

Метод выявления используемых пахотных земель по данным дистанционного зондирования со спутников

С.А. Баргалева, Е.А. Лупян, И.А. Нейштадт

Институт космических исследований РАН

117997 Москва, ул. Профсоюзная, 84/32

E-mail: beml@smis.iki.rssi.ru

В работе предложен метод автоматизированного детектирования используемых пахотных земель, основанный на анализе многолетних временных рядов данных спутниковых наблюдений с учетом априорных знаний о различиях в закономерностях сезонного развития сельскохозяйственной и естественной растительности. Разработанный алгоритм детектирования пахотных земель, обладающий свойством автоматической адаптивности к специфике наблюдаемых регионов, позволил с использованием данных прибора MODIS, установленного на спутнике TERRA, осуществить выявление пахотных земель в основных сельскохозяйственных регионах России. В статье приводятся результаты сравнения полученных на основе разработанного метода данных о пахотных землях с материалами государственной статистики по посевным площадям.

Введение

Получение объективной информации о сельскохозяйственном производстве является важным условием проведения эффективной аграрной политики во многих странах мира, а в качестве одного из необходимых источников данных для получения такой информации все большее значение приобретают спутниковые системы дистанционного зондирования. Исследовательские работы, направленные на создание методов обработки спутниковых данных для получения информации о состоянии сельскохозяйственных земель, в настоящее время ведутся в рамках ряда международных и российских программ. В России в последние годы эти работы активно проводятся в рамках программы информатизации сельского хозяйства, которая должна, в частности, обеспечить получение объективной оперативной информации о состоянии и использовании сельскохозяйственных земель. Для реализации этой задачи Главный вычислительный центр МСХ РФ и Институт космических исследований РАН (ИКИ РАН) ведут совместную разработку системы спутникового мониторинга сельскохозяйственных земель России [1], призванной обеспечить оперативный сбор данных о площадях пахотных земель и посевах различных типов, их состоянии и продуктивности.

Одной из важных задач разрабатываемой системы является выявление и оценка по спутниковым данным площадей используемых пахотных земель. Актуальность этой задачи, в частности, обусловлена значительным сокращением и изменением структуры пахотных земель России в течение последних десятилетий в силу влияния причин социально-экономического характера [2, 3]. Так с 1970 года посевная площадь в России уменьшилась на 39% при значительном росте за этот же период доли зерновых и технических культур (рис. 1). Существующая традиционная система сбора информации о состоянии сельскохозяйственных земель на уровне страны или её регионов, опирающаяся на данные комитетов государственной статистики не может быть признана полностью объективной и отвечающей современным требованиям по уровню полноты, достоверности и оперативности получаемых данных.

Пахотными называют регулярно обрабатываемые с целью получения продукции растениеводства земли [4]. При этом система государственного учета земель относит к категории пашни все предназначенные для пахоты сельскохозяйственные угодья, которые, формально оставаясь в данной категории, в силу ряда причин фактически могут не обрабатываться и не использоваться для растениеводства. Это порождает необходимость разработки объективных методов выявления территорий

используемых пахотных земель (ИПЗ). При этом одним из условий создания эффективной системы оперативного спутникового мониторинга сельскохозяйственной растительности является наличие алгоритмов и технологий обработки спутниковых данных, не предполагающих участия экспертов или минимизирующих необходимость их участия. При разработке таких алгоритмов также важно минимизировать объем необходимой опорной информации, получение которой часто сопряжено с проведением дорогостоящих и трудоемких наземных работ.

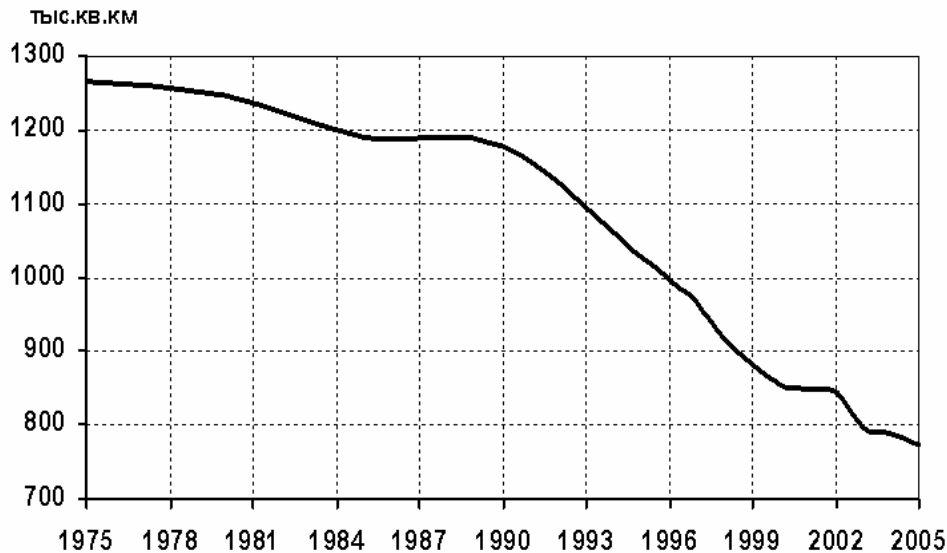


Рис.1. Динамика посевной площади в России за период 1975-2005 годов. Источник: Росстат, МСХ

Следует отметить, что хотя пахотные земли в периоды свободные от посевов визуально легко распознаются по спутниковым изображениям, построение эффективного автоматического алгоритма их картографирования на больших территориях сопряжено с необходимостью учета множества факторов. К факторам такого рода относятся региональные различия в динамике развития сельскохозяйственных культур, яркости почв, различия в метеорологических условиях в различные годы.

При этом часто алгоритмы классификации сельскохозяйственных земель по спутниковым данным предполагают выбор фиксированных пороговых значений экспертным путем, что, как правило, обеспечивает получение хороших результатов на уровне отдельных регионов, но приводит к снижению точности использования алгоритмов на больших территориях. Так описанная в работе [1] методика выявления используемых пахотных земель, опирающаяся на детектирование различных с.-х. культур, продемонстрировала возможность получения достаточно точных результатов на территории Ростовской области и Краснодарского края. Однако распространение данного метода на другие территории представляет значительную трудность ввиду необходимости учета агроклиматических региональных особенностей.

В настоящей работе представлен подход к построению автоматически адаптивных к региональным условиям алгоритмов, допускающих их применение на больших территориях без необходимости предварительной настройки параметров. Использование указанного подхода позволило создать алгоритм эффективного выделения ИПЗ в сельскохозяйственных регионах России на основе данных прибора MODIS.

Алгоритм детектирования используемых пахотных земель

Базовые принципы построения алгоритма. В качестве одного из необходимых условий при разработке алгоритма являлось обеспечение его автоматической адаптивности к различным региональным условиям. Обычно схема построения алгоритмов рассчитанных на непосредственное участие оператора в процессе обработки спутниковых данных предполагает осуществление следующих основных этапов:

- выбор множества классов распознавания;
- выбор набора данных и процедуры классификации;
- построение обучающих выборок (осуществляется оператором с учетом знания специфики региона и использованием наземных данных);
- обучение классификатора;
- классификация спутниковых данных.

Отметим, что два первых этапа в данной последовательности шагов выполняются один раз, а два последних из них, как правило, реализуются в виде автоматических процедур. Поэтому при создании автоматических алгоритмов одна из основных трудностей состоит именно в разработке процедуры построения обучающих выборок.

В решаемой нами задаче множество классов включает в себя лишь две категории земель, подлежащих автоматическому распознаванию, а именно ИПЗ и все остальные территории.

Выбор набора используемых спутниковых данных, а также разработка алгоритма классификации, включая построение обучающих выборок, основывалась на учете особенностей, отличающих характер внутрисезонной и межсезонной динамики зеленой биомассы на используемых пахотных землях от территорий, занятых естественной растительностью или длительное время не обрабатываемыми пахотными землями. К такого рода особенностям можно отнести следующие отличительные черты используемых пахотных земель:

- наличие существенных различий в продолжительности вегетационного периода, характерного для естественной растительности и некоторых сельскохозяйственных культур в силу проведения агротехнических мероприятий;
- наличие существенных различий межгодовой динамики зеленой биомассы вследствие ротации видов сельскохозяйственных культур.

Первая особенность может быть положена в основу выбора метода классификации и набора необходимых данных. Собственно отделение ИПЗ в конкретном регионе от других земель может быть произведено на основе порогового алгоритма, разделяющего территории по продолжительности вегетационного периода растительности. Необходимая в этом случае региональная адаптация значений порога в алгоритме может быть произведена на основе обучающей выборки.

Оценка продолжительности вегетационного периода может быть получена на основе анализа внутрисезонной динамики зеленой биомассы, характер поведения которой может быть описан, например, с использованием временных серий данных спектральных вегетационных индексов [5]. Наблюдение за динамикой значений вегетационного индекса для каждого пиксела позволяет оценивать даты наступления различных фенологических фаз. Так, например, считается, что максимум вегетационного индекса соответствует фазе цветения растений. Анализируя динамику вегетационного индекса для данного пиксела можно также оценивать продолжительность вегетационного периода. В силу проведения агротехнических мероприятий вегетационный период для сельскохозяйственных культур, как правило, короче аналогичного показателя для естественной растительности, произрастающей в этом же регионе. Указанные выше

различия в динамике вегетационного индекса для сельскохозяйственной и естественной растительности могут быть положены в основу детектирования ИПЗ.

Вторая из приведенных выше особенностей ИПЗ может быть положена в основу автоматического алгоритма построения обучающих выборок для конкретных регионов. На основе анализа меры сходства межгодовой динамики зеленой биомассы могут быть выбраны участки с высокой вероятностью соответствующие пахотным и не обрабатываемым землям. Поскольку при этом производится анализ изменчивости растительного покрова для каждого пиксела, это позволяет построить алгоритм, не зависящий от регионов наблюдения. Естественно, что для построения такого алгоритма выявления ИПЗ требуется, использование многолетних рядов спутниковых наблюдений. При этом можно предположить, что на некотором временном интервале, сопоставимом с продолжительностью севооборота в данном регионе, увеличение продолжительности наблюдений будет приводить к повышению точности результатов детектирования ИПЗ.

Детальное описание алгоритма выделения ИПЗ по спутниковым данным приводится ниже.

Входные данные. Созданный алгоритм рассчитан на использование данных наблюдений прибором MODIS со спутника Тетта. Прибор MODIS выполняет съемку в 36 спектральных каналах видимого и инфракрасного диапазона длин волн с пространственным разрешением 250м, 500м и 1км [6, 7] и обеспечивает ежедневное получение данных наблюдений для любого региона России. Получаемые данные спутниковых наблюдений свободно распространяются Геологической службой США (<http://lpdaac.usgs.gov/main.asp>). Разработанная в ИКИ РАН система [8] обеспечила автоматизированное получение данных и формирование на их основе композитные изображения, очищенные от влияния облаков, теней и снега [9]. Используемые для выявления обрабатываемых пахотных земель композитные изображения формировались по данным восьмидневных наблюдений и содержат значения коэффициента спектральной яркости (КСЯ) в красном (620-670 нм) и ближнем инфракрасном (841-876 нм) каналах с пространственным разрешением 250 м.

В качестве входных данных для алгоритма выявления ИПЗ используются многолетние временные ряды значений перпендикулярного вегетационного индекса PVI [5], инвариантного по отношению к отражательным свойствам почвенного покрова и обладающего высокой чувствительностью к изменениям объема зеленой биомассы растительности. Индекс PVI рассчитывается как расстояние от заданной точки до линии почв, измеряемое в двумерном пространстве значений КСЯ в красном и ближнем инфракрасном диапазонах спектра. Выражение для линии почв $R2 = 1.47 * R1 + 0.01$, где $R1$ и $R2$ – значения КСЯ в красном и ближнем инфракрасном каналах MODIS соответственно, было получено экспериментально по данным измерений спектрального отражения открытой почвы различного уровня яркости. Таким образом, используемое в данной работе выражение для расчета значений индекса PVI имеет следующий вид $PVI = -0.83 * R1 + 0.56 * R2 - 0.005$. Примеры кривых динамики индекса PVI для сельскохозяйственных культур и естественной растительности показаны на рис. 2.

Маскирование заведомо несельскохозяйственных земель. Одним из необходимых шагов реализации метода является построение маски территорий заведомо не являющихся ИПЗ. При построении такой маски могут использоваться различные вспомогательные картографические данные, а также результаты автоматизированной классификации спутниковых изображений MODIS. Обозначим множество пикселей входящих в эту маску M .

Построение признака классификации. В качестве признака классификации разработанный алгоритм использует длину вегетационного периода, которая определяется следующим образом. Рассмотрим ряд измерений pvi_i , $i = 1..N$ за один произвольный год для данного пиксела. Пусть максимальное значение этого ряда равно pvi_{max}^i и достигается при $i = i_{max}$, тогда обозначим $i = i_0$ максимальное i удовлетворяющее условиям $pvi_i < pvi_{max}^i / 2$ и $i < i_{max}$. Аналогично обозначим $i = i_1$ минимальное i такое,

что $pvi_i < pvi_{\max} / 2$ и $i > i_{\max}$. Обозначим $j = 1..J$ – номер рассматриваемого года. Величину $d_j = i_1 - i_0$, рассчитываемую для каждого из годов наблюдений, будем называть продолжительностью вегетационного периода (ВП). Следует отметить, что предложенная величина не соответствует используемому в биологии понятию вегетационного периода, однако может рассматриваться в качестве некоторой коррелирующей с ним относительной оценки. Как уже упоминалось ранее, для участков естественной растительности d_j будет иметь относительно большие и в то же время относительно мало изменяющиеся между различными годами значения. Для ИПЗ величина d_j будет иметь относительно небольшие значения и будет существенно меняться от года к году в зависимости от типа посевов. Введем минимальное за многолетний период наблюдения значение ВП $d_{\min} = \min_{j=1..J} d_j$ и примем эту величину в качестве признака классификации ИПЗ. Следует отметить, что хотя для ИПЗ d_{\min} меньше, чем для естественной растительности, указанная закономерность выполняется лишь при локальном характере сравнения, так как на продолжительность ВП существенное влияние оказывают климатические условия региона. Так, при сравнении величин d_{\min} для ИПЗ и естественной растительности из значительно различающихся по климатическим условиям регионов данная закономерность может нарушаться. Именно это обстоятельство приводит к необходимости разработки локально адаптивного алгоритма классификации.

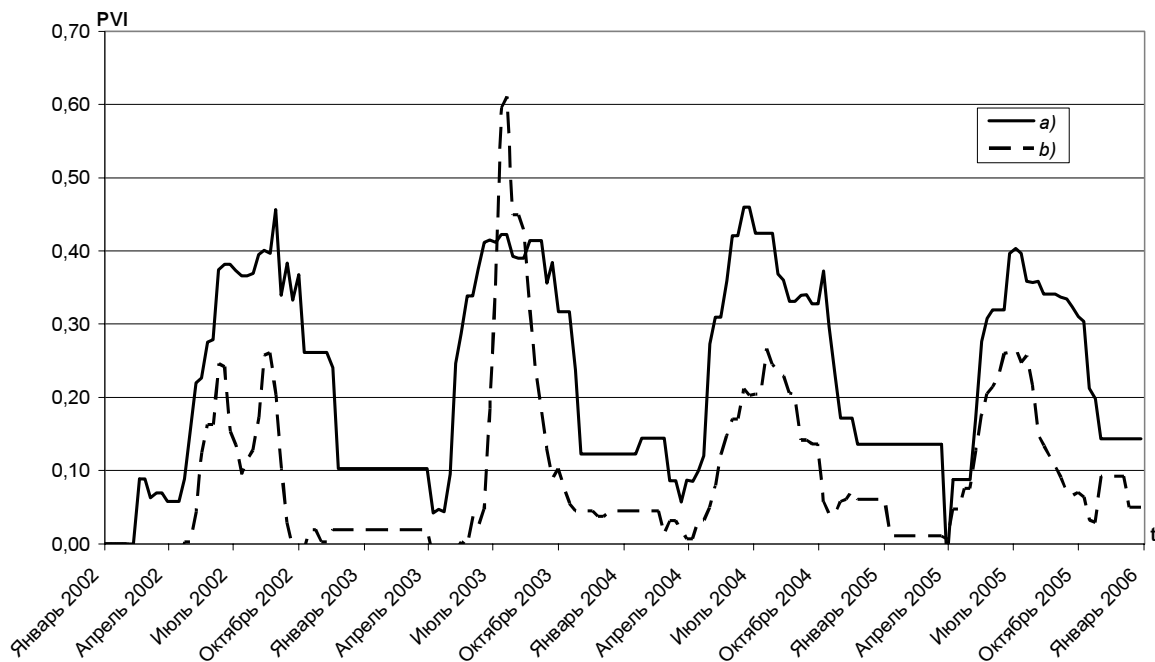


Рис.2. Примеры кривых вегетационного индекса PVI по данным MODIS для (a) естественной растительности и (b) сельскохозяйственных культур

Построение обучающих выборок. Целью второго этапа является автоматическое построение обучающих выборок для оценки значений признака классификации ИПЗ и естественной растительности. Обозначим множества пикселей этих выборок A и N соответственно. Указанные выборки могут строиться различным образом. Однако представляется эффективным использование значений коэффициента корреляции между годовыми рядами данных вегетационного индекса PVI. При этом предполагается, что вследствие ротации сельскохозяйственных культур на используемых пахотных землях, указанные участки будут характеризоваться более низкими, по сравнению с территориями, занятыми естественной растительностью, значениями коэффициента корреляции. Важно также отметить, что такой

подход в значительной степени адаптивен к изменяющимся в широком диапазоне климатическим условиям и, следовательно, обеспечивает высокий уровень региональной независимости разработанного алгоритма.

Для построения обучающих выборок для каждого пиксела в предлагаемом алгоритме рассчитываются коэффициенты корреляции r_i по всем возможным парным сочетаниям имеющихся годовых рядов значений. Существенные межгодовые различия динамики РVI для соответствующих ИПЗ пикселей, обуславливает в них относительно низкие величины коэффициентов корреляции. Согласно принятой методике, условием отнесения пиксела к множеству N было превышение медианы значений r_i пороговой величины 0,7. В случае же появления среди множества полученных значений r_i двух или более отрицательных величин и при этом пиксел не был отнесен ранее к маске M, ему приписывалась принадлежность к множеству A. С учетом анализа особенностей пространственной изменчивости условий в сельскохозяйственных регионах России, в качестве размера окна L было выбрано значение 100 км.

Адаптивная классификация. Рассмотрим цифровое изображение, каждому пикселу которого присвоена величина d_{\min} . Рассчитаем для каждого пиксела квадратного окна со стороной L км среднее значение и стандартное отклонение d_{\min} , используя при этом только значения d_{\min} пикселей из множества A и, отдельно, используя значения d_{\min} пикселей из множества N. Величина L должна быть достаточно большой чтобы на результат не оказывал влияние локальный ландшафт, и в тоже время достаточно малой, чтобы использовались данные только из близкой по климатическим условиям территории. Полученные величины обозначим соответственно $E_A, \sigma_A, E_N, \sigma_N$. Данный пиксел отнесем к классу ИПЗ, если для него выполнены условия $d_{\min} < E_A + \frac{E_N - E_A}{2(\sigma_A + \sigma_N)}$ и пиксел не принадлежит M.

Следует особо отметить, что предложенный регионально адаптивный алгоритм не предполагает использования опорной информации в качестве входных данных, а целесообразность привлечения материалов наземных наблюдений обусловлена лишь необходимостью верификации полученных результатов.

Результаты экспериментальной апробации разработанного алгоритма

Экспериментальная апробация разработанного алгоритма была проведена на территории, охватывающей основные сельскохозяйственные регионы России и включающей 56 субъектов РФ. Для проведения апробации использовались данные наблюдений MODIS за четыре последовательных года 2002-2005.

Для построения маски заведомо несельскохозяйственных земель привлекались дополнительные данные из следующих источников:

- цифровая модель рельефа GTOPO30 с пространственным разрешением 30'' (<http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>);
- топографическая основа DCW масштаба 1:1000000 (www.maproom.psu.edu/dcw/);
- результаты детектированных ночных огней по данным прибора DMSP (<http://www.ngdc.noaa.gov/dmsp/index.html>).

Полученная маска заведомо несельскохозяйственных земель включала в себя такие категории земного покрова, как леса, степи, болота, солончаки, открытые пески и другие лишенные растительности территории, водные объекты, крупные населенные пункты, участки овражно-балочной сети, а также территории, ограниченные высотой над уровнем моря (>1000м), крутизной склона (>5°) или удаленностью от ближайшего населенного пункта (>25км). Алгоритмы картографирования этих объектов [10] в настоящее время основываются на пороговых методах классификации и требуют на определенных этапах участия эксперта. Однако, будучи выполненным для заданного региона, этот этап обработки спутниковых

данных не требует регулярного повторения, а его результаты могут быть использованы достаточно продолжительное, измеряемое несколькими годами, время.

Использование разработанного алгоритма позволило по данным последовательных четырехлетних спутниковых наблюдений прибором Terra-MODIS выявить используемые пахотные земли и оценить их площадь в сельскохозяйственных регионах России, совокупная посевная площадь в которых составляет 92% всей посевной площади страны. Пример результатов детектирования пахотных земель приведен на рис. 3.

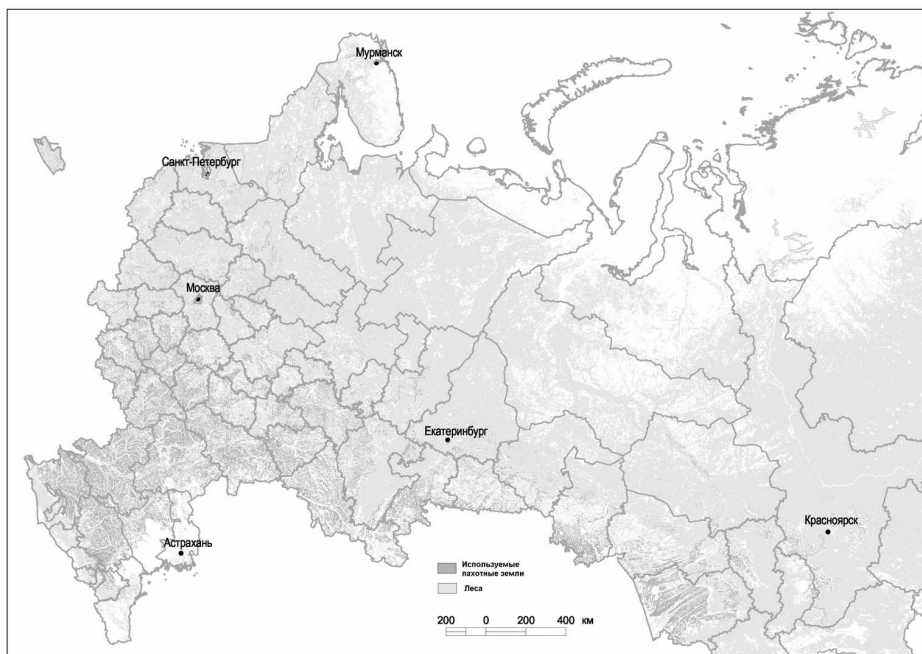


Рис. 3. Фрагмент карты используемых пахотных земель по данным MODIS

Анализ полученных результатов

Анализ полученных данных о площадях используемых пахотных земель был выполнен в сравнении с материалами государственной статистики [1, 11, 12]. Следует отметить, что данные государственной статистики также не могут рассматриваться в качестве абсолютно достоверных, однако в настоящее время они представляют собой единственный доступный источник данных о наличии пахотных земель на уровне субъектов РФ. Для сравнения были использованы данные о посевных площадях в разрезе субъектов РФ и данные о посевных площадях в разрезе административных районов на территорию Ростовской области и Краснодарского края. Результаты сравнения с имеющимися данными государственной статистики в разрезе субъектов РФ приведены на рис. 4. Результаты сравнения площадей пахотных земель с данными государственной статистики в разрезе административных районов показаны на рис. 5.

Разница в площадях между данными государственной статистики и результатами детектирования объясняется различными причинами для различных регионов. Наибольшая ошибка наблюдается в Волгоградской области, что, по-видимому, вызвано ошибочным детектированием большого количества заброшенных полей и степной растительности, что потребует учета при дальнейшем развитии алгоритма. Степная растительность детектируется как ИПЗ из-за того, что динамика развития растительности в степях тесно связана с количеством и временем выпадения осадков, и поэтому может иметь существенные отличия в различные по метеорологическим условиям годы. Это приводит к относительно низким значениям межгодовых коэффициентов корреляции PVI на участках степной растительности, и, как следствие, их ошибочному включению в обучающую выборку А сельскохозяйственной растительности.

В ряде субъектов РФ, расположенных вне зоны основного земледелия, использование алгоритма приводит к получению заниженных, по сравнению с данными государственной статистики, оценок площади ИПЗ. Это связано с относительно непродолжительным вегетационным периодом развития растительности в силу климатических условий в данных регионах, и, как следствие, сравнительно небольшим различиям в значениях d_{\min} для сельскохозяйственной и естественной растительности.

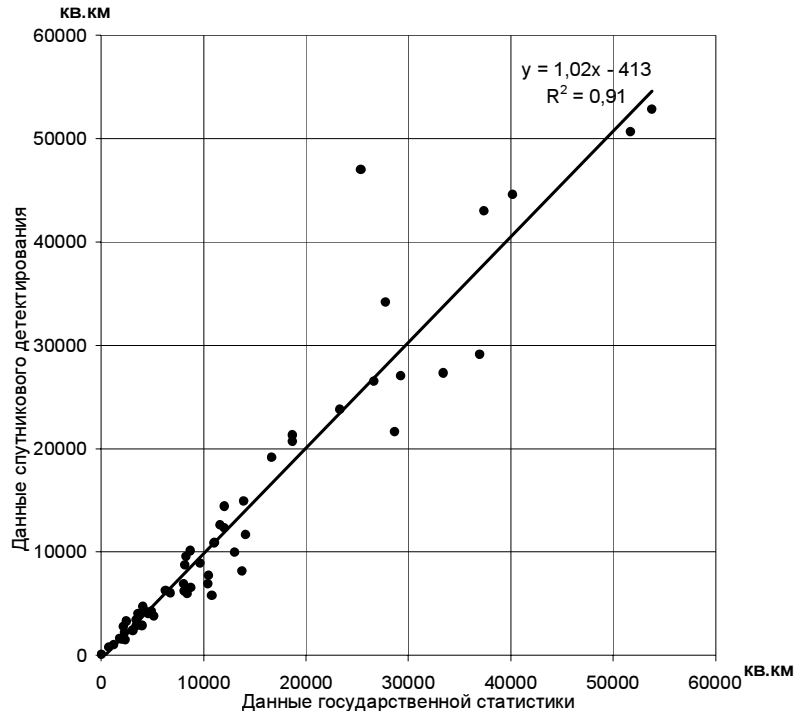


Рис. 4. Сравнение результатов детектирования ИПЗ по данным MODIS с данными государственной статистики по субъектам РФ

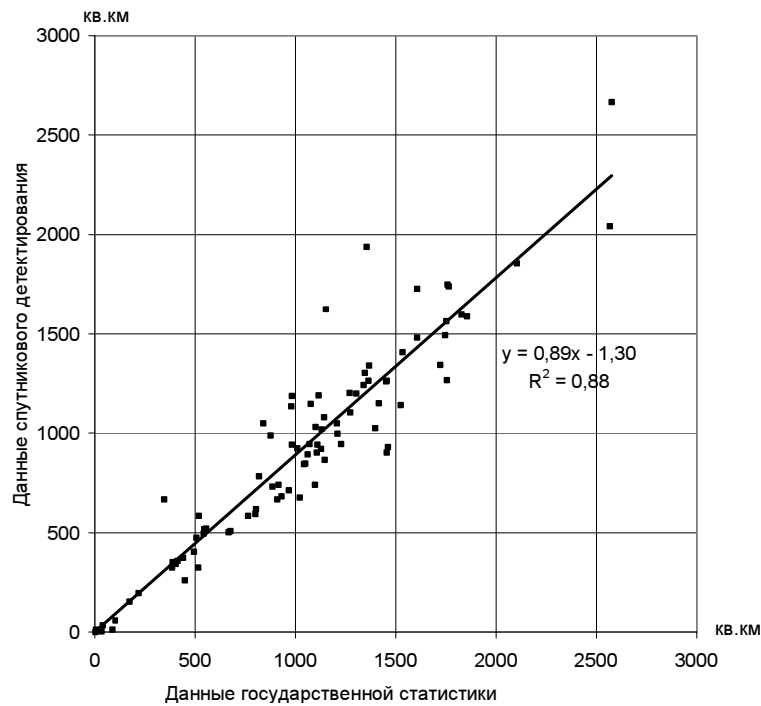


Рис. 5. Сравнение результатов детектирования ИПЗ по данным MODIS с данными государственной статистики по районам Ростовской области и Краснодарского края

Отметим также, что относительно невысокое пространственное разрешение используемых спутниковых данных (250м) приводит к достаточно большим погрешностям в оценке площадей ИПЗ, достигающих наибольших значений для относительно небольших территорий (хозяйство, район) и уменьшающихся при оценке площадей посевов на уровне субъектов РФ. Указанная ошибка может быть существенно снижена при использовании дополнительных данных о границах и площадях отдельных полей. В этом случае возможен совместный анализ данных, при котором посевы детектируются с использованием спутниковых изображений, а площади подсчитываются с учетом заранее известных границ полей, определенных по спутниковым данным более высокого разрешения, например Landsat-ETM+ или Метеор-3М/МСУ-Э.

Кроме того, проведенный анализ причин ошибочного детектирования используемых пахотных земель по данным Terra-MODIS указывает на потенциальную возможность повышения точности результатов за счет включения в алгоритм одновременно со значениями продолжительности вегетационного периода дополнительных признаков классификации, характеризующих различные аспекты сезонной динамики растительности.

Заключение

Описанные в статье метод обработки спутниковых данных обеспечивает принципиальную возможность детектирования используемых пахотных земель. Предполагаемое развитие метода будет направлено на дальнейшее повышение уровня региональной адаптивности алгоритма с целью обеспечения возможности его эффективного использования для территории всех аграрных регионов России, что является одним из важных шагов создания национальной системы сельскохозяйственного мониторинга.

Полученные в работе данные о площадях и пространственном распределении пахотных земель могут быть интересны для решения задач регулирования аграрного комплекса, а также для проведения региональных научных исследований динамики биосферы и климата. Следует также отметить, что полученные с помощью предлагаемого метода данные будут использованы при проведении и анализе результатов Всероссийской сельскохозяйственной переписи 2006 года (<http://www.perepis2006.ru>). Полученные в рамках настоящей работы данные включены в Информационную систему спутникового мониторинга бореальных экосистем TerraNorte [14] (<http://terranorte.iki.rssi.ru/>), создаваемую при поддержке РФФИ (грант 04-07-90263-в).

Литература

1. Барталев С.А., Луян Е.А., Нейштадт И.А., Савин И.Ю. Классификация некоторых типов сельскохозяйственных посевов в южных регионах России по спутниковым данным MODIS // Исследование Земли из космоса, 2006. №3. В печати.
2. Сельское хозяйство, охота и лесоводство в России. 2004: Стат. сб. // М.: Росстат, 2004. 478 с.
3. Агропромышленный комплекс России в 2004 году, отв. за выпуск: Акаткин Ю.М., Езепчук В.С., Корсун А.П. // М.: Россельхоз, 2005. 577 с.
4. Толковый словарь по сельскохозяйственной метеорологии // Санкт-Петербург: Гидрометеоиздат, 2002. 470 с.
5. Wiegand C.L., Richardson A.J., Escobar D.E., Gerbermann A.H. // Vegetation Indices in Crop Assessments // Remote Sensing of Environment, 1991, №35. P.105-119.
6. Earth Observation System (EOS) Data Products Handbook, Eds.: King M.D., Closs J., Spangler S., Greenstone R. // NASA Goddard Space Flight Center, Greenbelt Maryland, 2003. Vol. I. 258 p.

7. *Justice C.O., Townshend J.R.G., Vermote E.F. et al.* An overview of MODIS Land data processing and product status // *Remote Sensing of Environment*, 2002, №83. P.3-15.
8. *Бурцев М.А., Мазуров А.А., Нейштадт И.А., Прошин А.А.* Построение архива спутниковых данных для анализа динамики растительности // *данный сборник*.
9. *Нейштадт И.А.* Построение безоблачных композитных спутниковых изображений MODIS для мониторинга растительности // *данный сборник*.
10. *Нейштадт И.А., Барталев С.А., Еришов Д.М. и др.* Алгоритмы анализа данных спутниковых наблюдений TERRA-MODIS для мониторинга сельскохозяйственных земель // *Сб. науч. статей Международной научно-технической конференции, посвященной 225-летию МИИГАиК, М.: Геоинформатика, 2004. С.205-209.*
11. *Посевные площади, валовые сборы и урожайность сельхозкультур в хозяйствах всех категорий Ростовской области на 1 декабря 2003 года // Ростов-на-Дону: Ростовстат, 2004.*
12. *Площади, валовые сборы и урожайность сельскохозяйственных культур и многолетних насаждений // Краснодар: Краснодарстат, 2004.*
13. *Bartalev S.A., Belward A.S., Erchov D.V., and Isaev A.S.* A new SPOT4-VEGETATION derived land cover map of Northern Eurasia // *International Journal of Remote Sensing*, 2003. Vol. 24, No. 9. P.1977-1982.
14. *Уваров И.А., Барталев С.А., Егоров В.А., Лупян Е.А., Нейштадт И.А., Ховратович Т.С.* Структура и функциональные возможности информационной системы TerraNorte для поддержки спутникового мониторинга бореальных экосистем // *данный сборник*.