

УДК 631.8:519.237.5

ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ РЕГРЕССИОННОГО АНАЛИЗА В АГРОХИМИЧЕСКИХ ИССЛЕДОВАНИЯХ

© 2022 г. А. В. Пасынков¹, Е. Н. Пасынкова^{1,*}

¹Ленинградский научно-исследовательский институт сельского хозяйства “Белогорка” – филиал Федерального исследовательского центра картофеля им. А.Г. Лорха 188338 п. Белогорка, Гатчинский р-н, Ленинградская обл., Россия

*E-mail: pasynkova.elena@gmail.com

Поступила в редакцию 11.04.2022 г.

После доработки 02.06.2022 г.

Принята к публикации 12.07.2022 г.

Обсуждаются особенности и практические аспекты использования регрессионного анализа в агрохимических исследованиях. Показано, что линейные уравнения регрессии не могут являться корректным методом статистической обработки экспериментальных данных, т.к. в зависимости от того, в какой части биологического интервала независимых переменных (в начале, в середине или в конце) получены данные, уравнения могут показывать противоречивые результаты. В большинстве случаев зависимости между различными показателями в агрохимических исследованиях максимально точно отражают нелинейные уравнения: логарифмические, половинной степени, 2-го, 3-го порядка и т.п. Множественный регрессионный анализ с учетом ограничений (распределение выборок, мультиколлинеарность и т.д.) может быть использован при исследовании зависимостей урожайности и показателей качества от доз и соотношений удобрений, норм биологических и химических средств защиты растений, гидротермических условий, складывающихся в период вегетации, а также в изучении взаимосвязей биохимического состава и технологических качеств сельскохозяйственных культур.

Ключевые слова: регрессионный анализ, нелинейные уравнения, урожайность, удобрения, средства защиты растений, гидротермические условия, технологические качества.

DOI: 10.31857/S0002188122100088

ВВЕДЕНИЕ

Общеизвестна важная роль методов математической статистики при анализе данных, полученных при проведении полевых экспериментов и сопутствующих им широкого спектра наблюдений и аналитических работ. При этом важность статистической обработки экспериментальных данных не подвергалась сомнению в предшествующий период развития агрохимии, а в настоящее время роль статистической обработки данных значительно возросла. Это связано с накоплением и использованием сравнительно больших массивов экспериментальных данных, полученных в различных регионах России как при проведении серийных краткосрочных и особенно длительных стационарных опытов, проводимых в рамках Географической сети опытов с удобрениями под методическим руководством ВНИИ агрохимии им. Д.Н. Прянишникова.

Еще одной из причин потребности в статистической обработке является обеспокоенность сло-

жившимся положением редакторов научных журналов биологического и сельскохозяйственного профиля. Даже в комментариях к “Правилам для авторов” одного из ведущих сельскохозяйственных журналов нашей страны прямо указано, что “... большинство статей отклоняется по причине отсутствия статистической обработки” [1]. Трудно представить такие “откровения” у редакторов физико-математических журналов.

Известно несколько методов статистической обработки экспериментальных данных, используемых в агрохимических исследованиях. Анализ частоты использования методов статистической обработки данных, представленных в публикациях в последние годы, показал, что наряду с дисперсионным, регрессионный анализ является одним из основных ее методов. Поэтому в данной публикации частично обсуждаются теоретические, но в основном практические аспекты использования именно регрессионного анализа.

Регрессионный анализ, как метод статистической обработки экспериментальных данных основан на методе наименьших квадратов [2]. Более детально с сущностью метода наименьших квадратов, а также развитием и особенностями применения регрессионного анализа можно ознакомиться в капитальном труде, переведенном на русский язык и выдержавшем в СССР и современной России 3 издания [3]. В этой книге описываются методы подбора и исследования линейных и нелинейных регрессионных моделей различной степени сложности, а также, что очень важно в настоящее время, рассматриваются и обсуждаются практические аспекты их применения. Важность практического использования регрессионного анализа при статистической обработке экспериментальных данных в агрохимических исследованиях отмечали с середины XX в. и в нашей стране [4–6]. Цель работы – исследование особенностей использования регрессионного анализа в агрохимических исследованиях.

МЕТОДИКА ИССЛЕДОВАНИЯ

В представленной работе обсуждаются особенности и результаты использования регрессионного анализа экспериментальных данных, полученных авторами статьи, а также полученных при статистической обработке опубликованных данных других авторов.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Рассмотрим основные аспекты использования парного и множественного регрессионного анализа в агрохимических исследованиях. Наиболее часто используют уравнения парной линейной регрессии: $Y = a_0 + bX$, где Y – зависимая переменная, X – независимая переменная, a_0 – в частном случае [5], константа, показывающая, какую величину примет зависимая переменная в случае отсутствия действия независимой переменной, b – эмпирический коэффициент, который, как и a_0 рассчитывают с определенной ошибкой. Известно, что максимально точно линейные уравнения отражают связи между зависимой (Y) и независимой (X) переменной только в узком их (переменных) интервале варьирования. Однако если провести даже краткий и поверхностный анализ статей, поступающих в отечественные и зарубежные журналы биологического и сельскохозяйственного профиля, можно довольно быстро убедиться в широком применении уравнений парной линейной регрессии при анализе зависимостей, по-

лученных при проведении полевых и вегетационных экспериментов, а также аналитических работ.

В следующем примере показаны как положительные стороны в использовании линейных уравнений, так и возможные отрицательные последствия их применения. Рассмотрим зависимость экстрактивности зерна пивоваренного ячменя от содержания белка (рис. 1). Проведенный авторами статьи анализ собственных и литературных данных зависимостей экстрактивности зерна (Y , % а.с.в.) пивоваренного ячменя от содержания сырого белка ($X = N_{\text{общ}} \times 6.25$, % а.с.в.) с учетом географического принципа показал, что в северных и центральных областях Нечерноземья России между этими показателями наблюдается статистически значимая (при $p < 0.05$) линейная отрицательная (рис. 1а), южнее Нечерноземья – слабая отрицательная или ее отсутствие (рис. 1б), еще южнее – отсутствие или слабая положительная зависимость (рис. 1в). Необходимо отметить, что на рис. 1а и 1в графики линейного уравнения и уравнения 2-го порядка практически совпадают.

Объединение всех данных в один массив при общем числе наблюдений $n = 120$ и проведение графического и регрессионного анализа показало, что зависимость экстрактивности зерна от содержания белка носит нелинейный характер. При этом данная зависимость наиболее точно (по величине коэффициента детерминации R^2) описывается уравнением 2-го порядка с четко выраженной точкой экстремума (ТЭ) (в данном случае минимума функции), находящейся в пределах полученных экспериментальных данных (рис. 1г). Следует отметить, что при анализе только данных, представленных на рис. 1б и (или) 1в, можно прийти к заключению о том, что белковость зерна пивоваренного ячменя в данных (конкретных) почвенно-климатических условиях не имеет никакого значения для накопления экстрактивных веществ по причине отсутствия статистически значимых зависимостей между ними.

Приведенный выше пример показывает, что линейные уравнения регрессии не могут являться корректным методом статистической обработки полученных данных, т.к. в зависимости от того, в какой части биологического интервала (в начале, в середине или в конце) получены экспериментальные данные, уравнения будут показывать некорректные, взаимоисключающие и (или) противоречивые результаты. При этом линейные уравнения без использования графического анализа не будут показывать реально существующие зависимости (закономерности) и вполне вероятно будут вводить в заблуждение относительно степени

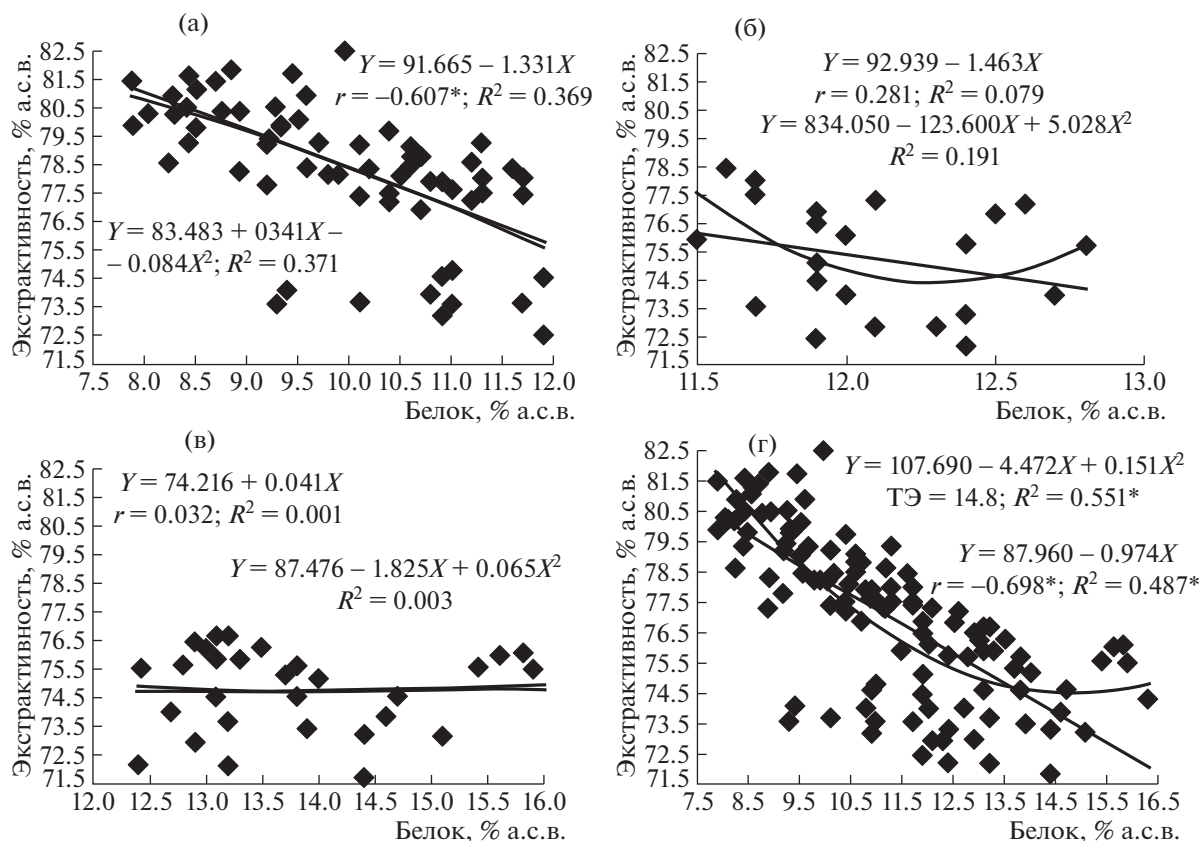


Рис. 1. Зависимость экстрактивности (Y) от содержания белка в зерне пивоваренного ячменя (X), ТЭ – точка экстремума.

тесноты связи между переменными и направленности их действия.

В настоящее время общепризнано, что в подавляющем большинстве случаев зависимости между различными показателями в биологических и агрохимических исследованиях максимально точно отражают нелинейные уравнения. Одними из наиболее часто применяемых в агрохимических исследованиях являются уравнения половинной степени: $Y = a_0 + bX - cX^{0.5}$ и 2-го порядка: $Y = a_0 + bX - cX^2$, которые используют при исследовании зависимостей урожайности и показателей качества сельскохозяйственных культур от доз минеральных и органических удобрений [6–8], норм применения биологических и химических средств защиты растений [9–12], а также от гидротермических условий, складывающихся в период вегетации (суммы осадков, суммы эффективных температур, ГТК, продолжительности межфазных периодов и т.п.) [13, 14]. Чаще всего уравнения половинной степени используют в тех случаях, когда точка экстремума не достигается по причине выхода величин зависимой переменной на плато, а 2-го порядка – когда точка

экстремума (минимума или максимума функции) объективно существует и поэтому может быть определена. При этом точка экстремума может находиться или не находиться в пределах данных, полученных при проведении экспериментальных и (или) аналитических работ. Ввиду того, что уравнения половинной степени и 2-го порядка в основном используют при статистической обработке данных, полученных при проведении многофакторных экспериментов, эффективность и особенности их применения будут рассмотрены ниже.

Следующим видом нелинейных уравнений, используемых в агрохимических исследованиях, служат уравнения 3-го порядка: $Y = a_0 + bX + cX^2 + dX^3$ с 2-мя точками экстремума. Причем обе точки экстремума или только одна из них может находиться в пределах полученных экспериментальных данных. В качестве примера уравнения 3-го порядка (рис. 2) представлена зависимость показателя “число падения” зерна озимой ржи от толщины зерновки [15]. При этом, как видим, обе точки экстремума (2.12 и 3.13 мм) в уравнении 3-го порядка находятся в пределах полученных данных.

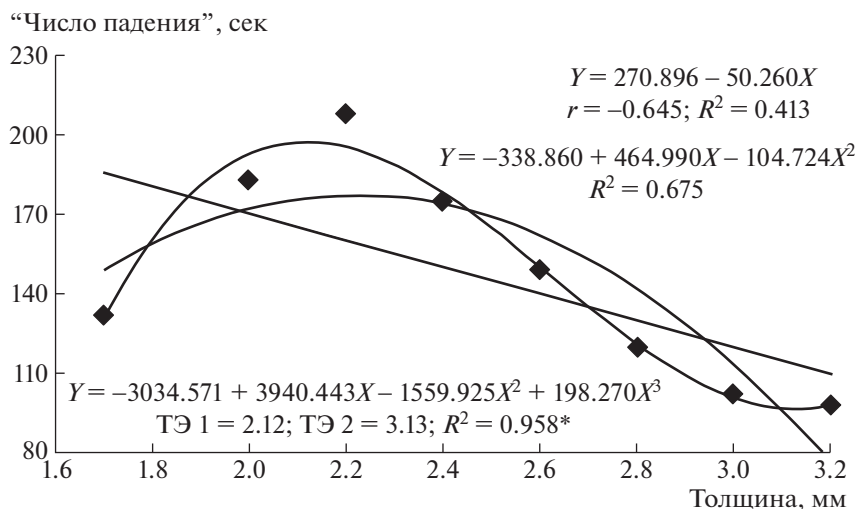


Рис. 2. Зависимость показателя “число падения” от толщины зерновки.

На рис. 2 также приведены линейное и уравнение 2-го порядка, коэффициенты корреляции (r) которых статистически незначимы на принятом в агрохимических исследованиях уровне значимости ($p < 0.05$). К тому же оба уравнения показывают и меньшую точность (по величине R^2), чем уравнение 3-го порядка. Изложенное выше и данные, представленные на рис. 1 и 2, подтверждают важность графического анализа полученных экспериментальных данных, а также необходимость расчета не только уравнения парной линейной регрессии, но и уравнений 2-го и 3-го порядка. Так как только в этом случае появляется определенная уверенность в корректном проведении статистической обработки данных и объективно вытекающих из этого заключений и выводов.

В научной литературе имеются сведения и об использовании в агрохимических исследованиях уравнений более высоких порядков (в частности, шестого) [16]. Как и в предыдущем случае, сначала авторы провели графический анализ полученных данных, а затем рассчитали уравнения, которые с максимальной точностью отражали наблюдаемые ими зависимости.

Логарифмические уравнения применяются в тех случаях, когда одна из переменных переведена в \lg или по своей сути является \lg . Известно, что показатель рН по своей сути является \lg [17]. По этой причине какие-либо зависимости, в которых данный показатель служит в качестве зависимой (Y) или независимой переменной (X) будут иметь следующий вид: $Y(\text{pH}) = a_0 + b \lg X$ или $\lg Y = a_0 + bX(\text{pH})$. Чаще всего в практической деятельности перевод в логарифмы необходим при

анализе зависимостей величины рН от содержания обменного Al^{3+} и подвижных форм Fe и Mn ($\lg \text{Al}^{3+}$, $\lg \text{Fe}$ и $\lg \text{Mn}$ соответственно), а также и других показателей почвенного плодородия от величины рН [18]. Необходимость перевода зависимой переменной в \lg может возникнуть и для выявления зависимости урожайности ($\lg Y$) от величины рН в контрольных (без внесения как минеральных, так и органических удобрений) вариантах полевых опытов. Необходимо отметить, что в случае применения удобрений зависимости урожайности (Y) от содержания элементов минерального питания в почве (X_1 , мг/кг) и находящихся в составе вносимых удобрений (X_2 , кг д.в./га) не имеют биологического смысла [7], т.к. определить вклад каждого из 2-х источников минерального питания растений (в рассматриваемом случае почвы (X_1) и удобрений (X_2)) в формировании урожая, показателей биохимического состава и технологических качеств возможно только при использовании схем опытов с удобрениями на нескольких контрастных фонах плодородия почвы, при этом более точно — при использовании изотопов соответствующих элементов минерального питания: ^{15}N , ^{32}P , ^{40}K и т.п.

В качестве примера использования регрессионного анализа данных, полученных авторами статьи при проведении длительного полевого опыта с известкованием [17, 18], показано (табл. 1), что содержание обменного Al^{3+} в почве (рис. 3а) и подвижных форм Fe и Mn статистически значимо связано с величиной pH_{KCl} , при этом более тесно — с содержанием Al^{3+} , менее тесно — с содержанием Mn. На рис. 3б представлена зави-

симость урожайности зерна ячменя ($\lg Y$), полученная в засушливый год, от величины рН.

Известно, что множественный регрессионный анализ (МРА) как статистический метод предполагает установление зависимостей от 2-х и более независимых переменных, причем в некоторых случаях независимые переменные могут иметь несколько градаций каждая. При этом в подавляющем большинстве случаев в агрохимических исследованиях необходим учет взаимодействия независимых переменных [10], что существенно усложняет проведение статистического анализа. По этой причине без использования современной компьютерной техники с соответствующим программным обеспечением невозможно учесть такое многообразие факторов, а также выявить степень их взаимодействия. При этом одновременно должны соблюдаться все обязательные ограничения [3] по корректному использованию МРА, которые будут рассмотрены ниже.

Метод МРА изначально был рекомендован, применялся (и применяется) при статистической обработке данных специально спланированных полнофакторных полевых опытов с удобрениями [2, 4], затем – с различными выборками из полнофакторных схем [4, 6, 9], а в последующий период – со средствами защиты растений при раздельном и комплексном их применении с удобрениями. Дальнейшее совершенствование схем многофакторных полевых опытов с комплексным применением средств химизации пошло по пути введения в их схемы факторов с прерывным шагом, в частности, доз удобрений и норм расхода химических средств защиты растений [9–12]. Это стало возможным в связи с развитием новых методических подходов к проведению полевых

Таблица 1. Зависимости величин рН (Y , ед.) от содержания обменного Al^{3+} и подвижных форм Fe и Mn (2021 г.)

Уравнение регрессии ($n = 7, p < 0.05$)	r	R^2
$Y = 4.233 - 0.552 \lg X (\text{Al}^{3+})$	-0.933*	0.870
$Y = 8.718 - 3.348 \lg X (\text{Mn})$	-0.820*	0.672
$Y = 7.184 - 3.348 \lg X (\text{Fe})$	-0.892*	0.796

Примечание. $n = 7$ – общее число наблюдений, *статистически значимо при $p < 0.05$. То же в табл. 2–5 и на рис. 1–8.

опытов при комплексном применении средств химизации [10] и развитием отечественной электронно-вычислительной техники, а затем и с появлением персональных компьютеров. При этом были созданы предпосылки для создания различных пакетов отечественных прикладных статистических программ, в частности, “Stat” (ВИУА им. Д.Н. Прянишникова, 1991, автор – Арутюнова Л.В.), что дало возможность проведения статистической обработки данных, полученных при проведении полевых опытов, в схемы которых были включены факторы как с непрерывным, так и с прерывным шагом [9–12].

Как было отмечено выше, полевые опыты могут быть как полнофакторные, а также представлять определенные выборки из них: 1/4, 1/8 или 1/9 [4, 6, 10]. В качестве примера полнофакторного полевого опыта можно привести четырехфакторный полевой опыт с озимой рожью, проведенный по схеме $4 \times 2 \times 2 \times 2$ в лаборатории агрохимии НИИСХ Северо-Востока (ныне ФАНЦ Северо-Востока, г. Киров) [19]. В 32-х вариантах опыта на постоянном фосфорно-калийном фоне изучали эффективность возрастающих доз азот-

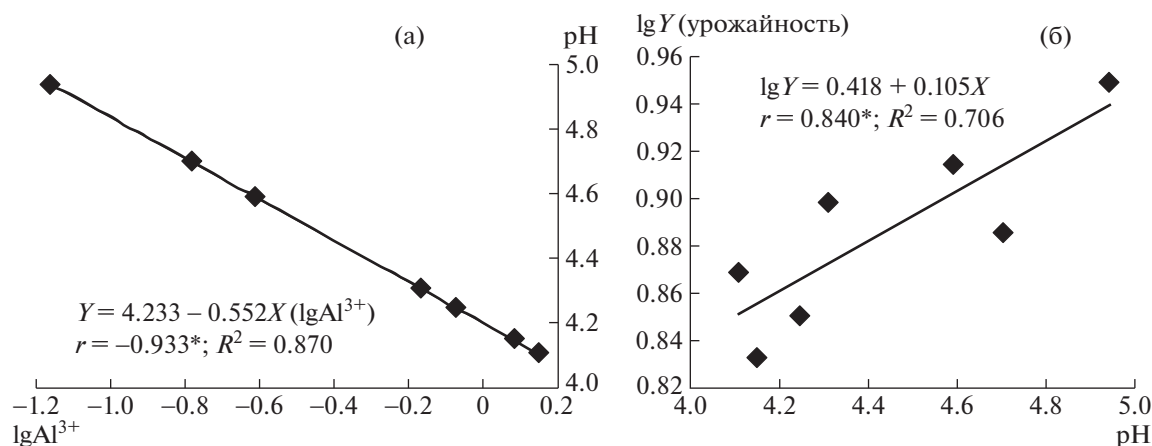


Рис. 3. Зависимость рН (Y , ед.) от содержания обменного Al^{3+} ($\lg X$) – (а) и урожайности зерна ячменя ($\lg Y$) от величины рН (X , ед.) – (б).

Таблица 2. Зависимость урожайности и показателей качества зерна (Y) озимой ржи от примененных средств химизации (X) (среднее за 3 года) [19]

Показатель (Y)	Уравнение регрессии ($n = 96$)	R^2
Урожайность, ц/га	$Y = 17.19 + 7.95N - 1.11N^2 + 5.56F + 4.34T + 0.93N \cdot T - 4.40K - 2.05N \cdot \Phi$	0.98*
ЧПС, шт./м ²	$Y = 280.2 + 35.08N + 0.50N \cdot F$	0.78*
Масса 1000 зерен, г	$Y = 25.60 + 2.24T - 0.73N - 0.18F$	0.90*
Натура, г/л	$Y = 686.87 + 7.44T - 6.41N$	0.77*
Белок, %	$Y = 8.45 + 0.94N - 0.23F \cdot K - 0.10N \cdot T$	0.89*
Крахмал, %	$Y = 49.66 + 1.03K \cdot T - 0.34N^2$	0.72*

Примечание. N – дозы азотных удобрений (в кодированных единицах: 0, 2, 3, 4), F – фундозол, K – кампозан, T – тилт (0 – без обработки, 1 – рекомендуемые производственные нормы), ЧПС – число продуктивных стеблей.

ных удобрений (N_0 , N_{60} до посева, N_{90} (N_{60} до посева + N_{30} весной) и N_{120} (N_{60} до посева + N_{30} весной + N_{30} в фазе трубкования)), фунгицидов фундозола и тилта, а также ретарданта кампозана (без обработки и их рекомендуемые производственные нормы). В табл. 2 представлены некоторые результаты статистической обработки данных, полученных при проведении упомянутого выше четырехфакторного опыта. Как видим, практически большинство зависимостей имели линейный характер. Исключение составляли зависимости урожайности зерна и содержания крахмала в нем от доз азотных удобрений. Например, с усилением азотного питания урожайность зерна возрастала, однако каждое последующее увеличение доз азота приводило к меньшему ее увеличению по сравнению с предыдущей дозой ($+N - N^2$). Содержание крахмала с увеличением доз азота снижалось, при этом каждая последующая доза в большей степени снижала содержание крахмала по сравнению с предыдущей ($-N^2$).

Многофакторный полевой опыт, представляющий собой выборку (1/4) из полнофакторного эксперимента ($6 \times 6 \times 6$) с 6-ю дозами азотных, фосфорных и калийных удобрений был заложен и его проводили также в НИИСХ Северо-Востока в 6-польном зернопаротравяном полевом севообороте: чистый пар–озимая рожь–ячмень с подсевом клевера–клевер одногодичного использования (сено)–яровая пшеница–овес [20]. Схема опыта включала 54 варианта в двукратной повторности. За единицу доз азотных (N_{aa}), фосфорных (P_{cd}) и калийных (K_x) удобрений под зерновые культуры приняты следующие величины (код. ед.): 000 – без удобрений, 111 – (NPK)60, 222 – (NPK)60, 333 – (NPK)90, 444 – (NPK)120 и 555 – (NPK)150. Под клевер внесение удобрений в опыте не предусмотрено.

Ниже в качестве примера статистической обработки данных приведено уравнение, отражаю-

щее зависимость содержания подвижного фосфора (Y , мг/кг) от доз азотных, фосфорных и калийных удобрений (код. ед.): $Y = 53.6 + 13.2P - 1.1P^2 + 1.2N \cdot P - 2.56N$, $R^2 = 0.809^*$ (рис. 4). Калийные удобрения на содержание подвижного фосфора достоверного влияния не оказали. Существенное влияние на содержание фосфора оказали фосфорные и азотные удобрения, а также их совместное применение. При этом с возрастанием доз фосфорных удобрений содержание фосфора в почве увеличивалось, однако каждое последующее их возрастание замедляло темпы роста содержания фосфора ($+P - P^2$). Применение азотных удобрений при одностороннем внесении снижало ($-N$), а совместное внесение азотных и фосфорных – повышало ($+N \cdot P$) содержание фосфора в почве. Необходимо отметить, что в целях наглядности на подобного рода рисунках, отражающих зависимость показателей плодородия почв от доз удобрений (рис. 4), возможно использование шкалы единиц доз удобрений в д.в. Однако при этом необходима обязательная ссылка на то, что независимые переменные при расчете уравнения регрессии были представлены в кодированных единицах.

Уравнение, отражающее зависимость содержания калия в почве (Y , мг/кг) от доз азотных, фосфорных и калийных удобрений (код. ед.), показало, что его содержание зависело от всех 3-х видов примененных в опыте удобрений: $Y = 154.835 + 17.404K - 1.339K^2 + 0.941N \cdot K + 0.918P \cdot K - 1.051N \cdot P$, $R^2 = 0.870^*$. Однако известно, что графическое изображение от 3-х независимых переменных (в данном случае N , P и K) в трехмерном пространстве невозможно. Поэтому в данном случае для построения графических изображений величина одной из независимых переменных (выбор ее определяется целью и задачами проведения исследований) фиксируется на одном уровне, а затем воспроизводится графиче-

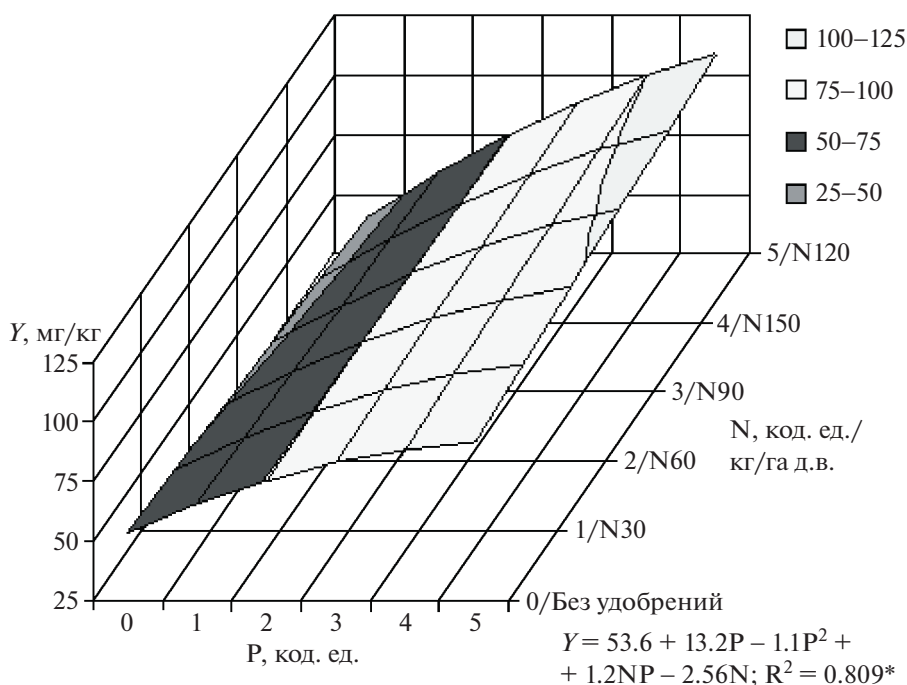


Рис. 4. Зависимость содержания подвижного фосфора в почве (Y) от доз азотных (N) и фосфорных удобрений (P).

ческое изображение уже от двух независимых переменных, как это представлено на рис. 4. При этом в тексте или подрисуночной надписи необходима обязательная ссылка на то, что величина одной из трех независимых переменных зафиксирована на определенном уровне.

МРА может быть использован и при изучении вопросов изменения содержания и фракционно-группового состава органического вещества (гумуса) в почве при проведении опытов с использованием полнофакторных схем или выборок из них с непрерывным или прерывным шагом независимых переменных. Проведение авторами статьи регрессионного анализа экспериментальных данных, опубликованных в работе [21], позволило рассчитать уравнения множественной нелинейной регрессии (полинома половинной степени и 2-го порядка), отражающие зависимость со-

держания в почве суммы гуминовых кислот от изученных доз минеральных удобрений с прерывным шагом и органических удобрений – с непрерывным шагом доз (табл. 3, 4, рис. 5).

Вызывает сожаление тот факт, что авторы публикации не применили множественный регрессионный анализ полученных ими же экспериментальных данных и не использовали графический анализ, который значительно облегчил бы интерпретацию и вызывал бы большее доверие к представленной работе, а также способствовал бы более объективному объяснению закономерностей, полученных авторами при проведении полевого эксперимента.

Важное практическое значение уравнений регрессии состоит в том, что их использование позволяет применить интерполяцию. Это дает возможность получить сведения о величине зависи-

Таблица 3. Зависимость содержания гуминовых кислот (Y , %) от доз минеральных (X_1) и органических удобрений (X_2)

Уравнение	Уравнение регрессии ($n = 12$)	R^2	ТЭ
Линейное (1)	$Y_1 = 28.125 - \text{const}$	–	–
Квадратичное (2)	$Y_2 = 24.406 + 12.184X_1 - 9.266X_1^2 + 0.776X_2^2$	0.650*	0.657
В степени 0.5 (3)	$Y_3 = 25.657 + 3.319X_1^{0.5} + 6.654X_2 - 7.554X_2^{0.5}$	0.629*	–

Примечание. X_1 (код. ед.): 0 – без удобрений, 0.33 – 1/3 дозы, 0.5 – 1/2 дозы, 1 – полная доза; X_2 (код. ед.): 0 – без удобрений, 1 – навоз 40 т/га, 2 – навоз 80 т/га [21]; ТЭ – точка экстремума в зависимости от доз минеральных удобрений. То же в табл. 4 и на рис. 5.

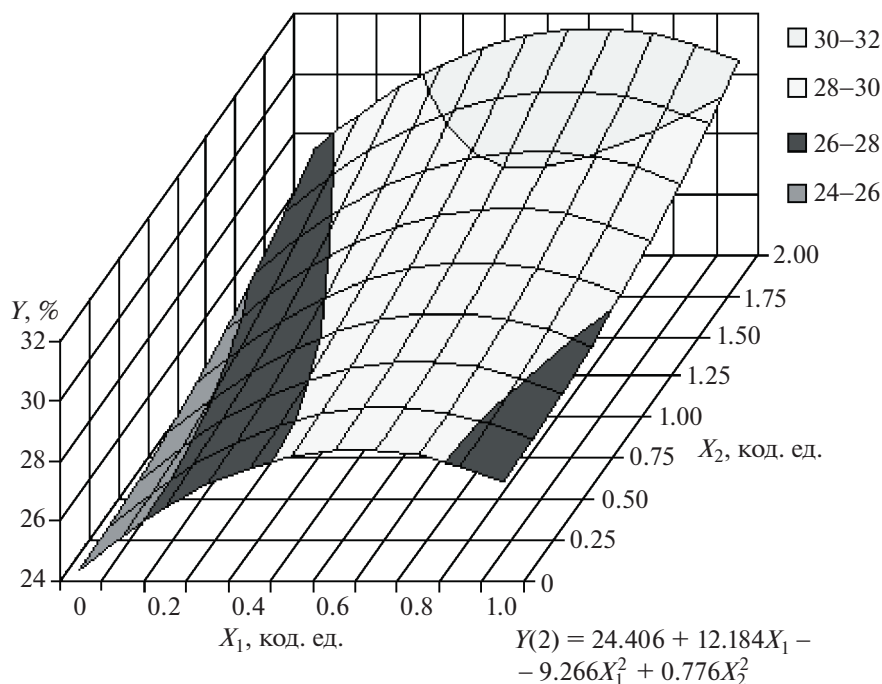
Таблица 4. Алгоритм и результаты сравнительной оценки точности уравнений

$Y_2 = 24.406 + 12.184X_1 - 9.266X_1^2 + 0.776X_2^2$						$Y_3 = 25.657 + 3.319X_1^{0.5} + 6.654X_2 - 7.554X_2^{0.5}$		
X_1	X_2	Y_3	Y_{p2}	$(Y_3 - Y_{p2})$	$(Y_3 - Y_{p2})^2$	$Y_{т3, п3}$	$(Y_3 - Y_{p3})$	$(Y_3 - Y_{p3})^2$
0	0	25.9	24.4	1.5	2.25	25.7	0.2	0.04
0.33	0	27.6	27.4	0.2	0.04	27.6	0.0	0.00
0.5	0	29.8	28.2	1.6	2.56	28.0	1.8	3.24
1	0	26.9	27.3	-0.4	0.16	29.0	-2.1	4.41
0	1	25.8	25.2	0.6	0.36	24.8	1.0	1.00
0.33	1	26.5	28.2	-1.7	2.89	26.7	-0.2	0.04
0.5	1	26.2	29.0	-2.8	7.84	27.1	-0.9	0.81
1	1	28.1	28.1	0.0	0.00	28.1	0.0	0.00
0	2	25.7	27.5	-1.8	3.24	28.3	-2.6	6.76
0.33	2	30.7	30.5	0.2	0.04	30.2	0.5	0.25
0.5	2	33.6	31.3	2.3	5.29	30.6	3.0	9.00
1	2	30.7	30.4	0.3	0.09	31.6	-0.9	0.81
Σ					24.76	Σ		26.36

Примечание. Y_3 – экспериментальное содержание гуминовых кислот, %; Y_p – расчетное содержание гуминовых кислот, % (расчитанное по уравнению (2) или (3) (табл. 3); $(Y_3 - Y_p)$ – отклонения экспериментальных величин содержания гуминовых кислот (Y_3) в почве от расчетных (Y_p); Σ – сумма квадратов отклонений экспериментальных величин от расчетных или $\Sigma(Y_3 - Y_p)^2$.

мой переменной (Y) для неосуществленных в опыте промежуточных вариантов, что значительно углубляет традиционный анализ опытных и аналитических данных [6]. Например, используя

уравнение 2-го порядка (табл. 3) можно определить содержание гуминовых кислот (Y) при внесении отсутствующих в опыте дозах минеральных (X_1) и органических удобрений (X_2), в частности:

**Рис. 5.** Зависимость содержания гуминовых кислот (Y) от доз минеральных (X_1) и органических удобрений (X_2).

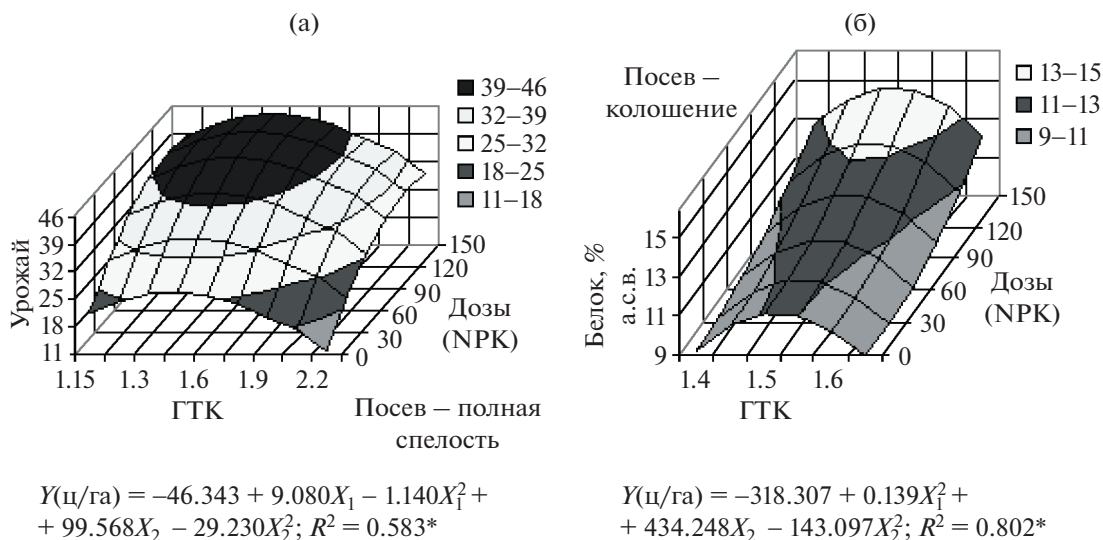


Рис. 6. Зависимость урожайности зерна пшеницы (а) и содержания белка в зерне ячменя (б) (Y) от доз минеральных удобрений (X_1 , код. ед.) и ГТК (X_2).

0.25 и 0.5, а также 0.75 и 1.25–27.1 и 29.5% соответственно.

МРА может быть применен и при изучении зависимостей урожайности и технологических качеств сельскохозяйственных культур от доз удобрений (код. ед.) и гидротермических условий (в частности, ГТК), складывающихся в период вегетации (посев–полная спелость) или, что особенно важно, в его основные межфазные периоды (рис. 6).

При этом в целях корректного исследования зависимостей урожайности и технологических качеств культур от гидротермических условий в период вегетации, рекомендуется учитывать температуру и уровень выпадения осадков ежедневно и связывать их не с календарными сроками (декада, месяц), а с датами наступления и продолжительностью основных фаз развития растений или в целом периода вегетации (посев (посадка)–полная спелость). Это связано с тем, что календарные сроки окончания осенней и возобновления весенней вегетации озимых культур и многолетних трав, а также сроки сева яровых зерновых и сроки посадки овощных и пропашных культур по годам не совпадают, а гидротермические условия летне-осеннего и весенне-летнего периодов вегетации, и в особенности их межфазных периодов, могут существенно отличаться от средних многолетних [22–24].

Исследования показали, что зависимости урожайности и технологических качеств зерновых культур от возрастающих уровней минерального питания и гидротермических условий межфаз-

ных периодов и всего периода вегетации имеют сложный нелинейный характер. Наиболее точно (по величине R^2) данные зависимости отражают уравнения множественной нелинейной регрессии, в частности, полинома половинной степени и 2-го порядка, с четко выраженными точками экстремума или областями оптимума. Максимальная урожайность зерна формируется при нормальном или избыточном увлажнении. При этом она наиболее тесно связана с гидротермическими условиями, складывающимися в период трубкование–цветение (яровая пшеница), трубкование–колошение (пивоваренный ячмень), трубкование–выметывание (пленчатый овес) и всего периода вегетации (посев–полная спелость) [13, 14, 24, 25] (рис. 6а). В максимальной степени (по величине R^2) показатели технологических качеств зерна пивоваренного ячменя (белок, экстрактивность и т.п.) определяется (в отличие от его урожайности [13]) гидротермическими условиями, складывающимися в вегетативный (посев–колошение; рис. 6б) и репродуктивный период (колошение–полная спелость) [14].

МРА может найти применение и в изучении взаимосвязей биохимического состава и технологических качеств культур. Например, его использование позволило получить уравнение множественной нелинейной регрессии, отражающее зависимость содержания сырой клейковины в зерне пшеницы (Y , %) от содержания белка ($X_1 = N_{\text{общ}} \times 5.7$, %) и массы 1000 зерен (X_2 , г): $Y = -41.928 + 0.081X_1^2 + 2.548X_2 - 0.028X_2^2$ (рис. 7). В разработанном уравнении все показатели каче-

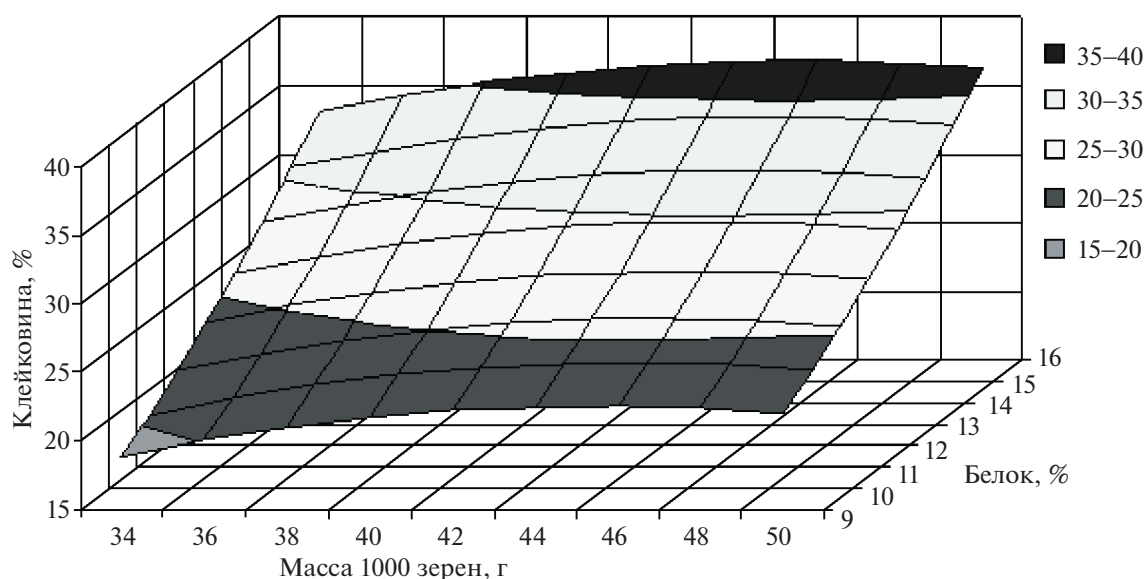


Рис. 7. Зависимость содержания сырой клейковины (Y , %) в зерне пшеницы от содержания белка (X_1 , %) и массы 1000 зерен (X_2 , г).

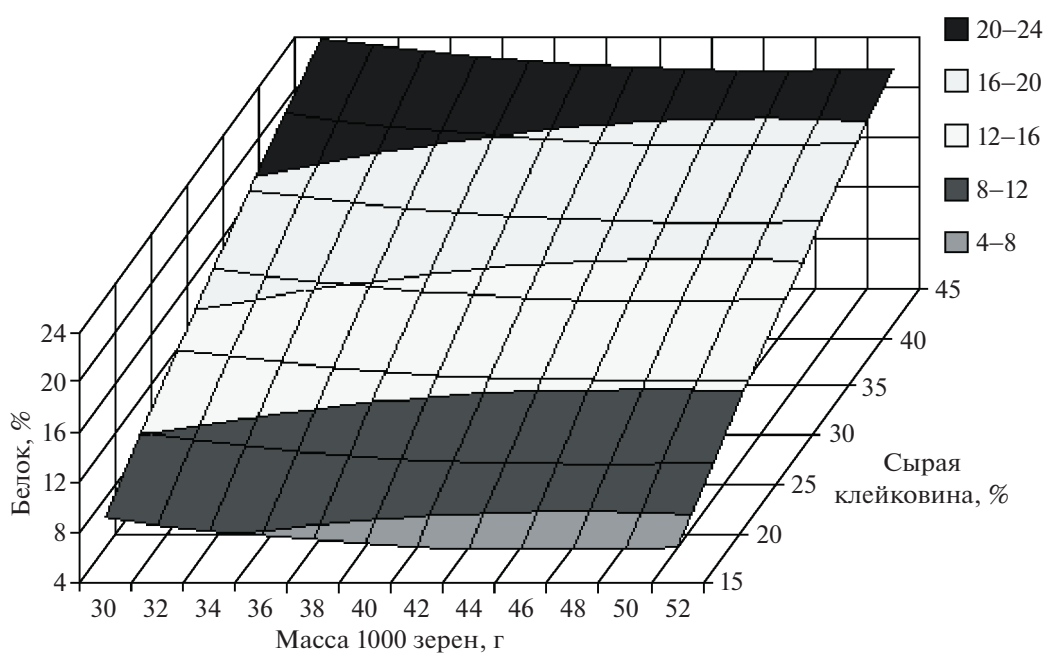


Рис. 8. Зависимость содержания белка (Y , % а.с.в.) в зерне пшеницы от содержания сырой клейковины (X_1 , %) и массы 1000 зерен (X_2 , г).

ства приведены на 12%-ную влажность или на воздушно-сухое вещество (в.с.в.) [26]. В последующий период была решена обратная задача и разработано уравнение 2-го порядка, отражающее зависимость содержания белка ($Y = N_{\text{общ}} \times 5.7$, %) в зерне пшеницы от содержания сырой клейковины (X_1 , %) и массы 1000 зерен (X_2 , г): $Y = 16.0570 + 0.6768X_1 - 0.0032X_1^2 - 0.7891X_2 + 0.0083X_2^2$ (рис. 8).

В разработанном уравнении содержание белка приведено на абсолютно сухое вещество (а.с.в.), а содержание клейковины и масса 1000 зерен – на 12%-ную влажность или на в.с.в. [27]. В дальнейшем был разработан алгоритм и представлены результаты проверки прогностических возможностей разработанных уравнений по независимым данным, т.е. используя экспериментальные дан-

Таблица 5. Алгоритм оценки точности прогноза уравнений

Данные X_1 , X_2 и Y_3 из работы [28], озимая пшеница сорта Московская 39					Данные X_1 , X_2 и Y_3 из работы [29], озимая пшеница 9-ти сортов					
$Y = -41.928 + 0.081X_1^2 + 2.548X_2 - 0.028X_2^2$					$Y = 16.0570 + 0.6768X_1 - 0.0032X_1^2 - 0.7891X_2 + 0.0083X_2^2$					
X_1	X_2	Y_p	Y_3	$(Y_3 - Y_p)$	X_1	X_2	Y_p	Y_3	$(N_3 - N_p)$	ДР
12.0	42.2	27.4	28.0	0.6	29.4	39.8	14.93	14.61	-0.056	0.157
13.6	42.0	30.7	27.8	-2.9*	30.8	41.3	15.43	15.49	0.010	0.162
12.8	43.4	29.2	30.1	0.9	32.1	38.5	16.41	15.90	-0.089	0.168
13.2	43.6	30.1	31.4	1.3	31.8	39.4	16.14	15.77	-0.064	0.166
13.2	44.5	30.1	30.2	0.1	29.2	41.1	14.68	14.46	-0.039	0.155
13.7	45.0	31.2	30.8	-0.4	32.2	38.2	16.50	15.94	-0.098	0.168
13.8	45.5	31.5	31.9	0.4	31.0	39.0	15.81	15.43	-0.067	0.163
13.7	45.3	31.2	30.4	-0.8	31.1	41.0	15.61	15.50	-0.019	0.163
13.9	46.1	31.7	30.7	-1.0	31.0	39.7	15.72	15.29	-0.075	0.162
13.8	46.0	31.5	31.6	0.1	$n = 9$		ЧВ = 0		ОП = 100%	
$n = 10$		ЧВ = 1		ОП = 90%	Примечание. X_1 – содержание сырой клейковины, %; X_2 – масса 1000 зерен, г; Y_p – расчетное содержание белка (расчет по разработанному уравнению регрессии), % а.с.в.; Y_3 – экспериментальное содержание белка ($N_{\text{общ}} \times 5.7$), % а.с.в.; $(N_3 - N_p)$ – отклонения экспериментальных величин содержания общего азота (N_3) в зерне пшеницы от расчетных (N_p), % а.с.в.; ДР – допускаемые расхождения (ГОСТ 10846-91), ±; n – общее число наблюдений; ЧВ – число величин, выходящих за пределы допустимых расхождений; ОП – оправдываемость (точность) прогноза, %					
Примечание. X_1 – содержание белка в зерне, %; X_2 – масса 1000 зерен, г; Y_p – расчетное содержание сырой клейковины (расчет по уравнению регрессии), %; Y_3 – экспериментальное содержание сырой клейковины, %; $(Y_3 - Y_p)$ – отклонения экспериментальных величин от расчетных, ±; n – общее число наблюдений; ±2.9* – отклонения выходят за пределы ±2%; ЧВ – число величин, выходящих за пределы ±2%; ОП – оправдываемость (точность) прогноза, %					Примечание. X_1 – содержание сырой клейковины, %; X_2 – масса 1000 зерен, г; Y_p – расчетное содержание белка (расчет по разработанному уравнению регрессии), % а.с.в.; Y_3 – экспериментальное содержание белка ($N_{\text{общ}} \times 5.7$), % а.с.в.; $(N_3 - N_p)$ – отклонения экспериментальных величин содержания общего азота (N_3) в зерне пшеницы от расчетных (N_p), % а.с.в.; ДР – допускаемые расхождения (ГОСТ 10846-91), ±; n – общее число наблюдений; ЧВ – число величин, выходящих за пределы допустимых расхождений; ОП – оправдываемость (точность) прогноза, %					

ные содержания белка, сырой клейковины и массы 1000 зерен, полученные другими авторами в опытах с иными сортами пшеницы и в иных, чем у авторов статьи, временных рамках и почвенно-климатических условиях (табл. 5). При этом оправдываемость прогноза обоих уравнений превысила 80% [26, 27]. Необходимо отметить, что в большинстве ГОСТов нашей страны используется понятие “допускаемое расхождение”, которое показывает максимальную аналитическую ошибку при определении какого-либо показателя (содержание белка, сырой клейковины, масса 1000 зерен, натура и т.п.).

Выше было отмечено существование ограничений по корректному использованию множественного регрессионного анализа. Первое – это проверка полученных экспериментальных или аналитических данных на нормальность распределения [3]. В тех случаях, когда распределение значимо отличается от нормального, рекомендуется использовать непараметрические методы [30–32]. Необходимость использования параметрического и непараметрического критериев обусловлена тем, что они имеют разные предпосылки для своего применения, а это существенно по-

вышает надежность результатов, если они совпадают по обоим критериям. Однако в некоторых случаях и непараметрические критерии могут оказаться неприменимыми, в частности, по причине малого количества опыто-лет [33, 34].

Второе – мультиколлинеарность или статистически значимая зависимость различных независимых переменных (X_1, X_2, \dots, X_n) между собой. Сравнительно быстро определить наличие или отсутствие мультиколлинеарности можно, используя матрицу коэффициентов корреляции между независимыми переменными. Для устранения мультиколлинеарности рекомендуется удалить из модели независимые переменные, ее вызывающие. При этом при проведении пошагового регрессионного анализа необходимо очень тщательно подходить к отсеву независимых переменных, обязательно учитывая биологический смысл изучаемых процессов, а в некоторых случаях, и физический смысл полученных оценок [33, 34]. Небольшой опыт применения авторами статьи множественного регрессионного анализа показал, что чаще всего мультиколлинеарность наблюдается в случае узкого биологического интервала варьирования независимых переменных.

Поэтому при планировании полевых, вегетационных и лабораторных экспериментов необходимо стремиться к тому, чтобы получить максимально возможный биологический интервал изучаемых факторов.

ЗАКЛЮЧЕНИЕ

Таким образом, приведенные примеры не охватывают даже малой части многообразных и сложных процессов, при описании которых могут быть использованы статистические методы, поэтому по мере количественного увеличения компьютерной техники и разработки программного обеспечения, возможности использования различных статистических моделей в естественных науках будут неуклонно расширяться. Следует лишь опасаться бездумного их использования и превращения науки в виртуальную компьютерную игру.

СПИСОК ЛИТЕРАТУРЫ

1. Правила для авторов. Требования к оформлению статей // *Зерн. хоз-во России*. 2019. № 4 (64). С. 77.
2. *Перегудов В.Н.* Метод наименьших квадратов и его применение в исследованиях. М.: Статистика, 1965. 340 с.
3. *Дрейпер Н.Н., Смит Г.* Прикладной регрессионный анализ / Пер. с англ. М.: Диалектика, 2007. 711 с.
4. *Перегудов В.Н.* Планирование многофакторных полевых опытов с удобрениями и математическая обработка их результатов. М.: Колос, 1978. 182 с.
5. *Доспехов Б.А.* Методика опытного дела / Изд. 2-е, перераб. и доп. М.: Колос, 1979. 416 с.
6. *Иванова Т.И.* Прогнозирование эффективности удобрений с использованием математических моделей. М.: Агропромиздат, 1989. 235 с.
7. *Державин Л.М.* Применение минеральных удобрений в интенсивном земледелии. М.: Колос, 1992. 272 с.
8. *Завалин А.А.* Удобрение сельскохозяйственных культур на осушаемых минеральных почвах. М.: ВИУА, 1995. 138 с.
9. *Петросян А.Г.* Влияние комплексного применения средств химизации на продуктивность озимой пшеницы и некоторые свойства дерново-подзолистой почвы: Автореф. дис. ... канд. с.-х. наук. М.: ВИУА, 1989. 24 с.
10. Программа и методика исследований в Географической сети опытов по комплексному применению средств химизации в земледелии. М.: ВИУА, 1990. 187 с.
11. *Цимбалит Н.И.* Взаимодействие и оптимизация применения минеральных удобрений и пестицидов в агробиоценозах Центрального района Нечерноземной зоны России: Автореф. дис. ... д-ра с.-х. наук. М.: ВИУА, 1993. 40 с.
12. *Цимбалит Н.И., Ладонин В.Ф., Алиев А.М., Цветкова Н.А.* Оптимизация сочетания азотных удобрений и пестицидов при возделывании озимой пшеницы // *Агрохимия*. 1996. № 8–9. С. 35–51.
13. *Пасынков А.В., Пасынкова Е.Н.* Урожайность зерна ячменя и ее зависимость от минерального питания и гидротермических условий в период вегетации // *Агрохим. вестн.* 2019. № 2. С. 33–38. <https://doi.org/10.24411/0235-2516-2019-10024>
14. *Пасынков А.В., Пасынкова Е.Н.* Зависимость показателей качества зерна пивоваренного ячменя от уровня минерального питания и гидротермических условий в период вегетации // *Агрохим. вестн.* 2020. № 3. С. 33–41. <https://doi.org/10.24411/1029-2551-2020-10036>
15. *Пасынков А.В., Андреев В.А., Завалин А.А., Пасынкова Е.Н.* Изменение показателей качества зерна озимой ржи при его фракционировании // *Достиж. науки и техн. АПК*. 2013. № 9. С. 36–40.
16. *Литвинович А.В., Лаврищев А.В., Буре В.М., Павлова О.Ю.* Изучение динамики изменения содержания подвижного железа в дерново-подзолистой легкой почве, мелиорируемой доломитом // *Агрохимия*. 2019. № 3. С. 44–53. <https://doi.org/10.1134/S0002188119030098>
17. *Небольсин А.Н., Небольсина З.П.* Теоретические основы известкования почв. СПб., 2005. 252 с.
18. *Небольсин А.Н., Небольсина З.П.* Известкование почв (результаты 50-летних полевых опытов). СПб.: ЛНИИСХ РАСХН, 2010. 354 с.
19. *Пасынков А.В., Ладонин В.Ф.* Эффективность средств химизации и энергетическая оценка их применения при возделывании озимой ржи // *Агрохимия*. 1996. № 6. С. 63–70.
20. *Пасынков А.В., Светлакова Е.В., Котельникова Н.В.* Влияние длительного применения минеральных удобрений на плодородие дерново-подзолистой почвы, продуктивность севооборота и качество зерна // *Агрохимия*. 2016. № 10. С. 38–47.
21. *Чеботарев Н.Т., Конкин П.И., Зайнуллин В.Г., Юдин А.А.* Изменение фракционно-группового состава и баланса гумуса под влиянием удобрений на дерново-подзолистой почве Евро-Северо-Востока // *Плодородие*. 2019. № 6. С. 25–28. <https://doi.org/10.25680/S19948603.2019.111.07>
22. *Минеев В.Г.* Основные направления исследований влияния погодно-климатических условий на эффективность удобрений // *Эффективность удобрений при различных погодных и климатических условиях*. Тр. ВИУА. 1985. С. 8–16.
23. *Минеев В.Г., Громыко О.И., Шербакова Н.И.* Изучение связи урожая яровых культур и эффективности удобрений с погодными условиями // *Погода и эффективность удобрений*. Бюл. ВИУА. 1985. № 75. С. 3–6.
24. *Завалин А.А., Пасынкова Е.Н., Пасынков А.В.* Зависимость урожая зерна яровой пшеницы от гидротермических условий межфазных периодов вегетации // *Плодородие*. 2010. № 4. С. 6–8.
25. *Пасынкова Е.Н.* Зависимость урожая зерна овса от доз азотных удобрений и гидротермических усло-

- вий межфазных периодов вегетации // Пробл. агрохим. и экол. 2010. № 4. С. 13–17.
26. Пасынков А.В., Завалин А.А., Пасынкова Е.Н. Совершенствование способа прогноза содержания сырой клейковины в зерне пшеницы // Рос. сел.-хоз. наука. 2020. № 2. С. 7–12.
<https://doi.org/10.31857/S2500-2627-2020-2-7-12>
 27. Пасынков А.В., Завалин А.А., Пасынкова Е.Н., Андреев В.Л. Сравнительная оценка различных методов прогноза содержания белка в зерне пшеницы // Рос. сел.-хоз. наука, 2021. № 3. С. 22–27.
<https://doi.org/10.31857/S2500262721030054>
 28. Кирпичников Н.А., Бижан С.П. Влияние фосфорных и цинковых удобрений в зависимости от известкования дерново-подзолистой почвы на урожай и качество зерна озимой пшеницы // Агрохим. вестн. 2020. № 3. С. 41–44.
<https://doi.org/10.24411/1029-2551-2020-10037>
 29. Колмаков Ю.В., Зелова Л.А., Ковтуненко А.Н., Кашуба Ю.Н. Источники высокого качества зерна озимой пшеницы // Зерн. хоз-во России, 2014. № 3. С. 46–48.
 30. Ивойлов А.В. Анализ данных агрономических исследований методами непараметрической статистики. Саранск: Изд-во Мордов. гос. ун-та им. Н.П. Огарева, 2000. 68 с.
 31. Якушев В.П., Буре В.М. Статистический анализ опытных данных. Непараметрические критерии. СПб., 2001. 61 с.
 32. Буре В.М. Методология статистического анализа опытных данных. СПб.: СПбГАУ, 2007. 141 с.
 33. Фрид А.С. Обоснование методических подходов к анализу данных многолетних опытов // Агрохимия, 2013. № 10. С. 75–96.
 34. Фрид А.С. Рекомендации к анализу данных многолетних полевых опытов (проект) // Агрохимия. 2022. № 3. С. 81–93.
<https://doi.org/10.31857/S0002188122030048>
 35. Небольсин А.Н., Небольсина З.П. Прогноз содержания легкоподвижного марганца в дерново-подзолистых почвах при известковании // Бюл. ВИУА. 2003. № 117. С. 111–113.

Features of the Use of Regression Analysis in Agrochemical Researches

A. V. Pasyнков^a and E. N. Pasyнкова^{a, #}

^aLeningrad Scientific Research Institute of Agriculture “Belogorka” –
Branch of Federal Research Centre of Potato Named after A.G. Lorch
ul. Institutskaya 1, Gatchinsky district, Leningrad region, p. Belogorka 188338, Russia

[#]E-mail: pasynkova.elena@gmail.com

Features are considered and practical aspects of the use of regression analysis in agrochemical research are discussed. It is shown that linear regression equations cannot be a correct method of statistical processing of experimental data, since depending on which part of the biological interval of independent variables (at the beginning, in the middle or at the end) data are obtained, the equations can show conflicting results. In most cases, the relationships between various indicators in agrochemical researches reflect non-linear equations as accurately as possible: logarithmic, half degree, second, third order, et all. Multiple regression analysis, taking into account limitations (sampling distribution, multicollinearity) can be used in the study of dependences of yield and quality indicators on doses and ratios of fertilizers, norms of biological and chemical plant protection products, hydrothermal conditions that develop in vegetation period, as well as in the study of the relationship between the biochemical composition and technological qualities of agricultural crops.

Key words: regression analysis, nonlinear equations, yield, fertilizers, plant protection products, hydrothermal conditions, technological qualities.