Исследование влияния сжатия данных с потерями алгоритмом LERC на применимость данных ДЗЗ для решения задач мониторинга

А. А. Прошин, Е. А. Лупян, М. А. Бурцев, К. А. Трошко, А. В. Кашницкий

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

Рассматривается влияние сжатия спутниковых данных ДЗЗ с потерей качества с помощью алгоритма LERC на качество и применимость в реальных задачах получаемых на их основе продуктов. Для оценки такого влияния в работе была оценена максимально допустимая ошибка, которую можно внести при сжатии, создана и сжата алгоритмом LERC тестовая выборка данных, представляющая собой временной ряд годовой продолжительности, после чего был проведён ряд сравнений. По исходному ряду и сжатому с потерями ряду сравнивались значения индекса NDVI (*англ.* Normalized Difference Vegetation Index) по отдельным сценам, качество восстановления ежедневного временного ряда и полученные по восстановленным рядам индексы NDVI, а также оценивалось качество решения задачи классификации объектов с использованием классификатора Random Forest по построенным временным рядам. Все эксперименты продемонстрировали почти полное совпадение расчётов между собой с ошибкой, не превышающей погрешности самих данных. Показано, что использование алгоритма LERC, позволяющего контролировать вносимую ошибку в данные, при условии соблюдения критериев сохранности качества данных, даёт возможность получать результаты обработки и классификации, сопоставимые с найденными по исходным данным, при выигрыше в объёме хранимых данных примерно в 2,5 раза.

Ключевые слова: ДЗЗ, сжатие данных, LERC, сжатие с потерями, Random Forest

Одобрена к печати: 23.10.2025
DOI: 10.21046/2070-7401-2025-22-5-63-72

Введение

Для решения различных задач, связанных с мониторингом разных объектов и процессов на поверхности Земли, одним из основных инструментов являются данные дистанционного зондирования Земли (ДЗЗ). На сегодня бурно развиваются возможности космических аппаратов (КА) ДЗЗ и их съёмочной аппаратуры, растёт пространственное разрешение данных и число доступных спектральных каналов, улучшается чувствительность приборов и другие их характеристики, что приводит к повышению информативности и росту размера данных. Также резко увеличивается и количество КА на орбите, данные с которых становятся доступными для использования (Лупян и др., 2018; Zhu et al., 2018). Помимо этого, бурно развиваются алгоритмы и методы обработки данных. Так, одним из наиболее востребованных типов данных для анализа и дальнейшей обработки становятся не исходные снимки, которые могут давать чистые наблюдения заданной территории со слабо предсказуемой периодичностью от нескольких дней до недель, что зависит от состояния облачности над территорией, а восстановленные ряды данных, способные предоставлять информацию с ежедневной периодичностью (Plotnikov et al., 2022). Конечно, такие ряды также занимают существенно больший объём по сравнению с исходными данными, на базе которых они были рассчитаны. В итоге сочетание выделенных факторов приводит к тому, что объёмы данных в архивах специализированных центров обработки и распространения ежегодно увеличиваются на петабайтные величины. Так, объём архивов центра коллективного пользования «ИКИ-Мониторинг» (http://ckp.geosmis.ru/) (Лупян и др., 2015, 2019) на сегодня превышает 9 Пб и ежегодно пополняется больше, чем на петабайт. В силу этого крайне актуальной становится проблема эффективного сжатия данных без потери их свойств и пригодности к дальнейшей обработке.

На сегодня реализовано большое количество алгоритмов сжатия, используемых для таких данных (Becker et al., 2016). Они делятся на алгоритмы с потерей и без потери качества. В областях, связанных с обработкой данных ДЗЗ, как правило, применяются алгоритмы без потери качества, чтобы избежать искажения данных при расчётах. Наибольшей степенью сжатия среди таких алгоритмов обладает JPEG2000 (*англ.* Joint Photography Expert Group, объединённая группа экспертов по фотографии), однако очень низкая скорость декомпрессии данных при чтении сильно сужает сферу его применения. Большинство остальных алгоритмов построено на базе алгоритмов Лемпеля – Зива – Уэлча и кодов Хаффмана (Oswal et al., 2016) и в среднем позволяет сжимать спутниковые данные примерно в полтора-два раза в зависимости от их типа и разрядности. Использование классических алгоритмов сжатия с потерей качества, таких как JPEG, даёт возможность добиться сжатия в 10–15 раз и более, но при этом приводит к существенным и неконтролируемым искажениям данных, что делает такие алгоритмы применимыми только для задач визуализации и представления данных, но не их обработки.

Тем не менее задача создания алгоритма сжатия с контролируемыми потерями качества оставалась актуальной, например, для эффективного хранения данных с избыточной точностью, и в 2015 г. компанией ESRI был представлен новый алгоритм сжатия LERC (*англ.* Limited Error Raster Compression, алгоритм сжатия растра с ограниченной ошибкой) (https://esri.github.io/lerc/) (Becker et al., 2015), который позволяет задать предельный порог вносимой в значение каждого пикселя ошибки при сжатии (MaxError), тем самым сохраняя радиометрические свойства изображения с заданной же погрешностью. Алгоритм основан на том, что данные разбиваются на небольшие блоки, в каждом из блоков маскируются незначащие пиксели, минимальное значение среди оставшихся значимых пикселей берётся за основу, вычисляется разность остальных значимых пикселей с ним, найденная разность делится на 2MaxError (где MaxError — заданный уровень допустимой ошибки) и округляется до целого, после чего полученный блок сжимается без потерь. В результате такого преобразования большая часть значений пикселей изменяется, но не более чем на величину установленного порога допустимой ошибки. Даже при очень небольших значениях ошибки он обеспечивает сжатие, которое в разы эффективнее стандартных алгоритмов без потерь (https://kokoalberti.com/articles/geotiff-compression-optimization-guide/), но может приводить к ухудшению радиометрического и структурного качества данных. На основании ранее проведённых исследований в ИКИ РАН была выработана универсальная методика определения максимальной ошибки LERC для различных типов спутниковых данных. Её использование позволяет сохранить на приемлемом уровне качественные показатели данных и сократить объёмы хранения примерно в три раза по сравнению с применением стандартных алгоритмов сжатия без потерь (Прошин и др., 2024). Однако для практического использования такого подхода не менее важно оценить, насколько сжатие с такой, номинально малозначимой ошибкой повлияет на точность результатов обработки сжатых данных и качество получаемых на их основе производных продуктов.

В работе оценивается влияние сжатия исходных канальных спутниковых данных ДЗЗ с помощью алгоритма LERC на качество и применимость для реальных задач, получаемых на их основе продуктов.

Методика исследования

Для оценки возможности применения алгоритма LERC к данным ДЗЗ и его влияния на качество и допустимость их использования в обработке проведён ряд тестов, для которых была сформирована тестовая выборка из данных приборов MSI (*англ.* Multispectral Imager, многоспектральное съёмочное устройство) спутников Sentinel-2 в красном и ближнем инфракрасном спектральных каналах по территории Рязанской области за 2024 г. Такой выбор обусловлен тем, что именно эти каналы наиболее часто используются в задачах анализа состояния растительности. На их основе рассчитывается один из наиболее популярных индексов, харак-

теризующих состояние растительности, — нормализованный разностный вегетационный индекс, он же NDVI (*англ.* Normalized Difference Vegetation Index). Он вычисляется по формуле NDVI = (NIR – RED)/(NIR + RED), где NIR — значение в ближнем инфракрасном (*англ.* near infrared) канале, RED — значение в красном канале, NDVI — значение результирующего индекса. Если значения входных параметров в этой формуле близки или малы, результат может сильно меняться даже при их небольших изменениях, вносимых при сжатии с потерями. Поэтому оценка поведения именно таких индексных продуктов представляет наибольший интерес при определении возможности применения сжатых с потерей данных в реальных задачах.

В рамках исследования для подобранной выборки был определён порог допустимой ошибки для каждого спектрального канала с использованием полученного в работе (Прошин и др., 2024) критерия. На основе данных, сжатых с определённым порогом потерь, рассчитаны значения индекса NDVI, которые были попиксельно сопоставлены со значениями NDVI, полученными по исходным данным, для оценки влияния вносимой в исходные данные ошибки на найденные индексы.

Затем были вычислены два ряда восстановленных ежедневных индексов NDVI на основе исходных и сжатых с потерями канальных данных соответственно (Plotnikov et al., 2022) за весь 2024 г. При визуальном сравнении индексов, полученных по сжатым и несжатым канальным данным, в некоторых случаях наблюдается заметная разница в мелкомасштабной текстуре, а также в значениях отдельных пикселей. Необходимо было установить, насколько наблюдаемые отличия влияют на качество получаемого по сжатым данным ежедневного индекса NDVI или же отличия находятся в пределах его ошибки.

Для того чтобы сравнить качество полученных восстановленных рядов индексов, их значения за даты и по территориям, по которым имелись безоблачные исходные фрагменты данных, были сопоставлены со значениями в соответствующих фрагментах несжатых данных. Существенно, что методика построения восстановленных рядов основана на полиномиальной интерполяции и колебания входных значений в определённых условиях потенциально могут привести к искажению результата восстановления. Таким образом, требовалось оценить, отличается ли точность значений восстановленного ряда данных от исходных или же отклонения лежат в пределах ошибки алгоритма интерполяции.

После оценки качества восстановления ряда требовалось оценить применимость полученных данных к реальным задачам. Одним из наиболее используемых классов задач, в которых применяются данные ДЗЗ, являются задачи классификации объектов и определения их характеристик, причём как попиксельно, так и интегрально на объект. В качестве примера такой задачи была взята задача определения основной культуры на заданных сельскохозяйственных полях. В её рамках проведены серии классификаций по восстановленным рядам, полученным на основе сжатых с потерями и неискажённых данных, результаты классификаций попиксельно сопоставлялись между собой. В качестве дополнительного критерия обе серии результатов были пообъектно сопоставлены с эталонной картой преобладающих культур, полученной на заданные поля экспертным способом.

Ниже приводится краткая информация о выработанном критерии для определения максимальной ошибки LERC, а затем описываются результаты работ по описанному плану исследования.

Критерий выбора допустимой ошибки LERC

Кратко рассмотрим критерий выбора допустимой ошибки, вносимой алгоритмом LERC, полученный ранее в работе (Прошин и др., 2024). При выборе пределов такой ошибки необходимо учитывать два фактора: сохранность радиометрических характеристик данных и сохранность их пространственной структуры. Для оценки с точки зрения первого фактора можно исходить из известных значений погрешности измерений съёмочных систем, а также алгоритмов обработки данных.

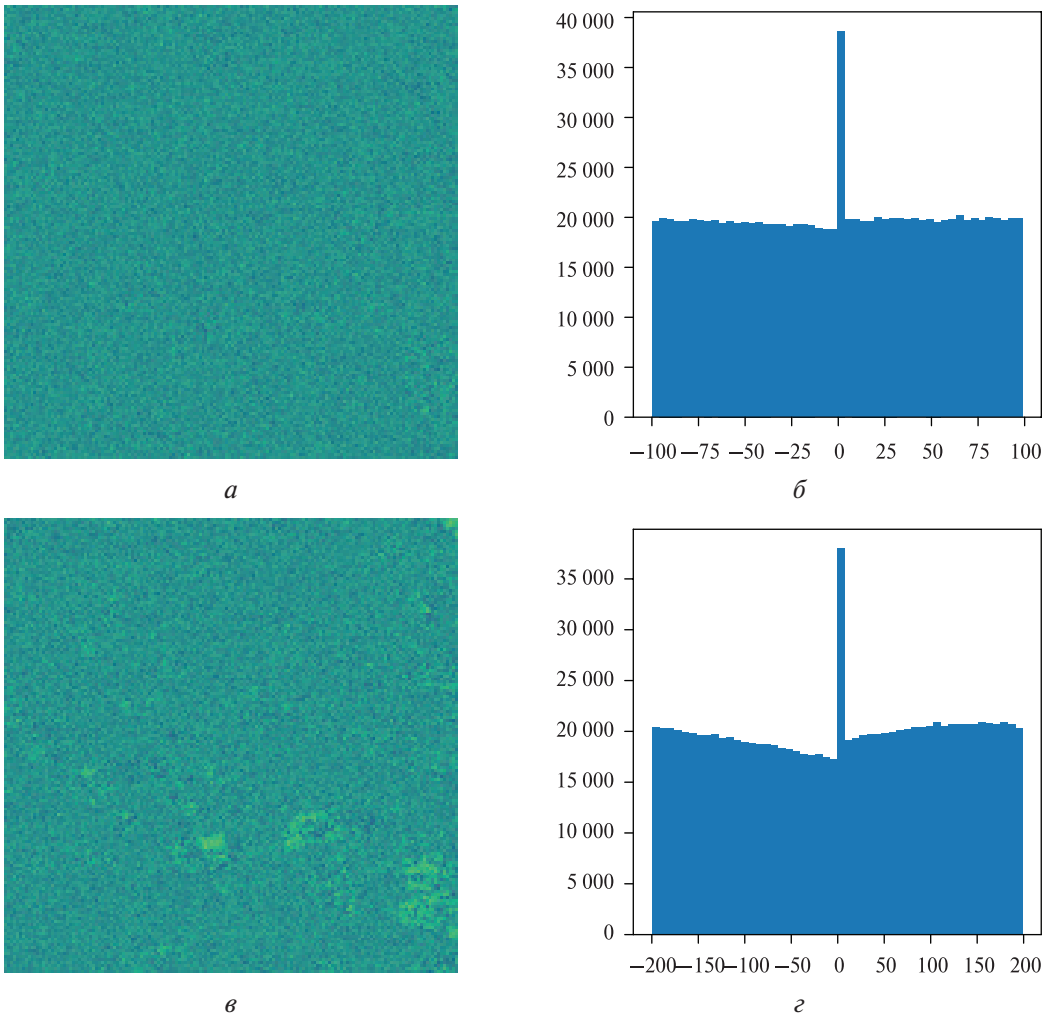

Рис. 1. Разности между сжатыми и несжатыми изображениями и их гистограммы: *а* — разность при ошибке в 1 %; *б* — гистограмма разности при ошибке в 1 %; *в* — разность при ошибке в 2 %; *г* — гистограмма разности при ошибке в 2 %

Что касается пространственной структуры, специфика работы алгоритма сжатия LERC, оперирующего блоками изображений фиксированного размера, такова, что искажения могут вноситься в текстуру изображения, т. е. в геометрические свойства объектов на нём. Для оценки влияния уровня вносимой ошибки на текстуру изображения в работе (Прошин и др., 2024) была проанализирована серия разностей между исходными и искажёнными изображениями с разной степенью искажений. Признаком того, что в искажённом изображении изменилась структура относительно исходного, стало наличие явной структуры в полученной разности. При отсутствии искажений структуры разность была близка к шуму. В качестве формализованного признака наличия привнесённой структуры на разностях анализировались гистограммы их значений. Изображению, близкому к шуму, соответствует симметричная гистограмма с практически равномерным распределением значений разностей во всём диапазоне от минус до плюс максимальной ошибки. Ниже рассматривается пример зависимости разницы между исходным и сжатым изображением и гистограммы её значений для данных восстановленного индекса от ошибки, для удобства приведённой в виде доли (в %) от диапазона значений. Для ошибки в 1 % разница изображений визуально похожа на белый шум, а гистограмма почти симметрична. Но уже для ошибки в 2 % в разнице начинают проявляться артефакты и гистограмма становится искажённой. На *рис. 1* представлены примеры таких разностей между изображениями и их гистограмм при ошибках в 1 и 2 % соответственно.

Для учёта обоих факторов был выработан универсальный критерий выбора допустимой максимальной ошибки, при которой выполнялись бы требования по сохранению радиометрической точности и пространственной структуры данных, т. е. вносимые искажения не влияли бы на результаты обработки и анализа сжатых изображений. В качестве такого критерия было взято среднее значение разностей между исходными и искажёнными данными, установленное по большой выборке из не менее чем 500 случайных фрагментов данных, полученных по разным территориям и за разные времена.

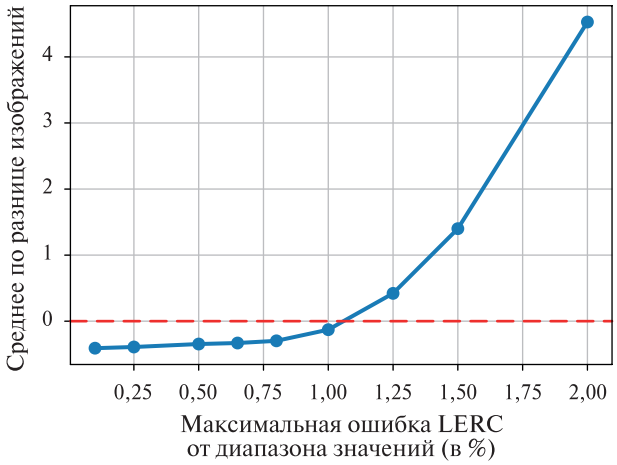

Рис. 2. Зависимость средних значений по разностям изображений от ошибки LERC

На *рис. 2* представлена зависимость агрегированного по тестовой выборке среднего значения по разницам между исходными и сжатыми изображениями от максимальной ошибки LERC. Видно, что при малых значениях ошибки LERC среднее уходит в отрицательную область. Такой эффект связан с наличием на исходном изображении однородных областей, для которых большинство значений попадает в диапазон от одной до двух ошибок от минимального значения в блоке и, соответственно, завышается при сжатии. Занижение значений в сжатом изображении при больших ошибках LERC связано с вкладом областей, для которых большинство пикселей находится в пределах одной ошибки от минимального значения в блоке. Хорошо видно, что после пересечения средним по разности нулевого значения скорость роста среднего в зависимости от ошибки существенно увеличивается, поэтому в качестве порогового значения для допустимой ошибки LERC было выбрано значение, при котором агрегированное среднее значение по разности ближе всего к нулю. В этом случае эффекты занижения и завышения значений в пикселях в результате применения алгоритма LERC максимально компенсируют друг друга. Более подробно обоснование критерия изложено в статье (Прошин и др., 2024).

Влияние сжатия канальных данных на применимость восстановленного индекса NDVI

Как было сказано в разд. «Методика исследования», в настоящей работе исследовалось влияние ошибки сжатия канальных данных алгоритмом LERC на получаемые значения индекса NDVI по исходным данным, качество восстановления ежедневных рядов индекса NDVI, а также применимость полученных результатов для решения задач классификации объектов.

Для созданной тестовой выборки за 2024 г. на основе приведённого выше критерия для канала NIR значение максимальной ошибки LERC было определено как 60 (сжатие примерно в 2,7 раза), а для канала RED — 40 (сжатие в 2,8 раза). Для учёта влияния шумов на данных, обусловленных наличием водных объектов, снега, облаков и их теней, все сравнения проводились только по чистым участкам территории. Для этого были использованы маски классификации SCL (*англ.* Scene classification, классификация сцены), входящие в состав стандартного атмосферно скорректированного продукта MSI L2A (*англ.* Level 2a, уровень обработки 2a) и позволяющие отделить чистые области на изображениях.

После фильтрации данных для чистых областей по выборкам исходных и искажённых данных были рассчитаны индексы NDVI и попиксельно сопоставлены между собой для каждой пары изображений в тестовых выборках. По результатам сопоставления были рассчитаны средние разности значений и их стандартное отклонение. Гистограмма полученных значений стандартного отклонения приведена на *рис. 3* (см. с. 68).

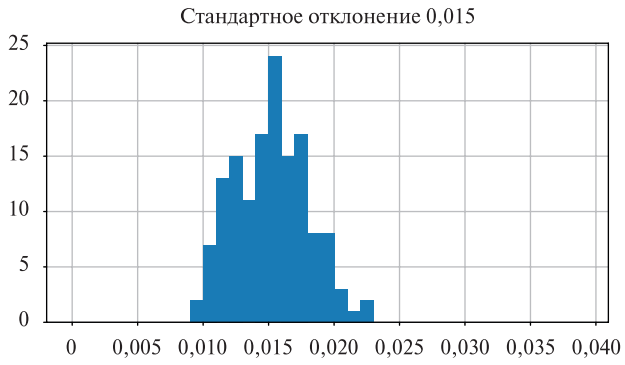

Рис. 3. Гистограмма стандартного отклонения по разницам индексов NDVI, полученным по исходным и сжатым алгоритмом LERC канальным данным

Из неё видно, что среднее стандартное отклонение по всем сценам равно примерно 0,015, что составляет 1,5 % от диапазона возможных значений и не превышает точности исходных данных, которая достигает от 3 до 5 % (см. отчёт о качестве данных продуктов уровня обработки L1C приборов MSI от марта 2023 г., https://sentinels.copernicus.eu/documents/247904/4868341/OMPC.CS.DQR.001.02-2023+-+i85r0+-+MSI+L1C+DQR+March+2023.pdf). Таким образом, можно считать, что сжатие исходных данных с ошибкой, подобранной по полученному критерию, не приводит к ухудшению найденных индексов NDVI.

После сопоставления между собой индексов NDVI, полученных по отдельным сценам данных, необходимо было оценить, как сжатие с потерями влияет на качество восстановления ежедневных рядов данных, так как именно такие ряды широко применяются в задачах мониторинга. По тестовой выборке для исходных данных (далее ряд NOLERC) и данных, сжатых с потерями (далее ряд LERC), были рассчитаны ряды таких восстановленных индексов NDVI и также попиксельно сопоставлены со значениями индекса NDVI, полученными по отфильтрованным исходным канальным данным. По результатам сопоставления были рассчитаны средние разности значений и их стандартное отклонение. В первом случае среднее по всем фрагментам стандартное отклонение составило около 0,034, т. е. 3,4 % от диапазона значений данных (*рис. 4*), а во втором — 0,036 или 3,6 % (*рис. 5*). Таким образом, использование сжатых с потерями канальных данных для получения восстановленного ряда практически не повлияло на качество восстановления, собственная погрешность которого составляет около 5 % (Plotnikov et al., 2022).

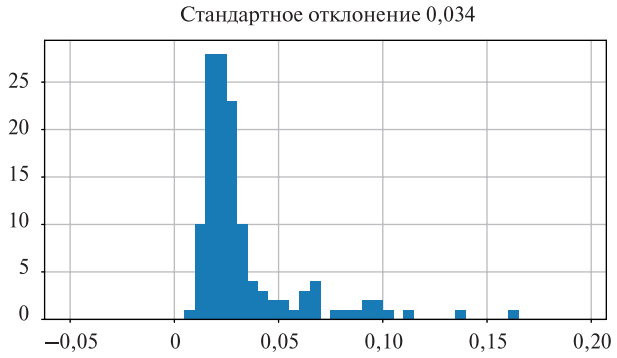

Рис. 4. Сравнение индексов по исходным канальным данным со значениями ряда NOLERC

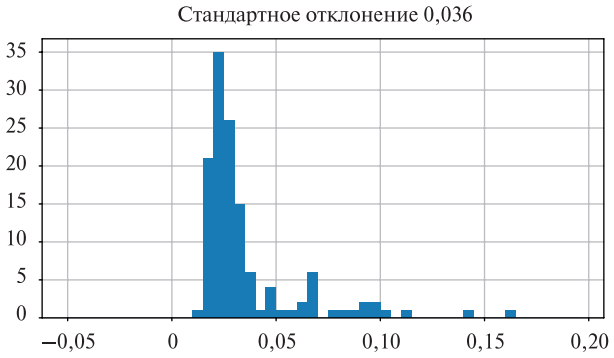

Рис. 5. Сравнение индексов по исходным канальным данным со значениями ряда LERC

Несмотря на то, что прямое сопоставление полученных рядов между собой показало их очень хорошее совпадение, необходимо было оценить, насколько сжатие с потерями может повлиять на решение задач классификации объектов и определения их характеристик на основе анализа временных рядов. В качестве такой характерной задачи, для которой применяются подобные ряды, выбрана задача классификации сельскохозяйственных полей для определения типа их использования, а также преобладающей культуры на них.

Для проведения такой классификации был использован метод, основанный на классификаторе RF (*англ.* Random Forest, метод случайного леса) (Денисов и др., 2024). Входными дан-

ными для него являются восстановленные ежедневные индексы NDVI с начала года до даты, на которую проводится классификация. Все имеющиеся данные за этот диапазон дат собираются в единый стек, прореживаются до каждого пятого дня, и на основе этого стека формируются векторы значений NDVI в каждом пикселе. Эти векторы и становятся основными признаками для работы классификатора. Классификация проводилась по территории Кораблинского района Рязанской области по данным за период с 20 апреля по 21 октября 2024 г.

Предварительно для проведения классификации на основе экспертного анализа спутниковых данных, доступных в системе «Вега» (цветосинтезированных изображений, в том числе многовременных, а также графиков сезонного хода NDVI), на часть района была сформирована векторная карта культур, использующаяся как для обучения, так и для оценки достоверности результатов классификации. В карту было включено 364 поля общей площадью 38 тыс. га, каждому из которых присвоен один из десяти классов: озимые зерновые, рапс озимый, яровые зерновые и зернобобовые (без кукурузы), соя, кукуруза, рапс яровой, подсолнечник, многолетние травы посева прошлых лет, чистый пар, неиспользуемые земли. В качестве обучающей выборки использовалось по 20 % от площади каждого класса. Для оценки достоверности результатов классификации была выполнена растеризация карты культур.

Непосредственно попиксельная классификация изображений проводилась в пределах границ полей с использованием специализированного инструмента, реализованного в системе «Вега» (Денисов и др., 2024). Для данных обоих типов (LERC и NOLERC) классификация была выполнена многократно для оценки устойчивости получаемых результатов.

Вследствие попиксельного сравнения итогов классификации с растеризованной картой культур были найдены следующие значения общей достоверности: для ряда LERC — 86,9–87,0 %, для ряда NOLERC — 86,8–87,0 %. Сопоставление результатов классификации по рядам LERC и NOLERC между собой показало совпадение на 95,2–95,3 %.

Также было проведено сравнение, насколько точно по этим рядам определяется основная культура сельскохозяйственных полей. Для этого результаты классификации преобладающих культур были интегрированы на поля и проведено сопоставление полученных интегральных оценок с эталонной картой. В этом случае результаты также оказались очень близкими: преобладающая культура совпала для 91,4 % полей, классифицированных по ряду LERC, и для 91 % — по ряду NOLERC.

Таким образом, на основе полученных результатов сравнений можно утверждать, что применение алгоритма сжатия LERC к данным ДЗЗ при условии соблюдения критериев подбора максимально допустимой ошибки не приводит к дополнительным ошибкам при их использовании в решении задач мониторинга, в том числе для задач восстановления рядов данных и классификации объектов.

Заключение

В работе представлена оценка применимости спутниковых данных, сжатых с потерей качества алгоритмом LERC, для проведения обработки и анализа на примере восстановления ежедневных рядов интерполированных индексов NDVI и их использования в задачах классификации объектов.

Результаты исследования показали, что при выборе максимально допустимой ошибки сжатия с применением критерия, обеспечивающего сохранение как радиометрической точности, так и пространственной структуры изображения, результаты тематической обработки существенно не изменяются и остаются в пределах своих погрешностей. Так, при формировании восстановленных ежедневных рядов индекса NDVI среднее отклонение от реальных измерений выросло с 0,034 до 0,036, т. е. изменилось на уровне шума. При проведении операций классификации объектов результаты также оказались практически идентичными для рядов, полученных по исходным и по сжатым с потерями данным.

Таким образом, установлено, что сжатие данных с потерями алгоритмом LERC при соблюдении описанного критерия позволяет существенно уменьшить объём хранимых

в архиве данных — примерно в 2,5 раза. Так, при использовании такого сжатия для архивов данных прибора MSI уровня обработки L1C, накопленных в архивах Центра коллективного пользования (ЦКП) «ИКИ-Мониторинг», их объём может сократиться с примерно 2,6 до 1 Пб.

Более того, исходя из особенностей проанализированного индекса, а именно зависимости вносимой в него ошибки от значений в каналах, можно утверждать, что влияние сжатия с потерями как на исходные канальные данные, так и на восстановленные ряды канальных данных гораздо меньше, поэтому их применимость для решения различных задач не вызывает сомнений. Как было показано в предыдущей работе (Прошин и др., 2024), представленный подход к сжатию данных с потерями может быть применён и к данным других приборов в архивах ЦКП «ИКИ-Мониторинг». При этом использование его позволит не только существенно сократить объёмы хранимых данных по текущим областям интереса ЦКП, но и снизить накладные расходы на операции обработки, передачи и предоставления данных, тем самым ускорив их, а также потенциально существенно расширить зону интереса ЦКП без радикального увеличения необходимых мощностей хранения.

Исследование было выполнено на базе использования возможностей ЦКП «ИКИ-Мониторинг» и при поддержке темы Министерства науки и высшего образования Российской Федерации «Большие данные в космических исследованиях: астрофизика, Солнечная система, геосфера» (№ 122042500019-6).

Литература

1. *Денисов П. В.*, *Кашницкий А. В.*, *Колбудаев П. А.*, *Лупян Е. А.*, *Матвеев А. М.*, *Плотников Д. Е.*, *Рукевич Б. П.*, *Толпин В. А.*, *Трошко К. А.* Реализация на базе возможностей системы Вега-PRO технологии классификации сельскохозяйственных культур на полях по данным спутников Sentinel-2 // Материалы 11-й Международ. науч. конф. «Региональные проблемы дистанционного зондирования Земли». Красноярск: СФУ, 2024. С. 40–46.
2. *Лупян Е. А.*, *Прошин А. А.*, *Бурцев М. А.*, *Балашов И. В.*, *Барталев С. А.*, *Ефремов В. Ю.*, *Кашницкий А. В.*, *Мазуров А. А.*, *Матвеев А. М.*, *Суднева О. А.*, *Сычугов И. Г.*, *Толпин В. А.*, *Уваров И. А.* Центр коллективного пользования системами архивации, обработки и анализа спутниковых данных ИКИ РАН для решения задач изучения и мониторинга окружающей среды // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 5. С. 263–284.
3. *Лупян Е. А.*, *Бурцев М. А.*, *Прошин А. А.*, *Кобец Д. А.* Развитие подходов к построению информационных систем дистанционного мониторинга // Современные проблемы дистанционного зондирования Земли из космоса. 2018. Т. 15. № 3. С. 53–66. DOI: 10.21046/2070-7401-2018-15-3-53-66.
4. *Лупян Е. А.*, *Прошин А. А.*, *Бурцев М. А. и др.* 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.
5. *Прошин А. А.*, *Лупян Е. А.*, *Бурцев М. А.* Особенности использования алгоритма сжатия изображений LERC для архивации данных ДЗЗ // Современные проблемы дистанционного зондирования Земли из космоса. 2024. Т. 21. № 6. С. 130–142. DOI: 10.21046/2070-7401-2024-21-6-130-142.
6. *Becker P.*, *Plesea L.*, *Maurer T.* Cloud optimized image format and compression // Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2015. V. 40. P. 613–615. DOI: 10.5194/isprsarchives-XL-7-W3-613-2015.
7. *Becker P.*, *Plesea L.*, *Maurer T.* Optimizing cloud based image storage, dissemination and processing through use of MRF and LERC // Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016. V. 41. P. 201–203. DOI: 10.5194/isprsarchives-XLI-B4-201-2016.
8. *Oswal S.*, *Singh A.*, *Kumari K.* Deflate compression algorithm // Intern. J. Engineering Research and General Science. 2016. V. 4. No. 1. P. 430–436.
9. *Plotnikov D. E.*, *Kolbudaev P. A.*, *Matveev A. M. et al.* Daily surface reflectance reconstruction using LOWESS on the example of various satellite systems // 2022 8th Intern. Conf. Information Technology and Nanotechnology (ITNT). IEEE, 2022. 5 p. DOI: 10.1109/ITNT55410.2022.9848630.
10. *Zhu L.*, *Suomalainen J.*, *Liu J. et al.* A review: Remote sensing sensors // Multi-purposeful Application of Geospatial Data / eds. Rustamov R. B., Hasanova S., Zeynalova M. H. IntecOpen, 2018. P. 19–42. DOI: 10.5772/intechopen.71049.

Study of the impact of lossy data compression using the LERC algorithm on the applicability of remote sensing data for monitoring tasks

A. A. Proshin, E. A. Loupian, M. A. Burtsev, K. A. Troshko, A. V. Kashnitskii

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

The paper examines the impact of compressing satellite remote sensing data with quality loss using the LERC algorithm on the quality and applicability in real-world tasks of the products obtained on their basis. To assess this impact, the maximum permissible error that can be introduced during compression was evaluated, a test data sample, representing a year-long time series, was created and compressed using the LERC algorithm, then a series of comparisons was conducted. The original and lossy compressed series were compared in terms of NDVI (Normalized Difference Vegetation Index) values for individual scenes, the quality of the daily time series restoration, and the NDVI indices obtained from the restored series were assessed. Also, the quality of solution of object classification task using the Random Forest classifier on the constructed time series was evaluated. All experiments showed almost complete agreement of the calculation results with an error not exceeding the accuracy of the data themselves. The results demonstrate that, provided the criteria for preserving data quality are met, the use of LERC, enabling controlled introduction of errors into the data, provides processing and classification results comparable to those obtained from the original data with approximately a 2.5-fold decrease in the volume of stored data.

Keywords: remote sensing, data compression, LERC, lossy compression, Random Forest

Accepted: 23.10.2025
DOI: 10.21046/2070-7401-2025-22-5-63-72

References

1. Denisov P. V., Kashnitskii A. V., Kolbudaev P. A., Loupian E. A., Matveev A. M., Plotnikov D. E., Rutkevich B. P., Tolpin V. A., Troshko K. A., Technology of crops classification according to Sentinel-2 satellite data based on the capabilities of the Vega-PRO system, *Materialy 11-i Mezhdunarodnoi nauchnoi konferentsii "Regional'nye problemy distantsionnogo zondirovaniya Zemli"* (Proc. XI Intern. Scientific Conf. "Regional Problems of Earth Remote Sensing" RPERS 2024), Krasnoyarsk: SFU, 2024, pp. 40–46 (in Russian).
2. Loupian E. A., Proshin A. A., Burtsev M. A., Balashov I. V., Bartalev S. A., Efremov V. Yu., Kashnitskiy A. V., Mazurov A. A., Matveev A. M., Sudneva O. A., Sychugov I. G., Tolpin V. A., Uvarov I. A., IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, V. 12, No. 5, pp. 263–284 (in Russian).
3. Loupian E. A., Burtsev M. A., Proshin A. A., Kobets D. A., Evolution of remote monitoring information systems development concepts, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2018, V. 15, No. 3, pp. 53–66 (in Russian), DOI: 10.21046/2070-7401-2018-15-3-53-66.
4. Loupian E. A., Proshin A. A., Burtsev M. A. et al., Experience of development and operation of the IKI-Monitoring center for collective use of systems for archiving, processing and analyzing of satellite data, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2019, V. 16, No. 3, pp. 151–170 (in Russian), DOI: 10.21046/2070-7401-2019-16-3-151-170.
5. Proshin A. A., Loupian E. A., Burtsev M. A., Features of LERC image compression algorithm application in Earth remote sensing data archiving, *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2024, V. 21, No. 6, pp. 130–142 (in Russian), DOI: 10.21046/2070-7401-2024-21-6-130-142.
6. Becker P., Plesea L., Maurer T., Cloud optimized image format and compression, *Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2015, V. 40, pp. 613–615, DOI: 10.5194/isprsarchives-XL-7-W3-613-2015.
7. Becker P., Plesea L., Maurer T., Optimizing cloud based image storage, dissemination and processing through use of MRF and LERC, *Intern. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2016, V. 41, pp. 201–203, DOI: 10.5194/isprsarchives-XLI-B4-201-2016.

8. Oswal S., Singh A., Kumari K., Deflate compression algorithm, *Intern. J. Engineering Research and General Science*, 2016, V. 4, No. 1, pp. 430–436.
9. Plotnikov D. E., Kolbudaev P. A., Matveev A. M. et al., Daily surface reflectance reconstruction using LOWESS on the example of various satellite systems, *2022 8th Intern. Conf. Information Technology and Nanotechnology* (*ITNT*), IEEE, 2022, 5 p., DOI: 10.1109/ITNT55410.2022.9848630.
10. Zhu L., Suomalainen J., Liu J. et al., A review: Remote sensing sensors, In: *Multi-purposeful Application of Geospatial Data*, Rustamov R. B., Hasanova S., Zeynalova M. H. (eds.), IntecOpen, 2018, pp. 19–42, DOI: 10.5772/intechopen.71049.