

ЛОКАЛЬНАЯ ДИАГНОСТИКА НАЛИЧИЯ ПОЛЯРНЫХ СИЯНИЙ НА ОСНОВЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ГЕОМАГНИТНЫХ ДАННЫХ

LOCAL DIAGNOSTICS OF AURORA PRESENCE BASED ON INTELLIGENT ANALYSIS OF GEOMAGNETIC DATA

А.В. Воробьев 

Геофизический центр РАН,
Москва, Россия, geomagnet@list.ru
Уфимский университет науки и технологий,
Уфа, Россия
Уфимский государственный нефтяной технический
университет,
Уфа, Россия

А.А. Соловьев 

Геофизический центр РАН,
Москва, Россия, a.soloviev@gcras.ru
Институт физики Земли им. О.Ю. Шмидта,
Москва, Россия

В.А. Пилипенко 

Геофизический центр РАН,
Москва, Россия, pilipenko_va@mail.ru
Институт физики Земли им. О.Ю. Шмидта,
Москва, Россия

Г.Р. Воробьева 

Уфимский университет науки и технологий,
Уфа, Россия, gulnara.vorobeva@gmail.com
Уфимский государственный нефтяной технический
университет,
Уфа, Россия

А.А. Гайнетдинова

Уфимский университет науки и технологий,
Уфа, Россия, gajnetdinova.aa@ugatu.su

А.Н. Лапин 

Уфимский университет науки и технологий,
Уфа, Россия, meccos160@yandex.ru

В.Б. Белаховский

Полярный геофизический институт РАН,
Апатиты, Россия, belakhovskiy@mail.ru

А.В. Ролдугин

Полярный геофизический институт РАН,
Апатиты, Россия, roldugin_a@pgia.ru

A.V. Vorobev

Geophysical Center RAS,
Moscow, Russia, geomagnet@list.ru
Ufa University of Science and Technology,
Ufa, Russia
Ufa State Petroleum Technological University,
Ufa, Russia

A.A. Soloviev

Geophysical Center RAS,
Moscow, Russia, a.soloviev@gcras.ru
Schmidt Institute of Physics of the Earth RAS,
Moscow, Russia

V.A. Pilipenko

Geophysical Center RAS,
Moscow, Russia, pilipenko_va@mail.ru
Schmidt Institute of Physics of the Earth RAS,
Moscow, Russia

G.R. Vorobeva

Ufa University of Science and Technology,
Ufa, Russia, gulnara.vorobeva@gmail.com
Ufa State Petroleum Technological University,
Ufa, Russia

A.A. Gainetdinova

Ufa University of Science and Technology,
Ufa, Russia, gajnetdinova.aa@ugatu.su

A.N. Lapin

Ufa University of Science and Technology,
Ufa, Russia, meccos160@yandex.ru

V.B. Belakhovsky

Polar Geophysical Institute RAS,
Apatity, Russia, belakhovsky@mail.ru

A.V. Roldugin

Polar Geophysical Institute RAS,
Apatity, Russia, roldugin_a@pgia.ru

Аннотация. Несмотря на существующее разнообразие подходов к мониторингу космической погоды и геофизических параметров в области аврорального овала, вопрос эффективного прогнозирования и диагностирования полярных сияний как особого состояния верхней ионосферы на высоких широтах остается практически открытым.

В работе исследуется и анализируется возможность локальной диагностики наличия сияний на основе интеллектуального анализа геомагнитных данных наземных источников. Оценивается значимость признаков переменных и их статистические взаимосвязи.

Так, например, применение байесовского вывода к данным геофизической станции «Ловозеро» за 2012–2020 гг. показало, что зависимость апостериорной вероятности наблюдения сияний в оптическом диапа-

Abstract. Despite the existing variety of approaches to monitoring space weather and geophysical parameters in the auroral oval region, the issue of effective prediction and diagnostics of auroras as a special state of the upper ionosphere at high latitudes remains virtually unresolved.

In this paper, we explore the possibility of local diagnostics of auroras through mining of geomagnetic data from ground-based sources. We assess the significance of indicative variables and their statistical relationship.

So, for example, the application of Bayesian inference to the data from the Lovozero geophysical station for 2012–2020 has shown that the dependence of a posteriori probability of observing auroras in the optical range on the state of geomagnetic parameters is logarithmic, and the degree of its significance is inversely

зоне от геомагнитных параметров имеет логарифмический характер, а степень значимости признака обратно пропорциональна невязке между эмпирическими данными и аппроксимирующей функцией.

Точность реализованного на основе метода случайного леса подхода к диагностике наличия полярных сияний составляет не менее 86 % при использовании нескольких локальных предикторов и ~80 % при использовании нескольких глобальных индексов геомагнитной активности, характеризующих возмущенность геомагнитного поля в авроральной зоне. В заключении рассматриваются перспективные пути улучшения метрик качества диагностических моделей и обсуждаются области их возможного применения.

Ключевые слова: полярные сияния, геомагнитные вариации, геомагнитные данные, аскаплоты, машинное обучение, интеллектуальный анализ данных, байесовский вывод, случайный лес.

proportional to the discrepancy between empirical data and approximating function.

The accuracy of the approach to diagnostics of aurora presence based on the random forest method is at least 86 % when using several local predictors and ~80 % when using several global geomagnetic activity indices characterizing the geomagnetic field disturbance in the auroral zone.

In conclusion, we discuss promising ways to improve the quality metrics of diagnostic models and their scope.

Keywords: auroras, geomagnetic variations, geomagnetic data, ascaplots, machine learning, data mining, Bayesian inference, random forest.

ВВЕДЕНИЕ

Как известно, наивысшие риски снижения уровня техносферной безопасности, связанные с эффектами воздействия космической погоды на объекты высокоширотной инфраструктуры (сбои систем коротковолновой радиосвязи и железнодорожной автоматики, возникновение дополнительных погрешностей магнитных инклинометров, отказы систем энергетического оборудования, снижение долговечности магистральных трубопроводов вследствие увеличения скорости их коррозии и пр. [Соколова и др., 2019; Птицына и др., 2008; Vorobev et al., 2022a; Soloviev et al., 2022; Пилипенко, 2021]) определяются в области аврорального овала — пояса интенсивной светимости, создаваемой вторжением в атмосферу электронов из околоземного пространства. Именно в данной области вследствие характерных для нее резких градиентов и высокого уровня турбулентности ионосферной плазмы регистрируются наиболее частые сбои фазы навигационного сигнала и экстремальные ошибки местоопределения [Захаров и др., 2020]. Вследствие этого погрешность высокоточной навигации GPS-приемников в режиме PPP (англ. Precise Point Positioning), эксплуатируемых в области высыпания авроральных электронов в ионосферу, может увеличиваться до пяти раз относительно фонового уровня [Yasyukevich et al., 2018, 2020].

Военные НАТО и пилоты ВКС РФ периодически докладывают о глобальных сбоях в системах приема GPS-сигнала, применяемых в зоне аврорального овала [<https://www.gpsworld.com/norway-finland-suspect-russia-of-jamming-gps>]. Так, например, космонавт-испытатель Токарев В.И. сообщает, что функциональный отказ штатных бортовых навигаторов при полете на малых высотах (200–500 м) в высокоширотных регионах является характерным откликом бортовой навигационной аппаратуры на возмущенное состояние космической погоды, снижающим безопасность эксплуатации военных, гражданских

и беспилотных летательных аппаратов в Арктическом регионе.

Следует отметить, что в периоды экстремальной геомагнитной активности (ГМА) вследствие смещения аврорального овала в область более низких широт обозначенные риски становятся актуальными и для технических объектов, эксплуатируемых на средних широтах.

Таким образом, рассматривая полярные сияния как естественный, а в некоторых случаях и единственно доступный индикатор состояния космической погоды, логично предположить, что достоверность прогноза данного явления коррелирует с уровнем техносферной безопасности за полярным кругом. По этой и иным причинам в последние десятилетия специалистами активно разрабатываются и совершенствуются модели аврорального овала, построенные, как правило, на основе данных многолетнего наблюдения пространственно-энергетических характеристик верхней ионосферы на высоких широтах.

Так, например, наиболее известной моделью такого рода является модель OVATION-Prime (OP) [Newell et al., 2014], базирующаяся на данных 21 года наблюдений спутниками DMSP потоков электронов и протонов разных энергий, принимающая на вход значения параметров солнечного ветра и межпланетного магнитного поля, регистрируемые в первой точке Лагранжа, и прогнозирующая вероятность возникновения сияний с точностью до ~77 % [Vorobev et al., 2022b; Machol et al., 2012]. Известны также прогностическая модель NORUSCA [<http://kho.unis.no/AuroraForecast.html>; Breedveld, 2020], разработанная специалистами норвежской обсерватории Kjell Henriksen (КНО), и диагностическая модель авроральных высыпаний (APM), предложенная в Полярном геофизическом институте (ПГИ) [Vorobjev, Yagodkina, 2005]. Последние модели принимают на вход набор геомагнитных индексов, формируя на выходе примерную геометрию и положение аврорального овала на мо-

мент регистрации входных параметров. К этому же типу можно отнести и модель, разработанную в Аляскинском университете в Фэрбенксе, США [<https://www.gi.alaska.edu/monitors/aurora-forecast>].

Среди средств прямого наземного наблюдения сияний по причине своей доступности широкое распространение получили камеры всего неба (all-sky-камеры) [Lebedinsky, 1961; Sigernes et al., 2014]. Однако эффективность таких наблюдений сильно зависит от условий окружающей среды (засветка небосвода, облачность, туман и пр.) и по самым оптимистическим оценкам не превышает 35–37 %. Существуют попытки и спутниковых оптических наблюдений аврорального овала, однако здесь в большинстве случаев данные фрагментарны, неоднородны, слабоструктурированы или недоступны.

Подводя итог, отметим, что известные на сегодняшний день модели параметров аврорального овала далеки от совершенства и находятся на этапе апробации. Таким образом, определенный теоретический интерес представляет уточнение и формализация взаимосвязей между вариациями геомагнитного поля (ГМП) и наличием сияний. При этом выраженным прикладным характером обладают решения в области разработки и модернизации подходов к диагностированию свойств верхней ионосферы на высоких широтах.

1. ИСХОДНЫЕ ДАННЫЕ, ИХ АНАЛИЗ И ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА

В работе в качестве основного источника данных о наличии полярных сияний используется обсерватория «Ловозеро» (LOZ), входящая в состав ПГИ и являющаяся практически единственной станцией на территории РФ, непрерывно и в течение долгого времени ведущей наблюдения и регистрацию полярных сияний, вариаций магнитного поля и прочих геофизических эффектов высоких широт, обусловленных процессами в магнитосфере, ионосфере и атмосфере Земли. Данные рассматривались за девять лет (2012–2020 гг.) — максимальный период открыто опубликованных результатов синхронных наблюдений полярных сияний и вариаций ГМП в исследуемой точке географического пространства: 67.97° N, 35.02° E (с. Ловозеро, Мурманская область, Россия).

Так, результаты оптических наблюдений сияний в окрестности LOZ традиционно представлены в виде наборов аскаплов (рис. 1) [Ягодкина и др., 2019], публикуемых на сайте ПГИ с 2009 г. [http://pgia.ru/lang/ru/archive_pgi].

Как показывает опыт, устоявшийся с 1970-х гг., формат представления результатов наблюдения сияний (см. рис. 1) в оригинальном виде практически неприемлем в задачах интеллектуального анализа больших объемов такого рода данных. По этой причине традиционные аскапловы предварительно конвертировались в электронные таблицы вида табл. 1.

В результате оцифровки 1035 аскаплов за 2012–2020 гг. (49680 эпизодов 30-минутных наблюдений) было установлено, что только по причине облачности

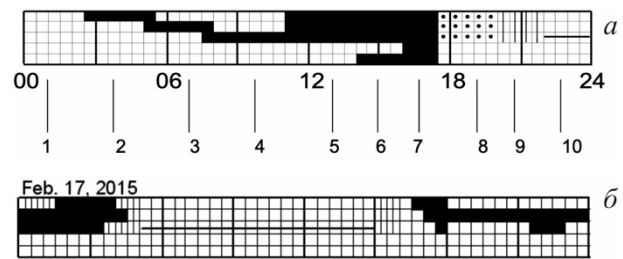


Рис. 1. Формат представления данных в виде аскаплота: 1 — сияние не наблюдается; 2 — сияние в северной области; 3 — сияние в зените; 4 — сияние на юге; 5 — сияние в зените, северной и южной областях; 6 — умеренное сияние в зените, кроме этого, свечение присутствует в северной и южной областях; 7 — сильное сияние в зените, кроме этого, свечение присутствует в северной и южной областях; 8 — частичная облачность; 9 — сплошная облачность; 10 — регистрация не проводилась (а); пример аскаплота обсерватории LOZ за 17.02.2015 г. [PGI Geophysical data, 2015] (б)

доля временных интервалов, не подходящих для наблюдения сияний, значительно превосходит долю благоприятных периодов. Так, например, за девять лет наблюдения полная или частичная облачность в зените относительно обсерватории LOZ препятствовала наблюдению небосвода в ~38.5 % случаев, в то время как доля подходящих для наблюдения фрагментов составила ~21.85 %.

Геомагнитные вариации (ГМВ) в месте наблюдения сияний (с. Ловозеро) регистрировались магнитометром уже другой станции LOZ, расположенной в той же географической локации, однако принадлежащей Мурманскому управлению по гидрометеорологии и мониторингу окружающей среды (Мурманское УГМС). Геомагнитные данные станции LOZ Мурманского УГМС доступны на сайте проекта SuperMag [<https://supermag.jhuapl.edu/mag>], который помимо сбора и хранения геомагнитных данных реализует также некоторые процедуры их предварительной обработки, например, исключение суточной составляющей вариаций ГМП, годового тренда и константы смещения [Gjerloev, 2012]. Подготовленные таким образом данные далее в работе обозначаются символом Δ (например, ΔZ_{LOZ}) и рассматриваются относительно локальной магнитной системы координат NEZ, предложенной SuperMag [Gjerloev, 2012].

В табл. 2 в терминах теории надежности [Vorobev et al., 2022a] приведены результаты оценки полноты временных рядов геомагнитных данных станции LOZ за исследуемый период.

Анализ отсутствующих фрагментов в геомагнитных данных показал, что кратковременные (до 5 мин) отказы магнитометра LOZ (с восстановлением) составляют ~81.5 % от общего числа отказов системы и ~0.27 % от всего времени неработоспособного состояния. Имеющие место в результате таких отказов пропуски здесь восстанавливались линейной интерполяцией без видимого ущерба точности информационного сигнала. Информация, утраченная по причине более длительных эпизодов неработоспособного состояния системы, исключалась из генеральной совокупности и не рассматривалась. Исключались также значения, имеющие выраженный аномальный характер на фоне соответствующих им выборок.

Таблица 1

Фрагмент оцифрованного аскаплота за 17.02.2015 г. (см. рис. 1, б)

Дата	UTC	Область проявления сияний относительно места наблюдения				
		север	зенит	юг	сияние в зените, а также на севере и юге	
					зенит (умеренное)	зенит (сильное)
17.02.2015	00:00	×	1	×	0	0
17.02.2015	00:30	×	1	×	0	0
17.02.2015	01:00	×	1	×	0	0
17.02.2015	01:30	1	1	1	0	0
...
17.02.2015	22:30	0	1	1	0	0
17.02.2015	23:00	0	1	0	0	0
17.02.2015	23:30	0	1	0	0	0

Примечание: 1 — сияние наблюдалось; 0 — сияние не наблюдалось; × — полная или частичная облачность.

Таблица 2

Показатели надежности магнитометра LOZ по данным за 2012–2020 гг.

T , мин	T_W , мин	T_W , %	T_F , мин	T_F , %	N_F	$\langle T2R \rangle$, мин	$\langle T2F \rangle$, мин
4734720	4104638	86.692	630082	13.308	632	996.97	6494.68

Примечание: T — наработка магнитометра LOZ; T_W и T_F — число информативных (общее время работоспособного состояния) и отсутствующих (общее время неработоспособного состояния) значений на выходе магнитометра LOZ за период T ; $\langle T2R \rangle$ и $\langle T2F \rangle$ — среднее время до восстановления рабочего состояния и до отказа магнитометра LOZ соответственно; N_F — число отказов магнитометра LOZ.

Таким образом, в результате предобработки для наблюдения сияний в зените относительно станции LOZ имеем 9408 событий, составляющих множество $GEN \supseteq (Cam \cap Mag)$, где Cam и Mag — подмножества данных наблюдений all-sky-камеры и магнитометра соответственно, 5430 из которых соответствуют отсутствию сияний (подмножество NEG), а 3978 — их наличию (подмножество POS), т. е. $GEN \supseteq (NEG \cup POS)$.

Помимо данных магнитометра LOZ, в качестве признаков с ненулевой значимостью предлагается рассматривать ряд индексов ГМА (SME , SML , SMU , SMR , PCN и др.), публикуемых с шагом дискретизации не менее 30 мин (шаг дискретизации аскаплов), а также параметризованную к значениям SME -индекса интегральную мощность свечения в авроральной зоне [Newell, Gjerloev, 2011]:

$$AP \sim AP_{SME} = 0.048 \cdot SME + 0.241 \sqrt{SME}, \quad (1)$$

где AP — интегральная мощность свечения в авроральной зоне, объединяющая четыре типа полярных сияний [Newell et al., 2010].

2. СТАТИСТИЧЕСКИЕ ВЗАИМОСВЯЗИ МЕЖДУ НАБЛЮДЕНИЕМ СИЯНИЙ В ВИДИМОМ СПЕКТРЕ И ГЕОМАГНИТНОЙ ОБСТАНОВКОЙ

Анализ значимости признаков переменных показал, что наиболее сильной связью с целевой функцией (наблюдение/отсутствие сияний в зените)

обладает модуль первой производной возмущенной составляющей ГМП по времени: $|d\Delta N_{LOZ}/dt|$, $|d\Delta Z_{LOZ}/dt|$ и $|d\Delta F_{LOZ}/dt|$, где $F = (N^2 + E^2 + Z^2)^{1/2}$. Сложившаяся ситуация, по-видимому, свидетельствует об определяющей связи сияний со среднемасштабными турбулентными и волновыми процессами в полярной ионосфере, находящей отражение в данных признаках. Среди значимых признаков второго порядка выделяется $|\Delta N_{LOZ}|$, очевидно, характеризующий связь сияний с интенсивностью аврорального электроджета. Среди глобальных предикторов, неплохую связь с наблюдением сияний демонстрируют индексы суббуревой активности, в частности, SME -индекс и производный от него параметр AP_{SME} .

На рис. 2 показаны распределения наиболее значимых локальных признаков. Видно, что в случае генеральной совокупности (GEN) статистика усредненных по 30 мин значений соответствует логнормальному закону (2) и обобщенному распределению Парето (3), характеризующим тяжелые хвосты, определяемые в области ~85.8 перцентиля (a).

$$PDF(x, s) = \frac{1}{sx\sqrt{2\pi}} \exp\left(-\frac{\log^2 x}{2s^2}\right), \quad (2)$$

$$PDF(x, c) = (1 + cx)^{-1-\frac{1}{c}}, \quad (3)$$

где s и c — параметры формы.

В периоды наблюдения сияний (данные подмножества POS) признаки проявляют аналогичный статистический характер. Сохраняются абсолютные значения границы тяжелых хвостов (6.3 ± 1.1 нТл/мин), вследст-

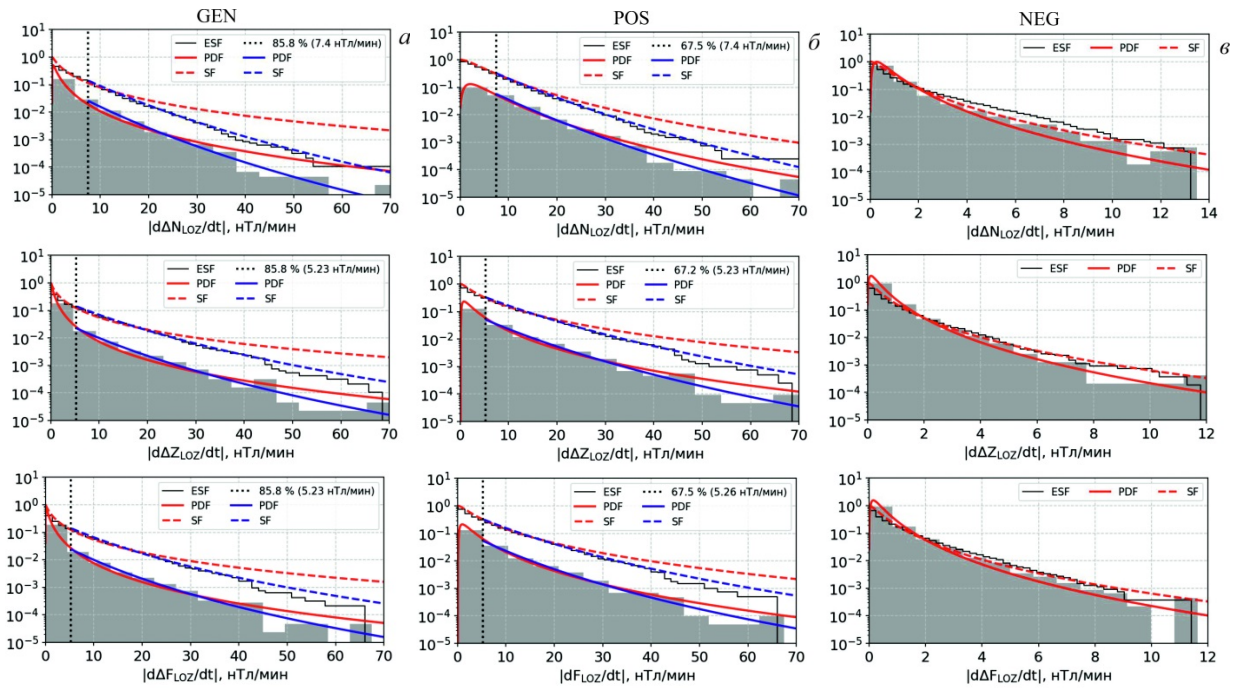


Рис. 2. Распределения локальных признаков: *a* — за весь период наблюдения; *б* и *в* — в периоды наличия и отсутствия полярных сияний в зените относительно станции LOZ соответственно. Красная сплошная и штриховая линии соответствуют функции плотности вероятности *PDF* и функции выживаемости *SF* логнормального закона распределения соответственно; синяя сплошная и пунктирная линии соответствуют *PDF* и *SF* обобщенного закона Парето соответственно; черная сплошная линия — эмпирическая функция выживаемости *ESF*

вие чего граница в относительных величинах опускается с ~85.8 до ~67.5 перцентиля (рис. 2, *a*, *б*). В отсутствие сияний (данные подмножества NEG) распределение значений признаков приближается к однородному и характеризуется исключительно логнормальным законом (рис. 2, *в*).

Кроме этого, статистика признаков в отсутствие сияний имеет максимальные показатели асимметрии и эксцесса, что соответствует наиболее тяжелым хвостам, в то время как асимметрия и эксцесс распределения тех же признаков подмножества POS минимальны. Сложившаяся ситуация указывает на то, что в случае выборок NEG значения признаков наиболее плотно сконцентрированы в окрестности средней величины, вследствие чего даже при малых приращениях признаков переменных вероятность наблюдения сияний резко увеличивается.

Отсюда можно сделать вывод, что геомагнитная обстановка в отсутствие полярных сияний более детерминирована, чем в периоды их наблюдения, т. е. имеет меньшую неопределенность. Другими словами, экстремальные ГМВ практически обеспечивают наблюдение полярных сияний, однако факт наблюдения сияний не гарантирует мощных вариаций ГМП и статистические является следствием таковых лишь в ~1/3 случаев (рис. 2, *б*).

На рис. 3 показаны сияния, наблюдаемые в зените станции LOZ при различном уровне ГМА относительно границы определения тяжелого хвоста (см. рис. 2, *б*). Так, при $|d\Delta N_{LOZ}/dt|$, $|d\Delta Z_{LOZ}/dt|$, $|d\Delta F_{LOZ}/dt| \ll 5$ нТл/мин наблюдаются слабо различимые, смещенные к северу сияния (рис. 3, *a*); при $|d\Delta N_{LOZ}/dt|$, $|d\Delta Z_{LOZ}/dt|$, $|d\Delta F_{LOZ}/dt| \sim 7$ нТл/мин вероятно существование характерной дуги (рис. 3, *б*), пе-

реходящей в активную дугу, спираль или вихрь при $|d\Delta N_{LOZ}/dt|$, $|d\Delta Z_{LOZ}/dt|$, $|d\Delta F_{LOZ}/dt| \gg 7$ нТл/мин (рис. 3, *в*).

Интересно, что статистика наиболее значимых глобальных признаков (рис. 4) структурно повторяет характер распределения локальных предикторов, а распределение AP_{SME} не противоречит известной интерпретации значений интегральной мощности свистости AP в авроральной зоне: при $AP < 20$ ГВт наблюдается слабое или визуально неразличимое сияние; при $20 \leq AP \leq 50$ ГВт наблюдение сияния возможно, однако на небольшом от него расстоянии; при $50 < AP \leq 100$ ГВт сияние можно наблюдать невооруженным глазом; $AP > 100$ ГВт соответствует экстремальной авроральной активности и значимому расширению аврорального овала [Vorobev et al., 2022b].

3. СИНТЕЗ И ВЕРИФИКАЦИЯ ДИАГНОСТИЧЕСКИХ МОДЕЛЕЙ

Как следует из рис. 2, 4, вероятность наблюдения сияний максимальна при превышении признаками некоторых референсных значений: $AP_{SME} > 70$ ГВт; $|d\Delta N_{LOZ}/dt|$, $|d\Delta Z_{LOZ}/dt|$, $|d\Delta F_{LOZ}/dt| > 13$ нТл/мин. В общем случае выполнение данного условия на интервале вероятного наблюдения сияний (рис. 5) дает возможность эффективной диагностики их наличия, например, в условиях полной или частичной облачности.

Однако из тех же рисунков следует и то, что большая доля регистрируемых сияний соответствует значениям предикторов ниже референсных, что вносит неопределенность в результат диагностики и нега-

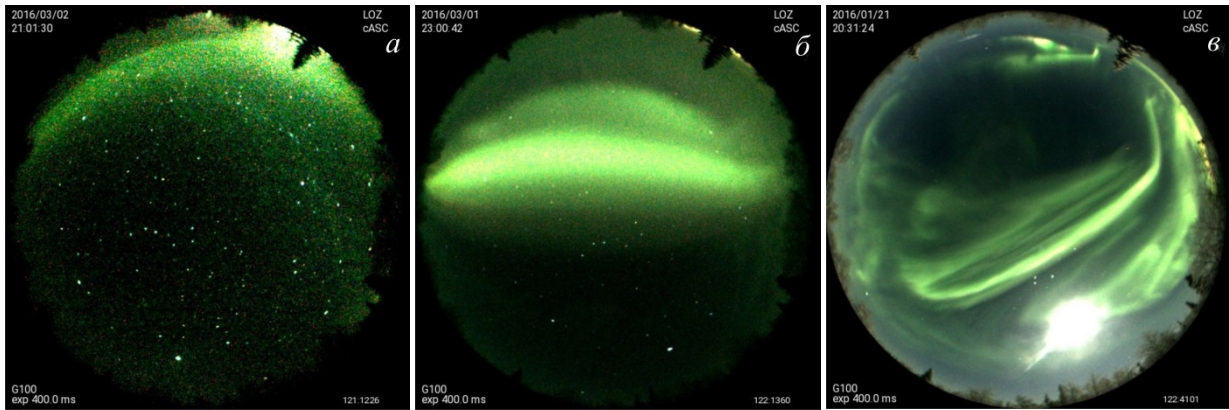


Рис. 3. Кадры камеры всего неба в обсерватории LOZ при $(|d\Delta N_{LOZ}/dt|, |d\Delta Z_{LOZ}/dt|, |d\Delta F_{LOZ}/dt|) \ll 5$ нТл/мин (а); $(|d\Delta N_{LOZ}/dt|, |d\Delta Z_{LOZ}/dt|, |d\Delta F_{LOZ}/dt|) \sim 7$ нТл/мин (б); $(|d\Delta N_{LOZ}/dt|, |d\Delta Z_{LOZ}/dt|, |d\Delta F_{LOZ}/dt|) \gg 7$ нТл/мин (в)

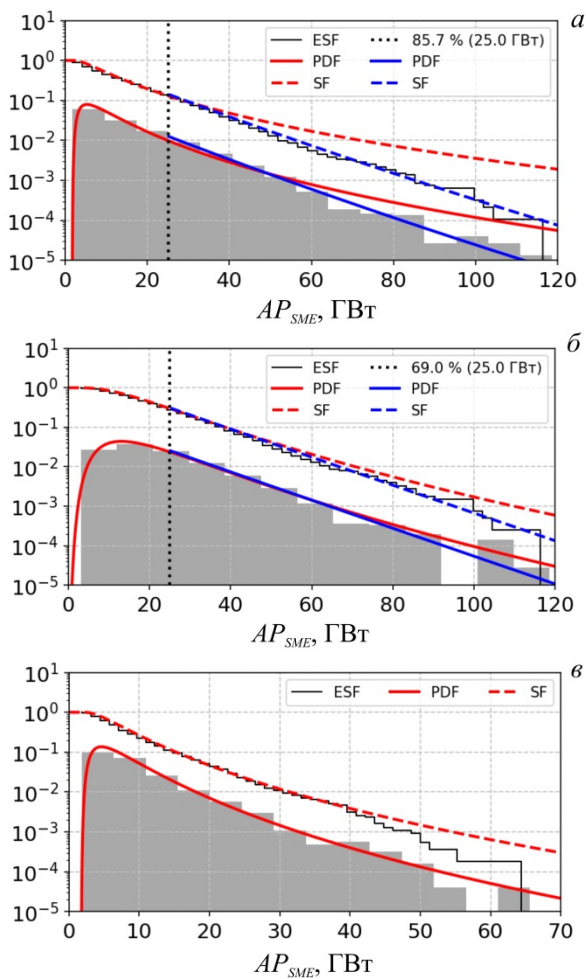


Рис. 4. Статистика параметризованной к значениям SME-индекса интегральной мощности светимости в авроральной зоне: а — за весь период наблюдения; б, в — в периоды наличия и отсутствия сияний в зените относительно станции LOZ соответственно

тивно сказывается на качестве полученных таким образом результатов. Выходом из сложившейся ситуации может служить применение более развитых методов бинарной классификации, например таких, как байесовский классификатор, логистическая регрессия, метод случайного леса и пр.

3.1. Диагностика наличия сияний на основе байесовского вывода

Рассмотрим базовый подход к диагностированию наличия полярных сияний по данным наземных магнитных станций на основе теоремы Байеса:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (4)$$

где $P(A)$ — априорная вероятность гипотезы A или априорное распределение; $P(A|B)$ — вероятность гипотезы A при наступлении события B (апостериорная вероятность); $P(B|A)$ — вероятность наступления события B при истинности гипотезы A ; $P(B)$ — полная вероятность наступления события B , определяемая в соответствии с выражением (5).

$$P(B) = \sum_{i=1}^N P(B|A_i)P(A_i), \quad (5)$$

где вероятности под знаком суммы известны или допускают экспериментальную оценку.

Тогда, в контексте решаемой задачи имеем

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\sim A)P(\sim A)}, \quad (6)$$

где $P(A|B)$ — вероятность того, что в окрестности станции LOZ сияния в зените наблюдаются при достижении признаковой переменной некоторого значения, т. е., например, при $|d\Delta N_{LOZ}/dt| \geq const$; $P(B|A)$ — вероятность того, что при наблюдении сияний в зените в окрестности LOZ значение параметра $|d\Delta N_{LOZ}/dt| \geq const$; $P(A)$ — вероятность наблюдения сияний в зените относительно LOZ; $P(B|\sim A)$ — вероятность того, что в отсутствие сияний в зените в окрестности станции LOZ значение $|d\Delta N_{LOZ}/dt| \geq const$; $P(\sim A)$ — вероятность отсутствия сияний в зените в окрестности станции LOZ.

В табл. 3 в качестве примера показан результат оценки апостериорной вероятности наблюдения сияний в зените LOZ относительно референсных значений параметра $|d\Delta N_{LOZ}/dt|$.

На рис. 6 показана функциональная зависимость апостериорной вероятности наблюдения сияний

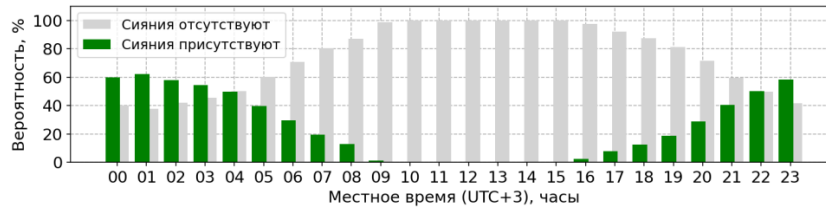


Рис. 5. Суточный ход вероятности наблюдения сияний в зените станции LOZ, рассчитанный на основе аскаплотов по данным, опубликованных ПГИ за 2012–2020 гг.

Таблица 3

Оценка апостериорной вероятности наличия сияний в зените относительно станции LOZ

$ d\Delta N_{LOZ}/dt $, нТл/мин	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6	≥ 7	≥ 8	≥ 9	≥ 10
$P(B A)$, %	91.23	80.22	68.9	59.55	50.73	42.28	35.14	29.49	24.66	20.66
$P(B \sim A)$, %	29.78	12.06	6.67	3.96	2.56	1.62	1.07	0.64	0.35	0.22
$P(A B)$, %	69.18	82.97	88.33	91.68	93.56	95.03	96.02	97.1	98.1	98.56

Примечание: $P(A) = 3978/9408 \approx 42.28\%$; $P(\sim A) = 5430/9408 \approx 57.72\%$.

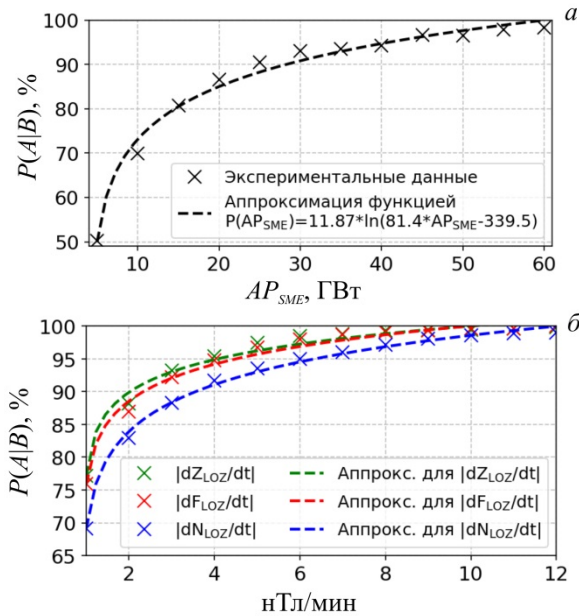


Рис. 6. Характер зависимости апостериорной вероятности наблюдения сияний в зените относительно станции LOZ от наиболее значимых глобальных (а) и локальных (б) признаков переменных

наиболее значимых предикторов. Наблюдаемая зависимость имеет выраженный логарифмический характер и в общем виде может быть аппроксимирована функцией (7). При этом уровень значимости признака обратно пропорционален невязке между эмпирическими данными и аппроксимирующей функцией.

$$P(A|B) \approx P(X) = a \ln(bX + c), \quad (7)$$

где X — признаковая переменная; a, b, c — соответствующие ей параметры формы; для $|dN_{LOZ}/dt|$ $a = 7.04$; $b = 1.32 \cdot 10^5$; $c = -1.14 \cdot 10^5$.

3.2. Диагностика наличия сияний на основе методов машинного обучения

Предварительная оценка метрик качества нескольких классических подходов к бинарной классификации в контексте решаемой задачи обнаружи-

вает некоторое преимущество метода случайного леса. Однако наряду с наилучшей точностью диагностирования данный метод практически не интерпретируем, т. е. в случае его реализации диагностическая модель для конечного пользователя будет представлять собой «черный ящик», генерирующий состояние целевой функции, никак не указывая при этом на механизмы, определяющие результат диагностики.

Оценка важности предикторов на основе критерия Джини [Witlox, 2017] для модели случайного леса позволила независимо от полученных ранее результатов ранжировать признаковые переменные и выявить параметры, вносящие вклад в достоверность результатов диагностики. Полученный результат хорошо соотносится с уже имеющимися данными и главным образом касательно признаков, обладающих наивысшей значимостью.

Таким образом, при отборе девяти наиболее значимых, не связанных между собой линейной зависимостью предикторов (табл. 4) и оптимизации гиперпараметров модели (число деревьев — 400, число случайных признаков для выбора расщепления — $\log_2(M) + 1 \approx 4$, где $M = 9$ — число признаков модели [Mantas et al., 2019], минимальное количество объектов в листьях — 3, максимальная глубина дерева — 8), точность диагностики наличия сияний составила не менее 86.3 % (табл. 4, 5).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (8)$$

где TP и TN — истинно-положительные и истинно-отрицательные диагнозы соответственно; FP и FN — ложно-положительные и ложно-отрицательные.

Диагностика наличия сияний на базе четырех наиболее значимых признаков (см. табл. 4) возможна с точностью от 85.7 %, время обучения модели при этом сокращается в ~ 4.1 раза и составляет 12.6 с. В случае использования в качестве входных параметров исключительно SME - и SMR -индексов доля ошибок модели не превышает 19.7 % (т. е. $Accuracy \geq 80.3\%$), а получаемый результат может быть актуа-

Таблица 4

Метрики качества диагностических моделей для различных наборов входных признаков

Входные признаки	Метрика	AUC^*	$Accuracy^{**}$
$ d\Delta N_{LOZ}/dt $, $ d\Delta E_{LOZ}/dt $, $ d\Delta Z_{LOZ}/dt $, $ d\Delta F_{LOZ}/dt $, $ \Delta N_{LOZ} $, $ \Delta E_{LOZ} $, $ \Delta Z_{LOZ} $, SME , SMR		0.927	0.863
$ d\Delta N_{LOZ}/dt $, $ d\Delta Z_{LOZ}/dt $, $ d\Delta F_{LOZ}/dt $, $ \Delta N_{LOZ} $		0.924	0.857
SME , SMR		0.879	0.803

Примечание: AUC (англ. Area Under Curve) — площадь, ограниченная ROC-кривой [Hand, Till, 2001] и осью доли ложных положительных классификаций; $Accuracy$ определяется выражением (8).

Таблица 5

Оценка качества диагностирования наличия сияний на тестовой выборке (25 % от GEN), при использовании на входе девять наиболее значимых признаков переменных

Сияния в зените Диагноз сияний	Наблюдались	Не наблюдались
	положительный	1227
отрицательный	182	803

лен для всего множества точек, находящихся на геомагнитной широте станции LOZ в ночное время.

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Проведенные исследования убедительно указывают на возможность эффективного диагностирования наличия полярных сияний посредством методов машинного обучения, статистического и интеллектуального анализа геомагнитных данных наземных источников. При этом полученные результаты обладают практической значимостью в задачах организации поддержки принятия решений при идентификации сияний путем неавтоматизированного анализа кеограмм и/или данных all-sky-камер. В перспективе область применения полученных результатов может быть расширена до оперативной диагностики состояния верхней ионосферы на высоких широтах, оценки рисков сбоя систем коротковолновой радиосвязи и экстремальных ошибок глобальных навигационных спутниковых систем, эксплуатируемых в Арктическом регионе.

По причине изначально достаточно слабых признаков точность полученных моделей относительно невелика и составляет ~86 %. В связи с этим дополнительных исследований в первую очередь требуют вопросы, касающиеся поиска более сложных и сильных синтетических предикторов, на существование которых однозначно указывают результаты анализа главных компонент [Jolliffe, 2002]. Также имеет смысл рассмотреть вопросы систематизации и комплексной обработки данных нескольких высокоширотных магнитометров и all-sky-камер в небольшом субрегионе (например, ограниченном обсерваториями «Ловозеро», «Апатиты» и «Верхнетулумский»).

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Имеющееся разнообразие подходов к мониторингу космической погоды и геофизических параметров в области аврорального овала не решает проблемы эффективного прогнозирования и диагностирования полярных сияний как особого состояния

верхней ионосферы на высоких широтах. При этом неизменный с 1970-х гг. формат представления данных с all-sky-камер в виде аскаплов на сегодняшний день практически утратил свою актуальность и требует глубокой модернизации, например, путем организации автоматизированной разметки данных, сокращения шага дискретизации и эволюции самого подхода к классификации наблюдаемых событий.

Метод рекурсивного исключения признаков [Kuhn, Johnson, 2019] и оценка критерия взаимной информации [Vaudot et al., 2019] указывают на то, что наиболее сильную связь с существованием сияний в оптическом диапазоне демонстрирует первая производная по времени северной и вертикальной составляющих вариаций ГМП, что, по-видимому, свидетельствует об определяющей связи сияний со среднemasштабными турбулентными и волновыми процессами в полярной ионосфере. Среди значимых признаков второго порядка можно выделить $|\Delta N_{LOZ}|$ и SME -индекс, очевидно характеризующие связь сияний с интенсивностью аврорального электроджета и уровнем суббуревой активности в целом.

Статистический анализ геомагнитных данных за девять лет (2012–2020 г.) указывает на то, что геомагнитная обстановка в отсутствие сияний имеет меньшую степень неопределенности, чем в периоды их наблюдения. Другими словами, экстремальные ГМВ обеспечивают наличие сияний в оптическом диапазоне, однако факт наблюдения сияний не гарантирует существования мощных вариаций ГМП и статистически является следствием таковых лишь в ~1/3 случаев.

Зависимость апостериорной вероятности наблюдения сияний от наблюдаемых признаков имеет выраженный логарифмический характер и в общем виде может быть аппроксимирован функцией типа $P(A|B) \approx P(X) = a \ln(bX + c)$, где X — признаковая переменная; a , b , c — соответствующие ей параметры формы (для $|dN_{LOZ}/dt|$ $a=7.04$; $b=1.32 \cdot 10^5$; $c=-1.14 \cdot 10^5$). При этом степень значимости признака

обратно пропорциональна невязке между эмпирическими данными и аппроксимирующей функцией.

Точность локального диагностирования наличия сияний по геомагнитным данным на основе метода случайного леса и ряда наиболее значимых признаков переменных составляет ~86 %. Диагностическая модель на базе глобальных геомагнитных индексов имеет ожидаемо меньшую точность ~80.3 %, однако получаемые на ее основе данные могут быть использованы при верификации известных глобальных диагностических моделей аврорального овала со схожим набором входных параметров (например, модель авроральных высыпаний [Vorobjev, Yagodkina, 2005]).

Исследование выполнено при поддержке Российского научного фонда (проект № 21-77-30010).

Авторы выражают благодарность рецензентам за внимательный анализ работы и большое число конструктивных замечаний, а также Токареву В.И. — космонавту-испытателю Центра подготовки космонавтов им. Ю.А. Гагарина за сведения об уникальном опыте пилотирования летательных аппаратов в воздушном пространстве Арктики в периоды экстремальной геомагнитной активности.

СПИСОК ЛИТЕРАТУРЫ

Захаров В.И., Чернышов А.А., Милох В., Джин Я. Влияние ионосферы на параметры навигационных сигналов GPS во время геомагнитной суббури. *Космические иссл.* 2020. Т. 60, № 6. С. 769–782. DOI: [10.7868/S0023420616010143](https://doi.org/10.7868/S0023420616010143).

Пилипенко В.А. Воздействие космической погоды на наземные технологические системы. *Солнечно-земная физика.* 2021. Т. 7, № 3. С. 72–110. DOI: [10.12737/2712-9640-2021-7-3](https://doi.org/10.12737/2712-9640-2021-7-3).

Птицына Н.Г., Тясто М.И., Касинский В.В., Ляхов Н.Н. Влияние космической погоды на технические системы: свои железнодорожной аппаратуры во время геомагнитных бурь. *Солнечно-земная физика.* 2008. Вып. 12. Т. 2, № 125. С. 360.

Соколова О.Н., Сахаров Я.А., Грицуненко С.С., Коровкин Н.В. Алгоритм анализа устойчивости энергосистем к геомагнитным бурям. *Изв. РАН. Энергетика.* 2019. № 5. С. 33–52. DOI: [10.1134/S0002331019050145](https://doi.org/10.1134/S0002331019050145).

Ягодкина О.И., Воробьев В.Г., Шекунова Е.С. Наблюдения полярных сияний над Кольским полуостровом. *Труды Кольского научного центра РАН.* 2019. Т. 10, № 8-5. С. 43–55. DOI: [10.25702/KSC.2307-5252.2019.10.8](https://doi.org/10.25702/KSC.2307-5252.2019.10.8).

Baudot P., Tapia, M., Bennequin D., Goillard J.-M. Topological information data analysis. *Entropy.* 2019. Vol. 21, iss.9. P. 869. DOI: [10.3390/e21090869](https://doi.org/10.3390/e21090869).

Breedveld M.J. *Predicting the Auroral Oval Boundaries by Means of Polar Operational Environmental Satellite Particle Precipitation Data.* Master Thesis. Arctic University of Norway. June 2020.

Gjerloev J.W. The SuperMAG data processing technique. *J. Geophys. Res.* 2012. Vol. 117, iss. A9. P. A09213. DOI: [10.1029/2012JA017683](https://doi.org/10.1029/2012JA017683).

Hand D.J., Till R.J. A Simple generalization of the area under the ROC curve for multiple class classification problems. *Machine Learning.* 2001. Vol. 45. P. 171–186. DOI: [10.1023/A:1010920819831](https://doi.org/10.1023/A:1010920819831).

Jolliffe I.T. *Principal Component Analysis.* Ser.: Springer Series in Statistics, 2nd ed., Springer, NY, 2002. XXIX. 487 p.

Kuhn M., Johnson K. *Feature Engineering and Selection: A Practical Approach for Predictive Models.* CRC Press, 2019. 298 p.

Lebedinsky A.I. Synchronous auroral registration by all-sky camera C-180 and patrol spectrograph C-180-S. *Ann. the Intern. Geophys. Year.* 1961. Vol. XI.

Machol J.L., Green J.C., Redmon R.J., et al. Evaluation of OVATION Prime as a forecast model for visible aurorae. *Space Weather.* 2012. Vol. 10, iss. 3. P. S03005. DOI: [10.1029/2011SW000746](https://doi.org/10.1029/2011SW000746).

Mantas C.J., Castellano J.G., Moral-García S., Abellán J. A comparison of random forest based algorithms: random credal random forest versus oblique random forest. *Soft Computing.* 2019. Vol. 23. P. 10739–10754. DOI: [10.1007/s00500-018-3628-5](https://doi.org/10.1007/s00500-018-3628-5).

Newell P.T., Gjerloev J.W. Substorm and magnetosphere characteristic scales inferred from the SuperMAG auroral electrojet indices. *J. Geophys. Res.* 2011. Vol. 116, iss. A12. P. A12232. DOI: [10.1029/2011JA016936](https://doi.org/10.1029/2011JA016936).

Newell P.T., Sotirelis T., Wing S. Seasonal variations in diffuse, monoenergetic, and broadband aurora. *J. Geophys. Res.* 2010. Vol. 115, iss. A3. P. A03216. DOI: [10.1029/2009JA014805](https://doi.org/10.1029/2009JA014805).

Newell P.T., Liou K., Zhang Y., et al. OVATION Prime-2013: Extension of auroral precipitation model to higher disturbance levels. *Space Weather.* 2014. Vol. 12, iss. 6. P. 368–379. DOI: [10.1002/2014SW001056](https://doi.org/10.1002/2014SW001056).

PGI Geophysical data. January, February, March 2015 / Ed. V. Vorobjev. Murmansk, Apatity: PGI KSC RAS. 2015.

Sigernes F., Holmen S.E., Biles D., et al. Auroral all-sky camera calibration. *Geoscientific Instrumentation, Methods and Data Systems.* 2014. Vol. 3, iss. 2. P. 241–245. DOI: [10.5194/gi-3-241-2014](https://doi.org/10.5194/gi-3-241-2014).

Soloviev A.A., Sidorov R.V., Oshchenko A.A., Zaitsev A.N. On the need for accurate monitoring of the geomagnetic field during directional drilling in the Russian Arctic. *Izvestiya. Physics of the Solid Earth.* 2022. Vol. 58. P. 420–434. DOI: [10.1134/S1069351322020124](https://doi.org/10.1134/S1069351322020124).

Vorobev A., Soloviev A., Pilipenko V., et al. An approach to diagnostics of geomagnetically induced currents based on ground magnetometers data. *Applied Sciences.* 2022a. Vol. 12, iss. 3. P. 1522. DOI: [10.3390/app12031522](https://doi.org/10.3390/app12031522).

Vorobev A.V., Soloviev A.A., Pilipenko V.A., Vorobeva G.R. Interactive Computer model for aurora forecast and analysis. *Solar-Terr. Phys.* 2022b. Vol. 8, no 2. P. 84–90. DOI: [10.12737/stp-82202213](https://doi.org/10.12737/stp-82202213).

Vorobjev V.G., Yagodkina O.I. Effect of magnetic activity on the global distribution of auroral precipitation zones. *Geomagnetism and Aeronomy.* 2005. Vol. 45. P. 438–444.

Witlox F. Gini Coefficient. *International Encyclopedia of Geography: People, the Earth, Environment and Technology.* 2017. DOI: [10.1002/9781118786352.wbieg0855](https://doi.org/10.1002/9781118786352.wbieg0855).

Yasyukevich Y., Astafyeva E., Padokhin A., et al. The 6 September 2017 X class solar flares and their impacts on the ionosphere, GNSS, and HF radio wave propagation. *Space Weather.* 2018. Vol. 16, iss. 8. P. 1013–1027. DOI: [10.1029/2018SW001932](https://doi.org/10.1029/2018SW001932).

Yasyukevich Y., Vasilyev R., Ratovsky K. Small-scale ionospheric irregularities of auroral origin at mid-latitudes during the 22 June 2015 magnetic storm and their effect on GPS positioning. *Remote Sensing.* 2020. Vol. 12, no 10. P. 1579. DOI: [10.3390/rs12101579](https://doi.org/10.3390/rs12101579).

Как цитировать эту статью:

Воробьев А.В., Соловьев А.А., Пилипенко В.А., Воробьева Г.Р., Гайнетдинова А.А., Лапин А.Н., Белаховский В.Б., Ролдугин А.В. Локальная диагностика наличия полярных сияний на основе интеллектуального анализа геомагнитных данных. *Солнечно-земная физика.* 2023. Т. 9, № 2. С. 26–34. DOI: [10.12737/szf-92202303](https://doi.org/10.12737/szf-92202303).