

Прогнозирование урожайности на основе многолетних космических наблюдений за динамикой развития вегетации

А.Б. Мурынин^{1,2}, В.Г. Бондур¹, В.Ю. Игнатьев^{1,3}, К.Ю. Гороховский¹

¹ Научный центр аэрокосмического мониторинга «Аэрокосмос», Москва, Россия

E-mail: office@aerocosmos.info

² Вычислительный центр им. А.А. Дородницына РАН, Москва, Россия

E-mail: amurynin@bk.ru

³ Московский физико-технический институт, Московская обл., Россия

E-mail: vladimir.ignatiev.mipt@gmail.com

Предлагается метод прогнозирования урожайности почв на основе многоспектральных космических изображений. Приводятся и анализируются результаты вычислительного эксперимента по прогнозированию урожайности различных сельскохозяйственных культур для различных областей. Проводится исследование по определению наиболее точной модели плодородия почв, результатом которого является модель, которая принимает значения вегетационных индексов за 5 месяцев, начиная с марта месяца и заканчивая июлем текущего года. На основании проведенного анализа делается вывод о точности осуществляемого прогноза и возможности расширения области применения полученной модели для прогнозирования урожайности сельскохозяйственных культур.

Ключевые слова: прогнозирование урожайности, множественная линейная регрессия, множественная нелинейная регрессия, вегетационный индекс.

Введение

Рассмотрим общие концепции, примененные для построения моделей плодородия. В основе моделей урожайности заложена зависимость между состоянием растительности в вегетационный период и собранным урожаем. Другими словами, если состояние растительности хорошее на протяжении всего периода роста, то вероятность получить хороший урожай увеличивается. В том же случае, если в какой либо промежуток произрастания состояние растительности отклонилось в сторону ухудшения, то вероятность получить хороший урожай уменьшается.

Предлагаемый метод прогнозирования можно описать следующим образом. Урожай определенной культуры на заданной территории должен достаточно достоверно предсказываться функцией, параметрами которой являются усредненные (по этой территории) значения вегетационного индекса в течение роста и созревания сельскохозяйственной культуры. Подобные методики используются и для оценки экологической обстановки (Bondur, 2011) Чем полнее история изменения индекса, тем точнее можно предсказывать урожай (Murynin et al., 2013). Соответственно, чем отдаленней период измерения индекса от момента созревания культуры, тем меньше будет предикативная сила модели. Исследование зависимости точности предсказания от отдаленности даты созревания является темой отдельной статьи. В данной работе дата предсказания фиксируется в июле месяце текущего года и в ней используются показания индекса в период март–июнь.

Концепция моделирования

Предложенная модель урожайности в общей формулировке выглядит как:

$$y_{kr} = f_{kr}(v(t), v(t+1), v(t+2), \dots),$$

где y_{kr} – прогнозируемое значение урожайности на окончание текущего сезона для территориального региона r и сельскохозяйственной культуры k ; f_{kr} – искомая функция прогноза урожайности для региона r и сельскохозяйственной культуры k ; $v(t)$ – значение индекса состояния растительности для данного участка посевов. Для данной модели в качестве единицы регионального деления выбраны административные области, так как информация об урожайности за предыдущие годы по отдельным культурам официально предоставляется службой государственной статистики Российской Федерации (РФ), и она находится в открытом доступе (http://www.gks.ru/bgd/regl/b11_14p/Main.htm). Наличие этой информации позволяет настроить свободные параметры модели на конкретный регион и сельскохозяйственную культуру.

Таким образом, среднее значение урожайности по административной области выглядит следующим образом:

$$\langle y_{kr} \rangle_{region} = \frac{\sum_{region} \sum_{region} f_{kr}(v(t, i, j), v(t+1, i, j), v(t+2, i, j), \dots) \cdot s(i, j)}{\sum_{region} \sum_{region} s(i, j)}. \quad (1)$$

На этапе обучения идет настройка модели и получение конкретных функций f_{kr} для каждой области и сельскохозяйственной культуры.

Обозначим среднее значение индекса состояния растительности по области в момент времени t как $\langle v(t) \rangle_{region}$. Среднее значение этого индекса по области вычисляется как:

$$\langle v(t) \rangle_{region} = \frac{\sum_{region} \sum_{region} v(t, i, j) \cdot s(i, j)}{\sum_{region} \sum_{region} s(i, j)}.$$

Поскольку на имеющихся в наличии данных статистики вариативность изменений урожая по отношению к амплитуде мала, то после разложения предложенной модели в ряд Тейлора основной вклад в точность прогноза будут давать линейные члены. В качестве упрощения нелинейными членами более высоких порядков можно пренебречь. В этом случае модель становится линейной, т.е. f_{kr} представляет собой линейную комбинацию от $v(t)$:

$$y_{kr} = \alpha_{0k} + \sum_{t=1}^T \alpha_{kr}(t) \cdot v(t),$$

формула (1) преобразуется:

$$\begin{aligned} \langle y_k \rangle_{region} &= \frac{\sum_{region} \sum_{region} \left(\alpha_{0k} + \sum_{t=1}^T \alpha_{kr}(t) \cdot v(t, i, j) \right) \cdot s(i, j)}{\sum_{region} \sum_{region} s(i, j)} = \\ &= \alpha_{0k} + \sum_{t=1}^T \left(\alpha_{jk} \cdot \frac{\sum_{region} \sum_{region} x_{NDVI}(t, i, j) \cdot s(i, j)}{\sum_{region} \sum_{region} s(i, j)} \right) = \alpha_{0k} + \sum_{t=1}^T \alpha_{kr}(t) \cdot \langle v(t) \rangle_{region}, \end{aligned} \quad (2)$$

где $\alpha_{kr}(t)$ – параметры модели для территориального региона r и сельскохозяйственной культуры k , определяемые посредством многопараметрической линейной регрессии; α_{0k} – свободный член модели; $s(i, j)$ – площадь пикселя изображения индекса состояния с координатами $s(i, j)$.

Информативными для прогнозирования урожайности считаются изображения, соответствующие месяцам, предшествующим началу сбора урожая. Имеется набор мультиспектральных космических изображений за годовой период. Также имеются статистические данные об урожайности. На основании имеющихся данных уточняются коэффициенты модели урожайности для различных сельскохозяйственных культур заданных областей.

На следующем этапе осуществляется применение полученных данных для участков внутри административной области. Для этого значения временного ряда показаний индекса усредняются для заданного участка и подставляются в формулу:

$$\langle y_{kr} \rangle_{area} = \alpha_{0k} + \sum_{j=1}^n \alpha_{jk} \cdot \langle v(t) \rangle_{area}.$$

Если нужно найти абсолютное значение собранного продукта для выбранного хозяйства, то, конечно, требуется использование карты посевов (Толпин и др., 2007). Однако необходимо подчеркнуть, что предлагаемые ниже модели не требуют построения карты посевных площадей, так как в основной концепции моделирования используются условия произрастания культур, которые, определяются по истории состояния растительности в целом по области, а не для отдельных культур. Наличие карты посевных площадей, вероятно, увеличило бы точность предсказания. Однако составление подробных карт – это трудоемкое и дорогостоящее мероприятие, требующее слаженного взаимодействия с агрохозяйствами исследуемой области. На сегодняшний момент не имеется общедоступной информации о территориях, занятых посевами, и картах посевов для отдельных хозяйств различных областей РФ. Из-за ограничения возможности оперативного получения информации имеющиеся карты посевов становятся неактуальными и непригодными для использования при краткосрочном прогнозировании.

Альтернативный подход, основанный на использовании существующих методов дистанционного зондирования для построения карты посевных площадей, может привести к обратному эффекту (ухудшению точности прогнозирования) из-за погрешности при определении посевной площади, как отдельных полей, так и для области в целом (Терехов, 2009). Этот вопрос не рассматривается в данной статье, поскольку является темой отдельного исследования.

Ниже приводится описание этапов развития регрессионной модели и иллюстрируется последовательное улучшение качества прогнозирования для всей совокупности культур.

Модель 1: глобальная линейная модель

Предположим, что почвенно-климатические характеристики мало изменяются для рассматриваемых областей РФ. При этом модель плодородия упрощается и становится одинаковой для рассматриваемых областей. Она будет выглядеть как:

$$y_{rk} = \alpha_k + \sum_{t=1}^T \alpha_k(t) \cdot \langle v(t) \rangle_r,$$

где k – индекс, указывающий на сельскохозяйственную культуру; r – индекс, указывающий на область (region) РФ; y_{rk} – оценка урожайности для данной области (r) и культуры (k); $\langle v(t) \rangle_r$ – усредненное значение индекса состояния растительности для отдельно взятой области РФ; α_k – свободный член модели; $\alpha_k(t)$ – настраиваемые параметры модели для отдельных интервалов времени года.

Параметры модели настраиваются на основе исторических данных об урожайности и соответствующих исторических значениях индекса.

Подстройка параметров модели осуществляется с помощью линейной регрессии. Анализ и сопоставление модели будут описаны в заключительной части этой статьи.

Модель 2: линейная модель для отдельных областей

В том случае, если предположение о минимальных различиях между рассматриваемыми областями неверно, необходимо возвратиться к более полной модели, описанной формулой (2), а именно:

$$y_{rk} = \alpha_{0k} + \sum_{t=1}^T \alpha_{rk}(t) \cdot \langle v(t) \rangle_r,$$

где $\alpha_{rk}(t)$ – параметры модели, которые теперь меняются не только от времени года (t), от культуры к культуре (k), но и от области к области (r). Изменения параметров в зависимости от области должны учитывать не только изменения в составе почвы, а и средние климатические условия в этом регионе. Остальные переменные этой формулы уже были описаны ранее.

Достоинство этой модели заключается в том, что выбранная формула позволяет тоньше настраивать параметры модели под особенности конкретной области.

Среди недостатков этой модели следует отметить тот факт, что настройка параметров требует гораздо больше исходных данных, доступных для обучения.

Модель 3: модель с мультипликативной поправкой для областей

В том случае, если количество данных, доступных для настройки индивидуальных моделей для каждой из областей, недостаточно, необходимо сократить количество настраиваемых параметров модели. Так, в частности, можно допустить, что основной вклад в различия по урожайности вносят следующие факторы:

- плодородие почв области;
- климатические различия между областями;
- количество солнечного излучения, зависящее от широты региона.

При этом для построения модели мы заведомо пренебрегаем временным смещением вегетативного сезона для различных областей западной части РФ.

Используя вышеизложенные допущения, получаем следующую формулу:

$$y_{rk} = C_{rk} \cdot \left(\alpha_k + \sum_{t=1}^T \alpha_k(t) \cdot \langle v(t) \rangle_r \right),$$

где k – индекс, указывающий на сельскохозяйственную культуру; r – индекс, указывающий на область (region) РФ; y_{rk} – оценка урожайности для данной области (r) и культуры (k); C_{rk} – коэффициент продуктивности области для конкретной сельскохозяйственной культуры; $\langle v(t) \rangle$ – усредненное значение индекса состояния растительности для отдельно взятой области РФ; α_k – свободный член модели; $\alpha_k(t)$ – настраиваемые параметры модели для отдельных интервалов времени года.

В данной статье эта модель была выбрана за базовую. Дальнейшие изменения связаны с более тонкими поправками.

Модель 4: трендовая модель с мультипликативной поправкой для областей

На протяжении нескольких последних десятилетий во всем мире наблюдается устойчивый и стабильный рост урожайности в расчете на единицу возделываемой площади (Fischer et al., 2009). Это связано с несколькими факторами. Прежде всего, стоит отметить прогресс генной инженерии в выведении усовершенствованных видов сельхоз культур. Прилученные семена более устойчивы к засухам, температурным перепадам и паразитам. Следующим фактором является более эффективное применение удобрений. Прогресс в области сельхозтехники позволил собирать выращенный урожай с меньшими потерями. Также усовершенствованные методы химической обработки позволяют лучше контролировать популяции вредителей.

Подобное увлечение урожайности называется трендовым. Целесообразно учитывать этот тренд в модели урожайности, так как велика вероятность того, что подобный тренд сохранится в ближайшие несколько лет.

Сделав допущение, что изменения урожайности линейно зависят от времени на рассматриваемом историческом промежутке, становится возможным записать формулу для прогнозирования долгосрочного увеличения урожайности. Так, средняя урожайность за текущий год может быть выражена из урожайности за начальный год посредством следующей формулы:

$$\frac{\langle y_{current} \rangle - \langle y_{start} \rangle}{\langle y_{start} \rangle} = \beta \cdot (Y_{current} - Y_{start}),$$

где $\langle y_{current} \rangle$ – средняя урожайность за текущий год $Y_{current}$; $\langle y_{start} \rangle$ – средняя урожайность в год начала наблюдений Y_{start} ; β – относительный годичный прирост урожайности.

Выразим $\langle y_{current} \rangle$ через остальные переменные:

$$\langle y_{current} \rangle = [1 + \beta \cdot (Y_{current} - Y_{start})] \cdot \langle y_{start} \rangle.$$

Получаем следующую уточненную формулу для модели урожайности:

$$y_{rk} = [1 + \beta \cdot (Y - Y_{start})] \cdot C_{rk} \cdot \left(\alpha_k + \sum_{t=1}^T \alpha_k(t) \cdot \langle v(t) \rangle_r \right), \quad (3)$$

где k – индекс, указывающий на сельскохозяйственную культуру; r – индекс, указывающий на область (region) РФ; y_{rk} – оценка урожайности для данной области (r) и культуры (k); Y – текущий год для которого производится оценка урожайности; Y_{start} – год начала наблюдений; β – относительный годичный прирост урожайности; C_{rk} – коэффициент продуктивности области для конкретной сельскохозяйственной культуры; $\langle v(t) \rangle_r$ – усредненное значение индекса состояния растительности для отдельно взятой области РФ; α_k – свободный член модели; $\alpha_k(t)$ – настраиваемые параметры модели для отдельных интервалов времени года.

Настройка моделей и прогнозирование

Общая схема настройки метода прогнозирования урожайности по многоспектральным изображениям с помощью многомерной линейной регрессии представлена на *рис. 1*.

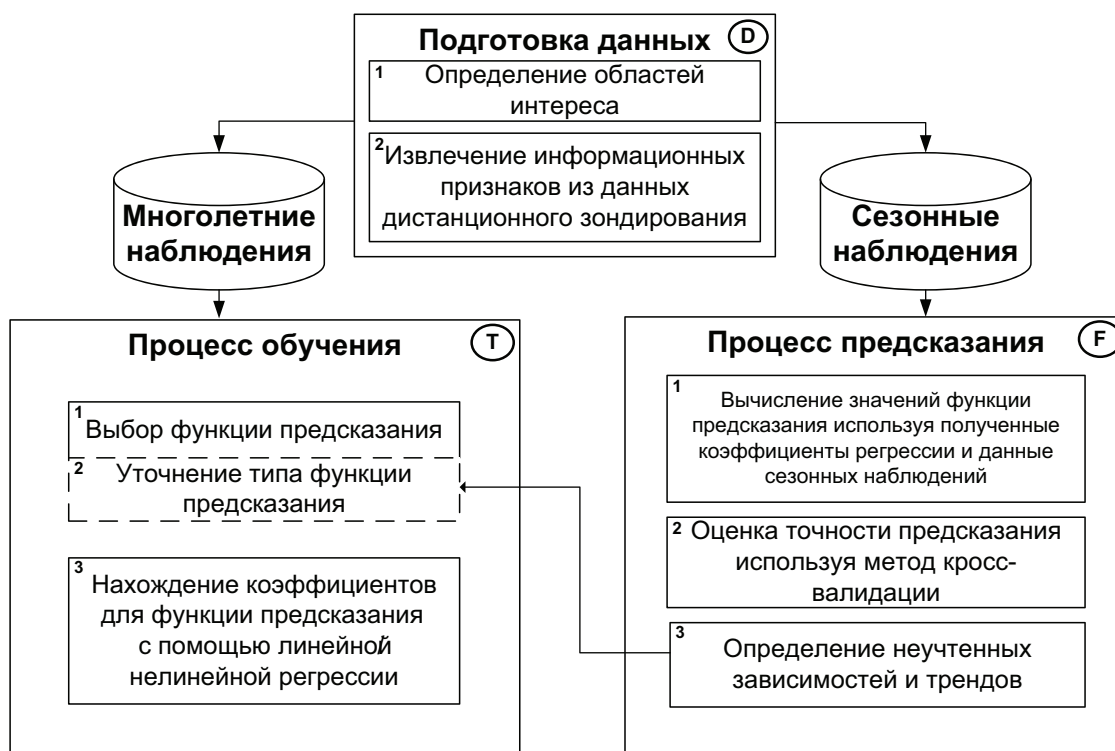


Рис. 1. Метод прогнозирования урожайности по многоспектральным космическим изображениям с помощью многомерной регрессии

Подстройка параметров выбранных моделей производилась с использованием композитных индексов NDVI за период с 2000 по 2009 г. Использовались шестнадцатидневные композитные изображения с разрешением 250 м со сканера MODIS за период в 10 лет (ftp://e4ftl01.cr.usgs.gov/MODIS_Composites).

Данные об урожайности за указанный промежуток времени были получены с интернет-сайта Федеральной службы государственной статистики. Настройка производилась для трех групп сельхозкультур: пшеница, картофель, овощи.

В качестве информативных использованы изображения, полученные за период времени пять месяцев, которые предшествуют началу сбора урожая (март–июль). Все множество исходных данных было разбито на два: обучающие и тестовое. Ввиду малого количества исходных данных в качестве проверочного множества были выбраны данные по урожайности для всех 14 рассмотренных областей для трех культур за 2009 г. Данные за остальные годы использовались для настройки модели.

Необходимость учитывать тренд может быть продемонстрирована на *рис. 2а, б, в*, где показаны погрешности прогнозирования в зависимости от года для данных участвовавших в обучении модели 3. Как видно из рисунков, для всех трех культур присутствует тренд или отклонение от реальных данных в зависимости от года. Наличие подобного тренда может быть связано с постепенным улучшением продуктивности, которое не зависит от общего состояния растительности. Эти графики как раз и послужили основой для введения тренда и использования модели 4. Стоит отметить, что подобная регулярная ошибка прогноза присутствует не только в модели 3, но также в моделях 1 и 2 для всех рассмотренных культур.

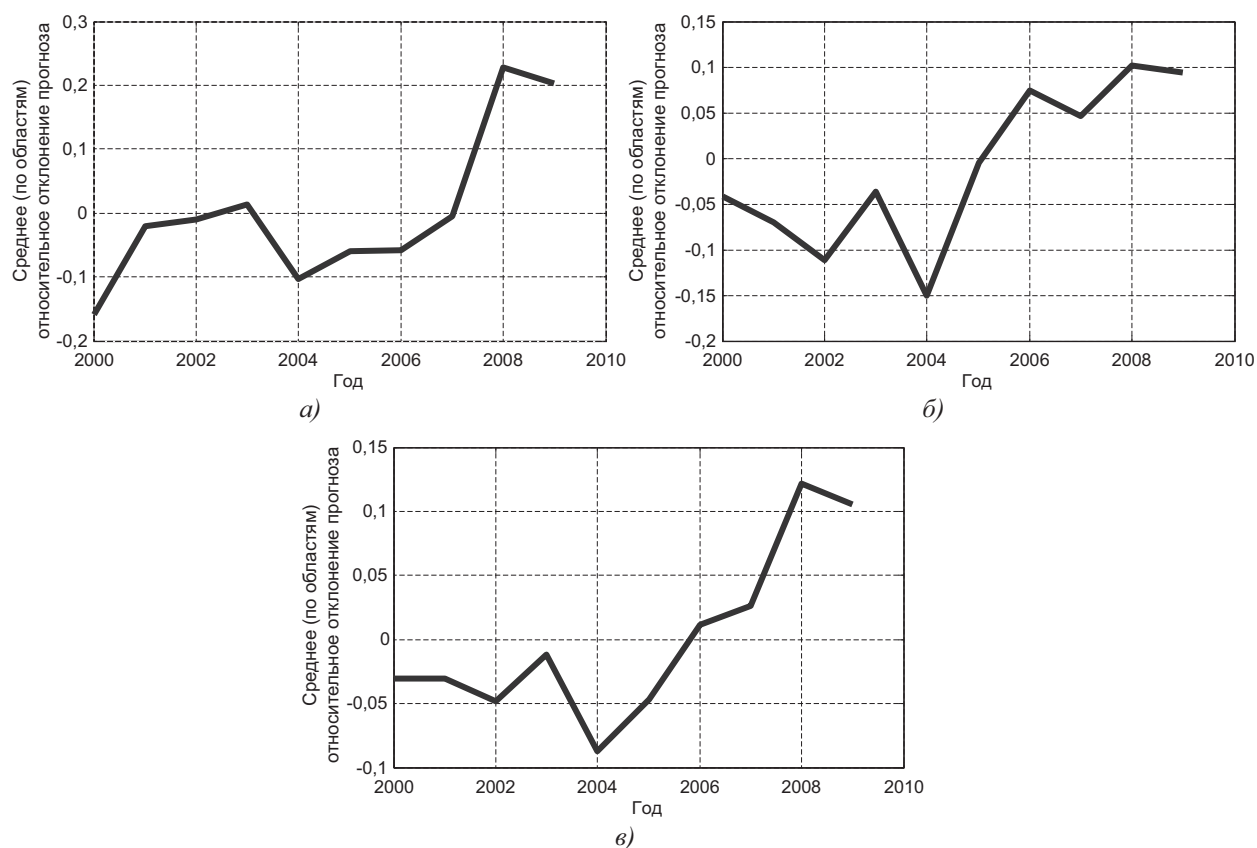


Рис. 2. Демонстрация наличия регулярной погрешности (тренда) в прогнозе с помощью модели 3 для различных культур:

(а) пшеница, (б) картофель, (в) овощи

Даже несмотря на то, что данные за 2010 и 2011 гг. показывают значительное отклонение от общей тенденции увеличения урожайности, ввиду количественного преобладания данных за 2000–2009 гг., подтверждающих наличие тренда, тренд существует для всех трех культур и, значит, улучшает точность предсказания. В свою очередь, 2010 г. является аномальным из-за засухи, поэтому рассмотрение задачи прогнозирования для аномальных лет также выходит за пределы проводимого исследования и является темой отдельной статьи.

Ниже приведены результаты настройки моделей и прогнозирования урожайности для трех групп культур. На *рис. 3* показаны ошибки прогнозирования для рассмотренных моделей. Как видно из рисунка, модель 4 (трендовая модель с мультипликативной поправкой для областей) показывает наилучшие результаты из всех рассмотренных.

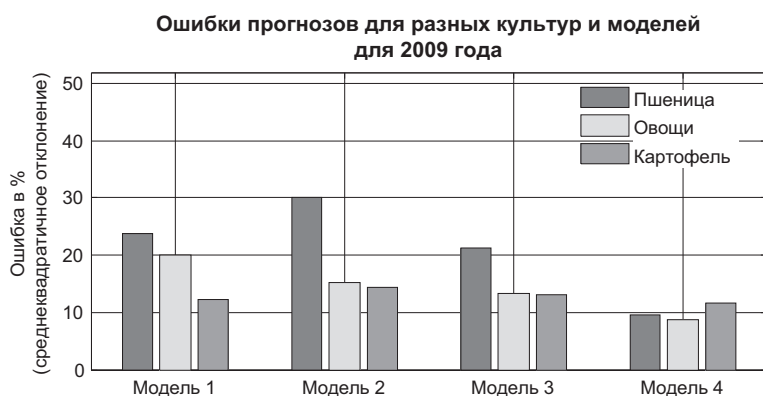


Рис. 3. Среднеквадратичное отклонение для всех рассмотренных культур и всех моделей

Определим статистическую значимость модели 4. Для этого вычислим коэффициенты детерминации (нескорректированный и скорректированный). Далее проверим значимость модели 4 при прогнозировании урожайности различных типов культур на основе скорректированного (истинного) коэффициента детерминации.

Для вычисления нескорректированного коэффициента детерминации R^2 использовалась формула:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2},$$

где y_i – наблюдаемые значения урожайности; f_i – значения урожайности, предсказанные моделью f ; \bar{y} – усредненные наблюдаемые значения урожайности.

В свою очередь скорректированный (истинный) коэффициент детерминации был вычислен с использованием следующей формулы:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{(n - 1)}{(n - k)},$$

где n – количество наблюдений участвовавших в обучении модели; k – количество степеней свободы модели (количество независимых параметров).

Для проверки статистической значимости модели используется критерий Фишера. Выдвигается гипотеза о равенстве R_{adj}^2 нулю. Статистика, связанная с коэффициентом детерминации, вычисляется по формуле:

$$F(k-1, n-k) = \frac{R_{adj}^2 (n-1)}{(1-R_{adj}^2)(n-k)},$$

где n – количество наблюдений участвовавших в обучении модели; k – количество степеней свободы модели (количество независимых параметров); R_{adj}^2 – скорректированный (истинный) коэффициент детерминации.

Полученное значение статистики сравнивается с табличным значением, которое является максимальным значением критерия под влиянием случайных факторов при текущих степенях свободы и уровне значимости для распределения Фишера.

В *таблице* приведены рассчитанные значения коэффициентов детерминации, которые характеризуют соответствие модели реальным данным для результатов на 2009 г. для каждой из рассматриваемых культур. Количество независимых наблюдений участвовавших в обучении во всех случаях равно произведению количества областей (14) на количество лет (9) взятых для исследования и составляет 126.

Таблица. Сравнение статистической значимости модели 4 для различных культур для результатов на 2009 г.

<i>Показатель</i>	<i>Пшеница</i>	<i>Овощи</i>	<i>Картофель</i>
Нескорректированный коэффициент детерминации R^2	0,65	0,88	0,85
Количество наблюдений n	126	126	126
Количество свободных параметров k	21		
Скорректированный (истинный) коэффициент детерминации R_{adj}^2	0,58	0,85	0,81
Значение статистики F	7,3	30,3	23,1
Табличное значение F – критерия на уровне значимости $\alpha = 0,05$	1,67		

Вычисление коэффициента детерминации проводилось с учетом числа параметров приведенной модели (Draper, Smith, 1998). Приближение коэффициента детерминации к 1 говорит о лучшем соответствии модели данным. Приведенные в *таблице* результаты позволяют заключить, что модель показывает лучшее соответствие данным для прогнозирования урожайности овощей и картофеля, чем для пшеницы. Это может быть следствием того, что значение урожайности для пшеницы за все года включает урожайности как яровой, так и озимой пшеницы; причем сбор урожая яровой и озимой пшеницы проводится в разное время.

Рассчитанные значения F – критерия для всех трех культур превышают табличное значение на заданном уровне значимости $\alpha = 0,05$, что подтверждает значимость построенной модели (отвергает выдвинутую гипотезу о ее незначимости).

На данный момент лучшая из представленных моделей позволяет осуществлять прогнозирование урожайности с 10%-ной погрешностью, что меньше стандартного отклонения показателя урожайности зерновых 17%.

На рис. 4а, б, в показаны примеры прогнозов и официальные данные статистики для рассмотренных регионов за 2009 г.

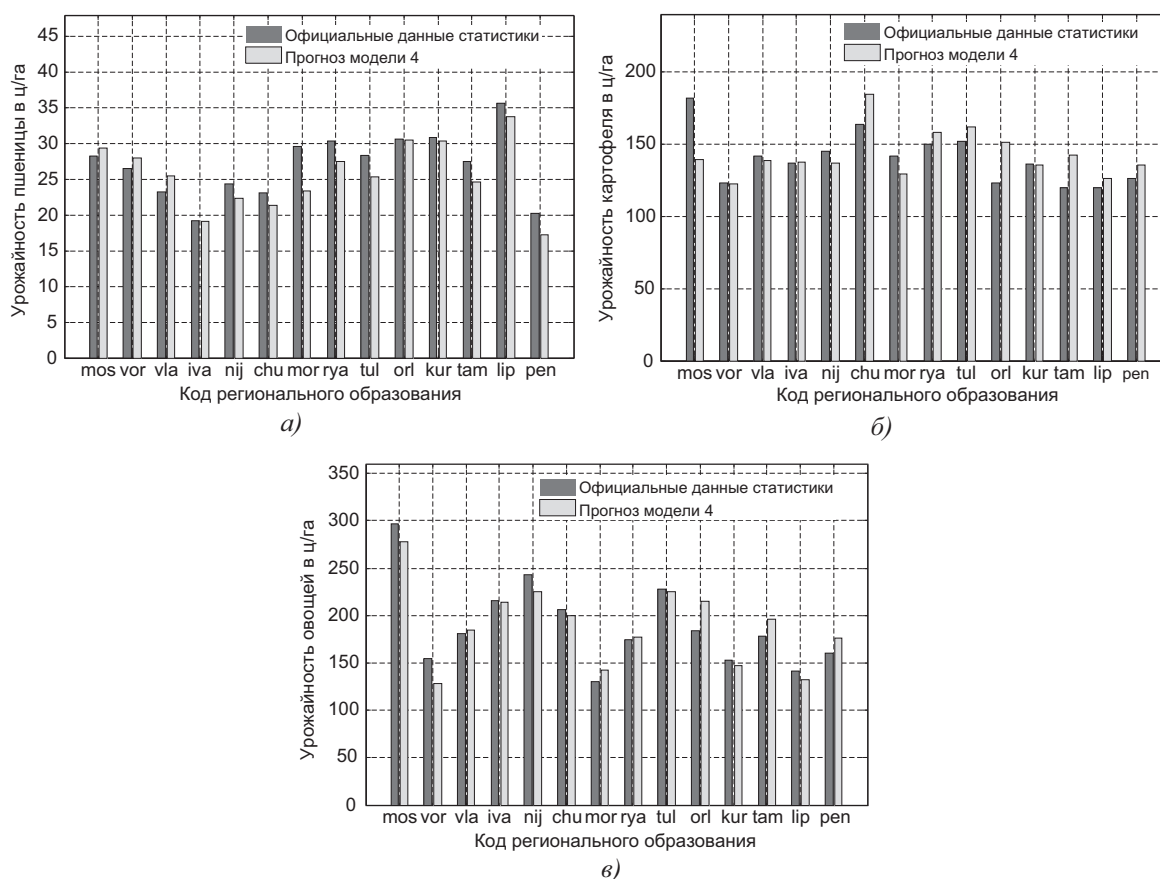


Рис. 4. Результаты прогноза и официальные данные статистики за 2009 год для различных областей страны для различных культур: (а) пшеница, (б) картофель, (в) овощи

Проинтерпретируем с точки зрения физического смысла коэффициенты модели 4 в формуле (3). Они разделяются на три подгруппы.

В первую подгруппу входит единственный коэффициент β , который описывает наличие тренда в модели.

Во вторую группу входят 14 коэффициентов C_{rk} , которые отвечают продуктивности каждой из областей соответственно. Стоит отметить, что эти коэффициенты носят относительный характер, поскольку показывают изменение продуктивности относительно значения продуктивности для Московской области, который принят за единицу. Таким образом, данная модель может быть использована для оценки относительной продуктивности областей.

Третья группа коэффициентов включает набор параметров $\alpha_k(t)$, определяющих вид зависимости между значениями вегетационного индекса и урожайностью. Коэффициенты этой группы имеют размерность центнеров/га, отнесенную к значению вегетационного индекса. В нашем приближении из упрощенного ряда Тейлора (формула (2)), следует, что представленные модели имеют накопительный характер, т.е. среднее значение вегетационного индекса во времени аддитивно воздействуют на величину урожайности.

В завершение проведем анализ коэффициентов регрессии. Среди интересных особенностей можно привести небольшой разброс коэффициентов плодородия областей C_{rk} и наличие временного тренда. На примере пшеницы видно, что общий тренд находится в районе 0,074, т.е. в общем урожайность увеличивается в среднем в 1,74 раза за 10 лет. Коэффициент плодородия находится в диапазоне: от 1,02 для самой плодородной области (Липецкой) до 0,59 для самой неплодородной области (Пензенской), что соответствует реальным показателям урожайности для этих областей. Полученные значения для коэффициентов подтверждают предположение о том, что различие в урожайности для различных областей может быть обусловлено конкретными условиями в каждой из областей (климатические условия, плодородие почвы и т.д.).

Заключение

Результаты проведенных вычислительных экспериментов продемонстрировали возможность прогнозирования урожайности по многоспектральным космическим изображениям с использованием реальных статистических данных по урожайности для различных сельскохозяйственных культур и разных областей страны.

Показано, что предложенный метод позволяет прогнозировать урожайность пшеницы, картофеля и овощей с приемлемой точностью.

Следует отметить, что осуществляемый прогноз урожайности является краткосрочным. Процесс сбора урожая занимает в среднем около месяца. При этом дата начала сбора может варьироваться в зависимости от территории, культуры и погодных условий. Возможность получения адекватного прогноза урожайности с заблаговременностью один месяц может быть использована при решении ряда практических задач, например при планировании закупок сельскохозяйственной продукции в случае малоурожайных лет.

В последующих исследованиях предполагается модификация построенных моделей с целью улучшения точности и общности при прогнозировании урожайности для большего количества областей.

Работа проводилась при финансовой поддержке Министерства образования и науки РФ (ГК 14.515.11.0003, 14.В35.21.0652) и РФФИ (проекты № 13-05-12019\13, №12-01-91162).

Литература

1. Терехов А.Г. Оценка точности спутниковой технологии определения площади сельскохозяйственных масок в схеме с использованием снимков различного разрешения / Седьмая Всероссийская Открытая конференция «Современные проблемы дистанционного зондирования Земли из космоса». Москва, Россия, Ноябрь 2009. С. 459–465.
2. Толтин В.А. и др. Оценка состояния сельскохозяйственных культур на основе межгодовой динамики с использованием данных MODIS // Современные проблемы дистанционного зондирования Земли из космоса. 2007. Т. 4. № 2. С. 380–389.

3. *Bondur V.* Aerospace methods and technologies for monitoring of oil and gas areas and facilities // *Izvestiya. Atmospheric and Oceanic Physics.* 2011. Vol. 47. No. 9.
4. *Draper N., Smith H.* Applied Regression Analysis. Wiley-Interscience, 1998. ISBN 0-471-17082-8.
5. *Fischer R., Byerlee E., Edmeades E.* Can technology deliver on the yield challenge to 2050 // Expert Meeting on How to Feed the World (Food and Agriculture Organization of the United Nations), June 2009.
6. *Murynin A. et al.* Analysis of large long-term remote sensing image sequence for agricultural yield forecasting. // *Image Mining. Theory and Applications. Proceedings of the 4th International Workshop on Image Mining.* Barcelona, Spain. February 2013. P. 48–55.

Yield forecasting on the basis of long-term space-based observations of the dynamics of vegetation

A.B. Murynin^{1,2}, V.G. Bondur¹, V.Y. Ignatiev^{1,3}, K.Y. Gorokhovskiy¹

¹ *Institute for Scientific Research of Aerospace Monitoring “AEROCOSMOS”, Moscow, Russia*

E-mail: office@aerocosmos.info

² *Dorodnicyn Computing Centre of RAS, Moscow, Russia*

E-mail: amurynin@bk.ru

³ *Moscow Institute of Physics and Technology, Moscow Region, Russia*

E-mail: vladimir.ignatiev.mipt@gmail.com

The method of yield forecasting on the basis of multispectral satellite images is proposed in this article. The results of computer simulation to predict crop yields for different crops of different areas are presented and analyzed. The study is conducted to determine the most accurate model of soil fertility, which results in a model that takes values of vegetation indices for 5 months, starting in March and ending in July months of this year. Based on the analysis concludes that the accuracy of forecasting and ongoing possibility of extending the scope of the resulting model for predicting crop yields.

Keywords: yield forecasting, multiple linear regression, multiple nonlinear regression, vegetation indices.