

IMF Bz Team

Neel Savani | (Pete Riley)

- **Team Recap:**
 - Definitions and formulation of the problem.
- **What's new:**
 - ROC Skill metrics showing promising initial results.
- **To the Future:**
 - Strategy for keeping up the momentum



Recap

Executive 'Tiger' team

- Small team of active participants from around the world:
 - National forecasters
 - Scientists
 - Other Team Leads
- Slack communication system
 - 18 people
 - 6 Countries
 - 10 Time Zones



Scientists

N. Savani	P. Riley
L. Mays	M. Owens
Y. Collado Vega	A. Vourlidas
S. Patsourakos	C. DeForest
A. Rouillard	S. Poedts
D. Shiota	E. Henley
C. Verbeke	N. Lugaz
R. Steenburgh	C. Dekonig
M. West	H. Singer

Forecasting agencies

US NOAA / SWPC
UKMO / MOSWOC
Japan NICT / SWx

Recap

Open Themes

- Draft Document sent to whole Community
 - Feb 2017
 - > 110 participants
 - 6 themes were discussed
 - Also found on CCMC site

Document Topics

1. Background Solar Wind
2. Core event selection
3. Magnetic What?
4. B Magnitude threshold
5. Time resolution
6. Validation Metric

IMF Bz at L1 Working Team

Leads: N. Savani, P. Riley (contact team leads/forum organizers to be added to the team)

Communications: ccmc-imf-bz@googlegroups.com (mailing list)

Participants: Eric Adamson · Nick Arge · Michael Balikhin* · Francois-Xavier Bocquet · Sean Bruinsma* · Yaireska Collado-Vega* · Pedro Corona-Romero* · Curt de Koning* · Manolis K. Georgoulis* · Edmund Henley · Bernard Jackson* · Leif Jøen · Christine Kay · Néaj Lugo · Anthony Mannucci* · Periyasamy K. Manoharan* · Slavek Mednik* · Marilena Miodo · Joseph Minner* · Christian Moseth · Kerin

Recap

Main conclusions

- Forecaster end result should work towards a single sentence that identifies 3 quantities:
 - A duration window for the forecast in the future
 - A field strength to exceed
 - An probability of uncertainty.

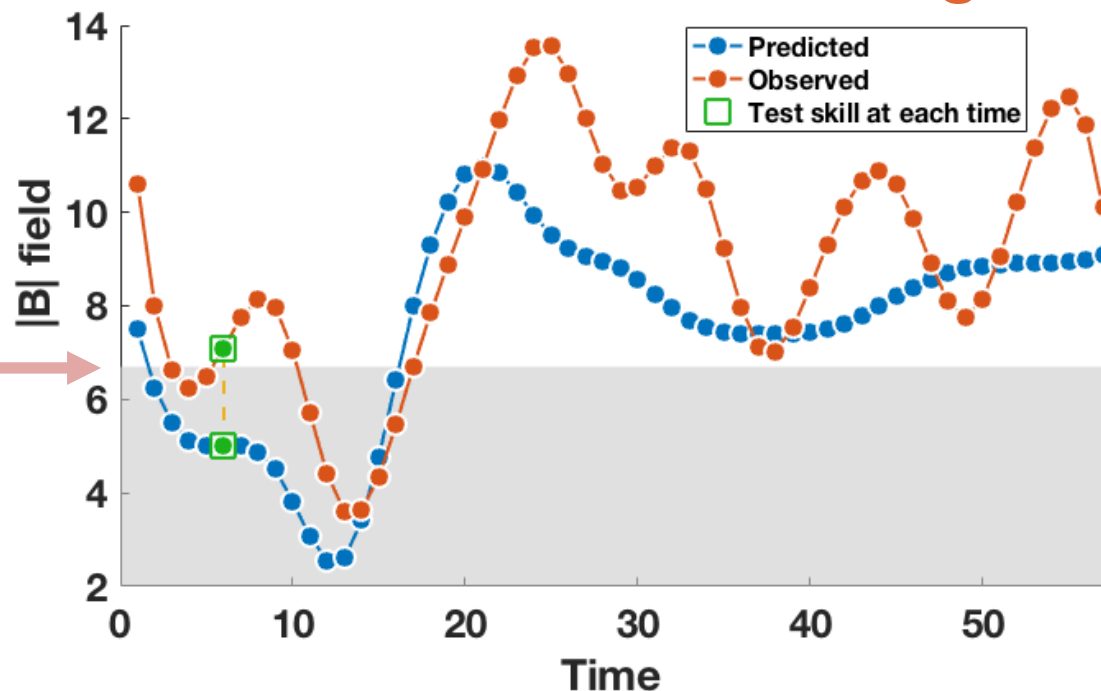
*“We forecast, in the next **24 hours** for a minimum of **60 minutes** the IMF Bz will drop below **-10nT** with **75% probability**.”*

What's New

ROC curve

- Work in progress.
 - Conversion of a deterministic forecast into probabilistic → i.e. use uncertainty.
 - Guidance taken from flare forecasting community

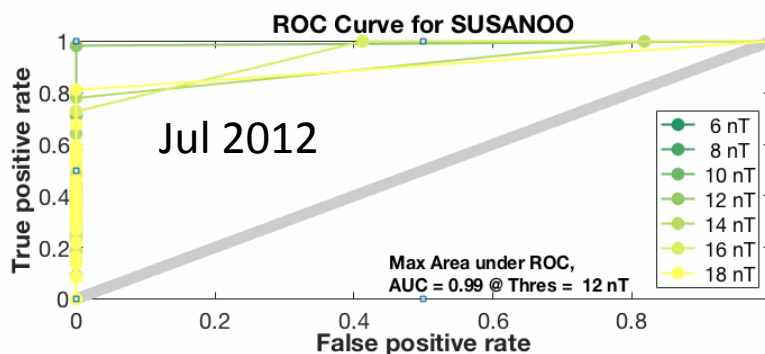
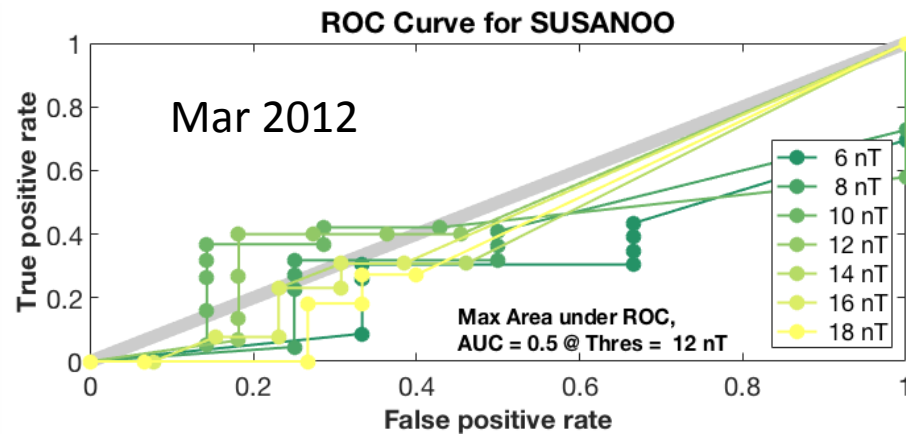
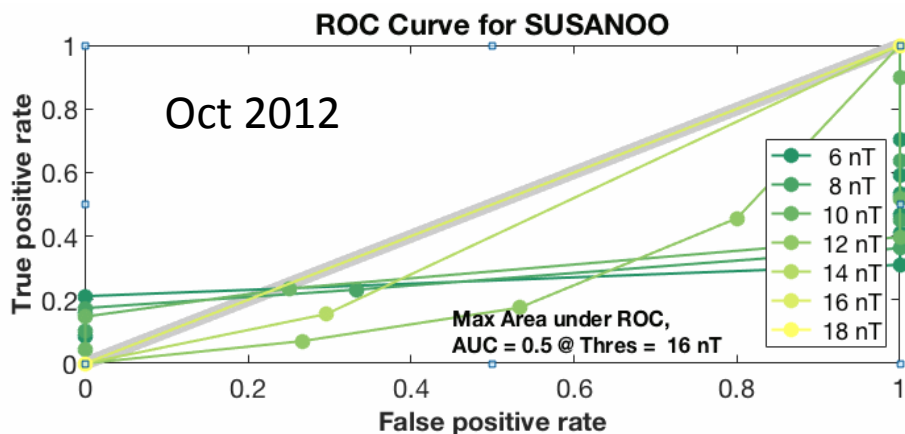
(Event Definition
Intensity, I.
[equivalent to
M-class boundary])



What's New

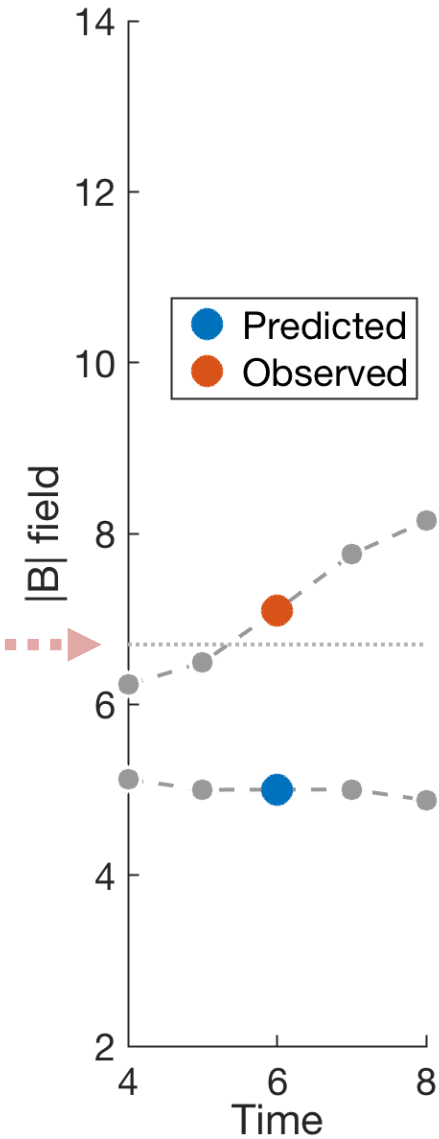
ROC curve

- Preliminary results using SUSANOO
- Variety of results shown:
 - lack of independence between points is the cause?
 - Period of analysis require more than CME time?



A report of the methodology will be sent to everyone for comment in the coming months

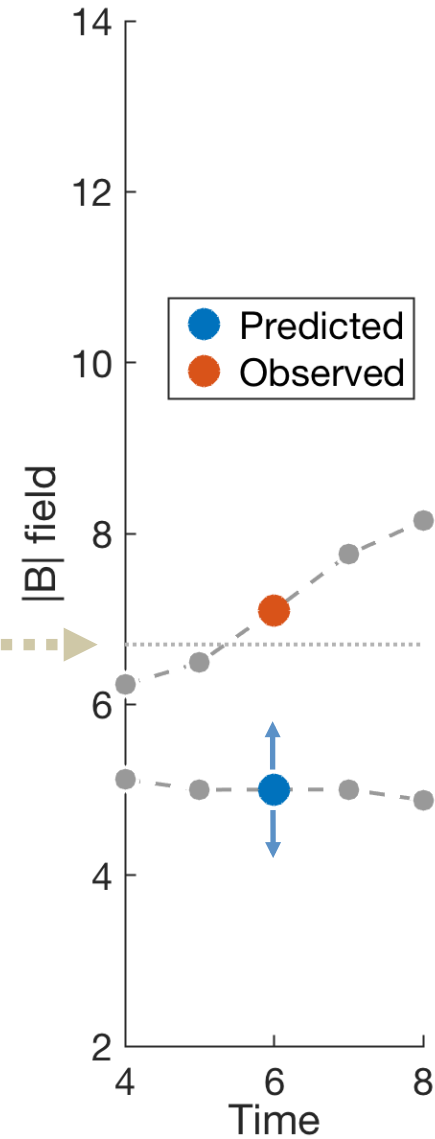
Event Definition
Intensity threshold, I
[equivalent to M-class
boundary]



Event Definition
Intensity threshold, I
[equivalent to M-class
boundary]

1. Each and every predicted
value is required to have an
uncertainty (error) associated

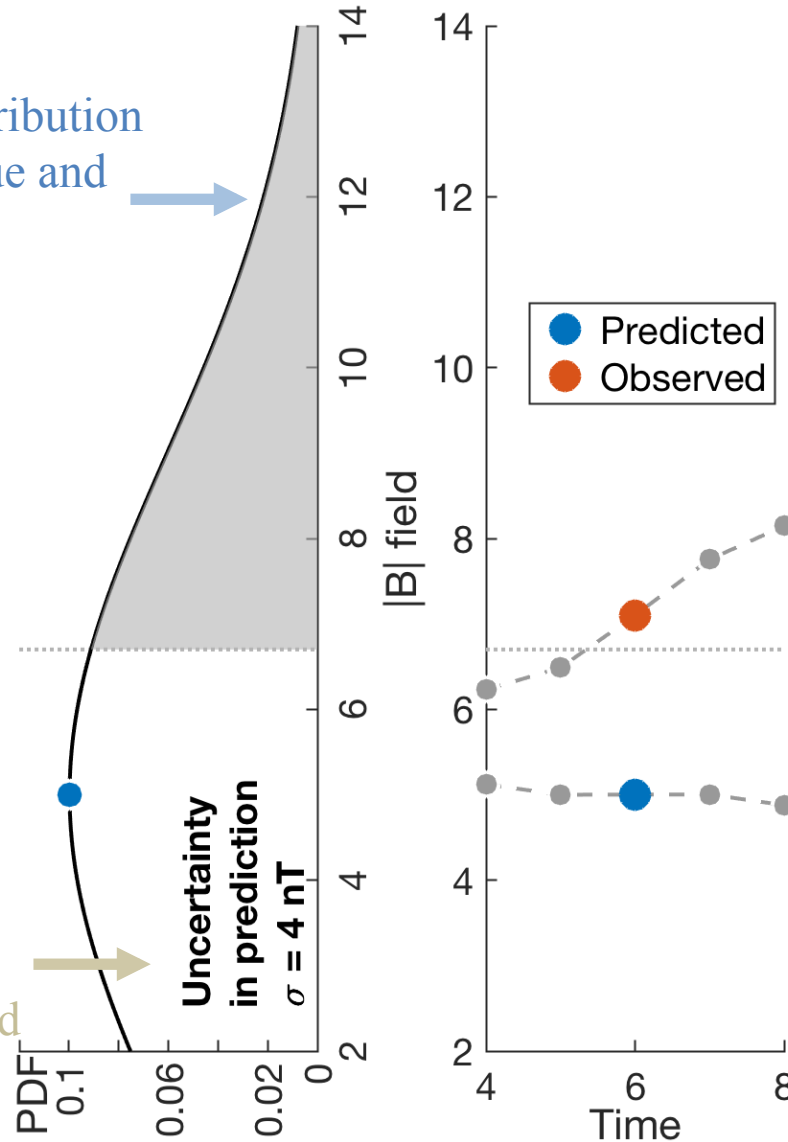
Uncertainty
in prediction
 $\sigma = 4 \text{ nT}$



2. Create a normal distribution from the predicted value and its uncertainty

Event Definition
Intensity threshold, I
[equivalent to M-class
boundary]

1. Each and every predicted value is required to have an uncertainty (error) associated

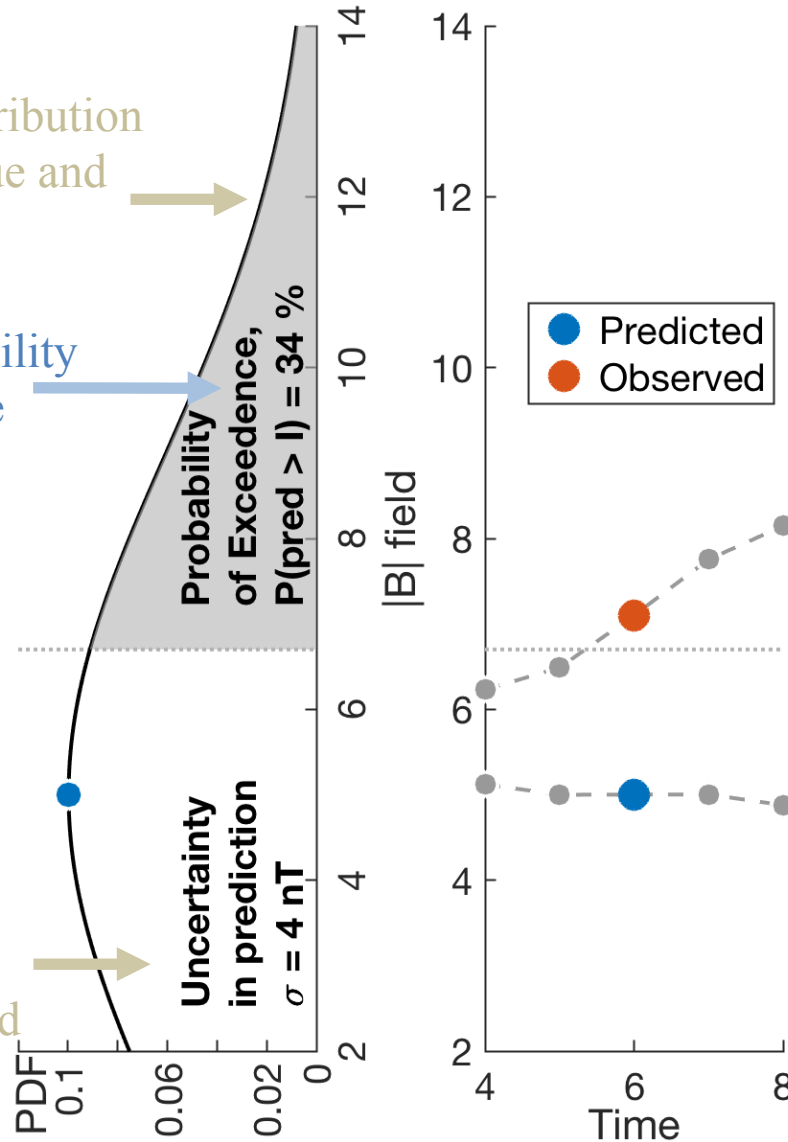


2. Create a normal distribution from the predicted value and its uncertainty

3. Calculate the probability that prediction is above event definition

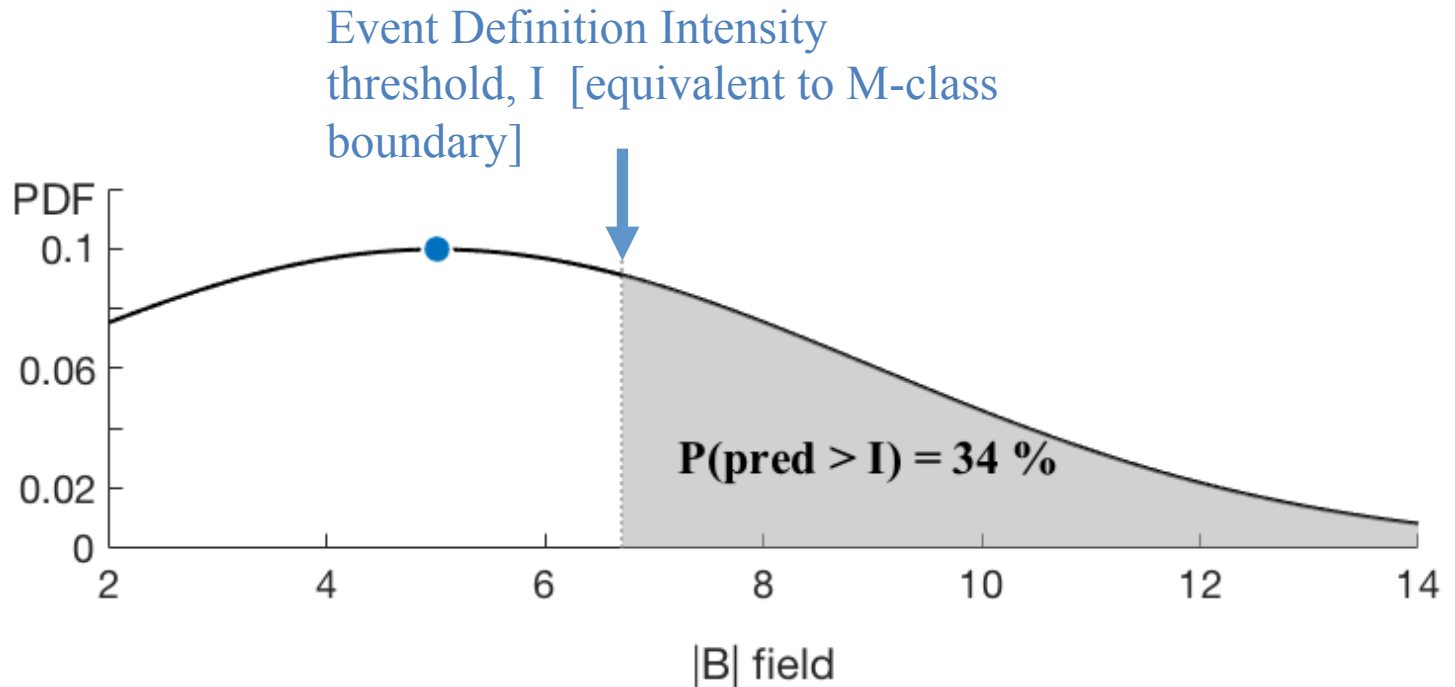
Event Definition
Intensity threshold, I
[equivalent to M-class boundary]

1. Each and every predicted value is required to have an uncertainty (error) associated



1. So, we have a fixed probability that our prediction is above Event Definition, I (which is equivalent to the M-class intensity boundary)

$$P(\text{pred} > I) = 34\%$$



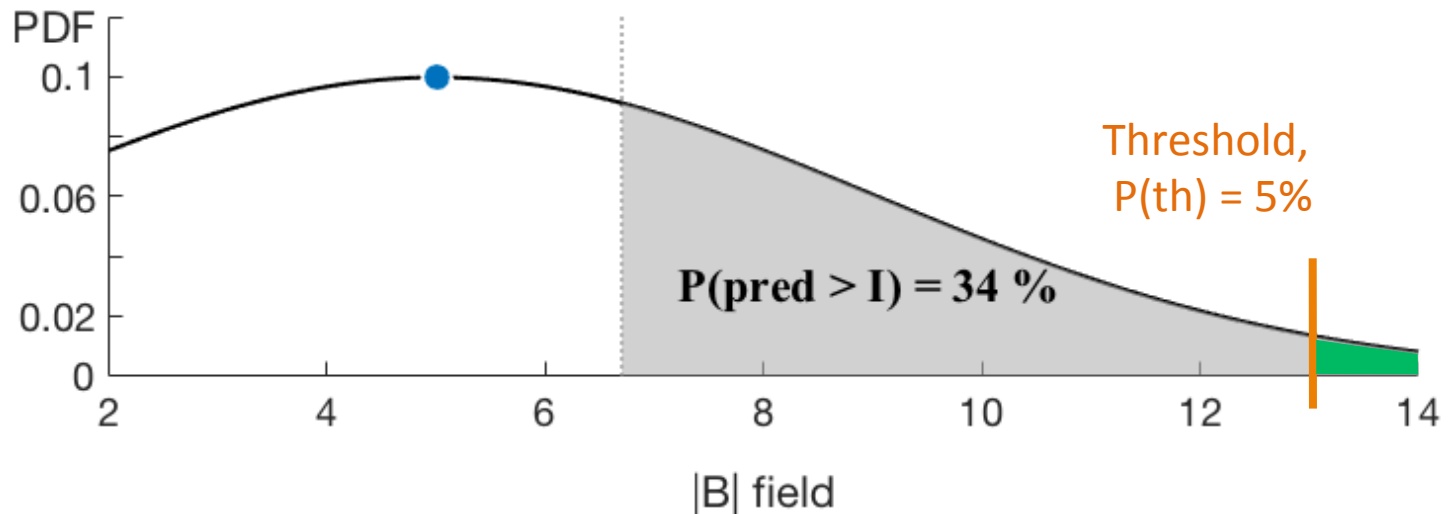
1. So, we have a fixed probability that our prediction is above Event Definition, I (which is equivalent to the M-class intensity boundary)

$$P(\text{pred} > I) = 34\%$$

2. Lets define a threshold, $P(\text{th})$, for which we require the prediction probability to exceed, in order to qualify as a ‘Yes’ predicted event.

$$P(\text{pred} > I) > P(\text{th}) \rightarrow \text{Prediction} = Y$$

$$P(\text{pred} > I) < P(\text{th}) \rightarrow \text{Prediction} = N$$



1. So, we have a fixed probability that our prediction is above Event Definition, I (which is equivalent to the M-class intensity boundary)

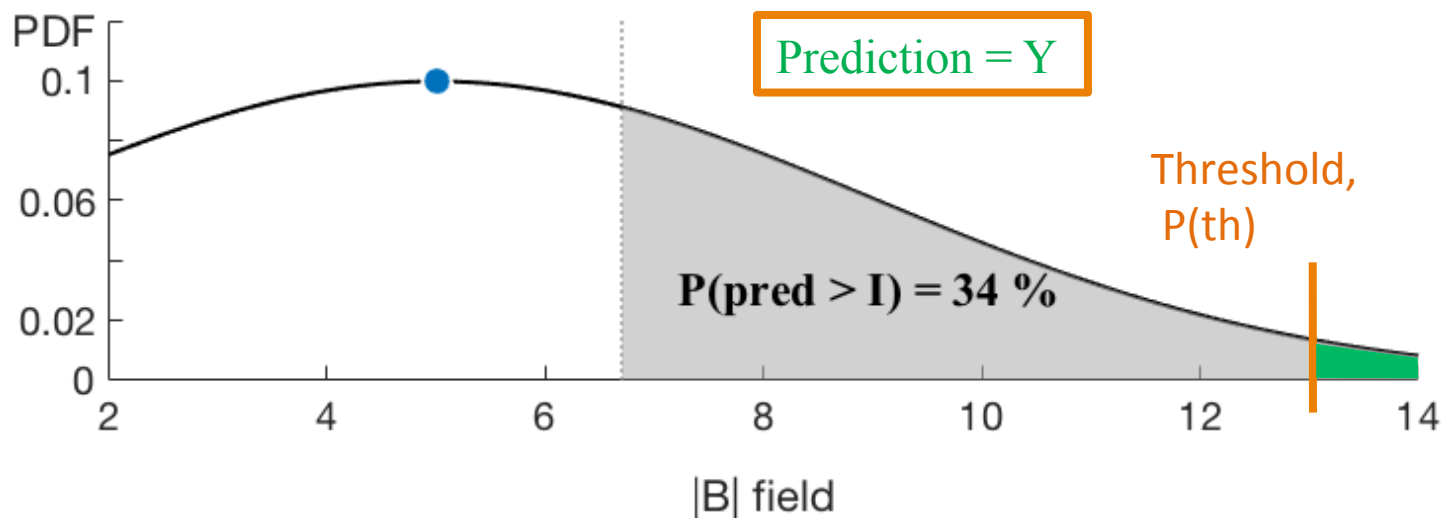
$$P(\text{pred} > I) = 34\%$$

2. Lets define a threshold, $P(\text{th})$, for which we require the prediction probability to exceed, in order to qualify as a ‘Yes’ predicted event.

$$P(\text{pred} > I) > P(\text{th}) \rightarrow \text{Prediction} = Y$$

$$P(\text{pred} > I) < P(\text{th}) \rightarrow \text{Prediction} = N$$

3. We can now manually change this threshold to vary across 0-100%. [This is equivalent to varying probability threshold for Flare forecasting].



1. So, we have a fixed probability that our prediction is above Event Definition, I (which is equivalent to the M-class intensity boundary)

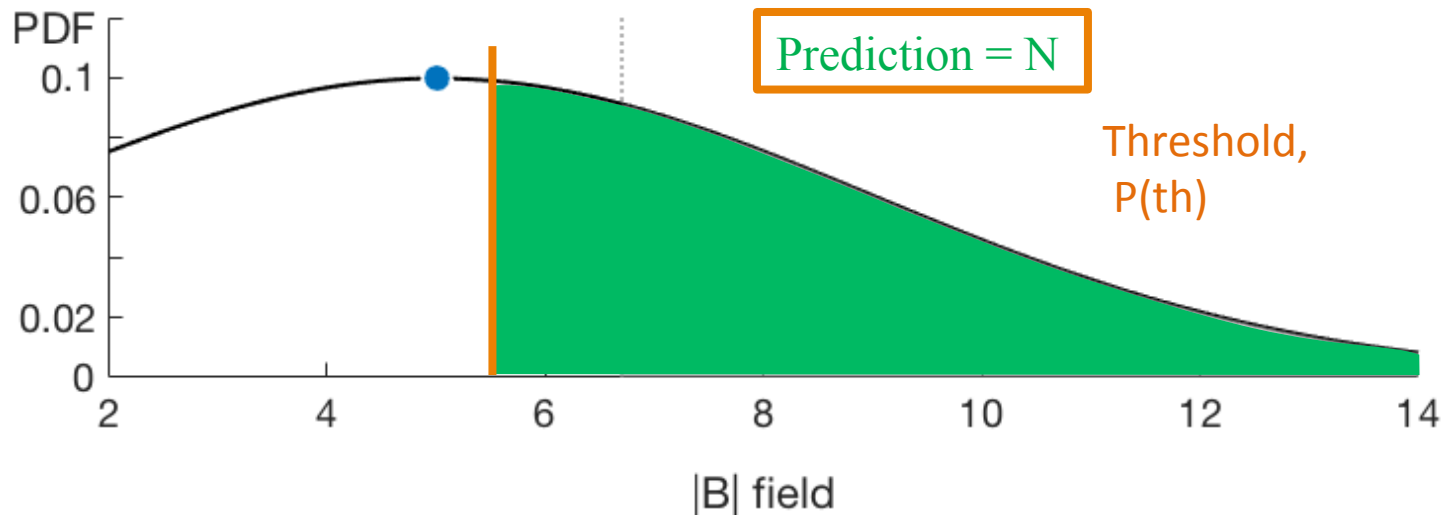
$$P(\text{pred} > I) = 34\%$$

2. Lets define a threshold, $P(\text{th})$, for which we require the prediction probability to exceed, in order to qualify as a ‘Yes’ predicted event.

$$P(\text{pred} > I) > P(\text{th}) \rightarrow \text{Prediction} = Y$$

$$P(\text{pred} > I) < P(\text{th}) \rightarrow \text{Prediction} = N$$

3. We can now manually change this threshold to vary across 0-100%. [This is equivalent to varying probability threshold for Flare forecasting].



1. So, we have a fixed probability that our prediction is above Event Definition, I (which is equivalent to the M-class intensity boundary)

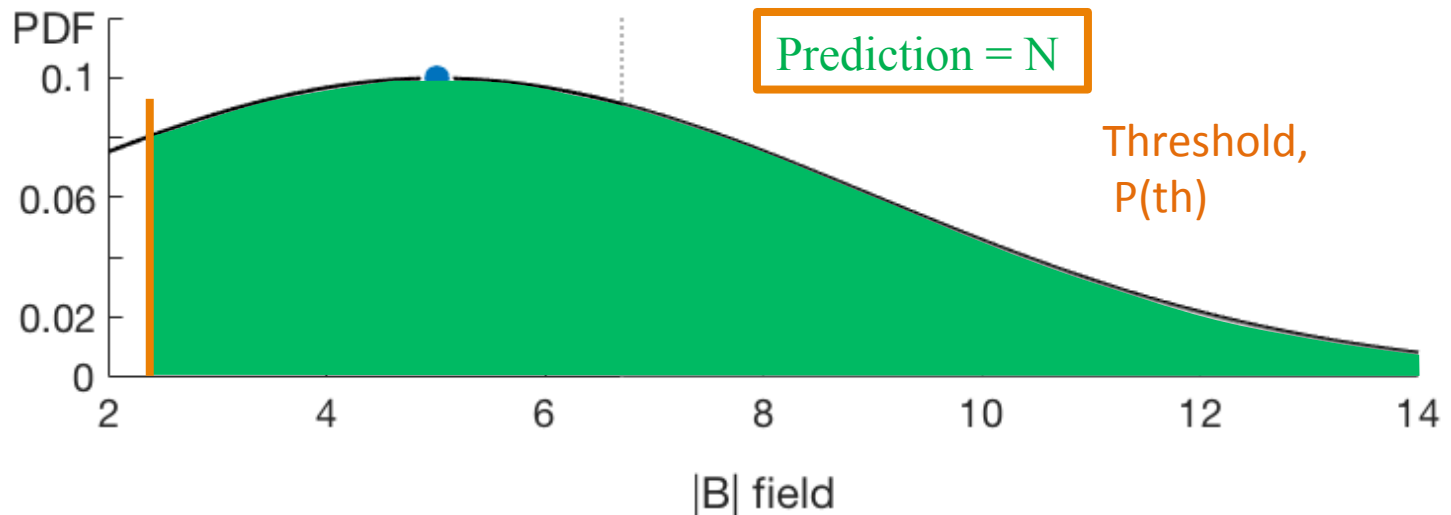
$$P(\text{pred} > I) = 34\%$$

2. Lets define a threshold, $P(\text{th})$, for which we require the prediction probability to exceed, in order to qualify as a 'Yes' predicted event.

$$P(\text{pred} > I) > P(\text{th}) \rightarrow \text{Prediction} = Y$$

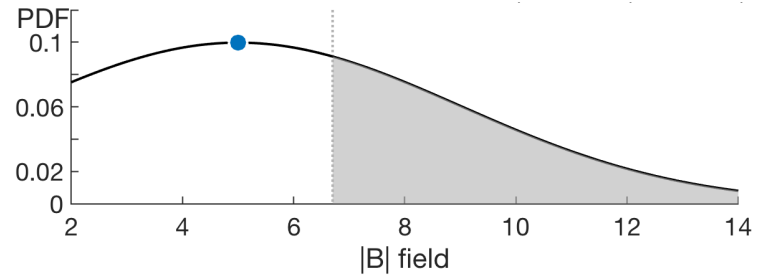
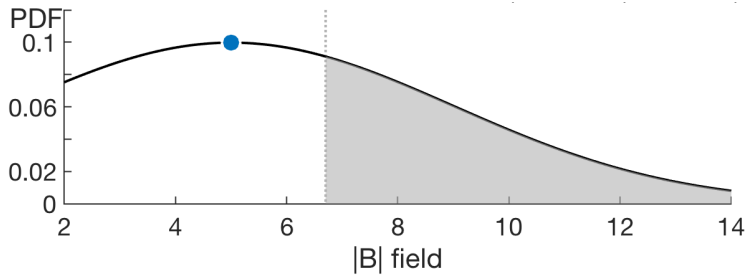
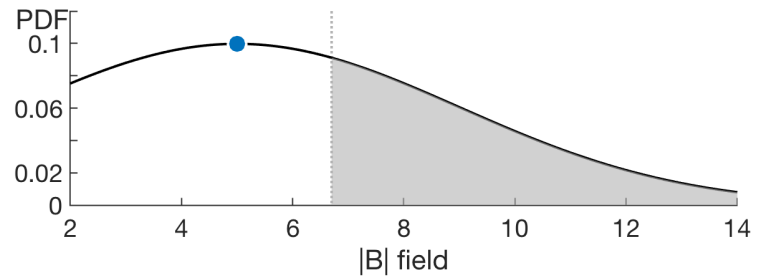
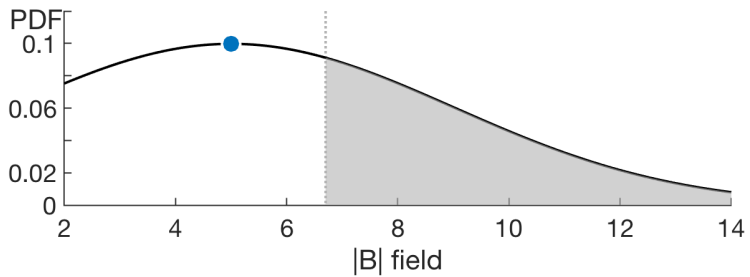
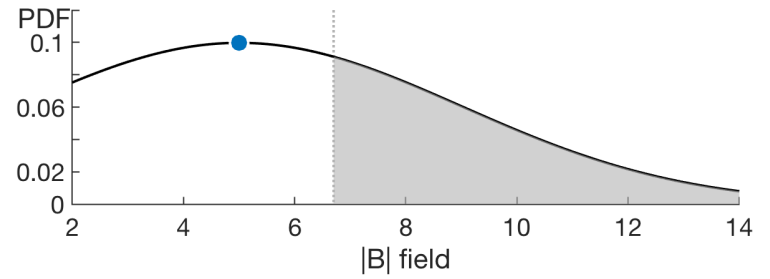
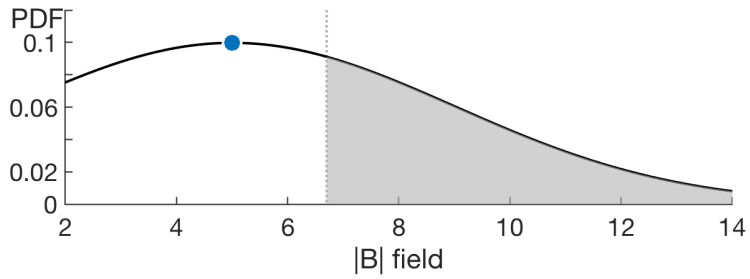
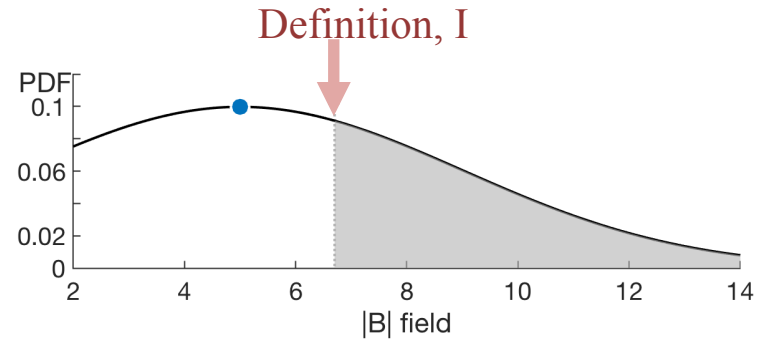
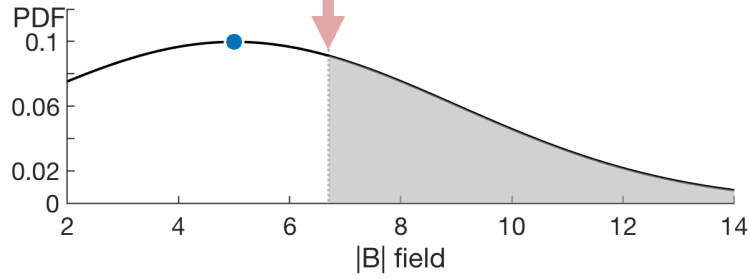
$$P(\text{pred} > I) < P(\text{th}) \rightarrow \text{Prediction} = N$$

3. We can now manually change this threshold to vary across 0-100%. [This is equivalent to varying probability threshold for Flare forecasting].



1. Now consider if the Observation was measured above/below Event Definition (I).

Event
Definition, I



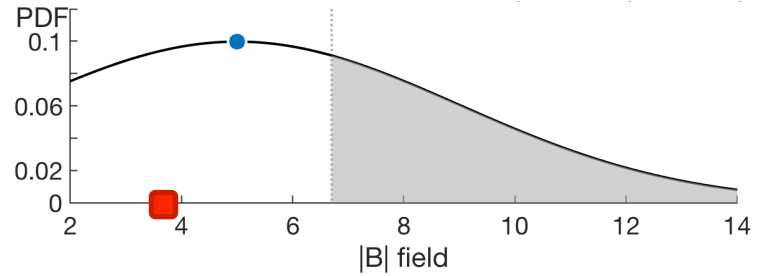
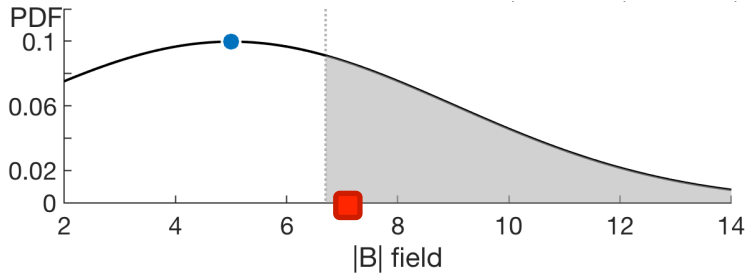
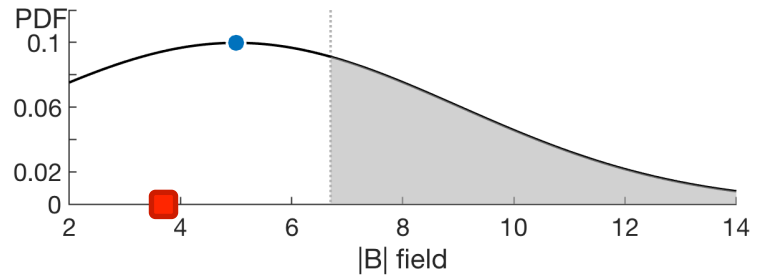
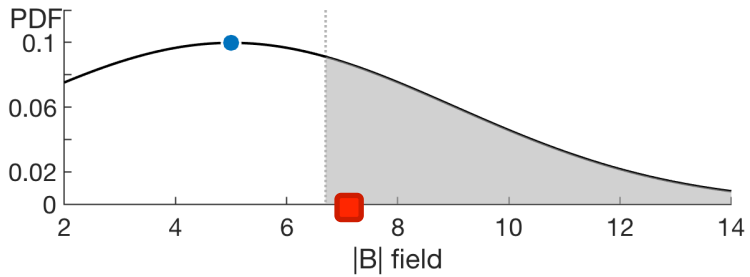
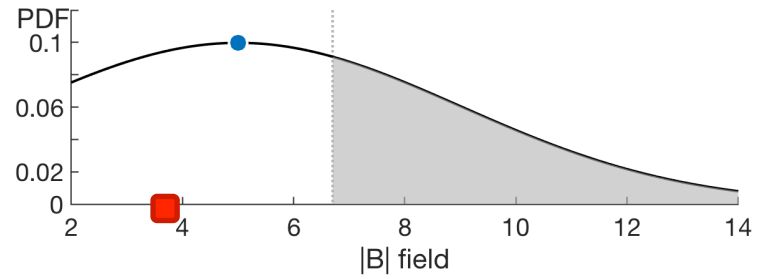
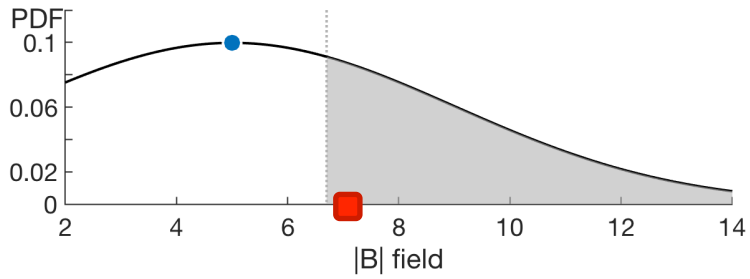
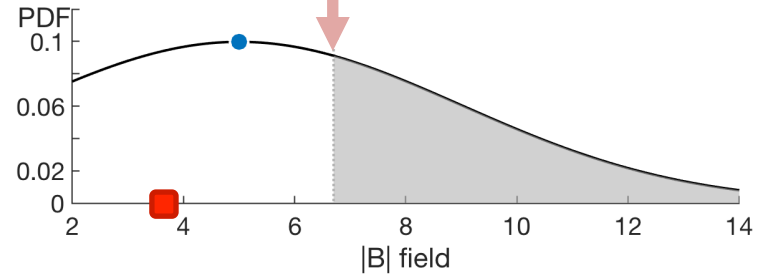
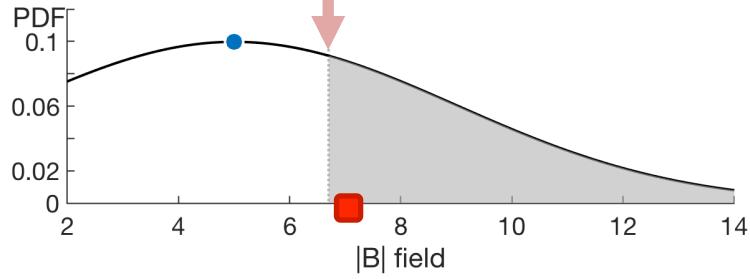
obs = Yes

1. Now consider if the Observation was measured above/below Event Definition (I).

obs = No

Event
Definition, I

Event
Definition, I



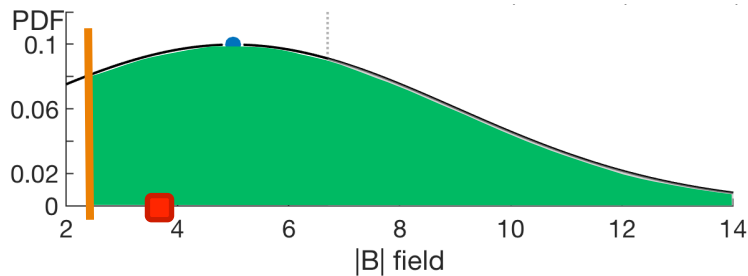
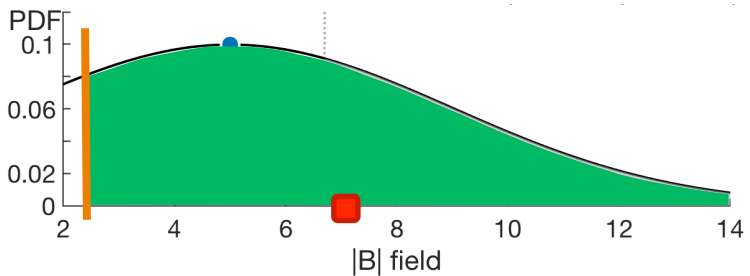
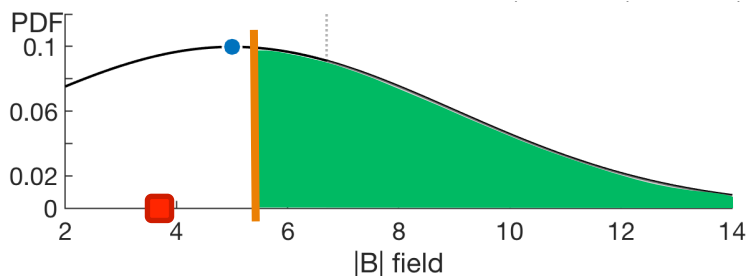
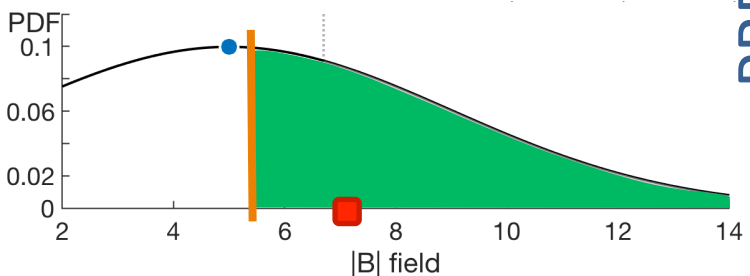
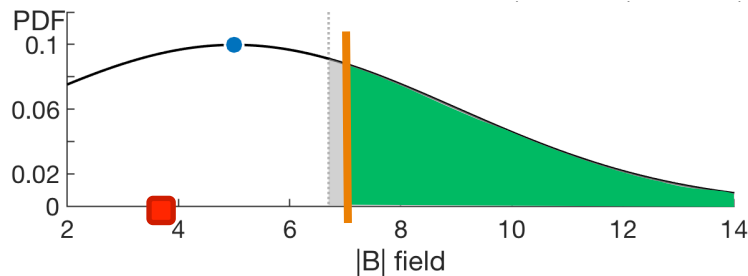
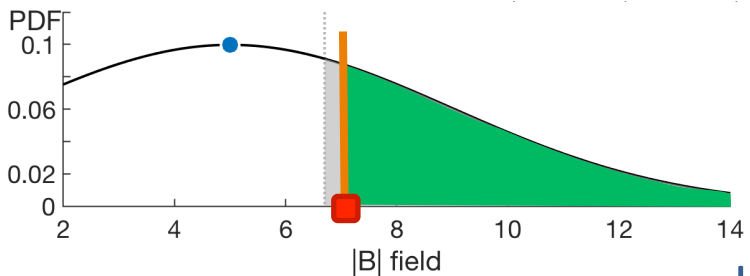
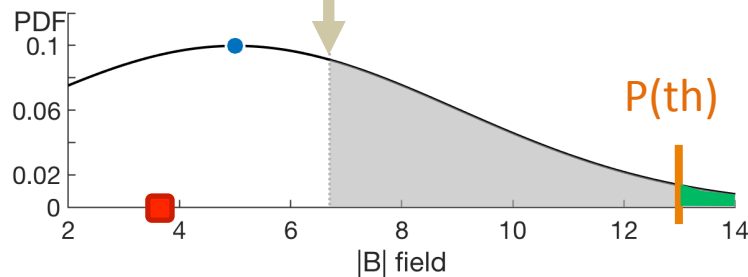
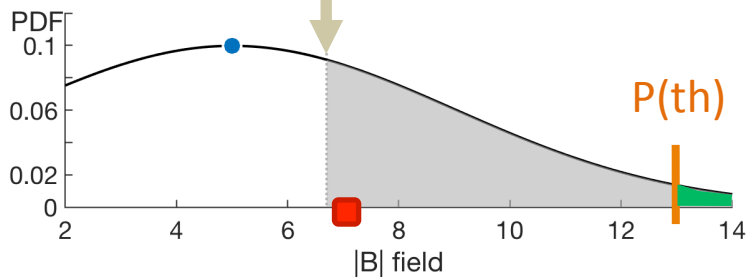
obs = Yes

2. Combine observation value with predicted value to generate the skill [Hit, Miss, etc.]

obs = No

Event Definition, I

Event Definition, I



OBS

PRE

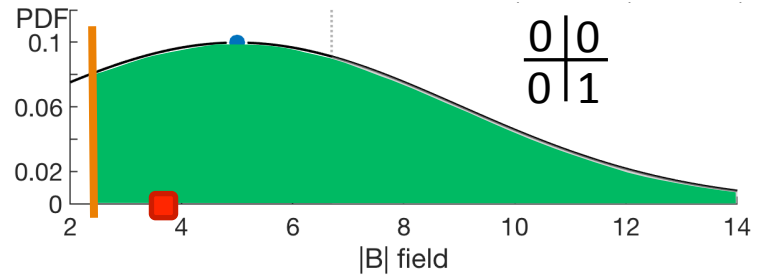
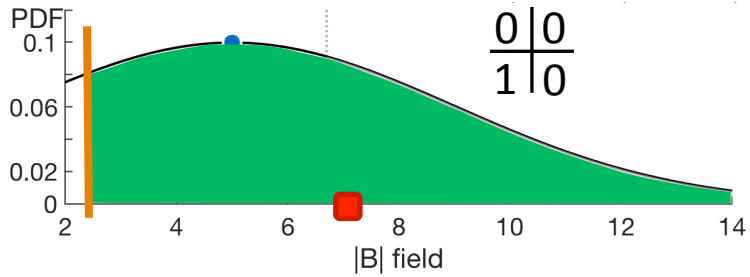
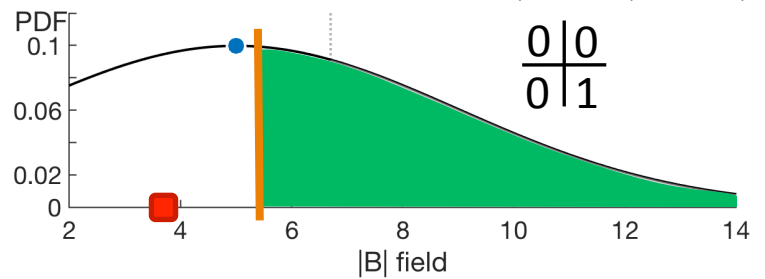
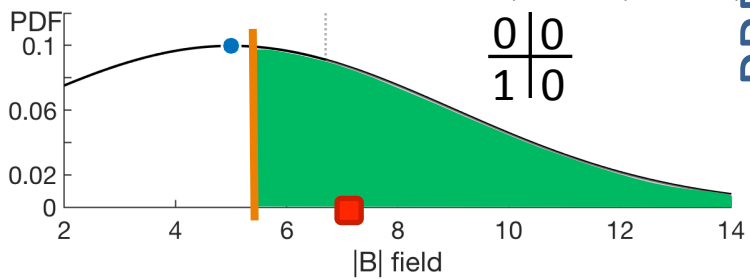
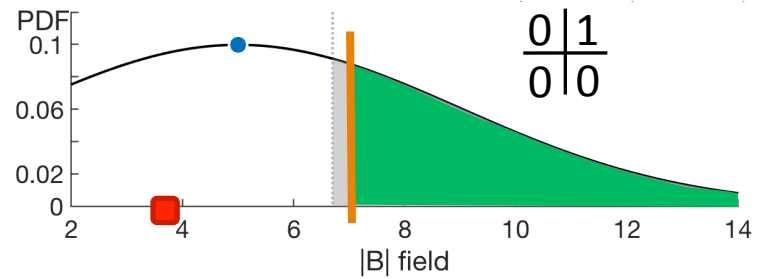
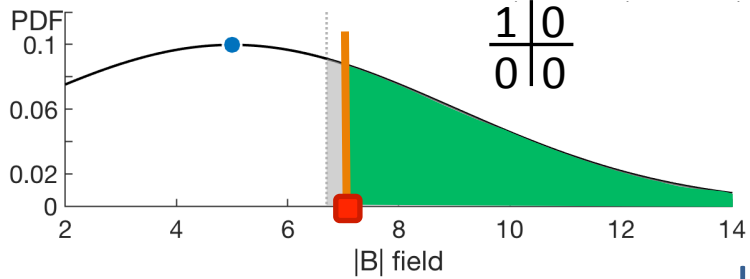
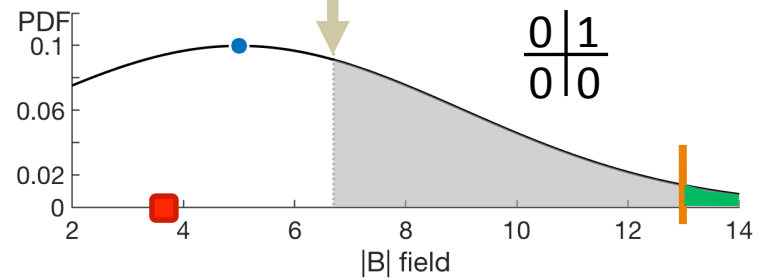
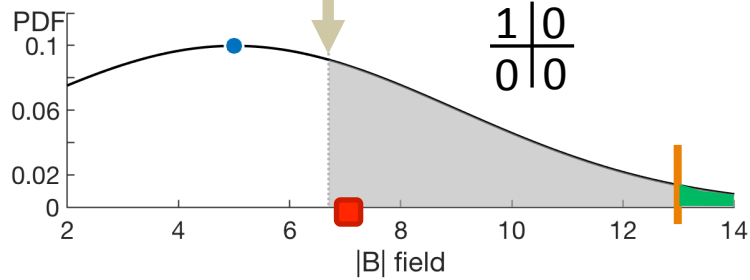
True	False
True	True
False	False

2. Combine observation value with predicted value to generate the skill [Hit, Miss, etc.]

obs = No

Event Definition, I

Event Definition, I



OBS

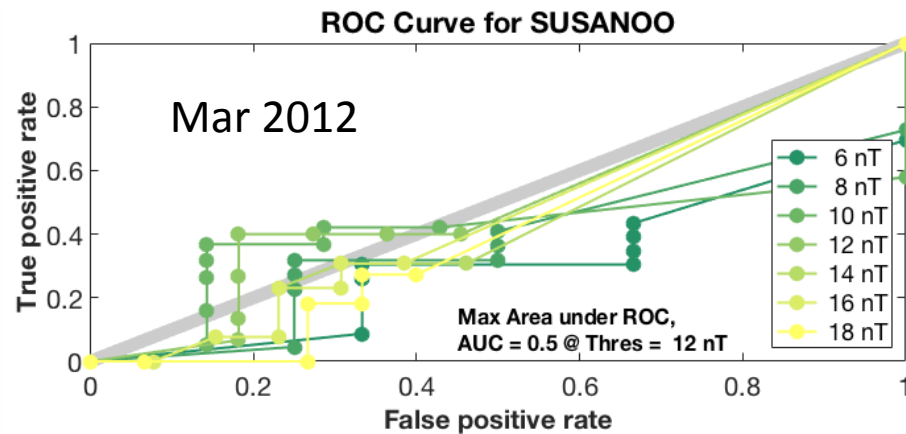
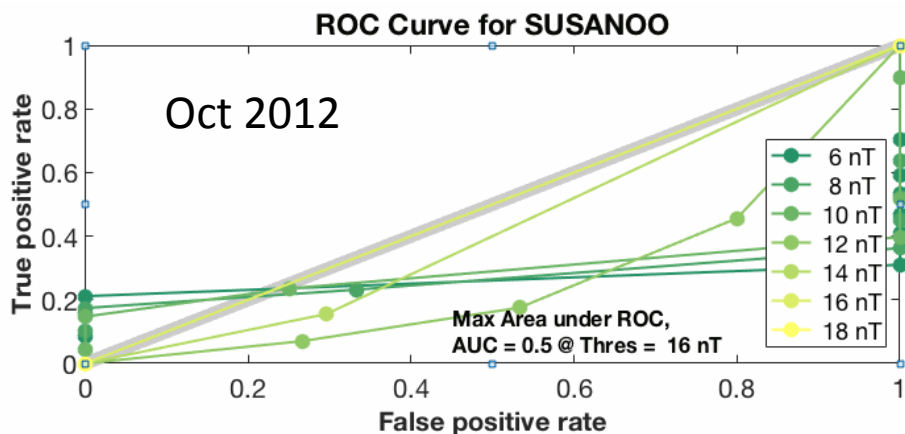
PRE

True	False
True	True
False	False

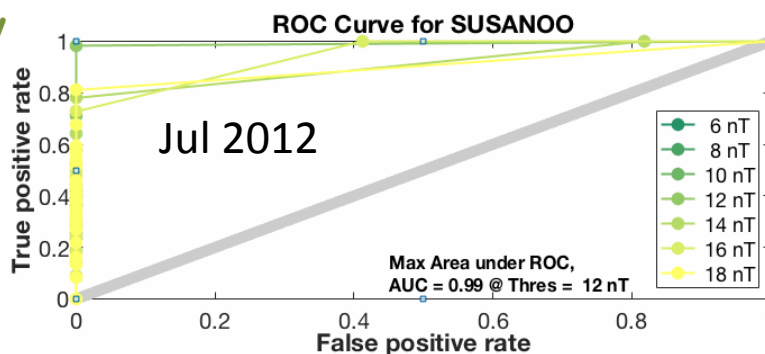
What's New

ROC curve

- Preliminary results using SUSANOO
- Variety of results shown:
 - lack of independence between points is the cause?
 - Period of analysis require more than CME time?



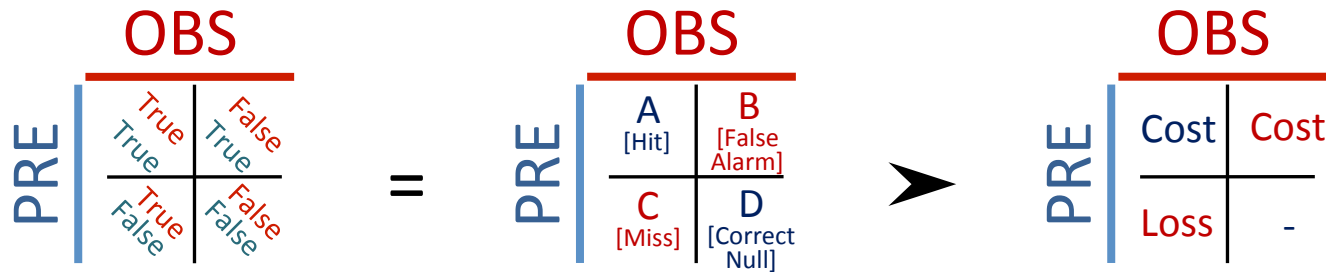
A report of the methodology will be sent to everyone for comment in the coming weeks, along with SWx Journal Special Issue



What's New

ROC curve → Cost Loss Curve

- **Cost/Loss** provides additional insight by enabling a weighting system within the values inside a contingency table.
 - E.g. 'Miss' has more end-user impact than 'False Alarm'
- 'Expense' is effectively the total impact from events exceeding the event definition, I



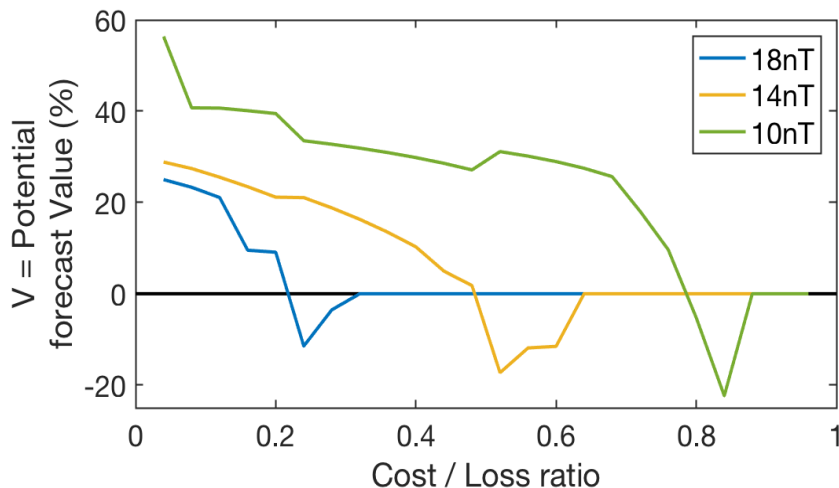
$$\text{Total Cost} = A.Co + B.Co$$

$$E = A.Co + B.Co + C.Lo \quad : \text{Total Expense}$$

$$P(th) = Co/Lo \quad : \text{Cost/Loss ratio}$$

$$E = Lo. [A. P(th) + B. P(th) + C] : \text{Total Expense}$$

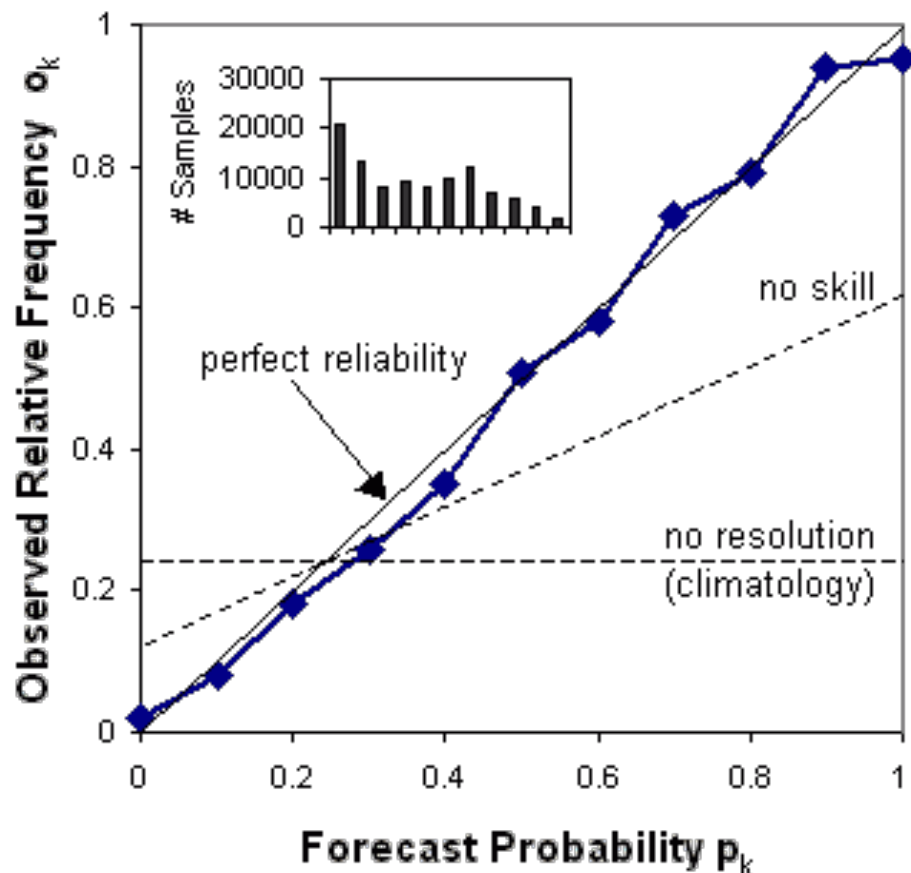
$$V = 100 * (Ec - E) / (Ec - E0) \quad : \text{Potential Value}$$



What's New

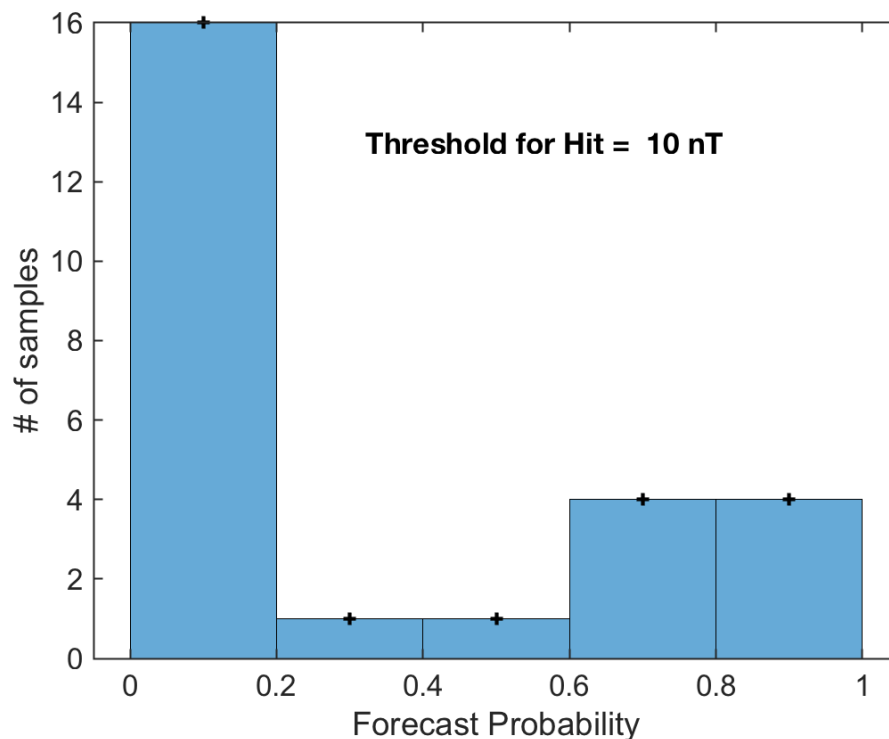
Add Attributes Curve

- The ROC is conditioned on the observations (i.e., given that Y occurred, what was the corresponding forecast?). It is therefore a good companion to the Attribute diagram, which is conditioned on the forecasts.



← Sample reliability Curve.

Statistics too low with a single CME. ↓



Future

Scientific metric of success

- **Currently working towards same metric to compare between models.**
 - ROC curve / (Cost-Loss Analysis)
 - + Attributes (Reliability) Diagram
 - Do scientists want a larger variety?
- **Benefits of these curves:**
 - Condenses several skill metrics to a simple visual – (ideal for scientist comparison)
 - ROC Curve can be converted to a single number (Area under the ROC) – (ideal for end-user comparison)
 - If other newly developed metrics are created, they themselves need a R2O process, similar to the science models.

Future

Scientific metric of success

- **Event list pipeline for an R2O approach.**
 - **Stage 1: 4 events of model benchmarking (code tweaking)**

DATE	GEM	CCMC DONKI						solarmonitor FLARE		
DONKI hyperlink	Chal.	SOURCE	AR	LON	LAT	SPEED	1/2 width	tilt	Class	Source Loc
2010-04-03T09:54	YES	S20E05	11059	8	7	620	26	80	B7.4	S22W03
2011-08-02T06:40	YES	N17W12	11261	15	4	900	35	-35	M1.4	N17W14.6
2012-07-12T16:54	NO	S14W02	11520	6	-13	1300	65	60	X1.4	S14W02
2013-07-09T15:09	NO	N20E15	N/A	-10	2	600	40	-20	PROMINENCE	

- **Stage 2: CCMC-SWPC MOU selection provided by SWPC**
- 33 historical periods with 36 CME input parameters
- Model software must remained fix
- Manual adjustments to the CME parameters are not allowed

Future

Main conclusions

- ROC + Attributes Curve is promising but needs more work for statistics.
- Standardise the CME Event list
- NEW scoreboard for IMF Bz on the CCMC site.

Real-time Forecasting Methods Validation: IMF Bz Scoreboard

CCMC is in the design and implementation phase of the "Bz Scoreboard" together with the international research community. The Bz scoreboard is designed as an automated system to evaluate skills for any predictions of the magnetic characteristics observed at L1.

The scoreboard will provide all international scientists and forecasters a single location where the community can test and prototype a variety of models than span the regime between fully operational to initial research ideas.

IMF Bz Scoreboard planning group:

Leads: Neel Savani (UMBC/NASA GSFC), Pete Riley (Predictive Science)

CCMC Facilitator: Leila Mays (NASA/GSFC)

