

WALIDACJA SYSTEMU WSPOMAGANIA DECYZJI ZEASOFT – MODELE PLONÓW

Andrzej S. Zaliwski, Anna Nieróbca

Zakład Agrometeorologii i Zastosowań Informatyki

Instytut Uprawy Nawożenia i Gleboznawstwa-Państwowy Instytut Badawczy w Puławach

Streszczenie. Dokonano walidacji regresyjnych modeli plonów kiszonki i ziarna systemu ZeaSoft z wykorzystaniem ok. 10 000 rekordów danych historycznych z doświadczeń odmianowych kukurydzy. Do automatyzacji przetwarzania i analizy tych zbiorów opracowano program komputerowy. Przy jego pomocy wygenerowano zestawienia danych doświadczalnych i korespondujących wyników pochodzących z modeli regresyjnych plonów ziarna i kiszonki, interpolowanych dla miejscowości przy pomocy modelu indeksu klimatycznego kukurydzy. Wstępna analiza wyników pozwoliła stwierdzić błąd systematyczny modelu plonów kiszonki wynoszący średnio 24% i błąd modelu plonów ziarna wynoszący średnio 11%.

Słowa kluczowe: system wspomaganie decyzji, kukurydza, model plonów, walidacja modeli

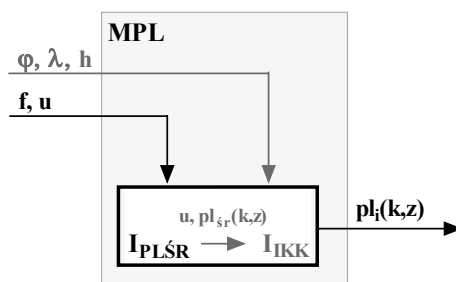
Wprowadzenie

Producenci kukurydzy mogą podnosić efektywność produkcji zasadniczo na dwa sposoby: obniżając koszty przez doskonalenie technologii produkcji oraz podnosząc wartość produkcji przez właściwy dobór odmian roślin do siedliskowych i ekonomiczno-organizacyjnych warunków produkcji [Krasowicz 2007]. Korzystny dobór odmiany w zależności od warunków siedliskowych jest uwarunkowany dysponowaniem dokładniejszymi informacjami o odmianie i o warunkach. Z myślą o dostarczaniu takich informacjach opracowano system wspomaganie decyzji ZeaSoft [Zaliwski 2009]. Dostarcza on informacji o efekcie ekonomicznym uprawy kukurydzy na ziarno i kiszonkę w zależności od typu wczesności kukurydzy (określonego liczbą FAO). Informacje takie mogą być generowane dla gospodarstwa o dowolnym położeniu geograficznym w kraju. Do obliczenia efektu ekonomicznego wykorzystywany jest zbiór danych pochodzących z kilku różnych modeli: technologii produkcji, nawożenia, indeksu klimatycznego kukurydzy (IKK), regresyjnych modeli plonów oraz modelu prawdopodobieństwa dojrzewania kukurydzy. Ustalono, że w systemie ZeaSoft istotnym problemem mogą być niedokładności modeli plonów, bowiem od wyników plonowania zależy końcowa ocena wariantów technologicznych. Wynikła stąd konieczność dokładniejszej analizy problemu i wskazanie sposobu jego rozwiązania.

Celem niniejszej pracy jest analiza przepływu informacji w modelach plonów systemu wspomaganie decyzji ZeaSoft i prezentacja procesu walidacji modeli.

Analiza przepływu informacji w modelu plonów

W systemie ZeaSoft model plonów (MPL) zawiera dwa zasadnicze algorytmy, $I_{PL\dot{S}R}$ i I_{IKK} (rys. 1). Z punktu widzenia przetwarzania informacji [Mazur 1970] $I_{PL\dot{S}R}$ jest zbiorem informacji o sposobie transformowania liczby FAO i kierunku użytkowania kukurydzy na plon średni (dla Polski). Jest on złożony z dwóch informacji, mających postać dwóch równań regresyjnych, jednego dla ziarna i drugiego dla kiszonki. Obliczenie plonu dla dowolnego punktu w Polsce umożliwia informacja I_{IKK} - model indeksu plonu kukurydzy, opracowany w Zakładzie Agrometeorologii i Zastosowań Informatyki IUNG-PIB. Informacja I_{IKK} wyraża sposób obliczenia plonu średniego kukurydzy o kierunku użytkowania „u” w zależności od położenia geograficznego określonego przez współrzędne φ, λ, h . Pozwala to na interpolację plonu średniego $pl_{sr}(k,z)$ dla gospodarstwa o dowolnym położeniu geograficznym w granicach kraju w celu otrzymania wartości plonu interpolowanego $pl_i(k,z)$.



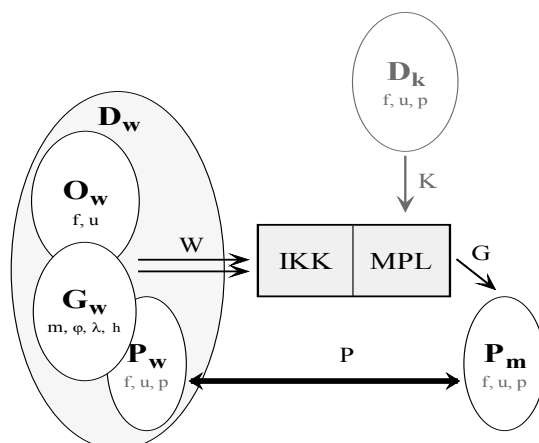
Rys. 1. Transformacja danych wejściowych na dane o plonie. Oznaczenia: MPL - model plonów, $I_{PL\dot{S}R}$ - informacja o sposobie obliczenia plonu średniego, I_{IKK} - informacja o sposobie obliczenia indeksu klimatycznego kukurydzy; f - liczba FAO, u - kierunek użytkowania (kiszonka lub ziarno), φ, λ, h - współrzędne geograficzne; $pl_{sr}(k,z)$, $pl_i(k,z)$ - odpowiednio: plon średni kiszonki lub ziarna, plon interpolowany kiszonki lub ziarna

Fig. 1. Transformation of input data into data on yields. Notation: MPL - yield model, $I_{PL\dot{S}R}$ - information how to calculate mean yield, I_{IKK} - information how to calculate maize climatic index; f - FAO number, u - utilisation mode, φ, λ, h - geographic coordinates; $pl_{sr}(k,z)$ - silage or grain mean yield, $pl_i(k,z)$, silage or grain interpolated yield, respectively

Przeprowadzenie walidacji

Walidacja modelu polega na wykazaniu, że w zakresie określonym celem modelowania informacje z modelu (obrazy) mogą zastąpić informacje z modelowanego systemu empirycznego (oryginały). Celem modelowania plonów w systemie ZeaSoft jest prognoza plonów dla gospodarstwa dowolnie położonego w granicach Polski. W momencie generowania prognozy oryginały (dane o plonie rzeczywistym, zebranych w gospodarstwie) jeszcze nie istnieją, bowiem prognoza jest wykonywana przed siewem, a wartości plonu rzeczywistego mogą być dostępne dopiero po zbiorze. Z tego powodu do celów walidacji modelu

plonów odwrócono kierunek przepływu danych, tzn. dla posiadanych danych o plonach rzeczywistych P_w zasymulowano prognozę plonów P_m (rys. 2). Wykorzystano tutaj zbiór danych historycznych o plonach D_w , pochodzący z polowych doświadczeń odmianowych COBORU.



Rys. 2. Przebieg walidacji. Oznaczenia: IKK - model indeksu klimatycznego kukurydzy, MPL - modele plonów; D_w - zbiór danych walidacyjnych pozyskanych z doświadczeń polowych; O_w - podzbiór danych odmianowych, G_w - podzbiór danych geograficznych, P_w - podzbiór danych o plonach; D_k - zbiór danych, użyty do konstrukcji modeli MPL; P_m - zbiór danych o plonach, wygenerowanych przez modele MPL do walidacji; dane: f - liczba FAO, u - kierunek użytkowania, m - miejscowość, p - plon, φ , λ , h - współrzędne geograficzne; procesy: K - konstrukcja modelu, W - wprowadzanie danych, G - generowanie zbioru wynikowego, P - porównanie plonów.

Fig. 2. Course of validation. Notation: IKK - maize climatic index model, MPL - yield models; D_w - validation data set acquired from field experiments: O_w - variety data subset, G_w - geographic data subset, P_w - yield data subset; D_k - data set used for MPL models construction; P_m - yield data set generated by model; data: f - FAO number, u - utilisation mode, m - locality, p - yield, φ , λ , h - geographic coordinates; processes: K - model construction, W - data input, G - result set generation, P - yields comparison.

Walidację poprzedziła konstrukcja dwóch modeli MPL (dla kiszonki i ziarna) na podstawie zbioru danych P_k (rys.2), co nastąpiło znacznie wcześniej, przy opracowaniu systemu ZeaSoft. Modele MPL są wynikiem agregacji danych, tzn. cały zbiór danych P_k zostaje zastąpiony dwoma informacjami uśrednionymi. Agregacja danych prowadzi z reguły do utraty precyzji, w tym przypadku jednak tak nie jest. Przy prognozowaniu plonu bowiem można jedynie stwierdzić, że przyszły plon w danej miejscowości z dużym prawdopodobieństwem znajdzie się w zakresie plonów uzyskiwanych w latach poprzednich. Jeżeli rozwarłość zakresu zmniejszymy tylko do jednego roku, prawdopodobieństwo trafności prognozy będzie niewielkie. Agregacja danych odpowiada zwiększaniu rozwarłości zbioru (zmniejszanie precyzji), ale zwiększa się jednocześnie prawdopodobieństwo zgodności

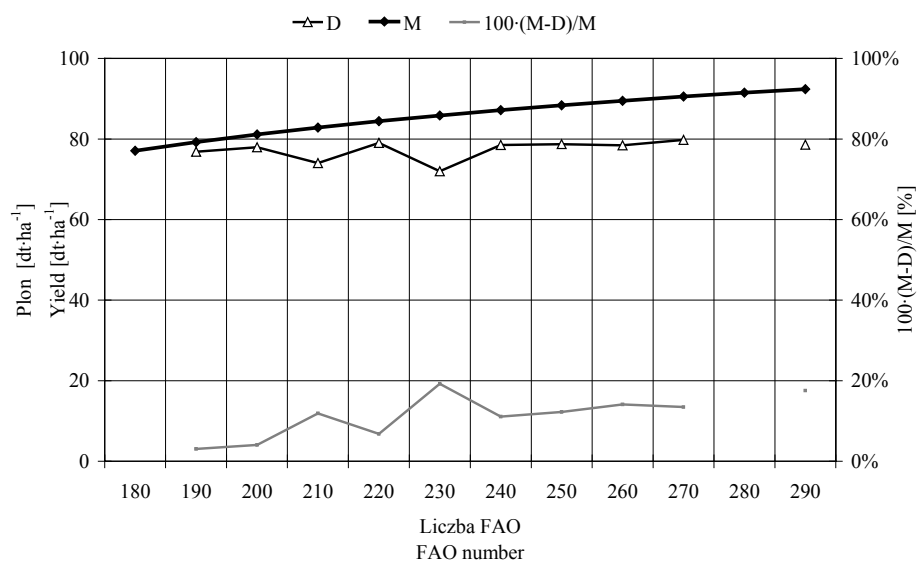
plonu rzeczywistego z prognozowanym (zwiększanie precyzji). Wynika z tego, że w naszym przypadku agregacja danych nie obniży zasadniczo trafności prognozy. Zbiór P_k musi być natomiast różny od zbioru walidacyjnego P_w , ponieważ informacja weryfikująca powinna pochodzić ze źródła niezależnego, innego niż informacja weryfikowana.

Przebieg walidacji można prześledzić na rys.2. Istotą walidacji jest porównanie zgodności plonów w zbiorach P_m (obrazy) i P_w (oryginały). Ich zgodność pozwoliłaby sądzić, że wartości plonu z modelowanego systemu empirycznego można zastąpić wartościami plonu z modelu. Zbiór P_m są to dane wygenerowane przy użyciu podzbiorów O_w i G_w , z zachowaniem odpowiedniości elementów tych zbiorów. Innymi słowy, odpowiadające sobie dane w zbiorach P_w , P_m , O_w i G_w dają się uporządkować w jeden rekord bazodanowy. Fakt ten wykorzystano przy automatyzacji procesu walidacyjnego, która okazała się niezbędna w związku z dużą liczebnością danych użytych do walidacji (ok. 10 000 rekordów).

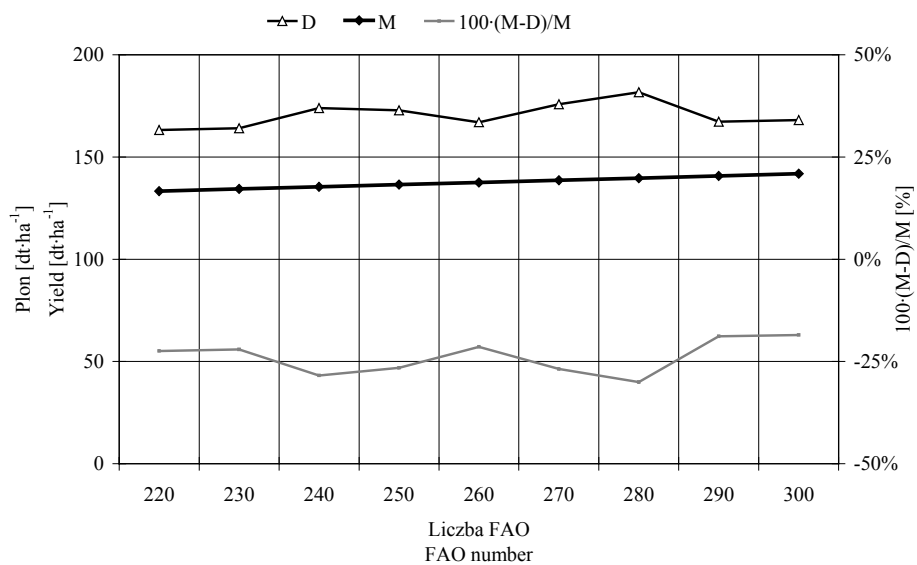
Do automatyzacji procesu walidacji opracowano program ZeaSoftValidation, zawierając w nim procedury do analizy i przygotowania danych z doświadczeń odmianowych kukurydzy. Oryginalne dane miały analogowy format opisowy, który wymagał digitalizacji. Dane po digitalizacji nadal zachowały cechy formatu opisowego. Analizą objęto dane zdigitalizowane w celu wygenerowania listy miejscowości i odmian. Na podstawie tych list skonstruowano następnie tabele z danymi pośrednimi. Ważnymi procedurami analitycznymi były algorytmy wyszukujące powtórzenia, błędy i luki w danych, co umożliwiło ich ręczną korektę i uzupełnienie. Analiza dotyczyła także danych przetworzonych w celu zapoznania się z danymi (np. z liczebnością powtórzeń przy podziale na różne klasy, zakresami klas, itd.). Przygotowanie danych polegało na wygenerowaniu rekordów dla zbioru G_w i ich zapisie do bazy danych. Rolą procedur przygotowujących dane było wyszukanie odpowiednich fragmentów tekstu w danych zdigitalizowanych, utworzenie rekordów i automatyczne uzupełnienie o współrzędne geograficzne na podstawie nazw miejscowości oraz liczby FAO na podstawie nazw odmian kukurydzy. Dane potrzebne do uzupełnienia tych rekordów były pobierane z tabel z danymi pośrednimi. Do dalszej analizy ze zbioru G_w wybrano tylko rekordy zgodne co do odmiany z modelami plonów.

Analiza i omówienie wyników

Program ZeaSoftValidation zawiera niezbędne algorytmy modeli IKK i MPL programu ZeaSoft w celu generowania rekordów zbiorów P_w i P_m - zestawień danych doświadczalnych i korespondujących wyników pochodzących z modeli regresyjnych plonów ziarna i kiszonki, interpolowanych dla miejscowości przy pomocy modelu IKK. Po wstępnej analizie wyników zbioru P_w i P_m podzielono na dane dotyczące ziarna i kiszonki, a rekordy pogrupowano wg liczby FAO, zgodnie z algorytmem modeli MPL. Obliczone średnie plony dla zbiorów P_w i P_m przedstawiono na rys. 3 i 4. Średnia obejmuje wszystkie miejscowości i lata doświadczeń, selekcję rekordów i obliczenie średnich wykonuje się przy pomocy zapytania SQL.



Rys. 3. Wyniki walidacji plonów ziarna. Oznaczenia: D - dane z doświadczeń, M - dane z modelu
 Fig. 3. Results of grain yields validation. Notation: D - data from field experiments, M - data from the model



Rys. 4. Wyniki walidacji plonów kiszonki. Oznaczenia: D - dane z doświadczeń, M - dane z modelu
 Fig. 4. Results of silage yields validation. Notation: D - data from field experiments, M - data from the model

Jak wynika z przedstawionych wykresów, zarówno modele jak i dane doświadczalne charakteryzują się dość szerokim zakresem liczby FAO. W przypadku odmian kukurydzy na ziarno wynosi ona 220-300, zaś odmian na kiszonkę 180-290. Wystąpił brak danych doświadczalnych dla odmian kukurydzy na ziarno o liczbie FAO 180 i 280 (rys. 3), związany z niezgodnością odmian w modelach i doświadczeniach. Niezgodność ta może wynikać z faktu, że prezentowana analiza nie wykorzystuje wszystkich dostępnych danych doświadczalnych, a także z wprowadzania nowych odmian do rejestru odmian i usuwaniem starszych. Program ZeaSoft został opracowany bowiem w 2004 roku, a dane do jego konstrukcji pochodzą z lat wcześniejszych. Dane doświadczalne natomiast pochodzą z lat 1999-2008. Lata pokrywają się więc tylko w niewielkim zakresie.

Analizując rys. 3 i 4 można stwierdzić, że zarówno plony ziarna jak i kiszonki z modelu odbiegają od plonów z doświadczeń. Różnica ta dla ziarna wynosi średnio ok. +11%. Natomiast plony kiszonki z modelu odbiegają od plonów z doświadczeń średnio o -24% wskazując na znaczący błąd systematyczny modelu plonów kiszonki.

Wnioski

Przeprowadzona walidacja modeli plonów ziarna i kiszonki kukurydzy upoważnia do stwierdzenia następujących wniosków:

1. Występujące luki w danych użytych do konstrukcji modeli względem danych doświadczalnych wskazują na potrzebę śledzenia zmian w rejestrze odmian kukurydzy i aktualizację modeli plonów przynajmniej raz na kilka lat.
2. W przypadku modelu ziarna plony z modelu odbiegały od plonów z doświadczeń średnio o +11%. Natomiast dla modelu kiszonki plony z modelu odbiegały od plonów z doświadczeń średnio o -24% wskazując na znaczący błąd systematyczny. Wyniki te sugerują potrzebę opracowania nowego modelu plonów kiszonki.

Bibliografia

- Krasowicz S.** 2007. Organizacyjne i ekonomiczne aspekty produkcji pasz. Wybrane elementy technologii produkcji roślinnej [online]. Studia i Raporty IUNG-PIB. Nr 9, s. 209-216. [dostęp 3-12-2010]. Dostępny w Internecie: www.iung.pulawy.pl/images/wyd/pib/zesz9.pdf.
- Mazur M.** 1970. Jakościowa teoria informacji [online]. WNT. Warszawa. [dostęp 25-03-2010]. Dostępny w Internecie: www.autonom.edu.pl/publikacje.html.
- Zaliwski A.S.** 2009. System wspomagania decyzji w wyborze odmiany kukurydzy (ZeaSoft) [online]. Studia i Raporty IUNG-PIB. Nr 16, s. 83-96. [dostęp 3-12-2010]. Dostępny w Internecie: www.iung.pulawy.pl/images/wyd/pib/zesz16.pdf.

Publikacja opracowana w ramach zadania 2.9 programu wieloletniego „Kształtowanie środowiska rolniczego Polski oraz zrównoważony rozwój produkcji rolniczej”.

VALIDATION OF THE ZEASOFT DECISION SUPPORT SYSTEM – YIELD MODELS

Abstract. A validation of the regression models of silage and grain yields of the ZeaSoft system was conducted using approximately 10 000 historical data records from maize variety field experiments. A computer program was developed in order to automate the processing and analysis of the data. With its aid experimental data sets and the corresponding results derived from the silage and grain yield regression models, interpolated for the locations with the model of climatic maize index, were generated. A preliminary analysis of the results indicate a systematic error of the silage yield model of about 24% and an error of the grain yield model of about 11%.

Key words: decision support system, maize, yield model, model validation

Adres do korespondencji:

Andrzej Zaliwski; e-mail: andrzej.zaliwski@iung.pulawy.pl
Zakład Agrometeorologii i Zastosowań Informatyki
Instytut Uprawy Nawożenia i Gleboznawstwa - Państwowy Instytut Badawczy
ul. Czartoryskich 8
24-100 Puławy