

Автоматизированная классификация растительного покрова средиземноморского ландшафта по космическим изображениям высокого пространственного разрешения с учётом спектрально-текстурных и топографических признаков

А. Хатиб, В. А. Малинников

*Московский государственный университет геодезии и картографии
Москва, 105064, Россия
E-mail: syriaheart@live.com*

В работе рассматриваются результаты оценки достоверности автоматизированной классификации растительного покрова средиземноморского ландшафта по алгоритмам случайного леса и максимума правдоподобия с учётом различных спектрально-текстурных и топографических признаков, извлекаемых из многозональных космических изображений высокого пространственного разрешения Landsat (OLI) и цифровой модели рельефа (ASTER GDEM 2). Показано, что из большого количества разнообразных спектрально-текстурных и топографических признаков с помощью алгоритма случайного леса можно выделить набор из 20 наиболее информативных признаков, при использовании которых общая точность классификации растительного покрова увеличивается на 3,7 % по сравнению со случаем использования всех 36 признаков, рассмотренных нами. Общая точность классификации растительного покрова по алгоритмам случайного леса и максимума правдоподобия при использовании набора из 20 наиболее информативных признаков составляет 87,3 и 86,4 % соответственно, а это различие в достоверности классификации имеет статистическую значимость при уровнях значимости 10 и 5 % по проведённому статистическому тесту Мак-Немара на основе распределения хи-квадрат с поправкой Йетса на непрерывность. Вместе с тем оценка минимальной точности распознавания всех выделенных классов растительного покрова показывает, что использование алгоритма случайного леса даёт более достоверные результаты, чем алгоритм максимума правдоподобия. Это демонстрирует эффективность применения алгоритма случайного леса для классификации растительного покрова средиземноморского ландшафта с учётом спектральных и топографических признаков, что согласуется с результатами исследований, проведённых на других территориях средиземноморского региона.

Ключевые слова: автоматизированная классификация, многозональные космические изображения, цифровая модель рельефа, растительный покров, средиземноморский регион, случайный лес, максимум правдоподобия, текстурные признаки, топографические признаки, информативность признаков, тест Мак-Немара, статистическая значимость

Одобрена к печати: 25.02.2021

DOI: 10.21046/2070-7401-2021-18-2-51-63

Введение

Являясь источником биологических ресурсов, растительный покров играет значимую экологическую и социально-культурную роль для человечества (Барталев и др., 2016). По статистическим данным ФАО (Продовольственная и сельскохозяйственная организация ООН, *англ.* FAO — Food and Agriculture Organization, <http://www.fao.org/faostat/en/#data>) за 2017 г., все страны, входящие в 10 ведущих стран-производителей оливок: Испания, Греция, Италия, Турция, Марокко, Египет, Сирия, Тунис, Алжир и Португалия, расположены в средиземноморском регионе. На этот регион приходится 95 % мирового валового сбора оливок. Кроме того, к 10 ведущим странам-производителям цитрусов относятся такие средиземноморские государства, как Сирия и Тунис. Выращивание фруктовых деревьев — источник дохода для многих сельских семей в этом регионе, что ограничивает миграцию населения из сельской местности в города. В дополнение к фруктовым деревьям площадь лесов средиземноморского региона в 2010 г. по данным ФАО составляла более 85 млн га. Хотя их площадь занимает

только 2 % от площади лесов мира, средиземноморские леса помогают в сохранении биоразнообразия и хранении углерода. Они являются источниками древесины, пробки, энергии и продуктов питания (State..., 2013). Таким образом, очевидно, что обеспечение достоверной информации о растительном покрове имеет большую значимость при решении ресурсно-природных задач средиземноморского региона.

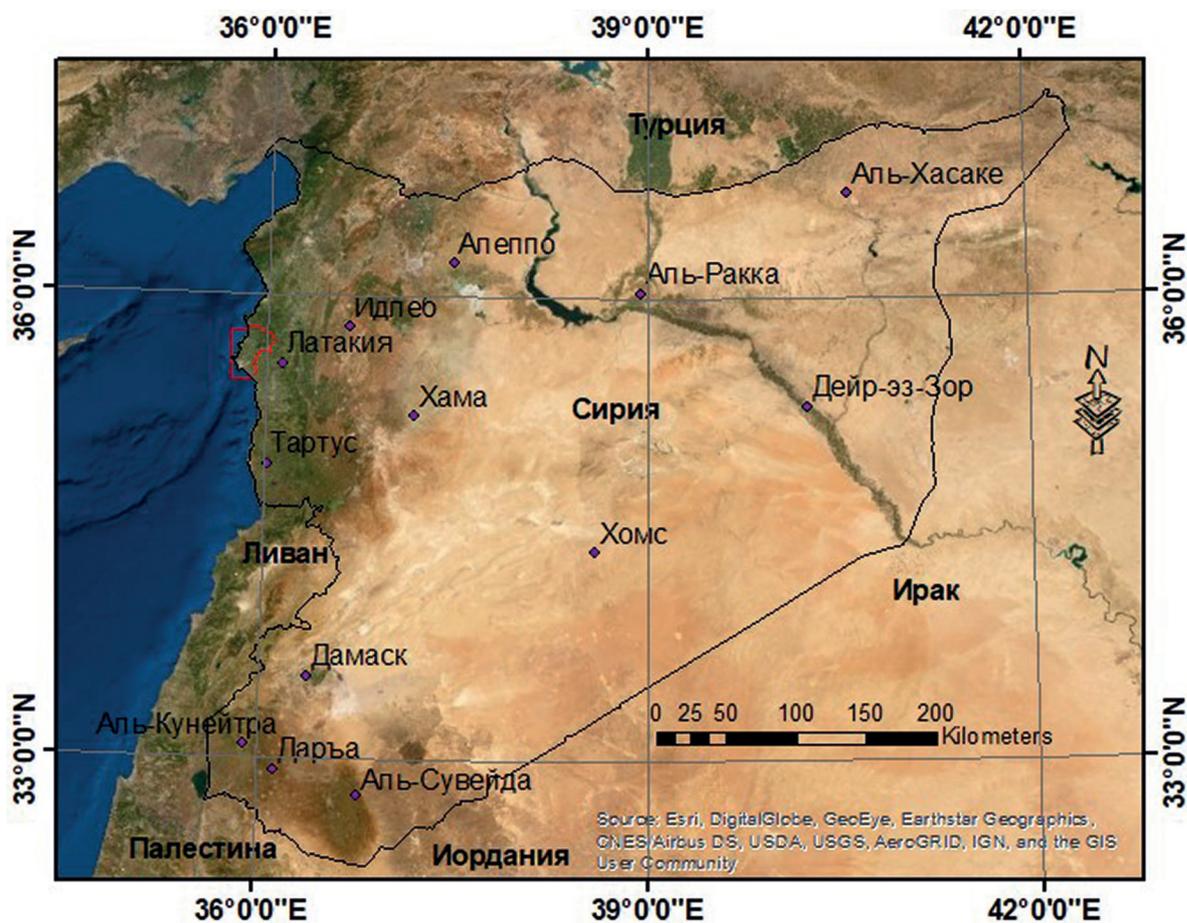
Современные технические возможности систем дистанционного зондирования Земли из космоса позволяют осуществлять наблюдение объектов земной поверхности, в том числе растительности, на всех уровнях (глобальном, региональном и локальном) охвата в зависимости от различных характеристик этих систем (Малинников и др., 2008; Савиных и др., 2000). На сегодняшний день разработка новых методов и способов обработки материалов космической съёмки в целях повышения достоверности классификации растительного покрова становится одной из главных научных задач совершенствования космического картографирования растительного покрова (Барталев, Лупян, 2013).

Несмотря на высокое временное разрешение космических изображений, таких как MODIS (*англ.* Moderate Resolution Imaging Spectroradiometer), уровень пространственного разрешения этих снимков недостаточен для их использования в целях автоматизированной классификации локально-региональных типов растительного покрова (Барталев и др., 2016). Космические изображения высокого пространственного разрешения (как Landsat) использовались для создания карт земной поверхности на глобальном уровне (Chen et al., 2015; Hansen et al., 2013), но их легенда не позволяет обеспечить информацию об экономико-значимых классах растительного покрова, имеющих региональные и/или локальные характеристики. Это, в дополнение к особенностям свободного доступа и качества данных, способствовало широкому использованию космических изображений Landsat при проведении исследований растительного покрова на локальном уровне во всех странах мира.

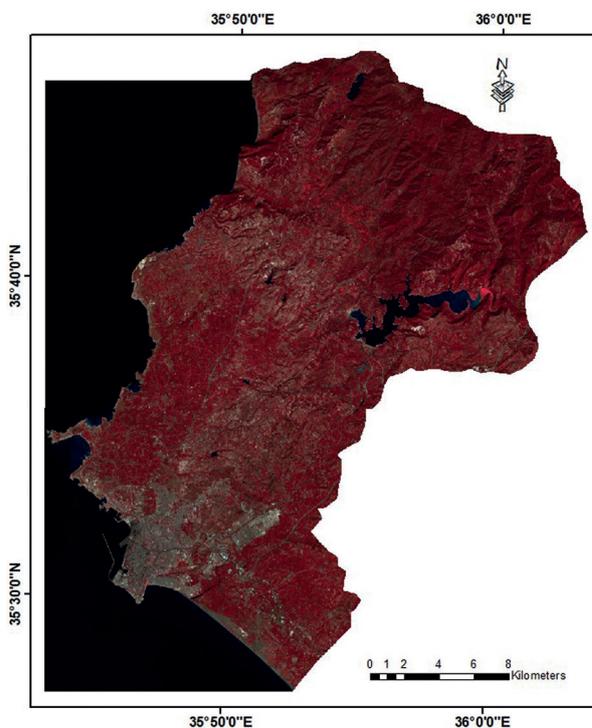
Смешение значений пикселей многозональных космических изображений различных объектов земной поверхности, в том числе типов растительного покрова, приводит к снижению достоверности дешифрирования исходных изображений (Introduction..., 2011). В целях повышения достоверности дешифрирования многозональных снимков Landsat исследователи использовали разнообразные дополнительные спектральные индексные изображения, такие как: вегетационные индексы (Терехин, 2012; Jin et al., 2018), соотношения зональных изображений (Krishna, 2009) и текстурные признаки (Jin et al., 2018; Rodriguez-Galiano et al., 2012). А также были использованы дополнительные географические признаки, например топографические (Гаврилюк и др., 2018; Хатиб, Малинников, 2019; Elumnoh, Shrestha, 2000; Frank, 1988; Franklin, 1987; Janssen et al., 1990), и признаки, полученные из почвенных карт (Corcoran et al., 2013). Поскольку сезонные вариации различных типов растительного покрова будут проявляться в значениях пикселей сезонных изображений, то некоторые исследователи использовали разносезонные изображения (Марчуков, Стыщенко, 2012; Плотников и др., 2018; Стыщенко, 2017). Но одновременное использование всех возможных спектральных и топографических признаков может привести к снижению достоверности дешифрирования космических изображений из-за проблемы, известной как «проклятие размерности» (Belgiu, Dragut, 2016).

В целях исследования достоверности автоматизированной классификации растительного покрова средиземноморского ландшафта по космическим изображениям высокого пространственного разрешения с учётом спектрально-текстурных и топографических признаков нами были поставлены следующие вопросы:

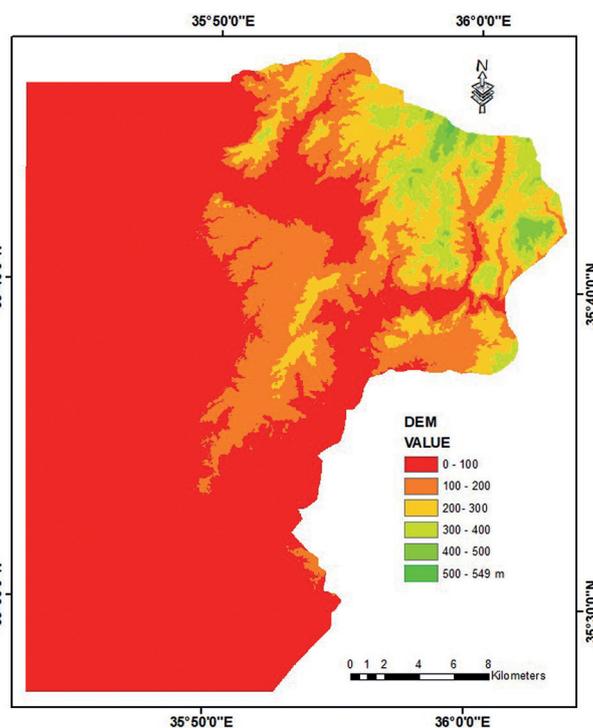
1. Какие из большого количества спектральных и топографических признаков наиболее информативны?
2. Как различаются результаты достоверности классификации растительного покрова, выполняемой при использовании набора из 20 наиболее информативных признаков и набора из всех возможных?
3. Как различаются результаты достоверности классификации растительного покрова по алгоритмам случайного леса и максимума правдоподобия при использовании набора из 20 наиболее информативных признаков?



а



б



в

Рис. 1. Исследуемая территория и исходные информационные данные: а — географическое положение исследуемой территории, выделенной красным полигоном; б — комбинация зональных изображений Landsat (OLI): NIR, R и G; в — цифровая модель рельефа ASTER GDEM

4. Есть ли статистически значимое различие в достоверности классификации растительного покрова по методу случайного леса при использовании набора из 20 наиболее информативных признаков и набора из всех возможных признаков на уровнях значимости 10 и 5 %?
5. Есть ли статистически значимое различие в достоверности классификации растительного покрова по алгоритмам случайного леса и максимума правдоподобия на уровнях значимости 10 и 5 %?

Поиск ответов на поставленные вопросы для исследования возможности обеспечения достоверной информации о растительном покрове сирийского средиземноморского ландшафта — научная задача данной работы.

Информационные компоненты системы классификации растительного покрова

Сирийский средиземноморский регион расположен в северо-западной части страны и характеризуется средиземноморским климатом, т. е. жарким сухим летом и влажной прохладной зимой (Sundseth, 2009). Он состоит из двух провинций: Латакии на севере и Тартуса на юге. *Исследуемая территория* — часть провинции Латакия (рис. 1а, см. с. 53), выбор которой в настоящей работе обусловлен тремя причинами: 1) существование на этой территории основных типов растительного покрова сирийского средиземноморского ландшафта (цитрусовые деревья, оливковые деревья и леса); 2) изменчивость высоты и уклона поверхности земли, что позволяет оценить эффективность использования дополнительных топографических признаков при классификации растительного покрова; 3) наличие для этой территории государственных данных о растительном покрове и качественных космических изображений Google Earth Pro (сервис Google Планета Земля), полученных в период 2017–2018 гг., что позволяет их использование в качестве справочных информационных данных при проведении контролируемых методов классификации.

Исходные информационные данные. Анализ архива многолетних данных космической системы Landsat информационного сервиса EarthExplorer (Геологическая служба США, <https://earthexplorer.usgs.gov>) в Сирийском средиземноморском регионе с 1984 г. показывает, что количество временных рядов с процентом облачности меньше 10 %, которые можно получить ежегодно, составляет примерно 2–5 изображений. Эти временные ряды ежегодно неоднородные, что ограничивает их использование. Одиночные многозональные космические изображения Landsat, полученные осенью, являются наиболее частыми. Кроме того, на наш взгляд, космические изображения, полученные в сухой сезон, позволяют наиболее полно дешифровать основные классы растительного покрова на исследуемой территории. Исходя из этого, в настоящей работе использовались данные Landsat (OLI (англ. Operational Land Imager)) на дату 6 октября 2018 г. (рис. 1б) с уровнем предварительной обработки «2». Это означает, что все геометрические и радиометрические искажения уже скорректированы (Zanter, 2018).

Кроме того, в настоящей работе применялась цифровая модель рельефа ASTER (англ. Advanced Spaceborne Thermal Emission and Reflection Radiometer) GDEM 2 (англ. Global Digital Elevation Model), продукт METI (англ. Ministry of Economy, Trade and Industry, Министерство экономики, торговли и промышленности Японии) и NASA (англ. National Aeronautics and Space Administration, НАСА — Национальное управление по аэронавтике и исследованию космического пространства) с пространственным разрешением 30 м (рис. 1в) на дату 17 октября 2011 г. Предполагается, что высоты поверхности земли на исследуемой территории существенно не изменились с 2011 по 2018 г.

Используемые признаки. В дополнение к оригинальным зональным космическим изображениям Landsat (OLI): синему (B: $\lambda = 0,452–0,512$ мкм); зелёному (G: $\lambda = 0,533–0,590$ мкм); красному (R: $\lambda = 0,636–0,673$ мкм); ближнему инфракрасному (NIR: $\lambda = 0,851–0,879$ мкм); коротковолновым инфракрасным (SWIR₁: $\lambda = 1,566–1,651$ мкм; SWIR₂: $\lambda = 2,107–$

2,294 мкм) — использовались следующие спектрально-текстурные и топографические признаки:

- 5 спектральных вегетационных индексов: нормализованный разностный вегетационный индекс NDVI (*англ.* Normalized Difference Vegetation Index), нормализованный разностный индекс влажности NDMI (*англ.* Normalized Difference Moisture), усовершенствованный вегетационный индекс EVI (*англ.* Enhanced Vegetation Index), почвенный вегетационный индекс SAVI (*англ.* Soil Adjusted Vegetation Index) и модифицированный почвенный вегетационный индекс MSAVI (*англ.* Modified Soil Adjusted Vegetation Index);
- все 15 возможных соотношений используемых оригинальных зональных изображений: B/G, B/R, B/NIR, B/SWIR₁, B/SWIR₂, G/R, G/NIR, G/SWIR₁, G/SWIR₂, R/NIR, R/SWIR₁, R/SWIR₂, NIR/SWIR₁, NIR/SWIR₂, SWIR₁/SWIR₂;
- 8 текстурных признаков, которые рассчитывались на основе первого главного компонента PC₁ с размером окна 3×3 пикселя: гистограммные признаки первого порядка — среднее T_{mean} и дисперсия $T_{variance}$; гистограммные признаки второго порядка: энтропия $T_{entropy}$, различие $T_{dissimilarity}$, второй момент (энергия) $T_{s.moment}$, корреляция $T_{correlation}$, однородность $T_{homogeneity}$, контрастность $T_{contrast}$;
- 2 топографических признака: высота DEM и наклон SLOPE.

Описание расчётных формул вегетационных индексов и используемых в них параметров приведено в информационном сервисе Landsat Surface Reflectance-Derived Spectral Indices (Геологическая служба США, <https://www.usgs.gov/land-resources/nli/landsat/landsat-surface-reflectance-derived-spectral-indices>). Подробная информация об используемых текстурных признаках содержится в публикациях (Чабан, 2016; Haralick et al., 1973).

Классификационная система. На космическом изображении определились следующие тематические классы: цитрусовые деревья, оливковые деревья, леса, травянисто-кустарниковая растительность, другие типы растительности, вода и водонепроницаемые поверхности. Предложенная система классификации была разработана таким образом, чтобы она позволяла обеспечить информацию об основных классах растительного покрова средиземноморского ландшафта (цитрусовые деревья, оливковые деревья и леса).

Создание обучающей выборки. На основе визуального дешифрирования космических изображений Google Earth Pro, полученных в период 2017–2018 гг., создавалась обучающая выборка из 659 тестовых участков небольшого размера (10–25 пикселей). Тестовые участки обучающей выборки распределились в разных частях изображения, чтобы обеспечить хорошую репрезентативность характеристик объектов земного покрова для всех используемых признаков. Количество пикселей обучающей выборки по классам распределилось следующим образом: цитрусовые деревья — 2211; оливковые деревья — 2253; леса — 2126; травяно-кустарниковая растительность — 1047; другие типы растительности — 1255; вода — 2314; водонепроницаемые поверхности — 2243.

Выделение наиболее информативных признаков

Для оценки информативности признаков использовался алгоритм случайного леса — один из наиболее известных ансамблей классификаторов машинного обучения. Этот ансамбль состоит из непараметрических классификаторов (деревьев решения). Каждое дерево вносит один голос при определении заданных классов, а классификация осуществляется путём голосования по всем деревьям (Breiman, 2001). Для выполнения этого алгоритма требуется определение двух параметров: число деревьев и число случайно используемых признаков в каждом дереве (Чистяков, 2013). Число деревьев было принято равным 500, поскольку экспериментально установлено, что значения ошибок стабильны относительно этого числа. Число случайно используемых признаков в каждом дереве было принято считать равным квадратному корню числа всех признаков, поскольку значение этого параметра рекомендуется во многих работах

(Aaron et al., 2018; Belgiu, Dragut, 2016). При выделении наиболее информативных признаков по алгоритму случайного леса в качестве контрольной выборки используется определённый пользователем процент из обучающей выборки. Для обеспечения стабильных результатов в настоящей работе были взяты 10, 20, 30 и 40 % из обучающей выборки.

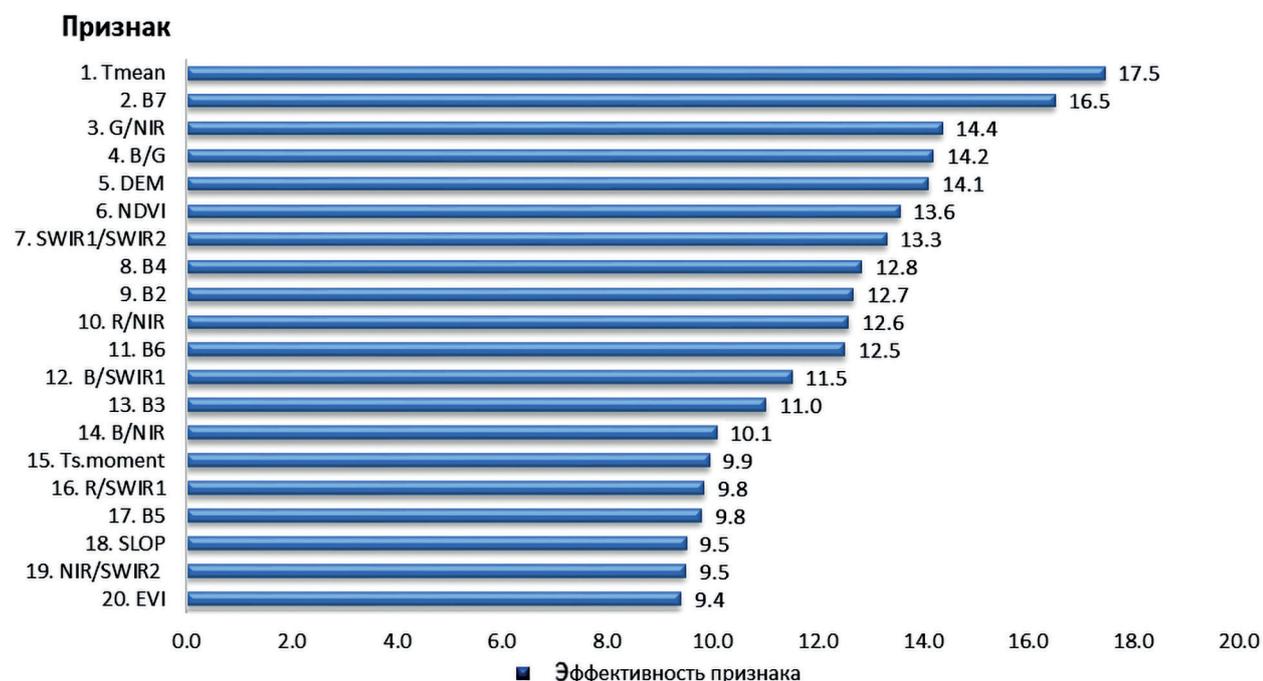


Рис. 2. Эффективность 20 наиболее информативных признаков по методу случайного леса

Эффективность набора из 20 наиболее информативных признаков по методу случайного леса представлена на рис. 2. В этот набор входят: все оригинальные зональные изображения — B, G, R, NIR, SWIR₁ и SWIR₂; спектральные вегетационные индексы — NDVI и EVI; соотношения оригинальных зональных изображений — B/G, B/NIR, B/SWIR₁, G/NIR, R/NIR, R/SWIR₁, NIR/SWIR₂ и SWIR₁/SWIR₂; текстурные признаки — T_{mean} и $T_{s.moment}$; топографические признаки — DEM и SLOPE.

Автоматизированная классификация растительного покрова

Для автоматизированной классификации растительного покрова в настоящей работе использовались следующие контролируемые пиксельные алгоритмы: непараметрический алгоритм случайного леса RF (*англ.* random forest) и параметрический алгоритм максимума правдоподобия ML (*англ.* maximum likelihood). Если мы хотим поговорить о значимости этих двух алгоритмов, может быть достаточным сказать, что они используются для распознавания растительного покрова по космическим данным в программном комплексе LAGMA-PLUS (*англ.* Locally Adaptive Global Mapping Algorithm), разработанном Институтом космических исследований РАН (ИКИ РАН) (Барталев и др., 2016).

Поскольку используемые алгоритмы — пиксельные, были выполнены постклассификационные процедуры для сглаживания результатов классификации путём отнесения мелких изолированных регионов (меньше чем 4 пикселя) к ближайшему классу. Результаты классификации растительного покрова на космических изображениях по алгоритму случайного леса с использованием набора из всех 36 признаков RF (36) и по алгоритмам случайного леса и максимума правдоподобия с использованием набора из 20 наиболее информативных признаков RF (20) и ML (20) представлены на рис. 3 (см. с. 57).

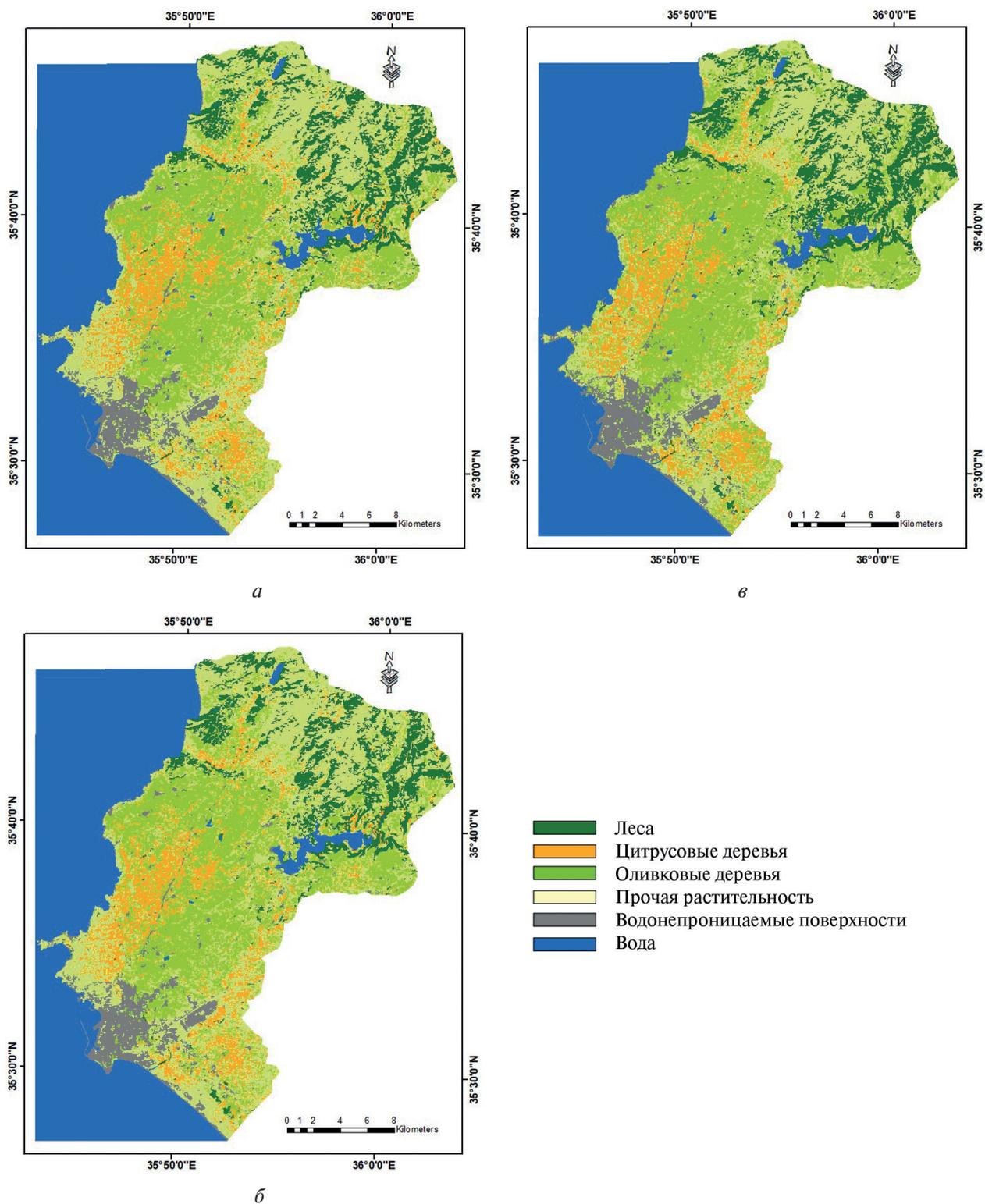


Рис. 3. Результаты классификации растительного покрова с помощью различных методов: *a* — RF (20); *b* — RF (36); *v* — ML (20)

Оценка достоверности классификации растительного покрова

В качестве справочных данных для оценки достоверности результатов классификации растительного покрова использовались космические изображения Google Earth Pro. Сначала создавалось 350 контрольных точек, случайно распределённых на классифицируемых изображениях,

в целях обеспечения около 50 контрольных точек для каждого класса (Congalton, 1991). Далее классы травяно-кустарниковой растительности и других типов растительности были объединены под названием «прочая растительность». Во избежание предвзятости результатов достоверности дешифрирования исходного изображения использовались только точки, расположенные на классах растительного покрова, поскольку большинство водных объектов и водонепроницаемых поверхностей классифицируются с высокой точностью. А также исключились точки, расположенные на границах между различными классами. На основе оставшихся контрольных точек были построены матрицы ошибок и рассчитаны показатели достоверности классификации по следующим формулам (Congalton, 1991):

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r x_{i+} x_{+i}}{N^2 \sum_{i=1}^r x_{i+} x_{+i}}, \quad p_{\text{общ}} = \sum_{i=1}^r x_{ii} / N, \quad p_{\text{пол}} = \sum_{i=1}^r x_{ii} / x_{i+}, \quad p_{\text{про}} = \sum_{i=1}^r x_{ii} / x_{+i},$$

где k — коэффициент Каппа; $p_{\text{общ}}$ — общая точность; $p_{\text{пол}}$ — точность пользователя i -го класса; $p_{\text{про}}$ — точность производителя i -го класса; N — общее число контрольных точек; r — число классов; x_{ii} — число контрольных точек в строке i и столбце i ; x_{i+} , x_{+i} — сумма контрольных точек в строке i и столбце i соответственно.

Построенные матрицы ошибок и результаты достоверности классификации растительного покрова представлены в *табл. 1*. Результаты оценки общей точности и коэффициента Каппа показывают, что использование набора из 20 наиболее информативных признаков даёт более достоверные результаты классификации растительного покрова по алгоритму случайного леса, чем использование набора из 36 признаков. Также они показывают, что при использовании набора из 20 наиболее информативных признаков алгоритмы случайного леса и максимума правдоподобия не дают большого различия в общей точности классификации растительного покрова.

Таблица 1. Сравнение результатов достоверности классификации растительного покрова

Классифицированные данные	Алгоритм (набор признаков)	Справочные данные				
		Оливковые деревья	Цитрусовые деревья	Леса	Прочая растительность	Точность пользователя, %
Оливковые деревья	RF (36)	65	0	0	9	88
	RF (20)	67	0	0	7	91
	ML (20)	67	1	0	12	84
Цитрусовые деревья	RF (36)	0	33	1	5	85
	RF (20)	0	36	1	4	88
	ML (20)	0	33	0	0	100
Леса	RF (36)	1	0	30	3	88
	RF (20)	1	0	30	3	88
	ML (20)	1	0	34	5	85
Прочая растительность	RF (36)	4	9	3	50	76
	RF (20)	2	6	3	53	83
	ML (20)	2	8	0	50	83
Точность производителя, %	RF (36)	93	79	88	75	
	RF (20)	96	86	88	79	
	ML (20)	96	79	100	75	

Общая точность: RF (36) = 83,6 %; RF (20) = 87,3 %; ML (20) = 86,4 %.

Коэффициент Каппа: RF (36) = 0,77; RF (20) = 0,83; ML (20) = 0,81.

Из *табл. 1* видно, что для класса оливковых деревьев наилучшая минимальная точность (минимальное значение между точностью производителя и пользователя) составляет 91 % при использовании RF (20). Для класса цитрусовых деревьев — 86 % при использовании RF (20). Для класса лесов — 88 % при использовании RF (20) и RF (36). Для класса прочей растительности — 79 % при использовании RF (20). Эти результаты показывают, что использование набора из 20 наиболее информативных признаков по алгоритму случайного леса даёт более достоверные результаты распознавания всех выделенных классов растительного покрова, чем по алгоритму максимума правдоподобия.

Оценка различия в достоверности классификации растительного покрова

Для оценки рассчитанных различий в достоверности классификации растительного покрова были построены матрицы сопряжённости размером 2×2 и проведён статистический тест Мак-Немара на основе распределения хи-квадрат с поправкой Йетса на непрерывность

по формуле (Foody, 2004): $\chi^2 = \frac{(|f_{12} - f_{21}| - 1)^2}{f_{12} + f_{21}}$, где f_{12} — число контрольных точек, которые

правильно классифицируются при использовании первого алгоритма и неправильно — при использовании второго; f_{21} — число контрольных точек, которые неправильно классифицируются при использовании первого алгоритма и правильно — при использовании второго.

Построенные матрицы сопряжённости и результаты статистического теста представлены в *табл. 2*. Из неё видно, что рассчитанное значение ($\chi^2 = 4,9$) также больше, чем критические табулированные значения по степени свободы $df = 1$ при уровнях значимости 10 % ($\chi^2_{\alpha = 10\%} = 2,71$) и 5 % ($\chi^2_{\alpha = 5\%} = 3,84$). Таким образом, тест Мак-Немара на основе распределения хи-квадрат с поправкой Йетса на непрерывность показывает, что действительно существует статистически значимое различие в достоверности классификации растительного покрова при использовании набора из 20 наиболее информативных признаков и набора из всех 36 признаков при заданных уровнях.

Таблица 2. Различия в достоверности классификации растительного покрова

RF (20)	RF (36)			ML (20)		
	f_1	f_2	χ^2	f_1	f_2	χ^2
f_1	177	9	4,9	154	32	7,2
f_2	1	26		13	14	

Аналогичным образом проведённый тест показывает, что действительно существует статистически значимое различие в достоверности классификации растительного покрова по алгоритмам случайного леса и максимума правдоподобия при использовании набора из 20 наиболее информативных признаков при уровнях значимости 10 и 5%, поскольку рассчитанные значения ($\chi^2 = 7,2$) больше, чем критические табулированные значения.

Заключение

Анализ результатов проведённого исследования, посвящённого автоматизированной классификации растительного покрова средиземноморского ландшафта по космическим изображениям на исследуемой территории с учётом спектрально-текстурных и топографических признаков, позволяет сделать следующие выводы:

1. Из большого количества разнообразных спектральных и топографических признаков по алгоритму случайного леса можно выделить набор из 20 наиболее информативных

признаков, позволяющих классифицировать растительный покров на исследуемой территории с высокой достоверностью: общая точность — 87,3 % и коэффициент Каппа — 0,83.

2. При использовании набора из 20 наиболее информативных признаков общая точность классификации растительного покрова по алгоритму случайного леса увеличивается на 3,7 % по сравнению со случаем использования всех 36 признаков, рассмотренных нами. Статистический тест Мак-Немара на основе распределения хи-квадрат с поправкой Йетса на непрерывность показывает, что рассчитанное различие в достоверности классификации растительного покрова имеет статистическую значимость при уровнях значимости 10 и 5 %. Это наглядно указывает на актуальность выделения наиболее информативных признаков при классификации растительного покрова по космическим изображениям именно в случае использования большого количества дополнительных спектральных признаков, которые не являются независимыми.
3. Хотя использование алгоритмов случайного леса и максимума правдоподобия не даёт большого различия в общей точности классификации растительного покрова: 87,3 и 86,4 % соответственно, результаты статистического теста Мак-Немара на основе распределения хи-квадрат с поправкой Йетса на непрерывность показывают, что рассчитанное различие в достоверности классификации имеет статистическую значимость при уровнях значимости 10 и 5 %. Вместе с тем минимальная точность классификации всех выделенных классов растительного покрова показывает, что использование алгоритма случайного леса даёт более достоверные результаты, чем алгоритм максимума правдоподобия. Это демонстрирует эффективность применения алгоритма случайного леса для классификации растительного покрова средиземноморского ландшафта с учётом спектральных и топографических признаков, что согласуется с результатами исследований (Gounaridis et al., 2016; Rodriguez-Galiano et al., 2012), которые проводились на других территориях средиземноморского региона.

Литература

1. *Барталев С. А., Лупян Е. А.* Исследования и разработки ИКИ РАН по развитию методов спутникового мониторинга растительного покрова // *Современные проблемы дистанционного зондирования Земли из космоса*. 2013. Т. 10. № 1. С. 197–214.
2. *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.
3. *Гаврилюк Е. А., Плотникова А. С., Плотников Д. Е.* Картографирование наземных экосистем Печоро-Илычского заповедника и его окрестностей на основе восстановленных мультивременных спутниковых данных Landsat // *Современные проблемы дистанционного зондирования Земли из космоса*. 2018. Т. 15. № 5. С. 141–153.
4. *Малинников В. А., Стеценко А. Ф., Алтынов А. Е., Попов С. М.* Мониторинг природной среды аэрокосмическими средствами: учеб. пособие. М.: МИИГАиК, 2008. 145 с.
5. *Марчуков В. С., Стыценко Е. А.* Дешифрирование растительного покрова с использованием спектрально-временных признаков // *Исслед. Земли из космоса*. 2012. № 1. С. 77–88.
6. *Плотников Д. Е., Колбудаев П. А., Барталев С. А., Лупян Е. А.* Автоматическое распознавание используемых пахотных земель на основе сезонных временных серий восстановленных изображений Landsat // *Современные проблемы дистанционного зондирования Земли из космоса*. 2018. Т. 15. № 2. С. 112–127.
7. *Савиных В. П., Малинников В. А., Цыпина Э. М., Сладкопевцев С. А.* География из космоса. М.: МИИГАиК, 2000. 244 с.
8. *Стыценко Е. А.* Возможности распознавания сельскохозяйственных угодий с использованием методики совместной автоматизированной обработки разносезонных многозональных космических изображений // *Современные проблемы дистанционного зондирования Земли из космоса*. 2017. Т. 14. № 5. С. 172–183.
9. *Терехин Э. А.* Информативность спектральных вегетационных индексов для дешифрирования сельскохозяйственной растительности // *Современные проблемы дистанционного зондирования Земли из космоса*. 2012. Т. 9. № 4. С. 243–248.

10. Хатиб А., Малинников В. А. Дешифрирование растительного покрова Средиземноморского региона на многозональных космических изображениях с учетом данных цифровой модели рельефа // Изв. высш. учеб. заведений. Геодезия и аэрофотосъемка. 2019. Т. 63. № 4. С. 432–439.
11. Чабан Л. Н. Методы и алгоритмы распознавания образов в автоматизированном дешифрировании данных дистанционного зондирования. М.: МИИГАиК, 2016. 77 с.
12. Чистяков С. П. Случайные леса: обзор // Тр. Карельского науч. центра РАН. 2013. № 1. С. 117–136.
13. Aaron E., Timothy A., Fang F. Implementation of machine-learning classification in remote sensing: an applied review // Intern. J. Remote Sensing. 2018. V. 39. No. 9. P. 2784–2817.
14. Belgiu M., Dragut L. Random forest in remote sensing: A review of applications and future directions // ISPRS J. Photogrammetry and Remote Sensing. 2016. V. 114. P. 24–31.
15. Breiman L. Random Forests // Machine Learning. 2001. V. 45. P. 5–32.
16. Introduction to Remote Sensing. Fifth Edition / ed. Campbell J., Wynne R. N.Y.: The Guilford Press, 2011. 662 p.
17. Chen J., Chen Jin, Liao A., Cao X., Chen L., Chen X., He C., Han G., Peng S., Lu M., Zhang W., Tong X., Mills J. Global land cover mapping at 30 m resolution: A POK-based operational approach // ISPRS J. Photogrammetry and Remote Sensing. 2015. V. 103. P. 7–27.
18. Congalton R. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data // Remote Sensing of Environment. 1991. V. 37. No. 1. P. 35–46.
19. Corcoran J., Knight J., Gallant A. Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota // Remote Sensing. 2013. V. 5. No. 7. P. 3212–3238.
20. Elumnoh A., Shrestha R. Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand // Photogrammetric Engineering and Remote Sensing. 2000. V. 66. No. 3. P. 297–304.
21. Foody G. Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy // Photogrammetric Engineering and Remote Sensing. 2004. V. 70. No. 5. P. 627–633.
22. Frank T. Mapping Dominant Vegetation Communities in the Colorado Rocky Mountain Front Range with Landsat Thematic Mapper and Digital Terrain Data // Photogrammetric Engineering and Remote Sensing. 1988. V. 54. No. 12. P. 1727–1734.
23. Franklin S. Terrain Analysis from Digital Patterns in Geomorphometry and Landsat MSS Spectral Response // Photogrammetric Engineering and Remote Sensing. 1987. V. 53. No. 1. P. 59–65.
24. Gounaridis D., Apostolou A., Koukoulas S. Land cover of Greece, 2010: a semi-automated classification using random forests // J. Maps. 2016. V. 12. No. 5. P. 1055–1062.
25. Hansen M., Potapov P., Moore R., Hancher M., Turubanova S., Tyukavina A., Thau D., Stehman S., Goetz S., Loveland T., Kommareddy A., Egorov A., Chini L., Justice C., Townshend J. High-Resolution Global Maps of 21st-Century Forest Cover Change // Science. 2013. V. 342. No. 6160. P. 850–853.
26. Haralick R., Shanmugam K., Dinstein I. Textural features for image classification // IEEE Trans. Systems, Man, and Cybernetics. 1973. V. 3. No. 6. P. 610–621.
27. Janssen L., Jaarsma M., van der Linden E. Integrating Topographic Data with Remote Sensing for Land-Cover Classification // Photogrammetric Engineering and Remote Sensing. 1990. V. 56. No. 11. P. 1503–1506.
28. Jin Y., Liu X., Chen Y., Liang X. Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong // Intern. J. Remote Sensing. 2018. V. 39. No. 23. P. 1–21.
29. Krishna B. Improving Landsat and IRS Image Classification: Evaluation of Unsupervised and Supervised Classification through Band Ratios and DEM in a Mountainous Landscape in Nepal // Remote Sensing. 2009. V. 1. No. 4. P. 1257–1272.
30. Rodriguez-Galiano V., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J. An assessment of the effectiveness of a random forest classifier for land-cover classification // ISPRS J. Photogrammetry and Remote Sensing. 2012. V. 67. P. 93–104.
31. State of Mediterranean Forests. FAO, 2013. 173 p.
32. Sundseth K. Natura 2000 in the Mediterranean Region. Luxembourg: Office for Official Publications of the European Communities, 2009. 12 p.
33. Zanter K. Landsat 8 Surface Reflectance Code (LASRC) Product Guide. Department of the Interior U.S., 2018. 34 p.

Automated classification of the vegetation cover of Mediterranean landscape using spectral-textural and topographic features of high spatial resolution satellite imagery

A. Khatib, V. A. Malinnikov

Moscow State University of Geodesy and Cartography, Moscow 105064, Russia
E-mail: syriaheart@live.com

The paper considers the results of assessing the reliability of the automated classification of Mediterranean vegetation by random forest and maximum likelihood algorithms using several spectral, texture and topographic features extracted from high spatial resolution multispectral satellite images Landsat (OLI) and digital elevation model (ASTER GDEM 2). The study shows that from a large number of various spectral-textural and topographic features, it is possible to select, with the random forest algorithm, a set of 20 most informative features the use of which increases the overall classification accuracy of vegetation cover by 3.7 % compared to the use of all 36 features. The overall accuracy of vegetation cover classification by random forest and maximum likelihood algorithms using a set of 20 most informative features is 87.3% and 86.4 %, respectively, and this difference in classification reliability is statistically significant at 10% and 5 % significance levels according to the McNemar statistical test based on the Chi-square distribution with Yates' correction for continuity. At the same time, the estimate of the minimum classification accuracy of all the identified classes of vegetation cover shows that the use of the random forest algorithm gives more reliable results than the maximum likelihood algorithm. This demonstrates the effectiveness of using the random forest algorithm to classify the Mediterranean vegetation, taking into account spectral and topographic features, which is consistent with the results of studies carried out in other territories of the Mediterranean region.

Keywords: automated classification, multispectral space images, digital elevation model, vegetation cover, Mediterranean region, random forest, maximum likelihood, textural features, topographic features, feature importance, McNemar's test, statistical significance

Accepted: 25.02.2021

DOI: 10.21046/2070-7401-2021-18-2-51-63

References

1. Bartalev S. A., Loupian E. A., R&D on methods for satellite monitoring of vegetation by the Russian Academy of Sciences' Space Research Institute, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2013, Vol. 10, No. 1, pp. 197–214 (in Russian).
2. Bartalev S. A., Egorov V. A., Zharko V. O., Loupian E. A., Plotnikov D. E., Khvostikov S. A., Shabanov N. V., *Sputnikovoe kartografirovaniye rastitel'nogo pokrova Rossii* (Land cover mapping over Russia using Earth observation data), Moscow: IKI RAN, 2016, 208 p. (in Russian).
3. Gavrilyuk E. A., Plotnikova A. S., Plotnikov D. E., Land cover mapping of the Pechora-Ilych Nature Reserve and its vicinity based on reconstructed multitemporal Landsat satellite data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2018, Vol. 15, No. 5, pp. 141–153 (in Russian).
4. Malinnikov V. A., Stetsenko A. F., Altynov A. E., Popov S. M., *Monitoring prirodnoi sredy aerokosmicheskimi sredstvami* (Monitoring the natural environment by aerospace tools), Moscow: MIIGAİK, 2008, 145 p. (in Russian).
5. Marchukov V. S., Stytsenko E. A., Interpretation of vegetation using spectral-temporal characteristics, *Issledovanie Zemli iz kosmosa*, 2012, No. 1, pp. 77–88 (in Russian).
6. Plotnikov D. E., Kolbudaev P. A., Bartalev S. A., Loupian E. A., Automated annual cropland mapping from reconstructed time series of Landsat data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2018, Vol. 15, No. 2, pp. 112–127 (in Russian).
7. Savinykh V. P., Malinnikov V. A., Tsykina E. M., Sladkoptsev S. A., *Geografiya iz kosmosa* (Geography from space), Moscow: MIIGAİK, 2000, 244 p. (in Russian).
8. Stytsenko E. A., Evaluation of the possibilities to classify agricultural lands using multi-seasonal satellite data processing, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2017, Vol. 14, No. 5, pp. 172–183 (in Russian).

9. Terekhin E.A., The efficiency of spectral vegetation indices for interpretation agricultural vegetation, *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2012, Vol. 9, No. 4, pp. 243–248 (in Russian).
10. Khatib A., Malinnikov V.A., Interpretation of Mediterranean vegetation in multi-channel space images using digital elevation model data, *Izvestiya vuzov "Geodesy and Aerophotosurveying"*, 2019, Vol. 63, No. 4, pp. 432–439 (in Russian).
11. Chaban L.N., *Metody i algoritmy raspoznavaniya obrazov v avtomatizirovannom deshifirovanii dannykh distantsionnogo zondirovaniya* (Pattern recognition methods and algorithms in automated decrypting of remote sensing data), Moscow: MIIGAİK, 2016, 77 p. (in Russian).
12. Chistyakov S.P., Random forests: an overview, *Trudy Karel'skogo nauchnogo tsentra RAN*, 2013, No. 1, pp. 117–136 (in Russian).
13. Aaron E., Timothy A., Fang F., Implementation of machine-learning classification in remote sensing: an applied review, *Intern. J. Remote Sensing*, 2018, Vol. 39, No. 9, pp. 2784–2817.
14. Belgiu M., Dragut L., Random forest in remote sensing: A review of applications and future directions, *ISPRS J. Photogrammetry and Remote Sensing*, 2016, Vol. 114, pp. 24–31.
15. Breiman L., Random Forests, *Machine Learning*, 2001, Vol. 45, pp. 5–32.
16. Campbell J., Wynne R., *Introduction to Remote Sensing, Fifth Edition*, New York: The Guilford Press, 2011, 662 p.
17. Chen J., Chen Jin, Liao A., Cao X., Chen L., Chen X., He C., Han G., Peng S., Lu M., Zhang W., Tong X., Mills J., Global land cover mapping at 30 m resolution: A POK-based operational approach, *ISPRS J. Photogrammetry and Remote Sensing*, 2015, Vol. 103, pp. 7–27.
18. Congalton R., A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, *Remote Sensing of Environment*, 1991, Vol. 37, No. 1, pp. 35–46.
19. Corcoran J., Knight J., Gallant A., Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota, *Remote Sensing*, 2013, Vol. 5, No. 7, pp. 3212–3238.
20. Elumnoh A., Shrestha R., Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand, *Photogrammetric Engineering and Remote Sensing*, 2000, Vol. 66, No. 3, pp. 297–304.
21. Foody G., Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy, *Photogrammetric Engineering and Remote Sensing*, 2004, Vol. 70, No. 5, pp. 627–633.
22. Frank T., Mapping Dominant Vegetation Communities in the Colorado Rocky Mountain Front Range with Landsat Thematic Mapper and Digital Terrain Data, *Photogrammetric Engineering and Remote Sensing*, 1988, Vol. 54, No. 12, pp. 1727–1734.
23. Franklin S., Terrain Analysis from Digital Patterns in Geomorphometry and Landsat MSS Spectral Response, *Photogrammetric Engineering and Remote Sensing*, 1987, Vol. 53, No. 1, pp. 59–65.
24. Gounaridis D., Apostolou A., Koukoulas S., Land cover of Greece, 2010: a semi-automated classification using random forests, *J. Maps*, 2016, Vol. 12, No. 5, pp. 1055–1062.
25. Hansen M., Potapov P., Moore R., Hancher M., Turubanova S., Tyukavina A., Thau D., Stehman S., Goetz S., Loveland T., Kommareddy A., Egorov A., Chini L., Justice C., Townshend J., High-Resolution Global Maps of 21st-Century Forest Cover Change, *Science*, 2013, Vol. 342, No. 6160, pp. 850–853.
26. Haralick R., Shanmugam K., Dinstein I., Textural features for image classification, *IEEE Trans. Systems, Man, and Cybernetics*, 1973, Vol. 3, No. 6, pp. 610–621.
27. Janssen L., Jaarsma M., van der Linden E., Integrating Topographic Data with Remote Sensing for Land-Cover Classification, *Photogrammetric Engineering and Remote Sensing*, 1990, Vol. 56, No. 11, pp. 1503–1506.
28. Jin Y., Liu X., Chen Y., Liang X., Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong, *Intern. J. Remote Sensing*, 2018, Vol. 39, No. 23, pp. 1–21.
29. Krishna B., Improving Landsat and IRS Image Classification: Evaluation of Unsupervised and Supervised Classification through Band Ratios and DEM in a Mountainous Landscape in Nepal, *Remote Sensing*, 2009, Vol. 1, No. 4, pp. 1257–1272.
30. Rodriguez-Galiano V., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J., An assessment of the effectiveness of a random forest classifier for land-cover classification, *ISPRS J. Photogrammetry and Remote Sensing*, 2012, Vol. 67, pp. 93–104.
31. *State of Mediterranean Forests*, FAO, 2013, 173 p.
32. Sundseth K., *Natura 2000 in the Mediterranean Region*, Luxembourg: Office for Official Publications of the European Communities, 2009, 12 p.
33. Zanter K., *Landsat 8 Surface Reflectance Code (LASRC) Product Guide*, Department of the Interior U.S., 2018, 34 p.