

УДК 553.3/4.04:528.8(571.63)

РАСПАКОВКА И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДИСТАНЦИОННОЙ ОСНОВЫ ФОРМАТА *HDF* В *PYTHON* (НА ПРИМЕРЕ ДАННЫХ *TERRA ASTER*)

Шевырёв С.Л., Борискина Н.Г., Шевырёва М.Ж., Горобейко Е.В.

ФГБУН Дальневосточный геологический институт Дальневосточного отделения
Российской академии наук, Владивосток, e-mail: boriskina2000@mail.ru

Изображения Земли из космоса часто используются в науках о Земле как источник информации о ландшафтах, горных породах, а также состоянии атмо-, гидро- и биосферы. Организация полевых работ для целей геологической съемки, картирования, а также прогноза и поисков месторождений полезных ископаемых требует проведения предварительной оценки (рекогносцировки) территории, при этом эффективность проводимых работ напрямую связана с качеством и актуальностью дистанционных материалов. Пользовательская подготовка космических данных нулевого и первого уровней обработки зачастую является задачей, затратной по времени и объему потребляемых вычислительных ресурсов, кроме того, пользовательские географические информационные системы (ГИС) могут не обладать достаточными возможностями для ее решения. Как правило, предоставляемые данные нулевого и первого уровней обработки представляют собой показатели излучения на сенсоре спектрорадиометра, испытавшего атмосферное рассеяние, значение которого отличается по отдельным каналам изображения. Это значит, что получение значений отражательной способности поверхности и вычисление значений ее температуры требует выполнения локальной коррекции. Сайты организаций, предоставляющих доступ к спутниковым данным, а также спецификации этих данных, содержат информацию для атмосферной коррекции изображения. Также необходимо отметить, что многозональные данные могут предоставляться в специализированных форматах, не поддерживаемых пользовательскими ГИС. В настоящей статье рассматривается алгоритм извлечения (распаковки) каналов *ASTER* из файлов иерархического формата данных (*HDF*), их подготовка, включающая атмосферную коррекцию, вычисление значений температуры поверхности для температурных каналов ночной съемки и сохранение их в популярном формате *GeoTiff* с использованием пользовательского скрипта *Python*, в основе которого лежит библиотека *GDAL*. Скрипт можно адаптировать для применения к другим космическим данным, кроме того, он может использоваться для обучения студентов, обучающихся по направлению «Науки о Земле» программированию на *Python*, работе с *GDAL* и основам геоинформатики.

Ключевые слова: *Gdal*, дистанционное зондирование Земли, *Python*, геологическое картирование, иерархический формат данных, коррекция дистанционного изображения

PYTHON UNPACKING AND PREPROCESSING OF REMOTE SENSING IMAGES IN HDF FORMAT ON A SAMPLE OF TERRA ASTER DATA

Shevyrev S.L., Boriskina N.G., Shevyreva M.Zh., Gorobeyko E.V.

Far East Geological Institute, Far Eastern Branch of Russian Academy of Sciences,
Vladivostok, e-mail: boriskina2000@mail.ru

Remote sensing images are often used in Earth science as the source of information on landscapes, rocks and conditions of atmo-, hydro- and biosphere. Preparation of field works for geological mapping and prospecting of mineral deposits requires preliminary assessment of the area. Efficacy of field works depends on quality and relevance of remote sensing data. User handling of zero and first level processing data is often time and computational power consuming task. Moreover, desktop geographic information systems (GIS) may not possess enough capabilities for solving of that task. In general, available data of zero and first levels of processing express values of radiation on spectroradiometer sensor, which were subjected to band-specific atmospheric scattering. Deriving of top atmospheric reflectance and surface temperature requires channel-wise correction. Websites of companies, which provide access to satellite data and their specifications, also offer information for atmospheric correction. Also, multiband data could be provided in specific formats, which are not supported by user GIS. Paper considers algorithms of data extraction (unpacking) of *ASTER* data from hierarchical data format (*HDF*) including atmospheric correction, computing of surface temperature (for night temperature bands) and saving output into popular *GeoTiff* format using *Python* script bases on *GDAL* library. Script could be adapted for application on other satellite data, moreover, described software could be used for teaching *Python* programming, work with *GDAL* and basics of geoinformatics to Earth science students.

Keywords: *GDAL*, remote sensing of the Earth, python, geological mapping, hierarchical data format, adjustment of remote sensing image

Наращивание объемов запасов и значений подсчитанных прогнозных ресурсов рудных и нерудных полезных ископаемых требует усиления роли технологий дистанционного зондирования Земли и, соответственно, привлечения новых источников

свободно распространяемых космических данных, а также и усиления эффективности применения существующих. Качество применяемых материалов космических съемок определяется степенью их предварительной подготовки. Данные продвинутого

космического радиометра температурного излучения (ASTER) первого уровня обработки распространяются источником в виде архивов в формате HDF (иерархический формат данных). Использование первого уровня обработки (L1) в оригинальных исследованиях позволяет вполне реализовать потенциал этого источника данных, а также получать спутниковые продукты максимального качества.

Целью настоящей работы является рассмотрение решения задачи распаковки изображений Terra ASTER из формата HDF, преобразования цифровых значений спектральной яркости в показатели отражательной способности, регистрируемой в верхней части атмосферы, приближенное вычисление значений температуры земной поверхности для ночных изображений, а также сохранение изображений в формате с географической привязкой (GeoTiff). В качестве средств для выполнения вычислений используются библиотека GDAL и язык программирования Python, методической основой служат официальные материалы от издателей изображений [1].

Материалы и методы исследования

Перенос региональных работ на территории со сложным рельефом и сложным геологическим строением усиливает роль дистанционных исследований, включающих геофизические методы, а также анализ и интерпретацию космических изображений Земли. Последнее требует проведения не только исследования данных и получения информации, непосредственно связанной с содержанием изображения, но и его подготовкой, связанной с выводом из форматов распространения и хранения.

Состав поверхностных комплексов и глубинное строение земной коры может быть изучено посредством анализа изображений продвинутого космического радиометра тепловой эмиссии и отражения (Advanced Spaceborne Thermal Emission and Reflection Radiometer, ASTER), установленного на космический аппарат Terra, запущенный NASA 18 декабря 1999 г. [2]. Изображения предоставляются в нескольких уровнях обработки, разрешения каналов различаются от 15 м для видимого и ближнего инфракрасного диапазонов (VNIR), 30 м для коротковолнового инфракрасного (SWIR) и 90 м для теплового инфракрасного (TIR) диапазонов, наборы данных включают материалы дневной и ночной

съемок. Архив космических изображений Геологической службы содержит данные, полученные за почти 23 года работы этого космического аппарата, доступ к которым предоставляется без ограничений [3]. Данные первого уровня обработки представляют собой исходные материалы, дающие исследователю наибольший простор действий в дальнейших преобразованиях изображения и получении спутниковых продуктов, и представляют собой данные показателей излучения на сенсоре, скорректированные относительно модели местности в виде 14 спектральных каналов, упакованных в иерархическом формате данных (Hierarchical Data Format, HDF) [2].

Формат HDF был разработан Национальным центром суперкомпьютерных вычислений (NCSA) для предоставления пользователям возможности хранить, передавать и обрабатывать научные данные в различных операционных системах [4]. В соответствии с информацией разработчиков, основные его особенности включают: стандартизованность и платформенную независимость, он может содержать научные данные и растровые изображения. Файл включает информацию о себе, является самоописательным, это значит, что каждому объекту, содержащемуся в файле, соответствует отдельный тег, а дополнительные модели данных могут быть добавлены как разработчиками, так и пользователями формата [4].

Открытие архивов HDF, извлечение данных съемок в виде цифровых матриц, а также получение метаданных, содержащих сведения о параметрах и условиях съемки, удобно производить в Python с помощью библиотеки абстракции географических данных (GDAL) [5]. Эта библиотека, распространяемая на основе свободной лицензии, позволяет работать с растровыми и векторными форматами пространственных данных [6]. Библиотека обладает API (Application Programming Interface) для работы с популярными языками программирования (C, C++, Python), позволяет создавать на ее основе утилиты и геоинформационные системы [7].

Материалы съемок ASTER поставляются в виде 14 спектральных каналов, в их числе полученные в видимом и ближнем инфракрасном (visible and near infrared, VNIR), коротковолновом инфракрасном (shortwave infrared, SWIR), и термальном инфракрасном (thermal infrared, TIR) диапазонах. Характеристики изображений спектральных каналов приведены в таблице.

Характеристики спектральных каналов ASTER [Working with ASTER L1T..., 2022]

Наименование	Описание	Разрешение	Размерность	Тип данных	Диапазон
VNIR_Band1	Ближний канал видимого и инфракрасного излучения 1 (0,52 to 0,60 мкм)	15 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
VNIR_Band2	Ближний канал видимого и инфракрасного излучения 2 (0,63 to 0,69 мкм)	15 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
VNIR_Band3N	Ближний канал видимого и инфракрасного излучения 3N (0,78 to 0,86 мкм)	15 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band4	Коротковолновой инфракрасный канал 4 (1,600 to 1,700 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band5	Коротковолновой инфракрасный канал 5 (2,145 to 2,185 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band6	Коротковолновой инфракрасный канал 6 (2,185 to 2,225 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band7	Коротковолновой инфракрасный канал 7 (2,185 to 2,225 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band8	Коротковолновой инфракрасный канал 8 (2,295 to 2,365 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
SWIR_Band9	Коротковолновой инфракрасный канал 9 (2,360 to 2,430 мкм)	30 м	Вт/м ² /ср/мкм	8-разрядное целое без знака	0 до 255
TIR_Band10	Термальный инфракрасный канал 9 (8,125 to 8,475 мкм)	90 м	Вт/м ² /ср/мкм	16-разрядное целое без знака	0 до 65535
TIR_Band11	Термальный инфракрасный канал 11 (8,475 to 8,825 мкм)	90 м	Вт/м ² /ср/мкм	16-разрядное целое без знака	0 до 65535
TIR_Band12	Термальный инфракрасный канал 12 (8,925 to 9,275 мкм)	90 м	Вт/м ² /ср/мкм	16-разрядное целое без знака	0 до 65535
TIR_Band13	Термальный инфракрасный канал 13 (10,250 to 10,950 мкм)	90 м	Вт/м ² /ср/мкм	16-разрядное целое без знака	0 до 65535
TIR_Band14	Термальный инфракрасный канал 14 (10,950 to 11,650 мкм)	90 м	Вт/м ² /ср/мкм	16-разрядное целое без знака	0 до 65535

В соответствии с имеющимися рекомендациями, для синтеза спутниковых продуктов и изучения ландшафтно-географических комплексов отражающей поверхности, показатели прошедшего атмосферу излучения на сенсоре радиометра, записанные в виде цифровых значений и хранящиеся в формате HDF, должны быть преобразованы в отражательную способность, регистрируемую в верхней части атмосферы (TOA Reflectance) [1].

Для распаковки и подготовки данных ASTER L1T был подготовлен скрипт Python [8], алгоритм и этапы работы которого отражены на схеме (рис. 1).

Скрипт распаковки HDF файлов, представляющий собой программу с открытым

исходным кодом на языке Python [8], помимо ключевой для него библиотеки `gdal` использует в работе библиотеки `numpy` (работа с матрицами и многомерными данными), `pandas` (открытие таблиц значений коэффициентов преобразования единиц и длин волн формата MS Excel), `scikit-image` для трансформации изображений, `datetime` и `time` для работы с форматами даты и времени съемки, `re` для анализа строковых переменных и `os` для удобной работы с файлами на диске. Скрипт не обладает визуальным интерфейсом, указание рабочего каталога, содержащего HDF файлы, осуществляется с помощью непосредственного редактирования файла программы с указанием значения переменной `*foldname*`.



Рис. 1. Структура алгоритма распаковки, преобразования и сохранения данных ASTER L1T в виде изображений с географической привязкой (GeoTiff)

По умолчанию других действий от пользователя не требуется, программа самостоятельно определяет параметры расчета на основе имен файлов (которые необходимо оставить по умолчанию, переименование не допускается) и метаданных. Запуск скрипта рекомендуется производить с помощью научных сред разработки Spyder или Jupyter, основанных на интерактивной консоли IPython.

Данные ASTER L1T хранятся в формате цифровых значений яркости изображений (digital numbers, DN). Для построения RGB изображений и спутниковых продуктов необходимо получить значения отражательной способности в верхней части атмосферы (Top Of the Atmosphere reflectance, TOA). Для ее получения необходимо предварительно вычислить показатель излучения на сенсоре [9]:

$$L = (DN - 1) \cdot usc(1), \quad (1)$$

где L_y – излучение в верхней части атмосферы, DN – значение яркости изображения, usc – коэффициенты преобразования единиц [9].

Вычисление отражательной способности в верхней части атмосферы выполняется с помощью следующей формулы [10]:

$$TOA = \frac{L d^2}{ESUN \cos(s)}, \quad (2)$$

где d – расстояние от Земли до Солнца, $ESUN_y$ – среднее солнечное экзосолнечное

излучение, полученное по данным [11], s – зенитный угол Солнца.

Значения usc для расчетов заимствуются из справочных таблиц [9], угол s извлекается из метаданных HDF файла. Для получения значения d можно воспользоваться формулой

$$d = (1 - 0,01672 \cdot \cos(0,9856 \times (doy - 4))), \quad (3)$$

где doy – порядковый номер дня в году, в программе он находится за счет парсинга имен файлов HDF.

Вычисление отражательной способности в верхней части атмосферы производится для дневных изображений («day», рис. 1). Ночные изображения («night», рис. 1) получаются в термальном инфракрасном диапазоне (TIRS, каналы 10–14), их обработка позволяет найти приближенные значения температуры земной поверхности. Для этой цели используется методика расчета, описанная [12]. Для вычисленных значений излучения в верхней части атмосферы L_y температура поверхности земли может быть найдена как

$$T = \frac{K_2}{\ln\left(\frac{K_1}{L_y} + 1\right)}, \quad (4)$$

где K_1 и K_2 – это коэффициенты, предопределенные эффективной длиной волны на сенсоре спутника [12].

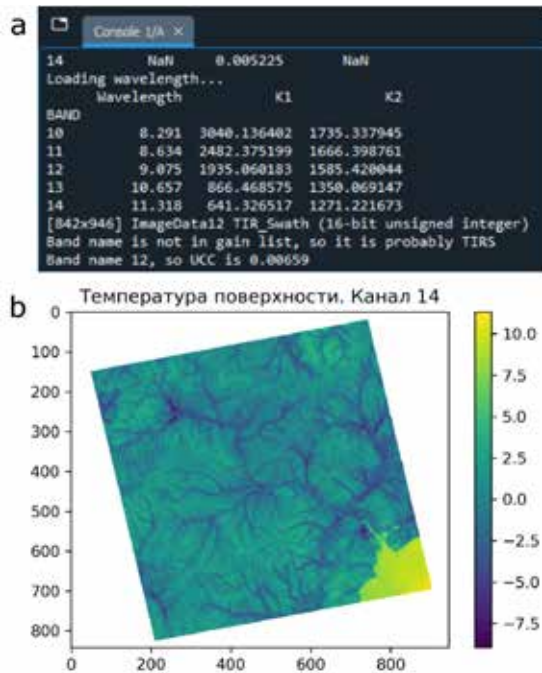


Рис. 2. Текстовый (а) и графический (b) выходы, сопровождающие работу программы распаковки HDF файлов ASTER LIT

Так как каналы TIRS и SWIR обладают меньшим разрешением изображений (30 и 90 м соответственно), выполняется программное повышение разрешения до VNIR (15 м) для облегчения дальней-

шего расчета пользователями спутниковых продуктов. Перед запуском программы необходимо указать расположение файлов, откорректировав значение переменной *foldname* в строке *foldname='ASTER_LIT_Night_Cloudless'*, при этом можно использовать как абсолютное, так и относительное указание расположения. Папка, путь к которой указан в *foldname*, должна содержать как минимум один HDF-файл. После успешного запуска программы в консоли Python можно увидеть вывод сообщений и эскизов изображений (рис. 2, а и b). В качестве демонстрации графического вывода использована пересчитанная сцена 14 канала ASTER LIT, полученная для Приморско-го края.

Создаваемые программой GeoTiff файлы имеют имя, задаваемое по шаблону: «ImageData_LIT.НОМЕР_КАНАЛА_ДИАПАЗОН_Swath_ДАТА_ВРЕМЯ_СУТОК_ПУТЬ_РЯД_ВИД_ИНФОРМАЦИИ.tiff». Здесь НОМЕР_КАНАЛА – это число, обозначающее номер, ДИАПАЗОН подразумевает диапазон значений (VNIR, SWIR, TIRS), ДАТА – дата съемки, ВРЕМЯ_СУТОК – соответственно дневное или ночное время, ПУТЬ и РЯД – указание расположения снимка, а ВИД_ИНФОРМАЦИИ – значение единиц измерения (для TIRS это градусы Цельсия).



Рис. 3. Геологическая карта острова Кунашир с указанием положения участка детальных исследований (составлено авторами по [14])

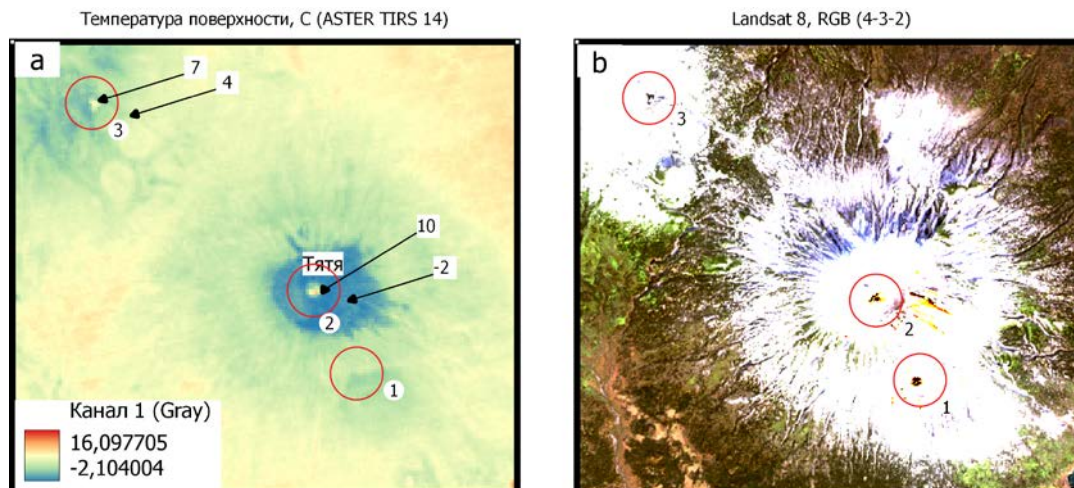


Рис. 4. Сопоставление фрагмента изображения температур поверхности (а), построенного на основе изображения ASTER L1T TIRS 14 (p200r585, дата съемки 8 июля 2015 г.) и изображения Landsat 8 OLI в естественных цветах (RGB 4–3–2, дата съемки 27 октября 2015 г.). Цифры в кружках – локальные температурные аномалии

Результаты исследования и их обсуждение

Тестирование приложения для распаковки данных ASTER L1T было выполнено для космического изображения вулкана Тятя, расположенного на севере острова Кунашир Большой Курильской гряды. Вулкан Тятя относится к числу активных вулканов острова и относится к типу Сомма-Везувий, на основании древнего его плейстоценового конуса была сформирована современная вулканическая постройка. К северо-западу от него находится вулкан Руруй. Тятя сложен базальтами, андезибазальтами и двупироксеновыми андезитами позднейших стадий извержений [13]. Геологическое строение территории и расположение участка исследований показано на схематической карте (рис. 3).

В подготовленной программе на языке Python нами была обработана сцена ASTER (p200r585), дата съемки 8 июля 2015 г. (рис. 4). Выделенные локальные участки повышенных температур рассматривались как аномальные и заверялись с сопоставлением снимка Landsat 8 OLI в псевдоестественных цветах (4–3–2), отражающего состояние местности на 27 октября 2015 г. с низким положением снеговой линии.

Визуальное сопоставление этих изображений позволило дать интерпретацию выделенным аномалиям. Объект, выделенный как участок таяния снега (рис. 4, б), номер 1 в кружке, на температурном изображении не отмечается, тогда как объекты 2 и 3 выделяются на обоих изображениях. Объект 2,

очевидно, соответствует жерлу кальдеры центрального конуса вулкана Тятя, его температура поверхности на снимке 10 и более градусов, тогда как он окружен областью отрицательных температур снеговой шапки. Объекту 3 соответствует температурная аномалия на склоне вулкана Руруй, возможно, сформированная одним из его паразитических конусов.

Заключение

Подготовленная программа распаковки HDF файлов дистанционных изображений ASTER L1T была апробирована на наборах данных, полученных из официального репозитория [3]. В ходе тестирования программы на материалах съемок площади, включающей вулкан Тятя, проведены поиски и заверка локальных участков повышенных температур, которые затем сопоставлялись с участками таяния снежного покрова на осеннем снимке.

Анализ получаемых с помощью программы температурных изображений ASTER может быть использован для выявления активных зон вулканических построек, а также для мониторинга динамики вулкана.

Методические наработки и исполняемый код, полученные при написании программы, являются практико-ориентированными и могут лечь в основу студенческих научных работ, а также использоваться при написании курсовых и дипломных исследований.

Помимо поиска горячих полей и источников программа может применяться для

поиска коренных выходов горных пород, а также ретроспективной оценки площадей лесных пожаров.

Список литературы

1. Working with ASTER L1T Visible and Near Infrared (VNIR) Data. [Электронный ресурс]. URL: <https://pdaac.usgs.gov/resources/e-learning/working-aster-l1t-visible-and-near-infrared-vnir-data-r/> (дата обращения: 29.08.2022).
2. Изображения Terra ASTER 1 уровня обработки. [Электронный ресурс]. URL: https://pdaac.usgs.gov/products/ast_11tv003/ (дата обращения: 29.08.2022).
3. EarthExplorer USGS. [Электронный ресурс]. URL: <https://earthexplorer.usgs.gov> (дата обращения: 29.08.2022).
4. Hierarchical data format (HDF). [Электронный ресурс]. URL: <https://asdc.larc.nasa.gov/documents/tools/hdf.pdf> (дата обращения: 29.08.2022).
5. GDAL documentation. [Электронный ресурс]. URL: <https://gdal.org/> (дата обращения: 29.08.2022).
6. GDAL license. [Электронный ресурс]. URL: <https://gdal.org/license.html> (дата обращения: 29.08.2022).
7. GDAL application programming interface (API). [Электронный ресурс]. URL: <https://gdal.org/api/index.html> (дата обращения: 29.08.2022).
8. Shevryev S.L. ASTER L1T HDF unpacker. [Электронный ресурс]. URL: https://github.com/SergeiShevryev/py_h2t (дата обращения: 29.08.2022).
9. Abrams M., Hook S., Ramachandram B. Aster user handbook. Version 2. 1999. [Электронный ресурс]. URL: https://asterweb.jpl.nasa.gov/content/03_data/04_Documents/aster_user_guide_v2.pdf (дата обращения: 29.08.2022).
10. Landsat 7 Science Data Users Handbook. [Электронный ресурс]. URL: https://landsat.gsfc.nasa.gov/wp-content/uploads/2016/08/Landsat7_Handbook.pdf (дата обращения: 29.08.2022).
11. Thome K.J., Biggar S.F., SLATER P.N. Effects of assumed solar spectral irradiance on intercomparisons of earth-observing sensors. Proceedings Volume 4540, Sensors, Systems, and Next-Generation Satellites V International Symposium on Remote Sensing (Toulouse, 12 December 2001). Toulouse, International Society for Optics and Photonics. 2001. P. 260–269. DOI: 10.1117/12.450668.
12. Ghulam A. How to calculate reflectance and temperature using ASTER data. [Электронный ресурс]. URL: <http://www.pancroma.com/downloads/ASTER%20Temperature%20and%20Reflectance.pdf> (дата обращения: 29.08.2022).
13. Мартынов Ю.А., Мартынов А.Ю., Чашин А.А., Рыбин А.В. Базальты вулкана Тятя: петрология и генезис (остров Кунашир, Курильская островная дуга) // Тихоокеанская геология. 2005. Т. 24. № 3. С. 22–31.
14. Shevryev S., Carranza E.J.M. Application of maximum entropy for mineral prospectivity mapping in heavily vegetated areas of Greater Kurile Chain with Landsat 8 data. Ore Geology Reviews. 2022. Vol. 142. P. 104758. DOI: 10.1016/j.oregeorev.2022.104758.