

Научная статья

Original article

УДК 631.559 : 338.27 : 004.9

doi: https://doi.org/10.55186/2413046X_2026_11_5_69

edn: GKMMYW

**УРОЖАЙНОСТЬ ОСНОВНЫХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ
КУЛЬТУР ПЕНЗЕНСКОЙ ОБЛАСТИ: АНАЛИЗ ПРЕДИКТОРОВ И
СРЕДНЕСРОЧНЫЙ ПРОГНОЗ**
**YIELD OF MAIN AGRICULTURAL CROPS IN THE PENZA REGION:
ANALYSIS OF PREDICTORS AND MEDIUM-TERM FORECAST**



Дубинин Алексей Викторович, аспирант кафедры финансов и информатизации бизнеса, ФГБОУ ВО Пензенский государственный аграрный университет, Пенза, E-mail: dubinin.aleksey@gmail.com

Тишин Максим Евгеньевич, кафедра математического обеспечения и применения ЭВМ, ФГБОУ ВО Пензенский государственный университет, Пенза, E-mail: guliverkoz@gmail.com

Федотова Марина Юрьевна, к.э.н., доцент кафедры финансов и информатизации бизнеса, ФГБОУ ВО Пензенский государственный аграрный университет, Пенза, E-mail: fedotova.m.y@pgau.ru

Dubinin Aleksey Viktorovich, postgraduate student of the Department of Finance and Informatization of Business, Penza State Agrarian University, Penza, E-mail: dubinin.aleksey@gmail.com

Tishin Maxim Evgenievich, Department of Computer Software and Applications, Penza State University, Penza, E-mail: guliverkoz@gmail.com

Fedotova Marina Yuryevna, Candidate of Economic Sciences, Associate Professor, Department of Finance and Business Informatization, Penza State Agrarian University, Penza, E-mail: fedotova.m.y@pgau.ru

Аннотация. В условиях возрастающей климатической нестабильности и политики импортозамещения среднесрочное прогнозирование урожайности приобретает стратегическое значение для регионального агропланирования. Пензенская область, занимающая лидирующие позиции в производстве сахарной свёклы и ряда других культур Приволжского федерального округа, испытывает дефицит верифицированных системных прогностических исследований. Цель работы — разработка и верификация комплекса прогнозных моделей урожайности пяти основных культур Пензенской области (зерновые и зернобобовые, сахарная свёкла, подсолнечник, картофель, овощи) на горизонте до трёх лет. Модели построены на основе Ridge-регрессии с L2-регуляризацией, экспоненциальным взвешиванием наблюдений и схемой валидации «расширяющее окно». Для каждой культуры проводился масштабный сеточный поиск (12000–25000 конфигураций) по пространству из 135 предикторов, включающих региональную статистику, спутниковые индексы MODIS NDVI, агроклиматические показатели и реанализ ERA5-Land за 1990–2025 гг. ($n = 36$). Автоматизированный отбор сформировал культурно-специфичные наборы из 17–31 признака с низким межкультурным перекрытием (коэффициент Жаккара 0,09–0,26) и агрономически интерпретируемым составом. Ретроспективный прогноз показал MAPE от 6,52% (сахарная свёкла) до 15,18% (подсолнечник). Ridge-регрессия превзошла Random Forest и XGBoost по всем культурам на 4,5–9,7 п.п., что подтверждает преимущество регуляризованных линейных моделей при малом объёме выборки. Точечный и интервальный прогноз на 2026–2028 гг. (bootstrap, $N = 2000$) свидетельствует об умеренном росте урожайности зерновых и зернобобовых и сахарной свёклы при стабилизации остальных культур.

Разработанный инструментарий обеспечивает точность прогнозирования от высокой до хорошей (по шкале Lewis) для четырёх из пяти культур и рекомендуется для интеграции в системы среднесрочного планирования регионального АПК — прежде всего для региональных органов управления и крупных агрохолдингов.

Abstract. Against the backdrop of increasing climatic instability and import substitution policies, medium-term crop yield forecasting is acquiring strategic importance for regional agricultural planning. Penza Oblast, which holds a leading position in sugar beet production and several other crops within the Volga Federal District, lacks verified systematic forecasting research. The aim of this study is to develop and validate a suite of predictive yield models for five major crops in Penza Oblast — grains and legumes, sugar beet, sunflower, potato, and vegetables — over a forecasting horizon of up to three years. The models are built on Ridge regression with L2 regularisation, exponential observation weighting, and an expanding window validation scheme. For each crop, an extensive grid search was conducted (12000–25000 configurations) across a space of 135 predictors comprising regional statistical data, MODIS NDVI satellite indices, agroclimatic indicators, and ERA5-Land reanalysis data for the period 1990–2025 ($n = 36$). Automated feature selection produced crop-specific predictor sets of 17–31 variables with low cross-crop overlap (Jaccard coefficient 0.09–0.26) and agronomically interpretable composition. Backtesting yielded MAPE values ranging from 6,52% (sugar beet) to 15,18% (sunflower). Ridge regression outperformed Random Forest and XGBoost across all crops by 4,5–9,7 percentage points, confirming the advantage of regularised linear models under small sample conditions. Point and interval forecasts for 2026–2028 (bootstrap, $N = 2000$) indicate moderate yield growth for grains and sugar beet, with stabilisation projected for the remaining crops. The developed forecasting framework achieves high to good accuracy (per the Lewis scale) for four out of five crops and is recommended for integration into regional medium-term agricultural planning

systems — primarily for regional agricultural authorities and large agro-industrial holdings.

Ключевые слова: прогнозирование урожайности, среднесрочный прогноз, Пензенская область, машинное обучение, гребневая регрессия, агрометеорологические предикторы, NDVI, MAPE, bootstrap-валидация

Keywords: crop yield prediction, medium-term forecast, Penza region, machine learning, Ridge regression, agrometeorological predictors, NDVI, MAPE, bootstrap validation

Введение. Прогнозирование урожайности сельскохозяйственных культур является одной из ключевых задач агрономической науки, имеющей прямое значение для обеспечения продовольственной безопасности, государственного планирования и оптимизации ресурсов. В условиях возрастающей климатической нестабильности и политики импортозамещения особое значение приобретают методы среднесрочного прогнозирования урожайности, позволяющие своевременно корректировать посевные площади и минимизировать риски в АПК.

В последние годы наблюдается значительный рост числа публикаций по данной тематике с применением методов машинного обучения [1, 2]. Наиболее часто используемые методы - «случайный лес» (Random Forest), искусственные нейронные сети, метод опорных векторов, градиентный бустинг, а также методы глубокого обучения. Вместе с тем для задач прогнозирования на региональном уровне методы классического машинного обучения нередко предпочтительнее: они требуют меньших объёмов данных, более интерпретируемы и менее склонны к переобучению [3]. Тема прогнозирования урожайности сельскохозяйственных культур находит отражение и в публикациях российских исследователей. Значимые результаты в этой области были достигнуты Иваньо Я.М., Архиповой М.Ю., Полковской М.Н. и другими.

Пензенская область занимает ведущее место в аграрном производстве Приволжского федерального округа (ПФО) и входит в топ-5 производителей сахарной свёклы в России. Среди работ по данному региону можно выделить следующие. В работе [4] достигнута средняя абсолютная процентная ошибка (MAPE) 9,2% при прогнозировании урожайности озимой пшеницы на основе агроклиматических показателей и NDVI. В работе [5] ошибка аппроксимации временного ряда урожайности сахарной свёклы в Пензенской области составила 13,8%; точность прогнозов MAPE - 0,5% на один год и 20,2% на два года вперед. В работе [6] дан прогноз урожайности масличных культур на 1-4 года вперед; ретроспективная оценка точности этих прогнозов показала MAPE в диапазоне 15–30% в зависимости от года. В работе [7] построены прогнозы валового сбора картофеля в Пензенской области на 2023 и 2024 г.г.; оценка MAPE составила 18,2% и 2,4%, соответственно. Поскольку работы [5-7] не содержат явных оценок точности прогнозов, значения MAPE рассчитаны нами ретроспективно по фактическим данным (ex-post MAPE) от Пензастата. Общая черта данных работ — ограниченный объём данных, простые регрессионные модели и анализ, как правило, одной культуры, что определяет необходимость в специализированных системных прогностических исследованиях.

Целью настоящей работы является разработка и верификация комплекса прогнозных моделей урожайности пяти основных сельскохозяйственных культур Пензенской области — зерновых и зернобобовых, сахарной свёклы, подсолнечника, картофеля и овощей — с использованием методики на основе Ridge-регрессии, описанной в [8]. Для достижения поставленной цели решались следующие задачи: 1) оценка межкультурных взаимосвязей урожайности; 2) определение культурно-специфичных наборов предикторов и количественная оценка их межмодельного сходства; 3) ранжирование предикторов по относительной важности; 4) верификация моделей на основе ретроспективного прогноза и оценка точности на горизонтах 1, 2 и 3 года; 5)

сравнение предложенного подхода с ансамблевыми алгоритмами (Random Forest, XGBoost); 6) построение точечного и интервального прогноза урожайности на 2026–2028 гг.

Материалы и методы. Для построения прогнозных моделей сформирован единый датасет на основе пяти источников, охватывающий период 1990–2025 гг. ($n = 36$). Объект наблюдения — ежегодная урожайность (т/га) пяти основных сельскохозяйственных культур Пензенской области по данным Территориального органа Федеральной службы государственной статистики по Пензенской области (Пензастат). Климатические данные с метеостанции Пенза (WMO 27962) получены из базы данных Всемирной Метеорологической Организации; спутниковые данные — из продукта MODIS MOD13A3 через Google Earth Engine; почвенные характеристики — из реанализа ECMWF ERA5-Land [9]. Исходный датасет включал 230 «сырых» признаков, расширенных до 300 производных; после фильтрации на корреляцию для перебора осталось 135 признаков.

Тесты стационарности (ADF и KPSS) однозначно указали на нестационарность всех рядов урожайности, что обусловило включение временных индексов в спецификацию моделей. Анализ матриц синхронных и лаговых кросс-корреляций выявил: наибольшую синхронную связь между урожайностью сахарной свёклы и картофеля ($r = 0,68$); значимую роль лаговой урожайности подсолнечника как предиктора для зерновых и зернобобовых ($r=0.39$), сахарной свеклы ($r=0.29$) и овощных культур ($r=0.28$), а также авторегрессионные эффекты для зерновых и зернобобовых, подсолнечника и овощей.

Для построения и оценки прогнозных моделей урожайности была использована методика, учитывающая особенности коротких временных рядов ($n=36$) и наличие структурных изменений в агротехнологиях [8]. Отличительными особенностями данной методики являются: использование Ridge-регрессии с L2-регуляризацией, экспоненциальное взвешивание

наблюдений с подбираемым параметром распада, схема валидации на основе «расширяющего окна», двухуровневая иерархическая схема отбора признаков. Подбор признаков и настройка гиперпараметров осуществляется с использованием масштабного сеточного поиска. Разработаны инструментальные программные средства поддержки данной методики на языке Python.

Результаты и их обсуждение. Результатом применения процедуры автоматизированного отбора из 135 кандидатов (12 000–25 000 конфигураций на культуру) стали культурно-специфичные наборы предикторов объёмом 17–31 переменной (таблица 1). В столбце «Культуры» указаны культуры, при прогнозировании урожайности которых указанные предикторы использовались. Приняты следующие сокращения для культур: З — зерновые и зернобобовые культуры; С — сахарная свекла; П — подсолнечник; К — картофель; О — овощи открытого грунта. В столбце «Ед.изм.» указывается единица измерения, диапазон значений или слово «индекс» для (безразмерных) индексов и коэффициентов.

Таблица 1. Результаты структурно-компонентного отбора предикторов в регрессионных моделях урожайности сельскохозяйственных культур

Предиктор	Описание предиктора	Ед. изм.	Культура
I. Спутниковые данные (MODIS NDVI)			
ndvi_jul_aug	Индекс вегетации за июль–август	0...1	С, К
ndvi_m06	Индекс вегетации за июнь	0...1	П
ndvi_m08	Индекс вегетации за август	0...1	К
ndvi_peak	Максимальное (пиковое) значение NDVI за сезон	0...1	З
ndvi_peak_anom	Отклонение пикового NDVI от многолетнего среднего	-0,3...0,3	З
ndvi_veget	Среднее значение NDVI за период вегетации	0...1	З, П, К
II. Агроклиматические показатели (метеостанции и расчетные индексы)			
evaporation	Суммарное фактическое испарение	мм	С, К
GTK	Гидротермический коэффициент Селянинова	индекс	П
heat_x_moisture	Индекс взаимодействия тепла и влаги	индекс	З

mean_temp_jul	Среднемесячная температура воздуха в июле	°С	К
mean_temp_sep	Среднемесячная температура воздуха в сентябре	°С	С, О
mean_temp_summer	Средняя температура воздуха за лето	°С	З
rf_sum_apr	Суммарное количество осадков в апреле	мм	С
rf_sum_spring	Суммарное количество осадков за весну	мм	П
rf_sum_summer	Суммарное количество осадков за лето	мм	З, С
SAT	Сумма активных температур (выше 10°С)	°С	З, С
veget_precip	Суммарные осадки за период вегетации	мм	П
water_deficit	Дефицит водного баланса (испаряемость минус осадки)	мм	К
III. Данные реанализа ECMWF (ERA5-Land)			
era5_spring_recharge	Весенняя влагозарядка почвы	мм	П
pet_mm_m09_abs	Потенциальная испаряемость в сентябре	мм	С
snow_water_eq_mm_m03	Водный эквивалент снежного покрова в марте	мм	З
soil_temp_7_28cm_m06	Температура почвы на глубине 7–28 см в июне	°С	П
IV. Авторегрессионные показатели (лаги и скользящие средние урожайности)			
p_grain_crops_lag1	Урожайность зерновых и зернобобовых в предыдущем году	т/га	О
p_grain_crops_ma3	Средняя урожайность зерновых и зернобобовых за 3 года	т/га	З
p_potato_lag1	Урожайность картофеля в предыдущем году	т/га	О
p_potato_lag2	Урожайность картофеля за два предыдущих года	т/га	К
p_sugar_beet_lag1	Урожайность сахарной свеклы в предыдущем году	т/га	О
p_sugar_beet_ma3	Средняя урожайность сах. свеклы за 3 года	т/га	З
p_sunflower_lag1	Урожайность подсолнечника в предыдущем году	т/га	С, П, О
p_sunflower_ma5	Средняя урожайность подсолнечника за 5 лет	т/га	П
p_vegetables_lag1	Урожайность овощных культур в предыдущем году	т/га	О
p_vegetables_lag2	Урожайность овощных культур за два предыдущих года	т/га	П, О
p_vegetables_ma3	Средняя урожайность овощей за 3 года	т/га	О
p_vegetables_ma5	Средняя урожайность овощей за 5 лет	т/га	О
V. Структурные показатели отрасли (посевные площади)			
log_s_potato_lag1	Логарифм посевной площади картофеля (t-1)	log(тыс.га)	К
s_grain_crops_lag1	Посевная площадь зерновых и	тыс. га	К

	зернобобовых в предыдущем году		
s_potato_lag1	Посевная площадь картофеля в предыдущем году	тыс. га	К, О
s_sugar_beet_lag1	Посевная площадь сахарной свеклы в предыдущем году	тыс. га	С, П
s_sunflower_lag1	Посевная площадь подсолнечника в предыдущем году	тыс. га	С
VI. Временные тренды			
time_index	Временной индекс (порядковый номер года)	ед.	З, С, П, К, О
time_index_sq	Квадратичный временной тренд	ед. ²	С, К, О

Матрица попарного сходства наборов предикторов (таблица 2) свидетельствует о высокой культурно-специфичности моделей: значения коэффициента Жаккара варьируются в диапазоне 0,09–0,26. Наибольшее сходство — у пар «сахарная свёкла — картофель» и «сахарная свёкла — овощи» ($J = 0,26$), наименьшее — у пары «картофель — подсолнечник» ($J = 0,09$).

Таблица 2. Матрица сходства наборов предикторов по культурам (коэффициент Жаккара)

	З	С	П	К	О
З	1.00	0.15	0.15	0.10	0.10
С	0.15	1.00	0.20	0.26	0.26
П	0.15	0.20	1.00	0.09	0.14
К	0.10	0.26	0.09	1.00	0.20
О	0.10	0.26	0.14	0.20	1.00

Анализ стандартизованных коэффициентов регрессии (β^*) и показателей относительной важности предикторов (таблицы 3–7) позволяет верифицировать физико-агрономическую адекватность построенных моделей. Во всех пяти моделях присутствует компонент временного тренда, отражающий долгосрочный технологический прогресс. Символ † означает нестабильный признак (меняет знак между фолдами).

Таблица 3. Коэффициенты регрессии и относительная важность предикторов модели урожайности зерновых и зернобобовых

Предиктор	β^*	Важность, %
time_index	+0.380	17.80
p_sugar_beet_ma3	+0.356	16.92
mean_temp_summer	-0.186	14.90
rf_sum_summer	-0.095	9.43
SAT	-0.107	6.92
ndvi_peak_anom	+0.200	6.23
ndvi_peak	+0.200	6.23
ndvi_veget	-0.254	5.52
heat_x_moisture	+0.143	5.14
snow_water_eq_mm_m03	-0.100	3.07
p_grain_crops_ma3	+0.184	0.54†
Свободный член	2.161	
Суммарная важность		92.7

Таблица 4. Коэффициенты регрессии и относительная важность предикторов модели урожайности сахарной свеклы

Предиктор	β^*	Важность, %
time_index_sq	+0.970	6.18
time_index	+0.797	5.94
pet_mm_m09_abs	-1.037	5.59
ndvi_jul_aug	+0.650	5.48
mean_temp_sep	-0.833	5.30
evaporation	+1.054	5.11
s_sunflower_lag1	+0.645	4.83
s_sugar_beet_lag1	+0.772	4.45
p_sunflower_lag1	+0.968	3.54†
rf_sum_summer	+0.638	1.91†
rf_sum_apr	-0.813	1.73†
SAT	-0.636	0.95†
Свободный член	34.829	
Суммарная важность		51.01

Таблица 5. Коэффициенты регрессии и относительная важность предикторов модели урожайности подсолнечника

Предиктор	β^*	Важность, %
p_vegetables_lag2	+0.178	11.30
soil_temp_7_28cm_m06	+0.154	8.02
s_sugar_beet_lag1	+0.239	7.34†
era5_spring_recharge	-0.091	6.93
time_index	+0.089	6.30
ndvi_veget	+0.092	6.01
rf_sum_spring	+0.135	3.74†
veget_precip	-0.070	3.72†
p_sunflower_lag1	-0.077	2.31†

p_sunflower_ma5	+0.079	2.20†
GTK	-0.072	2.17†
ndvi_m06	+0.064	0.75†
Свободный член	1.127	
Суммарная важность		60.79

Таблица 6. Коэффициенты регрессии и относительная важность предикторов модели урожайности картофеля

Предиктор	β^*	Важность, %
water_deficit	-0.260	8.42
ndvi_m08	+0.293	7.34
ndvi_veget	+0.252	6.76
ndvi_jul_aug	+0.226	6.69
s_grain_crops_lag1	-0.190	6.08
time_index	+0.212	5.40
time_index_sq	+0.199	5.35
p_potato_lag2	+0.240	5.26
mean_temp_jul	-0.192	5.11
s_potato_lag1	-0.163	4.00
log_s_potato_lag1	-0.152	3.76
evaporation	+0.213	3.30
Свободный член	12.553	
Суммарная важность		67.47

Таблица 7. Коэффициенты регрессии и относительная важность предикторов модели урожайности овощей

Предиктор	β^*	Важность, %
p_vegetables_lag1	+0.442	8.79
time_index	+0.431	8.62
time_index_sq	+0.357	7.22
p_sunflower_lag1	+0.342	6.87
p_vegetables_ma3	+0.360	6.73
mean_temp_sep	-0.271	6.44
p_grain_crops_lag1	+0.351	5.79
p_vegetables_ma5	+0.346	5.54
s_potato_lag1	-0.297	5.06
p_vegetables_lag2	+0.277	5.05
p_sugar_beet_lag1	+0.386	4.76†
p_potato_lag1	+0.294	4.17
Свободный член	17.411	
Суммарная важность		75.04

Модели различаются по доминирующим группам предикторов. Для зерновых и зернобобовых наибольший вклад вносят временной тренд и трёхлетнее скользящее среднее урожайности сахарной свёклы как прокси-

индикатор агрохозяйственной интенсивности. Отрицательная связь урожайности с летней температурой и суммой активных температур при положительном коэффициенте индекса тепла–влаги свидетельствует о том, что лимитирующим фактором является не теплообеспеченность, а её соответствие режиму увлажнения. Для сахарной свёклы ключевую роль играет водный баланс: высокое фактическое испарение положительно связано с урожайностью, тогда как сентябрьская жара и высокая испаряемость отражают водный стресс в период сахаронакопления. Модель картофеля характеризуется наибольшим представительством спутниковых данных: три переменных NDVI за июль–август обеспечивают суммарный вклад ~20,8%. Модель подсолнечника демонстрирует наиболее равномерное распределение вкладов между группами предикторов и включает ряд переменных с неочевидной на первый взгляд интерпретацией, в частности лаг урожайности овощных культур как ведущий предиктор (11,30%). Модель овощных культур — наиболее инерционная: восемь из двенадцати предикторов представляют авторегрессионные лаги и скользящие средние, а единственной климатической переменной является средняя температура сентября, что указывает на доминирующую роль агрохозяйственной инерции над погодными условиями.

На рисунках 1-5 представлена динамика фактической и прогнозной урожайности сельскохозяйственных культур в Пензенской области (1990–2025 гг.) и оценка точности ретроспективных прогнозов (2011–2025 гг.). В верхней части рисунков приведены фактические значения урожайности (синяя сплошная линия) и ретроспективные прогнозы модели (красная пунктирная линия). В нижней части рисунков — средняя абсолютная процентная ошибка (MAPE) по годам. Светло-розовыми вертикальными полосами выделены годы с гидротермическим коэффициентом (ГТК) Селянинова меньше 0,70, характеризующие засушливые условия вегетации.

На рисунке 3 виден структурный разрыв в ряде урожайности подсолнечника, обусловленный резкими изменениями в агротехнологиях.

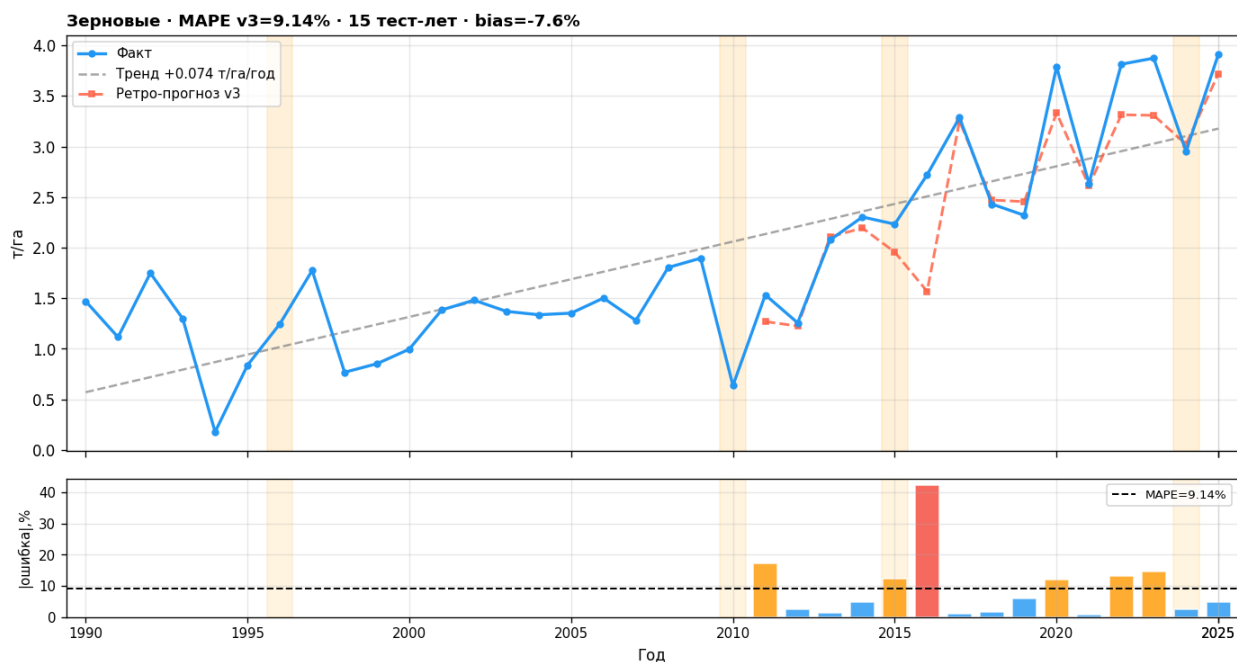


Рисунок 1. Динамика фактической и прогнозной урожайности зерновых и зернобобовых культур с оценкой ошибок ретроспективного тестирования модели.

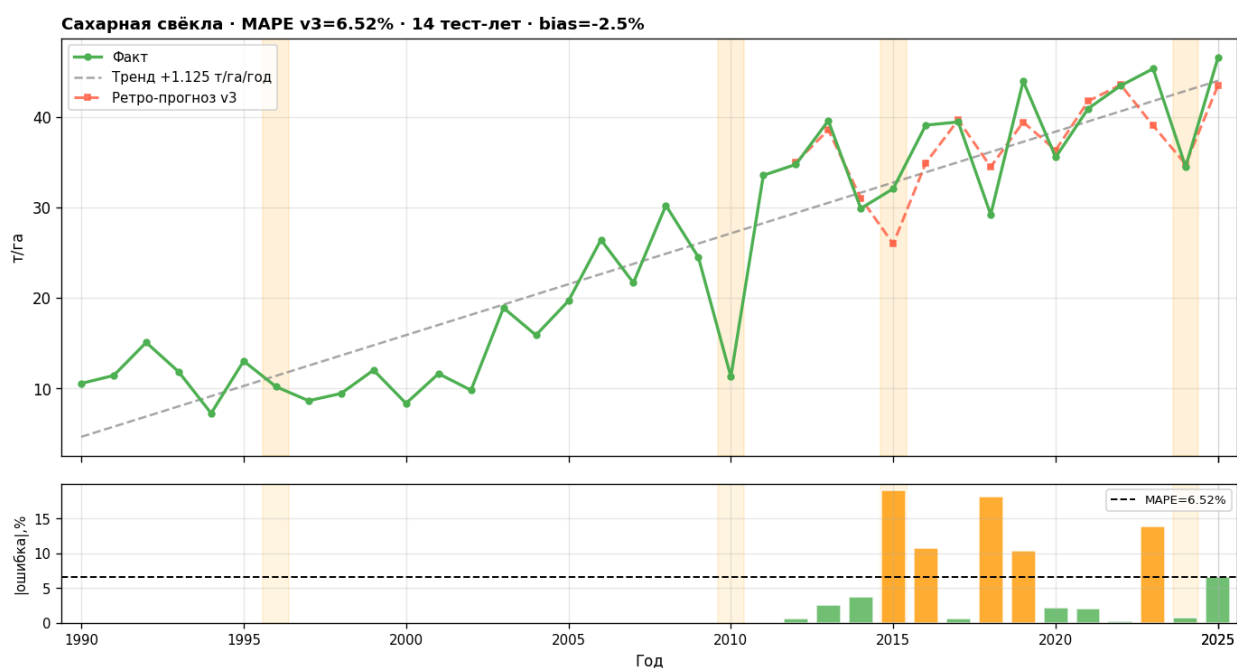


Рисунок 2. Динамика фактической и прогнозной урожайности сахарной свеклы с оценкой ошибок ретроспективного тестирования модели.

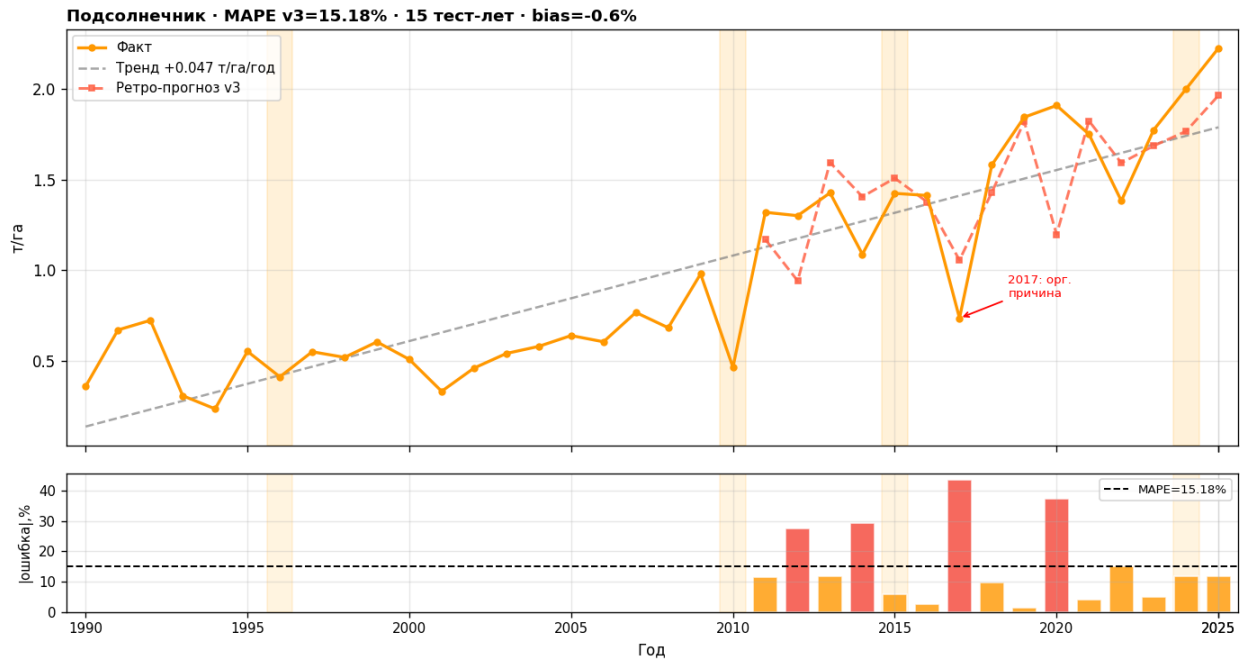


Рисунок 3. Динамика фактической и прогнозной урожайности подсолнечника с оценкой ошибок ретроспективного тестирования модели.

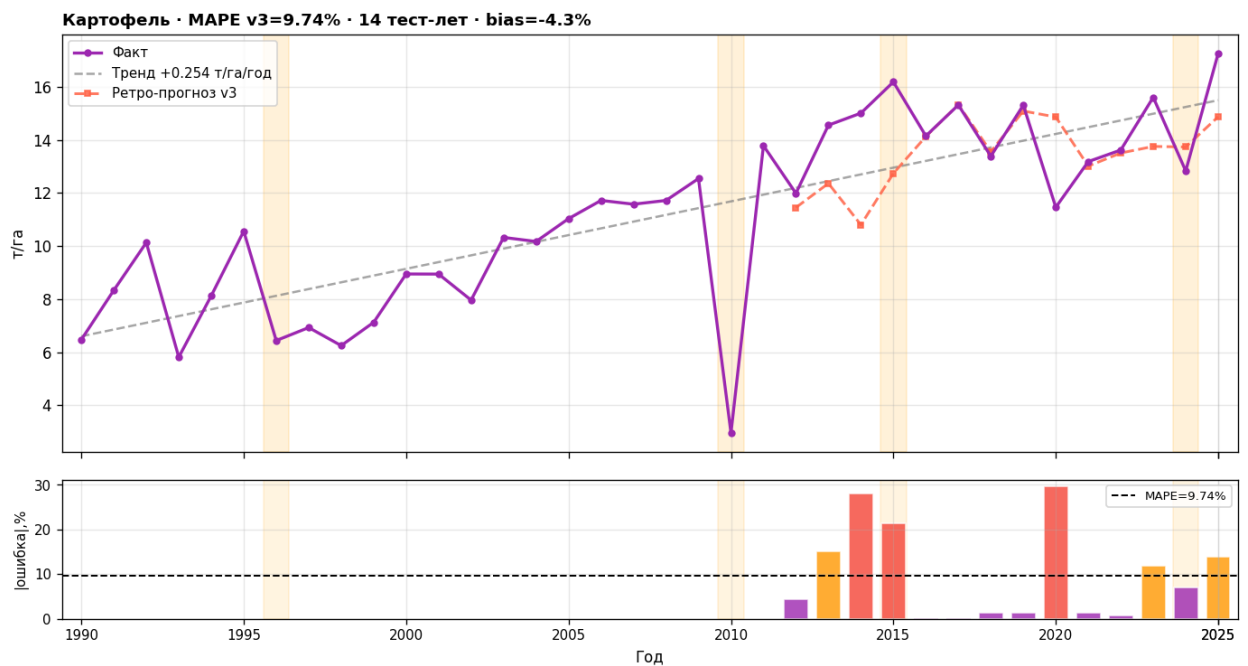


Рисунок 4. Динамика фактической и прогнозной урожайности картофеля с оценкой ошибок ретроспективного тестирования модели.

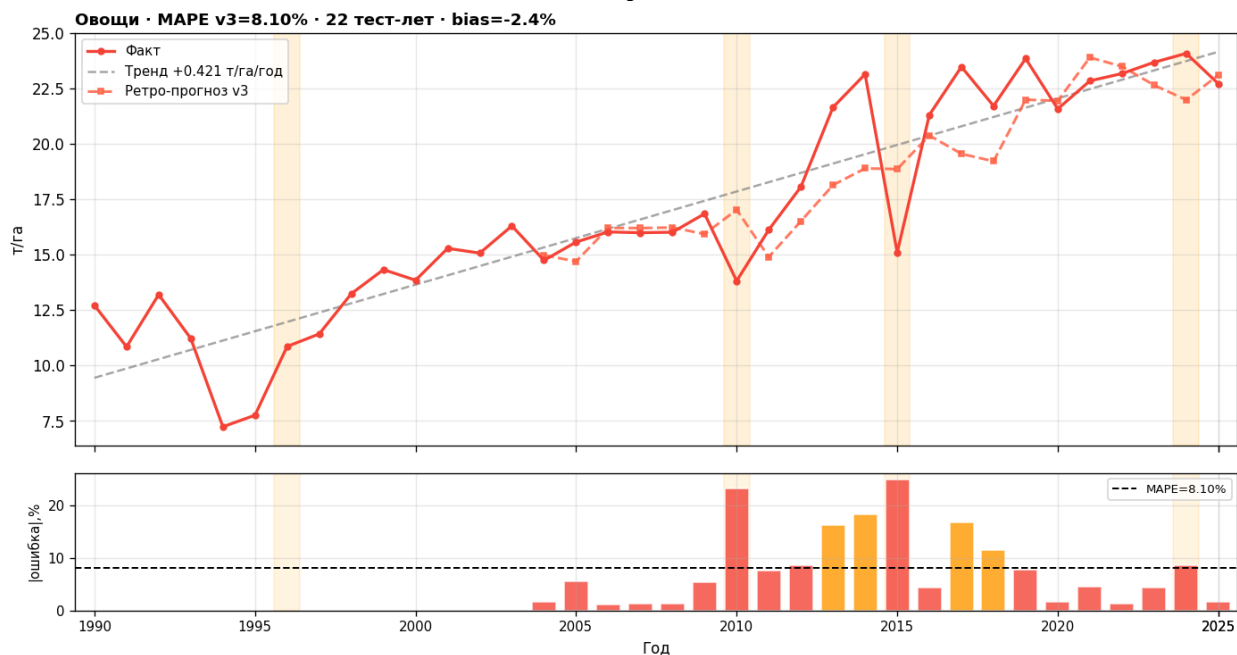


Рисунок 5. Динамика фактической и прогнозной урожайности овощей с оценкой ошибок ретроспективного тестирования модели.

Точность построенных регрессионных моделей оценивалась по показателю MAPE на двух уровнях. Первый — ретроспективный прогноз за период 2011–2025 г.г., полученный методом расширяющего окна (одношаговый бэкстест). Второй — прогностическая точность на горизонтах 1, 2 и 3 года, оценённая методом bootstrap-валидации. Результаты представлены в таблице 8.

Таблица 8. Точность прогнозных моделей урожайности сельскохозяйственных культур на ретроспективном горизонте и горизонтах планирования 1–3 года (MAPE, %)

Культура	Ретроспективный прогноз	1 год	2 года	3 года
Зерновые и зернобобовые	9.14%	18.32%	21.32%	15.80%
Сахарная свёкла	6.52%	12.64%	12.38%	10.95%
Подсолнечник	15.18%	27.83%	31.54%	29.48%
Картофель	9.74%	9.87%	11.36%	13.98%
Овощи	8.10%	12.47%	12.29%	12.46%

Для интерпретации значений MAPE использовалась шкала точности Lewis [10]: менее 10% — высокая точность, 10–20% — хорошая, 20–50% — удовлетворительная. Для четырёх из пяти культур — сахарной свёклы (6,52%), овощей (8,10%), зерновых и зернобобовых (9,14%) и картофеля (9,74%) — MAPE ретроспективного прогноза не превысила 10%. Для подсолнечника — 15,18% (хорошая точность).

При анализе прогнозных горизонтов установлена умеренная деградация точности, типичная для экстраполяционных моделей. Для зерновых и зернобобовых MAPE возрастает до 18,32–21,32% на горизонтах 1–2 года, однако снижается до 15,80% на трёхлетнем горизонте, что может указывать на захват моделью среднесрочной цикличности севооборотов. Картофель демонстрирует исключительную устойчивость: ошибки на горизонтах 1–2 лет (9,87% и 11,36%) близки к бэквесту вследствие высокой силы авторегрессионных лагов. Для овощей и сахарной свёклы MAPE стабилизируется на уровне 12,4% и снижается до 10,95% на третьем году соответственно, что отражает устойчивость долгосрочных трендовых компонентов. Наиболее выраженная деградация — у подсолнечника: MAPE достигает 27,83–31,54% на горизонтах 1–2 года (удовлетворительная точность), что обусловлено высокой чувствительностью к погодным условиям вегетационного периода.

Полученные значения MAPE сопоставимы с медианными показателями крупной европейской системы прогнозирования урожайности сельскохозяйственных культур MARS (2006–2015 г.г.): твёрдая пшеница — 10,15%, сахарная свёкла — 6,81%, картофель — 6,31%, подсолнечник — 8,93% [11]. Следует учитывать, что распределение MAPE логнормально, поэтому средние значения будут превышать приведенные медианные на 1–3%.

В целях верификации предложенного подхода проведена сравнительная оценка точности Ridge-регрессии и двух ансамблевых алгоритмов —

«случайного леса» (Random Forest) и градиентного бустинга (XGBoost) — на тех же наборах предикторов (таблица 9). Результаты демонстрируют стабильное превосходство Ridge по всем пяти культурам.

Таблица 9. Сравнительный анализ точности Ridge-регрессии и ансамблевых алгоритмов машинного обучения (по метрике MAPE, %)

Культура	Ridge	Random Forest	XGBoost	Преимущество Ridge, п.п.
Зерновые и зернобобовые	9.14	~14.2	~16.8	+5.1 / +7.7
Сахарная свёкла	6.52	~12.5	~14.1	+6.0 / +7.6
Подсолнечник	15.18	~21.3	~23.6	+6.1 / +8.4
Картофель	9.74	~17.8	~19.4	+8.1 / +9.7
Овощи	8.10	~12.6	~14.9	+4.5 / +6.8

Полученное соотношение результатов не является неожиданным. Ансамблевые методы требуют значительного объёма обучающих данных; при малом n они склонны к переобучению даже при кросс-валидации [12]. Ridge-регрессия с L2-регуляризацией целенаправленно вводит смещение ради снижения дисперсии — механизм, особенно эффективный при ограниченной выборке. Эмпирические свидетельства зафиксированы в прикладных исследованиях: Ridge демонстрировала устойчивую точность ($R^2 > 0,90$) на коротких временных рядах, тогда как ансамблевые методы утрачивали её при сокращении данных [13]. Аналогичные выводы получены в контексте прогнозирования урожайности: линейные регрессионные модели со спутниковыми индексами способны конкурировать с алгоритмами машинного обучения при работе с малыми наборами данных на региональном уровне [14, 15]. Таким образом, превосходство Ridge-регрессии обусловлено не принципиальными ограничениями ансамблевых методов, а несоответствием их вычислительных требований характеристикам доступных данных: малым n , высокой размерностью признакового пространства и временной структурой наблюдений.

На основе построенных регрессионных моделей методом bootstrap (N = 2000) получен прогноз урожайности пяти культур на период 2026–2028 гг. В качестве точечной оценки использована медиана bootstrap-распределения, интервальная оценка представлена 80% доверительным интервалом (ДИ). Результаты приведены в таблице 10.

Таблица 10. Прогноз показателей урожайности сельскохозяйственных культур на 2026–2028 гг., полученный методом Bootstrap (N=2000), т/га

Культура	2026	2027	2028	80% ДИ 2026	80% ДИ 2027	80% ДИ 2028
Зерновые и зернобобовые	3.59	3.73	3.89	[3.08-4.43]	[3.22-4.54]	[3.38-4.70]
Сахарная свёкла	44.61	45.54	45.83	[37.18-47.89]	[38.11-48.71]	[38.31-48.99]
Подсолнечник	2.06	2.06	2.15	[1.80-2.25]	[1.77-2.22]	[1.88-2.35]
Картофель	14.74	15.03	15.00	[14.11-15.76]	[14.40-16.08]	[14.34-16.02]
Овощи	24.94	24.64	25.02	[23.43-25.68]	[23.26-25.58]	[23.57-25.99]

Прогнозные значения урожайности для всех культур характеризуются стабильной или умеренно-восходящей динамикой. Наиболее выраженный рост демонстрируют зерновые и зернобобовые: с 3,59 т/га в 2026 г. до 3,89 т/га в 2028 г. (+8,4%), при этом ширина 80% ДИ (~1,35 т/га, или ~37% от медианы) отражает существенную межгодовую изменчивость урожайности данной группы. Сахарная свёкла прогнозируется с ростом на 2,7% (44,61→45,83 т/га); доверительный интервал асимметричен: нижняя граница отстоит от медианы на 7,3–7,5 т/га, верхняя — лишь на 3,2–3,3 т/га, что указывает на бóльшую вероятность умеренных сценариев по сравнению с рекордными. Подсолнечник демонстрирует практически горизонтальный прогноз (2,06–2,15 т/га), характеризующий исчерпание текущего технологического потенциала роста. Картофель и овощи образуют группу наиболее уверенно прогнозируемых культур: ширина их 80% ДИ составляет около 11% и 9–10% от медианы соответственно, а значения урожайности стабилизируются на достигнутом уровне (14,74–15,03 т/га и 24,64–25,02 т/га, соответственно). Наибольшая прогностическая неопределённость

сохраняется для зерновых и зернобобовых культур, наименьшая — для картофеля и овощей.

Выводы. В результате исследования разработан и верифицирован комплекс регрессионных моделей прогнозирования урожайности пяти основных сельскохозяйственных культур Пензенской области на горизонте до трёх лет. Подводя итоги, можно сформулировать следующие ключевые выводы.

1. Методика на основе Ridge-регрессии [8] показала высокую эффективность в условиях, типичных для региональных агроэкономических данных: малой выборки, высокой размерности признакового пространства и нестационарности временных рядов.
2. Автоматизированный отбор предикторов сформировал культурно-специфичные наборы переменных с низким межкультурным перекрытием и агрономически интерпретируемым составом, что свидетельствует о содержательной валидности применённой процедуры.
3. Точность ретроспективного прогноза соответствует категории высокой для четырёх из пяти культур и хорошей — для подсолнечника, что сопоставимо с результатами крупных европейских систем прогнозирования урожайности.
4. Деграция точности на горизонтах 1–3 года умеренна для большинства культур и выражена лишь для подсолнечника, прогнозирование которого остаётся наиболее сложной задачей вследствие высокой метеорологической изменчивости.
5. Регуляризованная линейная модель стабильно превзошла ансамблевые алгоритмы по всем культурам, что согласуется с теоретическим преимуществом Ridge при неблагоприятном соотношении объёма выборки и размерности признакового пространства.
6. Прогноз на 2026–2028 гг. характеризуют агропромышленный комплекс Пензенской области как устойчивую систему с умеренно-восходящими тенденциями по большинству культур.

7. Практическая ценность разработанных моделей наиболее высока для задач среднесрочного агропланирования на региональном уровне.
8. Перспективные направления исследований: повышение точности прогноза для подсолнечника, тестирование методики на данных других регионов ПФО и создание системы автоматического обновления моделей.

СПИСОК ИСТОЧНИКОВ

1. Shawon S.M. et al. Crop yield prediction using machine learning: An extensive and systematic literature review // Smart Agricultural Technology. – 2025. – Vol. 10. – Art. 100718. – DOI: <https://doi.org/10.1016/j.atech.2024.100718>.
2. Schauburger B., Jägermeyr J., Gornott C. A systematic review of local to regional yield forecasting approaches and frequently used data resources // European Journal of Agronomy. – 2020. – Vol. 120. – Art. 126153. – DOI: <https://doi.org/10.1016/j.eja.2020.126153>.
3. Jorvekar P.P., Wagh S.K., Prasad J.R. Predictive modeling of crop yields: a comparative analysis of regression techniques for agricultural yield prediction // Agricultural Engineering International: CIGR Journal. – 2024. – Vol. 26, No. 2. – P. 125–140.
4. Щербаков А.С., Богомазов С.В. Прогнозирование урожайности озимой пшеницы на основе агроклиматических показателей и вегетационного индекса NDVI в условиях Пензенской области // Нива Поволжья. – 2025. – № 3(75). – С. 10–11. – DOI: 10.36461/NP.2025.75.3.021.
5. Тиндова М.Г. Анализ динамики выращивания сахарной свеклы в РФ // Экономико-математические методы анализа деятельности предприятий АПК: материалы VII Международной научно-практической конференции. – Саратов, 2023. – С. 306–312.
6. Гурьянова Н.М., Майоркина Е.В. Прогнозирование валового сбора масличных культур Пензенской области // Сурский вестник. – 2020. – № 1(9). – С. 56–61.

7. Самандарзода И.Х. и др. Современное состояние и перспективы развития производства картофеля в Пензенской области // Нива Поволжья. – 2023. – № 3(67). – с. 4002. – DOI: 10.36461/NP.2023.67.3.021.
8. Тишин М.Е., Дубинин А.В. Методологический подход к прогнозированию урожайности сельскохозяйственных культур на основе Ridge-регрессии (на примере Пензенской области) // Сурский вестник. – 2026. – № 6. – (в печати).
9. Hersbach H. The ERA5 global reanalysis / H. Hersbach et al. // Quarterly Journal of the Royal Meteorological Society. – 2020. – Vol. 146. – P. 1999–2049. – DOI 10.1002/qj.3803.
10. Lewis C.D. Industrial and business forecasting methods. – London : Butterworth Scientific, 1982. – 143 p.
11. Velde M., Nisini L. Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015 // Agricultural Systems. – 2019. – Vol. 168. – P. 203–212. – DOI: <https://doi.org/10.1016/j.agsy.2018.06.009>.
12. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – 2nd ed. – New York : Springer, 2009. – 745 p. – DOI: <https://doi.org/10.1007/978-0-387-84858-7>.
13. Valenko D. et al. Agroclimatic Forecasting Under Degraded Sensor Data: A Robustness Benchmark of Machine-Learning Models // Applied Sciences. – 2025. – Vol. 16, No. 10. – Art. 5075. – DOI: <https://doi.org/10.3390/app16105075>.
14. Meroni M. et al. Yield forecasting with machine learning and small data: What gains for grains? // Agricultural and Forest Meteorology. – 2021. – Vol. 308–309. – Art. 108555. – P. 1–13. – DOI: <https://doi.org/10.1016/j.agrformet.2021.108555>.
15. Paudel D. et al. Machine learning for large-scale crop yield forecasting // Agricultural Systems. – 2021. – Vol. 187. – Art. 103016. – DOI: [10.1016/j.agsy.2020.103016](https://doi.org/10.1016/j.agsy.2020.103016).

References

1. Shawon S.M. et al. Crop yield prediction using machine learning: An extensive and systematic literature review // Smart Agricultural Technology. – 2025. – Vol. 10. – Art. 100718. – DOI: <https://doi.org/10.1016/j.atech.2024.100718>.
2. Schauburger B., Jägermeyr J., Gornott C. A systematic review of local to regional yield forecasting approaches and frequently used data resources // European Journal of Agronomy. – 2020. – Vol. 120. – Art. 126153. – DOI: <https://doi.org/10.1016/j.eja.2020.126153>.
3. Jorvekar P.P., Wagh S.K., Prasad J.R. Predictive modeling of crop yields: a comparative analysis of regression techniques for agricultural yield prediction // Agricultural Engineering International: CIGR Journal. – 2024. – Vol. 26, No. 2. – P. 125–140.
4. Shherbakov A.S., Bogomazov S.V. Prognozirovanie urozhajnosti ozimoy pshenicy na osnove agroklimaticheskix pokazatelej i vegetacionnogo indeksa NDVI v usloviyax Penzenskoj oblasti // Niva Povolzh`ya. – 2025. – № 3(75). – S. 10–11. – DOI: 10.36461/NP.2025.75.3.021.
5. Tindova M.G. Analiz dinamiki vy`rashhivaniya saxarnoj svekly` v RF // E`konomiko-matematicheskije metody` analiza deyatel`nosti predpriyatij APK: materialy` VII Mezhdunarodnoj nauchno-prakticheskoy konferencii. – Saratov, 2023. – S. 306–312.
6. Gur`yanova N.M., Majorkina E.V. Prognozirovanie valovogo sбора maslichny`x kul`tur Penzenskoj oblasti // Surskij vestnik. – 2020. – № 1(9). – S. 56–61.
7. Samandarzoda I.X. i dr. Sovremennoe sostoyanie i perspektivy` razvitiya proizvodstva kartofelya v Penzenskoj oblasti // Niva Povolzh`ya. – 2023. – № 3(67). – DOI: 10.36461/NP.2023.67.3.021.
8. Tishin M.E., Dubinin A.V. Metodologicheskij podxod k prognozirovaniyu urozhajnosti sel`skoxozyajstvenny`x kul`tur na osnove Ridge-regressii (na primere Penzenskoj oblasti) // Surskij vestnik. – 2026. – № 6. – (v pechati).

9. Hersbach H. The ERA5 global reanalysis / H. Hersbach et al. // Quarterly Journal of the Royal Meteorological Society. – 2020. – Vol. 146. – P. 1999–2049. – DOI 10.1002/qj.3803.
10. Lewis C.D. Industrial and business forecasting methods. – London : Butterworth Scientific, 1982. – 143 p.
11. Velde M., Nisini L. Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015 // Agricultural Systems. – 2019. – Vol. 168. – P. 203–212. – DOI: <https://doi.org/10.1016/j.agsy.2018.06.009>.
12. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – 2nd ed. – New York : Springer, 2009. – 745 p. – DOI: <https://doi.org/10.1007/978-0-387-84858-7>.
13. Valenko D. et al. Agroclimatic Forecasting Under Degraded Sensor Data: A Robustness Benchmark of Machine-Learning Models // Applied Sciences. – 2025. – Vol. 16, No. 10. – Art. 5075. – DOI: <https://doi.org/10.3390/app16105075>.
14. Meroni M. et al. Yield forecasting with machine learning and small data: What gains for grains? // Agricultural and Forest Meteorology. – 2021. – Vol. 308–309. – Art. 108555. – P. 1–13. – DOI: <https://doi.org/10.1016/j.agrformet.2021.108555>.
15. Paudel D. et al. Machine learning for large-scale crop yield forecasting // Agricultural Systems. – 2021. – Vol. 187. – Art. 103016. – DOI: [10.1016/j.agsy.2020.103016](https://doi.org/10.1016/j.agsy.2020.103016).

© Дубинин А.В., Тишин М.Е., Федотова М.Ю., 2026. Московский экономический журнал, 2026, № 5.