CME Arrival Time & Impact

Discussion Questions

Christine Verbeke, Leila Mays, Sandro Taktakishvilli

ESWW14 Topical Discussion Meeting

Event selection discussion questions

- ~100 events will consist a mix of fast and slow CMEs
- Keep track of multiple CME events for future subsets
- For models that have a training period, keep this separate from validation set
- **Choose a set of events to validate or time period?**

CME input parameters discussion questions

- Need a source for CME inputs. Use CME catalogs to reduce bias
- Input parameters for 2011 have been provided, 2012 and 2013 will be provided soon
- **Difficulties finding all necessary parameters from one single catalog, how to merge?**
- Interface with 3D CME Kinematics Working Team. **Are different people fitting different parts of the CME?**

Skill scores discussion

- CME arrival time:
  - Mean error, mean absolute error, RMSE
- CME arrival (yes/no):
  - Skill scores based on contingency tables:
    POD, POFD, FAR, Bias, Accuracy, CSI, ETS, HK, HSS
  - Vary hit definition (different user needs): 3, 6, 12, 18, 24, 36 hrs
  - Convert probabilistic to categorical using an agreed upon threshold
- Probabilistic CME arrival predictions:
  - Brier skill score, reliability diagram
- CME impact (mean or peak? v, B, n, T)
  - Correlation coefficient, other?
- Compute error bars for skill scores to allow comparison

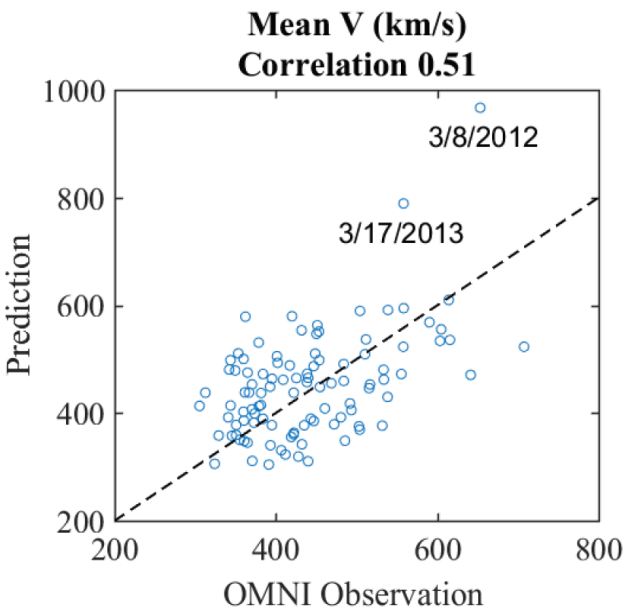
Skill score discussion questions

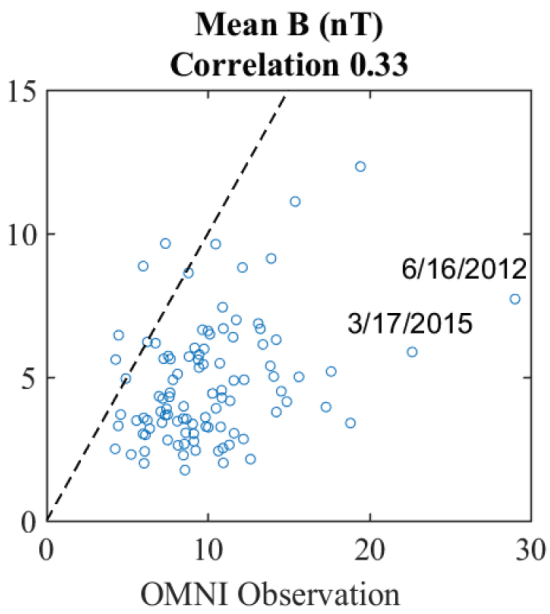
- Direct v, B, n, T time-series comparison?
- How will each model define the modeled CME arrival time and duration? Use the same algorithm for similar models? Model output submitter's decision? The metadata will keep track of human vs. algorithm identified model arrivals.
- How do we want to compare the time series of the modeled events to the observed time series? Multiple options:
  - No shift
  - Shift towards same arrival time
  - Shift towards density/velocity peak
  - Shift so that the performance of model is best (IMF Bz team choice)
- Is there a need for a combined skill score for user that can summarize overall model performance?

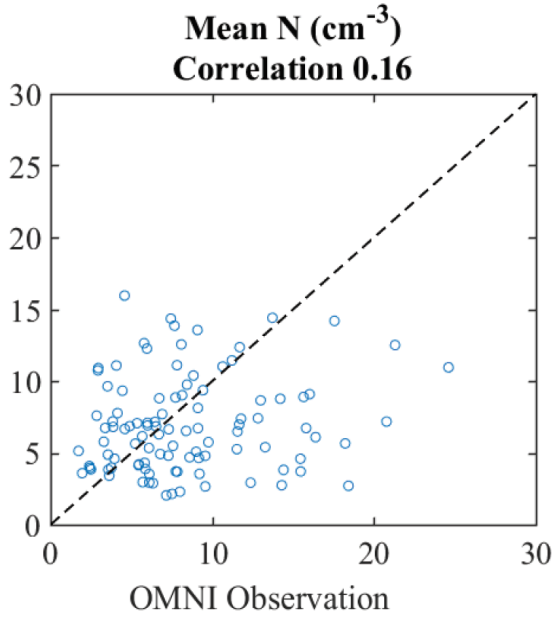
Quantifying CME impact

mean or peak? v, B, n, T

Correlation coefficient example

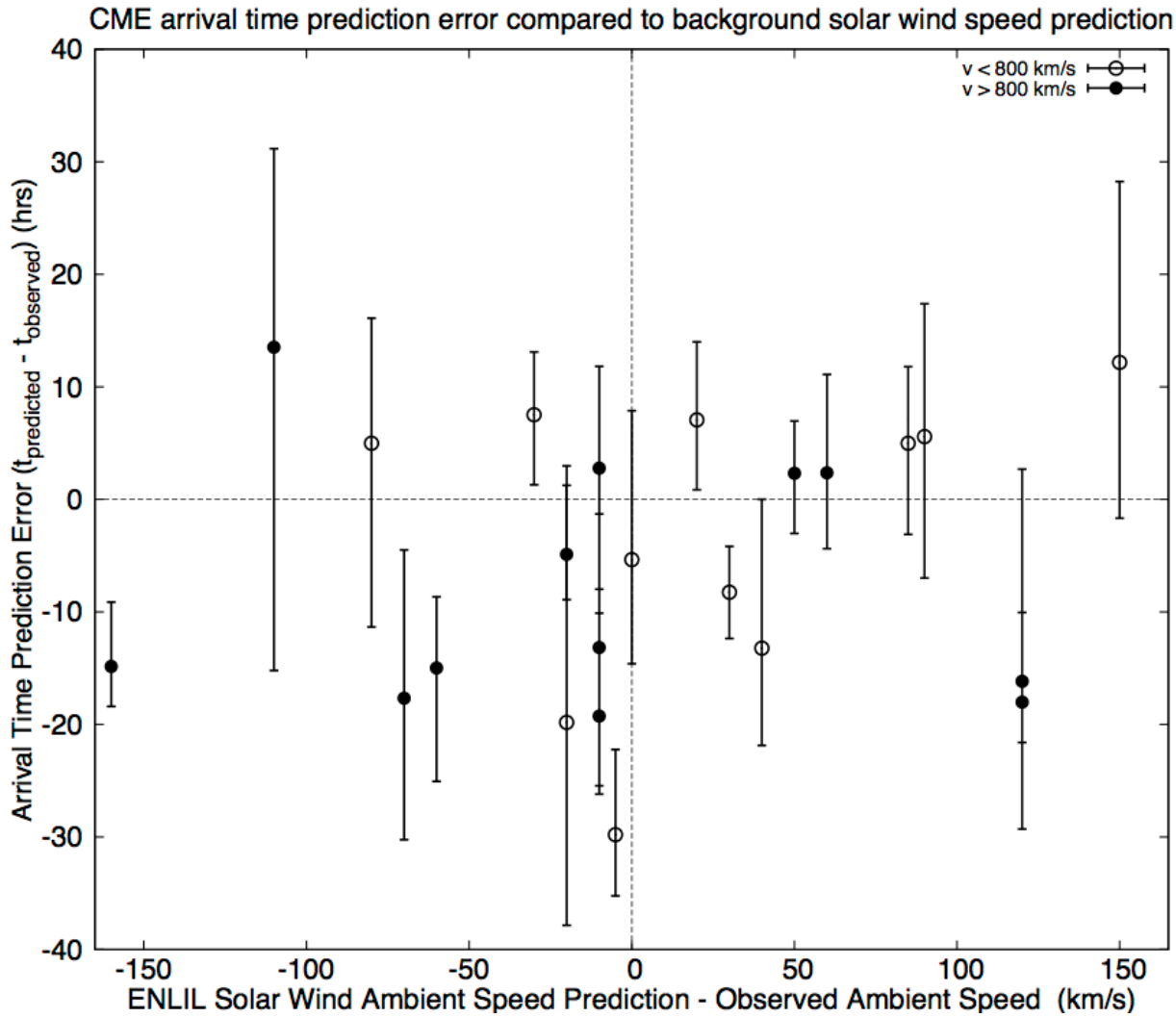






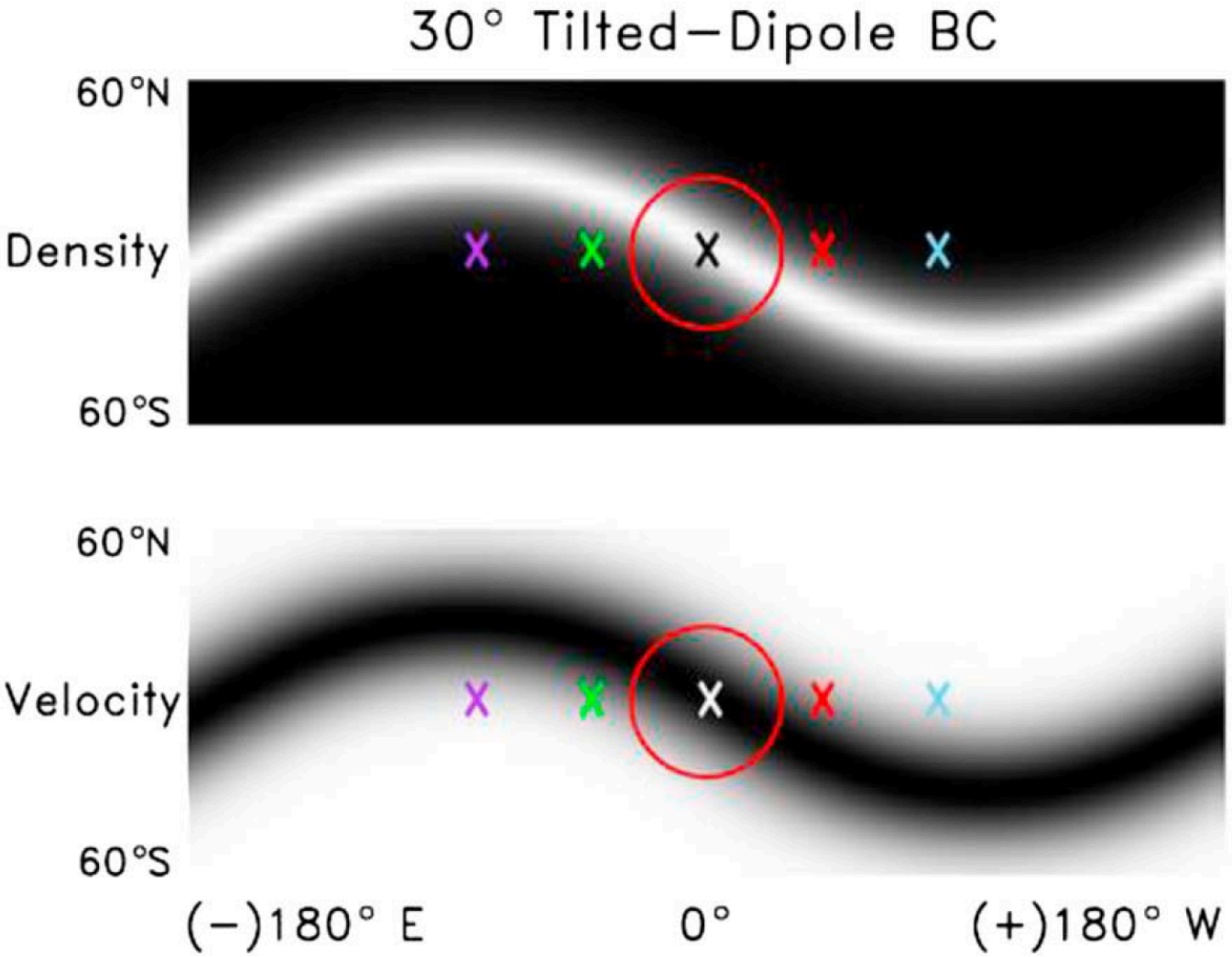
See poster by L.K. Jian in session 10

How best to quantify the effect of the background solar wind?



Example comparing the CME arrival time prediction error to the difference between the average predicted and observed solar wind speed 1 day before CME arrival

How best to quantify the effect of the background solar wind?

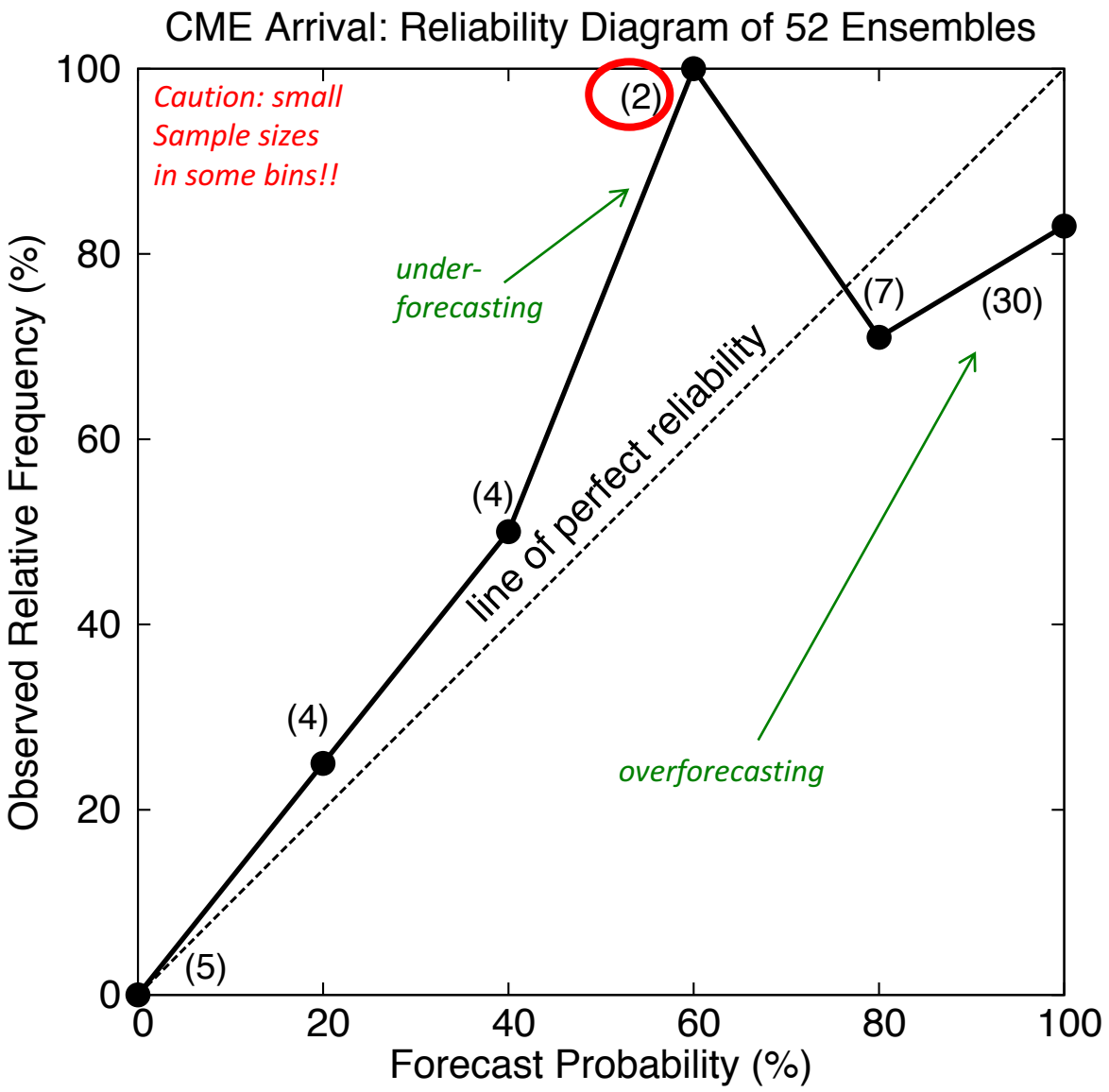


Pizzo, SW, 2015

Results

Score	MOSWOC	5% CL	95% CL	Score-board averg	5% CL	95% CL	A measure of...
Hits	**33**			**27**			Number of times a yes forecast was a yes occurrence.
Misses	**9**			**0**			Number of times a no forecast was a yes occurence.
False alarms	**6**			**12**			Number of times a yes forecast was a no occurence.
Correct rejections	**7**			**9**			Number of times a no forecast was a no occurrence.
Hit rate	**0.79**	0.68	0.88	**1**	1	1	Discrimination 1=perfect S=perfect. Ranges do not overlap.
False alarm rate	**0.46**	0.23	0.7	**0.57**	0.4	0.75	Discrimination S>M, however ranges overlap.
False alarm ratio	**0.15**	0.07	0.25	**0.31**	0.19	0.43	Reliability S is significantly higher than M. Ranges just overlap.
Probability of detection	**0.6**	0.49	0.69	**0.56**	0.46	0.69	
Probability of false detection	**0.12**	0.05	0.18	**0.25**	0.17	0.35	
Proportion correct	**0.73**	0.64	0.82	**0.75**	0.65	0.83	Accuracy Fraction of hits & correct rejections. Comparable for both.
Base rate	**0.76**	0.67	0.86	**0.56**	0.46	0.69	
Forecast rate	**0.71**	0.6	0.8	**0.8**	0.73	0.9	
Threat score	**0.69**	0.58	0.79	**0.69**	0.57	0.81	Accuracy 0=no skill, 1=perfect Comparable for both.
Bias score	**0.93**	0.79	1.09	**1.44**	1.24	1.76	Bias 1=perfect M<1 so under-forecasting. S>1 so over-forecasting. Ranges overlap.
Equitable threat score	**0.18**	0.04	0.34	**0.3**	0.16	0.47	Skill Accounts for hits occurring by chance in the threat score. 0=no skill, 1=perfect. Two approaches are comparable & ranges overlap.
Heidke score	**0.3**	0.07	0.51	**0.46**	0.27	0.64	Skill Fractional improvement over just chance. M slightly lower than S & ranges ovelap. Suggests some skill in both forecasting approaches.
Peirce score	**0.32**	0.08	0.57	**0.43**	0.25	0.6	Skill Similar to Heidke. The two approaches are comparable & ranges overlap.

Assessment: Confidence (likelihood) in CME arrival



- Example reliability diagram for CCMC/SWRC arrival time forecasts
- Underforecasting in the forecast bins between 40-80%
- Slightly overforecasting in the 80-100% forecast bins

Likelihood of CME arrival forecast verification: Brier Score

Using the forecast probability about the likelihood that the CME will arrive submitted on the scoreboard.

A method defining the mean squared probability forecast errors is the Brier Score:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

N = number of events,
p_i = forecast probability of occurrence for event i,
o_i = 1 if the event was observed to occur and 0 if it did not.

Ranges from 0 to 1, with 0 being a perfect forecast.

The Brier Skill Score (BSS) is the the Brier score relative to climatology

Note: confidence intervals should be computed for verification scores