

КЛАСТЕРНЫЙ АНАЛИЗ ДЛЯ ОЦЕНКИ КЛИМАТИЧЕСКИХ ФАКТОРОВ ПРИ ВОЗДЕЛЫВАНИИ ДВУХ УРОЖАЕВ КАРТОФЕЛЯ РАННИХ СОРТОВ

Я.П. Лобачевский³, ак. РАН, И.Н. Гаспарян¹, д.с.-х.н., А.Г. Левшин²

¹ФГБНУ ВНИИ агрохимии имени Д.Н. Прянишникова

²ФГБОУ ВО Российский государственный аграрный университет — МСХА имени К. А. Тимирязева

³Федеральный научный агроинженерный центр ВИМ

Проанализированы практики применения технологий по адаптации к изменениям климата. Приведен анализ развития методов факторного анализа. Отмечено увеличение в 2 раза за 4 года числа публикаций по новым технологиям интеллектуального анализа больших данных. Разработана и апробирована на базе многолетних экспериментальных данных методика применения технологий Data Mining с использованием аналитической платформы Loginot 7.2. Проведена кластеризация климатических факторов (средняя за декаду температура и средние за декаду осадки) по результатам шестилетних исследований по возделыванию двух урожаев 11 ранних сортов картофеля. Для согласования климатических характеристик с фазами развития картофеля применяли сквозную нумерацию декад вегетационного периода для первой и второй посадок. В качестве интегральной оценки результативности использовали данные по средней урожайности за 6 лет для первой и второй посадок. Приведена программа расчета, включающая анализ и заполнение пропусков, анализ и редактирование выбросов. Приведен подробный анализ результатов кластеризации. Установлены вероятности появления условий близких к трем кластерам. Определены для каждой посадки наиболее благоприятные кластеры, приведен анализ критических сочетаний факторов и предложены возможные агротехнические приемы для компенсации не благоприятных условий. Рекомендовано использовать вероятности появления кластеров для вероятностной оценки природно-климатических факторов при использовании алгоритмов теории игр с природой для обоснования выбора наиболее приспособленных сортов. При получении дополнительных данных методика допускает переобучение и возможность увеличения точности расчетов. Цель исследования – разработка методики кластеризации климатических факторов (средняя температура и средние осадки за декаду) за вегетационный период первой и второй посадок картофеля при возделывании двух урожаев ранних сортов за многолетний период.

Ключевые слова: анализ климатических характеристик, два урожая картофеля раннего, аналитическая платформа Loginot, Data Mining, k-means кластеризация, вероятности кластеров, центры кластеров, функция плотности распределения факторов для кластера.

Для цитирования: Лобачевский Я.П., Гаспарян И.Н., Левшин А.Г. Кластерный анализ для оценки климатических факторов при возделывании двух урожаев картофеля ранних сортов// Плодородие. – 2025. – №5. – С. 63-71. DOI: 10.25680/S19948603.2025.146.12.

Особенностью сельскохозяйственного производства в Российской Федерации является то, что более 70 % территории относится к зоне рискованного земледелия, включающая север европейской части, почти всю территорию Сибири и Дальнего Востока. Это предопределяет риски получения гарантированных урожаев сельскохозяйственных культур из-за климатических факторов и прежде всего температуры и влагообеспеченности. По данным Росгидромета, за последние 50 лет средняя температура на планете увеличилась на 1,2 °С, продолжительность вегетационного периода в целом по земледельческой зоне РФ – в среднем на 3–4 дня за каждые 10 лет. [1]. Повышение температуры создает условия для выращивания двух урожаев картофеля в условиях Московской области, использования сорта зерновых культур с разным сроком созревания. Это снижает напряженность уборочных работ, южные виды возделываемых культур (тыква, соя северного экотипа и др.) адаптируются в Центральной зоне РФ.

Для зоны рискованного земледелия велика опасность холодной весны и осени, недостатка или избытка осадков в важнейшие фазы развития, изменяется динамика внутри почвенных процессов. Увеличиваются риски от природных катаклизмов. К сожалению, из-за сложности

термодинамических процессов на планете, приемлемых методов и моделей прогнозирования климатических рисков на данный момент пока нет. При этом исследователь имеет огромную базу метеорологических наблюдений, но из-за большого объема традиционные методы анализа и сопоставления не позволяют получить достоверные знания из имеющейся информации.

В РГАУ-МСХА и далее во ВНИИ агрохимии проводили экспериментальные исследования по возделыванию двух урожаев картофеля в условиях Московского региона в течение 2017–2023 г., подтверждена эффективность рекомендуемой технологии. В качестве объектов исследования были выбраны 11 ранних сортов картофеля: Метеор, Жуковский ранний, Удача, Снегирь, Ред Скарлетт, Ривьера, Импала, Леди Клэр, Голубизна и Брянский ранний [1]. Осуществляли две посадки: в 3-й декаде апреля и 1-й декаде июля. По результатам исследования отмечена большая вариация урожайности по сортам и годам возделывания. В этих условиях важно обосновать выбор сортов для первой и второй посадок. С целью выбора сортов, наиболее адаптированных к вариациям климатических факторов, применили методы теории вероятностных игр с природой [2].

Отмечено, что получению наибольшей урожайности противостоит случайный характер появления различных почвенных, метеорологических и ландшафтных условий. При соблюдении одной технологии для одной и той же делянки разница в урожайности будет наблюдаться из-за различий климатических характеристик. Отмеченные особенности акцентируют внимание на наличии противоречий в стратегиях выполнения работ (выбор сорта для 1-й или 2-й посадки) и неопределенности внешних климатических условий.

Однако для этого необходимо иметь методику вероятностной оценки климатических характеристик. Поставленная задача решалась на основе статистических характеристик. Для приведенных опытных данных по урожайности 10 сортов за 6 лет разделение на качественные группы климатических условий на предварительном этапе для 1- и 2-й посадок рассматривали суммарную среднюю массу урожая за каждый год (табл. 1).

1. Суммарная средняя масса урожая за годы исследований (кг)

Урожай	Годы						Размах	d, Ширина класса
	2017	2018	2019	2020	2021	2022		
1-й	0,532	0,483	0,488	0,555	0,503	0,579	0,096	0,032
2-й	0,463	0,47	0,446	0,498	0,48	0,484	0,052	0,0173

Принимая гипотезу о распределении урожайности, подчиненной закону нормального распределения, определяли размах (разница между максимальным и минимальным значениями) для 1-й и 2-й посадки. Задаваясь тремя уровнями групп климатических условий (благоприятные, неблагоприятные и одна промежуточная группа) определяли ширину интервала для группы (табл. 2)

$$d = \frac{(m_{\text{максимум}} - m_{\text{минимум}})}{3}$$

2. Границы интервалов климатических групп (1-й урожай)

№ группы	Граница интервалов групп		Годы	Класс года	Вероятность
	левая	правая			
1-я	0,483	0,515	2018, 2019, 2021	Неблагоприятный	0,500
2-я	0,515	0,547	2017	Средний 1	0,167
3-я	0,547	0,579	2020, 2022	Благоприятный	0,333

Подсчитывали сколько лет попадали в каждый интервал и находили вероятности появления указанных групп погодных условий.

Вероятностная оценка по приведенной методике рассматривалась как предварительная. Полученные вероятности не были связаны с климатическими характеристиками.

При факторном анализе климатических характеристик в полевых опытах будем использовать средние за декаду данные температуры и осадков. Полевые опыты с 10 сортами картофеля раннего проводились в течение 6 лет – с 2018 по 2023 г. Косвенным показателем, оценивающим календарные сроки, можно использовать порядковый номер декады месяцев вегетационного периода (табл. 3).

В качестве оценки эффективности производственного процесса рассматривается средняя урожайность по испытываемым сортам для каждой посадки. Таким образом, в качестве исходной выступает многомерная матрица 6 x 9 для 1-й посадки и 6 x 8 для 2-й посадки. По

сути, получается кросс-таблица, отображающая многомерные данные на двухмерной плоскости [3].

Для группировки имеющихся данных с большой достоверностью могут быть использованы алгоритмы кластеризации динамических объектов наблюдения [4]. Процедура кластеризации должна удовлетворять следующим требованиям: отсутствие управляющих параметров; инвариантность относительно возможных преобразований координат; стабильность выявленных кластеров в течении экспертно-задаваемого времени.

3. Даты учитываемых данных для первой и второй посадок

1-я посадка	Апрель		Май			Июнь			Июль	
	Декада	3	1	2	3	1	2	3	1	2
Порядковый номер декады	1	2	3	4	5	6	7	8	9	
2-я посадка	Июль		Август			Сентябрь			-	
	Декада	2	3	1	2	3	1	2	3	-
Порядковый номер декады	1	2	3	4	5	6	7	8	-	-

Эффективность применения метода кластерного анализа ранее была определена при анализе миграции сельского населения в Республике Тыва [5]. Группировку проводили по общности 21 характеристики для 17 районов Республики по алгоритму k-means и нейронной сети Кохонена. В результате были обоснованы модельные районы, для которых в дальнейшем оптимизировались транспортные подсистемы для социальных служб.

Применение метода интеллектуального анализа данных позволило повысить эффективность работы по обнаружению нестандартных действий (в т.ч. мошеннических) [6].

В связи с этим, исследования по применению методов кластерного анализа для группировки климатических данных, основанного на технологиях искусственного интеллекта является своевременными и актуальными.

Методика. Системное описание теоретических и методологических основ кластерной концепции представлено в монографии [7]. Основным посылом книги, привлекающим за собой ряд методических особенностей предлагаемого подхода, является переосмысление феномена кластеров как самоорганизующихся систем в социально-экономическом пространстве.

В сфере интеллектуального анализа данных ведутся активные научные исследования. Анализ публикаций, индексированных в реферативной базе данных SCOPUS, показывает устойчивый рост количества научных работ по методам и технологиям факторного анализа (табл. 4) [8].

4. Количество публикаций в базе SCOPUS, тыс. [7]

Ключевые слова	Период				
	1991-1995	1996-2000	2001-2005	2006-2010	2011-2015
Data Analysis	221,1	333,9	536,7	906,2	1008,7
Machine Learning	1,9	2,96	9,3	26,4	43,5
Expert System	13,1	14,4	19,9	34,5	41,3
Data Mining	1,4	4,7	18,2	42,3	59,6
Big Data	1,3	2,6	5,1	10,5	28,2
Deep Learning	0,263	0,596	1,22	3,38	7,41

Приведенные данные показывают, что рост числа публикаций по новым технологиям анализа больших данных (Data Mining «добыча данных, интеллектуальный анализ данных, глубинный анализ данных, обнаружение знаний в данных», Big Data «большие данные», Deep Learning «глубокое обучение») практически удваивается за каждые 4 года. Для целенаправленного анализа большой информационной базы наиболее приемлемы методы Data Mining, реализованные компанией Loginom Company (РФ) в аналитической платформе Loginom [9, 10].

Loginom 7.2 — аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования. Loginom внесен в реестр российских программ.

Одним из методов представления данных и проведения оперативного анализа является технология OLAP (Online Analytical Processing) — оперативная аналитическая обработка данных. Такой подход позволяет в реальном времени генерировать описательные и сравнительные данные. OLAP-кубы представляют собой проекцию исходного куба данных на куб данных меньшей

размерности. При этом значения ячеек агрегируются, т.е. объединяются с применением функции агрегации — сумма, среднее количество, минимум, максимум. Такие проекции или срезы исходного куба демонстрируются на экране в виде кросс-таблицы [3].

Извлечение знаний из баз данных (Knowledge Discovery in Databases- KDD) представляет собой процесс поиска полезных знаний в «сырых данных». Метод включает вопросы подготовки данных, выбора информативных признаков, очистки данных, применение методов Data Mining (DM), постобработки данных и интерпретации полученных результатов. При использовании этих алгоритмов априори предполагают подчиненность анализируемых данных закону нормального распределения. В массиве данных не допускаются пропуски значений [10].

Массив данных импортируется из таблицы в формате Excel. Допустимо использовать базы данных, файлы Deductor Warehouse (прежняя версия программы), 1С запрос и др. В нашем примере использовали данные в формате Excel, которые загружали в первый узел 1-й программы (рис. 1).



Рис. 1. Программа для кластерного анализа

После того, как указаны входные и выходные поля, задаются методом нормализации данных в обучающей выборке. Цель нормализации значений полей — преобразование данных к безразмерному виду в диапазоне [0...1].

Введенные данные подвергаются процедуре предобработки: узел 2 — «заполнение пропусков» и узел 3 — редактирование выбросов. При заполнении пропусков использовали методы одномерного анализа — «Заменять случайными значениями». Редактирование выбросов позволяет выделить грубые ошибки и заменить их на смоделированное случайное значение.

В основе кластеризации алгоритмом k-means лежит принцип оптимальной в определенном смысле группировки множества данных на k кластеры. Алгоритм группирует данные в кластеры таким образом, чтобы целевая функция алгоритма разбиения достигала экстремума.

Алгоритм k-means состоит из четырех шагов [10]:

1. Первоначальное распределение объектов по кластерам. Задается число k, на первом шаге эти точки считаются центрами кластеров. Каждому кластеру соответствует один центр. Выбор начальных центров осуществляется случайным образом;

2. Случайным образом выбирается k записей, которые будут служить начальными центрами кластеров;

3. Для каждой записи исходной выборки определяется ближайший к ней центр кластера;

4. Производится вычисление центроида — центра тяжести кластеров. Это делается путем определения среднего для значений каждого признака всех записей в кластере.

Затем старый центр кластера смещается в его центроид. Таким образом, центроиды становятся новыми центрами кластеров для следующей итерации алгоритма.

Шаги 3 и 4 повторяют до тех пор, пока выполнение алгоритма не будет прервано, либо пока не будет выполнено условие в соответствии с некоторым критерием сходимости.

В качестве критерия сходимости, E используют сумму квадратов расстояний между центроидами кластера и всеми вошедшими в него точками (p):

$$E = \sum_{i=1}^k \sum_{p \in C_i} (p - m_i)^2,$$

где $p \in C_i$ — произвольные точки данных, принадлежащие кластеру C_i ; m_i — центроид данного кластера, k — число кластеров.

Выбор числа k может базироваться на теоретических соображениях или имитационных расчетах. Центр

кластера – средние значения переменных объектов, входящих в кластер [10].

Перед расчетом по приведенной программе (рис. 3) в узле 3 задается и реализуется алгоритм настройки «Переобучить узел» после чего дается команда на выполнение расчета.

В программном комплексе предусмотрены процедуры трансформации данных (группировка, разгруппировка, дополнение и объединение данных, свертка столбцов и другие сервисы) и исследования данных (автокорреляция, дубликаты и противоречия, корреляционный и факторный анализы). В отдельный модуль Data Mining включены алгоритмы: кластеризация,

кластеризация транзакций, линейная и логистическая регрессия, нейросеть (регрессия), самоорганизующаяся сеть и EM кластеризация.

В нашем исследовании использовали алгоритм кластеризации k-means (узел 3, рис.1). Предварительно каждый узел настраивается для согласования данных. Для визуализации результатов разбиения на кластеры, оценки центров и профиля кластеров использовали опции «диаграмма», «таблица» и «статистика» (рис.2). В качестве статистических характеристик использовали: минимальное и максимальное значения, среднее, среднее квадратическое отклонение и размах. Для температуры и осадков определяли сумму значений.

Разбиение на кластеры

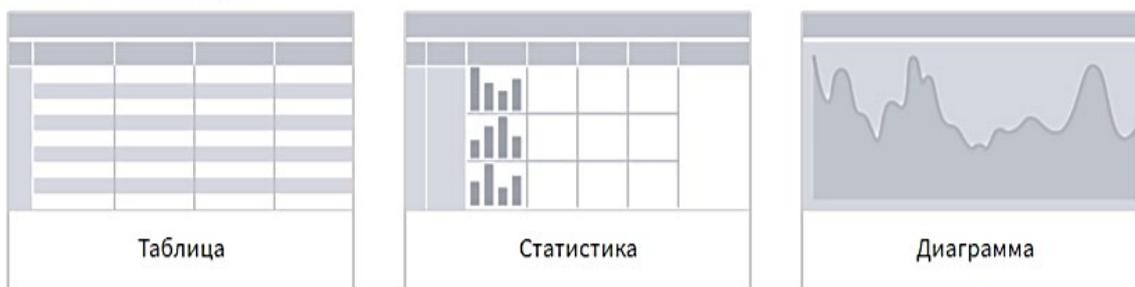


Рис. 2. Процедуры визуализации результатов

В качестве практического пособия можно воспользоваться учебными пособиями [8-10] и справочным сервисом программного комплекса.

Результаты и их обсуждение. Для кластеризации по алгоритму k-means важным моментом является задание числа кластеров. Это не формализованная процедура, поэтому были проведены расчеты для 4, 3 и 2 кластеров. В результате получили для двух кластеров вероятности 1 кластера 68%, а для 0 кластера – 32%. Достаточно большая неравномерность распределения данных. Для 4 кластеров получили соответственно: 3–24%, 1–22%, 0–13% и для 2–6,3%. Получается 1 лишний кластер, поэтому в дальнейших расчетах задавали 3 кластера.

При первой посадке в кластер 0 попали 13 точек, или 24%, кластер 1-21 точка, 39%, кластер 2-20 точек, 37% (табл. 5).

5. Распределение точек по кластерам для первой и второй посадок

Кластер	1-я посадка		2-я посадка	
	Число точек	Доля, %	Число точек	Доля, %
0	13	24	15	31
1	21	39	12	25
2	20	37	21	44
Всего	54	100	48	100

Для второй посадки в кластер 0 попали 15 точек, или 31%, в кластер 1 попало 12 точек (25%). Наибольшее количество точек 21 попало в кластер 2–44%. Точки распределились относительно равномерно, что подтверждает правильный выбор числа кластеров.

Расчеты по кластеризации (узел 3, рис. 1) проводили для соответствующих данных средних за декаду температур и осадков для календарных сроков, приведенных в таблице 3. Для декад установлены порядковые номера для соотношения данных с календарным сроком. Статистические характеристики анализируемых факторов для центров кластеров 1 и 2 посадки приведены в таблице 6.

6. Статистические характеристики для центров кластеров

Посадка	Метка	Минимум	Максимум	Среднее	Среднее квадратическое отклонение
1	Расстояние до центра кластера*	0,53	3,12	1,36	0,45
	Температура, °C	7,6	24,8	17,2	3,8
	Осадки, мм	0,1	122,3	31,0	27,3
	Урожайность, г/куст	452	599	510	52,5
2	Расстояние до центра кластера*	0,34	1,93	1,38	0,40
	Температура, °C	6,9	24,8	16,8	4,3
	Осадки, мм	0	80,6	21,4	21,2
	Урожайность, г/куст	441,9	498,1	468,0	21,1

* В нормированном виде.

Для анализа использовали средние значения урожайности по 10 сортам за 6 лет исследований. Средняя температура для 1- и 2-й посадок, но вариация для 2-й посадки больше на 0,5 °C. Распределение температуры подчиняется закону нормального распределения с эксцессом в сторону больших значений (рис 3). Характерной особенностью вариации температуры и осадков является то, что динамика изменения осадков существенно больше вариации температуры (рис. 4). Эта особенность относится и ко второй посадке.

Для 1-й посадки среднее количество осадков на 10 мм больше, а размах составил 122 мм, а для 2-й посадки размах на 42 мм меньше – 80,6 мм. Характер гистограммы для осадков и примерное равенство значений для среднего 31,0 мм и среднего квадратического отклонения 27,3 мм свидетельствует о подчиненности данных экспоненциальному закону распределения (рис. 5, 6).

Для 1-й посадки 47% попадают в интервал 0-13,4 мм и 20% в интервал 13,4-26,9 мм. Для 2-й посадки 41% значений попали в интервал 0,1-17,6 мм и 20% в интервал 17,6-35,0 мм.

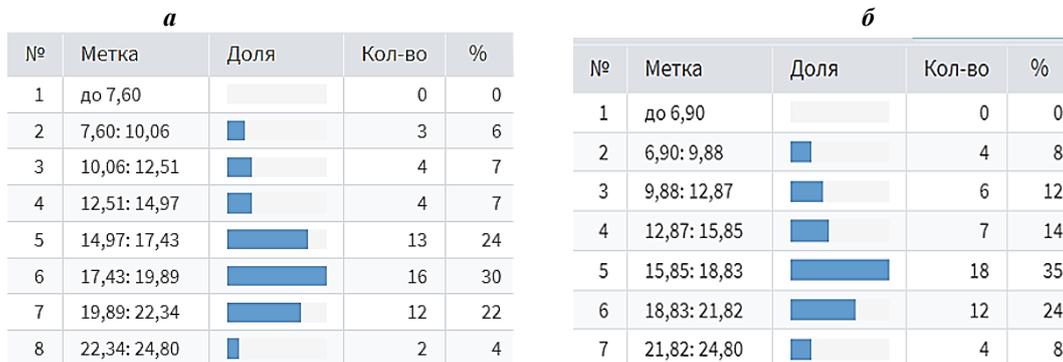


Рис. 3. Гистограммы для температуры при 1-й (а) и 2-й (б) посадках

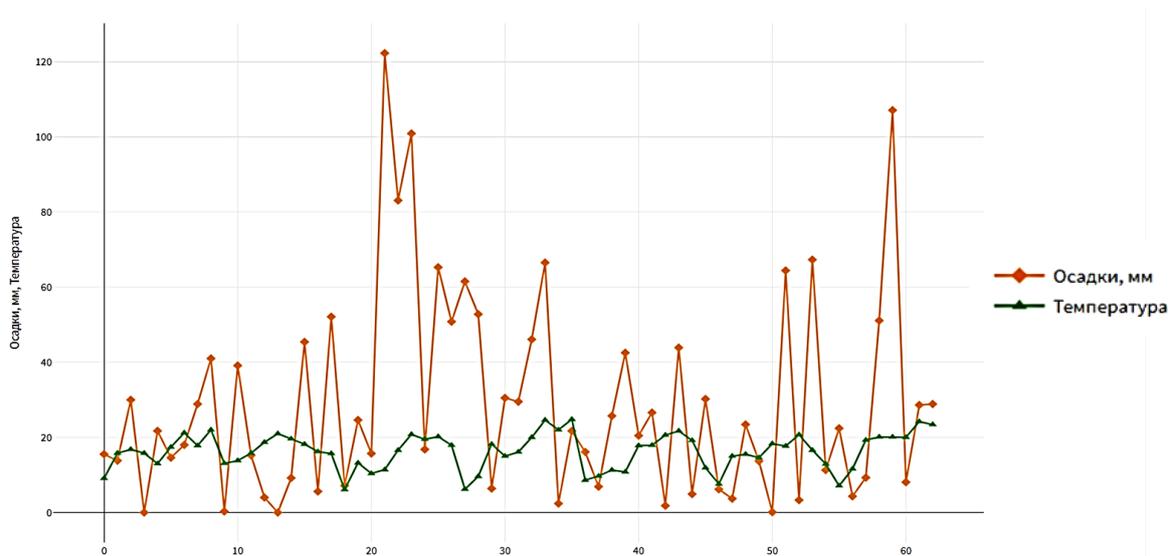


Рис. 4. Динамика температуры и осадков для 1-й посадки.

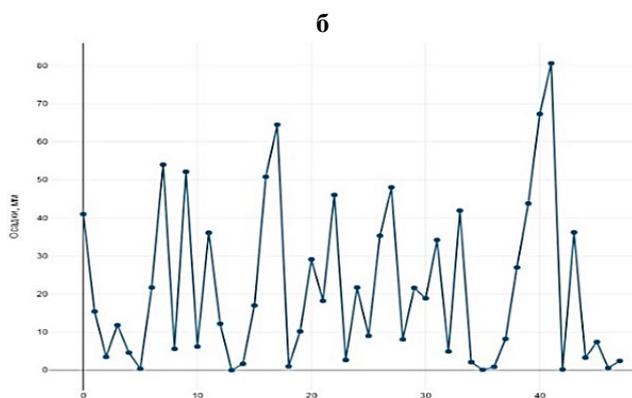


Рис. 5. Гистограмма (а) и диаграмма (б) осадков для 1-й посадки

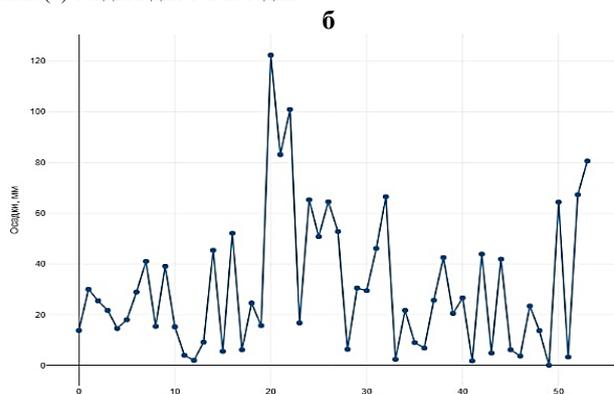


Рис. 6. Гистограмма (а) и диаграмма осадков (б) для 2-й посадки

Особо следует отметить, что урожайность при первой посадке была выше. Средняя урожайность для 1-й посадки – 510 г/куст, а для 2-й посадки – 468 г/куст. Для 1-й посадки существенно больше средние квадратические отклонения – 52,5 г/куст, а для 2-й посадки 21,1 г/куст.

Средние расстояния до центра кластера для 1-й и 2-й посадок практически одинаковые – 1,36 и 1,38 (в нормированном виде). Однако размах для 1-й посадки 1,59 существенно больше, чем для 2-й посадки – 2,59. Существенно различия и распределение значений на гистограмме. Для 1-й посадки значения смещены к верхней границе, а для 2-й посадки к нижней границе (рис. 7).

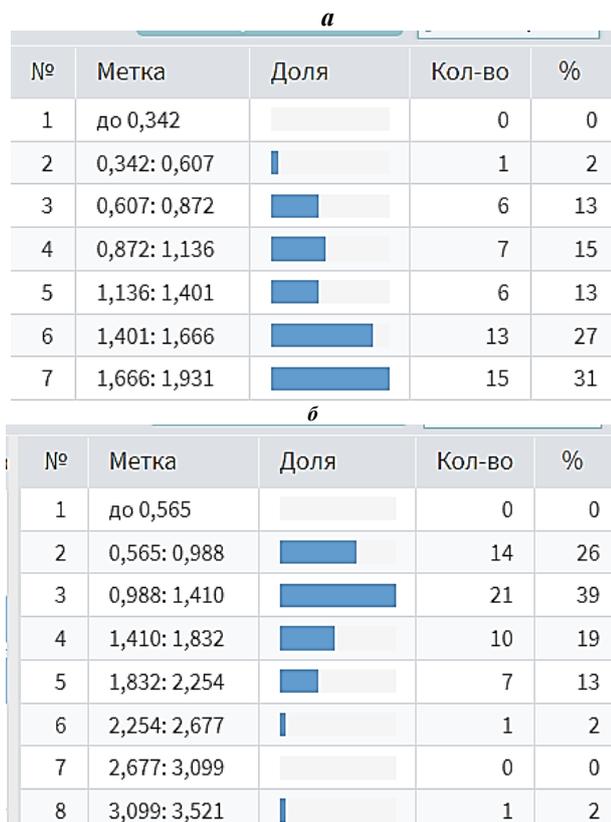


Рис. 7. Гистограмма распределения расстояний до центра кластеров для 1-й (а) и для 2-й (б) посадки

Центры кластеров для первой и второй посадок приведены в таблице 7, статистические характеристики для анализируемых факторов – в таблицах 8 (1-я посадка) и 9 (2-я посадка). Округленные значения декады приведены в таблице 7, а так как расчеты проводятся в нормированной форме, (0-1) то для целой величины (табл. 8 и 9) получены дробные значения декад для кластеров.

Кластер 0 для 1-й посадки средняя декада равна 6 (2-я декада июня). В это время происходят цветение и завязывание клубней. Этот кластер характеризуется средним уровнем температуры – 18,6 °С, сумма температур для кластера 242,2 °С, наибольшее количество осадков – 49,5 мм (стандартное отклонение – 37 мм, сумма осадков – 643,0 мм). Для этого кластера получена максимальная урожайность – 578,5 г/куст, с учетом этого его можно оценить как **благоприятный**.

Кластер 1 первой посадки соответствует фазе цветения (3-я декада июня), происходят формирование и рост клубней. Характеризуется повышенной температурой 19,6 °С (стандартное отклонение 2,5 °С, сумма температур 411,1 °С), средним количеством и неравномерным распределении осадков – 28,6 мм (стандартное

отклонение – 25,9 мм, сумма осадков 599,8 мм), получена наименьшая средняя урожайность – 475,0 г/куст. По результатам анализа 1-й кластер является **неблагоприятным**. Недостаток влаги можно компенсировать поливом (управляемый фактор) [11-13].

7. Средние значения факторов для центров кластеров при первой и второй посадках

Посадка	Кластер	Декада	Температура, °С	Осадки, мм	Урожайность, г/куст
1	0	6	18,6	49,5	578,5
	1	7	19,6	28,6	475,0
	2	2	13,8	21,5	502,3
2	0	7	11,8	20,5	468
	1	2	19,7	48,0	464
	2	4	18,8	7,0	471

8. Статистические характеристики профиля кластеров (1-я посадка)

Показатель	Кластер			Итого	Фактор
	0	1	2		
Среднее значение	6,23	6,86	2,25	5,0	Декада
Стандартное отклонение	1,96	1,56	1,07	2,61	
Среднее значение	18,6	19,6	13,8	17,2	Температура, °С
Стандартное отклонение	2,7	2,5	3,0	3,8	
Сумма	242,2	411,1	276,7	930,0	Осадки, мм
Среднее значение	49,0	28,6	21,5	31,0	
Стандартное отклонение	37	25,9	13,5	27,0	Урожайность, г/куст
Сумма	643,0	599,8	430,9	1674	
Среднее значение	578,5	475	502	510	Урожайность, г/куст
Стандартное отклонение	20	14	51	53	

9. Статистические характеристики профиля кластеров (2-я посадка)

Показатель	Кластер			Итого	Фактор
	0	1	2		
Среднее значение	7,2	2,25	3,86	4,5	Декада
Стандартное отклонение	0,77	1,22	1,56	2,32	
Среднее значение	11,8	19,7	18,8	16,8	Температура, °С
Стандартное отклонение	2,9	2,4	2,5	4,3	
Сумма	176,7	235,9	394,0	806,6	Осадки, мм
Среднее значение	20,5	48,0	7,0	21,4	
Стандартное отклонение	17,5	16,4	7,3	21,2	Урожайность, г/куст
Сумма	307,2	575,5	146,8	1029	
Среднее значение	468	464	471	468	Урожайность, г/куст
Стандартное отклонение	21	21	22	21	

Кластер 2 характеризуется низкой температурой 13,8 °С (стандартное отклонение 3,0 °С, сумма температур 276,7 °С). Растениям недостаточно тепла. Декада 2 (1-я декада мая) – это период прорастания и появления всходов, но из-за низкой температуры эта фаза развития затягивается. Возможная компенсация – пленочное укрытие [12]. Для этой фазы достаточное количество осадков – 21,5 мм (стандартное отклонение 13,5 мм, сумма – 430,9 мм). Кластер 2 является **средним** по продуктивности – 502,3 г/куст.

Для второй посадки кластер 0 – средняя декада – 7 (2-я декада сентября) соответствует фазе накопления массы, но из-за низкой температуры 11,8 °С потенциал растения не реализуется в полной степени при среднем

уровне осадков – 20,5 мм и как следствие невысокая урожайность – 468 г/куст. Вероятность этого кластера – 29%.

Кластер 1 декада 2 (3-я декада июля) характеризуется высокой температурой 19,7 °С и большим количеством осадков 48,0 мм (стандартное отклонение – 16,4 мм, сумма – 575,5 мм) наблюдается излишняя влага. В этот период происходит развитие всходов и закладка столонов. В итоге при благоприятных условиях закладывается излишнее количество столонов.

Для кластера 2 средняя 4-я декада – фаза цветения и накопления массы клубней (2-я декада августа), которая характеризуется небольшим количеством осадков – 7,0

мм (стандартное отклонение – 7,3 мм, сумма – 146,8 мм), поэтому при достаточно высокой температуре 18,8 °С существенно замедляется развитие растений. В итоге невысокий уровень урожайности – 471 г/куст. Кластер 2 наиболее вероятен и составляет 44%. При достаточной температуре недостаток влаги возможно компенсировать поливом. В итоге к **благоприятным** можно отнести **кластер 2, кластер 0 – неблагоприятный** и **кластер 1** (наименьшая вероятность – 24%) отнесем к среднему классу.

Функции плотности распределения факторов для каждого кластера (посадки 1 и 2) представлены на рисунке 7.

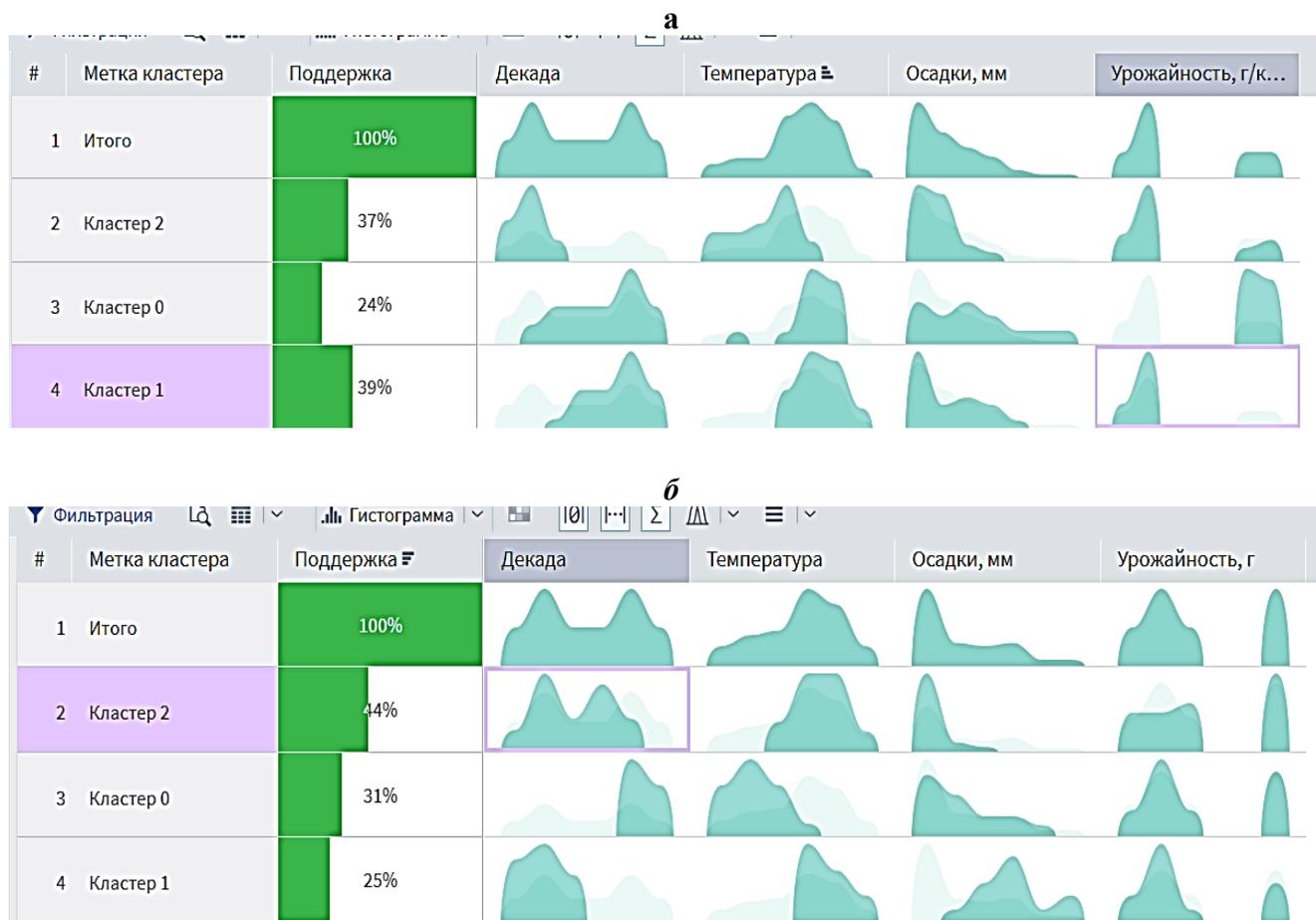


Рис. 7. Плотность функции распределения факторов для кластеров при посадках 1 (а) и 2 (б)

Для первой посадки функции плотности распределения существенно различаются по кластерам. Для кластера 2 декады смещены к левой границе диапазона (зоны посадки и начала развития), для кластера 0 – размещается в центре со смещением к правой границе (охватывает диапазон появления всходов и развитие растений) и для кластера 1 (вероятность 39%) функция плотности смещается к правой границе (фазы цветения и накопления биомассы).

Функция плотности для температуры в кластере 2 смещена к левой границе, в кластерах 0 и 1 – к правой границе. Функция плотности распределения осадков для всех кластеров смещена к левой границе диапазона. Для урожайности функция плотности четко выражена для кластера 0 (смещена вправо) и 1 (смещена влево), для кластера 2 – прерывистая (по виду отражает итоговое распределение).

Для 2-й посадки функция плотности распределения декады для кластера 2 – двухмодальная, для кластеров 0 и 1 смещение более четко определено: кластер 0 – смещение к правой границе; кластер 1 – к левой границе. Для температуры функция плотности также четко сориентирована: кластер 2 и 1 смещение к правой границе, а для кластера 0 к левой границе. Осадки для кластеров 2 и 0 смещены к левой границе, для кластера 1 – к правой. Функция плотности для урожайности прерывистая, масштабируемая с итоговым распределением.

Значимость фактора (в %) оценивается как разница – 1 минус вероятность нулевой гипотезы. Для непрерывных полей используется t-критерий Стьюдента, а для дискретных полей – критерий хи-квадрат. Общая значимость поля определяется по F-критерию Фишера [3]. Значимость для изучаемых факторов при 1 и 2 посадках показана на рисунке 8.

а

#	Метка кластера	Поддержка	Декада	Температура	Осадки, мм	Урожайность, г/к...
1	Итого	100%	80	84	37	69
2	Кластер 2	37%	65	62	28	47
3	Кластер 0	24%	71	75	40	81
4	Кластер 1	39%	64	62	26	46

б

#	Метка кластера	Поддержка	Декада	Температура	Осадки, мм	Урожайность, г
1	Итого	100%	79	90,4	59	66
2	Кластер 2	47%	49	59	40	41
3	Кластер 0	29%	77	73	45	51
4	Кластер 1	24%	60	82	72	54

Рис. 8. Значимость факторов для кластеров при 1 (а) и 2 (б) посадках

Значимость интегральных значений факторов имеет достаточно высокий уровень. Для сроков (декады) значимость равна 80 % и вариация для кластеров 64-71 % при 1-й посадке и 79 %, вариация 49-77 % при 2-й посадке. Значимость температуры: 1-я посадка интегральное значение 84 %, вариация по кластерам 62-74 %, 2-я посадка интегральное значение значимости 90,4 %, вариация 59-82 %.

Итоговые значения фактора урожайности 69–66 %, при этом вариация для кластеров 46-81 % 1 посадка и 41-54 % – 2-я посадка. Из-за большой вариации осадков (см. рис. 4) по сравнению с температурой итоговые значения важности фактора 37-59 %, для кластеров 26-40 % (1-я посадка) и 40-72 (2-я посадка).

Заключение. 1. При кластеризации климатических характеристик достаточно задавать количество кластеров, равное 3. Для анализируемых данных по средним за декаду температуре и осадкам на 1-й посадке получены вероятности появления кластеров: кластер 0 – 24 %, кластер 1 – 39, кластер 2 – 37 %; для 2-й посадки, соответственно, 0 – 31 %, 1 – 25 и 2 – 44%.

2. Кластерный анализ с использованием аналитической платформы Loginom 7.2.2 позволяет сгруппировать метеорологические данные в кластеры, оценить вероятность кластеров, статистические характеристики кластеров и сопоставить возможную урожайность, что позволяет дифференцировать условия на благоприятные, не благоприятные и средние.

3. Статистические характеристики кластеров позволяют сопоставить с фазами развития и оценить достаточность природно-климатических характеристик и при необходимости прогнозировать дополнительные агротехнические приемы для компенсации неблагоприятных отклонений: при недостатке влаги необходимость полива в определенные декады и применения пленочных покрытий при опасности низких температур.

4. Вероятности появления кластеров возможно использовать в качестве характеристики стратегий природы при обосновании сортов с использованием теории игр с природой.

Литература

1. *Лучшие* практики применения технологий по адаптации отрасли растениеводства к изменениям климата : монография / И.Н. Гаспарян и др.- М.: МЭСХ, 2024.- 196 с.
2. *Features of the choice of potato (Solanum tuberosum) varieties when growing two harvests in the temperate zone of Russia* / I. N. Gasparyana *, A. G. Levshina, b , O. N. Ivashovab , N. F. Deniskinab and Sh. V. Gasparyanb.// *Brazilian Journal of Biology*, 2024, vol. 84, e27827
3. *Deductor* Руководство аналитика. Версия 5.3 : Рязань, 1995-2013 Компания BaseGroup™ Labs <http://www.basegroup.ru/> (свободный доступ).
4. Кластеризация динамических объектов наблюдения/ Бобров К.А., Чекал Е.Г., Чичев А.А.1: // *Ученые записки УлГУ. Сер. Математика и информационные технологии. УлГУ.* // Электрон. журн. – 2019. – № 1. – С. 7-19.
5. *Ондар А.М.* Кластеризация районов Республики Тыва по уровню миграции сельского населения// *Доклады ТСХА: Сборник статей. Выпуск 291. Часть II.* - М.: Издательство РГАУ-МСХА, 2019. С. 170-174.
6. *Кластерный анализ: формирование индикатора риска для больших совокупностей учетной информации/* Е. Зверев, А. Никифоров // *Внутренний контроль в кредитной организации.* – 2018. - №3 (39). – С. 24-35.
7. *Марков Л.С.* Теоретико-методологические основы кластерного подхода. – Новосибирск: ИЭОПП СО РАН, 2015. – 300 с.
8. *Поручиков, М. А.* Анализ данных: учеб. пособие / М.А. Поручиков. – Самара: Изд-во Самарского университета, 2016. – 88 с.
9. *Прокопенко Н. Ю.* Аналитические информационные системы поддержки принятия решений [Текст]: учеб. пособие / Н.Ю. Прокопенко; Нижегород. гос. архитектур. – строит. ун-т – Н. – Новгород: ННГАСУ, 2020. – 142 с. ISBN 978-5-528-00395-5.
10. *Поллак, Г.А.* Аналитические информационные системы поддержки принятия решений на платформе Loginom Community: учебное пособие / Г.А. Поллак, И.А. Прохорова. – Челябинск: Издательский центр ЮУрГУ, 2022. – 229 с.
11. *Clustering for High Dimensional Data: Density based Subspace Clustering Algorithms/* Sunita Jahirabadkar, Parag Kulkarni// *International Journal of Computer Applications (0975 – 8887) Volume 63– No.20, February 2013.*
12. *Bairambekov Sh. B., Gar'yanova E. D., Gulyaeva G. V.* Use of temporary shelters when growing early potatoes in the Astrakhan region // *Potato growing.* – 2017. – P. 175-177.
13. *Evers, D., Boncheret, S., Hoffmann, L. and Hausman, J. F.* (2007). Physiological aspects of potato response to abiotic stress. *Belgian Journal of Botany*, 141-150.
14. *Butuzov A. E., Gasparyan I. N., Dyikanova M. E., Ivashova, O. N.* The influence of covering on the efficiency of early potato production in the Moscow region // *International technical and economic journal.* – 2018. – No. 3. – P. 15-20. – EDN XVLVCH

Ya.P. Lobachevsky³, Academician of the Russian Academy of Sciences, I.N. Gasparyan¹, A.G. Levshin²

¹*D.N. Pryanishnikov All-Russian Scientific Research Institute of Agrochemistry;
31a Pryanishnikova str., Moscow, 127434, Russia*

²*Russian State Agrarian University – Moscow Timiryazev Agricultural Academy,
127434, Moscow, 49 Timiryazevskaya str., Russia*

³*Federal Scientific Agroengineering Center VIM*

The article analyzes the practices of applying technologies for adaptation to climate change. An analysis of the development of factor analysis methods is presented. A twofold increase in the number of publications on new technologies for intellectual analysis of big data is noted. A methodology for applying Data Mining technologies using the Loginom 7.2 analytical platform was developed and tested on the basis of long-term experimental data. Clustering of climatic factors (average temperature per ten-day period and average precipitation per ten-day period) was carried out based on the results of six summer studies on the cultivation of two crops of 11 early potato varieties. To align climatic characteristics with potato development phases, a continuous numbering of ten-day periods of the growing season for the first and second plantings was used. Data on the average yield over 6 years for the first and second plantings were used as an integral assessment of performance. A calculation program is provided, including the analysis and filling of gaps, analysis and editing of outliers. A detailed analysis of the clustering results is provided. The probabilities of conditions similar to three clusters were determined. The most favorable clusters were identified for each planting, critical factor combinations were analyzed, and possible agronomic practices for compensating for unfavorable conditions were proposed. It is recommended to use the cluster occurrence probabilities for a probabilistic assessment of natural and climatic factors when using game theory algorithms with nature to justify the selection of the most adapted varieties. With additional data, the method allows for retraining and the possibility of increasing the accuracy of calculations. The objective of the study was to develop a method for clustering climatic factors (average temperature and average precipitation per ten-day period) for the growing season of the first and second potato plantings when cultivating two crops of early varieties over a multi-year period.

Keywords: analysis of climatic characteristics, two early potato crops, Loginom analytical platform, Data Mining, k-means clustering, cluster probabilities, cluster centers, factor distribution density function for a cluster.