Topical Discussion Meeting report

Name of the meeting: Flare forecasting: where are we and where should we be going?

Conveners: S. A. Murray (TCD), M. L. Mays (NASA/CCMC), K. D. Leka (NWRA), M. K. Georgoulis (AA), D. S. Bloomfield (UNN)

Date – Time – Room: 2018-11-06 17:15 MTC 00.15

Nr of participants: ~60

Objective of the TDM

This focus of this TDM was to highlight the ongoing work and future plans of the solar flare forecasting community. Updates on the NASA/CCMC Flare Scoreboard, International Solar Flare Prediction Working Team, and the PSTEP/ISEE operational forecast comparison workshop were presented. Open discussion followed about the best direction forward for forecasting efforts, including operational limitations and the novel research that needs to be undertaken to improve the current state-of-the-art.

Some discussion highlights

Leila Mays of NASA/CCMC kicked off the session with an overview of the Flare Scoreboard, a system which compares operational center flare forecasts around the world online in real time. She presented the current status as well as future updates, including the integration of forecast verification. This led to some discussion over what the meaning of 'score' is and what types of forecasts could realistically be compared to each other (different region definitions and time cadence for example).

Sophie Murray of Trinity College Dublin then showcased a new method for combining the Scoreboard forecasts using ensemble techniques. An average probability is already available on the Scoreboard, but weighted ensembles could be used in future. It was noted that ensembles have been used successfully in terrestrial weather forecasting (such as hurricane forecasts) and could prove useful as a more accurate starting point for human forecasters. Providing uncertainty estimates using these techniques would also be useful for forecasting centers.

KD Leka of NWRA then presented results of the PSTEP/ISEE forecasting comparison workshop, which has verified operational flare forecasts using the same verification techniques on the same data set for the first time. No one method really seems to outperform all others, although they are better than the climatological benchmark.

The end of the session allowed some more open discussion regarding what the best next steps forward are in the community. Different techniques were discussed, including the use of artificial intelligence on large spacecraft data sets (with the caveat that it is used correctly, as first-generation machine learning methods aren't performing all that better at the moment). There was a general consensus that eruption prediction is the next task once flare forecasting has improved sufficiently.

Main conclusion of the meeting

Lots done, lots more to do! Moving forward we should avoid replication and continue to work together as an international community for improving the current state-of-the-art.

Annexes

All presentation slides from the meeting are available at Presentation/discussion materials are available online: https://ccmc.gsfc.nasa.gov/assessment/topics/flare/esww15.php

Minutes of Meeting:

Flare Scoreboard Presentation (Leila Mays):

- Q: What is meant by real-time comparison?
    - A: Comparison between forecasts as and when the centers upload their probabilities. There were some ftp server issues but this is now back online so the 'real-time' aspect will be more official soon.
- Q: Scientists want AR forecasts rather than full disk
    - A: The purpose of the Scoreboard is an operational comparison rather than scientific, which generally has a focus on full-disk M- and X- class forecasts. These have particularly been chosen for verification purposes due to ease of comparison. However, the scoreboard does also provide active region forecasts for those centers who provide it.
- Q: flare "scoreboard" doesn't have a "score"
    - A: True currently, but we are working with the UK Met Office automated verification system to produce some real-time flare scoreboard stats. This will go online when completed.
- Q: Does the scoreboard accept forecasts at any time cadence?

    - A: Yes, but forecasts can be browsed using the arrows with a 1-hour time cadence for forecast issue times.

- Q: Combining regions could be an issue

    - A: Yes, for the scoreboard, regions are combined in the interface, but when you download the data, nothing has been combined and the user can apply their own combining algorithm.

    - Comment: For US SWPC forecasts the regions are defined by the US Air Force, and then NOAA numbers are assigned by SWPC if the regions persist.

Ensembles Presentation (Sophie Murray)

- Message: Weighted ensembles of Flare Scoreboard forecasts tend to outperform the forecasts on their own. Depending on the score you choose to optimize your ensemble weights, you get different results. Discussed constrained and unconstrained weights that can get rid of the effect of duplicate forecast "types".

- Comment: Find it interesting to see an ensemble of multiple models vs an ensemble generated by a single model, like hurricane forecasts.

- Q: For solar wind models, the ensembles are giving an estimate of uncertainty, is that the same for flare probability forecasts?

    - A: Yes, it's possible, but not trivial to extract uncertainties from the ensemble flare forecasts. Jordan is looking into this.

    - A (SWPC): Forecasters look at ensembles heuristically and throw out outliers, and look for the results that are more "tight" what to have more confidence in the forecasts.

- Q: But what if the outlier forecast is good or more sensitive?

    - A (SWPC): That's where the forecaster experience with the model over time is important and takes effort. Researchers can help forecasters by reporting model trends and biases.

- Q: Are the M class ROC curves similar to X? We noticed at SIDC that we are good at forecasting C flares, slightly worse at M, and bad at X.

    - A: The curves are definitely not as good. They also vary wildly between under and over forecasting for example for the reliability diagrams. This is common because of having less statistics from rare events.

- Q: Can an experienced forecaster beat models without using any tools?
    - A (SWPC): Yes, right now the tools are only valid near disk center anyway, so the forecaster can't rely on the tools heavily. The forecasters can forecast regions before they rotate on to the disk based on X-ray/EUV activity.

ISEE Workshop Presentation (KD Leka):

- Message: Hard to compare forecasting methods. This workshop focused on performance for a specific time period only. None of the methods have a skill score of 1, the Gini coeffecients are ok. Methods show good discrimination but not well with respect to climatology. No single metric answers your question, need to look at multiple metrics.
- Q: One method seems to have a very different reliability plot
    - A: They know why and the data may have been read it wrong, will be re-run in forecasting mode.
- Q: How are error bars for the reliability diagram determined?
    - A: Based on sample size.
- Q: Impressed by results, sets a benchmark for people to compare their methods to. Also because FLARECAST machinery is in place and this is the validation machinery.
    - A: Yes validation codes were published before and will be published again.
- Q/comment: But many users want eruptions to be predicted, not flares.
    - A: Yes, this can be done with similar methods using an event definition.
- Q/comment: Probably need more physics included in model to predict eruption based on conditions needed for energy release to get an eruption.
- Q/comment: Also need these models to successfully not predict a CME eruption when there are no CMEs (e.g. AR2192).
- Q/comment: Need to get the research to a point where we can predict eruptions.