

## Алгоритм автоматической фильтрации облачных данных для решения задач объектного дистанционного мониторинга

А. М. Константинова

*Институт космических исследований РАН, Москва, 117997, Россия  
E-mail: konstantinova.anouk@gmail.com*

Описывается подход к автоматической фильтрации облачных данных для обеспечения работы технологии объектного мониторинга, разработанной в Институте космических исследований РАН (ИКИ РАН). Данная технология создана для мониторинга и контроля динамики различных природных и антропогенных объектов (однородных образований). Она обеспечивает автоматическое формирование и анализ долговременных рядов различных характеристик наблюдаемых объектов, получаемых на основе данных спутниковых наблюдений. Одна из проблем формирования таких рядов заключается в организации автоматической фильтрации некорректных данных, возникновение которых связано в первую очередь с наличием на обрабатываемых спутниковых изображениях облачности, закрывающей объект. Организация такой фильтрации может быть затруднена в случаях, когда не удаётся обеспечить надёжное автоматическое выделение облачности. В работе представлен алгоритм, который позволяет осуществлять автоматическую фильтрацию и в такой ситуации. Он основан не на анализе наличия облачности в отдельных пикселях, а на интегральной оценке её присутствия в окрестности наблюдаемого объекта и поиске порогов процентного покрытия облачностью такой окрестности, при превышении которых значения характеристик объекта, рассчитанных по этим данным, почти заведомо будут сбийными. Проведён анализ и показана возможность выбора оптимальных пороговых значений, которые бы позволяли обеспечивать фильтрацию максимального числа сбийных наблюдений при минимальном числе пропусков (фильтрации) безоблачных данных. Приводятся примеры использования предложенного алгоритма фильтрации для формирования рядов наблюдений характеристик объектов, расположенных в регионах с различными условиями наблюдения. Также приведено сравнение результатов работы предложенного метода фильтрации с результатами других методов, в частности фильтрации выбросов в ряду значений характеристик объекта.

**Ключевые слова:** дистанционный мониторинг, природные объекты, технология, объектный мониторинг, фильтрация облачности, фильтрация выбросов, снежный индекс, ЦКП «ИКИ-Мониторинг»

Одобрена к печати: 25.08.2022  
DOI: 10.21046/2070-7401-2022-19-4-88-99

### Введение

Для задач мониторинга и изучения различных природных и антропогенных объектов в Институте космических исследований РАН (ИКИ РАН) была разработана специальная технология, основанная на автоматизированном расчёте различных характеристик таких объектов с использованием спутниковых данных (Константинова и др., 2021; Лупян и др., 2020). Технология рассчитана на организацию мониторинга объектов, представляющих собой «однородные» участки, состояние которых может описываться набором некоторых осреднённых по их площади характеристик. Такими объектами могут, например, быть поля, однородные участки леса, водные объекты и т.д. Созданная технология позволяет организовывать автоматизированную обработку и анализ временных рядов различных характеристик объектов, оценку которых можно получить на основе спутниковых данных.

Одна из задач, которую необходимо решать в рамках созданной технологии, состоит в построении рядов наблюдений характеристик объектов, очищенных от различных шумов, которые во многом связаны с наличием облачности на конкретных спутниковых изображениях в районах расположения объектов. В частности, описанная задача требует реализации методов автоматической фильтрации облачных наблюдений. Такая фильтрация

обычно выполняется на основе масок облачности (набора точек с большой вероятностью наличия облачности в них). При этом следует учитывать, что подобные маски в различных ситуациях и для различных данных могут содержать значительные ошибки, поэтому при оценке качества информации, которую можно получить по объекту с использованием конкретной сцены спутниковых данных, не всегда эффективно проводить попиксельный анализ маски облачности, поскольку в отдельных точках (пикселях) в ней могут содержаться значительные ошибки. Одним из вариантов исключения таких ошибок может быть анализ информации не в отдельных пикселях, покрывающих объект наблюдения, а характеристик облачного покрытия в окрестности объекта (например, процента покрытия облачностью окрестности). Фильтрация сбойных (облачных) наблюдений, проведённая на основе такого анализа, может в определённой степени нивелировать ошибки в построении масок облачности. Рассмотрению возможностей и особенностей именно такого подхода и посвящена настоящая работа.

### Предлагаемый подход

В рассматриваемом подходе предлагается заменить попиксельную фильтрацию облачности над объектом (результат наложения маски облачности на исходные данные) на фильтрацию «сбоев» в рядах вычисляемых по спутниковым данным различных характеристик объекта, основанную на использовании информации о проценте площади, занятой облачностью (ПЗО) в окрестности объекта на конкретных сценах спутниковых данных. В подходе предполагается, что, начиная с определённого порога ПЗО, можно с большой вероятностью сказать, что на рассматриваемой сцене объект будет в основном покрыт облачностью и получить качественную оценку необходимых характеристик невозможно, т. е. сцена не подходит для расчёта характеристик объекта. Использование такого подхода может нивелировать ошибки распознавания облачности и исключить из анализа те сцены, по которым информация, получаемая о наблюдаемых объектах, будет сильно зашумлена.

Алгоритм, основанный на данном подходе (фильтрации по порогу ПЗО), может быть сведён к следующей последовательности шагов:

- Для исследуемого объекта и рассматриваемой сцены спутниковых данных формируется маска облачности (это может быть стандартная маска облачности или результат детектирования облачности в пикселе любым другим способом).
- По сформированной маске рассчитывается ПЗО в окрестности анализируемого объекта.
- Если ПЗО превышает значение порога, то считается, что выбранная сцена непригодна для дальнейшего расчёта показателей объекта, и она не используется для формирования ряда характеристик наблюдаемого объекта.
- После проведённой таким образом отбраковки сцен формируется уже «очищенный» ряд данных по наблюдаемому объекту.

В рассматриваемом подходе ключевым вопросом становится выбор значения порога, на основе которого делается вывод, отнести ли сцену к облачной или нет при формировании ряда данных по конкретному объекту. Поэтому основная задача настоящей работы — анализ количества ошибок, которые возникают при вычислении характеристик объекта при различных уровнях облачности и условиях наблюдения территории, на которой располагается наблюдаемый объект. Такой анализ направлен в первую очередь на исследование зависимости количества ошибок (сбойных вычислений характеристик объекта) от покрытия облачностью района, в котором он располагается, и оценки возможности выбора оптимального порогового значения облачности, фильтрация по которому минимизирует количество ошибок в ряде наблюдений объекта. Анализ основан на использовании реальных рядов данных наблюдений различных объектов, для которых была осуществлена визуальная фильтрация спутниковых данных (изображений), чтобы исключить сбойные значения (вычисленные на облачных данных). Ниже представлены результаты проведённого анализа.

## Наборы данных, использованные в работе

Для решения обсуждаемой задачи использовались следующие данные:

- Описания объектов (границы объекта, состав вычисляемых характеристик, состав используемых спутниковых данных, временной период, за который выбираются спутниковые данные), расположенных на различных территориях. В работе использовался набор объектов, сформированных в рамках различных исследовательских проектов, в которых применяется технология объектного мониторинга (Константинова и др., 2021; Лупян и др., 2020; Мухамеджанов и др., 2022). Анализ проводился для объектов разных форм, разных типов (участки растительности, водные объекты), расположенных на географически разных территориях. Всего было выбрано 18 объектов за период с 2013 по 2021 г. На *рис. 1* представлена карта зон расположения выбранных объектов.

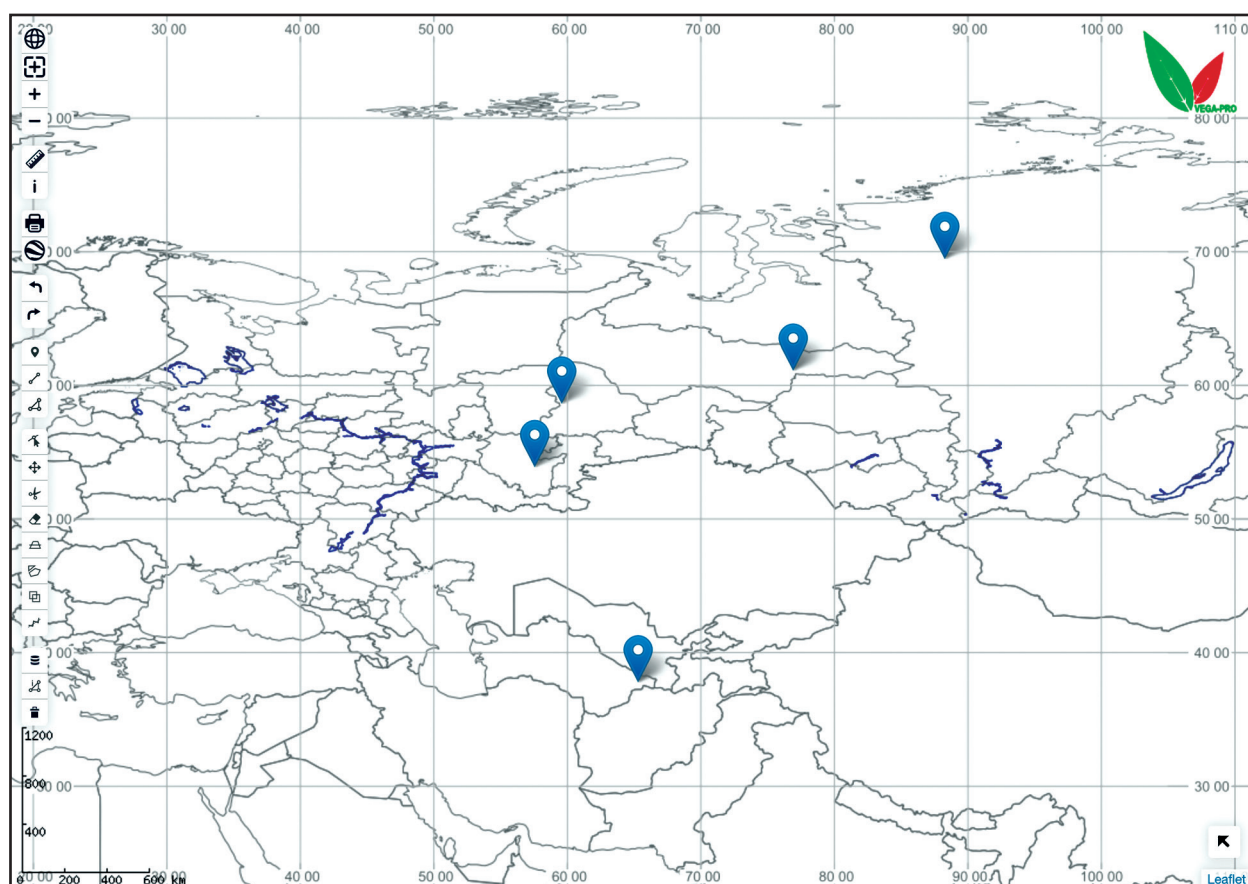


Рис. 1. Карта зон расположения анализируемых объектов

- Ряды различных характеристик выбранных объектов, автоматически сформированные по всем имеющимся для объектов спутниковым данным за заданные периоды наблюдения.
- Наборы спутниковых данных, которые использовались для формирования рядов характеристик выбранных объектов. Элементы спутниковых сцен по районам расположения выбранных объектов. Всего использовалось около 6300 таких элементов.
- Маски облачности по районам расположения выбранных объектов, сформированные для всех сцен, которые использовались для автоматического расчёта характеристик объектов.
- Набор элементов сцен, которые не позволили осуществить достоверный расчёт требуемых характеристик объектов. Эти списки были сформированы на основе визуального

анализа используемых спутниковых данных в окрестностях выбранных объектов и результатов автоматического расчёта их характеристик. То есть фактически это выборка элементов сцен, которые должны быть отфильтрованы при построении рядов характеристик объектов. В выборку вошло около 2400 таких элементов.

Таким образом, можно сказать, что использовался достаточно большой набор тестовых данных, сформированный на основе экспертного анализа и предназначенный для оценки возможности применения предлагаемого подхода и выбора оптимального ПЗО для проведения фильтрации облачных сцен при расчётах характеристик различных объектов.

В работе использовались данные следующих спутниковых приборов: MSI (*англ.* Multispectral Instrument) (Sentinel-2), OLI (*англ.* Operational Land Imager) (Landsat-8), ETM+ (*англ.* Enhanced Thematic Mapper Plus) (Landsat-7), TM (*англ.* Thematic Mapper) (Landsat-4, -5). Данные были получены из архивов Центра коллективного пользования (ЦКП) «ИКИ-Мониторинг» (Лупян и др., 2019; <http://ckp.geosmis.ru>). Также применялись стандартные маски облачности, получаемые по данным этих приборов. А именно продукты, созданные на основе алгоритмов Level-1C cloud detection algorithm для коллекции Sentinel-2 без атмосферной коррекции (Copernicus..., 2021a), Level-2A Algorithm для коллекции Sentinel-2 с атмосферной коррекцией (Copernicus..., 2021b) и Fmask (для коллекции Landsat) (Luo et al., 2020; Zhu et al., 2015). При этом маски для коллекций Sentinel-2 с атмосферной коррекцией и Landsat представляют собой изображения с несколькими классами (разные типы облаков, тени от них, снег, земная и водная поверхности и др.) (Baetens et al., 2019; Zhu et al., 2015). Маска для коллекции Sentinel-2 без атмосферной коррекции представляет собой изображение с тремя классами (безоблачные данные, плотные и перистые облака) (Coluzzi et al., 2018).

Следует отметить, что данные маски облачности могут содержать значительное число ошибок. Например, маски, построенные на основе алгоритма Fmask, имеют много ошибок в районах расположения водных объектов, особенно на мутной воде и воде с высокой концентрацией отложений (Zhu et al., 2015).

Поэтому для случаев (наборов объектов), когда стандартные маски имели значительные ошибки и не могли использоваться для расчёта ПЗО в окрестности анализируемых объектов, использовался специально реализованный алгоритм детектирования облаков. В качестве базы была взята фильтрация эффектов маскирующего влияния облаков по данным MODIS (*англ.* Moderate Resolution Imaging Spectroradiometer) из работы (Барталев и др., 2016), в основе алгоритма которой лежит использование данных измерений коэффициента спектральной яркости (КСЯ) в голубом канале  $R_2$  (456–523 и 450–510 мкм для MSI (Sentinel-2) и OLI (Landsat-8) соответственно) и нормализованного разностного индекса снега NDSI (*англ.* Normalized Difference Snow Index), определяемого для данных MSI (Sentinel-2) и OLI (Landsat-8) по формулам:

$$\text{NDSI} = \frac{R_3 - R_{11}}{R_3 + R_{11}}, \quad \text{NDSI} = \frac{R_3 - R_6}{R_3 + R_6},$$

где  $R_3$  — КСЯ в зелёном канале (542–577 и 530–590 мкм для MSI (Sentinel-2) и OLI (Landsat-8) соответственно),  $R_6$  и  $R_{11}$  — КСЯ в SWI-канале (*англ.* short wave infrared, коротковолновый инфракрасный) (1565–1655 и 1570–1650 мкм для MSI (Sentinel-2) и OLI (Landsat-8)).

В реализованном алгоритме детектирования облачности были использованы следующие критерии оценки:

- если значение КСЯ в голубом канале ниже выбранного порога, то пиксель относится к безоблачному, в противном случае он считается «потенциально облачным»;
- если в «потенциально облачном» пикселе значение индекса снега NDSI не превышает заданный порог, то пиксель считается облачным, в противном случае он относится к снежному покрову.



Порог для детектирования облачности по голубому каналу был определён экспериментальным путём и равен 0,25, а пороги детектирования снега по NDSI взяты из источников (Dozier, 1989; <https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm>) и равны 0,42 и 0,40 для MSI (Sentinel-2) и OLI (Landsat-8) соответственно.

### Выбор оптимального порога ПЗО для фильтрации облачных данных при расчёте рядов характеристик объектов

Оптимальным значением порога ПЗО следует считать величину, для которой при фильтрации ряда спутниковых наблюдений будет минимизирована интегральная ошибка (ИО), включающая в себя ошибки как первого, так и второго рода. В рассматриваемой задаче к этим ошибкам относятся пропуски (невключение в ряд) данных, на которых непосредственно над объектом нет облачности, и для объекта могут быть рассчитаны необходимые характеристики, а также облачные сцены, фильтрация которых не была проведена, и в ряду наблюдений остались сбойные данные. В предложенном подходе ИО не может быть сведена к нулю, поскольку всегда останется вероятность того, что даже на практически безоблачных данных именно на объект может попасть облачность. С другой стороны, сильное понижение порога ПЗО с целью минимизировать вероятность «закрытия» объекта облачностью приводит к ошибкам, связанным с фильтрацией безоблачных измерений (пропускам). Поэтому в предложенном подходе фактически не выполнена гипотеза компактности (Мерков, 2011), т. е. между облачными или безоблачными изображениями нельзя провести границу, безошибочно отделяющую изображения одного класса от другого в пространстве рассматриваемого признака классификации (рис. 2).

Оптимальное значение порога станет линейной границей между классами облачных и безоблачных изображений для одномерного пространства признаков классификации. Значение порога рассчитывается по формуле:

$$\varepsilon^* = \arg \min_{\varepsilon} \left\{ \sum_{i=1}^n \varphi_+(\varepsilon_{a_i} - \varepsilon) + \sum_{i=1}^m \varphi_-(\varepsilon_{b_j} - \varepsilon) \right\}; a_i \in A, b_j \in B, \quad (1)$$

где  $A$  — множество облачных изображений;  $B$  — множество безоблачных изображений;  $n = |A|$  — количество облачных изображений;  $m = |B|$  — количество безоблачных изображений;  $\varphi_+$  и  $\varphi_-$  — функции потерь (ФП) при классификации;  $\varepsilon_{a_i}$  — ПЗО на  $i$ -м облачном изображении;  $\varepsilon_{b_j}$  — ПЗО на  $j$ -м безоблачном изображении;  $\varepsilon^*$  — оптимальное значение порога ПЗО.

Разный вид ФП определяет разную математическую основу подхода к выбору оптимального значения порога облачности (рис. 3, см. с. 93):  $a$  — пороговая ФП обеспечивает минимум суммарного количества неправильно классифицированных изображений;  $b$  — линейная ФП обеспечивает минимум суммы расстояний процента облачности неправильно классифицированных изображений от значения порога;  $v$  — квадратичная ФП обеспечивает минимум дисперсии процента облачности неправильно классифицированных изображений относительно значения порога.

В результате анализа качества работы алгоритма была выбрана пороговая ФП для обеспечения условия минимума суммы ошибочно классифицированных облачных и безоблачных

изображений. Для этого случая формулу (1) можно свести к виду:

$$\varepsilon^* = \arg \min_{\varepsilon} \left\{ \frac{N_m(\varepsilon) + N_e(\varepsilon)}{N} \right\},$$

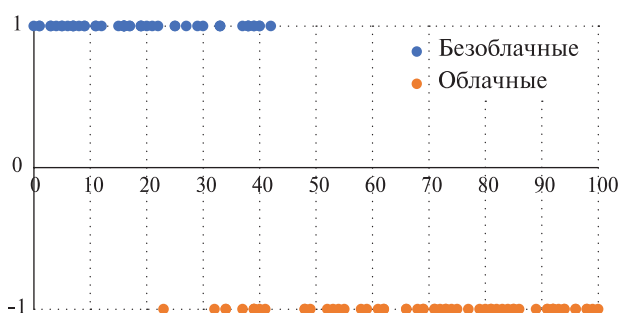


Рис. 2. Модельное распределение облачных и безоблачных изображений в зависимости от доли облачных пикселей в процентах в окрестности объекта

где  $N_m$  — количество пропущенных (missed) в результате фильтрации качественных изображений;  $N_e$  — количество лишних (extra) некачественных изображений, попавших в выборку после фильтрации;  $N$  — размер выборки.

То есть зависимость ИО от порога ПЗО будет описываться формулой:

$$f(\epsilon) = \frac{N_m(\epsilon) + N_e(\epsilon)}{N}$$

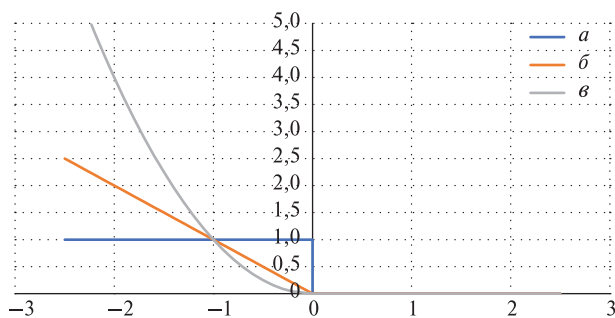


Рис. 3. Разные виды функций потерь

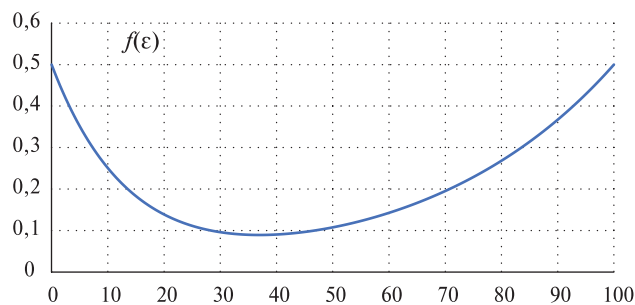


Рис. 4. Модельный вид функции интегральной ошибки в зависимости от порога ПЗО

Качественный вид функции ИО приведён на рис. 4. Он обусловлен тем, что при понижении порога растёт число «пропущенных» безоблачных данных, а при его повышении возрастают ошибки, связанные с неотфильтрованными сбойными (полученными по облачным измерениям) данными. Такой ход зависимости показывает, что может быть найдено оптимальное пороговое значение ПЗО для проведения фильтрации.

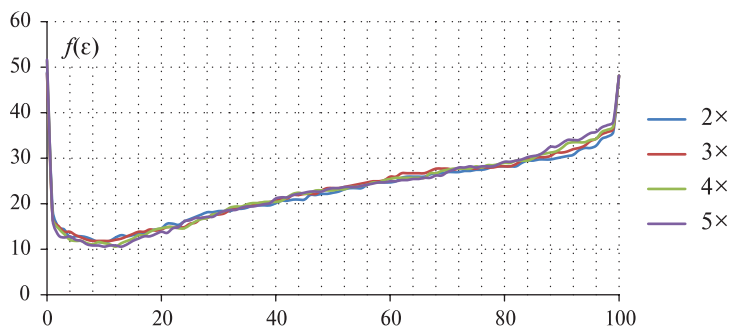


Рис. 5. Пример функции интегральной ошибки, сформированной для объектов, расположенных в зоне, для которой вероятность покрытия облачностью (число облачных дней в году) составляет 50 %.

Ряды сформированы на основе анализа данных наблюдений прибора MSI (спутники Sentinel-2)

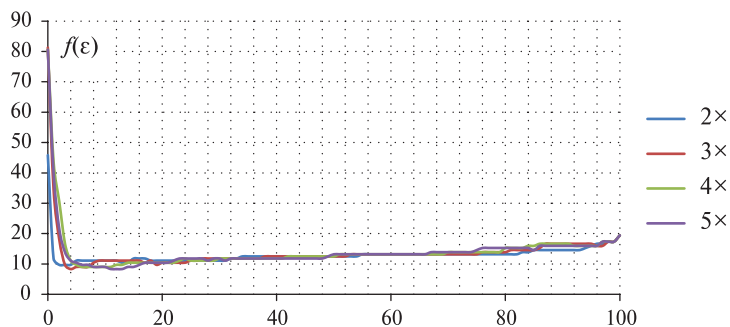
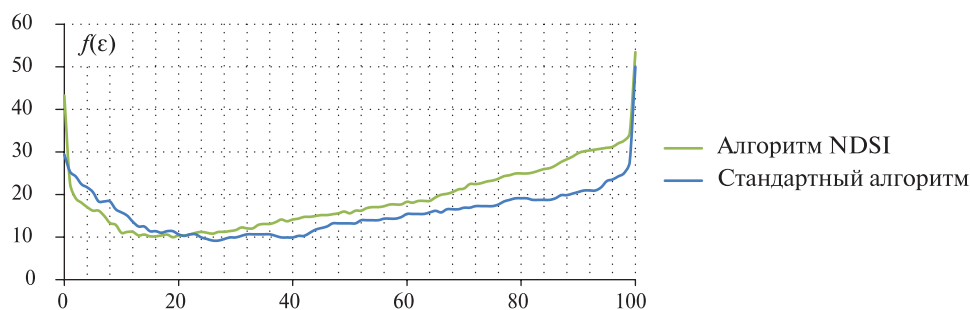


Рис. 6. Пример функции интегральной ошибки, сформированной для объектов, расположенных в зоне, для которой вероятность покрытия облачностью (число облачных дней в году) менее 20 %.

Ряды сформированы на основе анализа данных наблюдений прибора MSI (спутники Sentinel-2)

Для анализа особенностей выбора оптимального порогового значения ПЗО были проанализированы описанные выше тестовые данные. Анализ проводился для объектов, которые располагались в регионах с различными характеристиками облачного покрова (числом облачных дней в году). При этом оценка ПЗО осуществлялась по окрестностям (областям, окружающим объекты) различных размеров. Были проведены оценки для размеров окрестностей, которые в 2, 3, 4 и 5 раз превышали характерные размеры конкретных наблюдаемых объектов. Примеры оценок функции ИО приведены на *рис. 5* и *б* (см. с. 93), они были получены на основе анализа данных наблюдений прибора MSI (спутники Sentinel-2).

Также была проанализирована устойчивость зависимостей ИО от используемых методов построения масок облачности для оценки ПЗО. На *рис. 7* представлен пример зависимостей, полученных для рядов, формируемых на основе данных MSI с использованием стандартного алгоритма их обработки (Сорегнисус..., 2021b) и предложенного выше упрощённого алгоритма с использованием индекса NDSI. В целом проведённый анализ различных данных позволяет сказать, что получаемые зависимости ИО от порога ПЗО достаточно устойчивы к типу используемого алгоритма построения маски облачности.



*Рис. 7.* Функции интегральной ошибки для различных алгоритмов выделения облачности для проведения оценки ПЗО

Таким образом, выполненный анализ, проиллюстрированный графиками на *рис. 5–7*, позволяет сказать, что характер поведения полученных зависимостей соответствует модельной зависимости, вид которой приведён на *рис. 4*. При этом анализ различных наборов данных показал, что практически во всех ситуациях (исключая регионы, вероятность наблюдения облачности в которых близка к 100 %) можно выбрать такое пороговое значение, при котором ИО в получаемом ряде наблюдений будет составлять порядка 10 %. Также можно отметить, что при падении порога ПЗО ниже 10–20 % для всех объектов, вне зависимости от вероятности наблюдения облачности в районах их расположения, виден резкий рост ИО, связанный с пропуском безоблачных сцен. В то же время нарастание ИО от минимальных значений в области больших значений ПЗО происходит относительно плавно, даже для объектов, расположенных в районах с большой вероятностью наблюдения облачности. При этом из приведённых на *рис. 5* и *б* примеров хорошо видно, что оптимальные значения порога ПЗО для данных, полученных в регионах с большим процентом покрытия облачностью, также обеспечивают низкие ИО для регионов с низким процентом покрытия облачности. Исходя из вышеизложенного можно заключить, что достаточно универсальное значение ПЗО лежит в области от 10 до 20 %. Порог с такими значениями позволяет получить низкие ИО при наблюдении объектов, расположенных в различных регионах.

### Примеры применения предлагаемого алгоритма фильтрации

Ниже приведены примеры применения алгоритма фильтрации для различных объектов, расположенных в разных географических регионах: в Западной Сибири, на Южном Урале и в Центральной Азии (Узбекистан). В качестве размера окрестности был принят размер, равный

четырёх характерным размерам объекта, а значение порога облачности выбрано единым для всех объектов и равным 15 %. На рис. 8–10 (см. с. 96, 97) представлены очищенные вручную (тестовые) ряды характеристик объектов (оранжевый цвет), сформированные на основе данных коллекции Sentinel-2 без атмосферной коррекции, совместно с рядами характеристик, полученных на различных этапах фильтрации (синий цвет):

- до фильтрации;
- после применения «маскирования» данных (исключение облачных пикселей) на основе стандартной маски облачности (Cornefcus..., 2021a);
- после применения одного из стандартных алгоритмов фильтрации\* к «маскированному» на предыдущем шаге ряду данных;
- после фильтрации «маскированных» данных с использованием подхода по порогу ПЗО в районе наблюдаемого объекта.

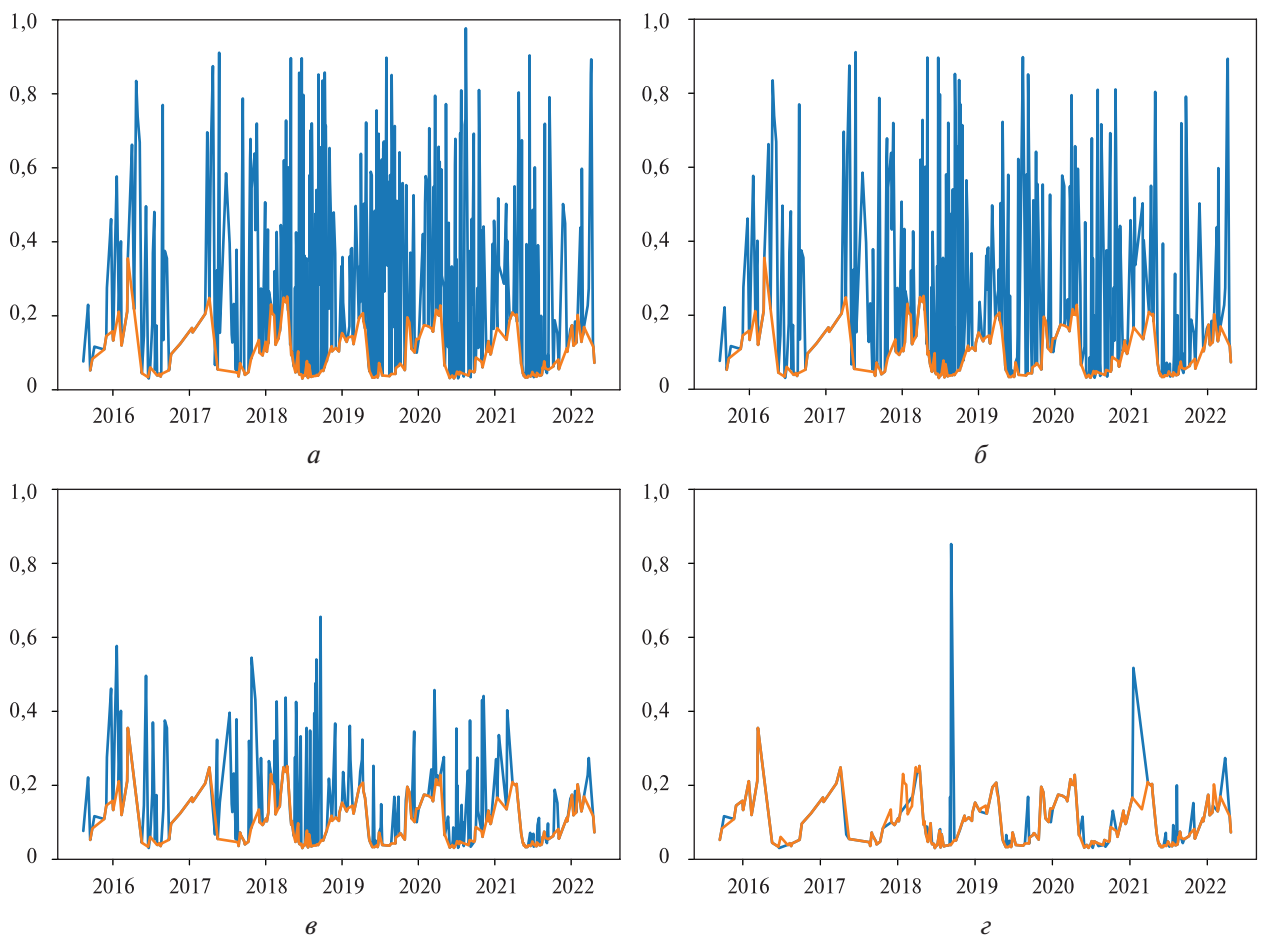


Рис. 8. Ряды со значениями конечного показателя для объекта, расположенного в регионе с большим числом облачных дней: а — до фильтрации; б — после «маскирования» по маске облачности; в — после фильтрации выбросов для «маскированных» данных; г — после фильтрации «маскированных» данных по порогу облачности. Оранжевый цвет — очищенные вручную ряды; синий цвет — ряды на различных этапах фильтрации

\* Фильтрация выбросов проводилась на основе расчёта верхней границы доверительного интервала значений показателя внутри скользящего окна и затем отбрасыванием всех значений показателя, превосходящих эту границу.



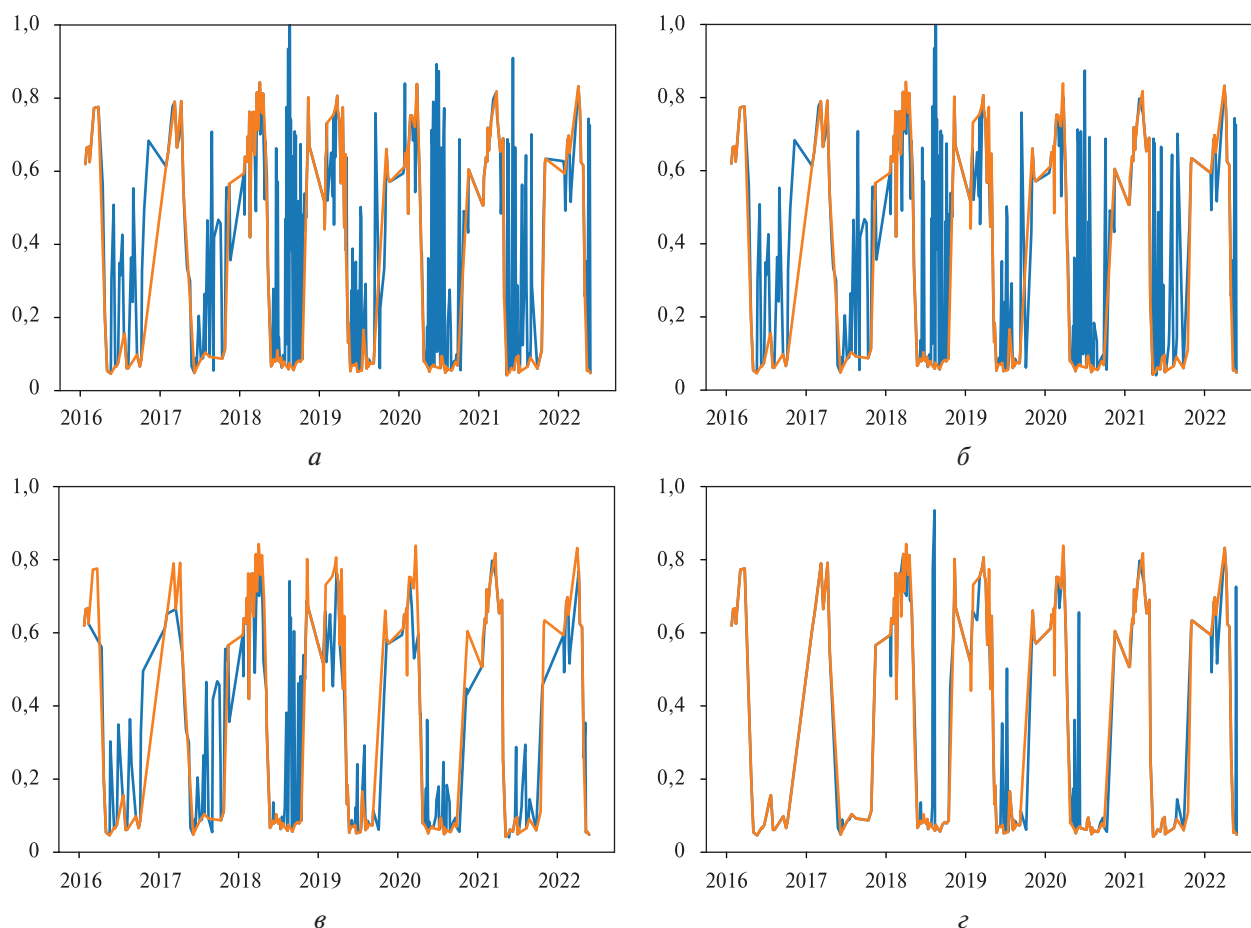


Рис. 9. Ряды со значениями конечного показателя для объекта, расположенного в регионе со значительным числом облачных дней (в легенде, представленной в подписи к рис. 8)

В таблице приведены доли суммарной ошибки (в процентах), полученной в результате применения двух способов фильтрации для рассматриваемых объектов.

Доли ИО, рассчитанной по результатам использования различных видов фильтрации в приведённых выше примерах

Объект	Количество изображений	Фильтрация выбросов по «стандартному алгоритму», %	Фильтрация по порогу, %
1	394	29	11
2	344	44	17
3	129	28	12

Таким образом, приведённые примеры подтверждают достаточно высокую эффективность работы предложенного алгоритма при выборе порога фильтрации из рассматриваемого диапазона (10–20 % облачности). При этом следует отметить, что ошибки в работе данного алгоритма фильтрации в целом соответствуют полученным выше оценкам. Приведённые примеры наглядно демонстрируют, что в случае, когда алгоритм выделения облачности работает со значительными ошибками, предложенный алгоритм фильтрации обеспечивает достаточно качественную фильтрацию сбойных (облачных) измерений и может давать лучшие результаты по сравнению со стандартными алгоритмами. Также следует отметить, что остающееся после работы предложенного алгоритма незначительное число сбойных измерений может быть вполне эффективно отфильтровано различными алгоритмами фильтрации выбросов.

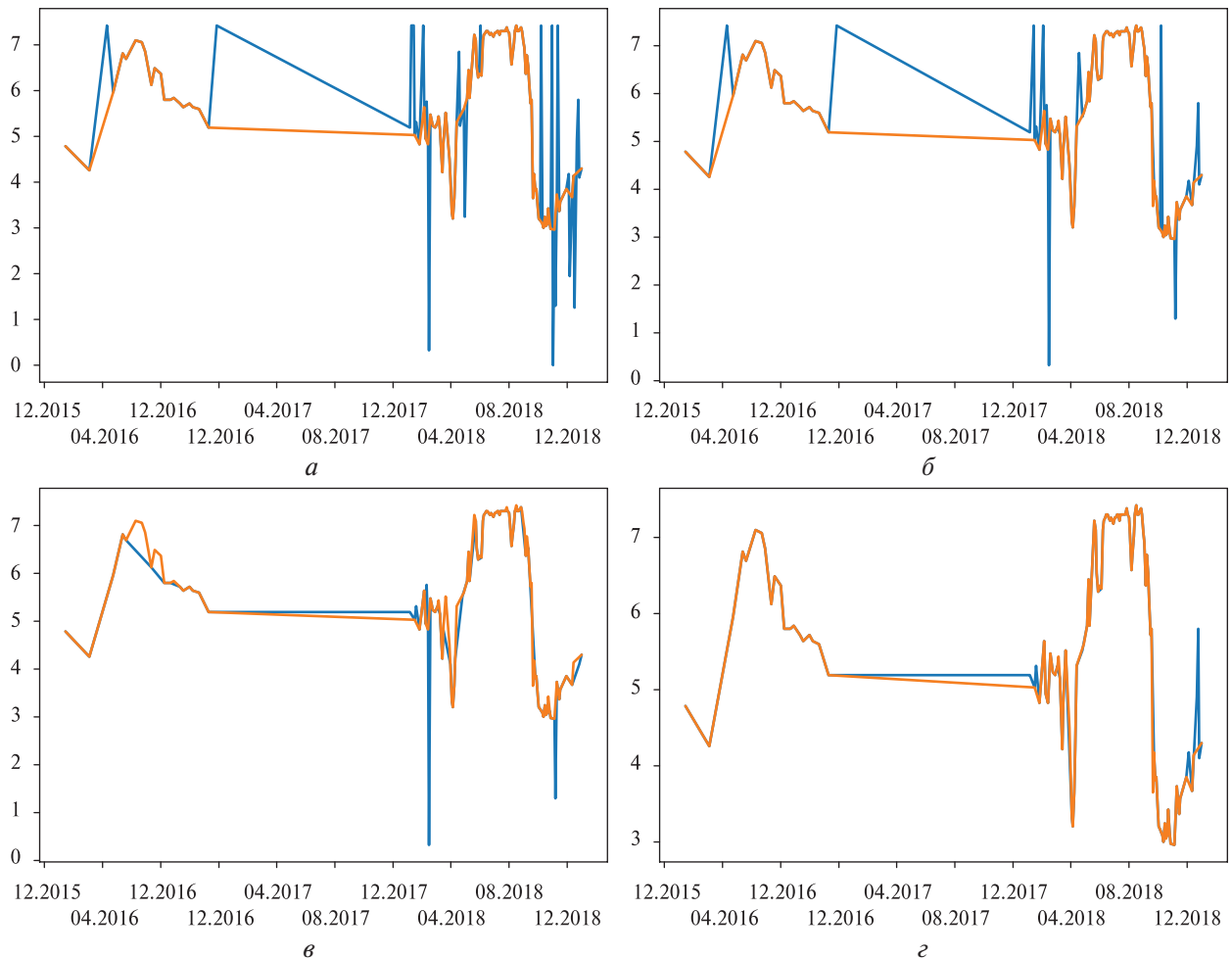


Рис. 10. Ряды со значениями конечного показателя для объекта, расположенного в регионе с малым количеством облачных дней (в легенде, представленной в подписи к рис. 8)

## Заключение

Таким образом, предлагаемый алгоритм фильтрации по порогу облачности показал хорошие результаты на различных выборках и позволяет проводить фильтрацию с ошибками порядка 10–20 %. Алгоритм является достаточно устойчивым к использованию различных подходов к построению масок облачности. Показано, что для практической реализации алгоритма может выбираться порог ПЗО в диапазоне 10–20 %, который позволяет получать хорошие результаты фильтрации в большом диапазоне регионов, в которых расположены наблюдаемые объекты.

Исследование выполнено с использованием возможностей ЦКП «ИКИ-Мониторинг» (Лупян и др., 2019), развиваемого и поддерживаемого в рамках темы «Мониторинг» Министерства науки и высшего образования Российской Федерации (госрегистрация № 122042500031-8). Работа выполнялась при поддержке проекта Российского фонда фундаментальных исследований (проект № 19-37-90114).

## 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. Константинова А. М., Балашов И. В., Кашицкий А. В., Лупян Е. А. Унифицированная технология дистанционного мониторинга природных и антропогенных объектов // Современные

- проблемы дистанционного зондирования Земли из космоса. 2021. Т. 18. № 4. С. 41–52. DOI: 10.21046/2070-7401-2021-18-4-41-52.
3. *Лулян Е. А., Прошин А. А., Бурцев М. А., Кашицкий 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., Сычугов 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.
  4. *Лулян Е. А., Константинова А. М., Балашов И. В., Кашицкий А. В., Саворский В. П., Панова О. Ю.* Разработка системы анализа состояния окружающей среды в зонах расположения крупных промышленных объектов, хвостохранилищ и отвалов // Современные проблемы дистанционного зондирования Земли из космоса. 2020. Т. 17. № 7. С. 243–261. DOI: 10.21046/2070-7401-2020-17-7-243-261.
  5. *Мерков А. Б.* Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
  6. *Мухамеджанов И. Д., Константинова А. М., Лулян Е. А., Умирзаков Г. У.* Оценка возможностей спутникового мониторинга динамики речного стока на примере анализа состояния реки Амударьи // Современные проблемы дистанционного зондирования Земли из космоса. 2022. Т. 19. № 1. С. 87–103. DOI: 10.21046/2070-7401-2022-19-1-87-103.
  7. *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.
  8. *Coluzzi R., Imbrenda V., Lanfredi M., Simoniello T.* A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses // Remote Sensing of Environment. 2018. V. 217. P. 426–443. DOI: 10.1016/j.rse.2018.08.009.
  9. Copernicus Sentinel-2 pre-Collection-1 MSI Level-1C (L1C) (2021a). European Space Agency, 2021. [https://doi.org/10.5270/S2\\_-d8we2fl](https://doi.org/10.5270/S2_-d8we2fl).
  10. Copernicus Sentinel-2 pre-Collection-1 MSI Level-2A (L2A) (2021b). European Space Agency, 2021. [https://doi.org/10.5270/S2\\_-6eb6imz](https://doi.org/10.5270/S2_-6eb6imz).
  11. *Dozier J.* Spectral signature of alpine snow cover from the Landsat Thematic Mapper // Remote Sensing of Environment. 1989. V. 28. P. 9–22. DOI: 10.1016/0034-4257(89)90101-6.
  12. *Luo Y., Guan K., Peng J., Wang S., Huang Y.* STAIR 2.0: A Generic and Automatic Algorithm to Fuse Modis, Landsat, and Sentinel-2 to Generate 10 m, Daily, and Cloud-/Gap-Free Surface Reflectance Product // Remote Sensing. 2020. V. 12(19). Art. No. 3209. 21 p. DOI: 10.3390/rs12193209.
  13. *Zhu Z., Wang S., Woodcock C.* Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images // Remote Sensing of Environment. 2015. V. 159. P. 269–277. DOI: 10.1016/j.rse.2014.12.014.

## Cloud data automatic filtering algorithm for object remote monitoring

**A. M. Konstantinova**

*Space Research Institute RAS, Moscow 117997, Russia  
E-mail: konstantinova.anouk@gmail.com*

The paper describes an approach to automatic filtering of cloud data to be used by the object monitoring technology developed at the Space Research Institute (IKI). The technology was created to monitor and control the dynamics of various natural and anthropogenic objects (homogeneous formations). It provides automatic creation and analysis of long-term series of various characteristics of observed objects obtained from satellite data. One of the problems in the creation of such series is the organization of automatic filtering of incorrect data, which is associated, first of all, with the presence of clouds covering the object in the satellite images. The organization of such filtering can be difficult in cases when it is not possible to provide a reliable automatic selection of clouds. The paper presents an algorithm that allows automatic filtering in such a situation. It is based on an integral analysis of cloudiness and the search of percentage thresholds of cloudiness coverage in the area of the observed object

instead of cloudiness analysis of individual pixels. The values of object characteristics calculated from data with cloudiness coverage percentage higher than the defined threshold would almost certainly be false. The paper provides an analysis and presents the possibility of optimal thresholds definition which allow to filter out a maximum number of false data along with losing a minimum number of cloudless data. The paper also gives examples of using the proposed filtering algorithm to derive series of characteristics of objects located in regions with different observation conditions. Also, the results of the proposed filtering method are compared with the results of other filtering methods, in particular, noise filtering in a series of object characteristics.

**Keywords:** remote sensing, natural objects, technology, object monitoring, cloud filtering, noise filtering, IKI-Monitoring Center for Collective Use

Accepted: 25.08.2022

DOI: 10.21046/2070-7401-2022-19-4-88-99

## References

1. Bartalev S., Egorov V., Zharko V., Loupian E., Plotnikov D., Khvostikov S., Shabanov N., *Sputnikovoe kartografirovaniye rastitel'nogo pokrova Rossii* (Land cover mapping over Russia using Earth observation data), Moscow: IKI RAN, 2016, 208 p. (in Russian).
2. Konstantinova A. M., Balashov I. V., Kashnitskiy A. V., Loupian E. A., Unified technology for remote monitoring of natural and anthropogenic objects, *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2021, Vol. 18, No. 4, pp. 41–52 (in Russian), DOI: 10.21046/2070-7401-2021-18-4-41-52.
3. Loupian E. A., Proshin A. A., Bourtsev M. A., Kashnitskiy 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., Sychugov 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, *Sovremennyye 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.
4. Loupian E. A., Konstantinova A. M., Balashov I. V., Kashnitskiy A. V., Savorsky V. P., Panova O. Yu., Development of a system for analyzing the state of environment in areas of large industrial facilities, tailings and dumps, *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2020, Vol. 17, No. 7, pp. 243–261 (in Russian), DOI: 10.21046/2070-7401-2020-17-7-243-261.
5. Merkov A. B., *Raspoznavaniye obrazov. Vvedeniye v metody statisticheskogo obucheniya* (Automatic cognition. Introduction to statistical learning methods), 2011, 256 p. (in Russian).
6. Muhamedzhanov I. D., Konstantinova A. M., Loupian E. A., Umirzakov G. U., Evaluation of satellite monitoring capabilities of stream runoff based on the Amu Darya River state analysis, *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2022, Vol. 19, No. 1, pp. 87–103 (in Russian), DOI: 10.21046/2070-7401-2022-19-1-87-103.
7. 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.
8. Coluzzi R., Imbrenda V., Lanfredi M., Simoniello T., A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses, *Remote Sensing of Environment*, 2018, Vol. 217, pp. 426–443, DOI: 10.1016/j.rse.2018.08.009.
9. *Copernicus Sentinel-2 pre-Collection-1 MSI Level-1C (L1C)* (2021a), European Space Agency, 2021, [https://doi.org/10.5270/S2\\_-d8we2fl](https://doi.org/10.5270/S2_-d8we2fl).
10. *Copernicus Sentinel-2 pre-Collection-1 MSI Level-2A (L2A)* (2021b), European Space Agency, 2021, [https://doi.org/10.5270/S2\\_-6eb6imz](https://doi.org/10.5270/S2_-6eb6imz).
11. Dozier J., Spectral signature of alpine snow cover from the Landsat Thematic Mapper, *Remote Sensing of Environment*, 1989, Vol. 28, pp. 9–22, DOI: 10.1016/0034-4257(89)90101-6.
12. Luo Y., Guan K., Peng J., Wang S., Huang Y., STAIR 2.0: A Generic and Automatic Algorithm to Fuse Modis, Landsat, and Sentinel-2 to Generate 10 m, Daily, and Cloud-/Gap-Free Surface Reflectance Product, *Remote Sensing*, 2020, Vol. 12(19), Art. No. 3209, 21 p., DOI: 10.3390/rs12193209.
13. Zhu Z., Wang S., Woodcock C., Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images, *Remote Sensing of Environment*, 2015, Vol. 159, pp. 269–277, DOI: 10.1016/j.rse.2014.12.014.