

Анализ влияния почвенных характеристик на результаты регрессионного моделирования урожайности озимой пшеницы по данным вегетационного индекса NDVI

А.В. Евстифорова¹, А.Ю. Денисова¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. В статье рассмотрена зависимость качества регрессионного моделирования урожайности озимой пшеницы по данным вегетационного индекса NDVI при наличии дополнительной априорной информации о плодородности почвы. Рассматриваемые в данной статье модели урожайности строятся на основе следующих групп факторов: факторы, полученные по значениям вегетационного индекса NDVI за сезон, биоклиматические данные, и данные сельхозтоваропроизводителей (фактическая урожайность, сорт и т.п.). В зависимости от набора используемых наиболее значимых факторов было выделено три модели урожайности, имеющие наибольший коэффициент детерминации. Эти модели являлись базой для сравнения с аналогичными моделями, построенными с учётом дополнительной априорной информации о плодородности почвы. При этом, в качестве характеристик плодородности почвы рассматривались балл бонитета почв, полученный по данным почвенных карт и справочников бонитетов почв, и данные агрохимического обследования. Проведённое сравнение исходных и дополненных моделей, показало, что качество модели на основе вегетационного индекса NDVI возрастает при использовании априорной информации о степени плодородности почвы.

1. Введение

Используемые в сельском хозяйстве методы прогнозирования урожайности требуют наличия большого количества измерений различных величин и наземных обследований, что существенно усложняет их оперативное применение на практике. Для автоматизации прогнозирования урожайности целесообразно использование геоинформационных систем (ГИС) и данных дистанционного зондирования Земли (ДЗЗ), поскольку они позволяют обойтись без трудоёмких полевых обследований и использовать вместо них пространственные и семантические данные о засеянной территории из единой базы данных. Применение ГИС и ДЗЗ обеспечивает необходимый уровень объективности данных и эффективный инструментарий для расчета параметров моделей урожайности сельскохозяйственных посевов, а также даёт возможность делать с их использованием прогнозы на основе имеющихся данных об участках.

Существующие методы использования данных ДЗЗ и ГИС для прогнозирования урожайности имеют тесную связь с методами, разработанными для этой задачи непосредственно в самой агрономической науке. Агрономические методы прогнозирования

урожайности строятся с использованием линейной регрессии или имитационного моделирования [1]. В первом случае, в качестве независимой переменной выбирается некоторый критический фактор: температура, влагообеспеченность, солнечная радиация и т.п., – и построенная модель иллюстрирует влияние данного фактора на урожайность. Выбор параметра модели производится экспертами в зависимости от того, влияние какого фактора на посев в данной конкретной ситуации сильнее, например, водный стресс играет существенную роль, а температурный нет. Прогноз с помощью таких моделей обычно формируется в начале сезона и представляет собой потенциальную урожайность, которая может быть достигнута за сезон. В аналогичных моделях, по данным ДЗЗ и ГИС, в качестве основного фактора, чаще всего выбирается суммарное значение вегетационного индекса за вегетационный период [2], дополнительно в модель могут быть включены и другие параметры, например, биоклиматические: климатическая зона, средняя температура за сезон и т.п.

Имитационное моделирование считается более точным, но и вместе с тем более трудоемким методом. Суть подхода заключается в оценке прироста сухой биомассы в течение всех фенологических стадий развития растения с помощью уравнений, связывающих биоклиматические параметры и прирост биомассы растительности. Результатом имитационного моделирования является оценка сухой биомассы растительности для каждой даты вегетационного периода. Среди имитационных моделей, для которых исследователи получили ряд результатов по использованию в них данных ДЗЗ и ГИС, находятся такие модели роста культур, как EPIC [3], WOFOST [4], SAFY [2] и т.п. Недостатком имитационных моделей является большое количество параметров, многие из которых не могут быть получены по данным ДЗЗ. Поэтому применение имитационных моделей по-прежнему требует значительного количества дорогостоящих полевых обследований и нетривиальной калибровки моделей для различных участков местности.

Таким образом, регрессионные модели на основе вегетационных индексов остаются наиболее перспективными средствами автоматизации прогнозирования урожайности в отличие от имитационных моделей. Современные исследования в данной области ориентированы на решение двух основных проблем: усовершенствование оценки по данным ДЗЗ конкретных факторов и изучение поведения моделей урожайности при наличии дополнительной информации, извлекаемой из доступных источников ГИС.

Целью настоящего исследования являлось построение регрессионной модели прогнозирования урожайности озимой пшеницы с использованием нормализованного разностного индекса вегетации NDVI [5] и изучение её поведения при включении в модель дополнительных факторов, отражающих плодородие почвы. Для характеристики плодородия почвы в модели использовались балл бонитета почвы и данные агрохимических обследований, обычно применяемые при прогнозировании урожайности в сельском хозяйстве. Однако несмотря на то, что плодородие почвы имеет важное значение для прогнозирования урожайности в сельском хозяйстве, данный факт не очевиден при использовании регрессионных моделей на основе NDVI.

2. Постановка задачи

Регрессионное моделирование применяется для поиска зависимости математического ожидания случайной величины y от набора факторов x_1, x_2, \dots, x_n . Функциональный вид зависимости определяется в соответствии с эвристическими предположениями или известными связями между наблюдаемыми независимыми параметрами и зависимой переменной. В настоящей статье в качестве зависимой переменной y выступает урожайность посевов озимой пшеницы, а функциональный вид зависимости выбран в виде линейной функции факторов со свободным членом:

$$y = \sum_{i=1}^n a_i x_i + a_0 \quad (1)$$

где $a_i, i = 0, \dots, n$ – коэффициенты модели.

Модель (1) строилась для нескольких наборов факторов. Первый базовый набор (далее БНФ) представлял собой совокупность факторов, доступных по данным производителей, метеорологических служб и данным ДЗЗ низкого разрешения, и включал в себя следующие показатели:

- 1) факторы, определяемые по ДЗЗ: сумма и максимум значения вегетационного индекса NDVI за период наиболее активной вегетации, а также длина периода вегетации и дата начала вегетации:

$$\xi_r = \sum_{t=r}^{T+r} \xi(t) \quad (2)$$

где $\xi(t)$ – это среднее для поля значение вегетационного индекса NDVI на дату t , r – относительная дата начала вегетационного периода, T – длина вегетационного периода.

- 2) биоклиматические факторы: максимальная и минимальная температура за сезон вегетации, максимальная и минимальная влажность за сезон вегетации, климатическая зона, в которой расположено поле.
- 3) данные сельхозтоваропроизводителей: сорт пшеницы, репродукция семян, наименование подразделения хозяйства, дата засева, географические координаты центра поля.

Второй набор факторов (ВНФ) включал в себя факторы, определённые по почвенным картам в ГИС, а именно: балл бонитета почвы и содержание гумуса. Балл бонитета почвы определялся по справочнику [6] для соответствующих нанесённым на карту типов почв.

Третий набор факторов (ТНФ) был построен как комбинация показателей, определяющих продуктивность почвы по данным агрохимических обследований, а именно: среднее содержание гумуса, фосфора, серы и калия в почве. Эти данные были получены по семантической информации карты участков по агрохимическим обследованиям в ГИС.

Настоящее исследование состояло из нескольких этапов:

- 1) Оценка оптимального подмножества БНФ (ОПБНФ) для линейной регрессионной модели (1) и построение базовой для сравнения модели, которую далее будем называть Модель 1.
- 2) Оценка модели по объединению ОПБНФ и ВНФ. Будем называть далее эту модель - Модель 2.
- 3) Оценка модели (1), построенной по ОПБНФ и ТНФ, далее Модель 3.
- 4) Сравнение качества прогнозирования, получаемого моделями, построенными на предыдущих этапах.

Построение всех моделей производилось с использованием метода наименьших квадратов [7]. Для характеристики качества моделей на этапе сравнения применялись коэффициент детерминации [7] и среднеквадратичная ошибка моделирования ε :

$$\varepsilon = \sqrt{\frac{1}{M} \sum_{j=1}^M (y_j - \hat{y}_j)^2} \quad (3)$$

где \hat{y}_j – полученное по модели значение урожайности для поля с номером j , y_j – фактическое значение урожайности для поля с номером j .

В следующем разделе более подробно описан процесс построения оптимального подмножества БНФ.

3. Построение оптимального подмножества базового набора факторов

Базовый набор факторов может быть определён различными способами в части факторов, рассчитываемых по данным ДЗЗ по формуле (2). А именно, основной вопрос заключается в определении параметров начала цикла вегетации r и длительности периода вегетации T . Для экспериментальной оценки этих параметров было рассмотрено два подхода.

Первый подход заключается в определении некоторого порога ξ_T значения вегетационного индекса. По данному порогу определяются минимальная τ_{\min} и максимальная τ_{\max} относительная дата пересечения графиком временного ряда $\xi(t)$ значения ξ_T , тогда начало цикла и длительность определяются следующим образом:

$$r = \tau_{\min}, T = \tau_{\max} - \tau_{\min} \quad (4)$$

На рисунке 1 приведен схематический график временного ряда NDVI и соответствующие первому подходу значения ξ_T , r и T .

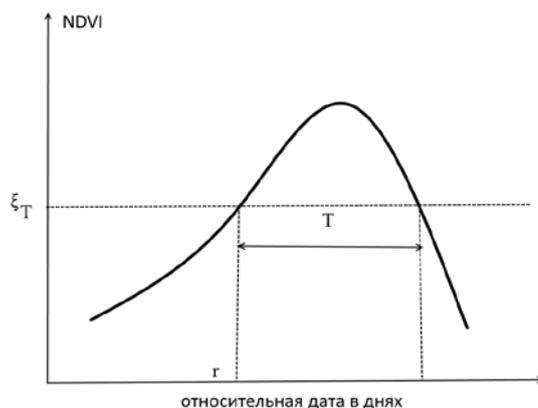


Рисунок 1. Схема оценки параметров факторов по ДЗЗ для первого подхода.

Второй подход заключается в определении периода вегетации как симметричного относительно максимума вегетационного индекса отрезка времени, т.е. если $\xi_{\max} = \max_t \xi(t)$ и $t_{\max} = \arg \max_t \xi(t)$, то r определяется как:

$$r = t_{\max} - T/2 \quad (5)$$

где T считается заранее заданным. Иллюстрация процесса определения факторов для второго подхода схематически изображена на рисунке 2.

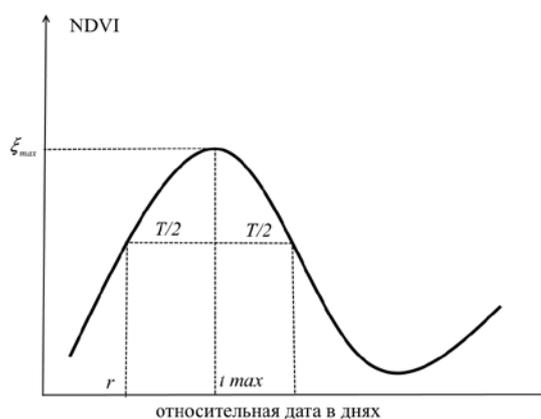


Рисунок 2. Схема оценки параметров факторов по ДЗЗ для второго подхода.

Предложенные способы определения параметров модели по данным ДЗЗ порождают разные подгруппы факторов. Факторами, соответствующими первому подходу, являются суммарное значение NDVI ξ_r для заданного ξ_T , дата начала сезона вегетации r и длительность периода вегетации T . Факторы, формируемые в результате второго подхода к определению признаков

по ДЗЗ, включают суммарное значение NDVI ξ_r , максимальное значение вегетационного индекса ξ_{\max} и полуширина периода вегетации $T/2$.

Для выбора оптимального подмножества БНФ были рассмотрены комбинации биоклиматических факторов и факторов по данным сельхозтоваропроизводителей с факторами по ДЗЗ полученными обоими подходами с различными параметрами. Полный перечень рассмотренных факторов ДЗЗ и названия соответствующих их использованию в БНФ моделей приведён в таблице 1.

Таблица 1. Перечень параметров для используемых расчёта факторов по ДЗЗ и названия соответствующих регрессионных моделей.

Название модели	Способ определения факторов по ДЗЗ	Параметры определения факторов по ДЗЗ
NDVI03	первый подход	$\xi_T = 0.3$
NDVI04	первый подход	$\xi_T = 0.4$
NDVI05	первый подход	$\xi_T = 0.5$
NDVI06	первый подход	$\xi_T = 0.6$
NDVI20	второй подход	$T/2 = 20$
NDVI30	второй подход	$T/2 = 30$
NDVI40	второй подход	$T/2 = 40$
NDVI50	второй подход	$T/2 = 50$
NDVI60	второй подход	$T/2 = 60$

Таким образом, варьируя порог для вегетационного индекса в первом подходе к определению факторов по данным ДЗЗ и длину полупериода вегетации во втором подходе, были получены 9 регрессионных моделей.

Определение оптимального подмножества БНФ производилось следующим образом:

- 1) Для начала оценивались коэффициенты всех девяти моделей по всему набору доступных факторов. Для этого категориальные независимые переменные (например, наименование подразделения хозяйства, сорт или репродукция семян) методом дихотомизации вводились в модель в виде наборов бинарных факторов, принимающих значение 0 или 1 в зависимости от выбранной категории.
- 2) Затем производилось центрирование факторов и нормализация дисперсии,
- 3) Нормализованные факторы и значения урожайности далее использовались в методе наименьших квадратов (МНК) для поиска неизвестных коэффициентов регрессии в выражении (1).
- 4) Для каждой из моделей производилась проверка адекватности построенной регрессионной зависимости путём проверки стандартных гипотез МНК [7]:
 - гипотеза о значимости коэффициентов регрессии (t-критерий Стьюдента),
 - гипотеза о статистической значимости уравнения регрессии в целом (F-критерий),
 - оценка коэффициента детерминации R^2 ,
 - анализ остатков. Остатки должны быть распределены нормально с нулевым математическим ожиданием и удовлетворять критериям независимости, отсутствия тренда и гетероскедастичности.
- 5) значимые факторы регрессионных моделей, прошедших проверку адекватности, принимались в качестве базовых наборов факторов.

Для определённости в дальнейшем сравнительном исследовании регрессионных моделей с использованием факторов почвы и без их использования в качестве БНФ были приняты

факторы модели с наивысшим коэффициентом детерминации среди всех моделей, признанных по итогам проверки адекватными.

Результаты оптимального подмножества базового набора факторов приводятся в разделе экспериментальных исследований.

4. Экспериментальное исследование

Для экспериментов были использованы данные выборочных хозяйств по Самарской области за 2015 в виде векторного слоя с границами посевов озимой пшеницы в ГИС. Сельскохозяйственными предприятиями для каждого поля были предоставлены данные о сорте пшеницы, репродукции семян, наименовании подразделения и фактическая урожайность посева за исследуемый период в центнерах на гектар. Данные о фактической урожайности используются для настройки регрессионной модели. Для этих полей с использованием пространственных запросов в ГИС были определены следующие параметры: климатическая зона по справочнику климатических зон Самарской области [8] и географические координаты центра поля. Значение климатической зоны представляло собой индекс от 1 до 3 в соответствии со значением: 1 - северная зона, 2 - центральная, 3 - южная климатическая зона.

Факторы по данным ДЗЗ, рассчитывались на основании временных рядов вегетационного индекса NDVI по снимкам MODIS космического аппарата Terra/Aqua с разрешением 250 метров в период с 1 апреля по 31 августа 2015 года. Для расчёта временных рядов NDVI применялся алгоритм, описанный в работах [9-11].

Климатические данные по температуре воздуха, относительной влажности и осадков были получены за весь период наблюдений. Они были собраны по архивам следующих метеостанций: село Кинель-Черкассы (№548621) с координатами 53°0'0"N 47°0'0"E, поселок Серноводск (№496568) с координатами 53°0'0"N 92°0'0"E, село Красноармейское (№824366) с координатами 52°0'0"N 7°0'0"E. Для отдельного поля климатические параметры определялись по ближайшей из рассмотренных метеостанций.

Почвенные параметры были получены из следующих источников. Балл бонитета определялся в соответствии с типом почвы на почвенной карте и таблице бонитетов типов почв в справочнике [6]. Данные агрохимического обследования были предоставлены ФГБУ Станцией агрохимической службы «Самарская».

Для построения базовых моделей применялась выборка из 127 полей для расчета Модели 1, из 20 полей для расчета Модели 2, из 32 полей для расчета Модели 3. Для расчета СКО моделирования использовалась вся выборка. Для нахождения коэффициентов множественной линейной регрессии и проверки адекватности моделей использовался программный продукт MATLAB, в котором проводились все вычисления.

Для определения базовой модели прогнозирования урожайности по данным NDVI, было получено 9 моделей, которые отличались суммой и максимумом значения вегетационного индекса NDVI за период наиболее активной вегетации, а также длиной периода вегетации и датой начала вегетации (2) (Таблица 2). Уровень значимости при проверке адекватности всех моделей составлял $\alpha = 0.05$.

Таблица 2. Коэффициент детерминации для моделей при выборе ОПБНФ

Название модели	Модель 1
NDVI03	0.7110
NDVI04	0.7162
NDVI05	0.6970
NDVI06	0.8035
NDVI20	0.7972
NDVI30	0.6999
NDVI40	0.6768
NDVI50	0.6405
NDVI60	0.7110

Наибольший коэффициент детерминации имеет модели NDVI06 и NDVI20. Однако для модели NDVI20 не выполняется критерий Лиллиефорса при проверке адекватности модели, что не позволяет использовать её для расчётов. Поэтому в качестве второй базовой модели была выбрана Модель 1, рассчитанная для БНФ с NDVI04 в качестве, факторов, рассчитанных по ДЗЗ.

Таким образом, в качестве ОПБНФ в дальнейшем рассматривались два набора значимых факторов Модели 1:

1) Для модели NDVI06: сумма и максимум значения вегетационного индекса NDVI за период наиболее активной вегетации, длина периода вегетации и дата начала вегетации, максимальная и минимальная температура за сезон вегетации, максимальная и минимальная влажность за сезон вегетации, репродукция семян, дата засева, географические координаты центра поля, остальные факторы оказались не значимыми.

2) Для модели NDVI04: сумма и максимум значения вегетационного индекса NDVI за период наиболее активной вегетации, длина периода вегетации и дата начала вегетации, максимальная и минимальная температура за сезон вегетации, максимальная и минимальная влажность за сезон вегетации, сорт пшеницы, наименование подразделения хозяйства, дата засева, географические координаты центра поля.

Далее были рассчитаны модели с объединенными ОПБНФ и ВНФ, а также ОПБНФ и ТНФ. Полученные четыре модели для двух рассматриваемых ОПБНФ при уровне значимости $\alpha = 0.05$, были также проверены на адекватность путём того же набора гипотез МНК. Все полученные модели для рассматриваемых выборок оказались адекватными. В таблице 3 представлены результаты оценки коэффициента детерминации и СКО ошибки моделирования для полученных моделей (Модели 2 и Модели 3) при наличии факторов, характеризующих плодородие почвы.

Таблица 3. Критерий детерминации и значение СКО для каждого вида моделей.

Название модели	Модель 2		Модель 3	
	R^2	СКО	R^2	СКО
NDVI06	0.9523	0.8866	0.8046	2.3931
NDVI04	0.9820	0.5447	0.8341	2.2051

Модели с использованием данных ВНФ показали лучшие результаты тестов, чем модели основе данных ОПБНФ и ТНФ, что позволяет сделать вывод о том, что использование почвенных параметров для прогнозирования урожайности повышает качество прогнозирования. Лучшим предсказанием с использованием ВНФ является модель на основе NDVI 04, как обладающая минимальным СКО и максимальным коэффициентом детерминации среди рассмотренных моделей.

Графики полученных регрессионных зависимостей для каждой из моделей приведены на рисунках 3-8, на которых красным цветом обозначен верхний доверительный интервал, зеленым – нижний доверительный интервал, синим – значение модели. Результаты экспериментов показали, что СКО моделирования без использования почвенных параметров составляет от 2.7 до 3 ц/Га и от 0.5 до 2.4 ц/Га с их использованием.

Таким образом, можно заключить, что для прогнозирования урожайности озимой пшеницы с использованием данных ДЗЗ целесообразно применять данные бонитировки почв, получаемые путём сопоставления справочников бонитировки с типом почвы по почвенным картам в ГИС. В качестве факторов ДЗЗ целесообразно выбирать параметры, рассчитываемые с помощью первого из описанных подходов к оценке суммарного значения NDVI за сезон с порогом 0,4.

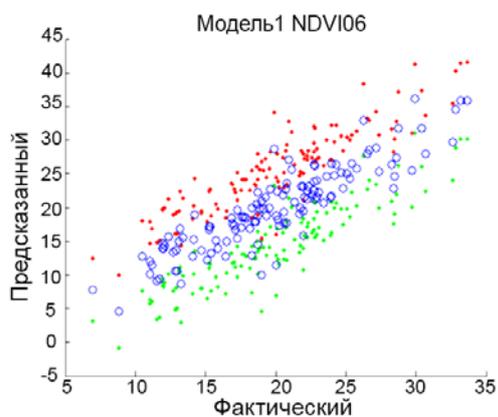


Рисунок 3. Диаграмма фактических и предсказанных значений урожайности для Модели 1 NDVI06.

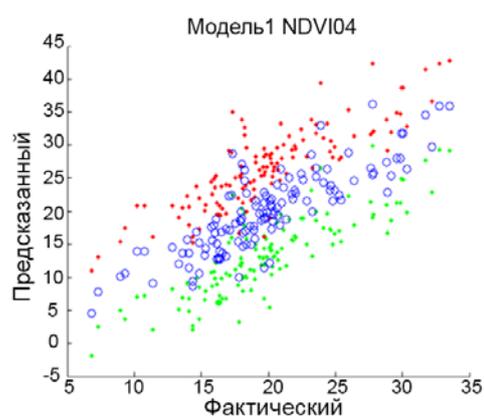


Рисунок 4. Диаграмма фактических и предсказанных значений урожайности для Модели 1 NDVI04.

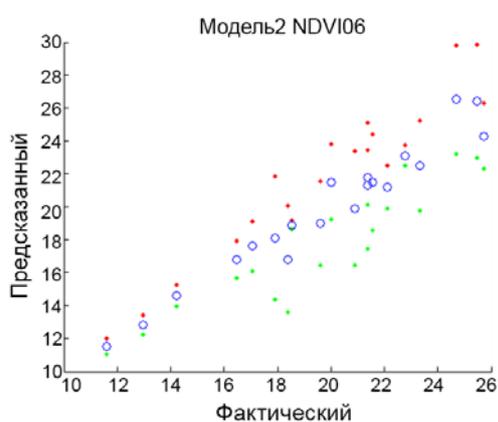


Рисунок 5. Диаграмма фактических и предсказанных значений урожайности для Модели 2 NDVI06.

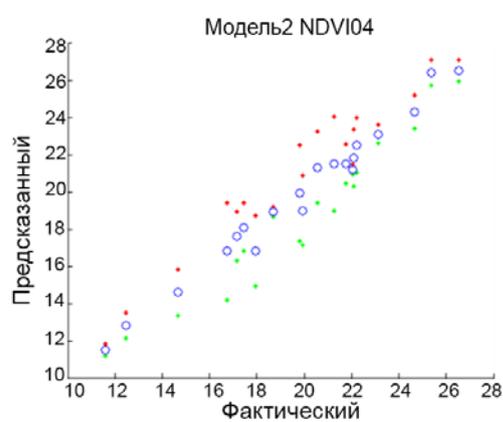


Рисунок 6. Диаграмма фактических и предсказанных значений урожайности для Модели 2 NDVI04.

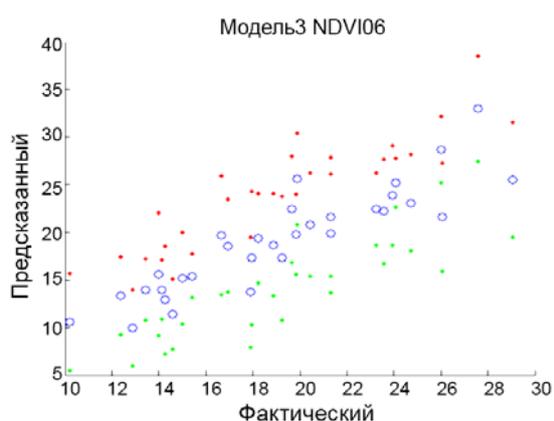


Рисунок 7. Диаграмма фактических и предсказанных значений урожайности для Модели 3 NDVI06.

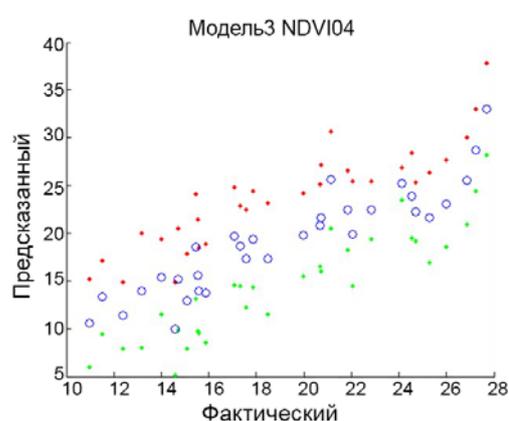


Рисунок 8. Диаграмма фактических и предсказанных значений урожайности для Модели 3 NDVI04.

5. Заключение

Проведённые в настоящей статье исследования касались вопроса использования дополнительных факторов, характеризующих плодородие почвы, в многофакторной регрессионной модели урожайности озимой пшеницы по данным ДЗЗ и ГИС. В качестве

данных ДЗЗ использовались значения суммарного NDVI за период вегетации, определяемые по временным рядам вегетационного индекса с использованием снимков низкого разрешения. В качестве данных ГИС выступали границы сельхозугодий, фактическая урожайность посевов и семантические данные о посевах, предоставленные по ряду хозяйств области. Кроме того, были использованы метеорологические факторы за рассмотренный сезон. В результате исследования был проведён отбор оптимального подмножества рассматриваемых факторов среди регрессионных моделей прошедших проверку адекватности и имеющих наибольший коэффициент детерминации. Полученное оптимальное подмножество базовых факторов комбинировалось с двумя группами факторов почвы, получаемых по различным источникам.

В результате сравнения регрессионных моделей с использованием почвенных факторов и без было установлено, что использование почвенных характеристик позволяет улучшить качество моделирования, при этом модели на основе балла бонитета почв являются более предпочтительными. Среднеквадратичная ошибка прогнозирования с использованием построенных моделей с использованием почвенных факторов составила от 0.5 до 2.4 ц/Га.

6. Благодарности

Настоящая работа была выполнена при поддержке грантов РФФИ № 16-37-00043 мол_а и № 16-29-09494 офи_м.

7. Литература

- [1] Каюмов, М.К. Программирование урожаев сельскохозяйственных культур / М.К. Каюмов. – М.: Агропромиздат, 1989. – 368 с.
- [2] Chahbi, A. Estimation of the dynamics and yields of cereals in a semi-arid area using remote sensing and the SAFY growth model / A. Chahbi, M. Zribi, Z. Lili-Chabaane, B. Duchemin, M. Shabou, et al. – London: Taylor & Francis Ltd., 2014. – Vol. 35(3). – P.1004-1028.
- [3] Брыксин, В.М. Использование модели биопродуктивности EPIC и космоснимков MODIS для прогнозирования урожайности зерновых культур / В. М. Брыксин, А. В. Евтюшкин // Современные проблемы дистанционного зондирования Земли из космоса. – М.: ИКИ, 2007. – Т. 4, № 2. – С. 189-196.
- [4] Ma, G. Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield / G. Ma, J. Huang, W. Wu, J. Fan, J. Zou, S. Wu // Mathematical and Computer Modelling. – 2013. – Vol. 58(3-4). – P. 634-643. DOI:10.1016/j.mcm.2011.10.038.
- [5] Quarmby, N.A. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction / N.A. Quarmby, M. Milnes, T.L. Hindle, N. Silleos // International Journal of Remote Sensing. – 2007. – P. 199-210. DOI: 10.1080/01431169308904332.
- [6] Трегубов, Б.А. Бонитировка почв пашни хозяйств Куйбышевской области / Б. А. Трегубов, Г. Г. Лобов, М. Г. Холина. – Куйбышев: Кн. изд-во, 1976. – 111 с.
- [7] Кобзарь, А.И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.
- [8] Васин, А.В. Формирование высокопродуктивных поливидовых агрофитоценозов кормовых культур в Среднем Поволжье / А. В. Васин. – Кинель, 2006. – 513с.
- [9] Воробьева, Н.С. Распознавание сельскохозяйственных культур по космическим снимкам с использованием алгоритма вычисления оценок / Н.С. Воробьева // Материалы Международной конференции и молодежной школы «Информационные технологии и нанотехнологии (ИТНТ-2015)», Самара, 2015. – С.83-88.
- [10] Vorobiova, N.S. Crops identification by using satellite images and algorithm for calculating estimates / N.S. Vorobiova // CEUR Workshop Proceedings. – 2016. – Vol. 1638. – P. 419-427. - DOI: 10.18287/1613-0073-2016-1638-419-427.
- [11] Воробьева, Н.С. Информационная технология раннего распознавания видов сельскохозяйственных культур по космическим снимкам / Н.С. Воробьева, В.В. Сергеев, А.В. Чернов // Компьютерная оптика. – 2016. – Т. 40, № 6. – С. 929-938. DOI: 10.18287/2412-6179-2016-40-6-929-938.

The research of soil characteristics influence on the results of regression modeling of winter wheat yield using NDVI vegetation index

A.V. Evstiforova¹, A.Y. Denisova¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The article considers the dependence of regression modeling of winter wheat yield on soil characteristics. The average value of the vegetation index, bioclimatic and agricultural data were used to calculate the primarily yield models. Best three models were chosen and then modified to take into account soil characteristics. We used soil score derived from soil maps and agrochemical inspection results as additional factors of modified models. The comparison of basic and modified models shown that winter wheat yield should be predicted with information about soil characteristics. The results of a research can be applied to forecasting yield of winter wheat at the regional level.

Keywords: regression modeling, NDVI vegetation index, winter wheat yield.