ANALYZING THE EFFICIENCY OF SOME SYSTEMS OF SHORT-TERM FLARE FORECASTING BASED ON OBSERVATIONS OF DIFFERENT LAYERS OF THE SUN'S ATMOSPHERE

© 2025 I. S. Kniazeva*, I. I. Lysov, E. A. Kurochkin, M. S. Korelov, N. G. Makarenko

Main (Pulkovo) Astronomical Observatory of the Russian Academy of Sciences, St. Petersburg, Russia

**e-mail: iknyazeva@gmail.com*

Received February 24, 2025

Revised April 23, 2025

Accepted June 17, 2025

Abstract. The task of operational forecasting of solar flares is an important task of solar physics. It is known to be a difficult task and the accuracy of the prediction is not great, with estimates of accuracy varying quite widely; this is the case in recent works that utilize modern machine learning techniques. Quite high figures are given to assess the quality of forecasts, and there is no validation of such models in the forecast mode, - therefore, it is difficult to draw a conclusion about the real effectiveness. In this paper, a comparative analysis of the real efficiency of forecasts of solar flares above class C and M for the period from 2009 to 2024 published by the Space Weather Forecasting Center on SolarMonitor and predictive criteria based on radio data and published by the Northwest Branch of the SAO RAS is carried out.

DOI: 10.31857/S00167940250718e3

1. INTRODUCTION

Operational forecasting of solar flares is one of the key challenges in solar physics and space weather [Bloomfield et al., 2012]. High-energy particle emissions and coronal mass ejections accompanying flares can have a significant impact on the Earth's atmosphere, magnetosphere, and ionosphere, causing geomagnetic storms, disruption of radio communications, and damage to satellites [Gallagher et al., 2002]. With the active development of space technology and the increase in the number of spacecraft, this problem is becoming increasingly important. In addition, accurate flare forecasts are also important from a scientific point of view, as they allow us to pre-tune observational instruments to study specific regions on the Sun where the event is expected, which improves the quality of the data obtained. Currently, the most common solar flare prediction methods are based on photospheric data, including measurements of the magnetic fields of active regions, their evolution and morphological characteristics [Abed et al., 2021; Popowicz, 2025;

Sinha et al., 2022]. These methods are traditionally used at space weather centers such as NOAA's SWPC[1] , and rely on large magnetogram databases from SOHO/MDI[2] and SDO/HMI[3] instruments. However, solar activity occurs not only in the photosphere but also in the upper atmosphere - the chromosphere and corona - so alternative forecasting methods use data on the behavior of the solar plasma at these levels.

One such method is the prediction of flare activity based on radio observations, in particular, RATAN-600 data [Bogod et al., 2018; Peterova et al., 2021]. This approach is based on the analysis of spectropolarimetric characteristics of active regions in the microwave range, which makes it possible to record preflare changes in the corona and chromosphere. In contrast to photospheric methods, radio prediction provides additional diagnostic capabilities, such as measuring coronal magnetic fields, analyzing anomalies in polarized radio emission, and detecting pre-event changes in the dynamics of active regions.

Observations in the radio band make it possible to detect early flare activity occurring in the chromosphere and corona. These processes are associated with enhanced magnetic fields and changes in the radio emission spectrum. In particular, the prediction model based on RATAN-600 data proposed by the group [Bogod et al., 2018] uses long-term spectropolarimetric observations in the 1-18 GHz range. According to the authors of the paper, this method has several advantages:

- High sensitivity to changes in coronal magnetic fields.
- Possibility of early detection of signs of flare activity.
- Independence from weather conditions, unlike optical observations.

The method is based on the Tanaka-Enomé criterion [Tanaka & Énomé, 1975], according to which the probability of a powerful flare is determined by the ratio of radio emission fluxes at different frequencies.

The development of each active region is in some sense a unique nonlinear process, as well as the realization of each flare event, and is usually of separate interest to researchers. Despite many years of attempts to find "precursors" of flares [Abramenko et al., 2003; Ishkov, 2023; Song et al., 2009], no reliable criteria have been found. In recent years, a large number of works using modern deep learning methods have appeared [Liu et al., 2022; Tang et al., 2021], and even prediction competitions are being organized[4] . Nevertheless, despite the exponential growth in the amount and

[1] https://www.swpc.noaa.gov/ftpdir/warehouse/

[2] http://soi.stanford.edu/sssc/doc/MDI_data_access.html

[3] http://hmi.stanford.edu/

[4] https://i4ds.github.io/SDOBenchmark/

complexity of data, and the development of methods for processing various data, prediction accuracy remains very low. Estimates of accuracy vary widely, ranging from random predictions to values indicating the high predictive power of models. Nevertheless, the practical application of such approaches in operational forecasts is still difficult. Thus, recent works that use modern machine learning techniques provide quite high figures for evaluating the quality of learning predictions [Liu et al., 2022; Tang et al., 2021], but there is no validation of such models in prediction mode, so it is rather difficult to conclude about the real effectiveness. Predictive validation refers to the long-term testing of a trained model on new data. As a rule, in works devoted to prediction [Leka et.all, 2019, Parkes et all, 2020, Nishizuka N. 2021], some fixed data set is used, which is split artificially into training and test samples, and quality metrics are given on the results of this small delayed test sample. That said, the forecast data is currently publicly available[5] , published by NOAA's SWPC forecast center and updated daily on the SolarMonitor resource[6] . The forecast represents the probability of realization of an event of a certain class, namely C, M and X, in the next 24 hours. Also available online is a lesser known system "Forecast"[7] , based on radio data from RATAN-600. The latter is not a forecasting system as such, as the authors publish the values of key radiation intensity indicators at certain frequencies and suggest whether there will be a flare or not in the next 1-3 days in binary form.

The aim of this paper is to perform a comparative analysis of the accuracy, reliability, and false alarm rate of publicly available solar flare forecasts based on data from different atmospheric layers for the period from July 2009 to October 2024. In particular, the predictive value of photospheric data published by SolarMonitor (SWPC NOAA) was quantitatively assessed in comparison to forecasts based on RATAN-600 radio observations.

2. DATA

This paper analyzes three data sources related to solar flare forecasting and registration: the SolarMonitor, RATAN-600, and GOES SWPC NOAA.

SolarMonitor data. SolarMonitor[8] is a web-based resource that aggregates solar activity data, including magnetograms, images of active regions, and flare forecasts based on photospheric characteristics.

The flare forecasts are published daily and contain probabilistic estimates of the occurrence of Class C, M, and X events for each active region (AR). These forecasts are generated based on

---

[5] https://solarmonitor.org/forecast.php

[6] https://solarmonitor.org/

[7] http://spbf.sao.ru/prognoz/tables.html

[8] https://solarmonitor.org/

several models:

- mcstat is a statistical model based on the frequency of flare realization in an area of a certain complexity class. The complexity of the region is labeled using the McIntosh classification system (McIntosh), the method is detailed in [Gallagher et al., 2002].

- mcevol - also a model using the McIntosh classification, but additionally taking into account the evolution of the active region during the last 24 hours.

- noaa - NOAA SWPC space weather center forecast based on expert analysis and classification of areas by type (McIntosh-classification).

Each model publishes the probability of Class C, M, and X outbreaks occurring in the next 24 hours. The data are presented in numerical format (0 to 100%) and updated daily.

RATAN-600 data. Radioastronomical observations of solar activity are carried out at RATAN-600 - the world's largest radio telescope located at the Special Astrophysical Observatory of the Russian Academy of Sciences (Russia). Within the framework of the "Forecast" project[9] , launched by the St. Petersburg branch of the SAO RAS, data of flux intensity at 3 GHz and 10 GHz are published daily in solar flux units (s.e.p.) (1 sfu = 10^4Jy), on the basis of which forecasts can be made. The data are updated daily and published on the SAO RAS server.

GOES SWPC NOAA data. GOES (Geostationary Operational Environmental Satellites) series satellites operated by NOAA provide the most accurate data on the Sun's X-ray emission and record flares in real time. The primary source of information is the GOES X-ray Event List, which contains data on flares that have occurred, including:

- Start, peak, and end times of the flare.
- The class of the flare (C, M, X) based on peak X-ray intensity in the range of 1-8 Å.
- Active region number (if the flare occurred in a registered AO).

Data are available on the SWPC NOAA website[10] and are updated in real time.

3. METHODS

To analyze the effectiveness of the forecasts published by SolarMonitor and the St. Petersburg branch of the SAO RAS, historical data on flares of class larger than C from the GOES X-ray Event List were used. The data were retrieved using a library written in Python ` sunkit_instruments`, part of the solar data library ecosystem `sunpy`[11] . For each actual record from the list, the current date and the previous day were taken - thus a list of dates was generated, with accuracy to the day, based on which the dataset for analysis was further generated. With this approach, the days of "quiet Sun",

[9] http://spbf.sao.ru/prognoz/

[10] https://www.swpc.noaa.gov/

[11] https://sunpy.org/

when there are no events on the whole disk for several days, were excluded from consideration. Then, for the obtained list of dates, the data were processed with SolarMonitor using the RatanSunPy package [Knyazeva et al., 2025], and a table of records of active regions and forecasts for them for the next day was generated.

For each active region, the forecast was compared with the actual events recorded within the next 24 hours after the forecast was published. Historical forecasts collected from July 2009 through December 2024 were used for the analysis. The probabilistic forecasts published on the SolarMonitor platform, namely three models: mcstat, mcevol, and noaa, were used to analyze the performance of solar flare forecasts. These forecasts represent the probability of occurrence of flares of different intensities (C, M, X) during the next 24 hours.

The GOES event list from NOAA's SWPC, which includes information on flares of class C and above, was used as a reference for actual events. For each active area (NOAA number) and each observation date, all events within the next 24 hours were searched. Based on this list, binary target variables were generated to determine whether at least one event in a given class occurred within 24 hours, as follows:

- Presence of at least one outbreak of a class higher than C: $Y_C(\geq C)$
- The presence of at least one outbreak of a class higher than M: $Y_M(\geq M)$

These target variables (binary) were used to evaluate the predictive ability of SolarMonitor forecasts.

When analyzing data from Ratan-600 we have to work with one-dimensional scans (for each frequency of observation), more details about data representation can be found in [Knyazeva et al., 2025], and the same area on the scan can contain several active areas, so one record from the database can contain a list of active areas. Therefore, to form the target variable, we also searched for events in any of these areas.

Metrics for assessing the quality of predictions. Three key metrics were used to quantify the predictive ability of the models:

1. Proportion of events detected (Probability of Detection, POD), which reflects how well the model detects outbreaks (sensitivity):

$$POD = \frac{TP}{TP + FN}$$

   Where:

   - TP (True Positive) - True Positive detection
   - FN (False Negative) - false pass.

2. False Alarm Ratio (False Alarm Ratio, FAR), which shows how often the model is wrong in predicting a flash that does not occur:

$$FAR = \frac{FP}{FP + TP}$$

3. True Skill Statistic (TSS). This metric reflects the difference between the probability of correctly predicting an event and the probability of a false alarm. Unlike other metrics, it is independent of the frequency of outbreaks and is the most objective measure of the quality of predictions:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

4. Brier Skill Score (BS). This metric is used to evaluate probabilistic predictions, is equivalent to the RMS error in a classification task:

$$BS = \frac{1}{N} \sum_{t=1}^{N} (f_t - o_t)^2$$

Where f_t is the predicted probability, o_t is the actual outcome of the event. The value can vary between 0 and 1, where 0 corresponds to a perfect prediction.

These metrics are chosen because of their interpretability in the context of rare event prediction. For example, POD is important for assessing the model's ability to detect outbreaks, but by itself does not indicate how often predictions are false. FAR, in turn, complements POD by indicating the number of false alarms. TSS is the most informative metric because it accounts for both model sensitivity and false alarm rate, which is particularly important when predicting events with rare realizations, such as Class X flares. The Brier Skill Score was only used to analyze data from the Solar Monitor because it is calculated based on probabilistic forecasts. It was previously used to analyze the subjective probabilistic assessment of the operational forecast by SWPC staff from 1996 to 2008 [Crown M.D., 2012].

Since forecasts represent probabilities of outbreak occurrence $p \in [0,100]$, and binary values are required to calculate performance metrics, some threshold for conversion to binary values ("there is an outbreak"/"no outbreak") is required. For this purpose, we applied threshold partitioning, whereby we used an adaptive threshold instead of a fixed conventional threshold (0.5), as a threshold that is too high may be inadequate in the case of rare events. Thresholds ranging from 10 to 50 in increments of 5 were chosen for analysis. Thus, for each such threshold, metrics for evaluating the quality of the predictions were calculated. To evaluate at which threshold value the forecast achieves an optimal balance between POD and FAR, metrics versus threshold curves were plotted (Figures 1 - 4). The POD and FAR plots show how the probability of event detection and the number of false alarms change as the threshold increases. The TSS plot shows the point of maximum value of this metric, which allows us to determine the best binarization threshold.

This analysis was performed separately for event predictions above Class C and above Class

M for all three models (mcstat, mcevol, noaa).

For the radio data, the event probabilities are not published, with no publicly available information on the probability of event prediction, but there is data for 3 and 10 GHz fluxes and a published methodology [Bogod et al., 2018] According to this methodology, the probability of an outburst depends on the flux value at 3 cm and the ratio of the fluxes at 3 and 10 cm (the Tanaka-Enome criterion); in particular, it is proposed to take the value of 10 s.e.p. at 3 cm (associated with magnetic activity in the corona) and the ratio of the fluxes at 3 and 10 cm equal to 0.8. In the methodology of [Bogod et al., 2018], there is a third criterion concerning polarization, but it is not available in the current version of the Forecast system, so it was not possible to test its significance within the framework of this approach. Since in this case there are no strict recommendations for the value of, we used a value grid for both the selection of the flux threshold (in the range 3-10 s.e.p.) and a grid for the flux ratio (in the range 0.4-1.1). Based on these thresholds, flare onset criteria were generated and then compared to real events in the same way as was done for the SolarMonitor data. We moved away from point-based guidelines for criterion construction to interval-based guidelines for experimental purposes, since these criteria represent some kind of heuristics, it was interesting to see how their modification would affect the quality of the prediction. The sensitivity of the complex and the large area of the South+Plosky system of RATAN-600 allows us to observe up to 0.1 s.e.p., so working with the range seems possible. As for the verification of higher levels (more than 10 s.e.p.), the verification of higher fluxes is planned in the future, but, presumably, the forecasting efficiency will be worse, because not all microwave sources in the AO give very large fluxes.

4. RESULTS

Assessment of the accuracy of SolarMonitor predictions. In this analysis, we collected data for the period from September 2009 through September 2024, sampling only those active regions that were observed on the Sun's disk at the time of the onset of a class C flare and the day preceding that time. A total of 2203 active regions were included in the database, with 8942 class C events, 1165 class M events, and 78 class X flares. The data analysis showed that the three SolarMonitor prognostic techniques (mcstat, mcevol, noaa) show similar results, but the noaa method shows the highest accuracy. However, it should be taken into account that data for noaa is available only from 2017, while mcstat and mcevol have been analyzed since 2009, which may affect the objectivity of the comparison. Graphs of the calculation of key metrics for mcstat, mcevol, and noaa are presented in Fig. 1-3. Evaluation of forecasts for RATAN-600 radio observations by the Tanaki-Enome criterion are located in Figure 4.

The TSS metric, which exceeds 0.5 in most solar flare prediction work (see the review paper [Bloomfield et al., 2012], and compared to results obtained with deep neural networks [Abduallah

Y et all, 2023]), reaches this value only when predicting M-class flares for the mcstat and noaa models at very low binarization threshold values of 10-20%. For the mcevol model, the TSS is below 0.5 at all threshold values.

In this case:

- The probability of detection (POD) for outbreaks above class M ranges from 22-75% for the mcstat and mcevol methods, while for noaa the probability is 18-79%. The maximum detection probability corresponds to a binarization threshold of 10% and the minimum probability corresponds to 50%.

- The false alarm rate (FAR) for high-power flashes (M and X classes) exceeds the detection probability, indicating a high false alarm rate.

For flashes above class C, the situation is similar. At a low threshold of 10%, the POD is 82%, 89%, and 93% for mcevol, mcstat, and noaa, respectively. However, the false alarm rate is lower, on the order of 70%, but still remains high. The TSS for C-class flares is lower than for M-class flares, indicating the lower predictive value of the models in predicting less powerful events.

Thus, SolarMonitor predictions show good sensitivity (high POD) but suffer from a high number of false positives, especially for powerful flares.

Evaluation of forecasts based on radio data (RATAN-600). Figure 4 shows the quality results of the forecasts based on a two-parameter grid of flux-based criteria. In order to simplify the visualization, in the upper row a fixed flux ratio parameter of 0.8 and flux magnitude at 3 cm was used, and in the lower row a fixed flux parameter equal to 4 s.e.p. was chosen and it is shown how the prediction changes with the change of flux ratio. Since in this work the task was to approximate the quality of publicly available criteria, a complete search of combinations of parameters was not done, 4 c.e.p. was chosen because, on the one hand, it is close to the recommended level in the original methodology, and on the other hand at this level TSS actually stops changing. At the same time, of course, the construction of a full-fledged model implies both calibration of parameters and analysis of all available data, not only flux ratios at selected frequencies. Prediction of flares on the basis of RATAN-600 radio data is inferior in terms of metrics to methods using SolarMonitor photospheric data, but in certain combinations of criteria (flux value at 3 cm and flux ratio at 3 and 10 cm) it is possible to achieve POD close to 80%, although the same level of false alarms remains. Based on the results obtained, the following conclusions can be drawn: although the radio prediction method is still inferior to photospheric methods, it demonstrates predictive power, even when using a minimal set of criteria. This confirms the prospect of further refinement of the method, especially with the application of machine learning.

5. DISCUSSION

As a result of the analysis carried out in this paper, it appears that the reliability of the solar flare realization prediction available to a wide range of researchers in solar physics is lower than the quality claimed in most scientific prediction papers. For strong flares, the probability of detecting an event of a class higher than class M approaches 80% only if we lower the decision threshold to a probability of 10%, which is much lower than the standardly used threshold of 50%, with a false alarm rate in this case exceeding 80%, the same threshold maximizing the combined TSS forecast quality parameter (0.55). In the case of radio data, the situation is more complicated, since there is no full-fledged prediction system yet - only heuristic methods based on only two approaches and taking into account only a small part of the available data are used. It is quite possible that training of a full-fledged model will allow us to achieve much higher values, and combining data from different layers of the atmosphere will significantly improve it. In addition, a correct comparison requires standardization not so much of the data for the forecast, but rather of the principles of the sample set for the forecast, distinguishing the points in time and the regions that are considered flare and non-flare.

6. CONCLUSIONS

In this paper, a comparative analysis of the accuracy, reliability, and false alarm rate of solar flare forecasts based on SolarMonitor photospheric data (SWPC NOAA) and RATAN-600 radio observations for the period from July 2009 to October 2024 has been performed. In general, the results show that the accuracy of the available forecasts is lower than stated in most scientific papers, especially at conventional binarization thresholds.

The analysis shows that SolarMonitor forecasts are highly sensitive, reaching probability of outburst detection (POD) up to 79% for class M events and above, but are accompanied by a significant number of false positives (more than 80%). The TSS metric slightly exceeds the value of 0.5 for class M events, which is true only for the mcstat and noaa models. This quality is obtained only at a low threshold of conversion to binary value: (10-20%). When the threshold is raised, a decrease in false alarms is observed, but other metrics decrease. The drop in TSS is more pronounced in class M events and the decrease in POD in class C events.

Forecasting based on RATAN-600 radio observations is inferior to SolarMonitor forecasts in terms of selected metrics, but has a number of advantages. The radio data reflect processes in the higher layers of the solar atmosphere, which can provide additional information on flare preparation mechanisms. In certain combinations of parameters, such as the flux at 3 cm and its ratio to the flux at 10 cm, the probability of flare detection can approach 80%. In this case, heuristics are used for prediction, which is a limitation. Nevertheless, prediction based on radio observations shows predictive power, which confirms its potential with further development, especially with the application of machine learning.

ACKNOWLEDGEMENTS

The authors thank the staff of the Northwest Branch of the SAO for supporting and organizing access to the data, as well as all the scientists maintaining the open databases, without whom this work would not have been possible

FUNDING

This work was supported by RNF grant #24-21-00476

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

REFERENCES

1. *Abed A.K., Qahwaji R., Abed, A.* The automated prediction of solar flares from SDO images using deep learning // Advances in Space Research. V. 67. N 8. P. 2544–2557. doi:10.1016/j.asr.2021.01.042. 2021.
2. *Abduallah Y., Wang J.T.L., Wang H., Xu Y.* Operational prediction of solar flares using a transformer-based framework // Sci Rep 13:13665 .2023
3. *Abramenko V.I., Yurchyshyn V.B., Wang H., et al.* Signature of an Avalanche in Solar Flares as Measured by Photospheric Magnetic Fields // The Astrophysical Journal. V. 597. N 2. P. 1135. doi:10.1086/378492. 2003.
4. *Bloomfield D.S., Higgins P.A., McAteer R.T.J., Gallagher, P.T.* Toward reliable benchmarking of solar flare forecasting methods // The Astrophysical Journal Letters. V. 747. N 2. P. L41. doi:10.1088/2041-8205/747/2/L41. 2012.
5. *Bogod V.M., Svidskiy P.M., Kurochkin et al.* A Method of Forecasting Solar Activity Based on Radio Astronomical Observations // Astrophysical Bulletin. V. 73. N 4. P. 478–486. doi:10.1134/S1990341318040119. 2018.
6. *Crown M.D.* Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look ‐ up table and forecaster ‐ issued probabilities // Space Weather, V. 10.N 6, LS06006. doi: 10.1029/2011SW000760.2012
7. *Gallagher P.T., Moon Y.-J., Wang H.* Active-Region Monitoring and Flare Forecasting – I. Data Processing and First Results // Solar Physics. V. 209. N 1. P. 171–183. doi:10.1023/A:1020950221179. 2002.
8. *Ishkov V.N.* Predicting Solar Flare Phenomena: Solar Proton Events // Bulletin of the Russian Academy of Sciences: Physics. V. 87. N 7. P. 942–944. doi:10.3103/S1062873823702763. 2023.
9. *Knyazeva I., Lysov I., Kurochkin E., et al.* A robust preprocessing pipeline for RATAN-600 solar radio observations data // Astronomy and Computing. V. 51. 100918.

doi:10.1016/j.ascom.2024.100918. 2025.
10. *Leka K.D., Park S.H., Kusano K., et al.* A Comparison of Flare Forecasting Methods. III. Systematic Behaviors of Operational Solar Flare Forecasting Systems // The Astrophysical Journal. V. 881 N.2.P.101-114 doi:/10.3847/1538-4357/ab2e11.2019
11. *Liu S., Xu L., Zhao Z., et al.* Deep Learning Based Solar Flare Forecasting Model. II. Influence of Image Resolution // The Astrophysical Journal. V. 941. N 1. P. 20. doi:10.3847/1538-4357/ac99dc. 2022.
12. *Nishizuka N., Kubo Y., Sugiura K., et al.* Operational solar flare prediction model using Deep Flare Net // Earth, Planets and Space. V. 73. N.1. P. 64. doi.org/10.1186/s40623-021-01381-9. 2021.
13. *Park S., Leka K., Kusano K, et al.* A Comparison of Flare Forecasting Methods. IV. Evaluating Consecutive-Day Forecasting Patterns // The Astrophysical Journal. 2020 V. 890. N. 124. P. 124. doi:org/10.3847/1538-4357/ab65f0 .2020.
14. *Peterova N.G., Topchilo N.A., Kurochkin E.A.* Increased Microwave Radiation Brightness as a Sign of Flare-Producing Active Regions Based on Observations of NOAA Active Region 12371 // Geomagnetism and Aeronomy. V. 61. N 1. P. S24–S35. doi:10.1134/S0016793222010145. 2021.
15. *Popowicz A.* Toward Reliable Solar Flare Forecasting: Exploring Multi-Wavelength Data from the Solar Dynamics Observatory // Preprints. 2025010895. doi:10.20944/preprints202501.0895.v1. 2025.
16. *Sinha S., Gupta O., Singh V., et al.* A Comparative Analysis of Machine-learning Models for Solar Flare Forecasting: Identifying High-performing Active Region Flare Indicators // The Astrophysical Journal. V. 935. N 1. P. 45. doi:10.3847/1538-4357/ac7955. 2022.
17. *Song H., Tan C., Jing J., et al.* Statistical Assessment of Photospheric Magnetic Features in Imminent Solar Flare Predictions // Solar Physics. V. 254. N 1. P. 101–125. doi:10.1007/s11207-008-9288-3. 2009.
18. *Tanaka H., Énomé, S.* The microwave structure of coronal condensations and its relation to proton flares // Solar Physics. V. 40. N 1. P. 123–131. doi:10.1007/BF00183156. 1975.
19. *Tang R., Liao W., Chen Z., et al.* Solar Flare Prediction Based on the Fusion of Multiple Deep-learning Models // The Astrophysical Journal Supplement Series. V. 257. N 2. P. 50. doi:10.3847/1538-4365/ac249e. 2021.

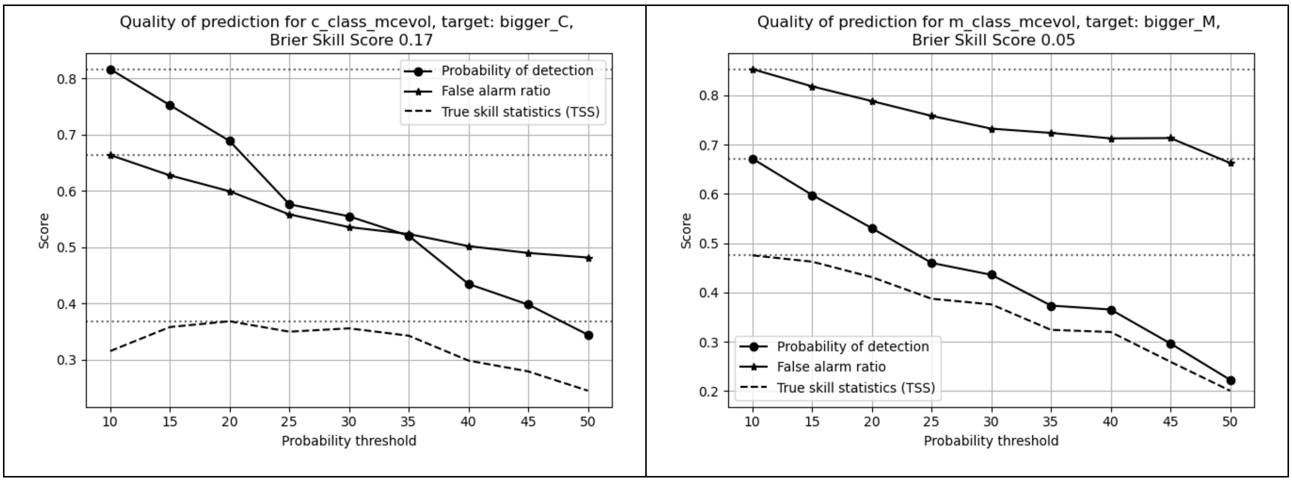
FIGURE CAPTIONS

Fig. 1. Prediction quality metrics of the mcevol model. Left for events of class M or higher, right for events of class C or higher.

Fig. 2. Prediction quality metrics of the mcstat model. On the left for class M or higher events, on the right for class C and higher events.

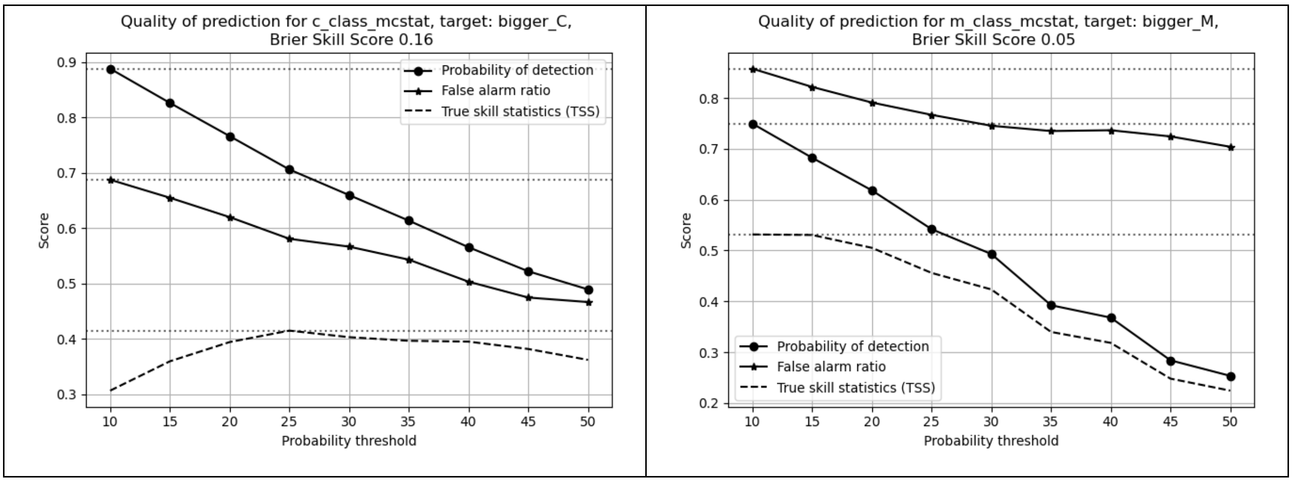
Fig. 3. Prediction quality metrics of the noaa model. Left for class M or higher events, right for class C and higher events.

Fig. 4. Results of the Tanaki-Enome criterion evaluation of radio observation-based prediction. The graphs on the left are for class M or higher events, on the right for class C or higher events. For the upper plots, a flux ratio (f_rel) value of 0.8 is recorded; for the lower plots, a flux at 3 cm (flux) value of 4 conventional units is recorded.



Quality of prediction for c_class_mcevol, target: bigger_C,
Brier Skill Score 0.17
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold
Quality of prediction for m_class_mcevol, target: bigger_M,
Brier Skill Score 0.05
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold

Fig. 1.



Quality of prediction for c_class_mcstat, target: bigger_C,
Brier Skill Score 0.16
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold
Quality of prediction for m_class_mcstat, target: bigger_M,
Brier Skill Score 0.05
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold

Fig. 2.

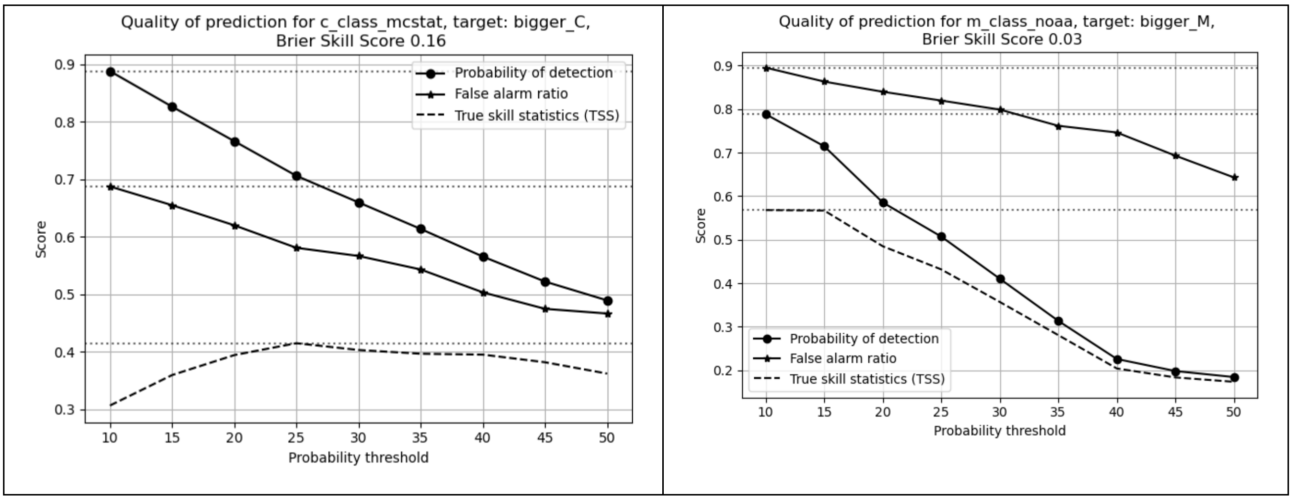

Quality of prediction for c_class_mcstat, target: bigger_C,
Brier Skill Score 0.16
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold
Quality of prediction for m_class_noaa, target: bigger_M,
Brier Skill Score 0.03
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Probability threshold

Fig. 3.

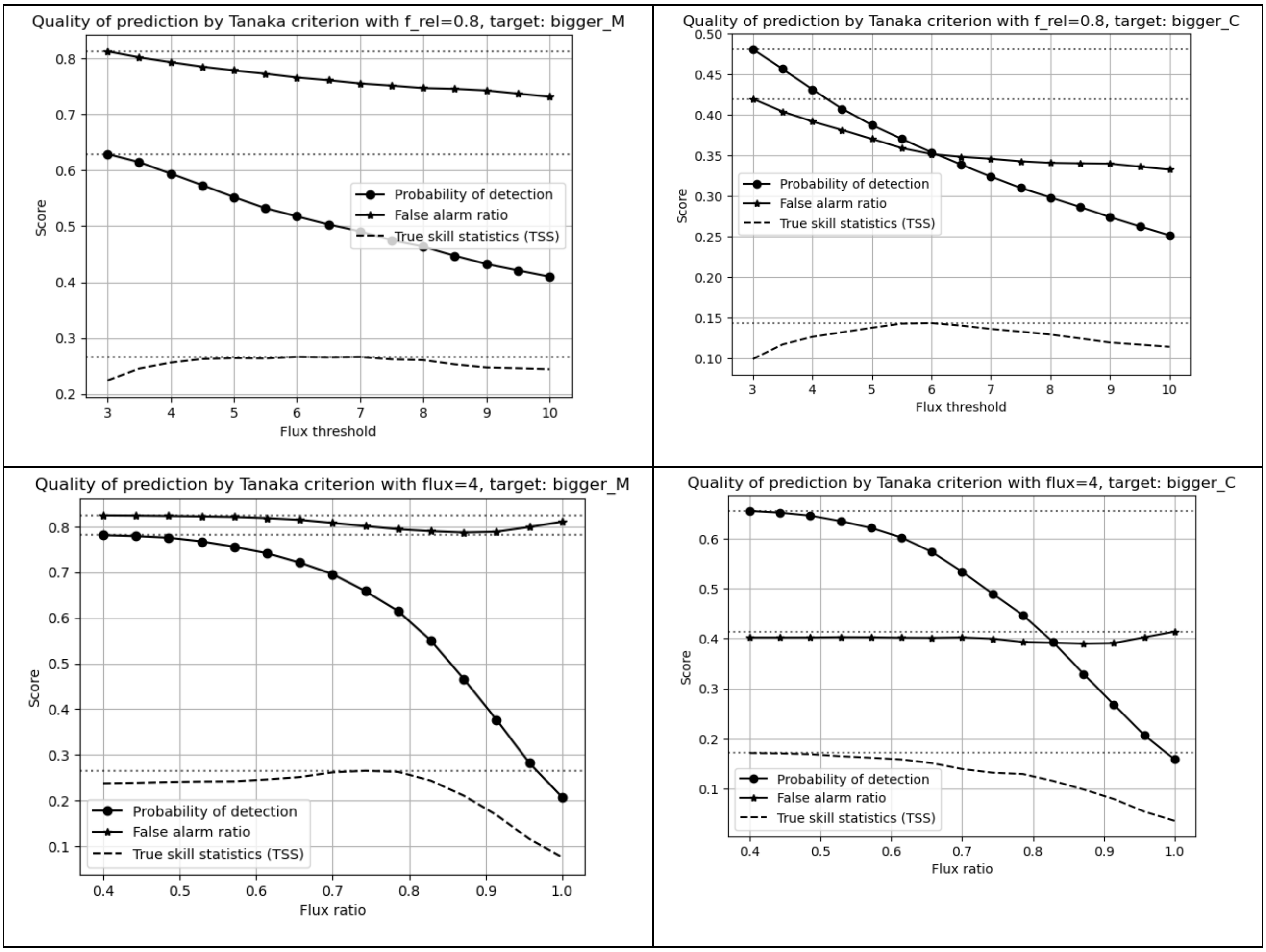

Quality of prediction by Tanaka criterion with f_rel=0.8, target: bigger_M
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Flux threshold
Quality of prediction by Tanaka criterion with f_rel=0.8, target: bigger_C
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Flux threshold
Quality of prediction by Tanaka criterion with flux=4, target: bigger_M
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Flux ratio
Quality of prediction by Tanaka criterion with flux=4, target: bigger_C
Probability of detection
False alarm ratio
True skill statistics (TSS)
Score
Flux ratio

Fig. 4.