AFRL-AFOSR-UK-TR-2019-0033

Improving Solar Eruption Forecasting using Active Region Evolution

Sophie Murray
UNIVERSITY OF DUBLIN, TRINITY COLLEGE
COLLEGE GREEN
DUBLIN, 2
IE

06/18/2019
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
18-06-2019

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
15 Sep 2017 to 14 Mar 2019

4. TITLE AND SUBTITLE
Improving Solar Eruption Forecasting using Active Region Evolution

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-17-1-0392

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Sophie Murray

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF DUBLIN, TRINITY COLLEGE
COLLEGE GREEN
DUBLIN, 2 IE

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
EOARD
Unit 4515
APO AE 09421-4515

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-UK-TR-2019-0033

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An active region tracking system has been successfully updated (Work Package 1) to provide an improved warning system for potential flares, and an ensemble system has been developed to provide an improved starting probability for flare forecasts (Work Package 2). Both outputs have been successful tested for operational use by the Met Office, and are openly available to the wider space weather community for science collaborative efforts.

15. SUBJECT TERMS
Solar Physics, Coronal Mass Ejection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		
Unclassified	Unclassified	Unclassified	SAR	

19a. NAME OF RESPONSIBLE PERSON
MILLER, KENT

19b. TELEPHONE NUMBER *(Include area code)*
011-44-1895-616022

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Improving Solar Eruption Forecasting using Active Region Evolution

Award No. FA9550-17-1-0392

Period of Performance: 2017-09-15 - 2019-03-14

PI: Dr Sophie A. Murray

Trinity College Dublin

Contents

Abstract

One essential component of operational space weather forecasting is the prediction of solar flares. With a multitude of flare forecasting methods now available online is still unclear which of these methods performs best, and none are substantially better than climatological forecasts. Space weather researchers are increasingly looking towards methods used by the terrestrial weather community to improve current forecasting techniques. This project has utilized software programming principles as well as state-of-the-art techniques used by the operational weather community in order to improve tools used by space weather forecasters to predict flares. An active region tracking system has been successfully updated (Work Package 1) to provide an improved warning system for potential flares, and an ensemble system has been developed to provide an improved starting probability for flare forecasts (Work Package 2). Both outputs have been successful tested for operational use by the Met Office, and are openly available to the wider space weather community for science collaborative efforts.

WP1: Automatic Tracking of Active Regions

The tracking of active regions (ARs), which can give rise to solar eruptive events such as solar flares and CMEs, is vital for the mitigation of extreme space weather events. This work package was focused on updating the successful Solar Monitor Active Region Tracker [SMART; Higgins et al., 2011]. The original system was developed in the IDL programming language (with problematic licensing for operations) and did not take full advantage of high-resolution Solar Dynamics Observatory (SDO) solar surface observations. Thus SMART was updated and converted into the open source Python 3 programming language, for easier use by Met Office operational forecasters.

The system was fully upgraded in two phases; the segmentation algorithm, then the tracking algorithm. The SMART algorithm inputs a solar surface magnetogram which then is segmented into ARs of interest, from which numerous magnetic properties of these regions are calculated (see Figure 1. The segmentation algorithm was converted line-by-line to ensure it functioned the same as its predecessor. The tracking routine however was quite outdated, so an improved algorithm was written in Python. A list of AR objects are created from SMART processed output files for whatever start and end times are chosen. Every possible AR pair from the two lists are compared after accounting for solar rotation. If any of the rotated ARs overlap they are re-designated as the same AR.

This project provided an excellent opportunity to ensure the updated SMART code was sufficiently robust for operational purposes (it previously being a science code simply modified for operational use). After rigorous validation procedures were completed (unit testing, bug fixing, documentation, etc), the code was then verified scientifically. First a well-known case study was chosen for analysis, namely NOAA AR 12673 from September 2017, which underwent rapid evolution across the solar disk and produced numerous M- and X- class flares from September 4th to 9th. The case study reproduced expected magnetic property evolution (see Figure 2, with rapid rises before the major flare events as well as typical declines towards the western limb.

The second scientific test was a long-term study of AR evolution using the full dataset of SDO from 2010 - 2018. Figure 3 shows daily detections by SMART over this nine year period compared to the SILSO (`http://www.sidc.be/silso/datafiles`) sunspot number. Note that both data are smoothed to show the trend using a convolution function to obtain an average of the spread of daily numbers. Both datasets clearly show the solar activity cycle, with discrepancies expected due to the use of visible light vs magnetogram data, as well as SMART often grouping multiple ARs together if they are close (particularly at solar maximum).

The new Python implementation of SMART is freely available on GitHub (`https://github.com/drsophiemurray/smart_python`), and has been successfully installed at the Met Office to replace its predecessor for operational use. The code can be used independently by the community with only access to the open source SDO data needed using any basic laptop with Python, however it has also been designed for integration with SolarMonitor.org. The website is currently undergoing an update (from IDL to Python), and once completed daily SMART output will go online in realtime. The new version of the code was also used for some initial scientific investigations into solar eruption forecasting [Murray et al., 2018], which will be continued in Award FA9550-19-1-7010 (Seeking the Source of Solar Eruptions, 2019 - 2022).

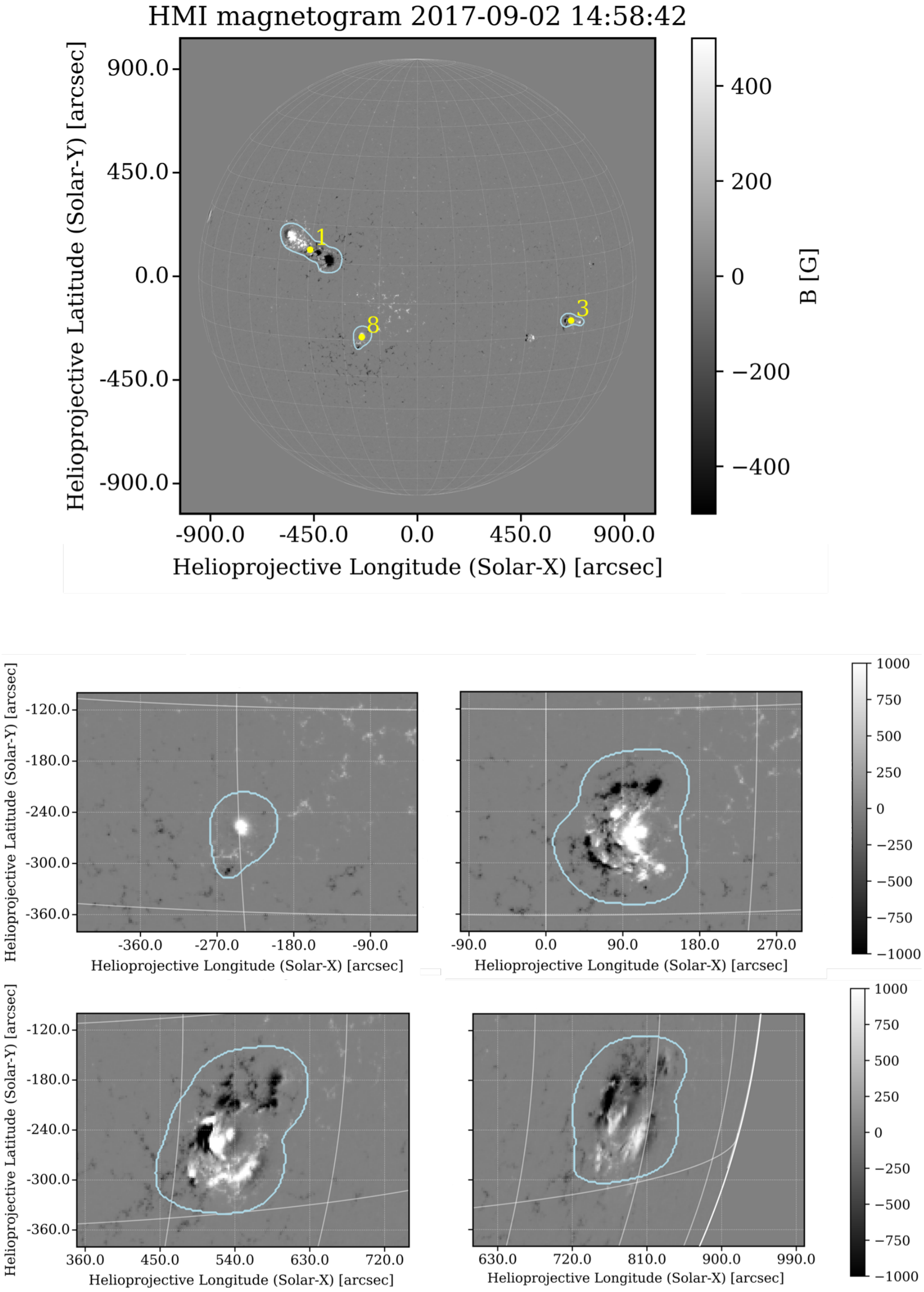

Figure 1: Top: HMI magnetogram on 2017 September 2 with SMART detection IDs labelled in yellow and detection boundaries in blue. Bottom: SMART region 8 from top panel (NOAA AR 12673) tracked by SMART throughout its rapid evolution across the solar disk.

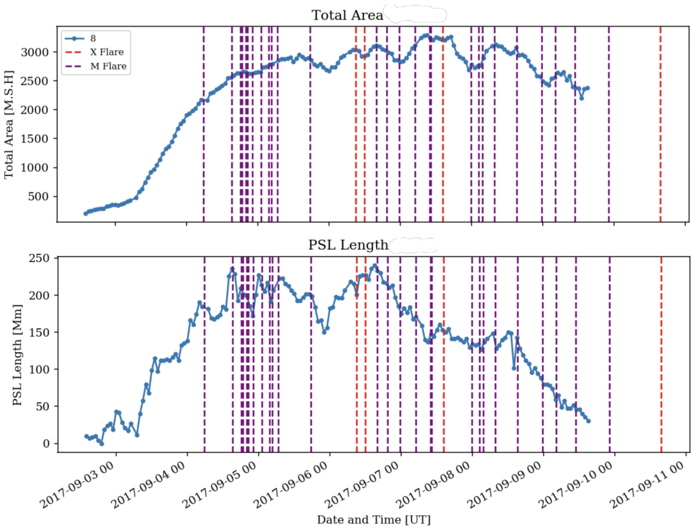

Figure 2: SMART detected total area (top) and polarity separation line length (bottom) of AR12673 from Figure 1. M- and X- class flare events are displayed with purple and red vertical lines respectively; 27 M- and 3 X-flares occurred during the SMART-detected evolution of the region.

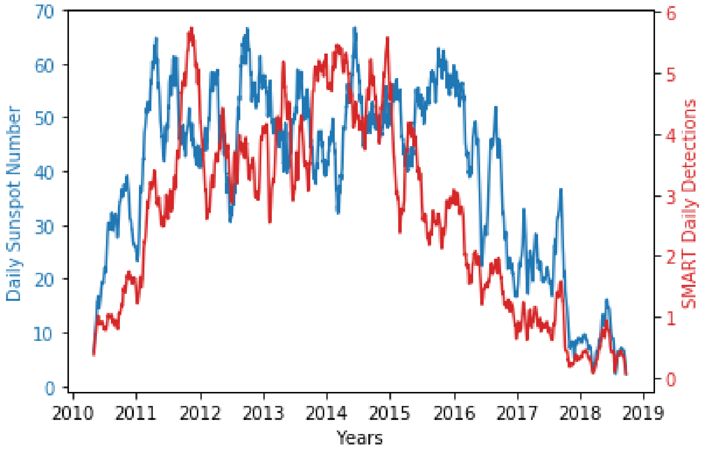

Figure 3: Daily SILSO sunspot number (blue) compared to SMART daily detections (red) from May 2010 to September 2018.

WP2: Ensemble Flare Prediction

Ensemble forecasting has been used in numerical weather prediction for many years as a way to combine different predictions in order to obtain a more accurate result, but it is fairly new in space weather [Murray, 2018]. This work packaged involved developing a flare forecasting ensemble system for operational use for the first time.

The work of Guerra et al. [2015] was expanded upon to create the system, who demonstrated the applicability of multi-model input ensemble forecasting for flare occurrence for a small data sample (four forecasting methods and 13 active regions). The authors proved that linearly combining probabilistic forecasts using combination weights based on the performance history of each method makes more accurate forecasts, and that combining forecasts intrinsically different (i.e., automated software versus human-influenced expert systems) improves the prediction in comparison to the case in which only forecasts of similar type are used. The code originally developed for this work was built upon to create an operational system. The forecasts that are made available by the NASA/CCMC Flare Scoreboard (`https://ccmc.gsfc.nasa.gov/challenges/flare.php`) were used as inputs to a six-member ensemble[1] (in particularly using ASAP, ASSA, MAG4, MOSWOC, NOAA, and SolarMonitor (MCSTAT)). Once the software system was sufficiently validated, similar to Work Package 1, it was then verified in the scientific sense.

To test usefulness of the ensemble system, a similar method to Guerra et al. work was used, however including further forecasting methods and a larger data sample. This WP also focused on analyzing full-disk forecasts rather than AR forecasts, which are used more widely by operational centers. Ensembles for M- and X- class flares were created by linearly combining the full-disk probabilistic forecasts from the Flare Scoreboard. Forecasts from each method are weighted by a factor that accounts for the method's ability to predict previous events, and several performance metrics (both probabilistic and categorical) are considered. The results can thus provide operation centres with an improved starting point to their forecasts that can be tailored to different end-user needs. For example, some users may want to mitigate false alarms while others may want to mitigate misses, therefore the most appropriate weighted metric ensemble can be chosen for each user.

Figure 4 presents the 24-hour probabilistic forecast data for M-class flares for the time period tested from 2014-2016, with histograms of values on the left and three-year time series on the right. Data is color-coded according

---

[1] Note that it was originally intended to use the outputs of the FLARECAST project (`https://api.flarecast.eu`), which would have provided an up to 20-member ensemble of machine learning outputs. However the realtime system is still not live as of April 2019. The final WP2 ensemble system has been flexibly designed such that any new method can be easily integrated if inputted using a standardized format.

to the forecasting method. All forecasts for M-class flares show similar characteristics: probability values range from almost 0.0 to $0.9 - 1.0$, with a decreasing frequency from low to high probability bins. Although, in case of MAG4, higher frequency is concentrated in the lowest-probability bin while some bins are empty. On the other hand, forecasts for X-class flares (not displayed[2]) show a variety of upper limit for probability ranges, between 0.25 (ASSA) and 0.90 (ASAP). There were in total 189 and 17 days between 1 January 2014 and 31 December 2016 that observed M- and X- class flares, respectively. These values yield climatological frequencies of 0.172 and 0.016 that were used to measure the skill of the ensemble forecasts compared to the originals.

Three different schemes for linearly combining forecasts were tested: internal linear combination, constrained, and unconstrained linear combination [see Guerra et al., for more details]. In the last two cases a group of 13 metrics (7 probabilistic and 5 categorical) were used as functions to be optimized and thus find the most optimal combination weights (see Table 1). The performance of the ensemble outputs were compared using several probabilistic performance metrics, following the operational flare forecasting verification measures suggested by Murray et al. [2017]. For example, Relativistic Operation Characteristic (ROC) curves and reliability diagrams are displayed in Figure 5 for a selection of M-class forecasts. Reliability diagrams indicate how close the forecast probabilities of an event correspond to the actual chance of observing the event, and are good companions to ROC curves since they are conditioned on forecasts (while ROC is conditioned on observations). ROC curves present forecast discrimination, and the area under curve the provides a useful measure of the discriminatory ability of a forecast.

The figure compares a selection of forecasts including the best original ensemble members and outputs, namely MOSWOC (human-edited, black line), MAG4 (automated, turquoise line), an average ensemble forecast of all original methods (blue line), and the best performing ensemble, which was created using an unconstrained NLCC weighting scheme (purple line). The ROC curve in the top row (with the better performing methods tending towards the upper left corner of the plot) highlights the clear improvement the ensembles have over the automated MAG4 method, with all other curves similarly good for M-class forecasts. In the reliability diagrams in the bottom row, methods should preferably be in the grey zone of 'skill' along the diagonal, and if they are tending toward the horizontal line they are becoming comparable to climatology. Here most methods are generally over-forecasting (lines lying to the right of the center diagonal line), except the best-performing NLCC ensemble method.

---

[2] The results of M-class flares are highlighted in this final report rather than X-class due to the rarity of events in the dataset.

Table 1: Performance metrics used during the optimization process. Each metric produces a different set of combination weights, or a different ensemble. In each case a label is shown in parenthesis. Categorical metrics are calculated using the 2×2 contingency table once probabilistic forecasts are turned into deterministic forecasts by choosing a threshold value.

Probabilistic	Categorical
Brier score (BRIER)	Brier score (BRIER_C)
Mean Absolute Error (MAE)	True Skill Score (TSS)
Linear Correlation coeff. (LCC)	Heidken Skill Score (HSS)
Nonlinear Correlation coeff. (NLCC)	Accuracy (ACC)
Reliability (REL)	Critical success index (CSI)
Resolution (RES)	Gilbert skill score (GSS)
ROC curve score (ROC)	

The ROC area for all forecasts and Brier score is presented in Table 2 for M-class forecasts. Brier score measures the mean square probability error, and can be broken into reliability, resolution, and uncertainty, which are also listed in the table. The scores in the table are listed in order of original input forecasts, then ensembles from probabilistic and categorical weighting schemes, with tesults for both constrained and unconstrained schemes are included. Most values are to be expected, with overall good Brier scores but poor resolution, and only a few resulting forecasts with a 'poor' ROC score (in the range 0.5 - 0.7). There is definitive improvement compared to MAG4, the best of the original automated forecasts.

It is particularly interesting to note how well the simple ensemble average performs in this work compared to the more complex weighting schemes. While ensemble averages will likely rarely outperform the human-edited forecasts, they are successful here in outperforming the best of automated methods. These could be a good starting point for forecasters when issuing forecasts operationally before additional information or more complex model results are obtained. However, the weighting schemes do provide a level of flexibility that simple averages cannot; they allow operational centers to tailor their forecasts depending on what measure of performance a user cares about the most. In this work only a selection of metrics are highlighted based on current standards used by the community, however the data and code is provided open access so the community can perform their own analysis. The code for the ensemble system is freely available on GitHub (`https://github.com/drsophiemurray/ensemble_system`).

The results of Work Package 2 demonstrate that multi-model input ensemble predictions of solar flares are versatile enough to be implemented and used in operational environments with metrics that satisfy user-specific needs. These results (figures and tables) are taken from a paper currently under preparation for the journal Space Weather [Guerra et al.], and the ensemble system itself has been tested for operational use by the Met Office. The ensemble average is currently available on the Flare Scoreboard and it is planned to integrate the weighting schemes and performance metric comparisons online in the future for community use. It is worth noting that an invited commentary on the importance of ensembles for space weather was also published during the project [Murray, 2018].

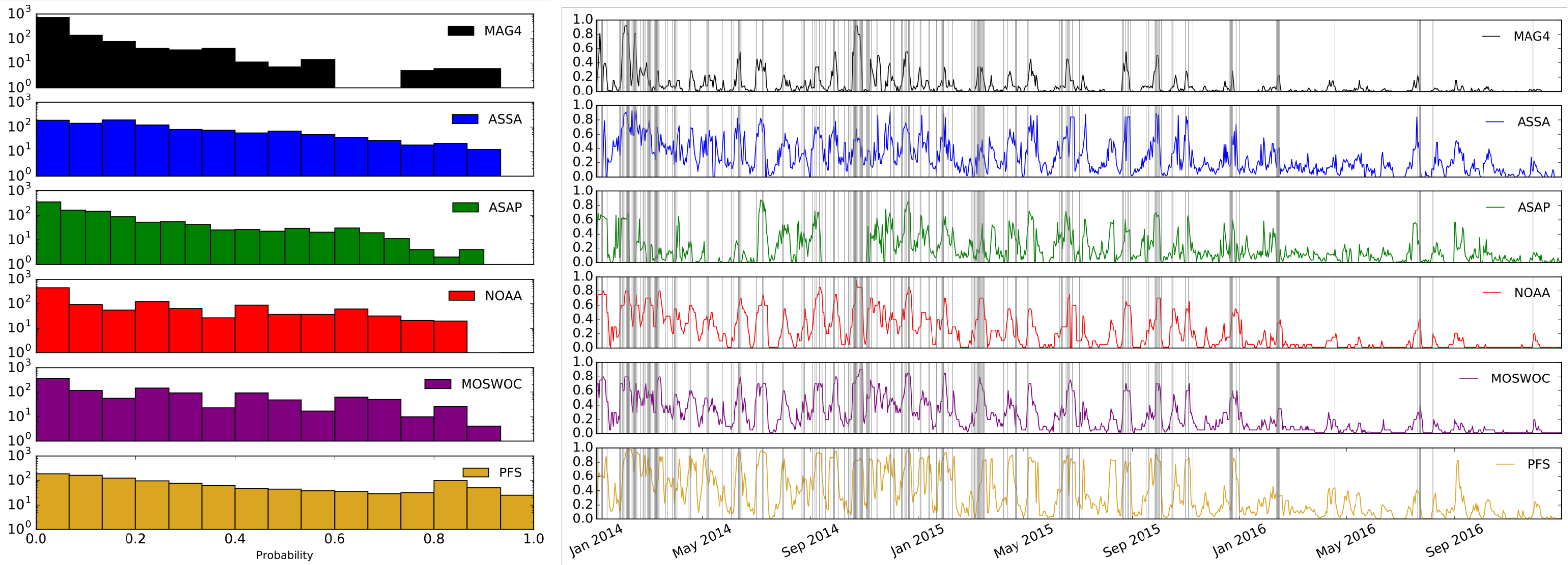

Figure 4: Probabilistic forecasts and events for M-class flares (histograms, left; time series, right). From the top, forecasting methods (color) are: MAG4 (black), ASSA (blue), ASAP (green), NOAA (red), MOSWOC (purple), and MCSTAT (gold). Vertical grey lines on the right signal positive events, i.e., days when at least one M-class flare was observed.

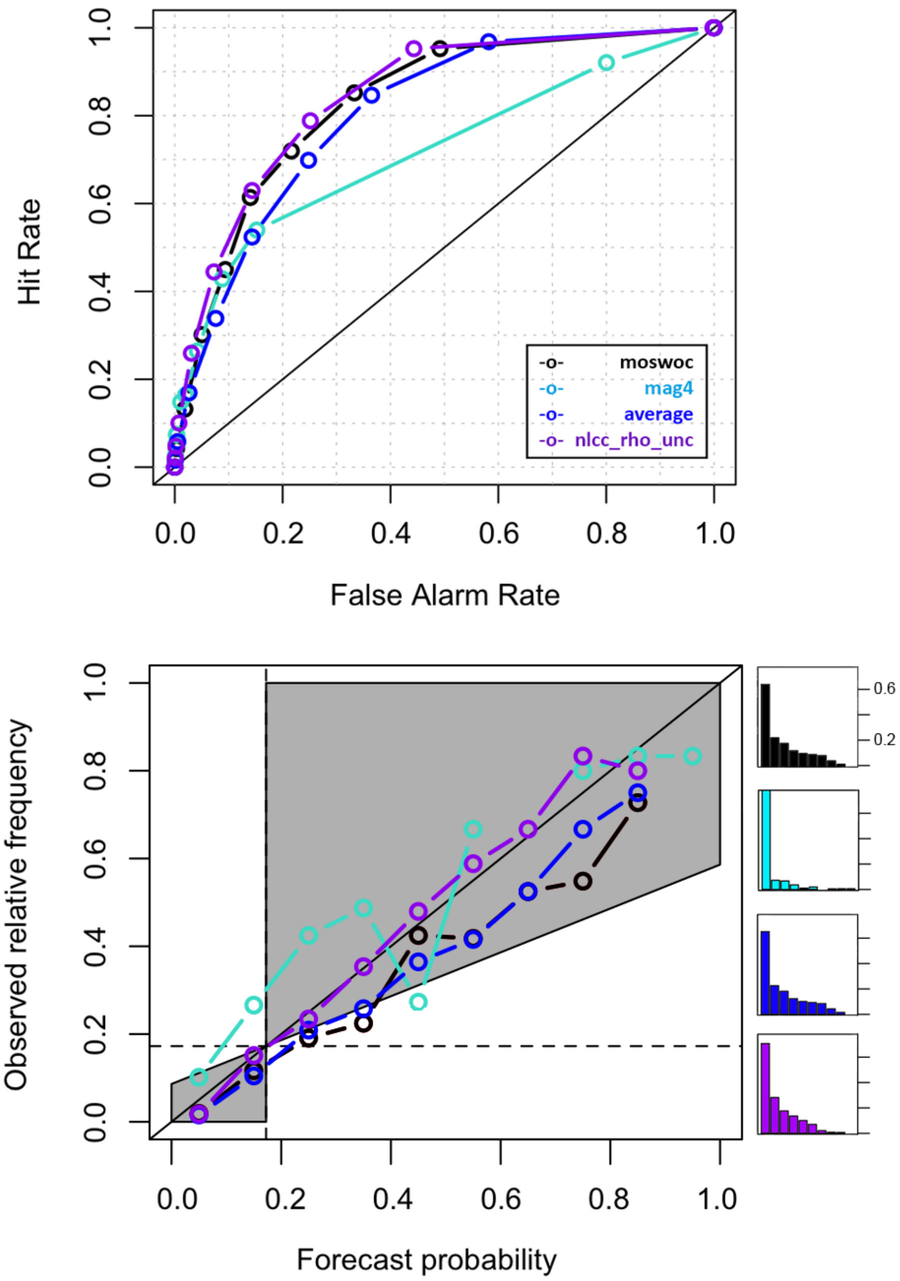

Figure 5: ROC curves (top) and reliability diagrams (bottom) for M-class flare forecasts, comparing the top performers overall. Note the center diagonal line in the ROC curves represents no skill, while for the reliability diagrams it indicates skill.

Table 2: Table with verification metrics for M-class flare forecasts. Note that there are 189 events and 907 non-events out of 1096 total records, and uncertainty for all cases is 0.143.

Forecast name	Brier score 1.0 - 0.0	Reliability 1.0 - 0.0	Resolution 0.0 - 1.0	ROC area 0.5 - 1.0	Group Rank	Overall Rank
ASAP	0.151	0.0163	0.0079	0.575	7	35
ASSA	0.150	0.0235	0.0167	0.738	5	34
MAG4	0.126	0.0064	0.0237	0.772	4	28
MCSTAT	0.183	0.0606	0.0200	0.769	6	33
MOSWOC	0.116	0.0056	0.0327	0.842	1	23
NOAA	0.116	0.0070	0.0335	0.838	2	25
Equal Weight	0.121	0.0046	0.0264	0.816	3	27
Brier	0.110	0.0009	0.0338	0.848	8	8
Brier_unc	0.107	0.0007	0.0368	0.853	2	2
LCC	0.109	0.0016	0.0355	0.848	6	6
LCC_unc	0.106	0.0019	0.0387	0.853	2	2
MAE	0.126	0.0064	0.0237	0.772	15	28
MAE_unc	0.127	0.0082	0.0244	0.811	15	30
NLCC_rho	0.110	0.0011	0.0334	0.849	7	8
NLCC_rho_unc	0.107	0.0007	0.0366	0.854	1	1
NLCC_tau	0.110	0.0018	0.0344	0.848	8	10
NLCC_tau_unc	0.109	0.0011	0.0351	0.856	2	4
REL	0.114	0.0013	0.0298	0.831	14	26
REL_unc	0.111	0.0008	0.0322	0.841	12	19
RES	0.110	0.0009	0.0332	0.841	11	17
RES_unc	0.114	0.0010	0.0322	0.832	13	24
ROC	0.109	0.0021	0.0357	0.847	8	10
ROC_unc	0.108	0.0010	0.0357	0.853	5	5
Accuracy	0.112	0.0023	0.0335	0.890	10	19
Accuracy_unc	0.126	0.0131	0.0297	0.625	11	31
Brier	0.111	0.0008	0.0327	0.891	8	19
Brier_unc	0.129	0.0156	0.0289	0.596	12	32
CSI	0.109	0.0013	0.0350	0.889	1	7
CSI_unc	0.111	0.0062	0.0376	0.630	2	12
GSS	0.110	0.0008	0.0338	0.839	3	12
GSS_unc	0.129	0.0221	0.0360	0.878	7	18
HSS	0.111	0.0033	0.0349	0.889	7	16
HSS_unc	0.108	0.0021	0.0372	0.620	5	15
TSS	0.111	0.0013	0.0348	0.856	3	12
TSS_unc	0.130	0.0227	0.0359	0.879	9	22

Presentations

Ensembles for space weather forecasting, Murray, S. A., April 2019 (invited), Space Weather Workshop, USA.

Fostering scientific collaboration for improved space weather forecasting, Murray, S. A., February 2019 (invited), Chapman Conference on Space Weather, Pasadena, USA.

Fostering scientific collaboration for improved space weather forecasting, Murray, S. A., et al, December 2018, AGU Fall Meeting, USA.

Learning from terrestrial weather: improved verification measures for space weather forecasting, Murray, S. A., et al, May 2018 (invited), Triennial Earth Science Summit, USA.

Ensemble flare forecasting: using numerical weather prediction techniques to improve space weather operations, Murray, S. A., et al, April 2018, EWASS 2018, UK.

Solar eruptive forecasting, Murray, S. A., January 2018 (invited), CFSA seminar, University of Warwick, UK.

Ensemble flare forecasting: using numerical weather prediction techniques to improve space weather operations, Murray, S. A. et al, December 2017 (invited), AGU Fall Meeting, USA.

Modelling ensemble forecasts of solar flares, Guerra, J. A., et al, September 2017, European Meteorological Society Annual Meeting, Ireland.

References

J. A. Guerra, S. A. Murray, D. S. Bloomfield, and P. T. Gallagher. Modeling Ensemble Forecasting of Major Solar Flares. *under preparation for submission to Space Weather*.

J. A. Guerra, A. Pulkkinen, and V. M. Uritsky. Ensemble forecasting of major solar flares: First results. *Space Weather*, 13:626–642, Oct. 2015. doi: 10.1002/2015SW001195.

P. A. Higgins, P. T. Gallagher, R. T. J. McAteer, and D. S. Bloomfield. Solar magnetic feature detection and tracking for space weather monitoring. *Advances in Space Research*, 47(12):2105–2117, Jun 2011. doi: 10.1016/j.asr.2010.06.024.

S. A. Murray. The Importance of Ensemble Techniques for Operational Space Weather Forecasting. *Space Weather*, 16(7):777–783, Jul 2018. doi: 10.1029/2018SW001861.

S. A. Murray, S. Bingham, M. Sharpe, and D. R. Jackson. Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather*, 15:577–588, Apr. 2017. doi: 10.1002/2016SW001579.

S. A. Murray, J. A. Guerra, P. Zucca, S.-H. Park, E. P. Carley, P. T. Gallagher, N. Vilmer, and V. Bothmer. Connecting Coronal Mass Ejections to Their Solar Active Region Sources: Combining Results from the HELCATS and FLARECAST Projects. , 293(4): 60, Apr 2018. doi: 10.1007/s11207-018-1287-4.