1. Motivation

It is common to observe that for given photospheric conditions – somehow favorable to the production of flares – the likelihood of observing a major flare can vary (sometimes significantly) from method to method. Figure 1 shows that the spread of values around the mean value is only small for very low or very high probabilities. For values in between, the probability of observing a flare can have a wide range of values. In such cases, a combination of such values will be closer to the real probability. A simple average often proves to be superior to any individual value. However, in most cases, the average is not the best performing combination. In this paper, we investigate the construction of several ensemble predictions in order to provide some guidance in an operational environment about the best performing combination according to specific needs of any end user.

2. Forecasting Methods and Data

The forecasting methods included in the ensembles are listed in Table 1. Full disk probabilistic forecasts for the occurrence of a M-class flare between 2013 and 2016 are used. Figure 1 displays time series (a) and histograms (b) for each method.

3. Ensemble Construction

Probabilistic forecasts from the ensemble members $P = \{P_{MAG4}, P_{ASSA}, P_{ASAP}, P_{NOAA}\}$ are linearly combined as

$$ P'(w;r) = \sum w_i P_i(x) $$

with $w = \{w_{MAG4},w_{ASSA},w_{ASAP},w_{NOAA}\}$. Combination weights are constrained to

$$ \sum \pi_i = 1 $$

Therefore, problem is reduced to determined w by optimization of a performance metrics. Table 2 shows the list of performance metrics employed.

3.1 Ensemble Construction

Table 3 lists the ensemble predictions according to the overall performance across the four metrics: ROC area (Figure 4), Reliability and Resolution (Figure 5), and Brier score.

3.2 Performance metrics used for the optimization during the ensemble construction. All metrics are well-known and widely used for validation of forecasts. Categorical metrics are calculated from the contingency table after applying a decision threshold to probabilistic forecasts.

3.3 Performance metrics used for the optimization during the ensemble construction. All metrics are well-known and widely used for validation of forecasts. Categorical metrics are calculated from the contingency table after applying a decision threshold to probabilistic forecasts.

4. Concluding Remarks

Table 3 lists the ensemble predictions according to the overall performance across the four metrics: ROC area (Figure 4), Reliability and Resolution (Figure 5), and Brier score.

4.1 Performance metrics used for the optimization during the ensemble construction. All metrics are well-known and widely used for validation of forecasts. Categorical metrics are calculated from the contingency table after applying a decision threshold to probabilistic forecasts.

Table 4 shows the ensemble of the results of the construction of combination weights and thresholds. To determine the average of forecasts cadence in the resulting ensemble.

Contact information: jotan.guerra@tcd.ie / jotan.guerra@gmail.com

2. Preliminary Results

Figure 5 shows the determined combination weights for probabilistic metrics (a) and categorical metrics (b). For probabilistic metrics, it is expected that ROC ensemble displays the best ROC curve. However, when the ensemble scores show a very similar ROC curve as the ROC ensemble. For categorical ensembles, optimizing TSS seem to produce the best ROC curve.