68), 83-108. (SJR₂₀₁₈=0,427, MNISW₂₀₁₆=5)

Przy czym, w ostatnim artykule zastosowałem połączenie analizy SVD z analizą skupień, a ponadto obliczyłem statystykę LOD (statystyka charakterystyczna dla analizy materiału genetycznego).

Zastosowanie procedury SVD (z opracowanymi dodatkowymi elementami) zostało wykonane przeze mnie w następujących artykułach:

5. Sienkiewicz-Paderewska D., **Paderewski J.** 2011. Możliwość zastosowania metody GGE biplot do oceny stanu i identyfikacji zagrożeń zespołu *Alopecuretum pratensis* opisanego na terenie Parku Krajobrazowego „Podlaski Przełom Bugu”. *Woda-Środowisko-Obszary Wiejskie*, 11 (3), 237–252. (MNISW₂₀₁₁=6)

6. Sienkiewicz-Paderewska D., **Paderewski J.** 2012. Habitat preferences of plant associations from *Molinio-Arrhenatheretea* class described in the 'The Bug Ravine Landscape Park' by GGE biplot analysis. *Grassland Science in Europe*, 17, 661-663. (MNISW₂₀₁₂=10)

7. Sienkiewicz-Paderewska D., **Paderewski J.** 2015. Habitat preferences of plant communities: new approach based on the GGE biplot analysis. *Polish Journal of Ecology*, 63, (3), 387-399. (IF₂₀₁₅=0,500, MNISW₂₀₁₅=15)

8. Sienkiewicz-Paderewska D., Narewska S., Narewski D., Olszewski T., **Paderewski J.** 2017. Floristic diversity and utilization value of the semi-natural grassland in the lower section of the Bug River Valley. (in English) *Łąkarstwo w Polsce (Grassland Science in Poland)*, 20, 121-147. (MNISW₂₀₁₇=7)

Zastosowanie **modelu AMMI do analizy plonu wielośrodowiskowego** wykonałem również (t.j. poza artykułami przedstawionymi jako główne osiągnięcia postępowania habilitacyjnego) w następujących pracach. W większości z nich należało rozwiązać więcej niż

dwa czynniki, podczas gdy metoda SVD (i w konsekwencji AMMI) została opracowana do analizy doświadczeń dwuczynnikowych.

9. Mądry 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 (2), 76-82. (IF₂₀₁₇=0,467, MNISW₂₀₁₇=20)
10. Mądry W., **Paderewski J.**, Gozdowski D., Rozbicki J., Golba J., Piechociński M., Studnicki M., Derejko A. 2013. Adaptation of winter wheat cultivars to crop managements and Polish agricultural environments. *Turkish Journal of Field Crops*, 18 (1), 118-127. (IF₂₀₁₃=0,641, MNISW₂₀₁₃=20)
11. Mądry W., Gacek E, **Paderewski J.**, Gozdowski D., Drzazga T. 2011. Adaptive yield response of winter wheat cultivars across environments in Poland using combined AMMI and cluster analyses. *International Journal of Plant Production*, 5 (3), 299-309. (IF₂₀₁₁=1,100, MNISW₂₀₁₁=13)
12. Mądry W., **Paderewski J.**, Rozbicki J., Gozdowski D., Golba J., Piechociński M., Studnicki M., Derejko A. 2012. Plonowanie odmian pszenicy ozimej w różnych środowiskach - jednoroczna seria PDOiR. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 263, 189-204. (MNISW₂₀₁₂=4)
13. **Paderewski J.**, Mądry W. 2012. Zastosowania modelu AMMI do analizy reakcji odmian na środowiska. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 263, 161-188. (MNISW₂₀₁₂=4)
14. **Paderewski J.**, Mądry W., Rozbicki J. 2010. Yielding of old and modern Polish wheat cultivars under different nitrogen inputs as assessed by combined using AMMI and cluster analyses. *Plant Breeding and Seed Science*, 62, 117-136. (MNISW₂₀₁₀=6)
15. Drzazga T., **Paderewski J.**, Mądry W., Krajewski P. 2009. Ocena rodzajów reakcji plonowania odmian pszenicy ozimej w doświadczeniach PDO na przestrzennie zmienne warunki przyrodnicze w kraju. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 253, 71-82. (MNISW₂₀₁₀=6)
16. **Paderewski J.**, Mądry W., Pilarczyk W., Drzazga T. 2008. Retrospektywne badanie reakcji plonu odmian pszenicy ozimej na warunki środowiskowe w miejscowościach za pomocą łącznej analizy AMMI i skupień : ocena postępu genetycznego w plonowaniu. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, 250, 87-106. (MNISW₂₀₁₀=6)

17. Mądry W., **Paderewski J.**, Drzazga T. 2006. Ocena reakcji plonu ziarna rodów hodowlanych pszenicy ozimej na zmienne warunki środowiskowe za pomocą analizy AMMI. *Fragmenta Agronomica*, 23 (4), 130-143. (MNISW₂₀₁₀=6)

18. **Paderewski J.**, Mądry W. 2006. Addytywno-multiplikatywny model AMMI do statystycznej analizy danych z serii doświadczeń genotypowych. *Colloquium Biometryczne*, 36, 125-148. (MNISW₂₀₁₀=6)

Pierwszym artykułem naukowym, w którym miałem wkład związany z analizą SVD jest publikacja napisana w 2004 roku, w której współpracowałem przy opisie teorii i znaczenia analizy PCA, wykorzystanej w tym artykule do charakterystyki zmienności genetycznej.

19. Sieczko L., Mądry W., Zieliński A., **Paderewski J.**, Urbaś-Szwed K. 2004. Zastosowanie analizy składowych głównych w badaniach nad wielo cechową charakterystyką zmienności genetycznej w kolekcji zasobów genowych pszenicy ozimej (*Triticulum durum* L.). *Colloquium Biometryczne*, 223-239. (MNISW₂₀₁₀=6)

Artykuły, w których wybrałem i wykonałem **inne techniki statystyczne:**

20. 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 multienvironmental trials. *Crop Science*, 57 (5), 2506-2516. (IF₂₀₁₇=1,635, MNISW₂₀₁₇=35)

Zaadaptowałem statystykę MSEP do projektu eksperymentu z tylko dwoma powtórzeniami. Napisałem program w języku R i wykonałem porównanie modeli za pomocą analizy MSEP.

21. Kozak M., **Paderewski J.**, Zieliński A. 2006. Numerical problems connected with accuracy of variance estimation. *Model Assisted Statistics and Applications*, 1 (3), 101-105. (SJR₂₀₀₆=0,101, MNISW₂₀₁₀=7)

Przekształciłem formuły dla estymatora wariancji dla populacji skończonej. Ta metoda drastycznie zmniejszyła czas obliczeń w stosunku do oryginalnych wzorów. Napisałem algorytm R i miałem udział w jego opisanii.

22. **Paderewski J.**, Sienkiewicz-Paderewska D. 2010. Multidimensional scaling for describing the risks for grasslands ecosystems located in the Polish Landscape Parks. *Grassland Science in Europe*, 15, 723-725. (MNISW₂₀₁₀=5)

Zastosowałem analizę NMDS (non-metric multidimensional saling analysis) niemetryczne wielowymiarowe skalowanie z 2-wymiarowym wykresem do analizy zagrożeń

występujących dla ekosystemów łąkowych na podstawie ankiet wypełnianych przez dyrekcje Parków Krajobrazowych.

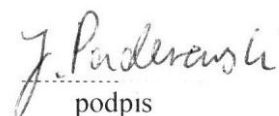
23. Sienkiewicz-Paderewska D., Narewska S., Narewski D., Olszewski T., **Paderewski J.** 2017. Ecological and syntaxonomical spectra of grass species of the grassland communities occurring in in the lower section of the Bug River Valley. (in English) *Łąkarstwo w Polsce* (Grassland Science in Poland), 20, 149-162. (MNISW₂₀₁₇=7)

Obliczyłem wybrane miary ekologiczne (technika statystyki opisowej). Zaprojektowałem i wykonałem wykresy w celu zobrazowania wyników.

Prace, w których miałem mniejszy udział, polegający głównie na **konsultacjach** i pisaniu małych fragmentów tekstu:

24. Kozak M., Bakinowska E., **Paderewski J.** 2009. Polish agricultural universities students' graphical perception. *Model Assisted Statistics and*, 4 (4), 281-286. (SJR₂₀₀₉=0,179, MNISW₂₀₁₀=7)

.....Warszawa.....
miejsowość


.....
podpis