

Технология создания безоблачных композитных изображений по данным спутников серии Sentinel-2

А. В. Кашницкий, М. А. Бурцев, А. А. Прошин

Институт космических исследований РАН, Москва, 117997, Россия

E-mail: kashnizky@gmail.com

В работе описывается технология получения безоблачных композитных изображений по данным спутников серии Sentinel-2, а также полученный с её помощью набор композитных изображений территории России за разные периоды. В основе технологии лежит маскирование облаков и их теней с помощью доработанного продукта классификации сцены алгоритмом Sen2Cor. При создании композитного изображения результирующее значение пикселя получается расчётом медианы по всем чистым измерениям за период, причём число участвующих в расчёте сцен подбирается в зависимости от конкретной ситуации. В результате с помощью описываемой технологии были созданы ежемесячные (с марта по ноябрь) и ежесезонные (весна, лето, осень, сезон вегетации, снежный период) безоблачные композитные изображения в шести спектральных каналах по данным спутников серии Sentinel-2 на территорию России за 2019–2022 гг. Применение созданных композитных изображений показало их преимущество при решении ряда задач по сравнению с использованием отдельных сцен. Набор композитных изображений территории России за разные периоды опубликован на ресурсах Центра коллективного пользования «ИКИ-Мониторинг» и предлагается настоящей работой для решения научных задач широкому кругу исследователей.

Ключевые слова: Sentinel-2, композитные изображения, детектирование облачности, безоблачные ежемесячные данные, зимний композит, ДЗЗ

Одобрена к печати: 31.10.2022

DOI: 10.21046/2070-7401-2022-19-5-76-85

Введение

Для изучения изменений, происходящих на поверхности Земли, во многих случаях необходимо получать регулярные и сопоставимые спутниковые данные дистанционного зондирования Земли (ДЗЗ). Особенно остро стоит вопрос получения периодической съёмки в задачах изучения растительности. Например, для получения карт растительного покрова или анализа его изменений необходима информация о динамике отражательной способности в различных диапазонах электромагнитного спектра каждой точки картируемой территории в течение сезона вегетации (Барталев и др., 2016). Надёжное регулярное получение такой информации по спутниковым данным может быть затруднено, в основном из-за влияния облачности. При этом для разработки автоматических методов анализа на большую территорию более применимыми могут быть полностью её покрывающие однородные бесшовные данные, в которых совсем отсутствует облачность и связанные с ней мешающие факторы. Одним из основных путей получения таких данных представляется создание безоблачных композитных изображений.

В работе описывается технология получения безоблачных бесшовных композитных изображений по данным спутников серии Sentinel-2, а также полученный на её основе набор ежемесячных и ежесезонных данных за разные периоды 2019–2022 гг. Этот набор данных был опубликован на ресурсах Центра коллективного пользования (ЦКП) «ИКИ-Мониторинг» (Лупян и др., 2019).

Исходные данные

Исходные данные прибора MSI (*англ.* Multispectral Instrument) спутников серии Sentinel-2, распространяемые в рамках программы Copernicus (*англ.* Copernicus Open Access Hub, <https://scihub.copernicus.eu/>) Европейского космического агентства (*англ.* European Space Agency — ESA) в настоящей работе для создания композитных изображений извлекаются из архивов ЦКП ИКИ-Мониторинг. Для этого ранее было организовано автоматическое поступление, обработка и сохранение данных спутников серии Sentinel-2 в архивы ЦКП «ИКИ-Мониторинг» (Прошин и др., 2019). В настоящее время в архивах ЦКП «ИКИ-Мониторинг» имеется информация спутников серии Sentinel-2 по территории Северной Евразии и ряду локальных участков по всему миру за период с 2015 г. по настоящее время. Пример покрытия данными (продукты всех уровней обработки) на лето 2019 г. приведён на *рис. 1*. Набор продуктов уровня обработки L2A, т.е. после проведения процедуры атмосферной коррекции, имеется в ЦКП «ИКИ-Мониторинг» только начиная с 2019 г. и только на территорию России.

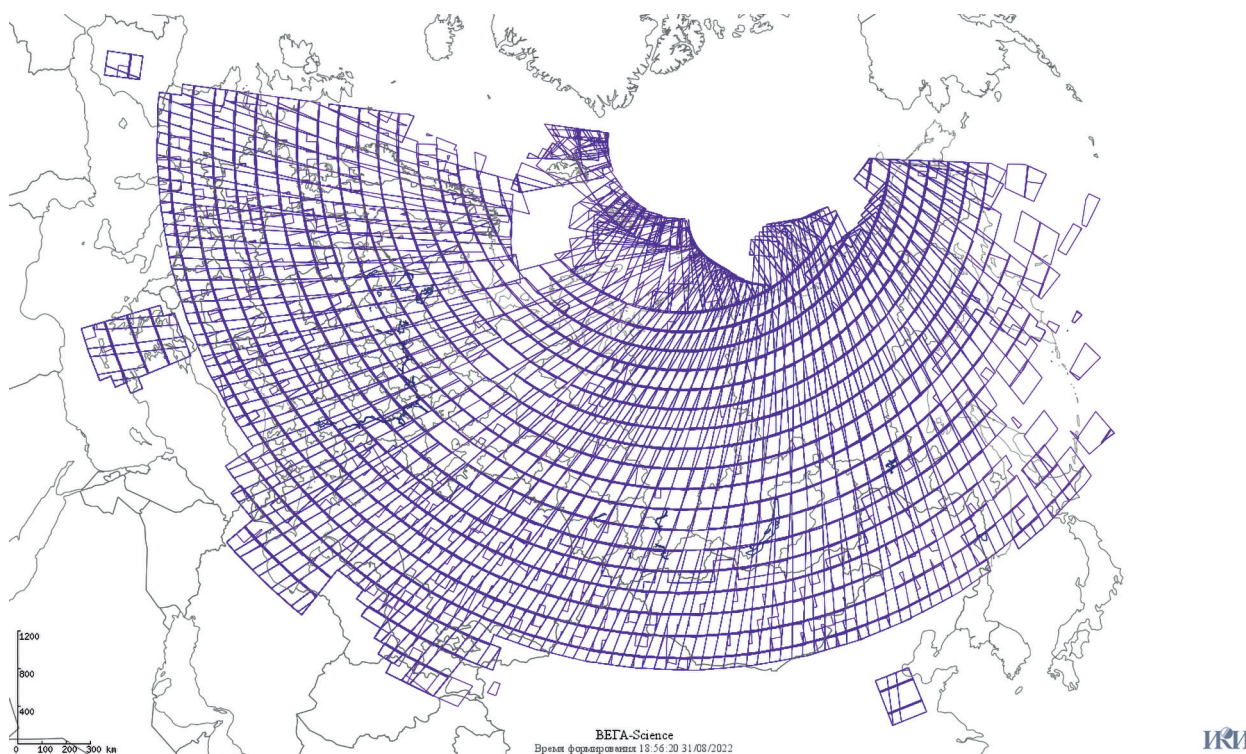


Рис. 1. Пример покрытия данными (продукты всех уровней обработки) спутников серии Sentinel-2 в архивах ЦКП «ИКИ-Мониторинг» на лето 2019 г.

Методы построения

В основе технологии лежит маскирование мешающих факторов с использованием стандартного продукта классификации сцены. Он получен алгоритмом Sen2Cor и подробно описан в технической документации (<https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm>) и ряде работ, например (Louis et al., 2016). Этот продукт, называемый в приведённой технической документации Scene Classification (далее — SC), для каждого пикселя данных указывает один из 11 классов, включая три класса облаков, тени от облаков, тёмные объекты, растительность, не растительность, воду, неклассифицированные пиксели, значение «нет данных» и сбой. Точность SC анализировалась в ряде работ. Например, в публикации (Baetens et al., 2019) указана общая точность 84 %. В исследовании

(Main-Knorn et al., 2017) получена общая точность по всем классам 74 ± 11 %, а при разделении между классами «чистая поверхность и вода» и «неиспользуемые пиксели» — уже 92 ± 4 %. Точность сильно зависит от состояния атмосферы, процента облачности, времени года, территории и других факторов. К тому же алгоритм Sen2Cor совершенствуется и точность SC меняется со временем. В настоящей работе не решалась задача по количественной оценке точности продукта SC, однако перед его применением проведена качественная оценка. Для различных периодов года и различных участков территории России за 2021 г. в экспертном режиме было проанализировано несколько сотен сцен и качество имеющегося для них продукта SC. Отметим основные замеченные проблемы на качественном уровне:

- наличие класса «неклассифицированные пиксели». Особенно часто встречается на урбанизированной территории. Также он присутствует на облачности или тенях от облаков;
- класс «тёмные объекты» часто перепутан с тенями от облаков;
- много проблем на границах классов, когда пограничные или соседние с границей классов пиксели классифицированы неверно. Проблема особенно актуальна для классов облачности;
- существуют одиночные ошибочные пиксели или линии пикселей в разных классах. Некоторые мелкие объекты, например в низинах или на урбанизированной территории, могут находиться в классах облачности непрерывно;
- проблема с классом «дымка» (в документации — THIN_CIRRUS). Часто дымка незначительно искажает значения яркости, однако под ней хорошо просматривается чистая поверхность;
- классы снега и облаков часто перепутаны, особенно это актуально зимой для снежных облачных фронтов.

Для устранения некоторых указанных проблем к каждой сцене дополнительно применяется ряд операций. Схема построения композитного изображения для бесснежного периода приведена на *рис. 2* (см. с. 79).

Поскольку исходные данные хранятся в тайлах размером около 100×100 км в универсальной поперечной проекции Меркатора UTM (*англ.* Universal Transverse Mercator), то для уменьшения временных затрат вся обработка ведётся в этой тайловой сетке. Композитное изображение создаётся по каждому тайлу независимо. Также независимо обрабатываются данные разных спектральных каналов. Для одного тайла отбирается не более определённого числа измерений за период с общим процентом облачности по тайлу не более заданного. Предельный процент облачности и максимальное число измерений задаются опционально в зависимости от периода. Например, для композитного изображения за сезон вегетации берётся 20 измерений с процентом облачности по тайлу не более 50 %. В случае, если за период есть больше указанного числа измерений, то выбираются те, у которых общий процент облачности по тайлу меньше. Остальные измерения не участвуют в дальнейшем расчёте.

В каждой сцене на первом этапе строится маска на основе продукта SC. Для бесснежного периода считаются чистыми классы «растительность», «не растительность», «вода», «тёмные объекты» и «неклассифицированные пиксели», а для снежного периода (зимний композит) ещё и класс «снег». Остальные классы считаются маской. Для снежного (зимнего) композитного изображения операция улучшения маски проводится в один этап, выполняется только стягивание одиночных групп пикселей. Для бесснежного периода к каждой маске применяются независимо морфологические операции расширения (на *рис. 2* «Уточнение “dilation”») и сужения (на *рис. 2* «Уточнение “erosion”»). После применения маски независимо для улучшений путём расширения и сужения считается медианное значение по всему массиву измерений. Если чистое измерение не было получено в ветке с применением операции расширения, то оно заполняется значением из ветки с применением операции сужения. Применение операции расширения в первую очередь необходимо для устранения проблем на границе класса облачности. Применение операции сужения необходимо для устранения одиночных ошибочных пикселей или ошибочных линий пикселей в разных классах. При отсутствии этого

этапа некоторые мелкие объекты были бы полностью исключены из композитного изображения. Размер расширения (10 пикселей) и сужения (1 пиксель) был подобран эмпирически при анализе сотен различных сцен за 2021 г. Результат оценивался в экспертном режиме на качественном уровне, без точной количественной оценки. Дополнительно исследовалась возможность учёта иной комбинации классов из продукта SC, например принятие тёмных объектов или класса неклассифицированных пикселей в качестве маски. Однако учёт иного набора классов в качестве чистых приводил к существенно худшему результату, особенно при числе безоблачных измерений за период менее пяти. На последнем этапе от оставшихся без данных пикселей дополнительно проводится отступление операцией расширения. Эта операция необходима, так как на границе мест отсутствия безоблачных измерений качество результирующего композитного изображения, полученного всего по нескольким сценам, получается низким.

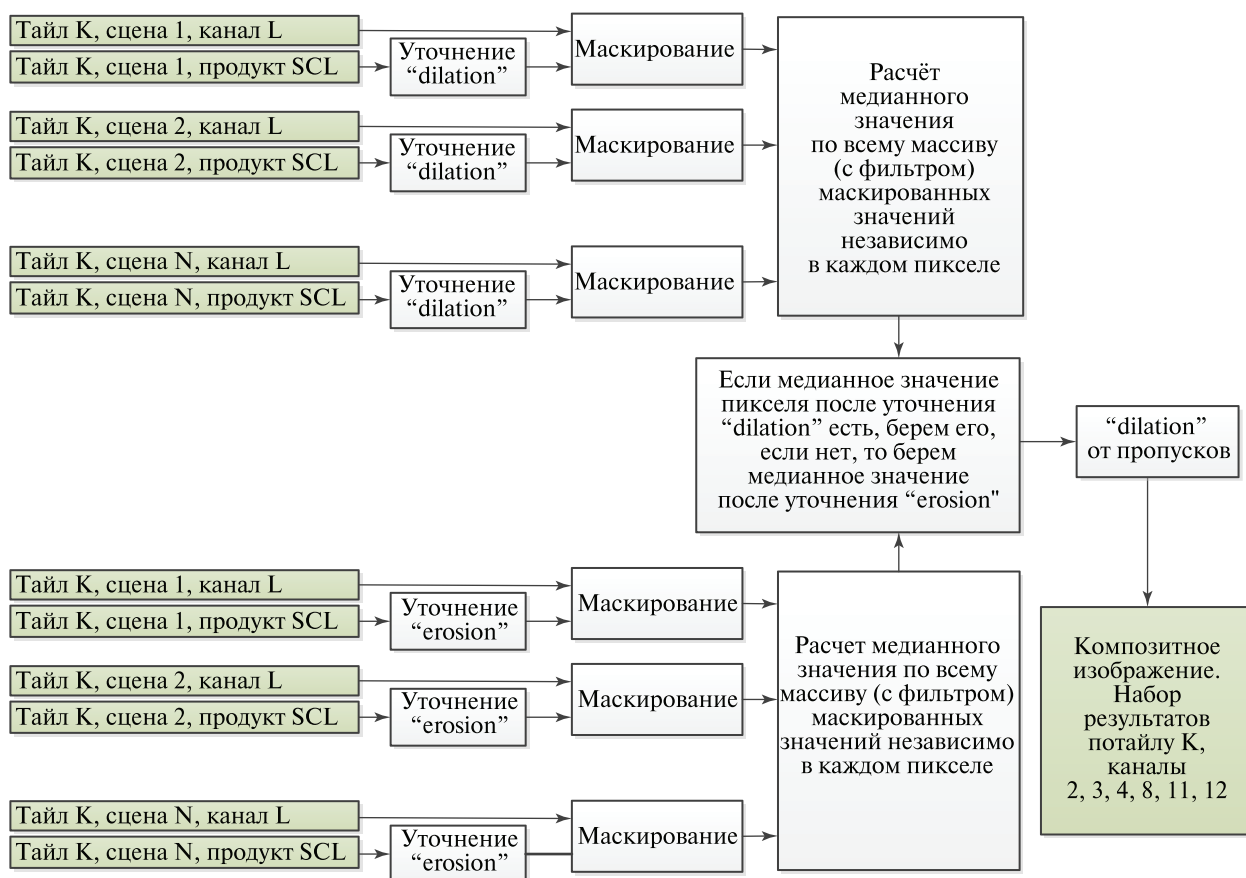
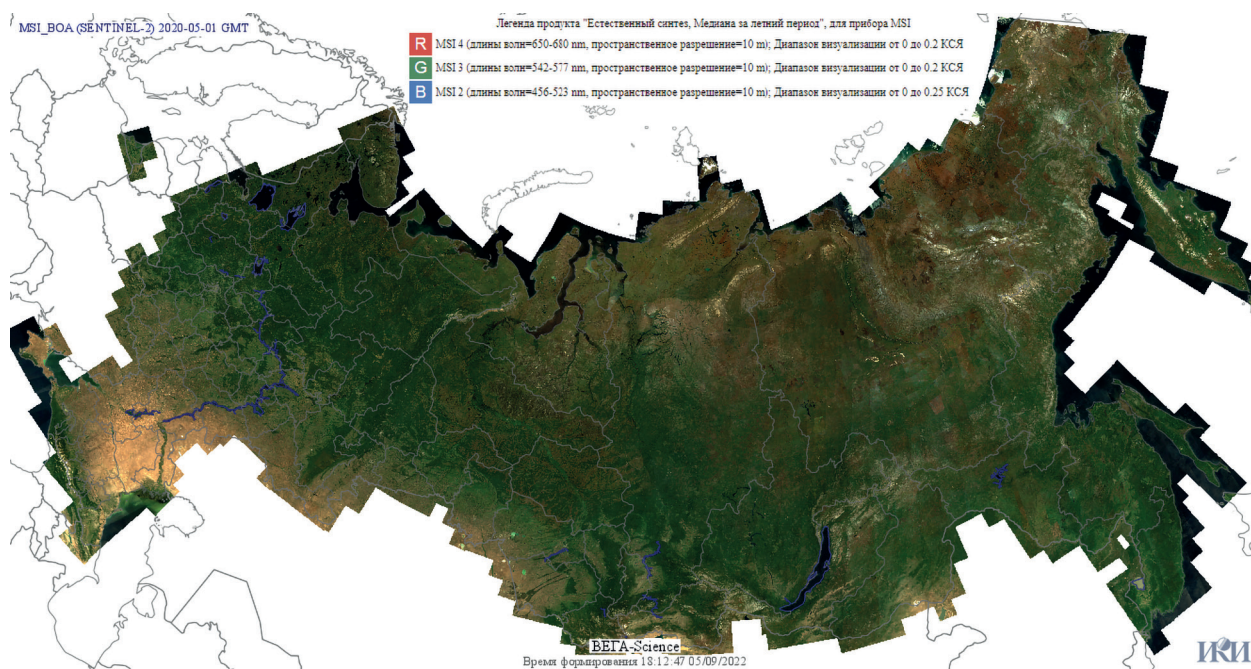


Рис. 2. Схема построения композитного изображения для бесснежного периода

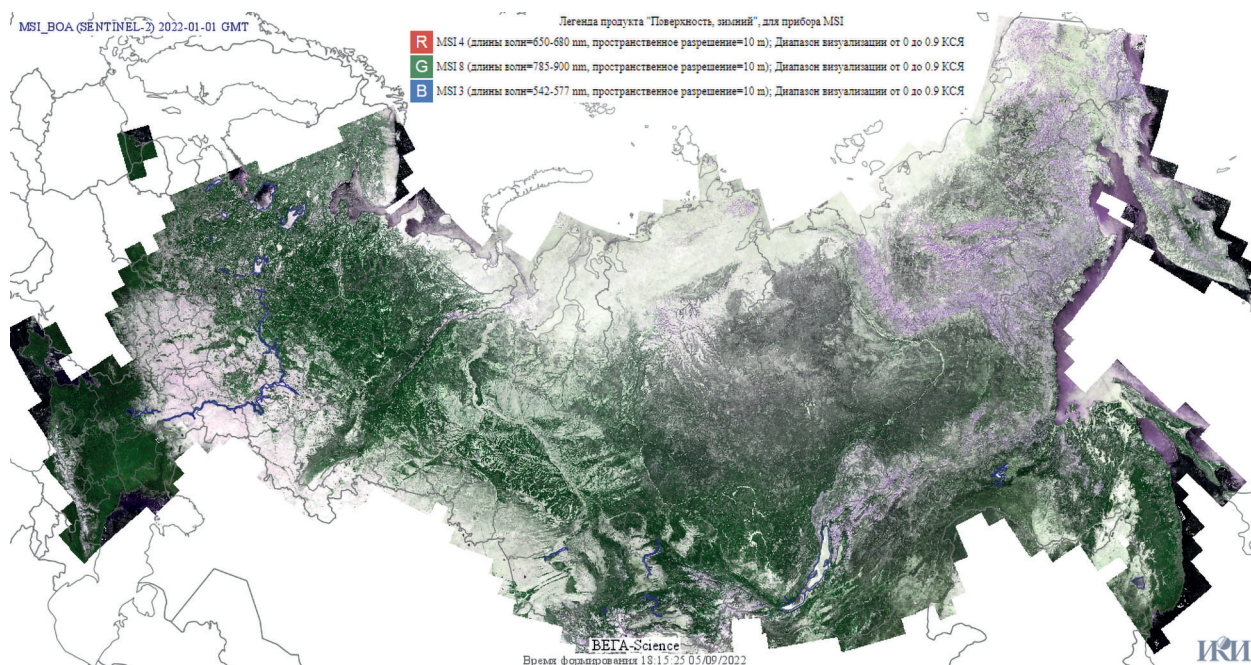
Отметим также особенности технологии и временные затраты на создание композитных изображений по описываемой технологии. Вся информация в ЦКП «ИКИ-Мониторинг» хранится в распределённом файловом архиве на множестве серверов на основе подходов, описанных в работе (Прошин и др., 2016). При создании композитного изображения данные за отобранные даты внутри периода выгружаются с сервера хранения непосредственно в оперативную память сервера, выполняющего обработку. Такой подход позволяет избежать каких-либо дополнительных временных затрат на копирование данных. В результате при расчёте в 8 потоков с использованием одного сервера из состава ЦКП «ИКИ-Мониторинг» (CPU: AMD EPYC 7413, VMware 8 core, 2645 MHz; Memory: 125 GiB) получено среднее время на создание композитного изображения за месячный период на территорию России около трёх дней (при обработке всего 2232 тайлов, в которых в сумме чуть более $270 \cdot 10^9$ пикселей, с фильтром на облачность в 75 % и лимитом безоблачных измерений не более 15).

Результаты

С помощью описанной технологии был создан набор безоблачных композитных изображений на территорию России. Список основных опубликованных композитных изображений приведён в *таблице*. В дальнейшем аналогичные продукты за новые периоды будут создаваться и пополнять архивы в автоматическом режиме. На *рис. 3* приведены примеры созданных композитных изображений.



а



б

Рис. 3а, б. Примеры созданных композитных изображений: а — за сезон вегетации с 1 мая по 31 сентября 2020 г. (синтез в естественных цветах, каналы Red-Green-Blue); б — за снежный период с 1 января по 31 марта 2022 г. (синтез в ложных цветах, каналы Red-NIR-Green) (рис. 3в см. на с. 81)

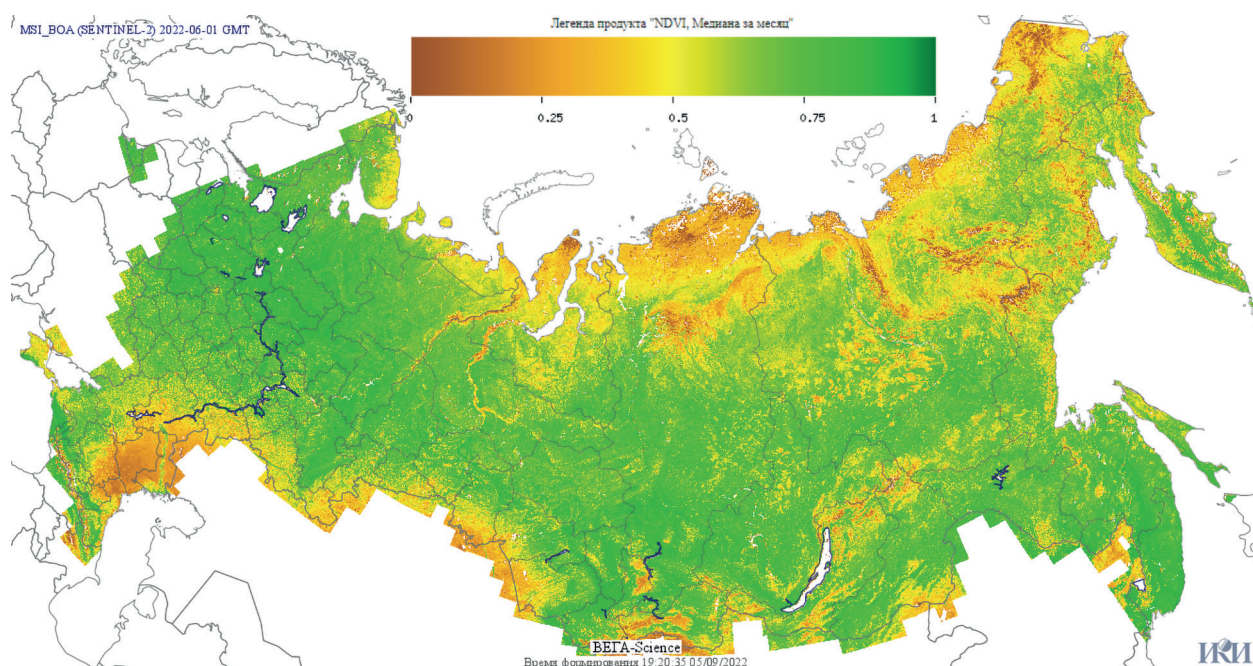


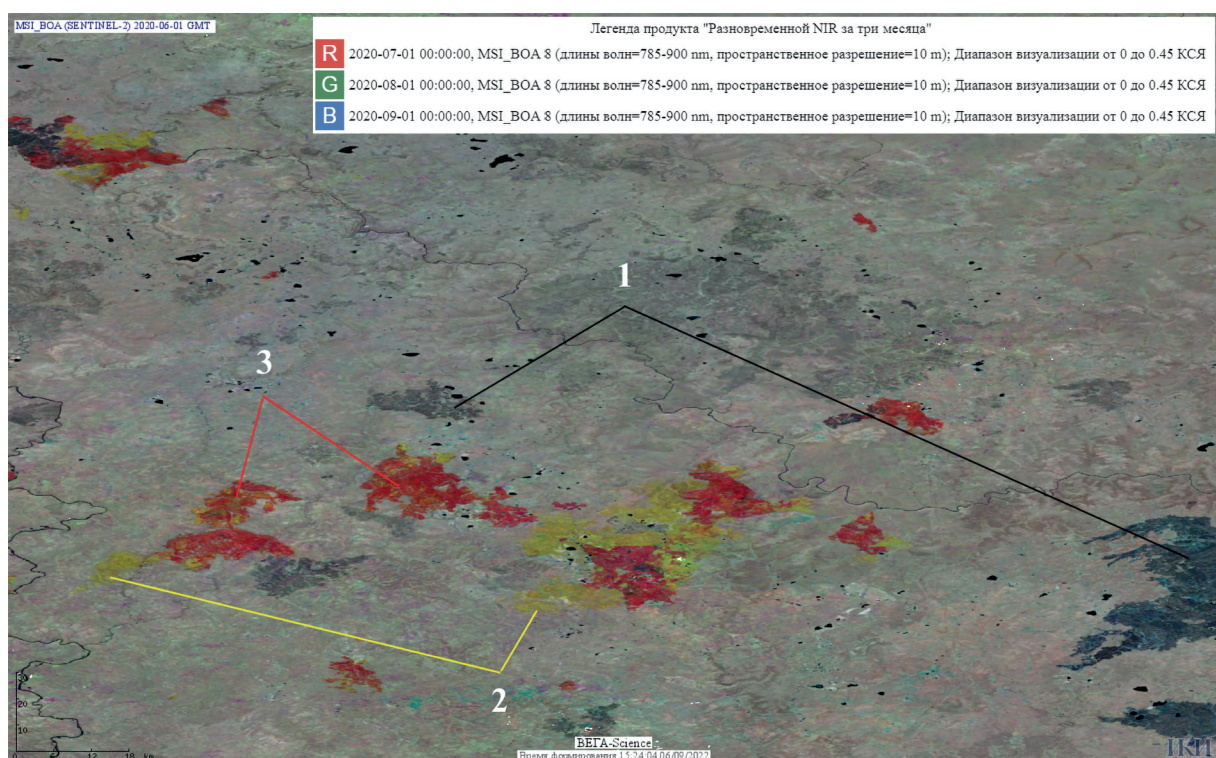
Рис. 3в. Примеры созданных композитных изображений: за июнь 2022 г. (продукт — нормализованный разностный вегетационный индекс NDVI (*англ.* Normalized Difference Vegetation Index)) (рис. 3а, б см. на с. 80)

Список основных опубликованных композитных продуктов

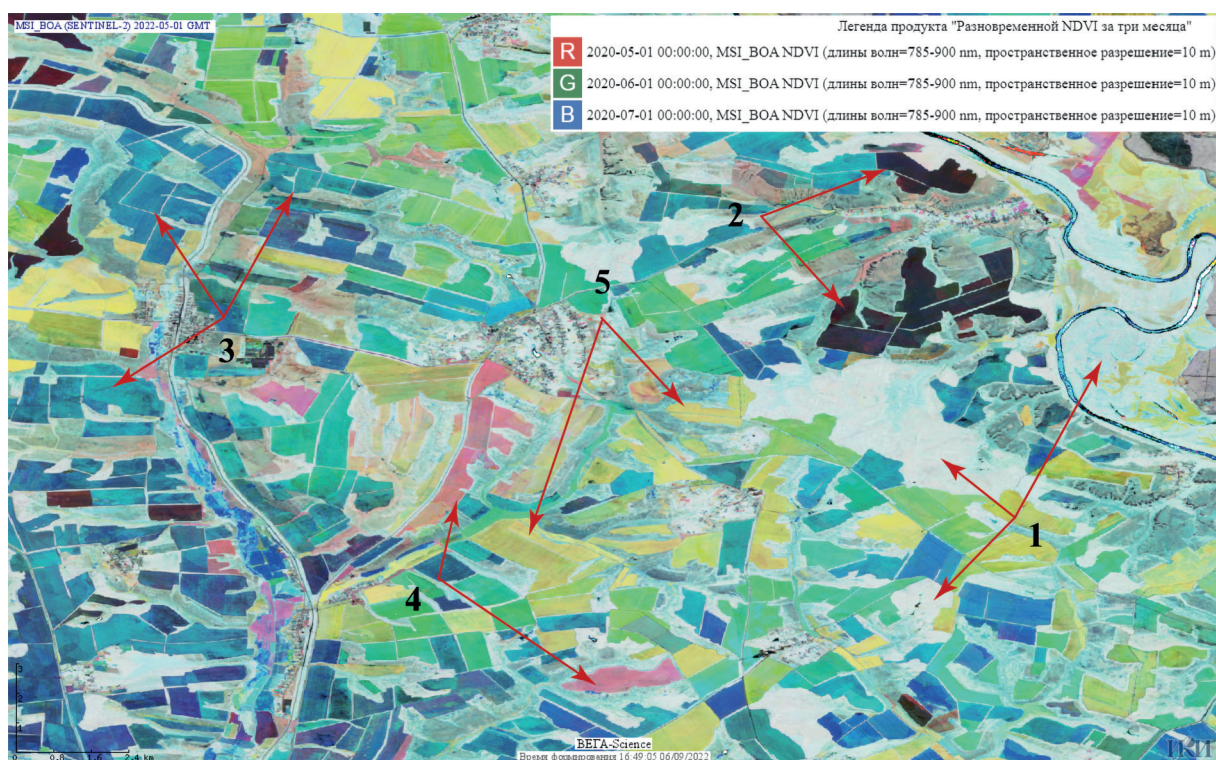
Название	Период
Ежегодный за сезон вегетации	За период с 1 мая по 30 сентября за 2019, 2020, 2021 гг.
Ежемесячный	Все месяцы с марта по ноябрь включительно за 2019, 2020, 2021, 2022 гг. (2022 г. — только по июль, на момент написания статьи)
Ежесезонный	Весенний (с 1 марта по 31 мая), летний (с 1 июня по 31 августа), осенний (с 1 сентября по 30 ноября) за 2019, 2020, 2021, 2022 гг. (2022 г. — только весенний, на момент написания статьи)
Снежный (зимний)	За период с 1 января по 31 марта за 2019, 2020, 2021, 2022 гг.

Реализована сезонная и месячная периодичности создания. Месячный период обусловлен, с одной стороны, достаточным обеспечением безоблачными данными каждой точки за этот период, а с другой стороны, тем, что скважность в один месяц позволяет выявить большинство особенностей развития растительного покрова в течение сезона вегетации. Выбор времени построения снежного (зимнего) композитного изображения обусловлен более вероятным наличием в этот период снежного покрова. Для создания композитных изображений из *таблицы* использованы стандартные продукты уровня обработки L2A, т.е. после проведения процедуры атмосферной коррекции. Все приведённые в *таблице* композитные изображения созданы в шести спектральных каналах: Blue (синий; канал 2 прибора MSI, соответствующий диапазону длин волн 456–523 нм, с пространственным разрешением 10 м/пиксель), Green (зелёный; канал 3, 542–577 нм, 10 м/пиксель), Red (красный; канал 4, 650–680 нм, 10 м/пиксель), NIR (*англ.* near infrared — ближний инфракрасный; канал 8, 785–900 нм, 10 м/пиксель), SWIR1 (*англ.* short wave infrared — коротковолновый инфракрасный; канал 11, 1565–1655 нм, 20 м/пиксель), SWIR2 (канал 12, 2100–2280 нм, 20 м/пиксель). Отметим, что с помощью описанной технологии возможно создание композитных изображений по информации в других спектральных каналах, а также по данным уровня обработки L1C.

Опубликованные композитные изображения доступны через систему «Beta-Science» (<http://sci-vega.ru/>) (Loupian et al., 2022). Также возможно предоставление отдельных наборов данных по запросу.



а



б

Рис. 4. Пример использования информационного продукта «синтез композитных изображений за три месяца»: а — синтез КСЯ (коэффициент спектральной яркости) в канале NIR по композитным изображениям за июль, август, сентябрь 2020 г., территория Якутии, различные цвета при синтезе показывают период появления повреждённых участков леса: 1 — присутствуют во всех месяцах (появились до июля), 2 — появились в августе, 3 — появились только в сентябре; б — синтез индекса NDVI по композитным изображениям за май, июнь, июль 2022 г., территория Воронежской обл., различные цвета при синтезе показывают тип использования: 1 — лес или неиспользуемые земли, 2 — земли под паром, 3 — яровые культуры, 4 — травы или сенокосы, 5 — озимые культуры

Кроме данных в отдельных спектральных каналах, в системе «Вега-Science» также предоставляется ряд других информационных продуктов, являющихся синтезом информации в отдельных каналах композитного изображения или результатом вычислений над ними. Такие информационные продукты получаются в режиме реального времени в момент запроса на основе ранее разработанной технологии предоставления виртуальных информационных продуктов (Прошин и др., 2016). Также предоставляются разновременные продукты на основе композитных изображений, например синтез за три месяца, который хорошо показывает динамику изменений земной поверхности во времени. На *рис. 4* (см. с. 82) приведён пример такого продукта и демонстрация возможного его использования: на *рис. 4а* — в задаче картографирования лесных гарей, различные цвета при синтезе явно показывают период появления повреждённых участков леса; на *рис. 4б* — в задачах анализа сельскохозяйственной растительности. Даже на простом продукте синтеза композитных изображений за три последовательных месяца (май, июнь, июль) возможно решение следующих задач в экспертном или автоматическом режиме: определение границ полей, выявление неиспользуемых земель, выявление пахотных земель, определение различных видов угодий и типов выращиваемых культур. Причём с помощью указанных особенностей в связи с непрерывным покрытием территории при использовании композитных изображений появляется возможность создания карт на большую территорию. Однако обсуждение пригодности полученных композитных изображений для автоматического создания карт выходит за рамки настоящей работы. Поэтому в данной статье ограничимся лишь указанием на возможность применения композитных изображений для создания карт как в экспертном, так и в автоматическом режиме.

Заключение

Представленная в статье технология позволяет достаточно оперативно создавать безоблачные композитные изображения на большую территорию. Полученный с её помощью набор ежемесячных и ежесезонных данных за разные периоды 2019–2022 гг. опубликован на ресурсах ЦКП «ИКИ-Мониторинг» и предлагается для решения научных задач широкому кругу исследователей.

Использование представленных в статье композитных изображений имеет ряд преимуществ по сравнению с данными отдельных сцен. Основными преимуществами выступают: отсутствие облачности и теней от облаков; однородность, непрерывность и бесшовность покрытия территории; отсутствие необходимости изучать множество отдельных сцен. Основной недостаток — это отсутствие точной привязки к определённой дате съёмки значения в конкретном пикселе, а также возможные пропуски экстремальных значений из-за сглаживания со значениями за другие даты.

Дальнейшим развитием работы может быть усовершенствование технологии для восстановления данных за длительные облачные периоды, усовершенствование масок облачности, а также применение созданных композитных изображений в задачах автоматического картографирования земной поверхности.

Работа проведена в рамках темы «Космос-Д» (госрегистрация № 122042500019-6). Получение и обработка спутниковых данных выполнялись с помощью возможностей Центра коллективного пользования «ИКИ-Мониторинг» (Лупян и др., 2019).

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., Кашницкий А. В., Балашов И. В., Барталев С. А., Константинова А. М., Кобец Д. А., Мазуров А. А., Марченков В. В., Матвеев А. М., Радченко М. В., Сычуглов И. Г., Толпин В. А., Уваров И. А. Опыт эксплуатации и развития центра коллективного пользования

- системами архивации, обработки и анализа спутниковых данных (ЦКП «ИКИ-Мониторинг») // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 3. С. 151–170. DOI: 10.21046/2070-7401-2019-16-3-151-170.
3. Прошин А. А., Лупян Е. А., Балашов И. В., Кашницкий А. В., Бурцев М. А. Создание унифицированной системы ведения архивов спутниковых данных, предназначенной для построения современных систем дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 3. С. 9–27. DOI: 10.21046/2070-7401-2016-13-3-9-27.
 4. Прошин А. А., Кашницкий А. В., Бурцев М. А., Брилль А. А. Организация получения, обработки и предоставления данных спутников серии Sentinel-2 в центре коллективного пользования «ИКИ-Мониторинг» // Материалы 17-й Всероссийской открытой конф. «Современные проблемы дистанционного зондирования Земли из космоса». 11–15 нояб. 2019. М.: ИКИ РАН, 2019. С. 117.
 5. Baetens L., Desjardins C., Hagolle O. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure // Remote Sensing. 2019. V. 11. Art. No. 433. DOI: 10.3390/rs11040433.
 6. Louis J., Debaecker V., Pflug B., Main-Knorn M., Bieniarz J., Mueller-Wilm U., Cadau E., Gascon F. Sentinel-2 Sen2Cor: L2A processor for users // Proc. Living Planet Symp. Prague, Czech Republic, 9–13 May 2016. ESA SP-740.
 7. Loupian E. A., Bourtsev M. A., Proshin A. A., Kashnitskii A. V., Balashov I. V., Bartalev S. A., Konstantinova A. M., Kobets D. A., Radchenko M. V., Tolpin V. A., Uvarov I. A. Usage Experience and Capabilities of the VEGA-Science System // Remote Sensing. 2022. V. 14. No. 1. Art. No. 77. DOI: 10.3390/rs14010077.
 8. Main-Knorn M., Pflug B., Louis J., Debaecker V., Müller-Wilm U., Gascon F. Sen2Cor for sentinel-2 // Proc. SPIE. 2017. V. 10427. P. 37–48. DOI: 10.1117/12.2278218.

Technology to create cloud-free composites from Sentinel-2 satellite data

A. V. Kashnitskii, M. A. Burtsev, A. A. Proshin

Space Research Institute RAS, Moscow 117997, Russia

E-mail: kashnizky@gmail.com

The paper describes the technology for obtaining cloudless composite images based on Sentinel-2 satellite data, as well as the set of composite images of the territory of Russia obtained with this technology for different periods. The technology is based on masking of clouds and their shadows using a modified scene classification Sen2Cor algorithm derived product. Resulting pixel values of the composite image are obtained by calculating the median of all net measurements for the period, and the number of scenes involved in the calculation is selected depending on the specific situation. As a result, a set of monthly (from March to November) and seasonal (spring, summer, autumn, growing season, snow period) cloudless composite images in six spectral bands was created using the described technology from Sentinel-2 series satellites data for the territory of Russia for 2019–2022. The use of the created composite images showed their advantage in solving a number of problems compared to using individual scenes. A set of composite images for the territory of Russia for different periods has been published on the resources of the IKI-Monitoring Center for Collective Use and is now available to the wide research community as an aid in solving scientific problems.

Keywords: Sentinel-2, composite images, cloud detection, cloudless monthly data, winter composite, remote sensing

Accepted: 31.10.2022

DOI: 10.21046/2070-7401-2022-19-5-76-85

References

1. 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).
2. Loupian E. A., Proshin A. A., Burtsev M. A., Kashnitskii A. V., Balashov I. V., Bartalev S. A., Konstantinova A. M., Kobets D. A., Mazurov A. A., Marchenkov V. V., Matveev A. M., Radchenko M. V., Syuhov I. G., Tolpin V. A., Uvarov I. A., Experience of development and operation of the "IKI-Monitoring" center for collective use of systems for archiving, processing and analyzing satellite data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2019, Vol. 16, No. 3, pp. 151–170 (in Russian), DOI: 10.21046/2070-7401-2019-16-3-151-170.
3. Proshin A. A., Loupian E. A., Balashov I. V., Kashnitskii A. V., Burtsev M. A., Unified satellite data archive management platform for remote monitoring systems development, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2016, Vol. 13, No. 3, pp. 9–27 (in Russian), DOI: 10.21046/2070-7401-2016-13-3-9-27.
4. Proshin A. A., Kashnitskii A. V., Burtsev M. A., Bril A. A., Organization of receiving, processing and providing data from Sentinel-2 satellites in the IKI-Monitoring shared use center, *Materialy 17-i Vserossiiskoi otkrytoi konferentsii "Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa"* (Proc. 17th All-Russia Open Conf. "Current Problems in Remote Sensing of the Earth from Space"), 11–15 Nov. 2019, Moscow: IKI RAN, 2019, p. 117 (in Russian).
5. Baetens L., Desjardins C., Hagolle O., Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure, *Remote Sensing*, 2019, Vol. 11, Art. No. 433, DOI: 10.3390/rs11040433.
6. Louis J., Debaecker V., Pflug B., Main-Knorn M., Bieniarz J., Mueller-Wilm U., Cadau E., Gascon F., Sentinel-2 Sen2Cor: L2A processor for users, *Proc. Living Planet Symp.*, Prague, Czech Republic, 9–13 May 2016, ESA SP-740.
7. Loupian E. A., Burtsev M. A., Proshin A. A., Kashnitskii A. V., Balashov I. V., Bartalev S. A., Konstantinova A. M., Kobets D. A., Radchenko M. V., Tolpin V. A., Uvarov I. A., Usage Experience and Capabilities of the VEGA-Science System, *Remote Sensing*, 2022, Vol. 14, No. 1, Art. No. 77, DOI: 10.3390/rs14010077.
8. Main-Knorn M., Pflug B., Louis J., Debaecker V., Müller-Wilm U., Gascon F., Sen2Cor for Sentinel-2, *Proc. SPIE*, 2017, Vol. 10427, pp. 37–48, DOI: 10.1117/12.2278218.