Метод автоматического распознавания сельскохозяйственных культур на основе спутниковых данных и имитационной модели развития растений

Д. Е. Плотников, С. А. Хвостиков, С. А. Барталев

Институт космических исследований РАН, Москва, 117997, Россия E-mail: dmitplot@d902.iki.rssi.ru

В работе рассматривается метод автоматического создания независимой и репрезентативной обучающей выборки на основе имитационной модели роста и развития растений WOFOST, параметризованной по историческим наземным и спутниковым данным, для распознавания посевов сельскохозяйственных культур текущего сезона вегетации. Предложенный метод позволил осуществить уверенное распознавание полей, занятых пятью различными культурами, без использования актуальной наземной информации об их расположении или априорных знаний о различиях их сезонного развития. Рассчитанная на основе наземных данных для пяти исследуемых культур общая точность распознавания составила 85 %. Рассмотрение временных серий модельных значений индекса листовой поверхности в качестве эталонов фазовых характеристик позволило использовать относительно простые критерии для определения типичных представителей выбранных культур на основе анализа их сезонной фенологии и дало возможность построить опорную выборку для обучения и дальнейшей классификации. Такое решение может быть использовано при распознавании широкого набора сельскохозяйственных культур на больших территориях, где основной проблемой является сложность или невозможность своевременного получения опорных данных текущего сезона вегетации для обучения классификатора другими методами.

Ключевые слова: распознавание культур, дистанционное зондирование, автоматическое построение обучающей выборки, объектно-ориентированная классификация, имитационное моделирование, WOFOST

Одобрена к печати: 01.08.2018 DOI: 10.21046/2070-7401-2018-15-4-131-141

Введение

Своевременная информация о размещении сельскохозяйственных культур обладает первостепенной важностью для оперативной оценки состояния посевов, контроля севооборота и прогнозирования урожайности. Использование данных дистанционного зондирования Земли имеет очевидные преимущества перед традиционными методами инвентаризации земель благодаря оперативности получения информации, объективности и пространственной детальности оценок. Автоматические методы обработки и анализа спутниковых данных позволяют минимизировать материальные и трудовые затраты, обеспечивая независимость, своевременность и повторяемость результатов, и имеют особую актуальность для оценки сельскохозяйственных земель России ввиду протяжённых размеров страны.

Дистанционный мониторинг широко распространён во многих важных сельскохозяйственных регионах и развитых странах, включая оценку состояния и прогнозирование урожайности культур, в том числе с использованием имитационных моделей роста и развития растений (de Wit et al., 2012; Ines et al., 2013; Tripathy et al., 2013). Дистанционная инвентаризация и контроль севооборотов актуальны в случаях недостаточности имеющейся информационной инфраструктуры, невозможности получения неискажённой и своевременной информации о размещении культур, а также необходимы для независимых национальных и наднациональных систем региональной и глобальной оценки и прогнозирования урожая, таких как MARS и CropWatch (Bingfang et al., 2014), или для систем информационного обеспечения в области продовольственного ценообразования, таких как AMIS и Crop Monitor.

Современные методы дистанционного распознавания культур текущего сезона вегетации зачастую предполагают использование априорной информации о сроках их сева, этапах

развития и уборки (Crop Calendar) (Foerster et al., 2012; Zafar, Waqar, 2014) либо используют устойчивые различия сезонной динамики дистанционно измеряемых характеристик растений (Massey et al., 2017; Palchoudhuri et al., 2018). Практически всегда используются преимущества временных серий спутниковых изображений, которые доказали свою эффективность для решения этих задач (Плотников и др., 2011; Foerster et al., 2012; Massey et al., 2017; Qian et al., 2017; Song et al., 2017). Использование обучающей выборки, описывающей пространственную и тематическую вариабельность характеристик объектов земной поверхности, является необходимым условием обеспечения требуемого уровня достоверности их распознавания (Foody et al., 2006) на основе параметрических и непараметрических классификаторов. Получение из независимых источников своевременной и пространственно распределённой на больших территориях информации о размещении культур, также как и накопление такого рода данных на основе наземных обследований или экспертного анализа спутниковых изображений сопряжено со значительными организационными трудностями, финансовыми и временными затратами и является трудновыполнимым.

Биофизическое моделирование позволяет имитировать продукционные процессы растений с использованием актуальной метеорологической информации, доступной в реальном времени (Jones et al., 2003). В частности, модель WOFOST (Diepen et al., 1989) воспроизводит процессы накопления, развития и отмирания биомассы побегов, плодов, листьев и корней, используя в том числе полученные в лабораторных условиях зависимости между структурными единицами растений с учётом сортовых особенностей. Последнее позволяет восстанавливать геометрические характеристики и архитектуру ключевых видов сельскохозяйственной растительности, предоставляя возможность их сопоставления с дистанционными измерениями со спутников. Способность имитационных моделей воспроизводить характеристики растений и динамику их сезонного развития продемонстрирована в ряде работ (Ines et al., 2013; Lobell et al., 2015; Tripathy et al., 2013) и успешно применялась для прогнозирования урожайности или принятия управленческих решений.

В настоящей работе рассматривается метод автоматического создания независимой и репрезентативной обучающей выборки на основе имитационной модели роста растений WOFOST, параметризованной по историческим наземным и спутниковым данным для регионального распознавания посевов сельскохозяйственных культур текущего сезона вегетации. Описываемый метод позволил осуществить уверенное распознавание полей, занятых пятью различными культурами, без использования актуальной наземной информации о расположении этих культур или каких-либо априорных знаний о различиях их сезонного развития.

Регион исследования и данные

Наземные данные были представлены результатами полевых обследований 380 полей шести хозяйств Курской области (рис. 1, см. с. 133), охватывающих период с 2007 по 2016 г., и включали в себя, помимо границ полей, названия гибридов сельскохозяйственных культур, даты наступления основных фенологических фаз и даты внесения удобрений, тип почвы и другие данные. Курская область является преимущественно сельскохозяйственным регионом, где представлены почти все важнейшие с точки зрения объёмов национальных валовых сборов культуры. Территория является однородной с точки зрения глобальной агростратификации GAES (Fischer et al., 2012), что означает минимальные различия почвенно-климатических условий и сельскохозяйственных практик в регионе исследования. На обследованных полях преобладают шесть культур: озимая пшеница, озимая рожь, кукуруза, подсолнечник, соя и ячмень. Другие обследованные поля, в частности находящиеся под паром, а также засеянные горохом, сахарной свёклой, люцерной и другими культурами, не рассматривались ввиду их незначительного количества. Озимая рожь была также исключена из рассмотрения по причине отсутствия для неё модельной параметризации. Таким образом, в работе исследовались пять культур: кукуруза, подсолнечник, соя, ячмень и озимая пшеница, для которых совокупная площадь посевов составляет более 80 % всей посевной площади региона интереса.

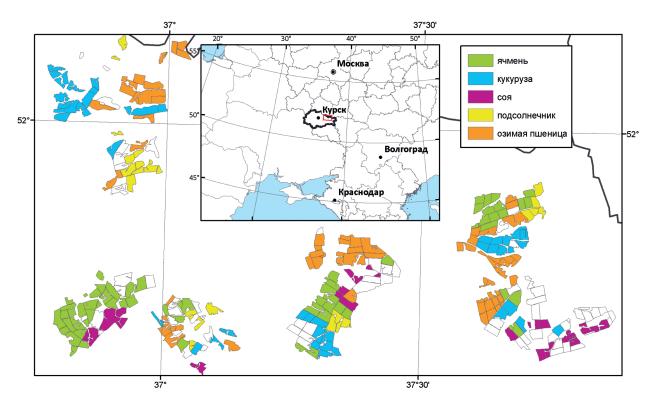


Рис. 1. Регион исследования, границы полей с наземными данными и расположение культур на них в 2016 г.

Данные дистанционного зондирования среднего пространственного разрешения, полученные спектрорадиометром MODIS со спутников Тегга и Аqua, были использованы для сопоставления с результатами моделирования. Разработанная в Институте космических исследований Российской академии наук (ИКИ РАН) технология предварительной обработки данных MODIS (продукт MOD09) предусматривает исключение измерений, находящихся под влиянием мешающих факторов, и создание безоблачных композитных изображений в информативных спектральных диапазонах длин волн за фиксированные интервалы времени (Барталев и др., 2016). В частности, в работе использовались временные серии четырёхдневных композитных изображений MODIS в ближнем инфракрасном диапазоне, на основе которых были получены временные серии измерений LAI (Leaf Area Index). Временные серии исторических спутниковых данных были использованы на этапе параметризации модели, а аналогичные данные текущего сезона применялись в процессе создания актуальной обучающей выборки.

Для выявления посевов сельскохозяйственных культур на основе созданной выборки в качестве признаков распознавания использовались сезонные временные серии спутниковых изображений высокого пространственного разрешения Landsat-8 (OLI), Sentinel-2A/B (MSI). Сезонные временные серии Landsat-8 были также использованы для выделения сезонно-однородных объектов растительного покрова на весь регион исследования разработанным ранее методом сегментации временных серий изображений (Плотников и др., 2018). В обоих случаях были выбраны информативные для дистанционного мониторинга растительности спектральные каналы, соответствующие красному, ближнему и среднему ИК-диапазонам длин волн, а также каналы «красного края» прибора MSI спутниковой системы Sentinel-2A/B. Всего для классификации было использовано 30 изображений, охватывающих период с апреля по сентябрь 2016 г.

На этапах калибровки и имитации модель использует метеорологическую информацию, содержащую измерения температуры воздуха и количества осадков, скорости и направлении ветра, которая обновляется каждые три часа. Для параметризации модели был использован временной ряд результатов математического реанализа измерений метеостанциями с 2011

по 2016 г., ассимилированных на узлы регулярной сети, а в случае недоступности такого рода данных использовалась информация из архива NCEP (Kalnay et al., 1996) с пространственным разрешением регулярной сети 1° и временным разрешением 6 ч. Информация о мощности падающего коротковолнового излучения и парциальном давлении водяного пара была получена из базы данных NCEP. Метеорологические данные посуточно агрегировались, причём измерения осадков и радиации суммировались, измерения давления осреднялись, для температуры воздуха использовалось минимальное и максимальное значение за сутки.

Параметризация модели и эталоны фазовых характеристик

Процесс создания региональной обучающей выборки для рассматриваемых культур состоял из двух этапов. На первом этапе наземная информация о расположении сельскохозяйственных культур и данные о динамике дистанционных измерений LAI за предыдущие годы были использованы для параметризации имитационной модели развития растений WOFOST. Параметры модели для каждой культуры были оптимизированы в процессе сопоставления временной серии дистанционных измерений LAI и соответствующих им во времени значений листового индекса зелёной биомассы растения, имитируемых моделью для местоположения поля с этой культурой. Ввиду ожидаемых различий абсолютных значений листового индекса, получаемых по некалиброванной модели и по данным дистанционных наблюдений, целью оптимизации являлась максимизация коэффициента корреляции Пирсона между этими двумя синхронизованными по времени наборами данных посредством вариации параметров модели. Дистанционные значения LAI рассчитывались с использованием измерений отражательной способности зелёного листа в ближнем инфракрасном канале на основе формулы

$$\mathrm{LAI}(t) = -\frac{\log \left(1 - \frac{R(t) - R_{\min}}{R_{\max} - R_{\min}}\right)}{0.6}, \ \mathrm{rge} \ R(t), \ R_{\min} \ \mathrm{u} \ R_{\max} \ - \ \mathrm{значения} \ \mathrm{коэффициента} \ \mathrm{спектральной}$$

яркости (КСЯ) в текущий момент времени, а также его пятилетние минимум и максимум соответственно. Отметим, что в зависимости от способа измерения LAI может включать в себя также площадь отмирающих и непригодных для фотосинтеза побегов и листьев — например, в случае измерения индекса на основе радиолокационных данных. Поскольку в настоящей работе индекс получен по данным измерений в ближнем ИК-диапазоне, под дистанционным измерением LAI растений здесь понимается листовой индекс именно зелёной биомассы, что соответствует характеристике, возвращаемой моделью. Оптимизация модели выполнялась с использованием генетического алгоритма Genoud. В процессе оптимизации настраивались модельные параметры, отвечающие за сезонную динамику LAI растения: суммы накопленных температур воздуха, соответствующие разным этапам фенологического развития, а также продолжительность жизни листьев.

На следующем этапе, когда оптимальные параметры модели для каждой из рассматриваемых культур установлены, становится возможным имитировать процесс фенологического развития структурных единиц растений и накопление ими зелёной биомассы с учётом агрометеорологических особенностей текущего сезона вегетации. При этом необходимо располагать информацией о датах сева моделируемых культур, получение которой в рамках этого исследования было осуществлено двумя способами: на основе анализа временных серий спутниковых данных и на основе совокупности метеорологических критериев для каждой культуры. Выяснилось, что оба метода способны достаточно точно оценить дату сева, которая оказалась близка к действительной.

Для установления принадлежности текущего объекта сельскохозяйственной растительности к одному из классов моделируемых культур осуществлялось сравнение временной серии дистанционных измерений LAI и соответствующей ей во времени серии прогнозируемых моделью значений (эталон фазовых характеристик). Примеры соответствия разносезонных временных серий спутниковых и модельных значений LAI представлены на рис. 2 (см. с. 135).

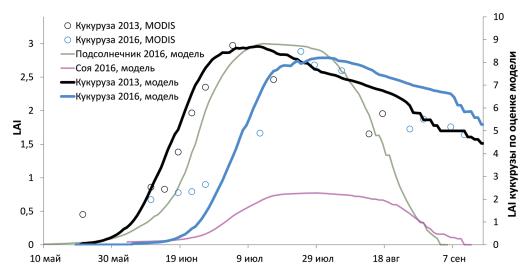


Рис. 2. Примеры прогнозируемых моделью значений и полученных на основе дистанционных наблюдений значений LAI для сезонов 2013 и 2016 гг.

Сравнение серий и выбор подходящего эталона происходили на основе одновременного учёта значений трёх параметров: величины корреляции R временной серии дистанционных измерений LAI с наиболее подходящим эталоном, величины отличия значения R от ближайшего аналогичного значения для другой культуры ΔR и количества пар значений N, на основе которых рассчитывалась корреляция. Параметр N может варьировать по причине естественного частичного отсутствия дистанционных измерений ввиду исключения зашумленных дистанционных измерений из-за влияния облачности и теней от неё. Для каждого параметра экспертным путём были найдены пороговые значения: минимальное, за пределами которого соответствие временной серии LAI рассматриваемой культуры модельному эталону признаётся недостоверным (нулевая достоверность), и максимальное, при превышении которого соответствие считается достоверным (единичная достоверность). Общая достоверность Conf отнесения рассматриваемой культуры к классу эталона оценивалась на основе произведения достоверностей, определённых для каждого из параметров: $Conf = Conf(R) \cdot Conf(\Delta R) \cdot Conf(N)$.

Создание обучающей выборки

Параметризация модели осуществлялась с использованием наземной и дистанционной информации за 2011—2014 гг. для 54 полей в пределах одного хозяйства для всех культур, кроме сои; модель этой культуры была параметризована на основе данных 2015—2016 гг. С помощью параметризованной модели для каждой из пяти рассматриваемых культур были получены региональные фазовые эталоны LAI сезона 2016 г.

Для оценки возможностей предлагаемого подхода эталоны LAI были использованы для создания обучающего набора данных только в пределах границ обследованных полей с наземной информацией о культурах, где являлось возможным сопоставление классов созданной выборки и культуры *in situ*. В результате класс эталона был достоверно присвоен 23 полям, причём для 22 из них было обнаружено совпадение классов, определённых по дистанционным и по наземным данным. Несмотря на высокую точность полученного обучающего набора данных, выборка, очевидно, не являлась репрезентативной для регионального обучения и распознавания культур.

Для увеличения числа кандидатов, подходящих для построения обучающей выборки, были задействованы все используемые пахотные земли региона исследования согласно актуальной карте, построенной в ИКИ РАН по данным MODIS (Bartalev et al., 2016; Waldner et al., 2016). В целях выделения сезонно-однородных объектов используемой пашни (участков сельскохозяйственных земель, занятых одной культурой) для дальнейшего сравнения с эталоном

была выполнена сегментация временных серий спутниковых изображений разработанным ранее методом (Плотников и др., 2018). Сегментация позволила получить 3676 объектов площадью более 30 га каждый, в дальнейшем рассматриваемых в качестве отдельных элементов обучающей выборки (полей). Сравнения фазовых эталонов со спутниковыми оценками позволили определить наиболее вероятную культуру и оценить значение критерия достоверности её определения для каждого выбранного объекта. Для каждой из пяти культур было выбрано 40 полей, обладающих максимальным для этой культуры значением достоверности с учётом их пространственного расположения для обеспечения необходимого географического охвата и репрезентативности выборки.

Результаты распознавания на основе обучающей выборки

Региональное распознавание культур выполнялось на основе полученной и описанной выше выборки, а также сезонных временных серий мультиспектральных спутниковых изображений высокого пространственного разрешения Landsat-8 (OLI) и Sentinel-2A (MSI) с помощью метода Random Forest (Breiman, 2001) на уровне отдельных пикселей. Использованный набор спутниковых данных содержал измерения КСЯ растений не только в красном и ближнем ИК-диапазонах (как 250-метровые каналы MODIS), но и в двух коротковолновых ИК (SWIR1 и SWIR2) диапазонах длин волн. Кроме этого, были задействованы три канала «красного края» (Vegetation Red Edge) сенсора MSI. Все указанные диапазоны длин волн используются для оценки зелёной биомассы растений и являются информативными при разделении культур. Результирующим классом поля считался преобладающий в его границах класс. Общая точность классификации оценивалась на основе 255 полей, которым по наземным данным была присвоена одна из пяти исследуемых культур, и составила 85%. Пространственное распределение ошибок распознавания представлено на рис. 3, где контурами обозначены все поля с наземной информацией, а цветом — поля одной из пяти исследуемых культур. Матрица ошибок классификации представлена в таблице. Можно видеть, что наибольшая ошибка связана с перепутыванием классов сои и подсолнечника.

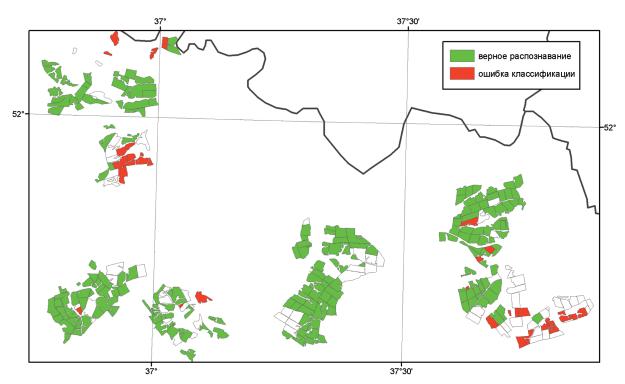


Рис. 3. Пространственное распределение ошибок классификации. Незаполненные контуры соответствуют обследованным полям, занятым остальными культурами

| Матрица ошибок результатов классификации на основе опорной выборки, |
|---|
| полученной по модельным эталонам |

| | Ячмень | Кукуруза | Соя | Подсолнечник | Пшеница | Producer accuracy |
|---------------|--------|----------|------|--------------|---------|-------------------|
| Ячмень | 69 | 0 | 0 | 0 | 1 | 0,99 |
| Кукуруза | 0 | 49 | 0 | 0 | 0 | 1,00 |
| Соя | 0 | 8 | 15 | 8 | 1 | 0,47 |
| Подсолнечник | 0 | 0 | 13 | 13 | 0 | 0,50 |
| Пшеница | 7 | 0 | 0 | 1 | 71 | 0,91 |
| User accuracy | 0,91 | 0,86 | 0,54 | 0,62 | 0,97 | 0,85 |

Обсуждение

Анализ результатов классификации показывает, что предложенный метод может быть достаточно успешно использован для распознавания сельскохозяйственных культур текущего сезона вегетации на основе спутниковых данных и имитационного моделирования сезонного развития растений.

Наибольшие ошибки распознавания приходятся на классы сои и подсолнечника, сезонные серии LAI которых сильно коррелируют и почти неотличимы в терминах используемой метрики. Однако абсолютные значения листового индекса этих культур заметно отличаются (см. рис. 2), что позволяет ожидать возможность их успешного разделения при использовании дополнительных метрик для совершенствования метода. При гипотетическом полном разделении этих классов общая точность классификации возрастёт до 93 %. Тем не менее достигнутое в рамках данного эксперимента значение общей точности 85 % оказывается сопоставимым и даже превышает значения, полученные ранее в аналогичных экспериментах, проведённых при обучении классификатора на основе только наземной информации текущего сезона вегетации (Плотников и др., 2011). Дальнейшее повышение точности и оперативности оценки возможно за счёт использования данных других спутниковых систем.

Результат распознавания культур для региона исследования на уровне выделенных сезонно-однородных объектов пахотных земель представлен на *puc. 4* (см. с. 138). Отметим, что он был получен при обучении классификатора только на классах ячменя, кукурузы, сои, подсолнечника и озимой пшеницы. Необходимо учитывать, что другие классы культурной растительности также имеются в регионе (суммарно менее 20 % посевной площади), однако приведённый результат даёт представление о пространственном распределении важнейших сельскохозяйственных культур на территории региона исследования.

Поскольку наиболее востребованным является получение оперативной информации о размещении культур текущего сезона вегетации, перспективным представляется исследование вопроса о возможности заблаговременного построения опорной выборки и зависимости её качества от даты построения. Используемые в настоящей работе дистанционные, метеорологические и модельные данные соответствовали интервалу времени с момента сева до даты уборки, отодвигая время создания обучающей выборки и карты расположения культур к началу осени текущего сезона вегетации. Дополнительные усилия планируется направить на изменение условий работы метода с целью обеспечения более оперативных оценок, а также на исследование влияния этих изменений на качество обучающей выборки и точность классификации.

Несмотря на то что регион исследования является однородным с точки зрения глобальной агростратификации GAES, дальнейшая работа на больших неоднородных территориях потребует локализованной параметризации модели, осуществляемой с использованием местных данных полевых обследований или экспертной интерпретации спутниковых изображений высокого пространственного разрешения и других вспомогательных данных.

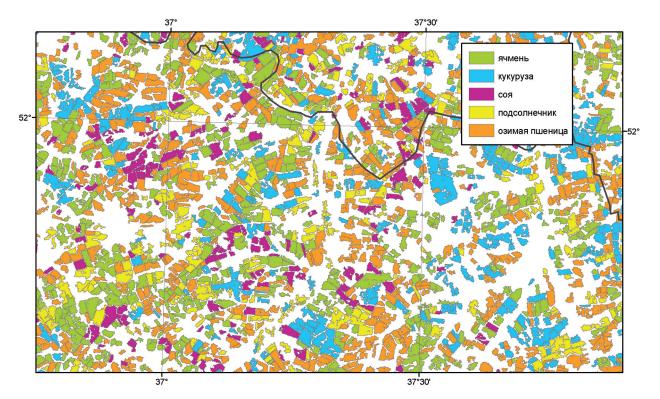


Рис. 4. Результат распознавания исследуемых культур на территории региона исследования

Кроме этого, для автоматического применения предложенного подхода на больших территориях необходимы надёжные методы независимого и своевременного определения дат сева и уборки урожая, которые могут быть основаны на использовании спутниковых, метеорологических и модельных индикаторов или их комбинаций.

Заключение

Предложенный метод может оказаться перспективным при разработке экономически эффективных и оперативных технологий распознавания сельскохозяйственных культур текущего сезона вегетации по данным дистанционного зондирования, в том числе на больших территориях ввиду своей относительной простоты и универсальности. Для его применения требуется ограниченный набор дистанционных и наземных данных, необходимых на этапах настройки имитационной модели. Рассмотрение временных серий модельных значений индекса листовой поверхности в качестве эталонов фазовых характеристик позволяет использовать простые критерии для выделения типичных представителей выбранных культур на основе анализа сезонной динамики спектрально-отражательных характеристик и построить достаточно точную опорную выборку для дальнейшей классификации. Такое решение может быть универсальным при распознавании широкого набора сельскохозяйственных культур на больших территориях, где основной проблемой является невозможность своевременного и единовременного получения опорных данных текущего сезона вегетации другими методами.

Исследования выполнены в рамках темы «Мониторинг» (госрегистрация № 01.20.0.2.00164) с использованием ресурсов Центра коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды (Лупян и др., 2015).

Литература

- 1. Барталев С.А., Егоров В.А., Жарко В.О., Лупян Е.А., Плотников Д.Е., Хвостиков С.А., Шабанов Н.В. Спутниковое картографирование растительного покрова России. М.: ИКИ РАН, 2016. 208 с.
- 2. Лупян Е.А., Прошин А.А., Бурцев М.А., Балашов И.В., Барталев С.А., Ефремов В.Ю., Кашниц-кий А.В., Мазуров А.А., Матвеев А.М., Суднева О.А., Сычугов И.Г., Толпин В.А., Уваров И.А. Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 247—267.
- 3. *Плотников Д. Е.*, *Барталев С. А.*, *Жарко В. О.*, *Михайлов В. В.*, *Просянникова О. И.* Экспериментальная оценка распознаваемости агрокультур по данным сезонных спутниковых измерений спектральной яркости // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 1. С. 199–208.
- 4. *Плотников Д. Е., Колбудаев П. А., Барталёв С. А.* Выделение сезонно-однородных областей на основе анализа временных серий спутниковых изображений // Компьютерная оптика. Самара: ИСОИ РАН, 2018. Т. 42. № 3. С. 447—456. DOI: 10.18287/2412-6179-2018-42-3-447-456.
- 5. Bartalev S.A., Plotnikov D. E., Loupian E.A. Mapping of arable land in Russia using multiyear time series of MODIS data and the LAGMA classification technique // Remote Sensing Letters. 2016. V. 7. No. 3. P. 269–278. DOI: 10.1080/2150704X.2015.1130874.
- 6. *Bingfang W., Jihua M., Qiangzi L., Nana Y., Xin D., Miao Z.* Remote sensing-based global crop monitoring: experiences with China's CropWatch system // Intern. J. Digital Earth. 2014. V. 7. Iss. 2. P. 113–137. DOI: 10.1080/17538947.2013.821185.
- 7. *Breiman L.* Random forests // Machine Learning. 2001. V. 45. Iss. 1. P. 5–32. DOI: 10.1023/A:1010933404324.
- 8. *de Wit A.*, *Duveiller G.*, *Defourny P.* Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations // Agricultural and Forest Meteorology. 2012. V. 164. P. 39–52.
- 9. *Diepen C.*, *Wolf J.*, *Keulen H.*, *Rappoldt C.* WOFOST: a simulation model of crop production // Soil Use and Management. 1989. V. 5. Iss. 1. P. 16–24.
- 10. Fischer G., Nachtergaele F., Prieler S., Teixeira E., Töth G., Velthuizen H., Verelst L., Wiberg D. Global Agro-Ecological Zones (GAEZ v3.0). Model Documentation. IIASA/FAO, 2012.
- 11. Foerster S., Kaden K., Förster M., Itzerott S. Crop type mapping using spectral-temporal profiles and phenological information // Computers and Electronics in Agriculture. 2012. V. 89. P. 30–40. DOI: 10.1016/j. compag.2012.07.015.
- 12. *Foody M. G.*, *Mathur A.*, *Sanchez Hernandez C.*, *Boyd D.* Training set size requirements for the classification of a specific class // Remote Sensing of Environment. 2006. V. 104. P. 1–14. DOI: 10.1016/j.rse.2006.03.004.
- 13. *Ines A. V. M.*, *Das N. N.*, *Hansen J. W.*, *Njoku E. G.* Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction // Remote Sensing of Environment. 2013. V. 138. P. 149–164. DOI: 10.1016/j.rse.2013.07.018.
- 14. Jones J. W., Hoogenboom G., Porter C. H., Boote K. J., Batchelor W. D., Hunt L. A., Wilkens P. W., Singh U., Gijsman A. J., Ritchie J. T. The DSSAT cropping system model // European J. Agronomy. 2003. V. 18. Iss. 3. P. 235–265. DOI: 10.1016/S1161-0301(02)00107-7.
- 15. Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Chellian M., Ebisuzaki W., Higgins W., Janowiak J., Mo K. C., Ropelewski C., Wang J., Leetmaa A., Reynolds R., Jenne R., Joseph D. The NCEP/NCAR 40-year reanalysis project // Bull. American Meteorological Society. 1996. V. 77. Iss. 3. P. 437–471. DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- 16. Lobell D. B., Thau D., Seifert C., Engle E., Little B. A scalable satellite-based crop yield mapper // Remote Sensing of Environment. 2015. V. 164. P. 324–333. DOI: 10.1016/j.rse.2015.04.021.
- 17. *Massey R.*, *Sankey T.*, *Congalton R.*, *Yadav K.*, *Thenkabail P.*, *Ozdogan M.*, *Sánchez Meador A.* MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types // Remote Sensing of Environment. 2017. V. 198. P. 490–503. DOI: 10.1016/j.rse.2017.06.033.
- 18. *Palchoudhuri Y.*, *Valcarce-Diñeiro R.*, *King P.*, *Sanabria-Soto M.* Classification of multi-temporal spectral indices for crop type mapping: a case study in Coalville, UK // J. Agricultural Science. 2018. V. 156. Iss. 1. P. 24–36. DOI: 10.1017/S0021859617000879.
- 19. Song Q., Hu Q., Zhou Q., Hovis C., Xiang M., Tang H., Wu W. In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest // Remote Sensing. 2017. V. 9. P. 1184. DOI: 10.3390/rs9111184.

- 20. *Tripathy R.*, *Chaudhari K. N.*, *Mukherjee J.*, *Ray S. S.*, *Patel N. K.*, *Panigrahy S.*, *Parihar J.* Forecasting wheat yield in Punjab state of India by combining crop simulation model WOFOST and remotely sensed inputs. // Remote Sensing Lett. 2013. V. 4. Iss. 1. P. 19–28. DOI: 10.1080/2150704X.2012.683117.
- 21. Waldner F., De Abelleyra D., Santiago V., Zhang M., Wu B., Plotnikov D.E., Bartalev S.A., Lavreniuk M., Skakun S., Kussul N. N., Le Maire G., Dupuy S., Jarvis I., Defourny P. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity // Intern. J. Remote Sensing. 2016. V. 37. Iss. 14. DOI: 10.1080/01431161.2016.1194545.
- 22. Zafar S., Waqar M. Crop Type Mapping by Integrating Satellite Data and Crop Calendar over Okara District, Punjab (Pakistan) // J. Space Technology. 2014. V. 4. Iss. 1. P. 21–25.

Method for automated crop types mapping using remote sensing data and a plant growth simulation model

D. E. Plotnikov, S. A. Khvostikov, S. A. Bartalev

Space Research Institute RAS, Moscow 117997, Russia E-mail: dmitplot@d902.iki.rssi.ru

This paper presents a method for automatic generation of representative and unbiased training set for in-season crop types mapping, based on WOFOST crop simulation model, previously parameterized using ground truth data, as well as meteodata and satellite historical data. The method provided confident mapping of agricultural fields, planted with five different crop types using no ground truth data or a-priori information about their in-season phenology. Calibration fields, most typical for each specific crop, were identified through comparison of their remote sensing-based seasonal time series with modeled LAI values, used as in-season temporal reference. Relatively simple criteria were then used to generate a training set of required fidelity. Overall mapping accuracy calculated using relevant ground truth data has reached 85 %. This kind of approach may be versatile for automated mapping of broad variety of crop types over large heterogeneous areas in cases when in-season ground truth timely information or other kind of calibration data is not available.

Keywords: crop mapping, remote sensing, automated training set building, object-oriented classification, simulation modeling, WOFOST

Accepted: 01.08.2018 DOI: 10.21046/2070-7401-2018-15-4-131-141

References

- 1. Bartalev S.A., Egorov V.A., Zharko V.O., Loupian E.A., Plotnikov D.E., Khvostikov S.A., Shabanov N.V., *Sputnikovoe kartografirovanie rastitel'nogo pokrova Rossii* (Land cover mapping over Russia using Earth observation data), Moscow: IKI RAN, 2017, 208 p.
- 2. Loupian E.A., Proshin A.A., Burtsev M.A., Balashov I.V., Bartalev S.A., Efremov V.Yu., Kashnitskii A.V., Mazurov A.A., Matveev A.M., Sudneva O.A., Sychugov I.G., Tolpin V.A., Uvarov I.A., Tsentr kollektivnogo pol'zovaniya sistemami arkhivatsii, obrabotki i analiza sputnikovykh dannykh IKI RAN dlya resheniya zadach izucheniya i monitoringa okruzhayushchei sredy (IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring), Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 2015, Vol. 12, No. 5, pp. 247–267.
- 3. Plotnikov D. E., Bartalev S. A., Zharko V. O., Mikhailov V. V., Prosyannikova O. I., Eksperimental'naya otsenka raspoznavaemosti agrokul'tur po dannym sezonnykh sputnikovykh izmerenii spektral'noi yarkosti (An experimental assessment of crop types recognisability using time-series of intra-seasonal spectral reflectance measurements by satellite sensor), *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2011, Vol. 8, No. 1, pp. 199–208.

- 4. Plotnikov D. E., Kolbudaev P. A., Bartalev S. A., Vydelenie sezonno-odnorodnykh oblastei na osnove analiza vremennykh serii sputnikovykh izobrazhenii (Identification of dynamically homogeneous areas with time series segmentation of remote sensing data), Komp'yuternaya optika, Samara: ISOI RAN, 2018, Vol. 42, No. 3, pp. 447–456, DOI: 10.18287/2412-6179-2018-42-3-447-456.
- 5. Bartalev S.A., Plotnikov D. E., Loupian E.A., Mapping of arable land in Russia using multiyear time series of MODIS data and the LAGMA classification technique, *Remote Sensing Lett.*, 2016, Vol. 7, No. 3, pp. 269–278, DOI: 10.1080/2150704X.2015.1130874.
- 6. Bingfang W., Jihua M., Qiangzi L., Nana Y., Xin D., Miao Z., Remote sensing-based global crop monitoring: experiences with China's CropWatch system, *Intern. J. Digital Earth*, 2014, Vol. 7, Issue 2, pp. 113–137, DOI: 10.1080/17538947.2013.821185.
- 7. Breiman L., Random forests, *Machine Learning*, 2001, Vol. 45, Issue 1, pp. 5–32, DOI: 10.1023/A:1010933404324.
- 8. de Wit A., Duveiller G., Defourny P., Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations, *Agricultural and Forest Meteorology*, 2012, Vol. 164, pp. 39–52.
- 9. Diepen C., Wolf J., Keulen H., Rappoldt C., WOFOST: a simulation model of crop production, *Soil Use and Management*, 1989, Vol. 5, Issue 1, pp. 16–24.
- 10. Fischer G., Nachtergaele F., Prieler S., Teixeira E., Toth G., Velthuizen H., Verelst L., Wiberg D., *Global Agro-Ecological Zones (GAEZ v3.0)*, *Model Documentation*. IIASA/FAO, 2012.
- 11. Foerster S., Kaden K., Förster M., Itzerott S., Crop type mapping using spectral-temporal profiles and phenological information, *Computers and Electronics in Agriculture*, 2012, Vol. 89, pp. 30–40, DOI: 10.1016/j.compag.2012.07.015.
- 12. Foody M.G., Mathur A., Sanchez Hernandez C., Boyd D., Training set size requirements for the classification of a specific class, *Remote Sensing of Environment*, 2006, Vol. 104, pp. 1–14, DOI: 10.1016/j. rse.2006.03.004.
- 13. Ines A. V. M., Das N. N., Hansen J. W., Njoku E. G., Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction, *Remote Sensing of Environment*, 2013, Vol. 138, pp. 149–164, DOI: 10.1016/j.rse.2013.07.018.
- 14. Jones J. W., Hoogenboom G., Porter C. H., Boote K. J., Batchelor W. D., Hunt L. A., Wilkens P. W., Singh U., Gijsman A. J., Ritchie J. T., The DSSAT cropping system model, *European J. Agronomy*, 2003, Vol. 18, Issue 3, pp. 235–265, DOI: 10.1016/S1161-0301(02)00107-7.
- Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Chellian M., Ebisuzaki W., Higgins W., Janowiak J., Mo K.C., Ropelewski C., Wang J., Leetmaa A., Reynolds R., Jenne R., Joseph D., The NCEP/NCAR 40-year reanalysis project, *Bull. American Meteorological Society*, 1996, Vol. 77, Issue 3, pp. 437–471, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- 16. Lobell D. B., Thau D., Seifert C., Engle E., Little B., A scalable satellite-based crop yield mapper, *Remote Sensing of Environment*, 2015, Vol. 164, pp. 324–33, DOI: 10.1016/j.rse.2015.04.021.
- 17. Massey R., Sankey T., Congalton R., Yadav K., Thenkabail P., Ozdogan M., Sánchez Meador A., MODIS phenology-derived, multi-year distribution of conterminous U.S. crop types, *Remote Sensing of Environment*, 2017, Vol. 198, pp. 490–503, DOI: 10.1016/j.rse.2017.06.033.
- 18. Palchoudhuri Y., Valcarce-Diñeiro R., King P., Sanabria-Soto M., Classification of multi-temporal spectral indices for crop type mapping: a case study in Coalville, UK, *J. Agricultural Science*, 2018, Vol. 156, Issue 1, pp. 24–36, DOI: 10.1017/S0021859617000879.
- 19. Song Q., Hu Q., Zhou Q., Hovis C., Xiang M., Tang H., Wu W., In-Season Crop Mapping with GF-1/WFV Data by Combining Object-Based Image Analysis and Random Forest, *Remote Sensing*, 2017, Vol. 9, p. 1184, DOI: 10.3390/rs9111184.
- 20. Tripathy R., Chaudhari K. N., Mukherjee J., Ray S. S., Patel N. K., Panigrahy S., Parihar J., Forecasting wheat yield in Punjab state of India by combining crop simulation model WOFOST and remotely sensed inputs, *Remote Sensing Lett.*, 2013, Vol. 4, Issue 1, pp. 19–28, DOI: 10.1080/2150704X.2012.683117.
- 21. Waldner F., De Abelleyra D., Santiago V., Zhang M., Wu B., Plotnikov D.E., Bartalev S.A., Lavreniuk M., Skakun S., Kussul N. N., Le Maire G., Dupuy S., Jarvis I., Defourny P., Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity, *Intern. J. Remote Sensing*, 2016, Vol. 37, Issue 14, DOI: 10.1080/01431161.2016.1194545.
- 22. Zafar S., Waqar M., Crop Type Mapping by Integrating Satellite Data and Crop Calendar over Okara District, Punjab (Pakistan), *J. Space Technology*, 2014, Vol. 4, Issue 1, pp. 21–25.