

**Artur Łopatka, Bożena Smreczak, Magdalena Łysiak, Piotr Koza,
Beata Suszek-Łopatka**

*Instytut Uprawy Nawożenia i Gleboznawstwa – Państwowy Instytut Badawczy
w Puławach*

PROGNOZA WPŁYWU EZŁ NA WŁAŚCIWOŚCI GLEBY I PLONY ZBÓŻ Z WYKORZYSTANIEM METOD UCZENIA MASZYNOWEGO*

Słowa kluczowe: Europejski Zielony Ład, gleba, plon, uczenie maszynowe, lasy losowe

Wstęp

W odpowiedzi na ogólnoświatowe problemy klimatyczne i środowiskowe w roku 2019 Komisja Europejska (KE) przedstawiła propozycję nowej strategii Europejskiego Zielonego Ładu (EZŁ) (ang. *European Green Deal*) (17). Jej celem jest budowa nowoczesnej, zasobooszczędnej i konkurencyjnej gospodarki, której wzrost zostanie oddzielony od wykorzystania zasobów naturalnych i która do roku 2050 osiągnie neutralność klimatyczną. W zakresie Wspólnej Polityki Rolnej (WPR) zapowiedziano m.in. że „Plany strategiczne będą musiały odzwierciedlać bardziej ambitne cele obejmujące znaczące ograniczenie stosowania chemicznych pestycydów i zagrożeń z nimi związanych, jak również stosowania nawozów i antybiotyków.”. Zapisy implementujące założenia EZŁ odnośnie stosowania nawozów i pestycydów w rolnictwie, w przedstawionej w roku 2020 przez KE strategii „od pola do stołu” (18), zapowiadają do 2030 r.:

- ograniczenie zużycia pestycydów o 50%;
- ograniczenie „strat składników pokarmowych o co najmniej 50%, przy jednoczesnym zapewnieniu, by nie doszło do pogorszenia żyzności gleby”, co doprowadzi do ograniczenia zużycia nawozów sztucznych o co najmniej 20%;
- przeznaczenie na rolnictwo ekologiczne co najmniej 25% gruntów rolnych w UE;
- objęcie 10% powierzchni użytków rolnych UE elementami krajobrazu o wysokiej różnorodności.

*Opracowanie wykonano w ramach zadania 1.2 pt. „Gleby użytkowane rolniczo” z dotacji budżetowej przeznaczonej na realizację zadań MRiRW w 2021 r.

W przedstawionym dla Polski w roku 2021 przez MRiRW Planie Strategicznym na lata 2023–2027 znalazł się ponadto w ramach ekoschematu „Opracowanie i przestrzeganie planu nawożenia” (25) zapis zapowiadający:

- dopłaty do wapnowania gleb o $\text{pH} \leq 5,5$ (od 2023 r.).

Wsparcie uzyskują również działania ograniczające straty składników pokarmowych, w tym plany nawozowe oparte na chemicznej analizie zasobności gleby oraz wdrożenie metod rolnictwa precyzyjnego. Wprowadzenie działań zarówno redukujących dawki środków produkcji, jak i zwiększających efektywność ich wykorzystania sprawia, że ocena efektu netto planowanych zmian zarówno w odniesieniu do właściwości gleb, jak i poziomu plonów nie jest oczywista. Bez takiej oceny utrudnione są jednak kolejne kroki w zakresie efektywnego uszczegółowionego wdrożenia założeń EZŁ do Planów Strategicznych i praktyki rolniczej.

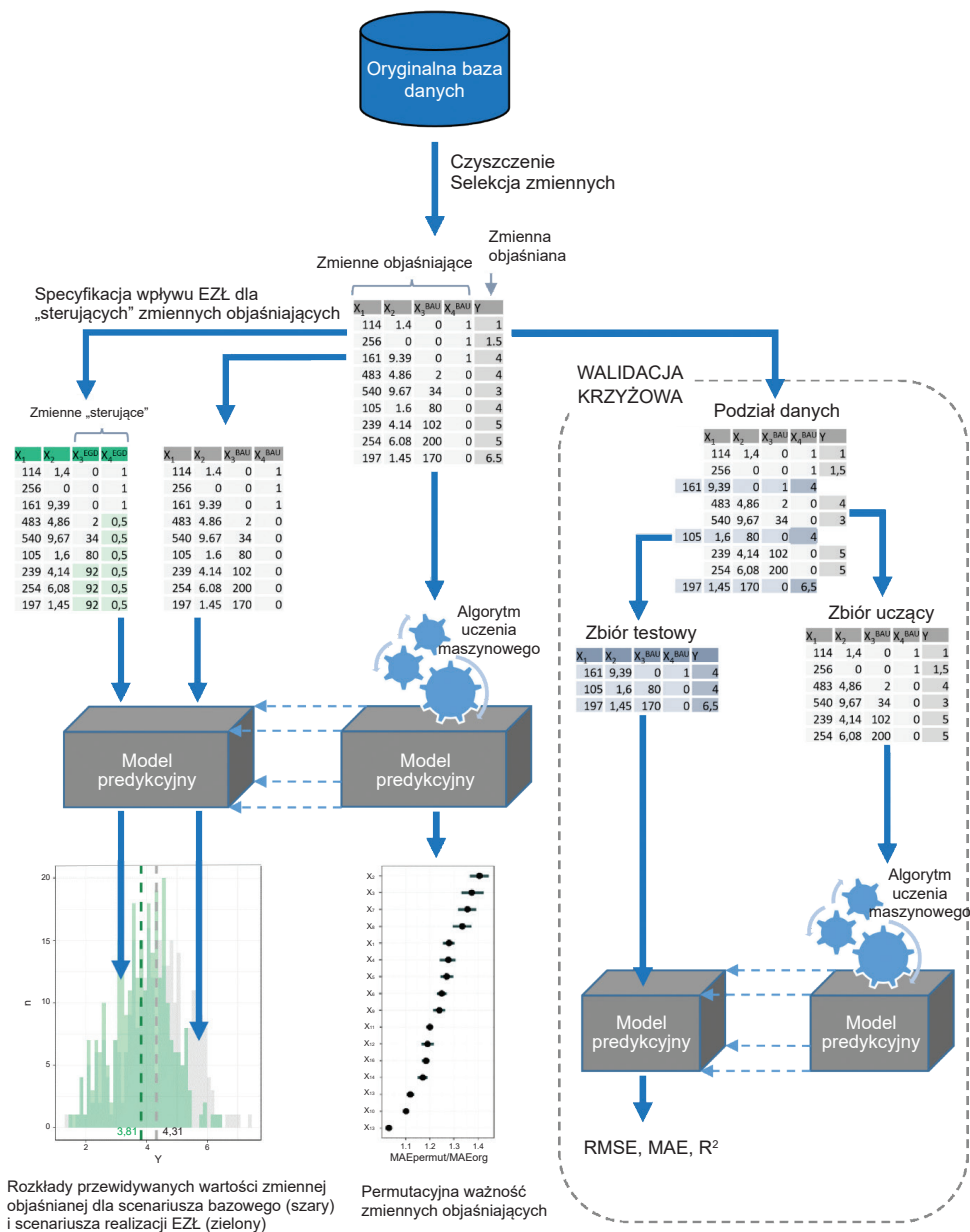
Szybkie tempo wprowadzania zmian w polityce rolnej kontrastuje z raczej powolnym procesem dyskusji, budowy, testowania i uzyskiwania konsensusu naukowego dla modeli wiążących zmiany polityki ze zmianami w środowisku. Przyspieszenie procesów konstrukcji użytecznych modeli jest obecnie możliwe dzięki dynamicznie rozwijającym się metodom sztucznej inteligencji i uczenia maszynowego (3, 31). Dzięki nim model ukierunkowany na prognozowanie, rozpoznawanie określonych wzorców lub podejmowanie właściwych decyzji tworzony jest w sposób automatyczny na podstawie zestawu tzw. danych uczących. Ponadto wybór metod uczenia maszynowego pozwala na ucieczkę od problemu arbitralności decyzji w zakresie konstrukcji algorytmów modeli, co jest szczególnie ważne w kontekście prognozowania efektów decyzji politycznych. Do ograniczeń metod uczenia maszynowego należy ich zależność jedynie od informacji zawartej w zbiorze uczącym z pominięciem większości dotychczasowego dorobku wiedzy dziedzinowej. Metody uczenia maszynowego wymagają więc większego nadzoru nad jakością i reprezentatywnością danych uczących dla badanego problemu, kontroli porównawczej nowych przewidywań z ustaleniami poczynionymi innymi metodami (7) i ostrożnego wnioskowania o relacjach przyczynowości w odniesieniu do użytecznych zmiennych objaśniających (35). Chociaż modele prognostyczne uzyskane metodami uczenia maszynowego są najczęściej trudniejsze w interpretacji niż klasyczne modele statystyczne ograniczenie to można pokonać, stosując tzw. metody agnostyczne (37, 24, 22, 2). Pozwalają one na interpretację wszelkich modeli predykcyjnych, niezależnie od różnic algorytmicznych, co jest szczególnie cenne w analizie porównawczej.

Przewidywanie wpływu EZŁ na gleby i plony będzie więc w przedstawionej pracy okazją do prezentacji zastosowań wybranych narzędzi uczenia maszynowego.

Metodyka

W zastosowanym procesie opracowania prognozy wpływu EZŁ na gleby i plony (rys. 1) wydzielić można pięć faz: 1) przygotowania danych historycznych zawierających zmienne objaśniające oraz zmienną objaśnianą; 2) wyboru i uczenia modeli na zestawie danych historycznych; 3) wykorzystania uzyskanych modeli predyk-

cyjnych do prognozowania wartości zmiennej objaśnianej z danych dla zmiennych objaśniających: a) historycznych w scenariuszu bazowym (SB) i b) syntetycznych różniących się od historycznych jedynie w przypadku zmiennych „sterujących” w scenariuszu realizacji Europejskiego Zielonego Ładu (EZŁ); 4) wykonania walidacji krzyżowej; 5) interpretacji wyników.



Rys. 1. Zastosowany proces prognozowania wpływu EZŁ na gleby i plony

Źródło: opracowanie własne

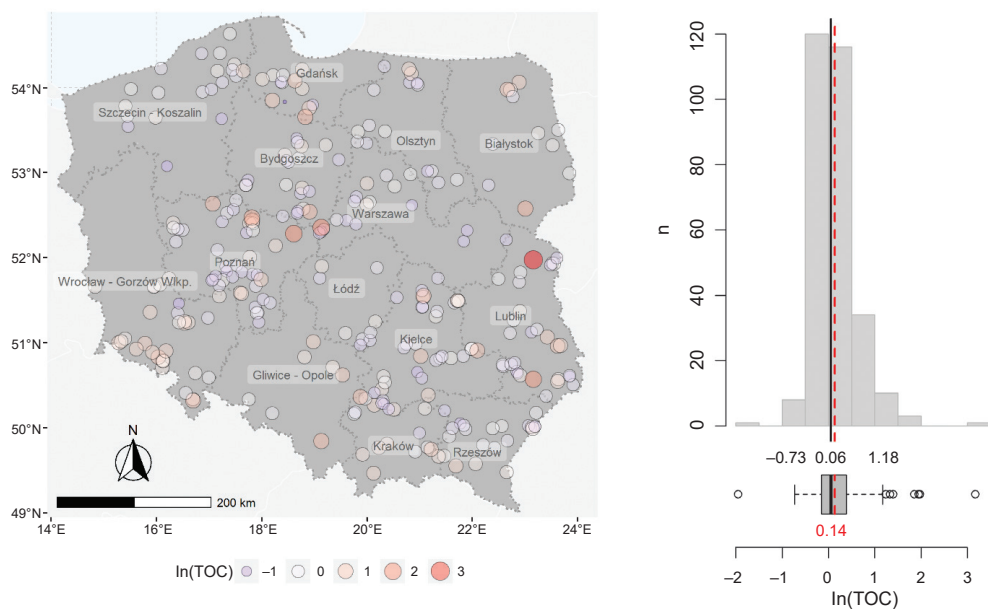
Wybrane fazy procesu prognozowania omówiono bardziej szczegółowo w kolejnych podpunktach.

1. Przygotowanie danych

Do oceny wpływu EZŁ na właściwości gleby i plony wykorzystano dane z monitoringu gleb prowadzonego przez IUNG-PIB na potrzeby oceny działań Programu Rozwoju Obszarów Wiejskich (PROW) w ramach zadania 1.3 PW 2016–2020 (32). Baza zawiera zarówno różnorodne parametry charakteryzujące gleby, jak i informacje o jej uprawie, plonie i stosowanej agrotechnice uzyskane metodą ankietową. W bazie nie zbierano jawnie informacji o poziomie stosowania środków ochrony roślin, jednak zbierano informacje o uprawianych roślinach, co pozwala na powiązanie z poszczególnymi lokalizacjami stosowanej dla wybranej uprawy w kraju przeciętnej dawki pestycydów (10). Ponadto baza umożliwia wskazanie działek, na których realizowano działanie „Rolnictwo Ekologiczne”, co wiąże się z zaniechaniem stosowania nawozów mineralnych i pestycydów. W bazie zawarte są dane z lat 2017 i 2018 odpowiednio dla 536 i 559 działek rozłożonych w przybliżeniu równomiernie na obszarze UR Polski. Ponieważ rodzaj uprawy wpływa na właściwości gleby, zdecydowano, że dalsze analizy dotyczyć będą jedynie zbóż podstawowych (pszenica, jęczmień, pszenżyto, żyto i owies), które są grupą upraw o największym areale w strukturze gruntów ornych w kraju i znacznym podobieństwie w zakresie agrotechniki i fizycznych właściwości plonu. Po usunięciu działek, dla których brakowało kluczowych informacji o lokalizacji, najważniejszych parametrach gleby, agrotechnice lub plonie uzyskano bazę zawierającą pełną informację dla obsianych zbożami podstawowymi 169 działek z roku 2017 i 125 z roku 2018. Wśród wybranych parceli działanie „Rolnictwo Ekologiczne” realizowano na 19 działkach z roku 2017 i 19 z roku 2018. W dalszej analizie dane z lat 2017 i 2018 wykorzystywano łącznie jako bazę dla $N = 294$ działki, a jedyną zmianą zależną od roku był klimatyczny bilans wodny.

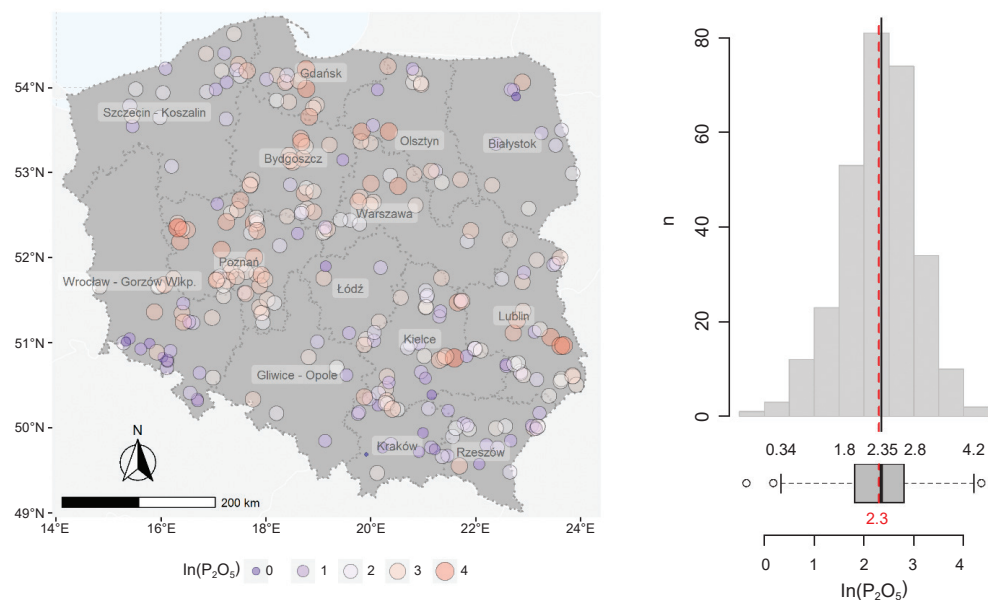
Jako zmienne objaśniane z bazy danych monitoringu PROW wybrano: TOC – zawartość węgla organicznego w glebie (rys. 2), P_2O_5 – zasobność gleb w przyswajalny fosfor (rys. 3), K_2O – zasobność gleb w przyswajalny potas (rys. 4), Mg – zasobność gleb w przyswajalny magnez (rys. 5), Y – plon zbóż podstawowych (rys. 6).

Ponieważ wstępna analiza wykazała, że wartości TOC, P_2O_5 , K_2O i Mg mają rozkład silnie skośny, dlatego dalsze analizy przeprowadzono na wartościach zlogarytmowanych, odpowiednio: $\ln(TOC)$, $\ln(P_2O_5)$, $\ln(K_2O)$ i $\ln(Mg)$. W przypadku jednej próbki, gdzie wartość zmierzona Mg wynosiła zero przyjęto, że wartość rzeczywista to połowa wartości ostatniej cyfry pomiaru, a więc $0,5 \text{ mg} \cdot 100 \text{ g}^{-1}$.



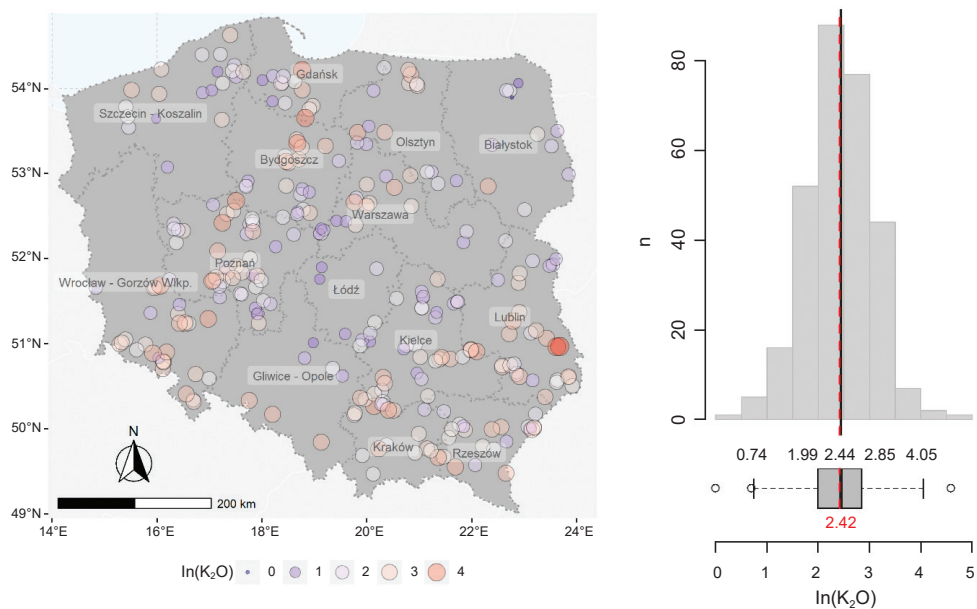
Rys. 2. Zmienność $\ln(\text{TOC})$ w analizowanej próbie działek z lat 2017–2018:
i) mapka (po lewej); ii) rozkład (po prawej)

Źródło: opracowanie własne



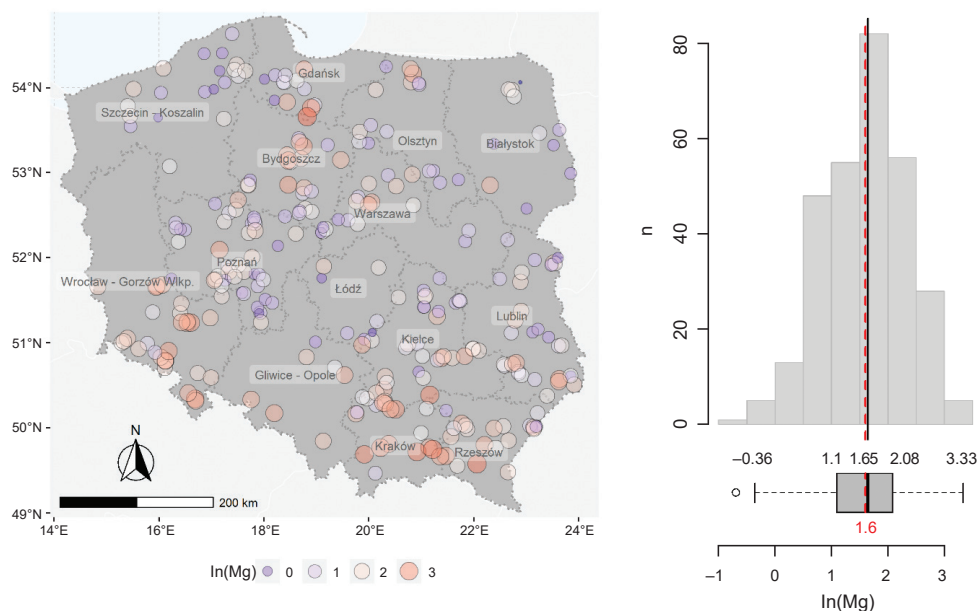
Rys. 3. Zmienność $\ln(\text{P}_2\text{O}_5)$ w analizowanej próbie działek z lat 2017–2018:
i) mapka (po lewej); ii) rozkład (po prawej)

Źródło: opracowanie własne



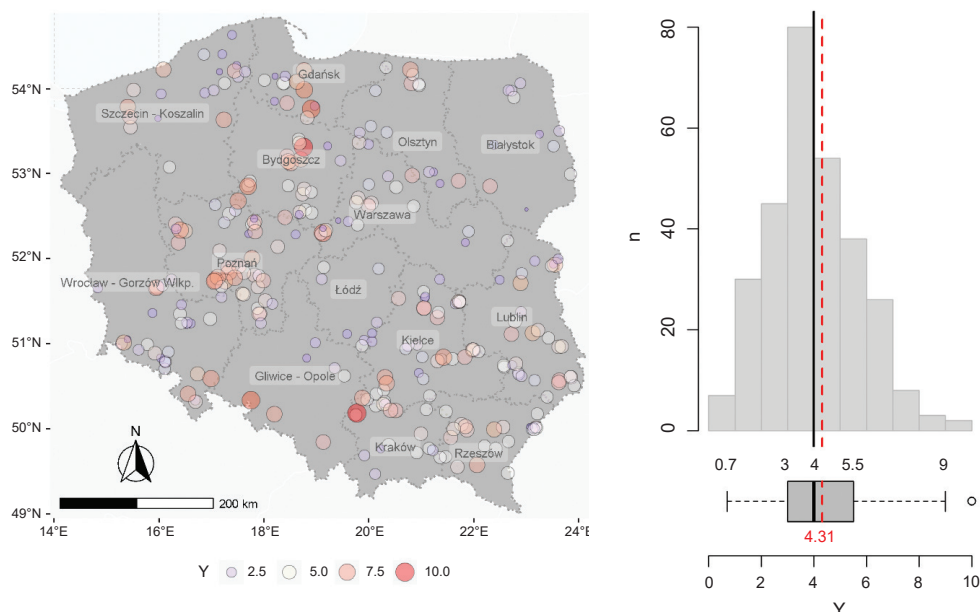
Rys. 4. Zmienność $\ln(K_2O)$ w analizowanej próbie działek z lat 2017–2018:
i) mapka (po lewej); ii) rozkład (po prawej)

Źródło: opracowanie własne



Rys. 5. Zmienność $\ln(Mg)$ w analizowanej próbie działek z lat 2017–2018:
i) mapka (po lewej); ii) rozkład (po prawej)

Źródło: opracowanie własne



Rys. 6. Zmienność plonu Y (t ha⁻¹) w analizowanej próbie działek z lat 2017–2018:
i) mapa (po lewej); ii) rozkład (po prawej)

Źródło: opracowanie własne

Wartości średnie parametrów glebowych w analizowanej próbie działek wynoszą odpowiednio dla: TOC – 1,36%, P_2O_5 – 12,9 mg·100 g⁻¹, K_2O – 13,8 mg·100 g⁻¹ i Mg – 6,34 mg·100 g⁻¹, podczas gdy średnie zmierzone dla 216 lokalizacji z roku 2015 w Monitoringu Chemizmu Gleb Ornych Polski (12) wynoszą odpowiednio dla: TOC – 1,12%, P_2O_5 – 15,4 mg·100 g⁻¹, K_2O – 17,0 mg·100 g⁻¹ i Mg – 7,35 mg·100 g⁻¹. Średni plon zbóż podstawowych na analizowanej próbie działek z lat 2017–2018 wynosił 4,31 t·ha⁻¹, podczas gdy średni plon zbóż podstawowych rejestrowany przez GUS z tego samego okresu dla całego kraju wynosił 3,72 t·ha⁻¹ (9).

W wyborze zmiennych objaśniających kierowano się ich znaczeniem dla procesów glebowych i poziomu plonowania oraz dostępnością przestrzennych baz danych o dostatecznej rozdzielczości. Starano się przy tym ograniczać wprowadzanie zmiennych o podobnym znaczeniu lub w oczywisty sposób silnie skorelowanych. Łączny zestaw 16 wybranych zmiennych objaśniających przedstawiono w tabeli 1.

Tabela 1

Wykorzystane w analizie zmienne objaśniające

	Symbol	Definicja zmiennej objaśniającej i jej wpływ na właściwości gleby oraz plony		Zakres zmienności				Źródło danych
				minimum	mediana	średnia	maximum	
Lokalizacja	LAT	szerokość geograficzna (°)	zmienne dodane jako substytut zmiennych istotnych, lecz pominiętych w modelu, ograniczają niekorzystne efekty niestacjonarności i autokorelacji przestrzennej (1)	49,5	51,7	51,8	54,6	(32)
	LONG	długość geograficzna (°)		14,8	19,8	19,6	23,9	(32)
Rzeźba terenu	ALT	wysokość n.p.m. (m) w rastrze o rozdzielczości 20 m; wpływ na procesy glebowe i ekonomikę produkcji rolnej		4	160	193	587	(8)
	SL	nachylenie terenu (%); obliczone na podstawie zmiennej ALT w rastrze o rozdzielczości 20 m; duże nachylenia ograniczają możliwości mechanizacji uprawy i zwiększają narażenie na erozję wodną		0,0	3,4	6,1	32,4	obliczenia własne na danych (8)
	TWI	topograficzny indeks wilgotności (-); obliczony na podstawie zmiennej ALT w rastrze o rozdzielczości 20 m za pomocą programu SagaGIS wg aktualnych rekomendacji (19); topograficzna składowa wilgotności gleby		5,1	8,3	8,8	18,7	obliczenia własne na danych (8)
Klimat	CWB	klimatyczny bilans wodny (mm) będący różnicą opadu i ewapotranspiracji potencjalnej w okresie od 21 marca do 30 września (4); w zależności od roku pobrania próbki gleby i określenia wysokości plonu dane z roku 2017 lub 2018; klimatyczna składowa wilgotności gleby; niskie wartości redukują plonowanie i przyspieszają mineralizację glebowej materii organicznej		-545	-187	-211	204	obliczenia własne na danych (15)
	GDD5	suma temperatur efektywnych (dzień °C) dla okresu wegetacyjnego (dni z temp. >5°C) będąca sumą dobowych wartości różnic średniej temperatury dziennej i temperatury bazowej równej 5°C; uśredniona dla lat 1971–2015; niskie wartości obniżają stabilność plonowania		1450	1871	1846	2062	(13)

cd. tab. 1

	Symbol	Definicja zmiennej objaśniającej i jej wpływ na właściwości gleby oraz plony	Zakres zmienności				Źródło danych
			minimum	mediana	średnia	maximum	
Gleby	AWC	polowa pojemność wodna gleby (mm) w warstwie korzeniowej obliczona na podstawie danych cyfrowej mapy glebowo-rolniczej w skali 1:25000 i pomiary zdolności retencyjnej gleb (33); wysokie AWC ogranicza skutki suszy	92	164	165	345	(14)
	CLAY	zawartość frakcji ilastej (%) w warstwie ornej gleby; decyduje o strukturze gleby, jej pojemności retencyjnej i sorpcyjnej	0,0	3,0	3,7	11,8	(32)
	pH	odczyn gleby (pH) oznaczony w H ₂ O; wpływa na przyswajalność składników pokarmowych i zanieczyszczeń	4,5	6,3	6,3	8,6	(32)
Użytkowanie	PEST	średni poziom dawki pestycydów (kg·ha ⁻¹) dla poszczególnych zbóż (10); w przypadku realizacji na działce działania „Rolnictwo Ekologiczne” przyjęto wartość 0 (25)	0,00	0,80	0,72	1,30	(32, 10)
	Nmin	poziom nawożenia mineralnego azotem w przeliczeniu na czysty składnik (kg N·ha ⁻¹ UR)	0	80	77	211	(32)
	Pmin	poziom nawożenia mineralnego fosforem w przeliczeniu na czysty składnik (kg P ₂ O ₅ ·ha ⁻¹ UR)	0	40	39	436	(32)
	Kmin	poziom nawożenia mineralnego potasem w przeliczeniu na czysty składnik (kg K ₂ O·ha ⁻¹ UR)	0	50	48	300	(32)
	LU	obsada zwierząt w gospodarstwie (sztuki duże·ha ⁻¹ UR); wpływa na poziom dawek nawozów organicznych	0,00	0,12	0,53	8,08	(32)
	PA	powierzchnia działki rolnej (ha); małe działki są przeważnie użytkowane ekstensywnie i częściej odłogowane	0,4	2,1	4,2	98	(32)

Źródło: opracowanie własne

2. Wybór i uczenie modeli

Analizowany w artykule problem prognozowania wybranej wielkości na podstawie wskazanego zestawu cech jest w dziedzinie uczenia maszynowego klasyfikowany jako uczenie nadzorowane. Ponieważ prognozowana wielkość jest ciągłą, zalicza się je do problemów typu regresja. Do najczęściej stosowanych metod uczenia maszynowego

dla problemów typu regresja należą: regresja liniowa (LM), regresja liniowa regularyzowana (Ridge, LASSO), uogólnione modele liniowe (GLM), drzewa decyzyjne (DT), lasy losowe (RF), algorytm k-najbliższych sąsiadów (k-NN), maszyna wektorów nośnych (SVM) oraz sieci neuronowe (NN). Wśród nich regresja liniowa jest metodą klasyczną i prostą, przez co stanowi dogodny punkt odniesienia dla metod bardziej skomplikowanych. Z tego powodu regresję liniową wybrano jako model referencyjny. Od regresji prostsze w interpretacji są drzewa decyzyjne, lecz pojedyncze drzewo decyzyjne jest podatne na nadmierne dopasowanie do danych uczących i nie sprawdza się dobrze w roli modelu wykrywającego rzeczywiste relacje w danych, tzn. relacje użyteczne w przypadku ekstrapolacji problemu poza obszar danych treningowych. Rozwiązaniem tego ograniczenia jest uśrednianie wielu drzew decyzyjnych trenowanych na niezależnych podzbiorach danych uczących wykorzystane w lasach losowych. Lasy losowe dziedziczą po drzewach decyzyjnych niezależność od skali zmiennych, a ponadto ich trening i testowanie na wielokrotnie losowanych ze zwracaniem podzbiorach niezależnych prowadzi do oszacowania błędu modelu (ang. *Out-of-bag error*) podobnego do uzyskiwanego w teście walidacji krzyżowej. Jakość przewidywań lasów losowych jest zazwyczaj dość wysoka, porównywalna z metodami o dłuższym i bardziej skomplikowanym sposobie uczenia, jak np. sieci neuronowe. Wymienione cechy sprawiają, że lasy losowe stosowane są często jako model pierwszego zastosowania (34). Ich wykorzystanie jest popularne w hydrologii (34), mapowaniu gleb (11, 38) i predykcji plonów (16, 27). Z tego powodu lasy losowe zostały wybrane jako drugi model w prezentowanej analizie. Trening modeli regresji liniowej (LM) i lasów losowych (RF) przeprowadzono w programie R (29) na zestawie danych historycznych odpowiednio: za pomocą funkcji wbudowanej i za pomocą pakietu „randomForest” (21). Lasy losowe trenowano przy ustawieniach domyślnych (ntree = 500, mtry = liczba zmiennych/3).

3. Prognozowanie dla scenariusza bazowego i scenariusza EZŁ

Do prognozowania wartości zmiennej objaśnianej z danych dla zmiennych objaśniających użyto uzyskanych na etapie uczenia modeli predykcyjnych: a) historycznych w scenariuszu bazowym (BAU) i b) syntetycznych różniących się od historycznych jedynie w przypadku zmiennych „sterujących” w scenariuszu realizacji Europejskiego Zielonego Ładu (EZŁ).

W scenariuszu realizacji EZŁ nie uwzględniano działań wspomagających wprowadzanie nowych technologii w gospodarstwach, przyjęto jedynie, że najważniejsze założenia dotyczące nawozów mineralnych i pestycydów mogą być przepisane w formie warunków dotyczących wybranych, wymienionych poniżej w punktach zmiennych objaśniających – „sterujących”:

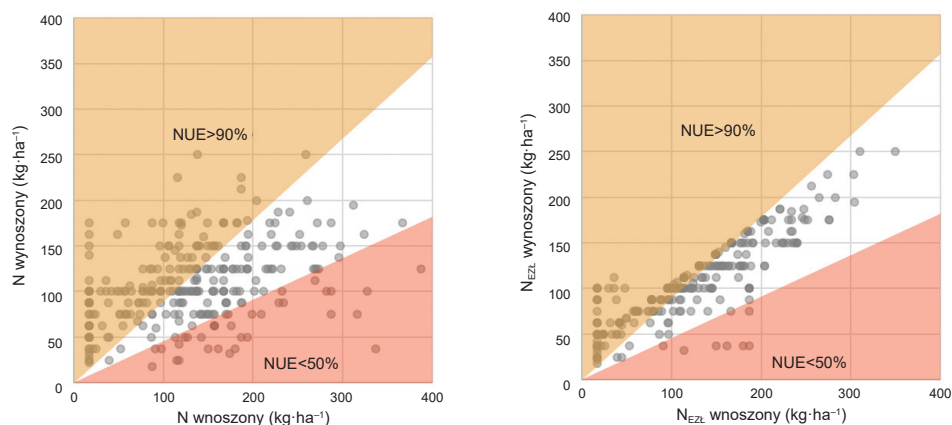
- **PEST** – poziom dawki pestycydów

Zadeklarowana w strategii „od pola do stołu” redukcja zużycia pestycydów o 50% została w Planie Strategicznym przyjętym przez MRiRW częściowo zaimplementowana w realizacji ekoschematów odnoszących się do integrowanej produkcji roślin i ochrony biologicznej, tak że szacowana redukcja zużycia pestycydów w Polsce wyniesie 5% (25). Przyjęto więc, że dla scenariusza realizacji EZŁ dla wszystkich gospodarstw, które nie realizowały pakietu „Rolnictwo Ekologiczne” zmienna PEST wynosić będzie 0,95 wartości zmiennej PEST dla danych aktualnych oraz danych dla scenariusza bazowego. Założenie to oznacza że restrykcje związane z redukcją stosowania pestycydów ponoszone będą przez wszystkie gospodarstwa w tym samym stopniu. Ze względu na szacunkowy charakter danych o poziomie stosowania środków ochrony roślin w poszczególnych lokalizacjach przyjęcie bardziej realistycznych założeń nie jest możliwe.

- **Nmin** – poziom nawożenia mineralnego azotem (rys. 7)

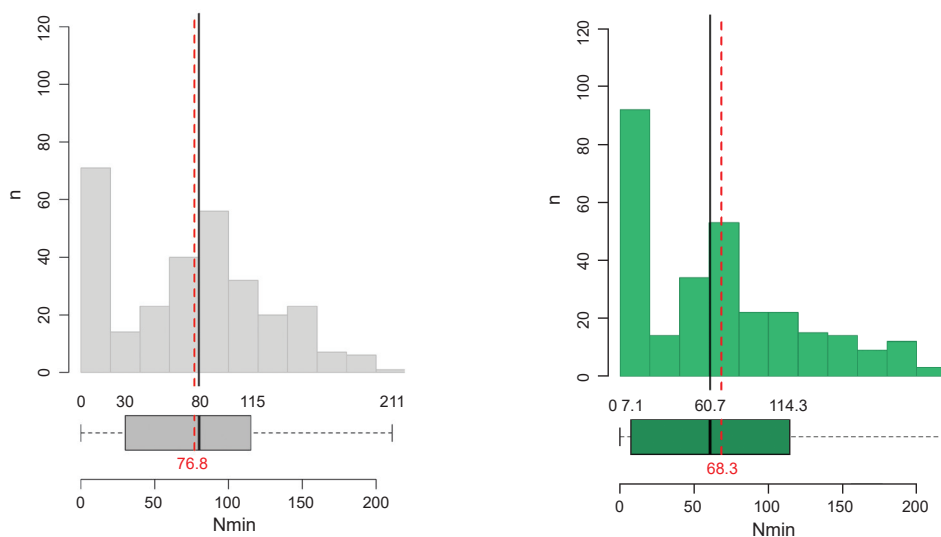
W Krajowym Planie Strategicznym przewidziano liczne interwencje – ekoschematy, przyczyniające się do realizacji zapowiedzianego w strategii „od pola do stołu” celu redukcji strat składników pokarmowych (25). Ponieważ wymienione interwencje są dobrowolne i nie zawierają jawnie zapisanych wymogów redukcji nawożenia, na potrzeby prowadzonej analizy przyjęto, że głównym źródłem redukcji będzie wdrażanie wprowadzonego ustawą w 2020 r. „Programu działań mających na celu zmniejszenie zanieczyszczenia wód azotanami pochodzącymi ze źródeł rolniczych oraz zapobieganie dalszemu zanieczyszczeniu” (5), określanego dalej jako program azotanowy. Pomimo że zapisy programu azotanowego wprowadzające obowiązek sporządzania i stosowania uproszczonego bilansu azotu dla określenia jego dawki w nawozach obejmują gospodarstwa duże (>100 ha) lub gospodarujące intensywnie, to jednak przyjęto, że w perspektywie roku 2030 sposób ten znajdzie powszechne zastosowanie.

Poziom dawek azotu w nawozach mineralnych dla scenariusza EZŁ uzyskano więc w wyniku zastosowania uproszczonego bilansu azotu do wykorzystywanych w analizie danych historycznych przy upraszczającym założeniu, że deklarowany oczekiwany poziom plonów wykorzystywany w obliczeniach uproszczonego bilansu pozostaje bez zmian. Obliczony poziom dawek azotu mineralnego w scenariuszu EZŁ jest średnio równy $68,3 \text{ kg N} \cdot \text{ha}^{-1}$, a więc o 11% niższy od wyjściowego równego $76,8 \text{ kg N} \cdot \text{ha}^{-1}$ (rys. 8), przy czym redukcja strat azotu (bilansu dla lokalizacji, gdzie jest on dodatni) wynosi 57%, a efektywność jego wykorzystania mierzona wskaźnikiem NUE zdefiniowanym jako stosunek azotu wynoszonego w plonie do azotu wnoszonego (26) wzrasta z 81% do 87% (rys. 7).



Rys. 7. Zależność pomiędzy ilością azotu wnoszonego i wynoszonego w uprawie zbóż podstawowych: i) aktualnie i dla scenariusza bazowego (po lewej), ii) po wprowadzeniu EZŁ (po prawej). Kolorami zaznaczono obszary niepożądanego poziomu wykorzystania azotu (26) prowadzącego bądź do zubożenia gleby ($NUE > 90\%$) lub jego nadmiernych strat i zagrożeń środowiskowych ($NUE < 50\%$)

Źródło: opracowanie własne

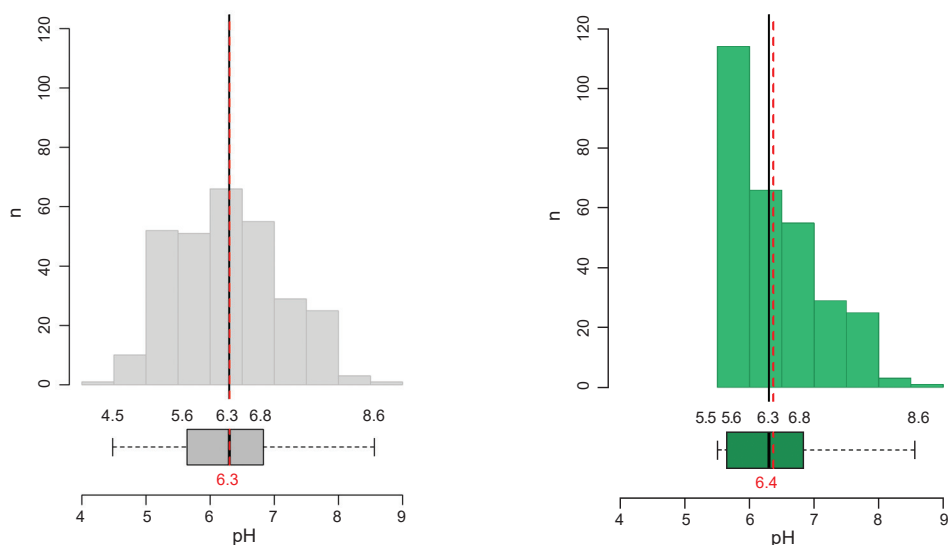


Rys. 8. Rozkłady w analizowanej próbie działek z lat 2017–2018: i) poziomu nawożenia mineralnego azotem (po lewej, kolorem szarym), ii) zakładanego poziomu nawożenia po wprowadzeniu EZŁ (po prawej, kolorem zielonym)

Źródło: opracowanie własne

- **pH** – odczyn gleb (rys. 9)

Przyjęto, że warunek objęcia dopłatami gleb o odczynie pH mniejszym lub równym 5,5 (w KCl) prowadzić będzie do ustabilizowania odczynu gleb objętych dopłatami na poziomie 5,5 (w H_2O).



Rys. 9. Rozkłady w analizowanej próbie działek z lat 2017–2018: i) poziomu pH gleb (po lewej, kolorem szarym), ii) zakładanego poziomu pH po wprowadzeniu EZŁ (po prawej, kolorem zielonym)
Źródło: opracowanie własne

4. Walidacja krzyżowa

Oryginalne błędy modelu regresji liniowej wyznaczone są na danych treningowych, więc są oceniane zbyt optymistycznie, natomiast oryginalne błędy modelu lasów losowych są błędami OOB, a więc są oceniane podobnie restrykcyjnie jak w przypadku walidacji krzyżowej. W celu uzyskania błędów porównywalnych modele poddano 2 testom walidacji krzyżowej:

- **LOO** (ang. *Leave-one-out*) – klasyczny sprawdzian krzyżowy, w którym model trenowany jest na całym zbiorze z pominiętym jednym elementem, a następnie model predykcyjny prognozuje wartość pominiętego elementu; proces jest powtarzany dla wszystkich elementów zbioru. Sprawdzenie jest restrykcyjne, pod warunkiem, że dane nie wykazują silnej autokorelacji, np. przestrzennej (30, 28). Autokorelacja przestrzenna obniża wartość predykcyjną pojedynczych pomiarów, ponieważ informacje w danych z sąsiednich lokalizacji są w znacznym stopniu zreplikowane, co w konsekwencji powoduje niedoszacowanie rzeczywistych błędów modelu (6).
- **CVB** – sprawdzian k-krotny blokowy – z uwagi na dużą wrażliwość wyników pomiarów parametrów glebowych na sposób poboru i przygotowania próbek do analiz, dane podzielono na bloki odpowiadające przestrzennemu zasięgowi działania jednostek pobierających i analizujących próbki gleb (jeden lub kilka OSChR, razem 14 bloków).

Testy walidacji krzyżowej wykonano w programie R za pomocą pakietu „caret” (20).

5. Interpretacja

Interpretując wyniki procesu uczenia modeli, dokonano oceny ich zdolności predykcyjnej, w tym: stopnia, w jakim modele wyjaśniają zmienność cechy (R^2); zmiany pierwiastka błędu średniokwadratowego (RMSE) i średniego błędu bezwzględnego (MEA) pomiędzy walidacją LOO a CVB oraz globalnej autokorelacji reszt mierzonej testem Morana.

Analizując wpływ scenariusza na zmienną objaśnianą, starano się wyjaśniać go pozycją na wykresie ważności zmiennych, a także znakiem współczynnika regresji liniowej dla zmiennych „sterujących”, jeśli zależność pomiędzy zmienną sterującą a objaśnianą jest monotoniczna.

Dla modeli regresji liniowej i lasów losowych porównano permutacyjną ważność zmiennych będącą miarą agnostyczną, tzn. możliwą do zastosowania w dowolnym typie modelu, obliczaną jako iloraz średniego błędu bezwzględnego dla modelu z losowo przesortowaną zmienną objaśniającą i modelu ze zmienną oryginalną. W przypadku regresji liniowej jako miarę ważności zmiennej i informacji o znaku współczynnika regresji podano dodatkowo wartości statystyki testowej t będącej ilorazem wartości współczynnika regresji do jego błędu standardowego.

Modele oceniano pozytywnie, gdy rankingi ważności zmiennych dla obu modeli dla najważniejszych zmiennych były zgodne, ponieważ świadczy to o tym, że wykrywany układ zależności jest wyraźny i poprawnie rozpoznawany niezależnie od modelu.

Dla modeli zmiennych zlogarytmowanych, a więc zmiennych o rozkładzie zbliżonym do rozkładu log-normalnego, zmiany związane z symulacją wprowadzenia EZŁ charakteryzowano za pomocą średniej geometrycznej, mniej podatnej na zmiany nielicznych wysokich wartości zmiennej objaśnianej.

Wykresy ważności zmiennych wykonano w programie R za pomocą pakietu „ggplot2” (36).

Wyniki i dyskusja

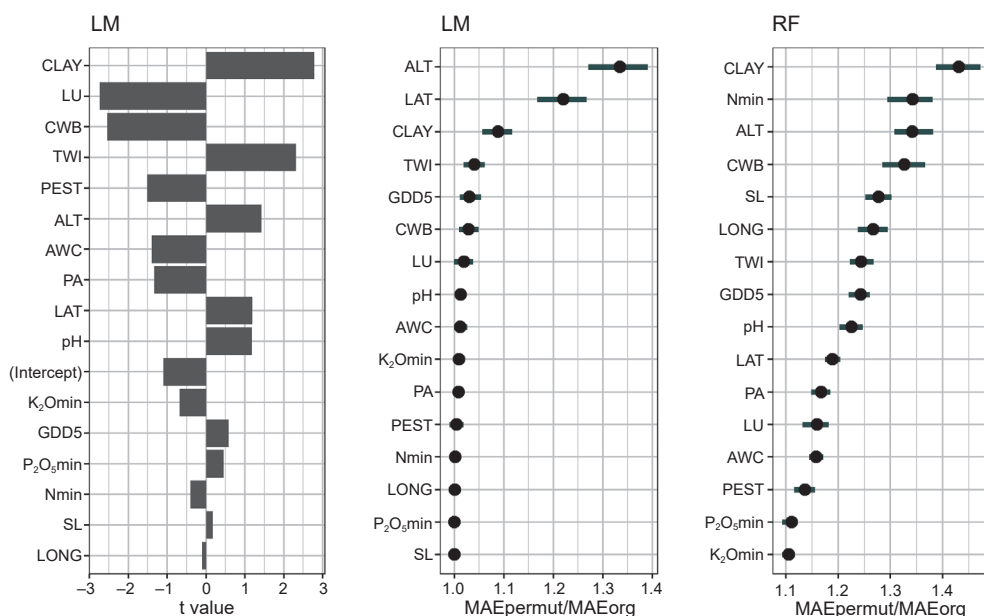
a. TOC – zawartość węgla organicznego w glebie

Wskutek realizacji scenariusza EZŁ średnia geometryczna TOC zmieni się wg modelu regresji liniowej z 1,15% na 1,16% (o 0,93%), a według modelu lasów losowych z 1,16% na 1,17% (o 0,75%). Zarówno model regresji liniowej, jak i las losowy mają niewielką zdolność predykcyjną, a ich reszty wykazują istotną autokorelację (tab. 2). Wśród zmiennych w obu modelach rozpoznanych jako istotne są (rys. 10): zawartość iłu CLAY i wysokość n.p.m. ALT. Las losowy silnie wyróżnił znaczenie nawożenia azotem, co jest przyczyną słabszej reakcji na scenariusz EZŁ.

Tabela 2
Porównanie jakości predykcyjnej modeli dla logarytmu zawartości węgla organicznego $\ln(\text{TOC})$

	Regresja liniowa (LM)			Las losowy (RF)		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Org.	0,45	0,32	0,123	0,47	0,33	0,065
LOO	0,48	0,34	0,041	0,47	0,33	0,069
CVB	0,49	0,36	0,063	0,48	0,36	0,051
Moran's	I = 0,135 p-value = 0,05			I = 0,021 p-value = 0,38		

Źródło: opracowanie własne



Rys. 10. Ważność zmiennych objaśniających w modelach $\ln(\text{TOC})$ wg: i) współczynników t dla regresji (po lewej); ii) metody permutacyjnej dla regresji (środek) oraz iii) metody permutacyjnej dla lasu losowego (po prawej)

Źródło: opracowanie własne

b. P₂O₅ – zasobność gleb w przyswajalny fosfor

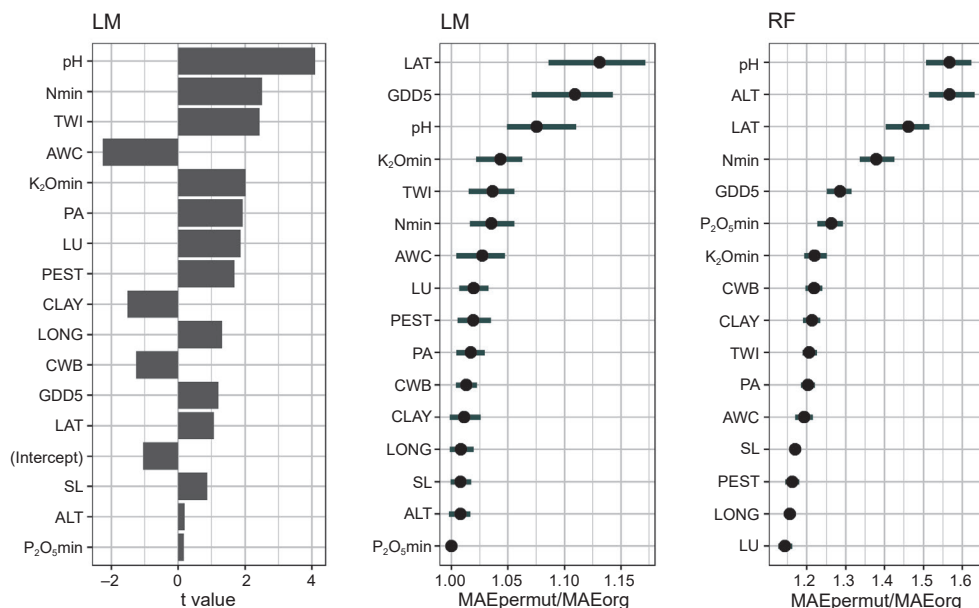
Wskutek realizacji scenariusza EZŁ średnia geometryczna P₂O₅ zmienia się według modelu regresji liniowej z 10,0 na 9,8 mg·100 g⁻¹ (o -1,0%), a według modelu lasów losowych z 9,9 na 9,7 mg·100 g⁻¹ (o -1,9%). Oba modele w testach krzyżowych wyjaśniają około 20% zmienności $\ln(\text{P}_2\text{O}_5)$, a stopień autokorelacji reszt jest istotny lecz dość niski (tab. 3). Wśród najważniejszych wspólnych dla obu modeli zmiennych znajduje się zmienna „sterująca” pH (rys. 11).

Tabela 3

Porównanie jakości predykcyjnej modeli dla logarytmu zawartości przyswajalnego fosforu $\ln(P_2O_5)$

	Regresja liniowa (LM)			Las losowy (RF)		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Org.	0,57	0,44	0,414	0,61	0,46	0,329
LOO	0,60	0,47	0,338	0,61	0,47	0,329
CVB	0,58	0,46	0,201	0,62	0,49	0,155
Moran's	I = 0,027 p-value = 0,36			I = 0,028 p-value = 0,35		

Źródło: opracowanie własne



Rys. 11. Ważność zmiennych objaśniających w modelach $\ln(P_2O_5)$ wg: i) współczynników t dla regresji (po lewej); ii) metody permutacyjnej dla regresji (środek) oraz iii) metody permutacyjnej dla lasu losowego (po prawej)

Źródło: opracowanie własne

c. K_2O – zasobność gleb w przyswajalny potas

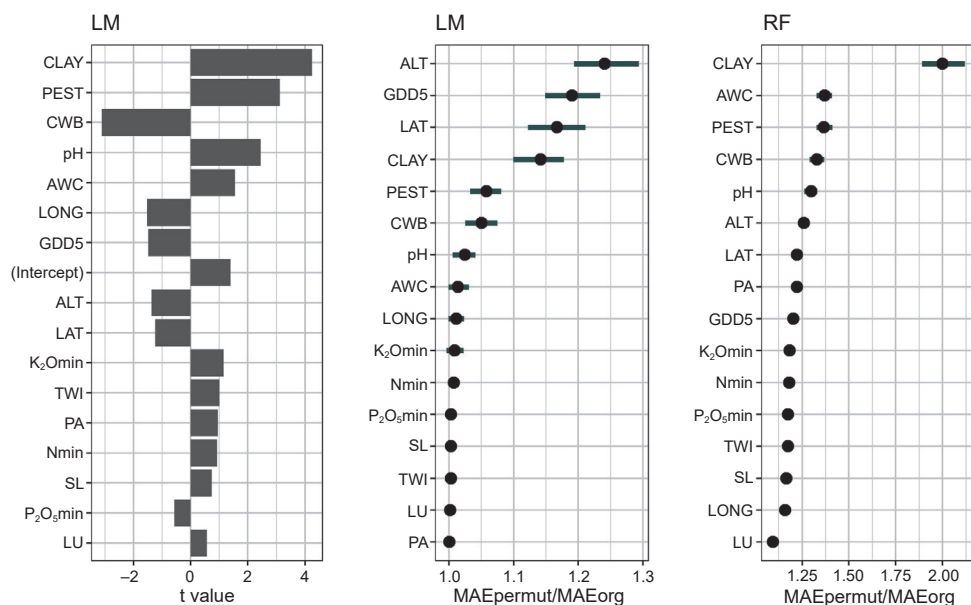
Wskutek realizacji scenariusza EZŁ średnia geometryczna K_2O zmieni się według modelu regresji liniowej z 11,2 na 11,1 $mg \cdot 100 g^{-1}$ (o -0,93%), a według modelu lasów losowych z 11,1 na 11,2 $mg \cdot 100 g^{-1}$ (o 0,85%). W testach krzyżowych model regresji wyjaśnia około 27%, a las losowy około 35% zmienności $\ln(K_2O)$, przy czym reszty tego modelu nie wykazują autokorelacji (tab. 4). Pomimo rozbieżności w rankingach (rys. 12) podobnie istotna jest zawartość iłu CLAY, co znajduje potwierdzenie w rozkładzie tej cechy na mapach (23).

Tabela 4

Porównanie jakości predykcyjnej modeli dla logarytmu zawartości przyswajalnego potasu $\ln(K_2O)$

	Regresja liniowa (LM)			Las losowy (RF)		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Org.	0,53	0,41	0,342	0,53	0,41	0,329
LOO	0,56	0,44	0,266	0,53	0,41	0,339
CVB	0,58	0,46	0,278	0,55	0,44	0,378
Moran's	I = 0,051 p-value = 0,26			I = 0,008 p-value = 0,45		

Źródło: opracowanie własne



Rys. 12. Ważność zmiennych objaśniających w modelach $\ln(K_2O)$ wg: i) współczynników t dla regresji (po lewej); ii) metody permutacyjnej dla regresji (środek) oraz iii) metody permutacyjnej dla lasu losowego (po prawej)

Źródło: Opracowanie własne

d. Mg – zasobność gleb w przyswajalny magnez

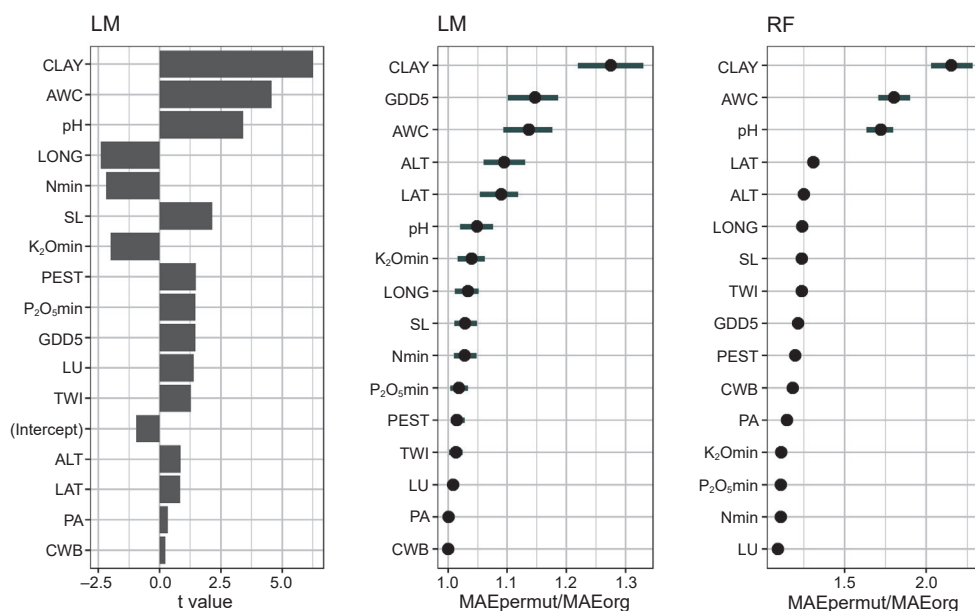
Wskutek realizacji scenariusza EŻŁ średnia geometryczna Mg zmieni się według modelu regresji liniowej z 5,0 na 5,1 $\text{mg} \cdot 100 \text{ g}^{-1}$ (o 1,8%), a według modelu lasów losowych z 4,9 na 5,1 $\text{mg} \cdot 100 \text{ g}^{-1}$ (o 4,3%). Oba modele wyjaśniają około połowy zmienności $\ln(\text{Mg})$ (tab. 5) i zidentyfikowały, że najważniejsze zmienne (rys. 13) to frakcja iłu CLAY, pojemność wodna profilu glebowego AWC i odczyn pH. Wpływ frakcji ilastej na rozkład zmienności Mg jest widoczny na mapach (23). Oba modele

mają ponadto bardzo zgodne pozytywne reakcje na wprowadzenie EZŁ związane z wysoką podobną pozycją zmiennej sterującej pH.

Tabela 5
Porównanie jakości predykcyjnej modeli dla logarytmu zawartości przyswajalnego magnezu $\ln(\text{Mg})$

	Regresja liniowa (LM)			Las losowy (RF)		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Org.	0,51	0,41	0,491	0,51	0,41	0,507
LOO	0,55	0,44	0,424	0,51	0,41	0,516
CVB	0,59	0,47	0,396	0,54	0,44	0,489
Moran's	I = 0,069 p-value = 0,19			I = 0,08 p-value = 0,16		

Źródło: opracowanie własne



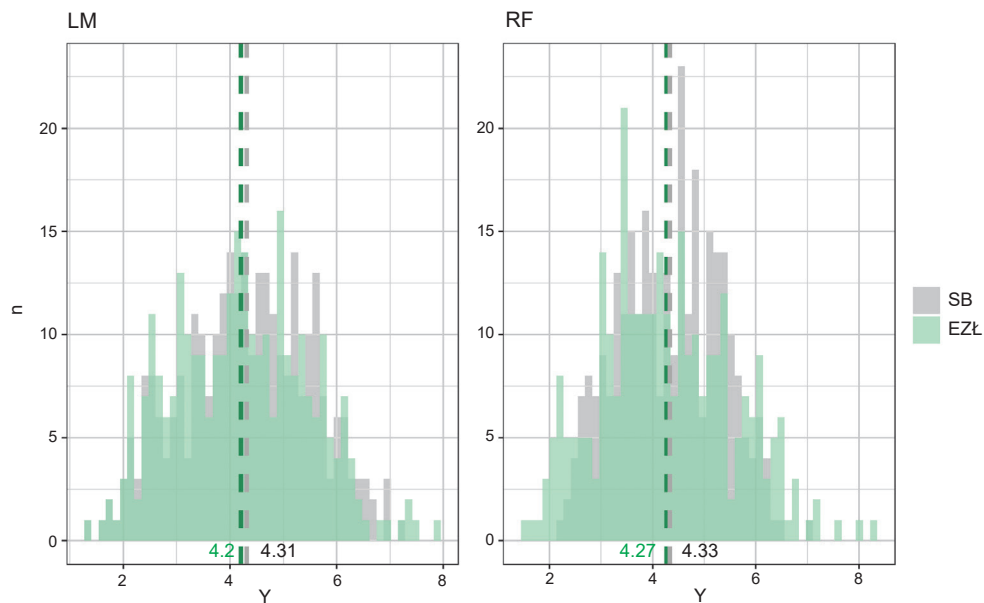
Rys.13. Ważność zmiennych objaśniających w modelach $\ln(\text{Mg})$ wg: i) współczynników t dla regresji (po lewej); ii) metody permutacyjnej dla regresji (środek) oraz iii) metody permutacyjnej dla lasu losowego (po prawej)

Źródło: opracowanie własne

e. Y – plon zbóż podstawowych

Wskutek realizacji scenariusza EZŁ średnia arytmetyczna plonu zbóż podstawowych Y (rys. 14) zmieni się według modelu regresji liniowej z 4,31 na 4,20 t·ha⁻¹ (o -2,4%), a według modelu lasów losowych z 4,33 na 4,27 t·ha⁻¹ (o -1,5%). Oba

modele wyjaśniają około 40% zmienności plonów (tab. 6) i wspólnie identyfikują jako najważniejsze zmienne – poziom stosowania środków ochrony roślin PEST i poziom nawożenia azotem Nmin – główną zmienną „sterującą” odpowiedzialną za ich ujemną reakcję na scenariusz EZŁ (rys. 15).



Rys. 14. Zmiana rozkładu plonów Y pomiędzy scenariuszem bazowym (histogram szary) a scenariuszem realizacji EZŁ (histogram zielony) przewidywana przez modele:
i) regresji liniowej (po lewej); ii) lasu losowego (po prawej)

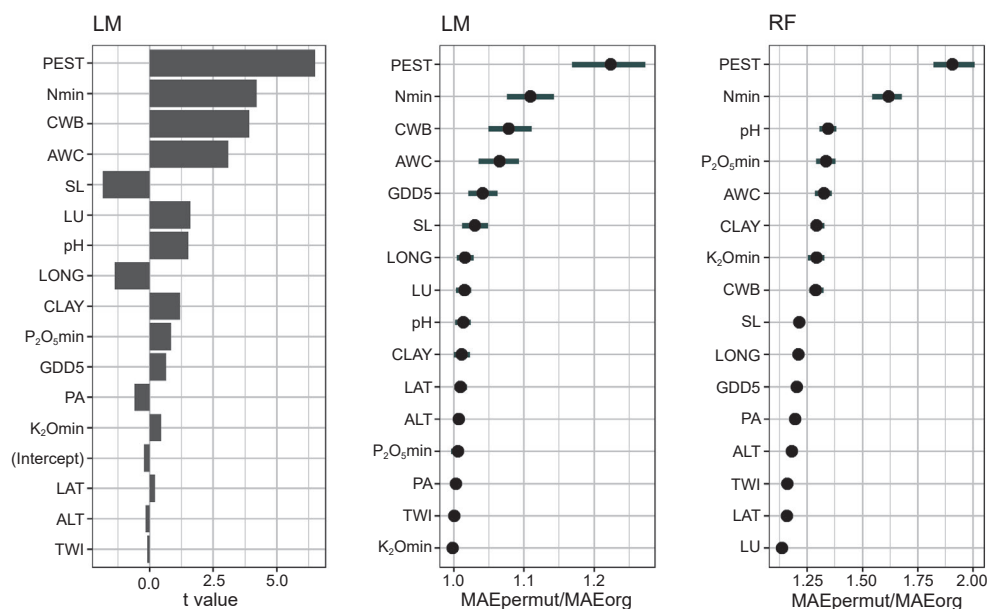
Źródło: opracowanie własne

Tabela 6

Porównanie jakości predykcyjnej modeli dla plonów zbóż podstawowych Y

	Regresja liniowa (LM)			Las losowy (RF)		
	RMSE	MAE	R ²	RMSE	MAE	R ²
Org.	1,26	0,98	0,484	1,34	1,03	0,409
LOO	1,34	1,04	0,419	1,34	1,04	0,417
CVB	1,44	1,11	0,399	1,43	1,12	0,417
Moran's	I = 0,13 p-value = 0			I = 0,151 p-value = 0		

Źródło: opracowanie własne



Rys. 15. Ważność zmiennych objaśniających w modelach plonu Y wg: i) współczynników t dla regresji (po lewej); ii) metody permutacyjnej dla regresji (środek) oraz iii) metody permutacyjnej dla lasu losowego (po prawej)

Źródło: opracowanie własne

W przypadku obu rozpatrywanych modeli o zasobności gleb w przyswajalny potas, fosfor i magnez w największym stopniu decydowała zmienna sterująca – odczyn gleb pH. O zmienności plonu dla obu modeli w największym stopniu decydowały poziomy stosowania pestycydów i PEST oraz nawożenia Nmin. Rola wymienionych czynników jest w przypadku modelowanych właściwości gleb i plonu dobrze znana i zgodna z oczekiwaniami.

Pomimo że model regresji liniowej ma ograniczoną przydatność w modelowaniu zależności nieliniowych, to w przedstawionej analizie przewaga modelu lasów losowych nie była duża, a w przypadku modeli dla materii organicznej i przyswajalnego fosforu model liniowy wykazywał niewielką przewagę predykcijną. Nieliniowości w analizowanych procesach z pewnością są obecne np. w przypadku reakcji plonów zbóż na zawartość frakcji ilastej CLAY w glebie, gdzie plony są niższe zarówno dla gleb bardzo lekkich, jak i bardzo ciężkich, czy też w relacji pomiędzy zastosowaną dawką nawozów Nmin a plonem, gdzie przy wysokich dawkach w doświadczeniach nawozowych dochodzi do wysycenia reakcji na wzrost dawek zgodnie z prawem malejących przychodów. Niewielki wpływ nieliniowości na zróżnicowanie jakości przewidywań testowanych modeli prawdopodobnie wiąże się z: i) ograniczonym, ekonomicznie uwarunkowanym zakresem rzeczywistego wykorzystania gleb o skrajnych właściwościach (np. bardzo lekkich czy bardzo ciężkich); ii) rozrzutem

wyników pomiarów właściwości ґleb, związanym z ich dużą zmiennością lokalną (efekt samorodka); iii) obserwacją plonów w gospodarstwach prowadzących produkcję rynkową, a więc optymalizujących produkcję, co sprawia, że obserwowana zależność pomiędzy dawką czynnika produkcji a plonem nie jest efektem przypadkowych zmian poziomu nawożenia w poszczególnych gospodarstwach, lecz raczej zróżnicowania optymalnego nawożenia pomiędzy gospodarstwami położonymi w różnych warunkach ґlebowych (im lepsze ґleby, tym wyższe nawożenie); iv) zbyt małą bazą danych na potrzeby poprawnej identyfikacji zależności nieliniowych przez typowe metody uczenia maszynowego, jak lasy losowe.

Podsumowanie

Proгнозы впливу wprowadzenia zasad ЕЗЛ на властивості ґleb і плони збóż podstawowych в Polsce wykonane за pomocą дву моделей uczenia maszynowego: regresji liniowej і lasów losowych posiadały zadowalającą zdolność predykcyjną в однієнні до позитивного впливу ЕЗЛ на poziom przyswajального магнезу в ґlebach (wzrost о 2–4%) oraz lekkiego negatywnego oddziaływania на плони (spadek о 1,5–2,4%). Model lasów losowych wykazywał bardzo niewielką przewagę predykcyjną над простым modelem regresji liniowej. Z wyjątkiem modeli dla przyswajального potasu, przewidywania obu modeli były zgodne zarówno co do najważniejszych zmiennych, jak і przewidywanego kierunku zmiany. Niewielkie bezwzględne zmiany przewidywanych cech ґleby і plonów в stosunku do scenariusza bazowego należy wiązać з прийнятым założeniem о niewielkim wymuszeniu в scenariuszu ЕЗЛ redukcji poziomu stosowanych pestycydów і kompensacją spadków plonów powodowanych ograniczeniem nawożenia poprzez efekty wapnowania ґleb.

Literatura

1. Behrens T., Schmidt K., Viscarra Rossel R.A., Gries P., Scholten T., MacMillan R.A.: Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, 2018, **69**: 757-770. <https://doi.org/10.1111/ejss.12687>
2. Bieчек P., Burzykowski T.: *Explanatory Model Analysis: Explore, Explain and Examine Predictive Models*, 1st ed., Chapman and Hall/CRC, 2021, p. 324. <https://doi.org/10.1201/9780429027192> ; <https://pbiecek.github.io/ema/>
3. Breiman L.: Statistical modeling: The two cultures. *Statistical Science*, 2001, **16(3)**: 199-215. <http://www.jstor.org/stable/2676681>.
4. Doroszewski A., Jadczyk J., Kozyra J., Pudełko R., Stuczyński T., Mizak K., Łopatka A., Koza P., Górski T., Wróblewska E.: Podstawy systemu monitoringu suszy rolniczej. *Woda-Środowisko-Obszary Wiejskie*, 2012, t.12, **2(38)**: 77-91.
5. Dziennik Ustaw Rzeczypospolitej Polskiej, 14 lutego 2020, poz. 243. Rozporządzenie Rady Ministrów z dnia 12 lutego 2020 r. в sprawie przyjęcia „Programu działań mających на celu zmniejszenie zanieczyszczenia wód azotanami pochodzącymi ze źródeł rolniczych oraz zapobieganie dalszemu zanieczyszczeniu”. https://eur-lex.europa.eu/legal-content/FI/ALL/?uri=CELEX:71991L0676POL_281852

6. Fletcher R., Fortin M.J.: Spatial ecology and conservation modeling: Applications with R. Springer Nature Switzerland AG, 2018. <https://doi.org/10.1007/978-3-030-01989-1>
7. Greener J.G., Kandathil S.M., Moffat L. et al.: A guide to machine learning for biologists. Nature Reviews Molecular Cell Biology, 2021. <https://doi.org/10.1038/s41580-021-00407-0>
8. GUGiK: Numeryczny model terenu dla Polski o rozdzielczości 5 m.
9. GUS: Rocznik Statystyczny Rolnictwa 2020, Warszawa, 2020. <https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-statystyczny-rolnictwa-2020,6,14.html>
10. GUS: Rolnictwo w 2019 r., Warszawa, 2020. <https://stat.gov.pl/obszary-tematyczne/rolnictwo-lesnictwo/rolnictwo/rolnictwo-w-2019-roku,3,16.html>
11. Hengl T., Nussbaum M., Wright M.N., Heuvelink G.B.M., Gräler B.: Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 2018. <https://doi.org/10.7717/peerj.5518>
12. IUNG, GIOŚ: Monitoring chemizmu gleb ornych Polski, 2015. https://www.gios.gov.pl/chemizm_gleb/
13. IUNG: ONW – Płatności dla obszarów z ograniczeniami naturalnymi lub innymi szczególnymi ograniczeniami, 2017, <http://onw.iung.pulawy.pl/biofizyczne/klimat>
14. IUNG: SMSR – System Monitoringu Suszy Rolniczej, Mapa polowej pojemności wodnej gleb w strefie korzeniowej, 2019. <https://susza.iung.pulawy.pl/>
15. IUNG: SMSR – System Monitoringu Suszy Rolniczej, mapy KBW, 2017 i 2018. <https://susza.iung.pulawy.pl/>
16. Jeong J.H., Resop J.P., Mueller N.D., Fleisher D.H., Yun K., Butler E.E., et al.: Random Forests for Global and Regional Crop Yield Predictions. PLoS ONE, 2016, 11(6). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4892571>
17. KE: Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Komitetu Ekonomiczno-Społecznego i Komitetu Regionów. Europejski Zielony Ład, Bruksela, 2019, COM(2019) 640 final. <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A52019DC0640>
18. KE: Komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady, Komitetu Ekonomiczno-Społecznego i Komitetu Regionów. Strategia „od pola do stołu” na rzecz sprawiedliwego, zdrowego i przyjaznego dla środowiska systemu żywnościowego, Bruksela, 2020, COM(2020) 381 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0381>
19. Kopecký M., Macek M., Wild J.: Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition, Science of The Total Environment, 2021, 757. <https://doi.org/10.1016/j.scitotenv.2020.143785>
20. Kuhn M.: Building predictive models in R using the caret package. Journal of Statistical Software, 2008, 28(5): 1-26. <http://dx.doi.org/10.18637/jss.v028.i05>
21. Liaw A., Wiener M.: Classification and Regression by randomForest. R News 2002, 2(3): 18-22. <https://cran.microsoft.com/snapshot/2018-03-27/web/packages/randomForest/randomForest.pdf>
22. Lucas T.C.D.: A translucent box: interpretable machine learning in ecology. Ecological Monographs, 2020, 90(4). <https://doi.org/10.1002/ecm.1422>
23. Łopatka A., Siebielec G., Smreczak B.: Zasobność gleb w podstawowe składniki oraz zanieczyszczenie gleb i wód azotanami. Studia i Raporty IUNG-PIB, 2020, 64(18): 77-90. <https://doi.org/10.26114/sir.iung.2020.64.05>
24. Molnar C.: Interpretable machine learning. A Guide for Making Black Box Models Explainable, 2019. <https://christophm.github.io/interpretable-ml-book/>
25. MRiRW: Plan Strategiczny dla Wspólnej Polityki Rolnej na lata 2023-2027, Warszawa, lipiec 2021. <https://www.gov.pl/attachment/621ffbec-72ea-4699-8b23-f81cd52971c6>
26. Oenema O., Brentrup F. et al: Nitrogen Use Efficiency (NUE) – an indicator for the utilization of nitrogen in agriculture and food systems. EU Nitrogen Expert Panel, 2015.
27. Partridge T.F., Winter M.W., Liu L., Kendall A.D., Basso B.: Mid-20th century warming hole boosts US maize yields. Environmental Research Letters 2019, 14: 114008. <https://doi.org/10.1088/1748-9326/ab422b>
28. Ploton P., Mortier F., Réjou-Méchain M. et al.: Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications, 2020, 11: 4540. <https://doi.org/10.1038/s41467-020-18321-y>

29. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org/>.
30. Roberts D.R., Bahn V., Ciuti S., Boyce M.S., Elith J., Guillera-Arroita G., Hauenstein S., Lahoz-Monfort J.J., Schröder B., Thuiller W., Warton D.I., Wintle B.A., Hartig F., Dormann C.F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 2017, **40**: 913-929. <https://doi.org/10.1111/ecog.02881>
31. Sarker I.H.: Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, 2021, **2**: 160. <https://doi.org/10.1007/s42979-021-00592-x>
32. Smreczak B.: Program Wieloletni IUNG-PIB na lata 2016-2020; Zadanie 1.3: Monitorowanie różnych parametrów środowiska glebowego dla właściwej oceny WPR, 2017. http://pw.iung.pl/index.php?option=com_content&view=article&id=182Itemid%3D99&Itemid=99
33. Ślusarczyk E.: Określenie retencji użytecznej gleb mineralnych dla prognozowania i projektowania nawodnień. *Melioracje Rolne*, 1979, **3(53)**.
34. Tyralis H., Papacharalampous G., Langousis A.: A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 2019, **11(5)**: 910. <https://doi.org/10.3390/w11050910>
35. Wadoux A.C., Samuel-Rosa A., Poggio L., Mulder V.L.: A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science*, 2020, **71**: 133-136. <https://doi.org/10.1111/ejss.12909>
36. Wickham H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
37. Zhao Q., Hastie T.: Causal Interpretations of Black-Box Models, *Journal of Business & Economic Statistics*, 2021, **39(1)**: 272-281. <https://doi.org/10.1080/07350015.2019.1624293>
38. Zhogolev A., Savin I.: Soil Mapping Based on Globally Optimal Decision Trees and Digital Imitations of Traditional Approaches. *ISPRS International Journal Geo-Information*, 2020, **9(11)**: 664. <https://doi.org/10.3390/ijgi9110664>

Adres do korespondencji:

mgr Artur Łopatka
Zakład Gleboznawstwa Erozji i Ochrony Gruntów
IUNG-PIB
ul. Czartoryskich 8,
24-100 Puławy
tel. 81 47 86 781
email: artur@iung.pulawy.pl

AUTOR	ORCID
Artur Łopatka	0000-0002-6977-4464
Bożena Smreczak	0000-0001-8972-8636
Piotr Koza	0000-0002-0243-7631
Beata Suszek-Łopatka	0000-0002-8398-9912