

ДВАДЦАТЬ ТРЕТЬЯ МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ
"СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ
КОСМОСА"

(Методы и алгоритмы обработки спутниковых данных)

Москва, ИКИ РАН, 10-14 ноября 2025 г.



**Применение машинного обучения для разделения природных пожаров и газовых факелов по
атрибутивным данным спутниковых информационных продуктов**

Н.В. Феоктистова, О.С. Воронова

**Научно-исследовательский институт аэрокосмического мониторинга
«АЭРОКОСМОС»**

Россия, г. Москва, Гороховский пер., д.4, 105064
Тел.: 632-16-54, тел./факс: 632-11-78; E-mail:
office@aerocosmos.info

ВВЕДЕНИЕ

Глобальные климатические изменения, наблюдаемые в последнее время, являются серьезным вызовом для России (Третий оценочный доклад, 2022) и мира в целом (Отчет IPCC, 2022). Природные пожары влияют на запас углерода в лесных экосистемах и, следовательно, на углеродный баланс, а также являются источниками климатически активных газов, в том числе таких как углекислый газ CO₂, угарный газ CO, метан CH₄ и др.

Прогрессирующая динамика интенсивности и частоты возникновения природных пожаров характерна для бореальных лесов Сибири и Дальнего Востока (Бондур и др., 2024; Bondur et al., 2023; Ponomarev et al., 2023). Суммарные объемы эмиссий климатически активных углеродсодержащих газов от пожаров в Сибири и на Дальнем Востоке за последнее десятилетие колеблются от 95 до 150 млн тонн (Бондур и др., 2024).

Важнейшими переменными для изучения лесных пожаров являются мгновенная мощность излучения пожара (FRP) и интегрированная по времени энергия излучения пожара (FRE), которые могут регистрироваться по спутниковым данным (Li et al., 2018). Спутниковые оценки этих параметров позволяют получать такие характеристики пожаров, как интенсивность, расход биомассы, а также оценивать объемы эмиссий климатически активных углеродсодержащих газов (Giglio et al., 2016; Li et al., 2018).

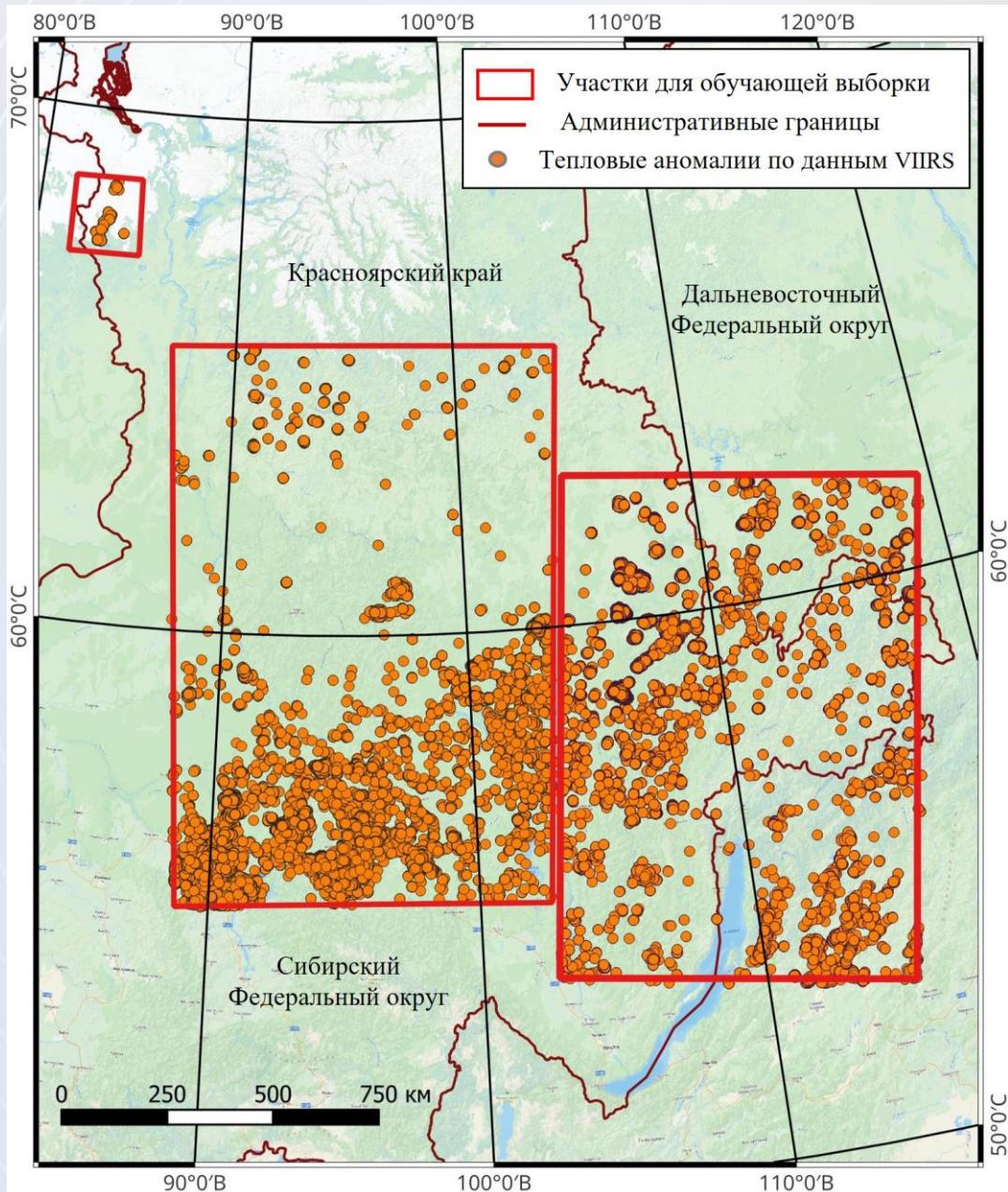
Газовые факелы, также являются источниками интенсивного выделения тепла, и могут быть с высокой степенью достоверности обнаружены путем регистрации из космоса тепловых аномалий в инфракрасном диапазоне спектра (Li R., 2021, Zhizhin M., 2021). В спутниковых информационных продуктах (например MODIS, VIIRS) эти объекты часто идентифицируются как природные пожары, что может приводить к неверной оценке пожарной активности и необходимости привлечения дополнительной информации при расчете пожарных эмиссий климатически активных газов. Таким образом, возникает необходимость дополнительного анализа спутниковых данных о тепловых аномалиях и более точного разделения их по типам источников.

ТЕРРИТОРИЯ ИССЛЕДОВАНИЯ

Для проведения исследований была выбрана территория Дальневосточного и Сибирского федеральных округов. Для этих регионов характерны интенсивные природные пожары (Бондур и др., 2018, 2024). Кроме того, в этом регионе находится много крупных месторождений нефти и газа с множеством газовых факелов, регистрируемыми по спутниковым данным как тепловые аномалии.

Добыча нефти и газа в Красноярском крае сосредоточенная в основном в пределах проекта «Восток Ойл» и на Ванкорском кластере. Основные запасы сосредоточены на месторождениях, таких как Ванкорское, Тагульское и Сузунское, а также новых месторождениях, включая Новоогненное, где запасы оцениваются в десятки миллионов тонн нефти и миллиардов кубометров газа.

В Якутии интенсивно растет добыча нефти, происходит открытие новых и доразведка существующих крупных газовых месторождений (Слепцова, 2024). При этом, Якутия является одним из наиболее пожароопасных субъектов России, что обусловлено как резко континентальным климатом с характерным жарким и засушливым летом, так и наличием обширных массивов светлохвойных лесов, занимающих большую часть территории (Протопопова, Габышева, 2018; Бондур и др., 2024).



ИСПОЛЬЗУЕМЫЕ ДАННЫЕ

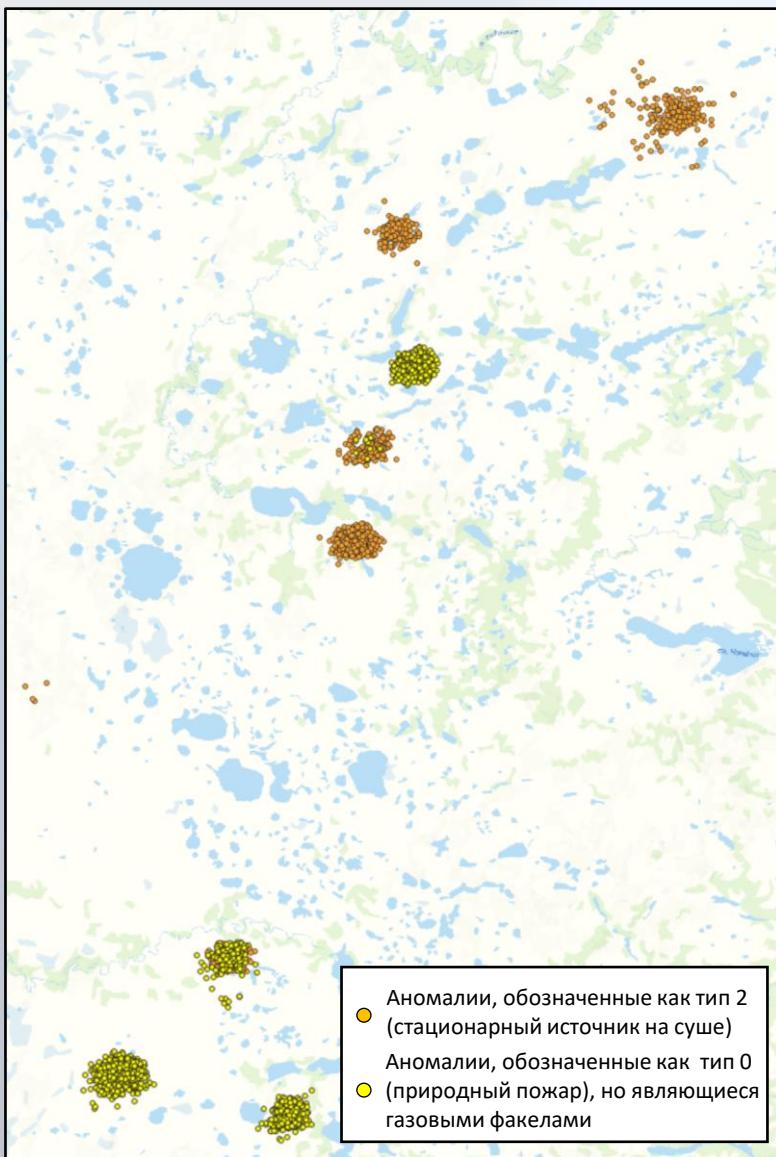
В настоящей работе использовался комбинированный информационный продукт MCD14 с пространственным разрешением 1 км, который содержит данные о тепловых аномалиях, со спутников Terra и Aqua (Giglio et al., 2016), а также информационный продукт VIIRS Fire and Thermal Anomalies для с пространственным разрешением 375 м, получаемых со спутников NOAA-20 (JPSS-1) и NOAA-21 (JPSS-2), а также со спутника Suomi-National Polar Orbiting Partnership (S-NPP), являющегося совместным проектом NASA/NOAA (Giglio, 2024).

Эти спутниковые информационные продукты предоставляют для каждого «горячего» пикселя информацию о времени обнаружения, координатах, дате, уровне достоверности, радиационной мощности (FRP) и яркостной температуре в среднем и тепловом ИК диапазонах (Giglio et al., 2016; Giglio, 2024). Данные признаки использовались нами при машинном обучении.

Признаки, используемые для машинного обучения, на основе атрибутивных данных спутниковых информационных продуктов MCD14 и VIIRS Fire and Thermal Anomalies

Признак	Описание для MODIS	Описание для VIIRS
Яркостная температура в среднем ИК диапазоне (brightness)	Яркостная температура пикселя пожара каналы 21/22 (3.96 мкм), градусы Кельвина.	Яркостная температура пикселя в канале 14 (3.55 - 3.93 мкм), Кельвин.
Яркостная температура в тепловом ИК диапазоне (bright_t31)	Яркостная температура пикселя в канале 31 (11.03 мкм), градусы Кельвина.	Яркостная температура пикселя в канале 15 (10.5 - 12.4 мкм), градусы Кельвина.
Достоверность (confidence)	Значение основано на наборе промежуточных величин, используемых в алгоритме обнаружения пожара. Оно предназначено для оценки качества отдельных пикселей горячих точек/пожаров. Оценки достоверности находятся в диапазоне от 0 до 100% и присваиваются одному из трех классов пожара (пожар с низкой достоверностью, пожар с номинальной достоверностью или пожар с высокой достоверностью).	Значения достоверности устанавливаются на низком (l), номинальном (n) и высоком (h) уровнях. Пиксели дневных пожаров с низкой (l) достоверностью обычно связаны с областями солнечного блика и более низкой относительной температурной аномалией (<15 К) в канале 14. Пиксели с номинальной достоверностью (n) – без солнечных бликов в течение дня и отмечены сильной (>15 К) температурной аномалией как в дневных, так и вочных данных.
Мощность излучения огня, МВт, (frp)	Интегрированная по пикселю мощность излучения пожара FRP в МВт.	Интегрированная по пикселю мощность излучения пожара FRP в МВт.
День/Ночь (daynight)	D - Дневной пожар, N - Ночной пожар	D - Дневной пожар, N - Ночной пожар
Дата съемки (julian_day)	Дата съемки преобразована в юлианский день	Дата съемки преобразована в юлианский день

ОСОБЕННОСТИ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ

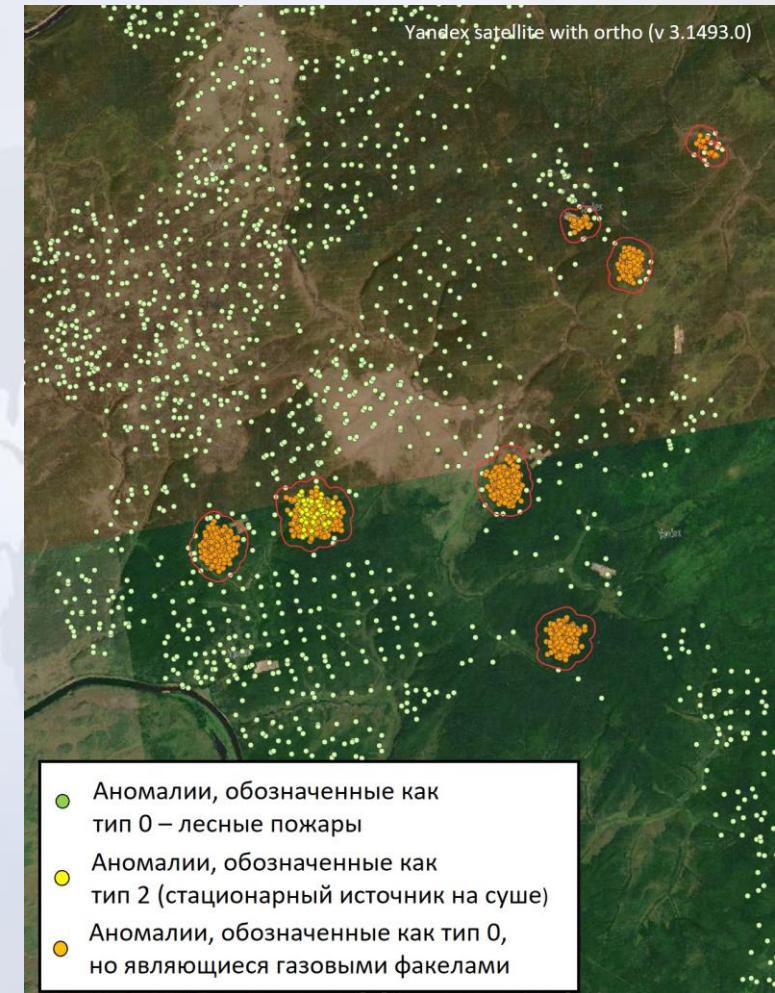


Ванкорское нефтегазоконденсатное месторождение в
Туруханском районе, Красноярский край

С использованием алгоритмов обнаружения тепловых аномалий по данным спутниковых приборов MODIS и VIIRS, определяют 4 типа тепловых аномалий:

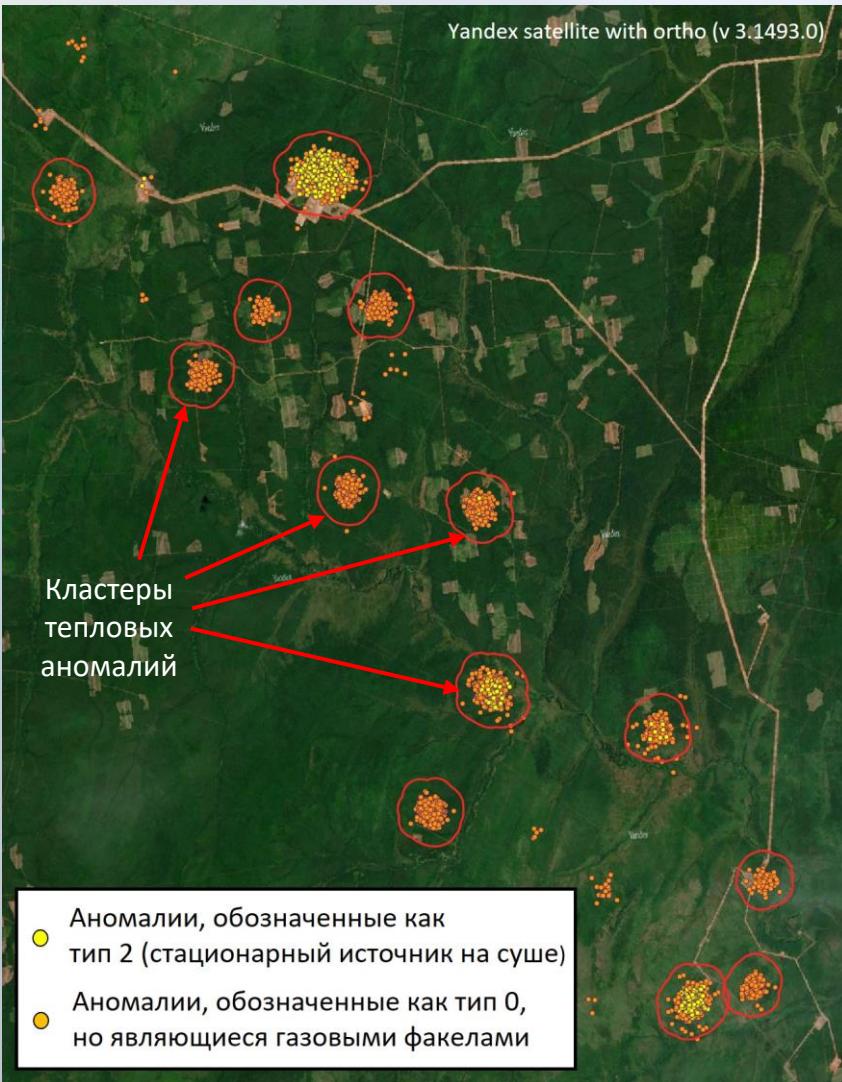
- 0 - предполагаемый природный пожар,
- 1 - действующий вулкан,
- 2 - стационарный источник на суще (как правило крупные антропогенные объекты, в том числе газовые факелы),
- 3 - блики на водной поверхности (включая все обнаружения над водой).

Однако, эта классификация не позволяет надежно разделять тепловые аномалии, связанные с природными пожарами и антропогенными источниками теплового излучения такими, как газовые факелы.



Даниловское нефтегазоконденсатное месторождение,
Иркутская область

ОСОБЕННОСТИ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ



1. Для обучения моделей были выбраны тепловые аномалии на территории исследуемого района в 2021-2023 гг. Разметка данных проводилась в два этапа. **На первом этапе** выделялись крупные кластеры тепловых аномалий, связанных с газовыми факелами. Как видно из рисунка, тепловые аномалии, соответствующие газовым факелам, объединенные за несколько лет, имеют более высокую пространственную плотность чем тепловые аномалии, характерные для природных пожаров. Это позволяет обнаружить крупные и мощные антропогенные тепловые источники с помощью операции кластеризации DBSCAN. Однако этот метод дает достаточно большое количество ложных срабатываний, а также не выделяет в кластеры объекты с малым количеством тепловых аномалий. Поэтому, **на втором этапе** разметка, сделанная на первом этапе, уточнялась с привлечением спутниковых данных высокого разрешения (слой Yandex satellite with ortho в среде QGIS, Google Earth).

2. Далее для улучшения результатов классификации и выделения класса лесных пожаров были дополнительно привлечены метеопараметры на момент съемки, полученные из глобальной системы ассимиляции данных о Земле GLDAS (Rodell et al., 2024), а также из наборов данных реанализа ERA5 (Muñoz, 2019) и CFSR: Climate Forecast System Reanalysis (Saha S., et al.). После анализа пропусков и нулевых значений в метеоданных, были выбраны 4 основные метеорологические характеристики: скорость ветра (wind), температура воздуха (temp_air), влажность (humid) а также осадки (precip) из разных наборов данных.

МАШИННОЕ ОБУЧЕНИЕ

Для классификации тепловых источников: газовых факелов и природных пожаров на территории Сибирского федерального округа по атрибутивным данным спутниковых информационных продуктов, применялись модели машинного обучения Случайный лес (Random Forest) и Категориальный бустинг (Categorical Boosting, CatBoost). При обучении моделей сначала использовались параметры, содержащиеся в атрибутивных данных спутниковых информационных продуктов о тепловых аномалиях MODIS и VIIRS, такие как достоверность данных (confidence), время суток, в которое зарегистрированы данные (daynight), значения яркостных температур в среднем и тепловом ИК диапазоне (brightness и bright_31), а также значения радиационной мощности (frp).

Далее для улучшения качества обучения выбранных моделей водились дополнительные признаки, такие как метеопараметры на момент обнаружения тепловых аномалий.

Перед этапом машинного обучения проводилась проверка на мультиколлинеарность. Такая проверка в контексте классификации необходима для выявления значительной корреляции между независимыми переменными (признаками), используемыми для обучения моделей. При наличии сильной корреляции, оценки параметров модели становятся чувствительными к небольшим изменениям в данных.

Оценка производительности моделей осуществлялась с использованием следующих метрик: F1-Score, ROC-AUC, PR-AUC.

Размеры массивов данных для машинного обучения

	MODIS		VIIRS	
класс:	0	1	0	1
кол-во тепловых аномалий	55555	49974	147467	94319

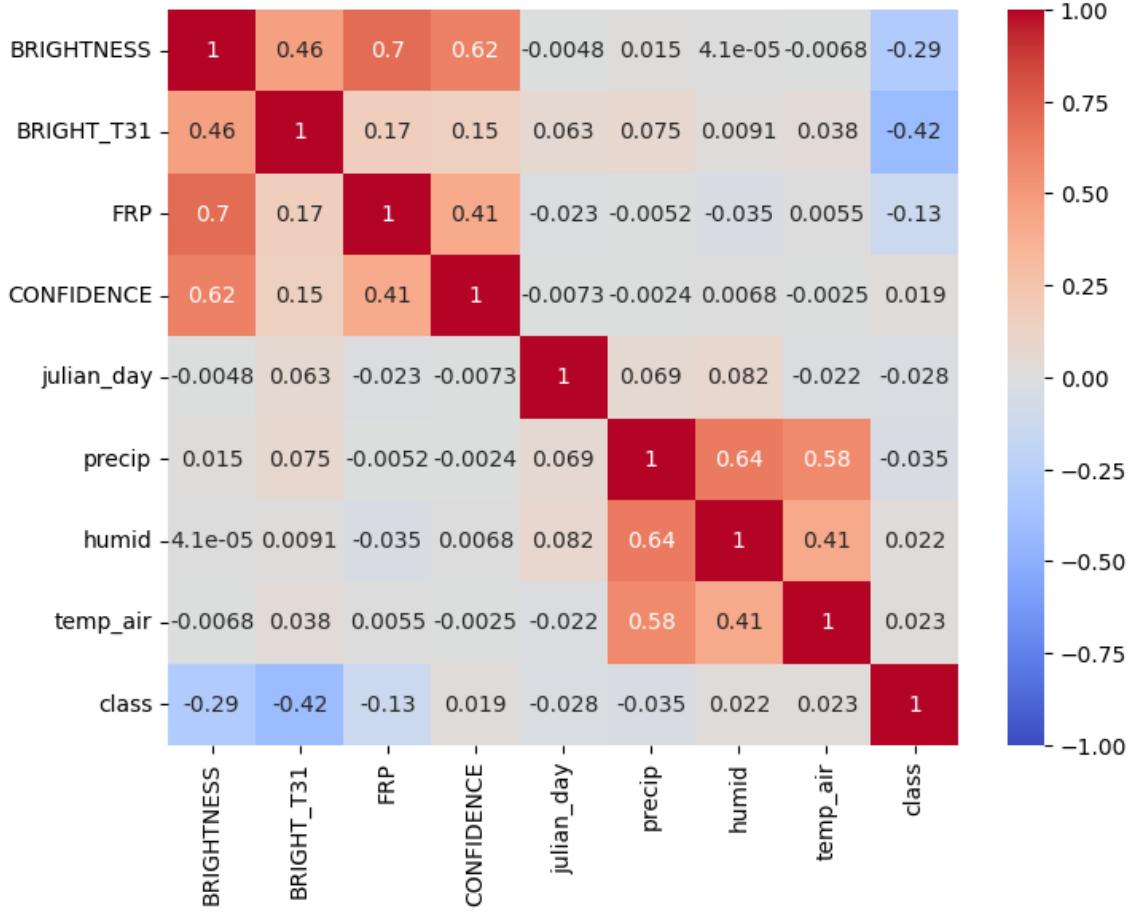
Количество тепловых аномалий, регистрируемых по данным приборов VIIRS, значительно больше, чем количество аномалий, регистрируемых приборами MODIS (Li et al., 2018). Это связано с разницей в размерах пикселей двух датчиков, а также с большим количеством спутников с прибором VIIRS, функционирующих на орбите (с 2022 г.).

Массивы данных были разделены на обучающую и тестовую выборки в процентном соотношении 70% на 30% соответственно.

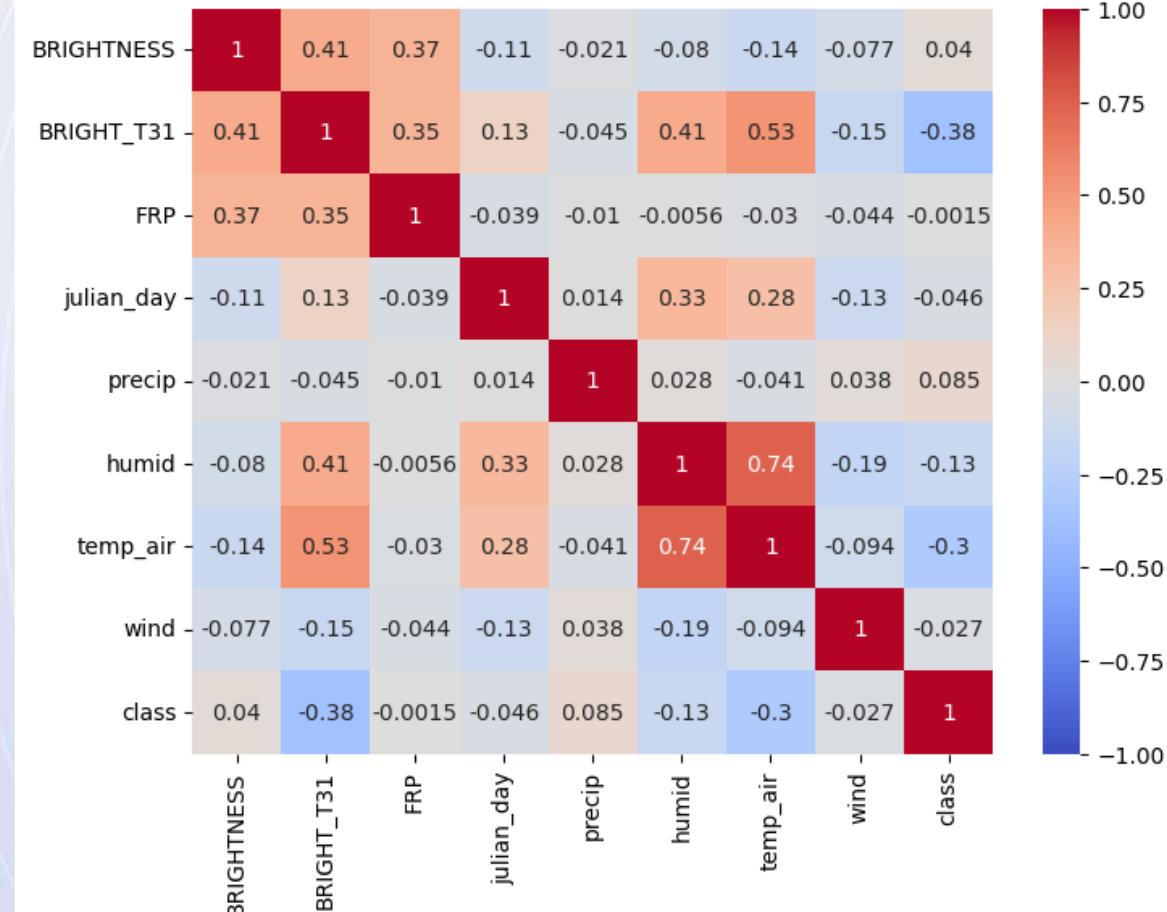
МАШИННОЕ ОБУЧЕНИЕ

Проверка признаков на мультиколлинеарность

Корреляция признаков для данных прибора MODIS



Корреляция признаков для данных прибора VIIRS



Используемые признаки для данных прибора MODIS имеют сильную корреляцию (0.62-0.70) внутри группы характеристики тепловой аномалии и метеопараметры. Самое высокое значение коэффициента корреляции (0.74) было выявлено у пары параметров влажность воздуха (humid) и температура воздуха (temp_air) для данных прибора VIIRS. При этом для выбранных признаков все коэффициенты корреляции находились внутри диапазона (-0.75 – 0.75). Поэтому эти переменные были оставлены для моделирования.

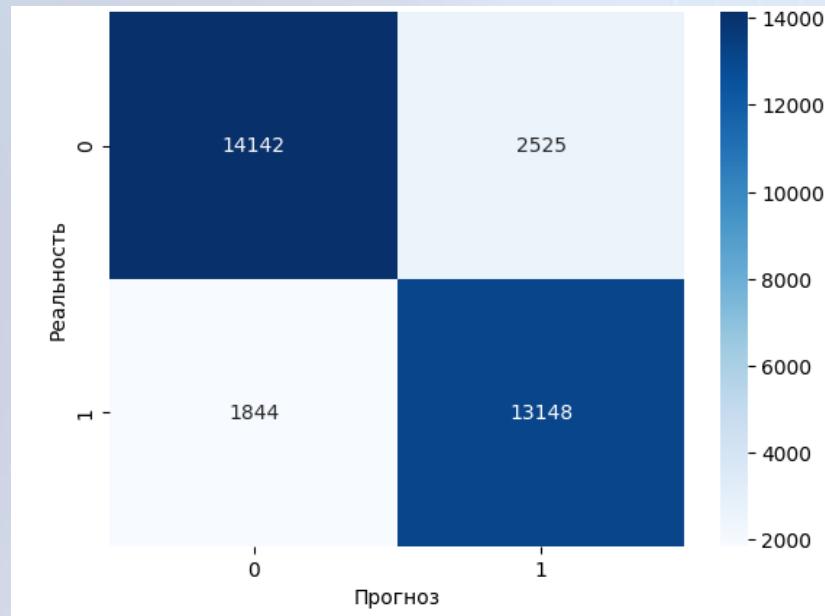
МАШИННОЕ ОБУЧЕНИЕ. РЕЗУЛЬТАТЫ

Показатели качества обучения выбранных моделей с использованием дополнительных признаков.

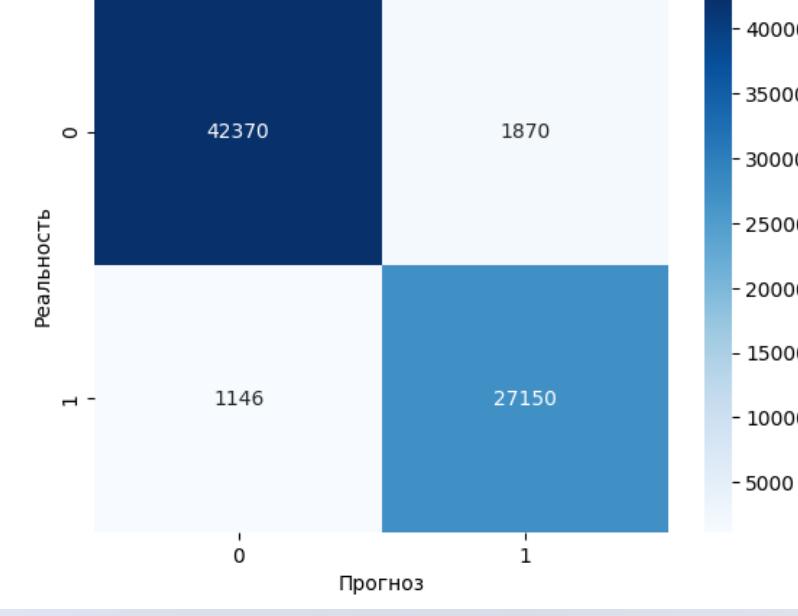
Название модели	С основными признаками (MODIS)			С дополнительными признаками (MODIS)		
	F1-score	ROC-AUC	PR-AUC	F1-score	ROC-AUC	PR-AUC
RF	0.796	0.883	0.893	0.799	0.886	0.837
CatBoost	0.840	0.923	0.873	0.858	0.937	0.887

Название модели	С основными признаками (VIIRS)			С дополнительными признаками (VIIRS)		
	F1-score	ROC-AUC	PR-AUC	F1-score	ROC-AUC	PR-AUC
RF	0.745	0.879	0.792	0.791	0.914	0.832
CatBoost	0.794	0.915	0.830	0.947	0.991	0.955

Матрица ошибок для лучшей модели на тестовых данных MODIS



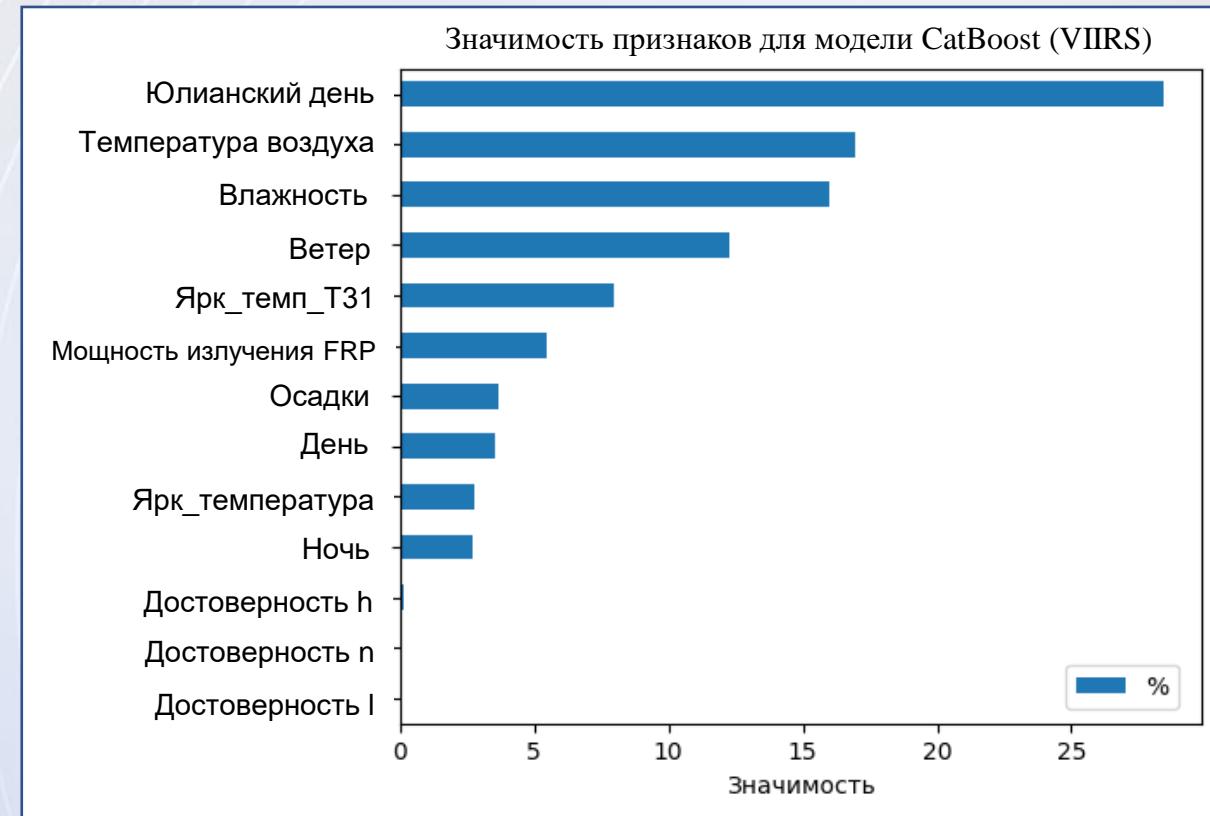
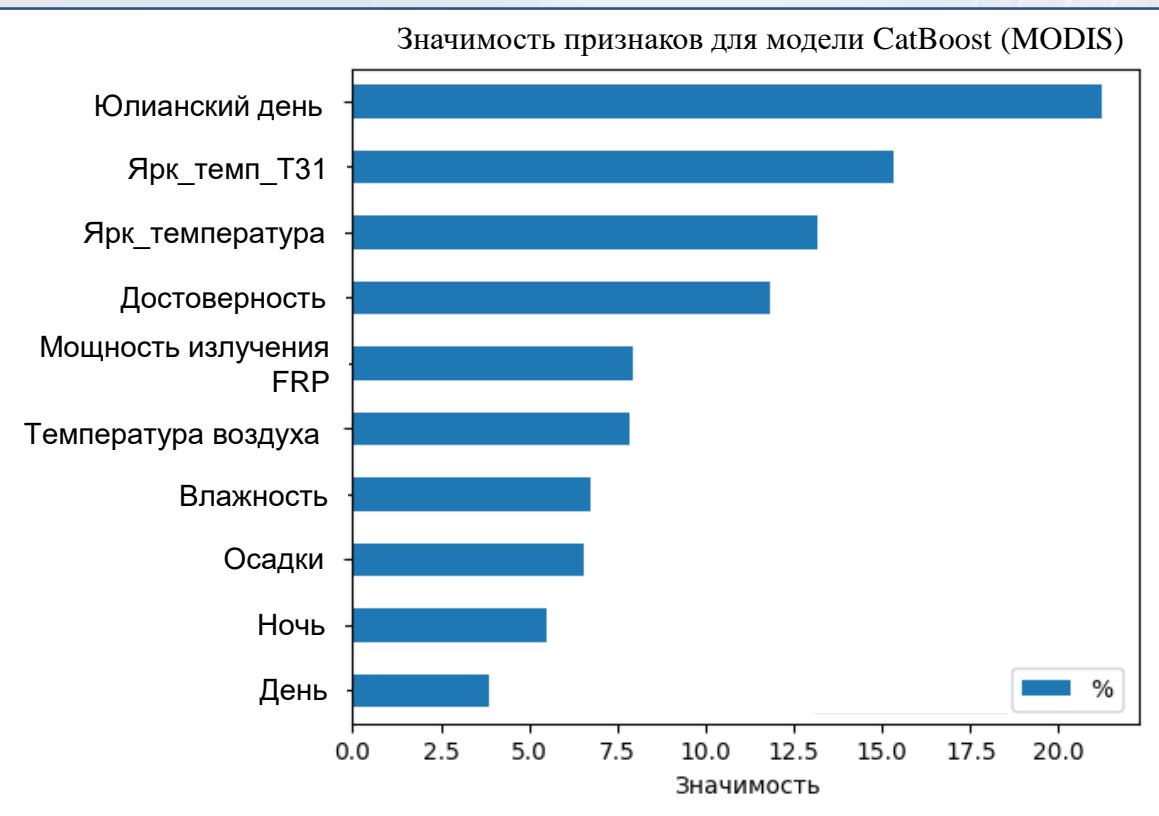
Матрица ошибок для лучшей модели на тестовых данных VIIRS



В обоих случаях лучшие показатели качества у модели CatBoost. При этом, введение дополнительных признаков – метеопараметров улучшило качество моделей для данных MODIS на 1,5-2%, а для данных VIIRS на 3,5 – 15%.

Модель для данных MODIS правильно классифицировала 14142 природных пожара и 13148 газовых факелов, при этом количество ложноположительных меток было предсказано в 2525 случаях (15%), а количество ложноотрицательных предсказано в 1844 случаях (12%). Модель для данных VIIRS позволила правильно классифицировать 42370 природных пожаров и 27150 газовых факелов, при этом количество ложноположительных меток составило 1870 (4%), а количество ложноотрицательных – 1146 (4%).

МАШИННОЕ ОБУЧЕНИЕ. РЕЗУЛЬТАТЫ



Оценка важности признаков показала, что для модели CatBoost для данных MODIS более важными признаками являются характеристики тепловых аномалий, а в случае модели для данных VIIRS наоборот - более важными признаками являются - метеопараметры. Это может быть связано с большим набором метеопараметров для данных VIIRS и в целом, большим набором данных для обучения. В обоих случаях самым важным признаком стал юлианский день, что связано с определенной сезонностью пожаров на исследуемой территории.

ВЫВОДЫ

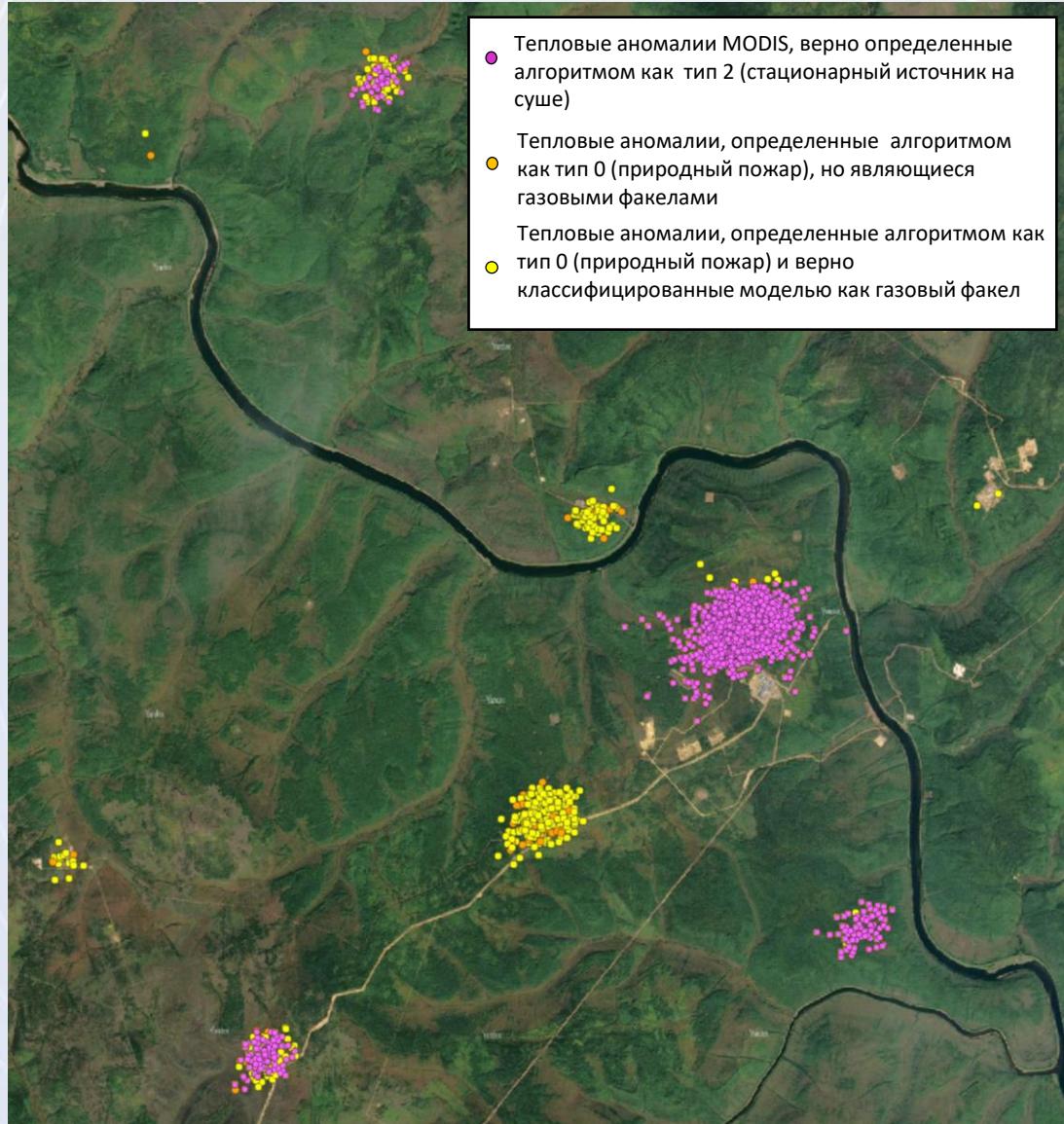
В данной работе предложены модели машинного обучения для классификации тепловых аномалий, регистрируемых из космоса, вызванных природными пожарами и газовыми факелами, на территории Дальневосточного и Сибирского федеральных округов. При обучении моделей использовались характеристики тепловых аномалий, содержащиеся в атрибутивных данных информационных продуктов

Проведено улучшение этих моделей путем введения дополнительных признаков (метеопараметров). Это позволило улучшить метрики качества F1-score, ROC-AUC и PR-AUC на 1,5-2% для данных MODIS и на 3-15% для данных VIIRS.

Лучшие модели CatBoost верно классифицировали 88% газовых факелов в случае данных MODIS и 97% для данных VIIRS.

Полученные результаты применения методов машинного обучения продемонстрировали возможность с высокой точностью разделить тепловые аномалии, вызванные природными пожарами и антропогенными объектами - газовыми факелами. Это важно при анализе пожарной активности в исследуемом регионе и позволит повысить точность оценок пожарных эмиссий климатически активных газов.

Пример результатов классификации тепловых аномалий по данным MODIS. Куюбинское нефтегазовое месторождение, Эвенкийский муниципальный район Красноярского края.



ИСТОЧНИК ФИНАНСИРОВАНИЯ

Исследование выполнено в НИИ «АЭРОКОСМОС» в рамках проекта №124021200003-1.

СПИСОК ЛИТЕРАТУРЫ

1. Бондур В.Г., Гордо К.А. Космический мониторинг площадей, пройденных огнем, и объемов эмиссий вредных примесей при лесных и других природных пожарах на территории Российской Федерации // Исследование Земли из космоса. 2018. № 3. С. 41-55. <https://doi.org/10.7868/S020596141803003X>
2. Бондур В.Г., Зима А.Л., Феоктистова Н.В. Долговременный спутниковый мониторинг различных типов природных пожаров и эмиссий климатически - активных газов и аэрозолей от них на территории России и ее крупных регионов / Исследование Земли из космоса. 2024. № 5. С. 19–34. <https://doi.org/10.31857/S0205961424050021>
3. Протопопова В. В., Габышева Л. П. Пирологическая характеристика растительности в лесах Центральной Якутии и ее динамика в постпожарный период. Природные ресурсы Арктики и Субарктики, 25, 3, 2018, pp. 80-86. <https://doi.org/10.31242/2618-9712-2018-25-3-80-86>
4. Слепцова М. И. Роль углеводородов в Якутии // Физико-технические проблемы добычи, транспорта и переработки органического сырья в условиях холодного климата. 2024, (1), 103-107, <https://doi.org/10.24412/cl-37255-2024-1-103-107>
5. Третий оценочный доклад об изменениях климата и их последствиях на территории Российской Федерации // Под ред. В.М. Катцова (Росгидромет). СПб.: Наукоемкие технологии, 2022. 676 с.
6. Anejionu Obinna C.D., Blackburn G. Alan, Whyatt J. Duncan. Detecting gas flares and estimating flaring volumes at individual flow stations using MODIS data // Remote Sensing of Environment Volume 158, 1 March 2015, Pages 81-94. <https://doi.org/10.1016/j.rse.2014.11.018>
7. Bondur V.G., Gordo K.A., Voronova O.S., Zima A.L., Feoktistova N.V. Intense Wildfires in Russia over a 22-Year Period According to Satellite Data. Fire 2023, 6, 99. <https://doi.org/10.3390/fire6030099>
8. IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA. 3056 pp. <https://doi.org/10.1017/9781009325844>
9. Giglio L., Schroeder W., Justice C. O. The Collection 6 MODIS active fire detection algorithm and fire products. Remote Sensing of Environment, 2016, Pp 31-41, ISSN 0034-4257, <https://doi.org/10.1016/j.rse.2016.02.054>.
10. Giglio L. VIIRS/JPSS1 Active Fires 6-Min L2 Swath 375m V002. 2024, distributed by NASA EOSDIS Land Processes Distributed Active Center, <https://doi.org/10.5067/VIIRS/VJ114IMG.002>

СПИСОК ЛИТЕРАТУРЫ

11. Li F., Zhang X., Kondragunta S., Csiszar I. Comparison of fire radiative power estimates from VIIRS and MODIS observations. *Journal of Geophysical Research: Atmospheres*, 2018, 123, 4545–4563. <https://doi.org/10.1029/2017JD027823>
12. Li R., Tao, M., Zhang, M., Chen, L., Wang, L., Wang, Y., et al. Application potential of satellite thermal anomaly products in updating industrial emission inventory of China. *Geophysical Research Letters*, 2021, 48, e2021GL092997. <https://doi.org/10.1029/2021GL092997>
13. Ponomarev E.I., Zabrodin A.N., Shvetsov E.G., Ponomareva T.V. Wildfire Intensity and Fire Emissions in Siberia. *Fire* 2023, 6, 246. <https://doi.org/10.3390/fire6070246>
14. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: unbiased boosting with categorical features // Computer Science. 2019. Pp.23 <https://doi.org/10.48550/arXiv.1706.09516>
15. Rodell M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., Toll, D. The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*. 2004. Vol. 85, No. 3, pp. 381-394. DOI: [10.1175/BAMS-85-3-381](https://doi.org/10.1175/BAMS-85-3-381)
16. Muñoz S. J. ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). 2019. DOI: [10.24381/cds.68d2bb30](https://doi.org/10.24381/cds.68d2bb30)
17. Saha S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, 91, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
18. Zhizhin M, Matveev A, Ghosh T, Hsu F-C, Howells M, Elvidge C. Measuring Gas Flaring in Russia with Multispectral VIIRS Nightfire. *Remote Sensing*. 2021; 13(16):3078. <https://doi.org/10.3390/rs13163078>