USING MACHINE LEARNING TO CREATE A CATALOG OF SOLAR FLARES FROM OBSERVATIONS ON THE SIBERIAN RADIOHELIOGRAPH

© 2025 Yu. N. Shamsutdinova, D. V. Rozhkova, L. K. Kashapova*, A. V. Gubin

Institute of Solar-Terrestrial Physics, Siberian Branch, Russian Academy of Sciences, Irkutsk, Russia.

**e-mail: lkk@iszf.irk.ru*

Received February 24, 2025

Revised April 29, 2025

Accepted June 17, 2025

Abstract. The paper presents and discusses the results on the creation of the Siberian Radioheliograph (SRH) solar flare catalog using machine learning methods. The high sensitivity of the tool, as well as the use of time profiles of the sum of correlation coefficients of antenna pairs (correlation curves) to search for events allowed us to include in the catalog weak events that are weakly discernible on time profiles of the radiation flux. To select candidate events, we proposed and tested a technique that allows us to determine the beginning, maximum, and end of a solar flare (event) by analyzing the derivative of the time profile given by a numerical function. Since the purpose of the catalog is to select broadband events, a criterion was introduced that allows automatic selection of an event depending on the simultaneous response at several frequencies. To clarify the solar nature of the events and the quality of the observational data, the Support Vector Method (SVM) was applied in a test mode. The amount of observational data acquired by the SVM in the second half of 2023 and for 2024 provided extensive material for both model training and testing. It was applied to time profiles obtained over the 9-10 GHz band to divide into "flare", "background" and "artifact" classes.

Keywords: *solar flares, microwave radiation, machine learning methods, catalogs and classification*

DOI: 10.31857/S00167940250723e5

1. INTRODUCTION

The advent of new highly sensitive instruments that observe the Sun simultaneously at multiple frequencies and have high temporal and spatial resolution has led to a significant increase in the amount of information to be processed. There is a need for efficient automated selection of events (solar flares) and their cataloging in large data streams. Currently, techniques for automated detection and identification of solar flares are used for observations by several space-based and

ground-based instruments.

Some of the best known observations are the time profiles of soft X-ray emission recorded by instruments aboard the Geostationary Operational Environmental Satellite (GOES) system[1] . The methodology used allows us to identify solar flares by analyzing the temporal profiles of the radiation flux in the spectral range of 1-8 Å. The results of analyzing the time profiles of soft X-ray radiation allow not only to identify events, but also to determine the class of events (flares) according to the GOES classification (A, B, C, M, X). Operational information on the occurrence of a flare with the ability to estimate its power, as well as catalogs of solar flares, created on the basis of GOES data, are in demand both for research in solar physics and geophysics, for example, studying the effect of solar activity on the Earth's magnetosphere and ionosphere, forecasting space weather phenomena. However, the GOES catalog in some cases, due to the peculiarities of the algorithm, may miss weak, slowly developing events or flares with a weakly pronounced maximum.

Another example of a catalog of solar flares observed in the X-ray range is the solar flare database of the Konus-Wind (KW) instrument[2] [Lysenko et al., 2022]. This catalog contains data on events recorded by the KW instrument in trigger mode with high temporal and spectral resolution. However, since only the events that were registered in the trigger mode are selected, the catalog is obviously incomplete and does not include slowly evolving events for which, for some reason, the threshold level necessary for triggering the trigger was not reached. Thus, there is a need to create another catalog of events observed outside the trigger mode.

The best known catalogs of solar flare observations in the microwave range are those based on the Nobeyama radio polarimeter data[3] . This catalog may also be incomplete, missing short or faint events, as well as events with an insufficiently broad spectral range.

The observations of the Siberian Radioheliograph (SRG, [Altyntsev et al., 2020]) make it possible to record events of different power and duration in the range 3-24 GHz with high spectral and temporal resolution. The temporal resolution of about 3.5 seconds and the high sensitivity of the SRG to the radiation flux allows the registration of both short and rather narrow-band events and weak flares, which are often missed in many solar flare catalogs. The high sensitivity can result in artifacts being sampled along with real events. With large data streams, visual inspection of the results obtained is inefficient, and it becomes necessary to involve machine learning methods to refine the type of selected events. Therefore, the task of our research was to develop an original

[1] https://www.swpc.noaa.gov/products/goes-x-ray-flux

[2] https://www.ioffe.ru/LEA/kwsun/index.html

[3] https://solar.nro.nao.ac.jp/norp/html/event/

technique for automatic detection of events in multi-frequency data obtained with the help of AWG and to test the possibilities of its correction using machine learning methods.

2. PRIMARY SELECTION OF CATALOG EVENTS

The event selection method used in this work is based on analyzing the derivative of the time profile represented as a continuous numerical function. Since the evolution of a solar flare is described as a sudden increase in flux followed by a continuous decline, we can use the derivative function to determine the start, maximum, and end of the event. First, for each spectral band or wavelength range, the standard deviation (sigma) of the derivative of the time profile in the event-free time interval is calculated. Then, values of the derivative that fall outside a positive and negative threshold, defined as three sigma, are selected. The use of a three sigma threshold is justified by the fact that in a normal distribution of measurements from a physical event, approximately 99.7% of the data are within three sigmas of the mean. Thus, if the derivative falls outside this threshold, it may indicate the presence of an event rather than random fluctuations or noise. The gradient of these sampled values is then calculated to identify areas with significant changes, indicating the start and end points of the event. Since the average temporal resolution of multi-wavelength AWG observations is 3-4 seconds, thus the shortest burst of physical nature cannot last less than three time samples or about 10-12 seconds. Since a burst lasting 12 seconds may also be associated with noise, for confident event detection we increase this interval by a factor of about three and assume that the event is a burst on the time profile that lasts longer than 50 seconds. The exact moment of maximum is defined as the maximum value of the time profile between the found moments of the beginning and end of the event. The algorithm is implemented in the Python programming language.

You can also use the sign function from the NumPy standard library to search for events, analyzing, as in the first algorithm, the derivative of the time profile. The function returns the value 1 if the number is positive; 0 if the number is zero; - 1 if the number is negative. With the help of the sign function we can determine the key points of the event: the beginning, the maximum and the end. Event maxima are found by assuming that the derivative goes from a positive value to a negative value. The beginning of the event is defined as the point where the derivative changes sign from negative to positive, indicating a sharp increase in the signal on the time profile. In addition, the derivative of the function in the time interval from the beginning of the event to its maximum should monotonically increase, reaching its maximum value in the middle of this time interval. The end of the event is the moment where the derivative also changes sign from negative to positive, but the derivative of the function in the time interval from the maximum of the event to its end should monotonically fall, reaching the minimum value. In addition, if the time interval between the moments when the event maxima occur is less than four minutes (240 seconds), then such events

are treated as a single event. In order to remove noise that the algorithm may erroneously select as an event, we also set a threshold of three sigma values. We consider that if the value of the time profile derivative goes beyond this set threshold, it corresponds to an event.

The purpose of developing the algorithms described above was to analyze observational data obtained with the Siberian Radioheliograph (SRH). The SRG is an interferometer consisting of three T-shaped independent antenna arrays with frequency ranges of 3-6, 6-12, 12-24 GHz. All antenna arrays record fluxes at 16 frequencies [Altyntsev et al., 2020]. One of the types of data obtained from the AWG are correlation curves, which are the time profile of the sum of complex covariances calculated for different pairs of antennas. As shown in [Lesovoy and Kobets, 2017] on the example of the AWG prototype in the 4-8 GHz range for compact sources, which include solar flares, the sensitivity of correlation curves in terms of flux density reaches 0.01 solar flux units (s.e.p.), which makes it possible to detect weak bursts that are difficult to discern on the intensity time profiles. The AWG is a more sensitive instrument than the 4-8 GHz prototype. However, the ability to detect weak events using correlation curves is also higher than using temporal intensity profiles. Therefore, in this paper, we used correlation curves for event recognition and selection.

Since the AWG is a multi-wavelength instrument, the algorithms developed were augmented with the ability to automatically select events based on simultaneous response at multiple frequencies. In most cases, a narrowband burst may be an artifact. In order to exclude such events, a detected spike on the time profile is considered an event only if there is a response on at least 15 frequencies, including at least 5 responses in two different spectral bands in which the AWG is observing. The choice of the minimum number of frequencies is connected with the necessity to ensure also the broadband nature of the selected events. It is known that the 3-6 GHz AWG antenna array has the minimum spectral resolution and, accordingly, observes in the narrowest band - 3 GHz. In order to select events whose radiation is observed at a smaller spectral width, a criterion of quasi-simultaneous responses at 15 frequencies was introduced. However, instrumental problems leading to false bursts can be observed on all 16 frequencies of one of the antenna arrays at once, so to minimize the number of false events, the criterion that the response must be observed on at least two arrays simultaneously was introduced. The moment of maximum spikes on time profiles at different frequencies relating to the same event may vary. These time variations may be caused by the characteristics of the observations at the AWG or by physical processes occurring during the outburst. While the difference between event maxima will not be greater than 4 seconds due to frequency overshooting during the AWG observations, the delay due to physical processes may result in a larger variation in the time values. According to the standard model of solar flare and gyrosynchrotron flare emission, high frequencies (above 6 GHz) are associated with acceleration processes. At the same time, emission at low frequencies is formed near the loop top and can be

associated with processes such as chromospheric acceleration, which usually gives a delay of about 3-5 min in reaching the maximum of the emission intensity. Therefore, we assumed that event maxima at different frequencies represent the same event if the moments of their registration are within 200 seconds, based on general ideas about solar flare dynamics. An example of the performance of the two algorithms for the observations on May 6, 2024 is shown in Fig. 1. For each day, a diagram was plotted plotting the position of the moment of maximum from the frequency. Those events that satisfied the broadband event criterion are highlighted with vertical lines. It can be seen that the first method selected 24 events and the second method selected 33 events. Overall, the first method selected 3420 events for the year 2024, while the second method shows the presence of 6920 events. This strong difference may have several reasons. The first is that the second method was more sensitive to weak bursts, and it was able to detect bursts at frequencies that were not available to the first method. As a result, more events were identified as broadband. The second reason may be due to the fact that the second method splits a single event into multiple bursts, and is more sensitive to fine temporal structure. An example of such an event is the burst that occurred on May 6, 2024. The first method perceived this burst as a single event with a maximum at 01:06 UT (shown by the circle in Fig. 2). The second algorithm, in turn, identified three events with maxima at 00:52 UT, at 01:03 UT, and at 01:06 UT (shown by crosses in Fig. 2). Physically, these could be either a single event or a combination of events that occurred in different parts of the solar disk. To answer this question, we need to construct an image and analyze the position of the sources. But this is a problem with all catalogs of observations without spatial resolution. A preliminary conclusion about the reasons for the differences can be made by analyzing the duration distribution of events detected by different methods. The results of this analysis are shown in Fig. 3.

In Fig. 3 shows that the number of events detected by the first method, 567 flashes, which is 16.6%, lasted less than five minutes. The number of flashes lasting between 5 and 10 minutes is only 611 (17.9%). The most significant portion, namely 2,186 flashes (63.9%), were between 10 and 100 minutes in duration. Events that lasted more than 100 minutes represent only 2% of the total. The second method, which in turn detected significantly more events over the same time period, shows a different distribution in terms of duration. Of these, 1215 events, representing 17.5%, lasted less than five minutes, which is close to the results of the first method. However, the number of events lasting between 5 and 10 minutes is much higher, amounting to 4155 (60%). At the same time, 1550 flashes (22.5%) lasted more than 10 minutes, which is less than that of the first method for the category of 10 to 100 minutes. The result of comparing the duration distribution of events sampled by different methods may indicate that the first algorithm selects outbursts by combining outbursts into long duration events, while the second algorithm treats each outburst as a

separate event in the range of 5 to 10 minutes.

3. USING MACHINE LEARNING TO RECOGNIZE BURSTS AND INSTRUMENTAL ARTIFACTS

Obviously, microwave observations are not free from artifacts that can be mistaken for flashes during selection. This may be caused by nonlinear dependencies in the calculation of correlation curves (failure of one of the antennas), as well as by the presence of artifacts caused by instrumental errors or by penetration into the radiation pattern of spacecraft radiating in the observation range of the instrument (e.g., geostationary artificial Earth satellite). In addition, some of the events may be related to variations in the background radiation. Recognizing the types of such profiles requires methods capable of considering temporal dependencies and extracting features for classification.

In our work, we analyze the possibilities of using machine learning to identify solar flares and artifacts using the Support Vector Machine (SVM) method [Cortes and Vapnik, 1995] as an example. The algorithm we have chosen belongs to the type of algorithms "with a teacher" and is well suited for the task of classifying univariate data, especially in the presence of a small dataset. SVM performs well on nonlinear data characterized by time profiles. Through the use of kernel functions such as polynomial, sigmoid and radial basis function (RFB), the data is transformed into a higher dimensional space where linear separation becomes possible. SVM shows good results in anomaly and fault detection [Widodo, Yang 2007], also widely used in solar physics. For example, the authors of [Hao et al., 2024] applied SVM to the task of solar flare classification (no flares, weak flares, strong flares) using data from the RHESSI mission and solar spectra from the HARPS-N instrument.

For training, we used events observed in 2023-2024 in the frequency range from 9.0 to 10.2 GHz, selected by the algorithm that was described in Section 2. The choice of frequency is due to the fact that gyrosynchrotron emission of a non-thermal nature is confidently recorded at this frequency for most solar flares. This is due to the fact that in most cases the microwave emission from solar flares in this frequency range is the emission from an optically thin source, which means that there are no additional effects associated with heating and matter motion that complicate and distort the shape of the time profiles characteristic of solar flares. This can lead to an artificial increase in the subtypes of time profile shape and complicate the training and recognition of solar flares.

Before using the data for analysis by machine learning methods, a number of pre-processing and preparation steps are necessary. In the first step, time profiles averaged over the frequency range 9.0-10.2 GHz were obtained. Next, each time profile was normalized by intensity. To make the data suitable for training, we applied resampling to make the number of elements in each time profile equal. To do this, we first determined the maximum dimensionality of the temporal profile

in the entire dataset, and then we resampled each temporal profile to this dimensionality using interpolation. To apply interpolation for each time profile, we first calculate the time step Δt required to form a new time series. Δt is defined as the ratio of the difference between the last and the first time point to the number of desired time points after interpolation (in our case, the maximum dimensionality of the time profile). Next, we created a new time series (vector) with a uniform step Δt. We perform linear interpolation of the time profile onto the coordinate grid defined by the new time series.

In this way, we have increased the number of elements of each time profile to the maximum dimensionality among time profiles. Since each value of each element represents a feature, the number of elements in each profile was reduced using a decimation method with a filter coefficient equal to the ratio of the initial maximum profile dimension to its dimension after interpolation. In this paper, each univariate array has 120 elements, which corresponds to the number of features. In total, we used a dataset consisting of 2653 events from June 2023 to August 2024 for training and testing. Using machine learning algorithms "with a teacher" requires data partitioning, which allows the model to be trained based on known examples. The partitioned time profiles are the input values, while their labels serve as outputs. In our work, we distinguished three classes: 'flash', 'background', and 'artifacts'. Examples of temporal profiles for each class, as well as the preprocessing process to bring them to a common dimensionality, are shown in Fig. 4.

The distribution of data across classes is presented in Fig. 5, which shows that the data is unbalanced. The class "flash" is 1591, "background" is 932, and "artifacts" is 130. To solve this problem, we used the *RandomOverSampler* method of the *imblearn* library, which is based on the SMOTE (Synthetic Minority Oversampling Technique) algorithm. SMOTE creates synthetic data for minority of classes by interpolating existing data points, which helps to balance the classes and improves the performance of machine learning models. After applying the method, the values of all classes became equal to the value of the largest class, 1591.

For classification tasks involving three or more classes, the scikit-learn Support Vector Classification (SVC) module is used based on the One-vs-All and One-vs-One approaches, which involves creating a separate classifier for each class [Rifkin and Klautau, 2004]. This algorithm finds optimal separating hyperplanes between classes in a multidimensional feature space, which makes it effective for multiclass problems. The table summarizes the hyperparameters optimized for models with different kernels using the *GridSearch* module from the *scikit-learn* library.

To visualize the efficiency of the algorithm, we use an error matrix, a table where the actual class values are plotted on the rows and the predicted values are plotted on the columns. If the predicted class corresponds to the actual one, the prediction outcome is called True Positive (TP) and True Negative (TN). Examples of classes for which the prediction outcome is false are called

False Positive (FP) and False Negative (FN), which means that the model made errors. The main diagonal represents correct predictions (True Positive and True Negative) and the reverse diagonal represents errors. Fig. 6 shows the error matrices for the three SVM models with different kernels. The error matrix of the model with the "RBF" kernel shows the best result of correctly predicted classes on the diagonal, indicating the model's ability to accurately predict most of the data points in each class correctly, especially when compared to the model with the "sigmoid" kernel. This confirms that the model has learned to correctly classify events using temporal profiles.

To evaluate the performance of the models, we used a set of metrics used in [Hao et al., 2024]. The authors used the following metrics: aggregate accuracy measures the average performance of the model across all classes, showing how accurately it predicts the correct classes for all examples in the dataset (labeled "accuracy" in Fig. 7); categorical accuracy measures the proportion of correctly classified data for each class (labeled "flash", "background", "noise" in Fig. 7). Fig. 7 demonstrates that the SVC model with radial-basis function performed the best for both aggregate accuracy and categorical accuracy.

Fig. 8 shows the metrics of precision (accuracy), completeness (recall) and F1-measure for each model. The chart also shows the best result for the SVC model with RBF kernel. The obtained scores confirm that the radial basis function allows the support vector method to better handle non-linear data and separate classes with distribution structure in feature space.

4. DISCUSSION

This work tested automatic methods for selecting solar events observed on the Siberian Radioheliograph to create a catalog of broadband microwave bursts and classify them. Both algorithms using the classical approach (analyzing the derivative of the time profile) and one of the machine learning methods were applied. To increase sensitivity and detect weaker events, time profiles of the sum of correlation coefficients between pairs of antennas, or correlation curves, rather than radiation fluxes at each frequency were used. This also reduced the influence of noise.

The primary selection of candidate events was performed using a technique that allows us to determine the beginning, maximum, and end of a solar flare (event) by analyzing the derivative of the time profile given by a numerical function. Since the purpose of the catalog is to select broadband events, a criterion was introduced to automatically select an event depending on the simultaneous response at several frequencies. To identify bursts at each of the frequencies, two methods using close algorithms were used - one is a unique one created to determine the start and end moments of a burst, and the other is a standard method from one of the Python libraries. However, when comparing the selection results of the two methods, it was found that the method using the standard function selected almost twice as many events as the other. At the same time, the pattern of distribution of selected events by duration changed. Most of the events selected by the

formally productive method were between 5 and 10 minutes in duration, while the contribution of longer events decreased.

We attribute this to the fact that the first method relies on searching for rapid changes in derivative values to find the start and end moments of the signal burst. The limitation of this method is its inability to resolve a complex composite structure, but it successfully selects simple pulsed events and classical long duration solar flares. In contrast, the second method is based on finding two extrema between the beginning and the end of the burst. It reveals subtle details of the complex temporal structure of the burst. In doing so, it can break a single event into several separate bursts. There is no contradiction in this from the point of view of solar flare physics, but additional analysis is required when investigating a specific event. However, the information obtained at this stage on solar microwave events from AWG data during the second half of 2023 and for 2024 has provided extensive material for both model training and testing.

To clarify the solar nature of the events and the quality of the observational data, the Support Vector Classification (SVC) algorithm from the scikit-learn library, based on the classical support vector method (SVM), was used in the test mode. It was applied to time profiles obtained in the 9-10 GHz band to divide into classes "flash", "background" and "artifact" for events selected by the method developed by the authors. The performance of SVM models with different cores was compared. It was shown that the radial-basis function allows the support vector method to handle nonlinear data better and to separate classes with distribution structure in feature space. It is for this model that it was possible to select the parameters that allowed the best separation of the three classes. However, the peculiarity of the SVM method is that it does not involve estimating the probability of an event falling into one of the classes. Therefore, in order to increase the reliability of the results while maximizing the automation of the selection and classification of events, it is intended to further complement the methods used with other approaches, such as Convolution Neural Network (CNN), which allow to use as input data also dynamic spectra. Dynamic spectra contain more information because they reflect the radiation dynamics not only in time, but also in the frequency domain, which allows us to observe the time-frequency evolution of the flare and to identify characteristic features of the microwave radiation structure that are poorly discernible on one-dimensional time profiles. Thus, CNNs will allow us to extend the classification based on microwave flare features.

In addition to the methods discussed above, we can consider the use of recurrent neural networks with Long Short-Term Memory (LSTM). Such networks are specially designed to work with sequential data and are able to efficiently detect hidden dependencies and long-term relationships in the data. An important advantage of LSTM is its ability to memorize and take into account the previous state of the sequence when deciding on the class of the current observation. As

the results of using classical methods for initial event selection have shown, bursts selected by the algorithm may be attributed to artifacts, while long-lasting events with complex structure may be overlooked by the catalog. Applying LSTM to time profiles obtained over an entire observing day will allow noise and artifacts to be eliminated from consideration in the first step, but still find long-lasting events with complex microwave emission structure.

Microwave radiation is an important tool for studying both the processes occurring in the solar atmosphere and for numerical estimation of plasma parameters. The broad spectral range of the SRG from 3 to 24 GHz, as well as its high sensitivity, allows us to record radiation from both powerful flares and events that can be classified as micro- and nano-flares. The specified selection parameters were aimed at detecting solar flares of different power with broadband radiation in the microwave range. It is events of this type that are considered to be indicators of electron acceleration processes and energy propagation during solar flares. A test version of the catalog can be found at https://stmw.iszf.irk.ru/catalogue/.

The identification of a large number of weak flares with a simple structure provides statistically significant material for expanding the database on the parameters and dynamics of the flare plasma, which in turn can be used to improve numerical models of flares. For example, only observations for the second half of 2023 and for 2024 have provided statistically significant material for constructing the mean profile of an impulsive (simple flare) [Motyk et al. 2025]. Information on weak flares with the possibility of spectral analysis also provides extensive material for studies related to the assessment of the contribution of weak flares to the heating of the solar atmosphere. In addition to solving problems related to the analysis of events observed in the 3-24 GHz range, the catalog can be used to identify events observed in the hard X-ray range by instruments without spatial resolution, as well as provide primary information on the microwave response of geoeffective solar flares.

ACKNOWLEDGEMENTS

The results were obtained using the unique scientific facility of the Siberian Radioheliograph. The authors thank the staff of the Siberian Radioheliograph for providing the data and assistance in their processing.

FUNDING

The study was supported by the Russian Science Foundation grant No. 24-22-00315, https://rscf.ru/project/24-22-00315/.

CONFLICT OF INTERESTS

The authors of this paper declare that they have no conflict of interest.

REFERENCES

1. *Altyntsev A.T., Lesovoy S.V., Globa M.V. and others.* The Multi-wave Siberian Radioheliograph // Solar-terrestrial physics. V. 6, No. 2, P. 37-50, 2020.
2. *Lesovoy S.V., Kobets V.S.* Correlation curves of the Siberian Radioheliograph // Solar-terrestrial Physics. V. 3, No. 1, P. 17-21, 2017.
3. *Motyk I.D., Kashapova L.K., Rozhkova D.V.* Average time profiles of microwave radiation from solar flares: morphology and application // Astronomical Journal, in print 2025
4. *Cortes C., Vapnik V.* Support-Vector Networks // Machine Learning, V. 20, P. 273-297, 1995.
5. *Hao N., Flagg L., Jayawardhana R.* Detecting and Classifying Flares in High-resolution Solar Spectra with Supervised Machine Learning // The Astrophysical Journal. V. 973, P.7, 2024.
6. *Lysenko A.L., Ulanov M.V., Kuznetsov A.A. et al.* KW-Sun: The Konus-Wind Solar Flare Database in Hard X-Ray and Soft Gamma-RayRanges // The Astrophysical Journal Supplement Series. V. 262, P. 32, 2022.
7. *Rifkin R., Klautau A.* In Defense of One-Vs-All Classification // Journal of Machine Learning Research, V. 5, P. 101-141, 2004.
8. *Widodo A., Yang B.-S.* Support vector machine in machine condition monitoring and fault diagnosis // Mechanical Systems and Signal Processing, V. 21, № 6, P. 2560–2574, 2007.

Table. Hyperparameters for models with different kernels

Kernel	Hyperparameters
"sigmoid".	C = 1000; gamma = 0.001
"poly"	C = 0.1; gamma = 1.0; degree = 2
"RBF"	C = 10.0; gamma = 1.0.

FIGURE CAPTIONS

Fig. 1. Example of flare detection in the 3-24 GHz range for the May 6, 2024 observations. Circles, black circles, and gray circles denote the frequency ranges 3-6, 6-12, and 12-24 GHz, respectively. The upper panel shows a plot of the microwave bursts detected from the AWG data sampled using the first method. Bottom panel: bursts selected by the second method. The vertical lines show the selected events.

Fig. 2. Example of a solar flare with a complex temporal structure

Fig. 3. Duration distribution statistics of the events sampled by the AWG for the year 2024.

Fig. 4. Examples of time profiles and their preprocessing steps to bring them to a single size

Fig. 5. Distribution of AWG data by class.

Fig. 6. Error matrix for SVC model with sigmoid kernel (top figure), with polynomial kernel (middle figure), with radial-basis function (bottom figure).

Fig. 7. Comparison of models with different kernels in terms of aggregate and categorical accuracy.

Fig. 8. Evaluation of models with different kernels on accuracy (Precision), completeness (Recall) and F1-measure metrics.

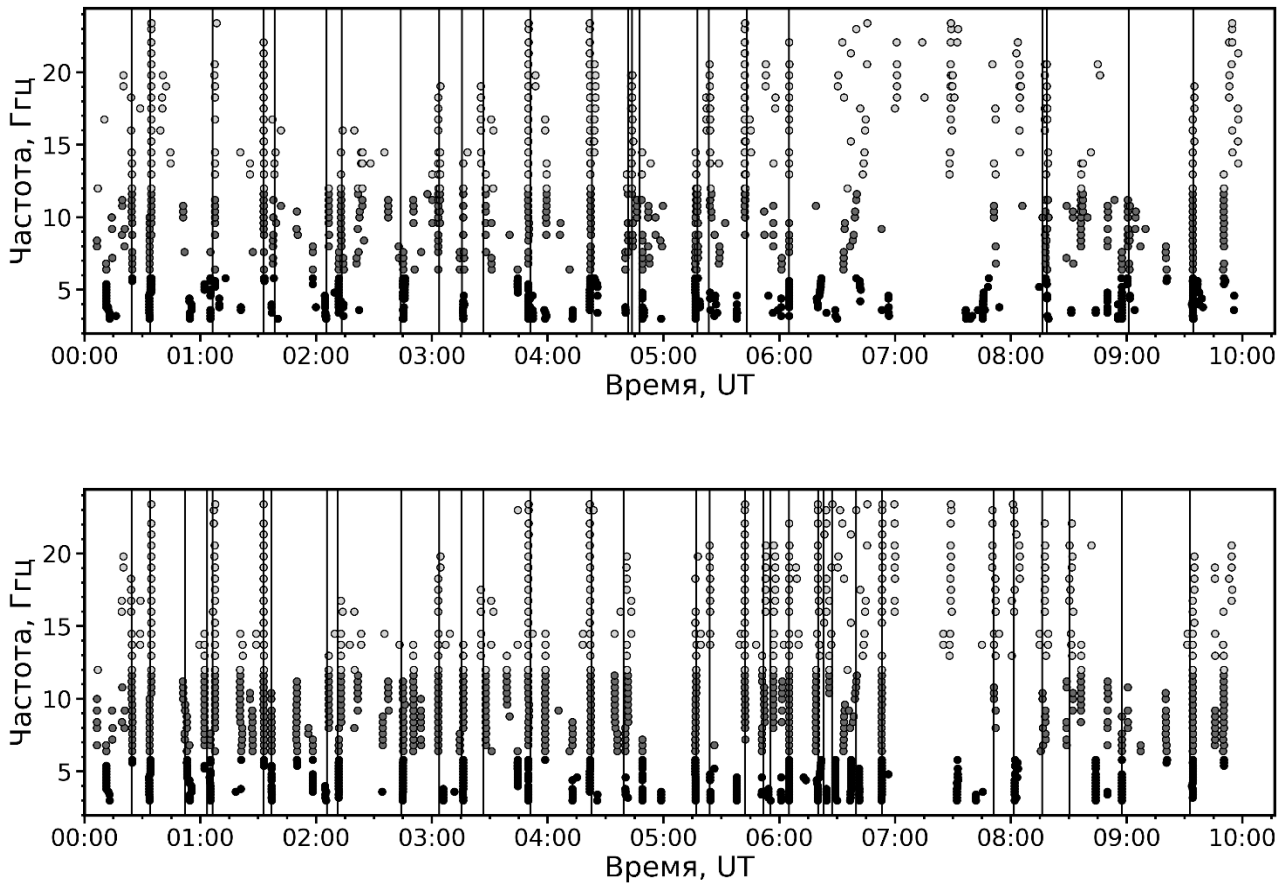

Fig. 1.

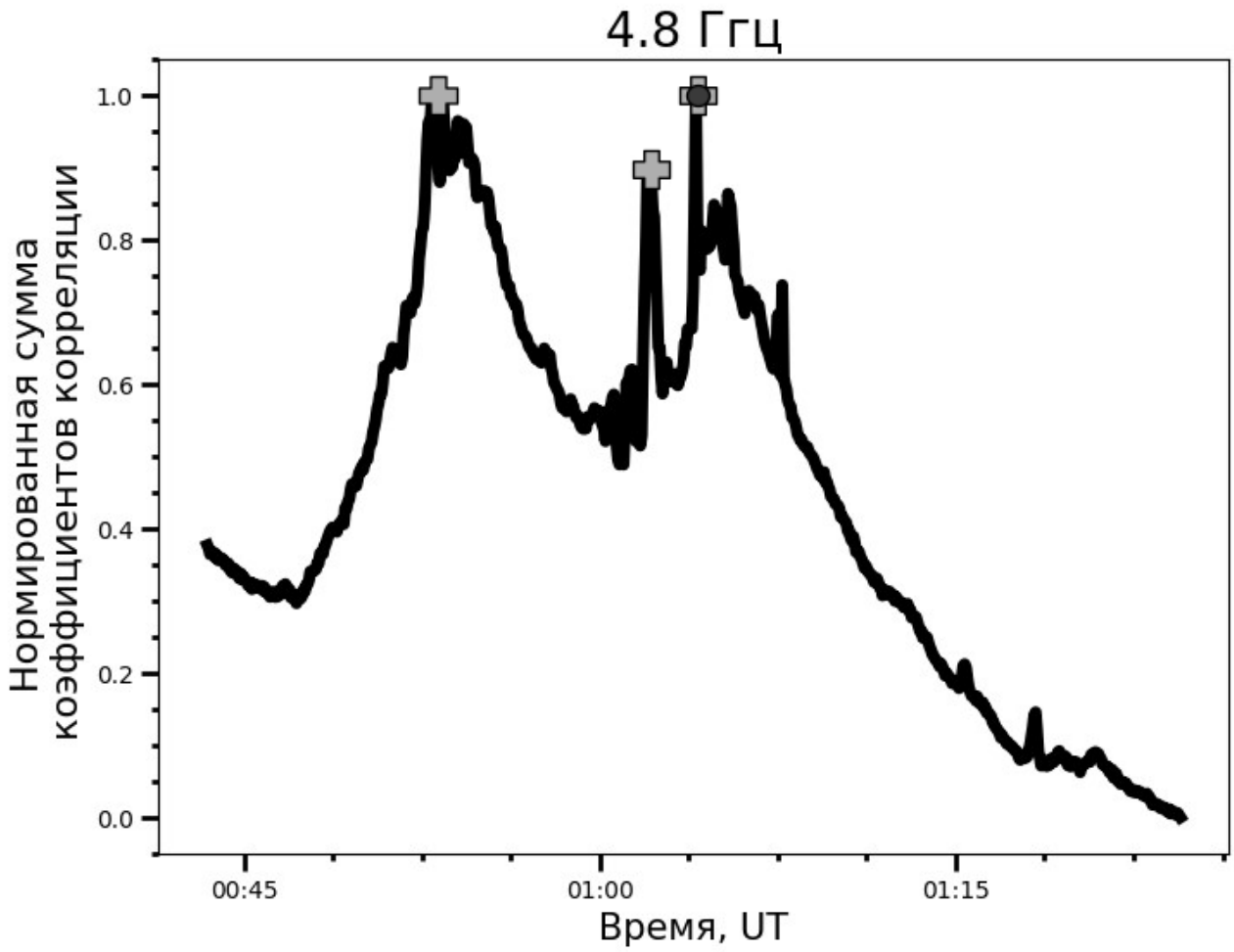
4.8 Ггц
Нормированная сумма коэффициентов корреляции
1.0
0.8
0.6
0.4
0.2
0.0
00:45
01:00
01:15
Время, UT

Fig. 2.

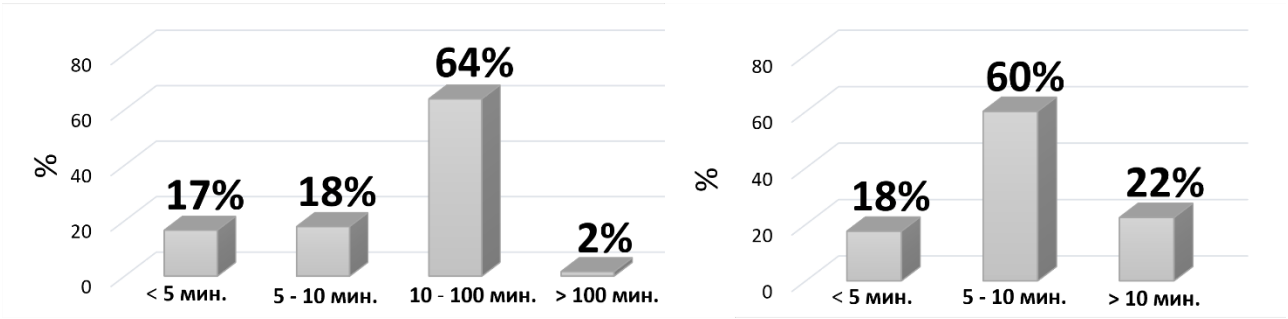
%
80
60
40
20
0
17%
18%
64%
2%
< 5 мин.
5 - 10 мин.
10 - 100 мин.
> 100 мин.
%
80
60
40
20
0
18%
60%
22%
< 5 мин.
5 - 10 мин.
> 10 мин.

Fig. 3.

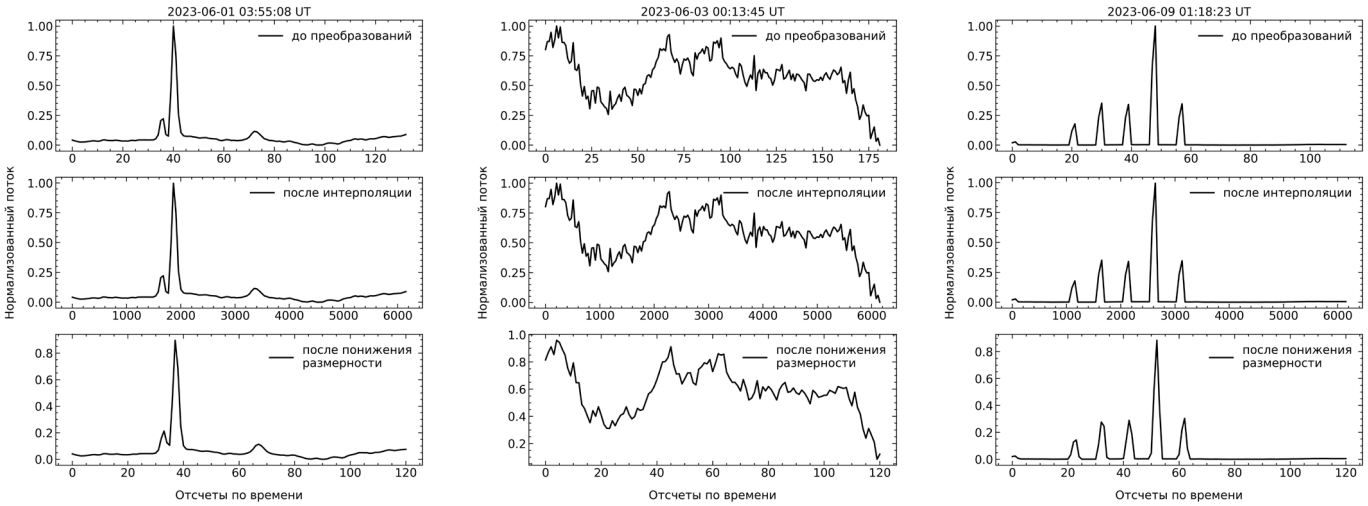
2023-06-01 03:55:08 UT
2023-06-03 00:13:45 UT
2023-06-09 01:18:23 UT
до преобразований
после интерполяции
после понижения размерности
Нормализованный поток
Отсчеты по времени

Fig. 4.

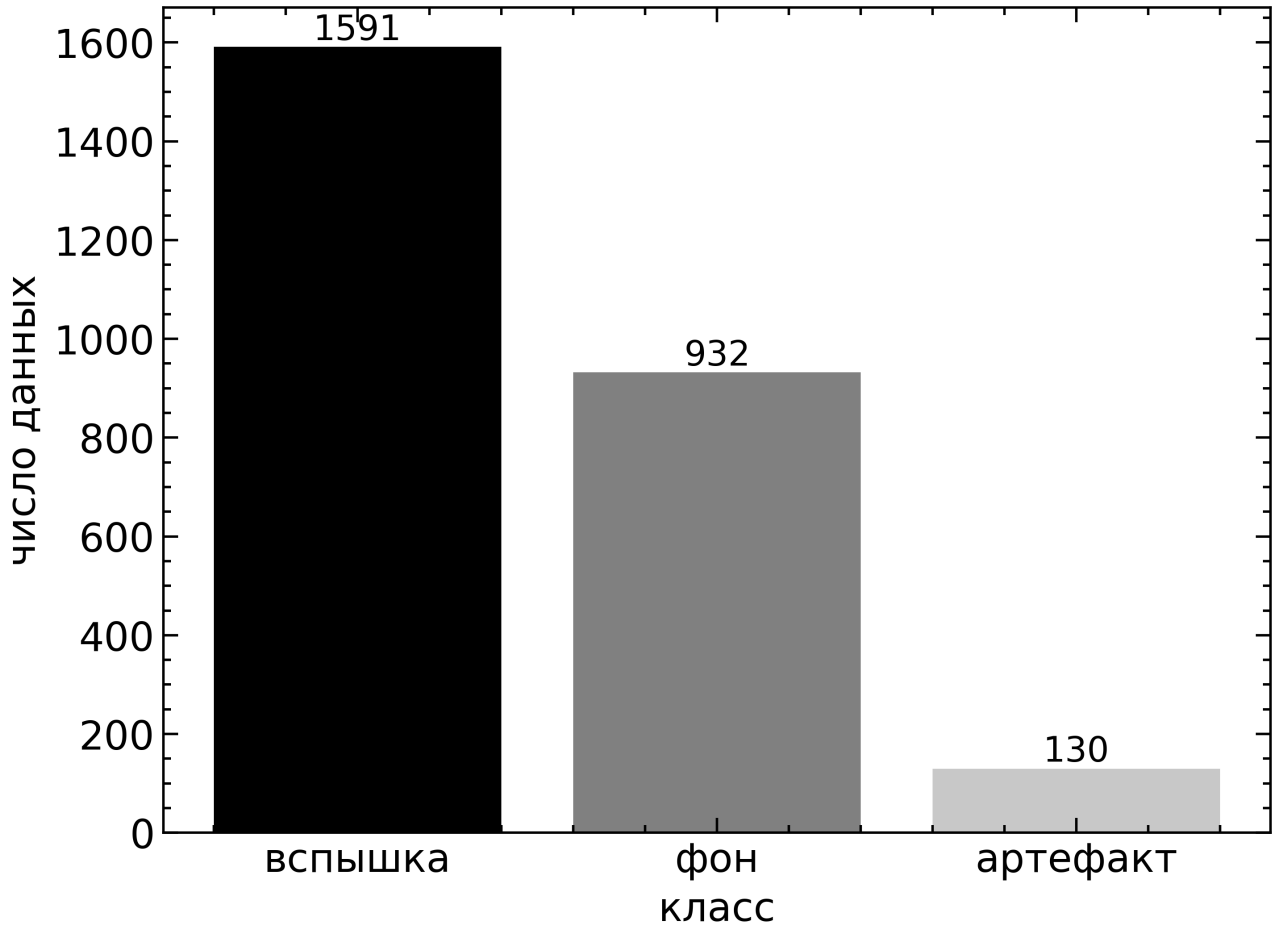
1591
932
130
число данных
вспышка
фон
артефакт
класс

Fig. 5.

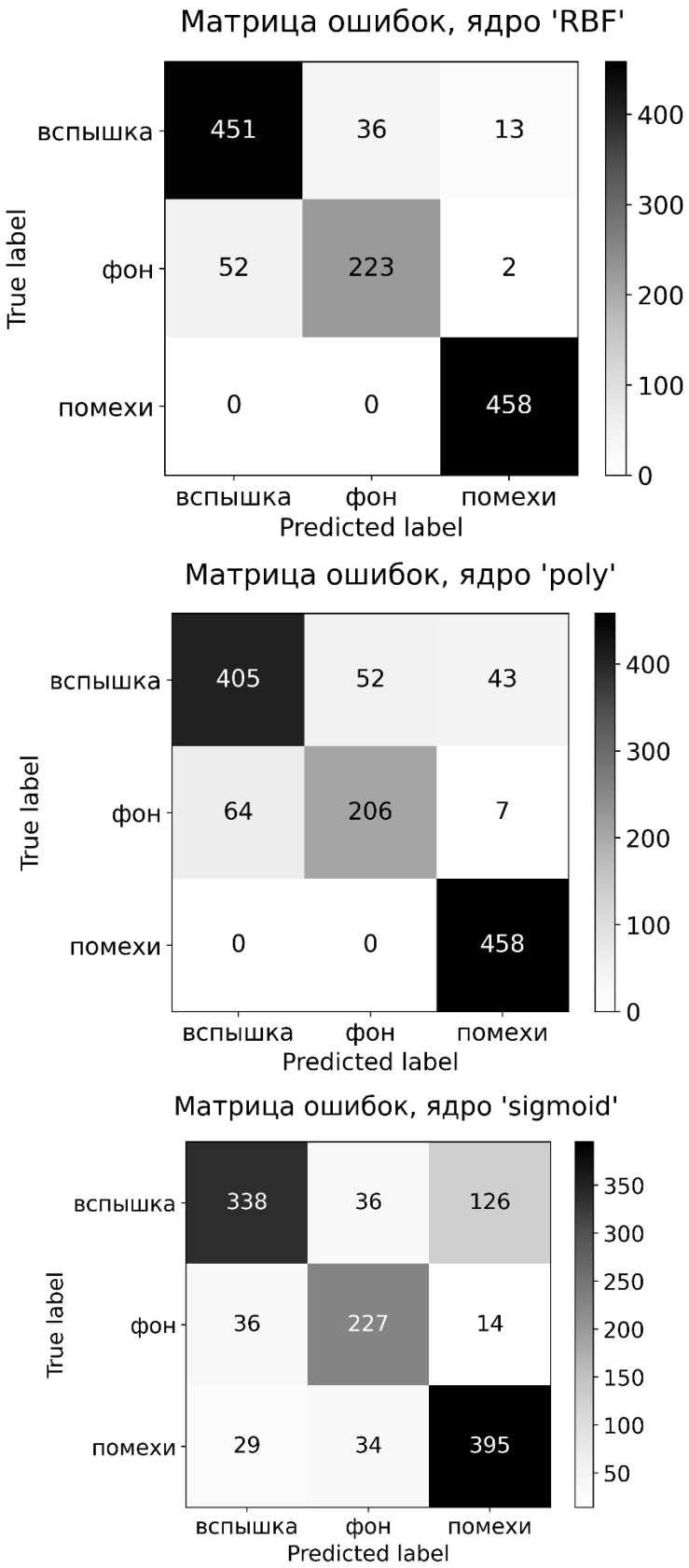
Матрица ошибок, ядро 'RBF'
вспышка
фон
помехи
True label
451
36
13
52
223
2
0
0
458
400
300
200
100
0
вспышка
фон
помехи
Predicted label
Матрица ошибок, ядро 'poly'
вспышка
фон
помехи
True label
405
52
43
64
206
7
0
0
458
400
300
200
100
0
вспышка
фон
помехи
Predicted label
Матрица ошибок, ядро 'sigmoid'
вспышка
фон
помехи
True label
338
36
126
36
227
14
29
34
395
350
300
250
200
150
100
50
вспышка
фон
помехи
Predicted label

Fig. 6.

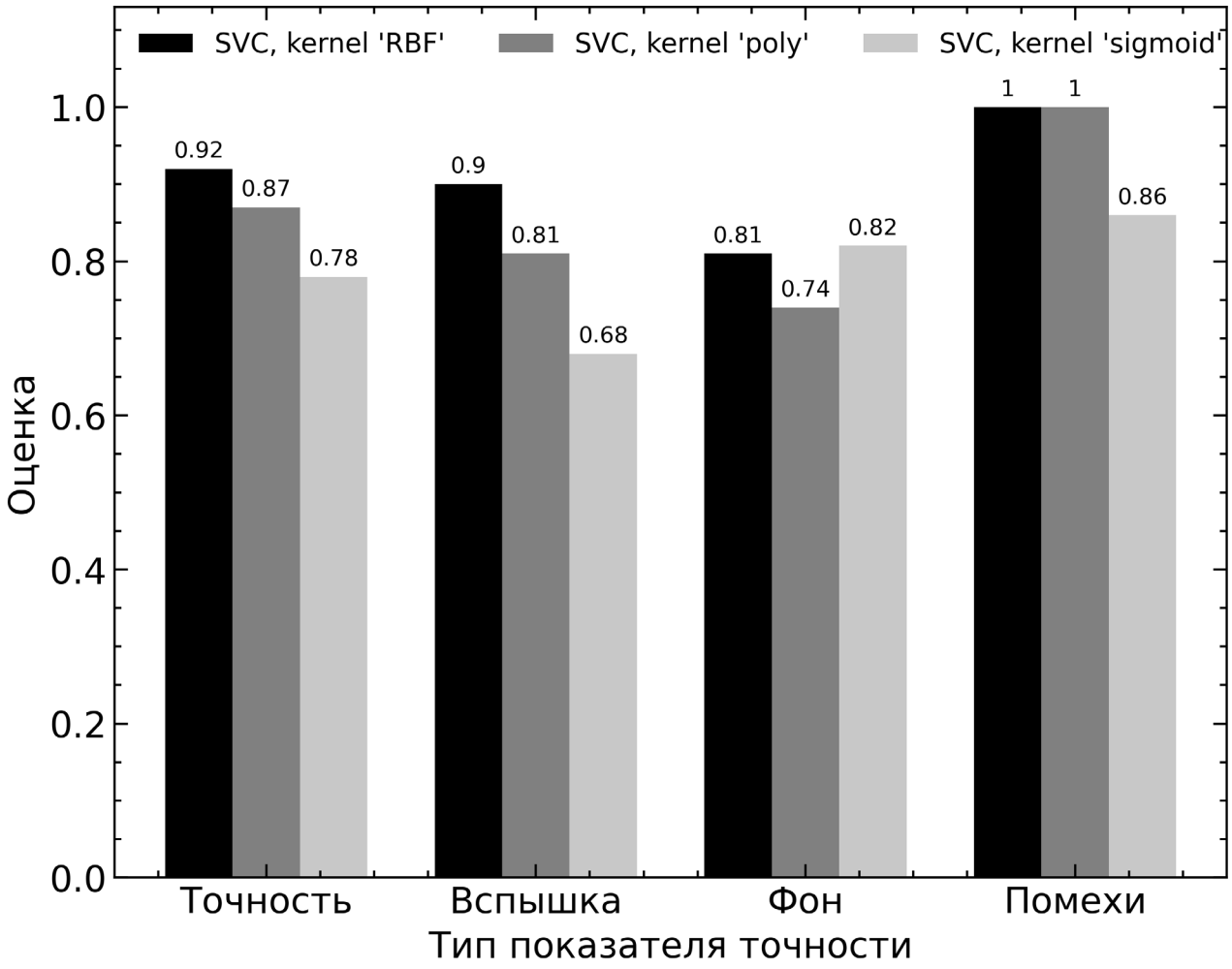
SVC, kernel 'RBF'
SVC, kernel 'poly'
SVC, kernel 'sigmoid'
Оценка
1.0
0.8
0.6
0.4
0.2
0.0
0.92
0.87
0.78
0.9
0.81
0.68
0.81
0.74
0.82
1
1
0.86
Точность
Вспышка
Фон
Помехи
Тип показателя точности

Fig. 7.

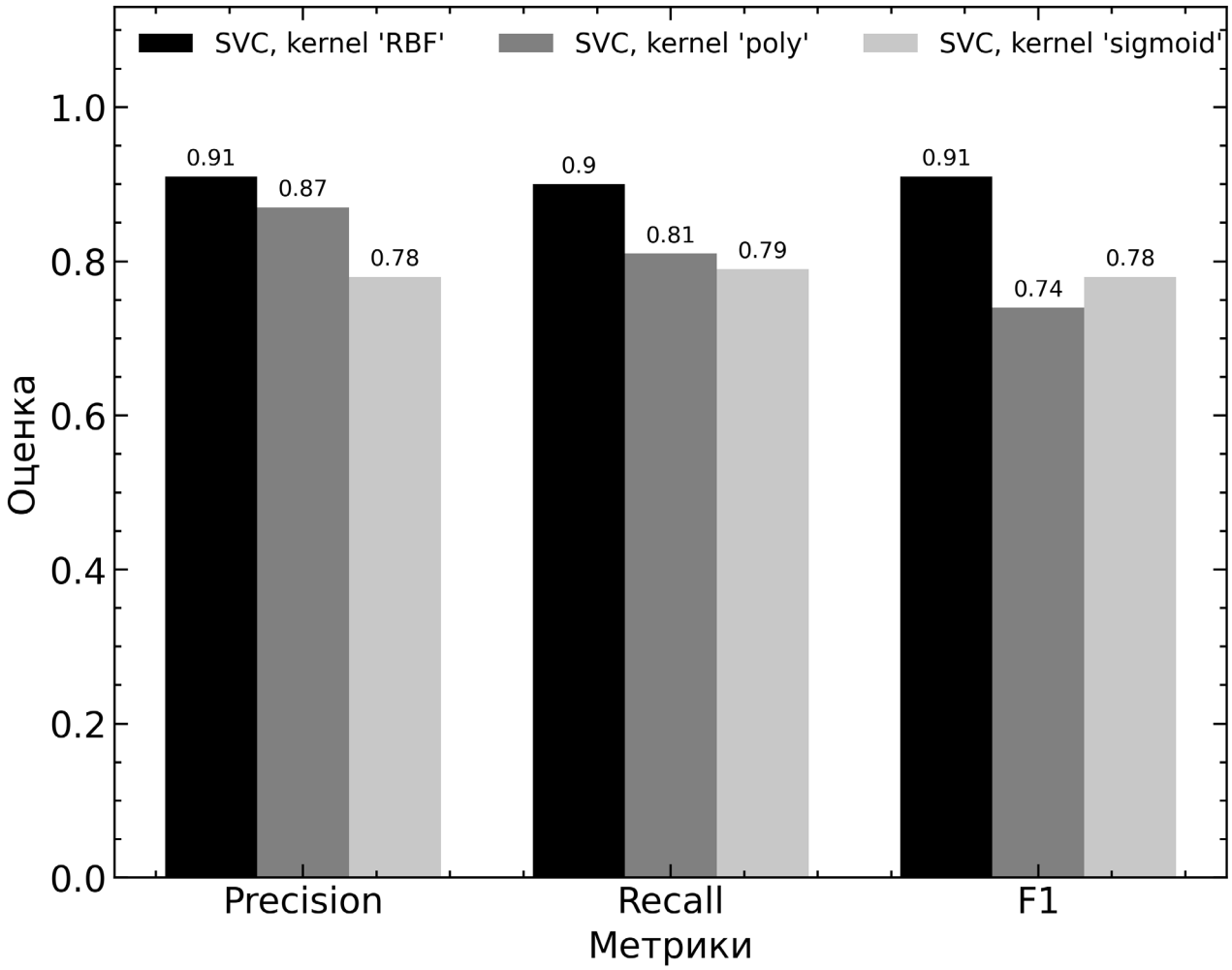
SVC, kernel 'RBF'
SVC, kernel 'poly'
SVC, kernel 'sigmoid'
Оценка
1.0
0.8
0.6
0.4
0.2
0.0
0.91
0.87
0.78
0.9
0.81
0.79
0.91
0.74
0.78
Precision
Recall
F1
Метрики

Fig. 8.