

Development of Seasonal NDVI Profiles References for Main Agricultural Crops

Sergey A. Khvostikov, Sergey A. Bartalev
Space Research Institute, Russian Academy of Sciences, Moscow, Russia
khvostikov@d902.iki.rssi.ru

Abstract

NDVI vegetation index can describe crop dynamic and is widely used in agriculture monitoring. This paper presents method for evaluation of NDVI seasonal profiles references for different crops using big archive of field data with significant level of errors. The NDVI reference profiles are created at the level of administrative districts using available field data and represent typical vegetation index dynamic of each crop in current year. The crop references of neighbor regions are additionally cross-compared to filter incorrect data. This paper presents a method for verification of field data using evaluated NDVI seasonal profiles. This method was tested for more than 400000 fields and expert analysis showed accuracy score at 92%.

Keywords: NDVI seasonal profiles, crop references, clusterization, verification

ПОСТРОЕНИЕ ЭТАЛОНОВ СЕЗОННОЙ ДИНАМИКИ NDVI ДЛЯ ОСНОВНЫХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР

С.А. Хвостиков, С.А. Барталев

Институт космических исследований РАН, Москва, Россия
khvostikov@d902.iki.rssi.ru

Вегетационный индекс NDVI широко используется в задачах сельскохозяйственного мониторинга и позволяет охарактеризовать динамику развития посевов различных культур. В работе рассматривается метод построения эталонов динамики индекса NDVI основных сельскохозяйственных культур на основе большого объема наземных данных, содержащих значительный уровень погрешности. Метод построения эталонов применяется на уровне административных районов для построения локальных эталонов культур. Для построения эталона определяется типичная динамика культуры и выполняется ее сравнение с соседними эталонами для исключения районов с высоким уровнем погрешности наземных данных. Также в работе описывается методика верификации входных данных о выращиваемых на полях культурах на основе полученных эталонов, позволяющая оценить корректность наземной информации. Метод верификации был проверен на более чем 400000 полей, а экспертная оценка его точности показала, что уровень погрешности методики составляет 8%.

Ключевые слова: сезонный временной ряд NDVI, эталоны культур, кластеризация, верификация

Введение

Дистанционное зондирование Земли (ДЗЗ) обеспечивает регулярное получение информации о состоянии сельскохозяйственной растительности в масштабах страны. Данные ДЗЗ позволяют определять площади и положение полей, засеянных разными культурами, оценивать их состояние в течение вегетационного сезона и используются для прогнозирования урожайности. Часто для оценки состояния с.-х. растительности используются вегетационные индексы, такие как нормализованный разностный вегетационный индекс NDVI. Временной ряд значений NDVI характеризует динамику развития культур, и может использоваться в задачах картографирования, оценки состояния и прогнозирования урожайности. Данная работа направлена на построение локальной эталонной динамики индекса NDVI различных с.-х. культур.

Метод построения эталонов

В работе используются данные, входящие в состав Единой федеральной информационной системы о землях сельскохозяйственного назначения (ЕФИС ЗСН) Министерства сельского

хозяйства Российской Федерации, включающей информацию о культурах более чем для 400000 полей в различных регионах страны в период 2017-2018 годов. Первичный анализ этой информации показал, что в данных о выращиваемых культурах имеется значительный уровень погрешностей, а в отдельных районах число полей с неверно указанной культурой может превышать 50%. Это обстоятельство потребовало разработки специального метода автоматической обработки большого объема входных данных о культурах со значительным уровнем погрешностей, и построения по ним эталонов сезонной динамики NDVI сельскохозяйственных культур с последующей фильтрацией некорректных данных.

Для определения достоверности входных данных о культурах на их основе строятся локальные эталоны временной динамики NDVI. Эталон культуры оценивается на уровне административного района по всем полям, засеянным данной культурой. Для всех полей с использованием ЦКП «ИКИ-Мониторинг» [1] был получен восстановленный ряд значений вегетационного индекса NDVI по данным MODIS [2]. Пример вегетационного индекса NDVI всех полей, засеянных озимой пшеницей, на территории одного административного района приведен на рисунке 1а. Рисунок показывает наличие большого количества полей, для которых максимум NDVI достигается в начале августа, что вероятно свидетельствует о том, что они засеяны не озимой пшеницей, а поздними яровыми культурами.

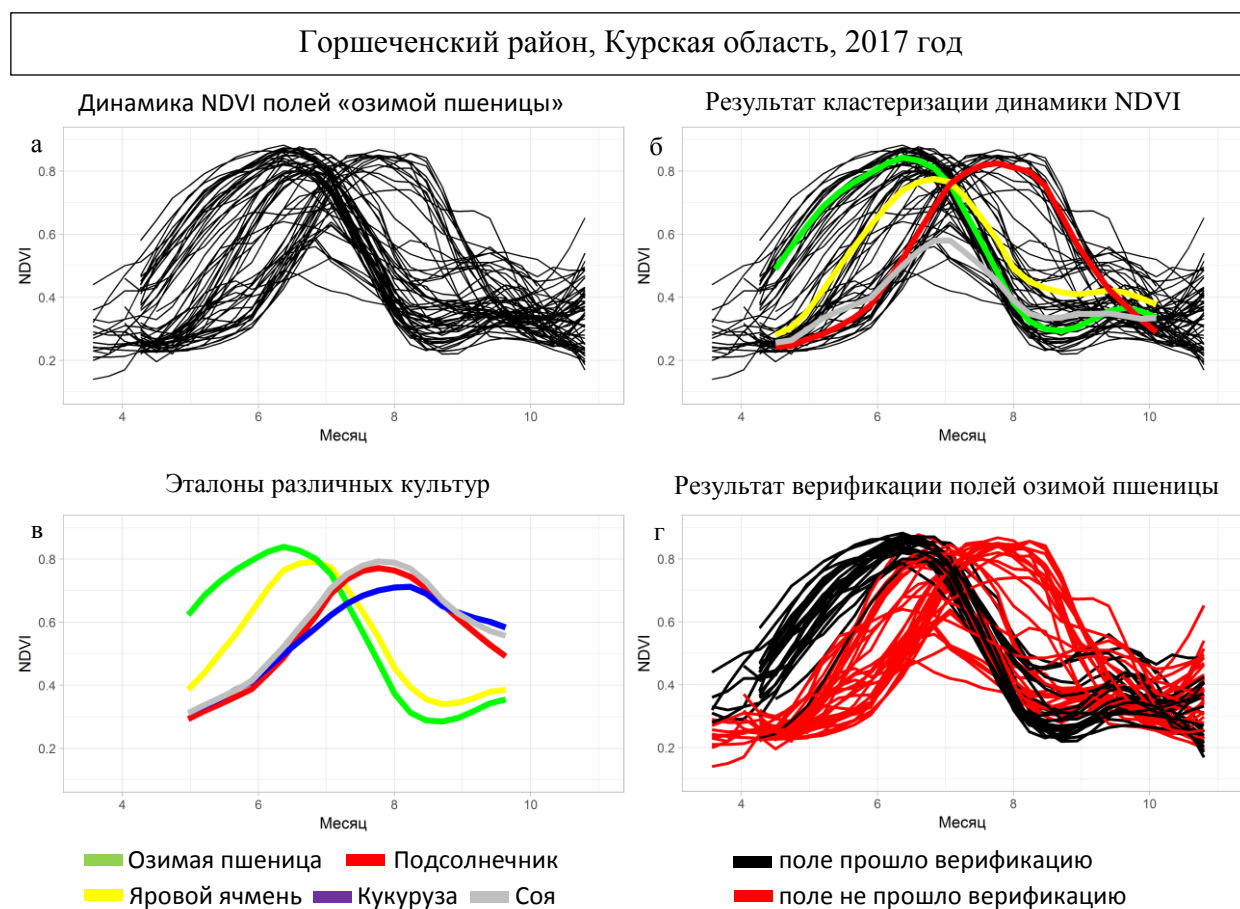


Рисунок 1. Этапы построения эталонов и последующей верификации полей озимой пшеницы на примере выборочного административного района

Все поля одной культуры в заданном районе используются для построения локального эталона сезонной динамики NDVI. В случае, если в районе присутствует небольшое количество полей данной культуры (менее 300), то в анализе также используются и данные соседних районов. Кластеризация временных рядов NDVI позволяет выявить типичные для района варианты динамики вегетационного индекса, часть из которых связана с погрешностью входных

данных, вероятно ассоциированных с другой культурой. Кластеризация выполнялась методом k-средних, а для определения числа кластеров использовался критерий, основанный на минимальном расстоянии между средними всех кластеров, которое не должно было быть меньше 0,1. На рисунке 1б приведен пример кластеризации сезонных рядов NDVI для полей, ассоциированных с посевами озимой пшеницы, на котором видно 4 характерных типа динамики вегетационного индекса: озимые (кластер 1), ранние (кластер 2) и поздние (кластер 3) яровые, а также угнетенные посевы (кластер 4).

При этом считается, что самый большой из полученных кластеров соответствует динамике целевой культуры, и на его основе можно построить ее эталон. Принимается, что эталон задан многомерным нормальным распределением, каждое измерение которого характеризует значение NDVI в определенный момент времени:

$$Etalon = N(\mu, \Sigma), \mu = (\mu_1, \mu_2, \dots, \mu_n), \Sigma = \begin{bmatrix} \Sigma_{11} & \dots & \Sigma_{n1} \\ \dots & \Sigma_{ii} & \dots \\ \Sigma_{1n} & \dots & \Sigma_{nn} \end{bmatrix}$$

где μ_i – среднее значение NDVI в определенную неделю, а Σ_{ij} – ковариация между значениями NDVI в соответствующие недели.

Кроме этого на основе полей данного кластера оценивается их типичный уровень однородности. Для этого для каждого поля рассчитывается вариация индекса NDVI в его границах для каждой недели. Эта вариация усредняется по всем полям и по всем неделям для получения характерного локального значения неоднородности. Также в методе используется понятие неразличимых эталонов, основанное на расстоянии Бхаттачария между ними:

$$D = \frac{1}{8} (\mu_1 - \mu_2) \Sigma^{-1} (\mu_1 - \mu_2)^T + \frac{1}{2} \ln \left(\frac{\det \Sigma}{\det \Sigma_1 \det \Sigma_2} \right)$$

где $\det \Sigma$ – определитель матрицы ковариации, $\Sigma = \Sigma_1 + \Sigma_2$, а $\mu_1, \mu_2, \Sigma_1, \Sigma_2$ – вектора средних и матрицы ковариаций нормальных распределений соответствующих эталонов. Если расстояние между двумя эталонами менее 2,5, то они считаются неразличимыми.

Анализ данных показал, что в отдельных районах недостоверной информации больше, чем достоверной, что приводит к построению некорректных эталонов и необходимости дополнительного сравнения эталонов между районами. Предполагается, что эталоны соседних районов должны быть похожи, или неразличимы, и вводится критерий, по которому если эталон неотличим от не менее, чем 66% полей с данной культурой в соседних районах, то он считается корректным. В противном случае локальный эталон заменяется на другой эталон соседнего района, для которого консенсус в 66% выполняется. Если консенсусный эталон не был найден, то считается, что в окрестности данного района есть несколько характерных вариантов динамики NDVI данной культуры, и локальный эталон остается без изменений. Полученные эталоны могут использоваться для верификации наземных данных.

Метод верификации наземных данных о культуре

Полученные локальные эталоны характеризуют типичную динамику NDVI для данной культуры в заданном административном регионе. Пример локальных эталонов NDVI приведен на рисунке 1в. Они могут использоваться для верификации данных о культуре на поле и фильтрации полей с неверно указанной информацией. В работе предлагается метод верификации данных, основанный на сравнении вегетационного индекса на поле со всем локальными эталонами. Сравнение основано на расстоянии Махаланобиса, вычисляемого по формуле $D = \sqrt{(x - \mu) \Sigma^{-1} (x - \mu)^T}$. Если культура ближайшего к полю эталона соответствует культуре, указанной в наземных данных, то поле проходит верификацию. Также поле проходит

верификацию, если ближайший к нему эталон неотличим от эталона указанной культуры. Дополнительно, для исключения выбросов вводится лимит на максимальное расстояние Махаланобиса, выбранный таким образом, чтобы 95% элементов нормального распределения попадало в его границы и проходили верификацию. Кроме того неоднородные поля также исключаются на этапе верификации. Поле считается неоднородным, если вариация значения NDVI в его границах значительно превышает типичную вариацию, оцененную при построении эталона.

Прошедшие верификацию поля в дальнейшем могут использоваться в качестве обучающей выборки в задачах распознавания и картографирования с.-х. культур, для оценки их состояния и прогнозирования урожайности. Пример результатов верификации с использованием предложенного метода приведен на рисунке 1г. Методика верификации была применена к выборке объемом более 400000 полей с данными о посевах различных культур в 2017-2018 годах. Анализ результатов верификации показал, что 300000 полей прошли верификацию успешно. 20000 полей не прошли верификацию по причине недостатка статистических данных для построения эталона. Также 6000 полей не прошли верификацию по причине их неоднородности, и 80000 полей не прошли верификацию по причине несоответствия динамики их индекса NDVI эталону.

Для выполнения детального анализа точности методики был сформирован выборочный набор полей. Поля выбирались в разрезе пар район-культура, всего было выбрано 240 таких пар (150 для 2017 года и 90 для 2018 года), на которые приходится 37907 полей (9,5% выборки). Для каждого выбранного района и культуры были выведены временные ряды динамики вегетационного индекса NDVI. Для выполнения экспертного анализа цветами были обозначены итоги верификации каждого поля. Также для упрощения экспертного анализа выводилось не более 60 случайно выбранных полей для каждого района. Дополнительно на каждом изображении был приведен график эталонной динамики данной культуры. Экспертный анализ каждого из изображений позволяет определить число полей, динамика которых соответствует или не соответствует ожидаемой динамике для данной культуры. Сравнение результатов экспертного анализа и результатов применения разработанного метода позволяет определить уровень погрешности метода для каждого из исследуемых районов и культур. Экспертный анализ 37907 тестовых полей показал, что для 2992 из них наблюдается ошибка в результатах верификации. Таким образом, точность предложенного метода составила 92%.

Работа выполнена в рамках темы «Мониторинг» (госрегистрация № 01.20.0.2.00164).

References

- [1] Loupian E.A., Proshin A.A., Burtsev M.A., Balashov I.V., Bartalev S.A., Efremov V.Yu., Kashnitskiy A.V., Mazurov A.A., Matveev A.M., Sudneva O.A., Sychugov I.G., Tolpin V.A., Uvarov I.A. 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, 12(5), pp. 263-284. (In Russian).
- [2] Plotnikov D.E., Miklashevich T.S., Bartalev S.A. Using local polynomial approximation within moving window for remote sensing data time-series smoothing and data gaps recovery, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2014, 11(2), pp. 103-110. (In Russian).