ASTR-6000 Seminar
COLLAGE: Coronal Heating, Solar Wind, & Space Weather

March 17, 2022
Space Weather Forecasting: how, and how well, can we predict space weather?

Dr. Thomas Berger
Dr. Steven Cranmer

Outline

1. What is a forecast?
2. Operational space weather forecasting observations and models (US-centric)
3. NOAA space weather forecasts, nowcasts, and reports
4. Forecast evaluation metrics: how well do current forecasts perform?

Review: Classifying space weather phenomena & impacts

NOAA/SWPC scales

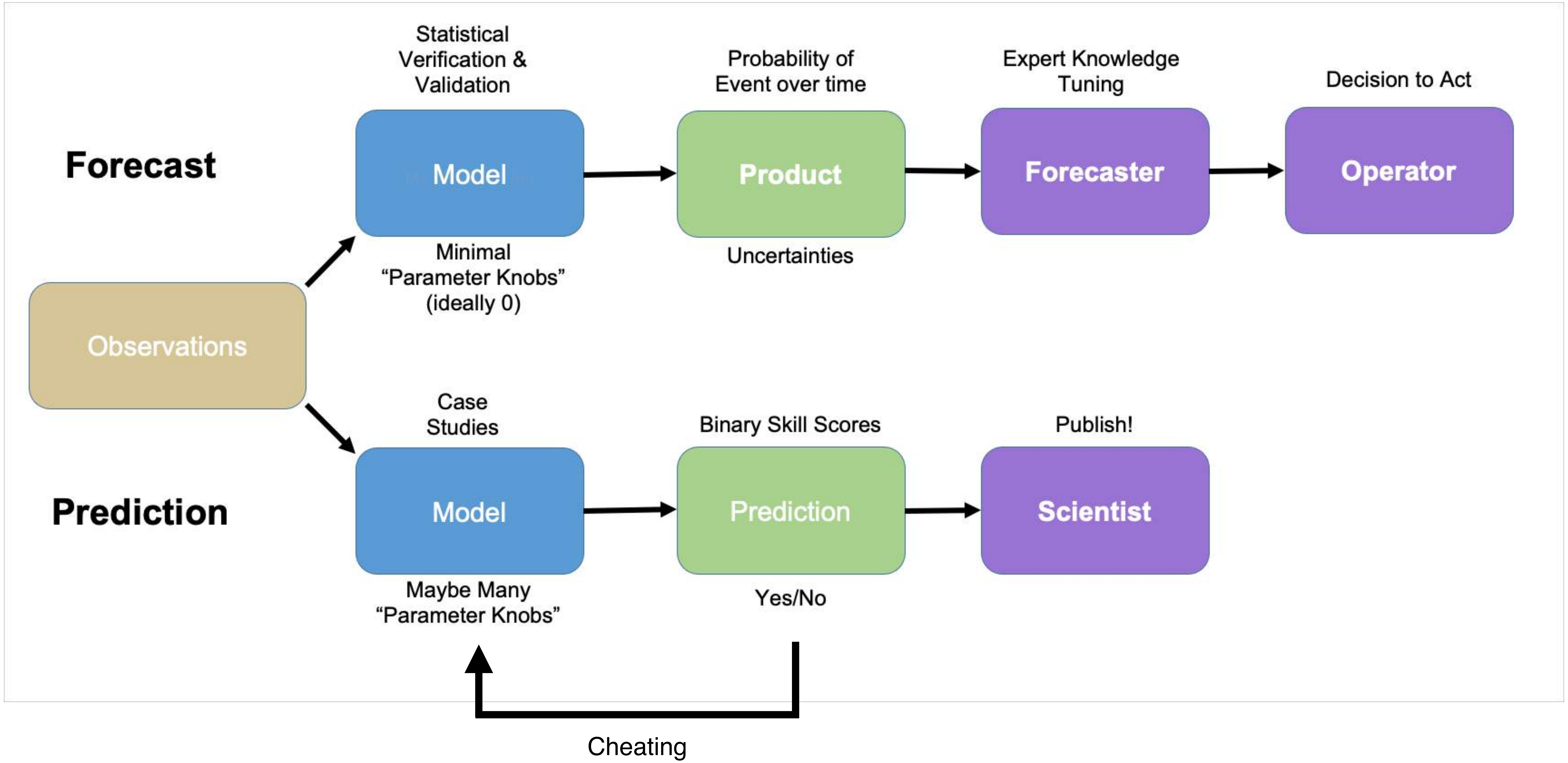
Scale	Description	Effect	Physical measure	Average Frequency (1 cycle = 11 years)
R 5	Extreme	**HF Radio:** Complete HF (high frequency) radio blackout on the entire sunlit side of the Earth lasting for a number of hours. This results in no HF radio contact with mariners and en route aviators in this sector. **Navigation:** Low-frequency navigation signals used by maritime and general aviation systems experience outages on the sunlit side of the Earth for many hours, causing loss in positioning. Increased satellite navigation errors in positi	X20 (2 x 10-3)	Less than 1 per cycle
R 4	Severe	**HF Radio:** HF contact lost du **Navigation:** (hours. Minor d		
R 3	Strong	**HF Radio:** Wi of Earth. **Navigation:** L		
R 2	Moderate	**HF Radio:** Lin **Navigation:** [		
R 1	Minor	**HF Radio:** We **Navigation:** L		

Scale	Description	Effect	Physical measure (Flux level of >= 10 MeV particles)	Average Frequency (1 cycle = 11 years)
S 5	Extreme	**Biological:** Unavoidable high radiation hazard to astronauts on EVA (extra-vehicular activity); passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** Satellites may be rendered useless, memory impacts can cause loss of control, may cause serious noise in image data, star-trackers may be unable to locate sources; permanent damage to solar panels possible. **Other systems:** Co and position errors n	10^5	Fewer than 1 per cycle
S 4	Severe	**Biological:** Unavoid high latitudes may b **Satellite operation** problems may cause **Other systems:** Bla errors over several d		
S 3	Strong	**Biological:** Radiatio flying aircraft at high **Satellite operation** panel are likely. **Other systems:** De		
S 2	Moderate	**Biological:** Passeng risk. **Satellite operation** **Other systems:** Sm locations possibly aff		
S 1	Minor	**Biological:** None. **Satellite operation** **Other systems:** Mir		

Scale	Description	Effect	Physical measure	Average Frequency (1 cycle = 11 years)
G 5	Extreme	**Power systems:** Widespread voltage control problems and protective system problems can occur, some grid systems may experience complete collapse or blackouts. Transformers may experience damage. **Spacecraft operations:** May experience extensive surface charging, problems with orientation, uplink/downlink and tracking satellites. **Other systems:** Pipeline currents can reach hundreds of amps, HF (high frequency) radio propagation may be impossible in many areas for one to two days, satellite navigation may be degraded for days, low-frequency radio navigation can be out for hours, and aurora has been seen as low as Florida and southern Texas (typically 40° geomagnetic lat.).	Kp = 9	4 per cycle (4 days per cycle)
G 4	Severe	**Power systems:** Possible widespread voltage control problems and some protective systems will mistakenly trip out key assets from the grid. **Spacecraft operations:** May experience surface charging and tracking problems, corrections may be needed for orientation problems. **Other systems:** Induced pipeline currents affect preventive measures, HF radio propagation sporadic, satellite navigation degraded for hours, low-frequency radio navigation disrupted, and aurora has been seen as low as Alabama and northern California (typically 45° geomagnetic lat.).	Kp = 8, including a 9-	100 per cycle (60 days per cycle)
G 3	Strong	**Power systems:** Voltage corrections may be required, false alarms triggered on some protection devices. **Spacecraft operations:** Surface charging may occur on satellite components, drag may increase on low-Earth-orbit satellites, and corrections may be needed for orientation problems. **Other systems:** Intermittent satellite navigation and low-frequency radio navigation problems may occur, HF radio may be intermittent, and aurora has been seen as low as Illinois and Oregon (typically 50° geomagnetic lat.).	Kp = 7	200 per cycle (130 days per cycle)
G 2	Moderate	**Power systems:** High-latitude power systems may experience voltage alarms, long-duration storms may cause transformer damage. **Spacecraft operations:** Corrective actions to orientation may be required by ground control; possible changes in drag affect orbit predictions. **Other systems:** HF radio propagation can fade at higher latitudes, and aurora has been seen as low as New York and Idaho (typically 55° geomagnetic lat.).	Kp = 6	600 per cycle (360 days per cycle)
G 1	Minor	**Power systems:** Weak power grid fluctuations can occur. **Spacecraft operations:** Minor impact on satellite operations possible. **Other systems:** Migratory animals are affected at this and higher levels; aurora is commonly visible at high latitudes (northern Michigan and Maine).	Kp = 5	1700 per cycle (900 days per cycle)

What is a Forecast?

Hint: a forecast is not just a prediction…



The ART of forecasting

Accurate + Reliable + Timely

“Forecasting is a necessary but not sufficient condition for success.”
Sir Mark Walport, UK Chief Science Advisor

To be useful to anyone, a forecast must be

Accurate

Definition depends on application but is generally based on

Time - “how close can you get on arrival time?”

Magnitude - “how strong will it be?”

Reliable

Definition depends on application but is generally based on **consistency over time** and **low False Alarm Rate.**

Can systems operators take actions based on well-tested justifications of performance?

Timely

Definition is generally independent of application and is based on **time to deliver forecast** relative to time to impact.

Usually this is “As soon as possible” - ASAP.

“There is no value in a forecast. There is only value in how a forecast is used.”
Tim Palmer, Royal Society Research Professor, Oxford

Products: how a forecast gets communicated

Watch ⟶ Warning ⟶ Alert

Watch: "Something has been detected *or modeled* and *may* or *may not* cause an event."

- Generally issued on the basis of an observation that is consistently known to cause events, e.g., a CME leaving the Sun in the direction of Earth, or a large coronal hole rotating into the Sun-Earth line.
- Note that a Watch is not a definitive prediction of occurrence - it is only stating a *possibility* of occurrence. Threshold for issuance is subjective, e.g., forecaster judges CME is Earth-directed from preliminary observations.

Warning: "Something has been detected or predicted and *will very likely* cause an event."

- Generally issued on the detection of an event at an upstream location, e.g. the detection of a shock wave at the L1 Lagrangian point.
- A Warning usually comes with a forecasted magnitude, e.g., "G3 Warning", but is often updated as conditions/measurements change.

Alert: "An event is in progress."

- Based on measured levels of activity at the location of interest, e.g. ground-based magnetometers on Earth.
- An alert is the initial/provisional statement of the timing and magnitude of an event. May be refined after the fact.

Current space weather Watch, Warning, Alert capabilities

An obvious lack of magnetic eruption products

Event	Watch	Warning	Alert
Eruption ("Flare")			✱[1]
Radiation Storm		✱[2]	✱[3]
CME Geomagnetic Storm	✱[4]	✱[5]	✱[6]

1. Based on passing M 1.0-level X-ray threshold in GOES XRS instrument.
2. 15—30 minutes based on flare magnitude and location.
3. Based on passing 10 MeV proton threshold in GOES SEISS instruments.
4. ±10 hours accuracy on CME arrival time, but this is not issued with the Watch. HSS and CIR events are not issued Watch products, only Warning on solar wind speed increase at DSCOVR at L1.
5. 15—45 minutes based on CME or HSS or CIR detection at DSCOVR at L1.
6. Based on detection of magnetic anomaly in USGS and Canadian ground-based magnetometer network.

Solar Magnetic Eruptions: Earthquake or Volcano?

Precursors are everything

Solar Magnetic Eruptions: Earthquake or Volcano?

Precursors are everything

Active region magnetic field "turn over time scale" ~ 30 hours

Contrast: what is a “nowcast”?

Nowcast is what is happening at the moment of product output

Nowcast = specification of *current conditions* relevant to a particular operation or event.

Examples:

- 10 MeV proton flux at GEO during a Solar Energetic Particle event.
- 1—8 Å X-ray irradiance during a solar flare event.
- Rate of change of TEC index over a geographical location.
- “Real-time” Kp index calculated from a magnetometer network.

Requirement: **Low-latency, “real-time”, observations**.

Latency requirements vary by mission, but are typically on the order of ***seconds or minutes***.
Example: GOES XRS (X-ray irradiance) and SEISS (proton flux) latency = ***3 seconds*** from ground-station to NOAA/ SWCP forecast office.

Related: **All Clear** announcement = official statement of event termination and return of safe conditions.

Does not currently exist. Many people think that this should be a goal of space weather forecasting offices. However, legal liability is a major issue and it is unlikely that any forecasting office will develop an official product of this type.

The Value of Nowcasting

Recognition of solar radio burst averts WW III

23 May 1967 Solar Radio Flare

Signal was originally interpreted as Russian jamming prior to a nuclear attack

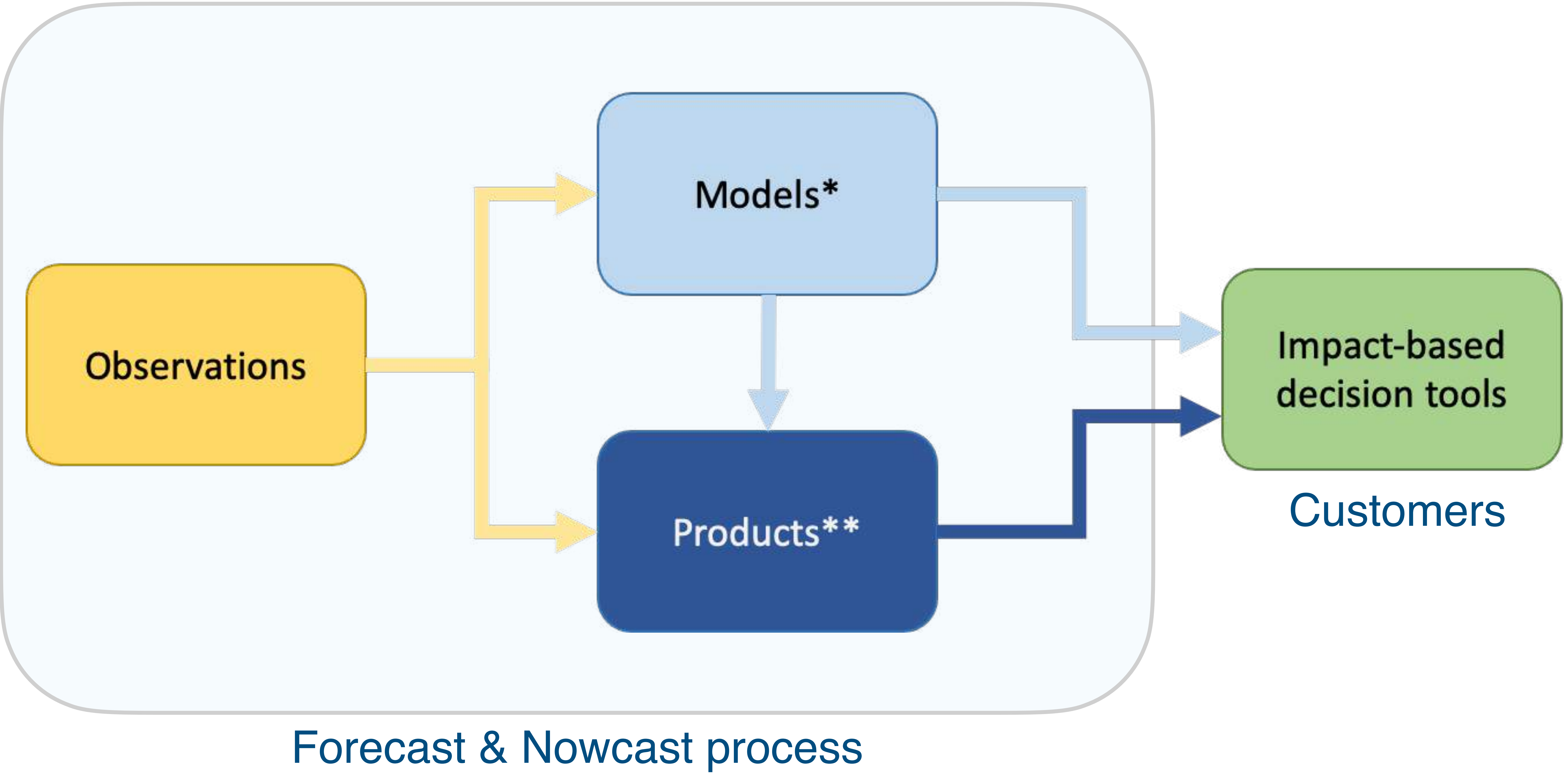
Ballistic Missile Early Warning System (BMEWS)

Over-the-Horizon radar system in Alaska

Sun was low in the Eastern sky at time of radio burst

Knipp, D.J. *et al.* (2016) *Space Weather*, 14(9), pp. 614–633. doi:10.1002/2016SW001423.

Space weather forecasting information flow

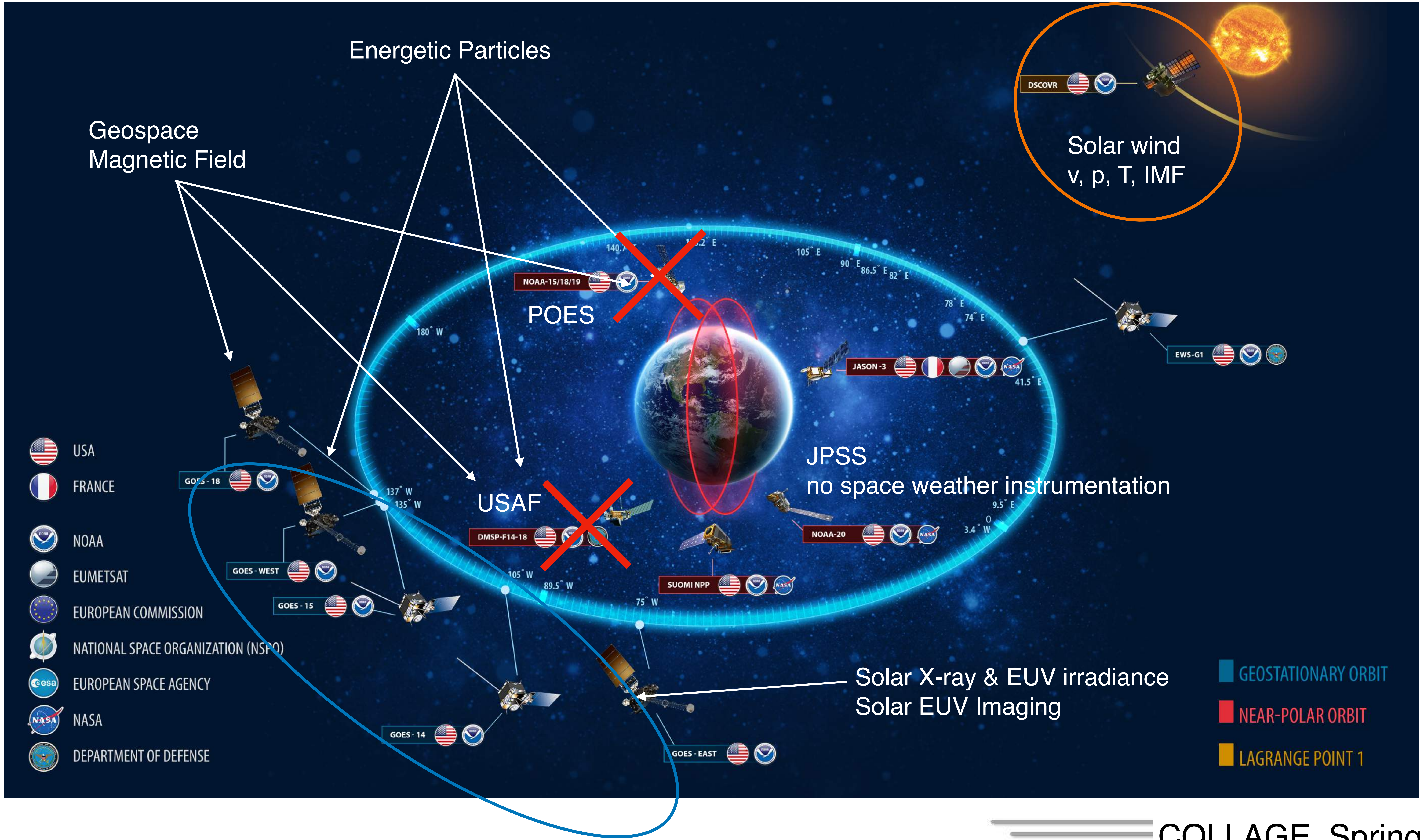


* Complex, multi-input, empirical, physics-based, or machine learning environmental prediction model
** Watches, Warnings, Alerts, nowcasting data such as real-time Kp index, etc.

Knipp, D.J. *et al.* (2016) *Space Weather*, 14(9), pp. 614–633. doi:10.1002/2016SW001423.

NOAA operational space weather observations

Civil space weather system



https://www.nesdis.noaa.gov/s3/2022-03/NOAASatelliteSystem_2022.03.14.png

USSF operational space weather observations

Military space weather system - formerly USAF

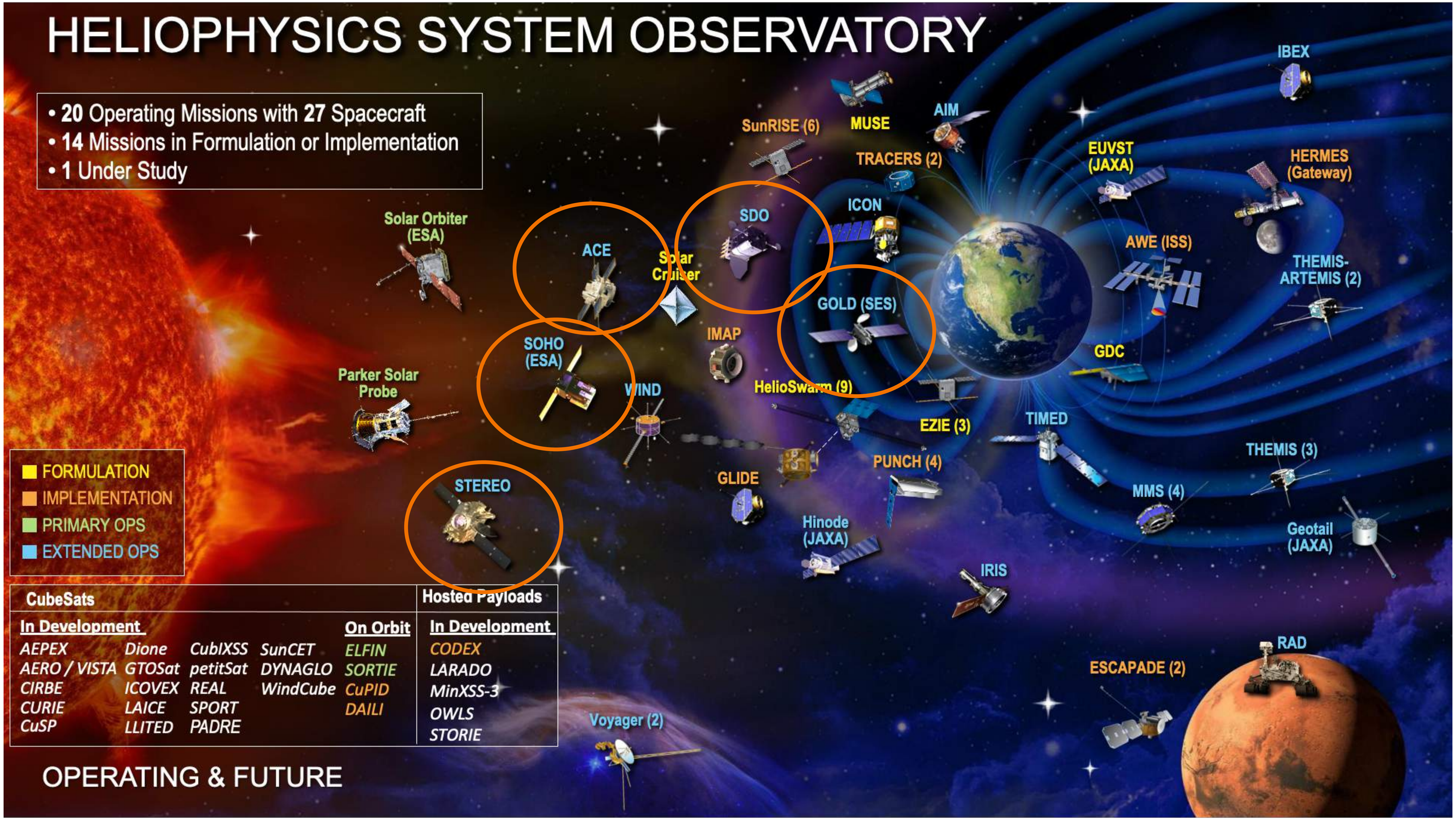
- **DMSP polar-orbiting weather satellites (see previous slide)**
- **HASDM Calibration Objects for satellite drag modeling**
  - Not really an “observation system”
  - high repeat-rate tracking of known/steady Ballistic Coefficient satellites/objects.
- **Solar Optical Observation Network**
  - Hα telescope
- **Solar and Electro-Optic Network**
  - Ha full-Sun imaging
  - Solar radio monitoring
- **SCINDA Network**
  - GPS scintillation monitoring

Solar Optical Observation Network (SOON)
Hα Telescope
Kirtland AFB, NM

Solar & Electro-Optic Network (SEON)
Radio telescope
Sagamore Hill, MA

NASA missions that contribute space weather observations

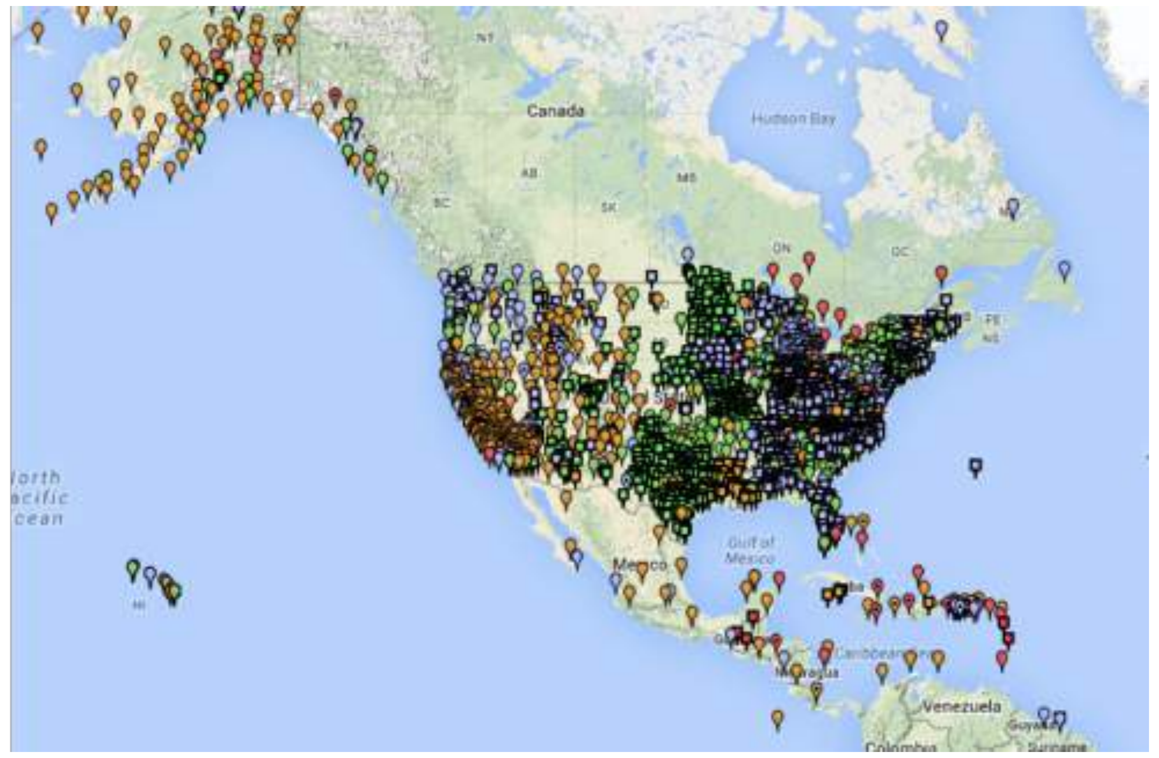
Research satellites providing space weather data used in operations



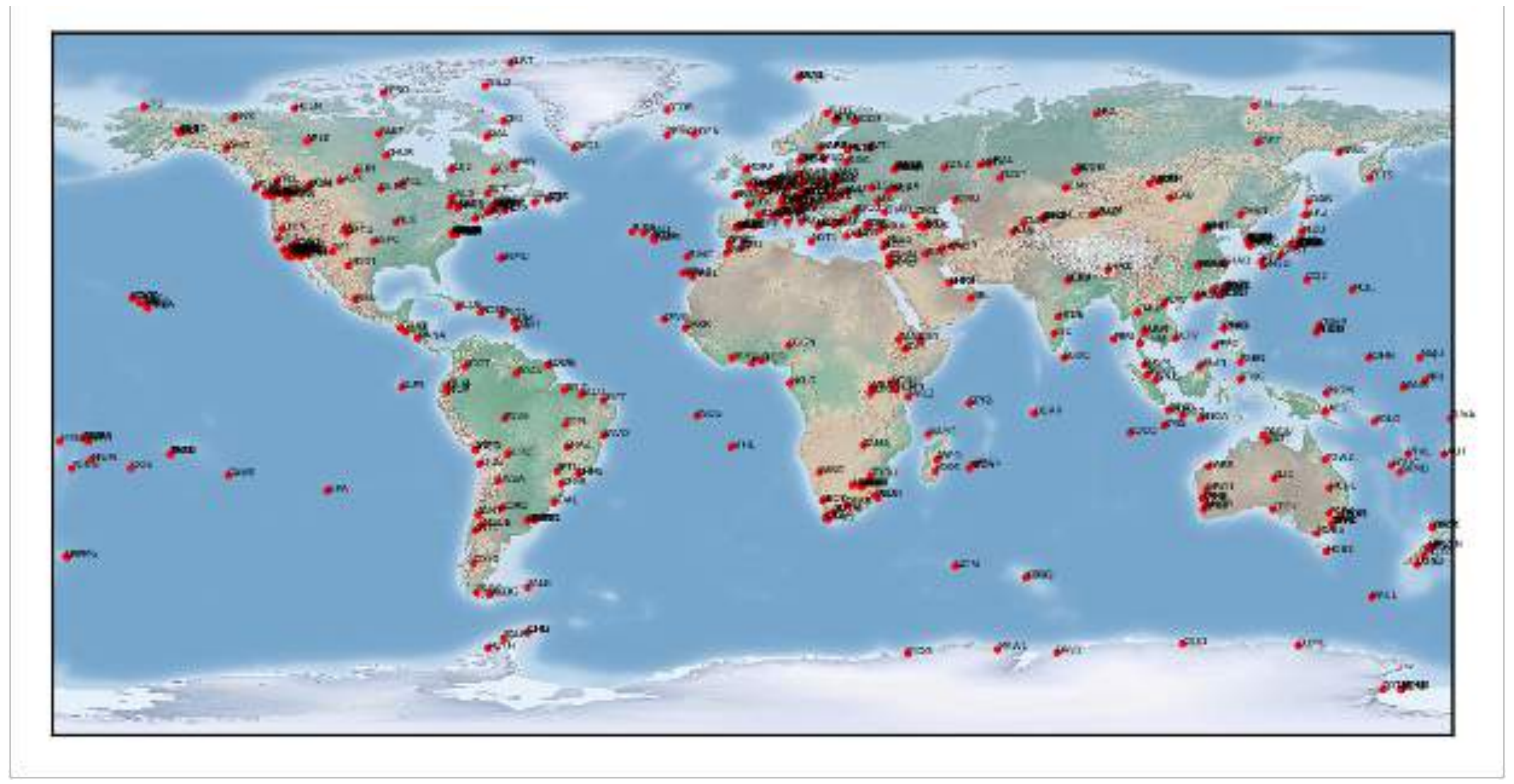
https://science.nasa.gov/heliophysics/mission-fleet-diagram

Civil ground-based space weather measurements

USGS Magnetometer Network
Geomagnetic storm data

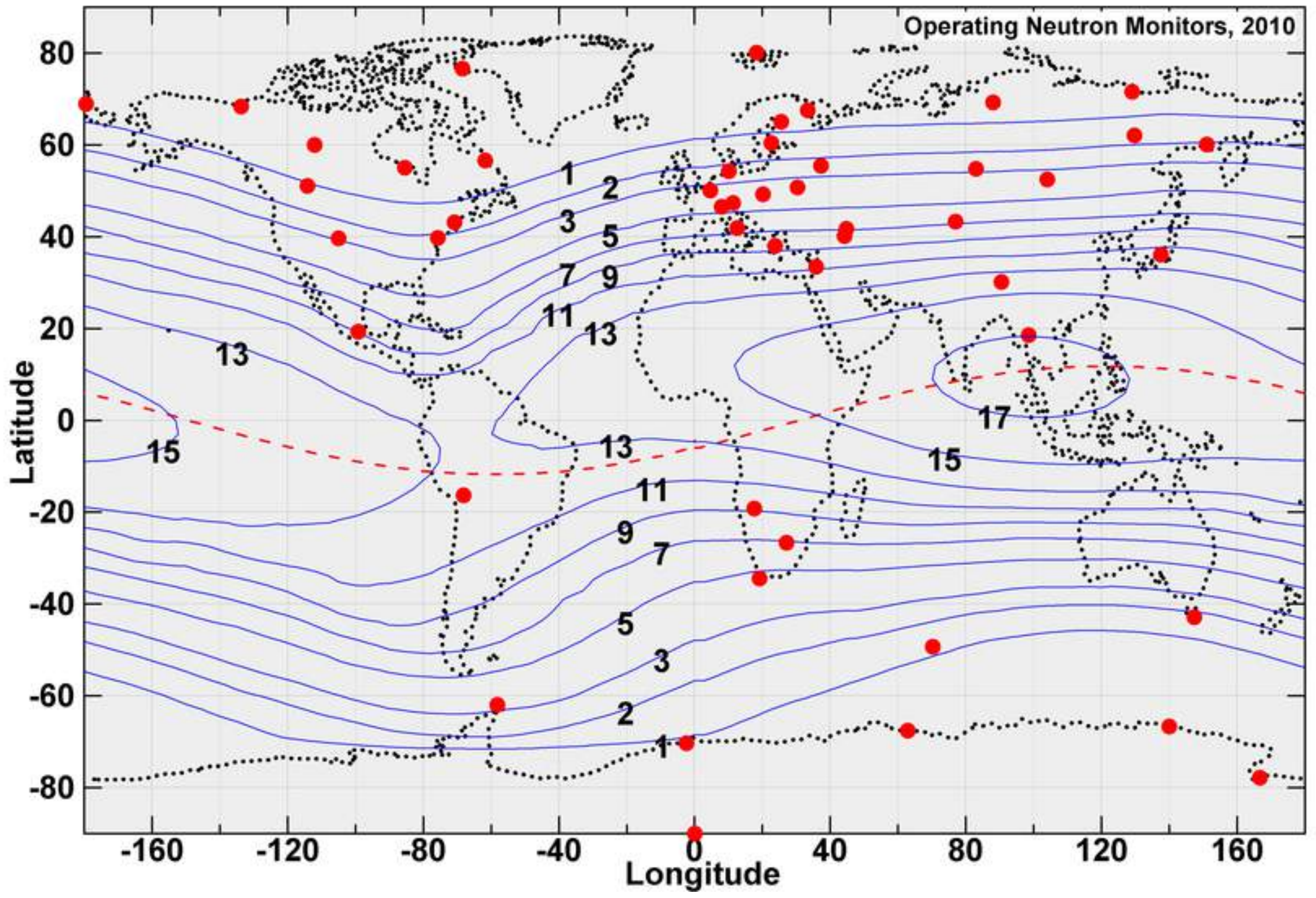


USCG CORS GPS Network
Ionospheric TEC data



IGS RTIG GPS Network
Ionospheric TEC data

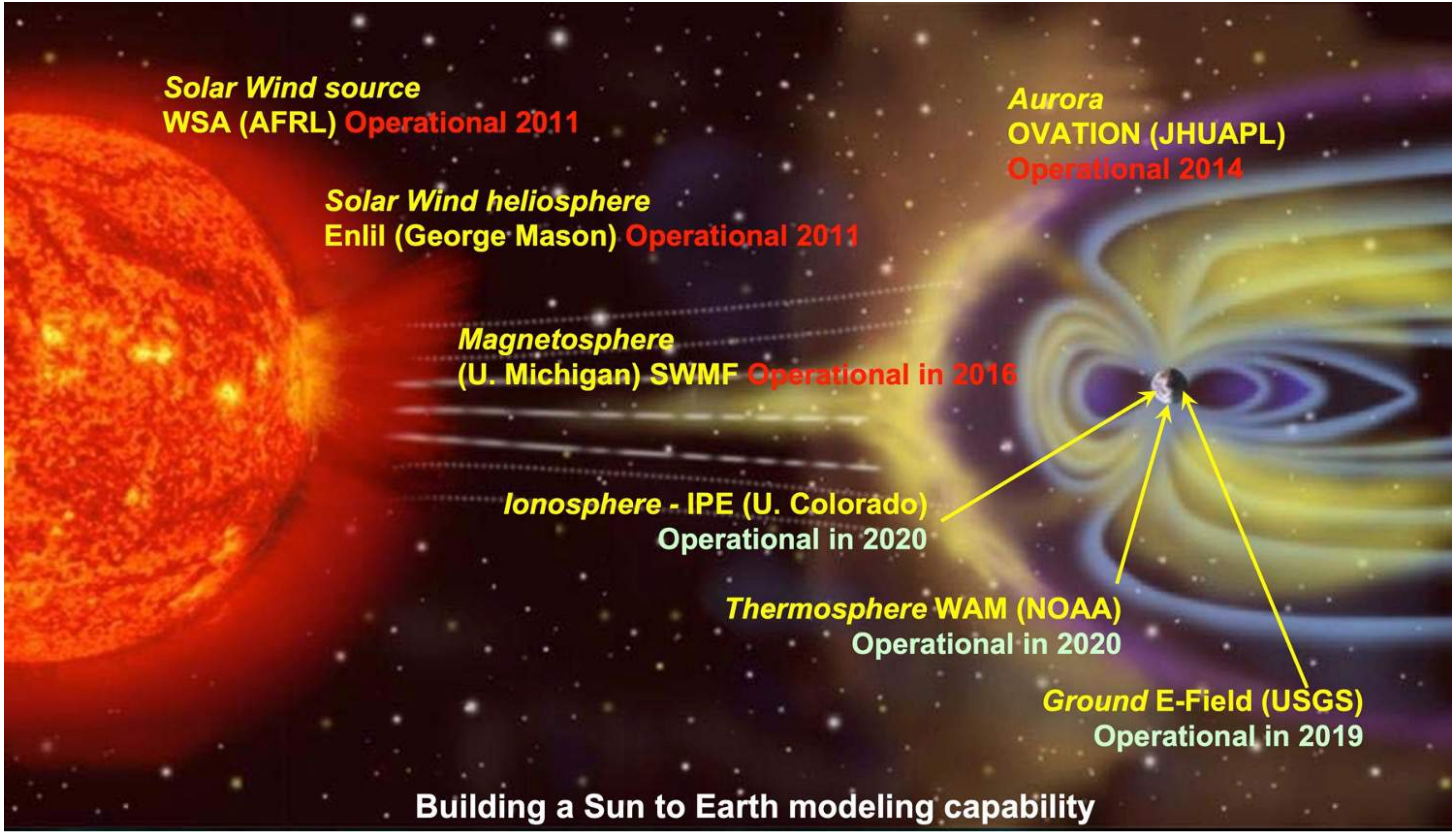
NSF Global Oscillations Network Group (GONG)
Solar magnetograms and Hα images



Neutron monitor network
SEP event aviation radiation dose calibration

Civil operational space weather models/products

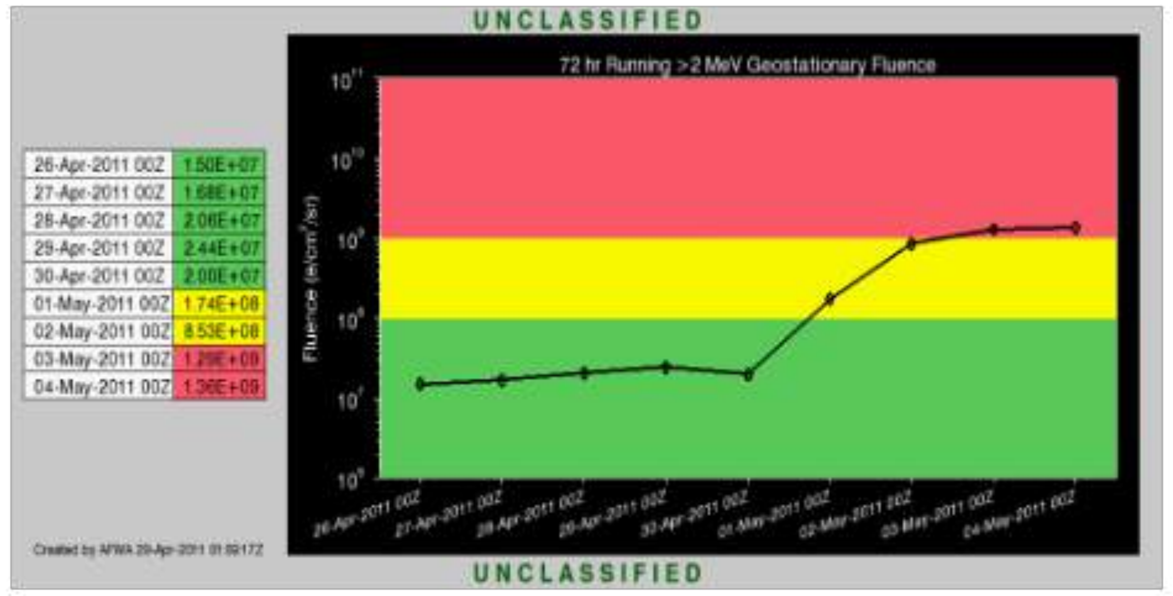
'Sun to mud' approach built from research heritage



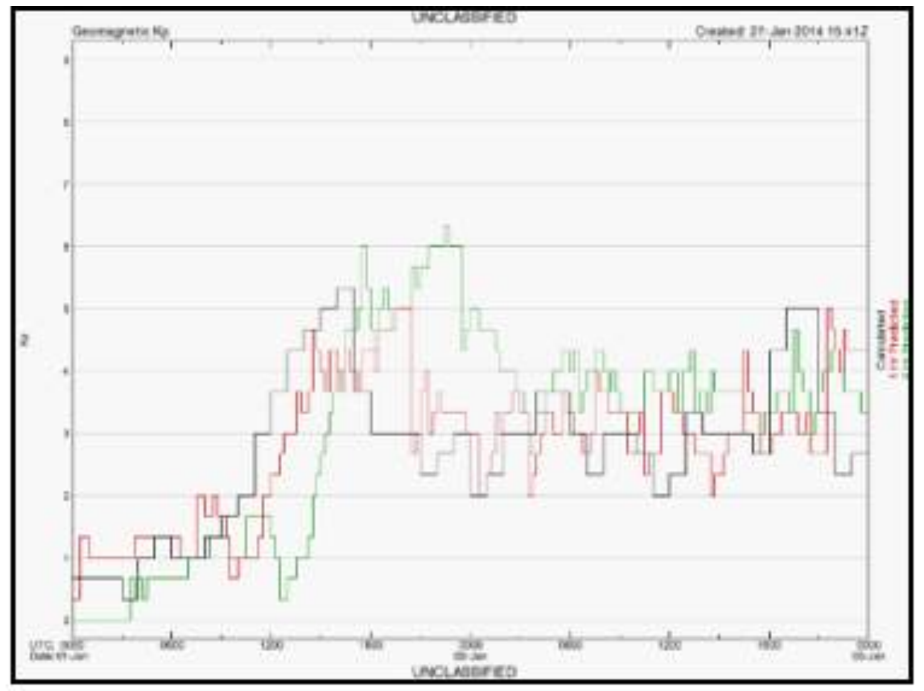
Adapted from E. Talaat, NAS SWx Infrastructure Workshop I, 16-June-2020

DOD operational space weather models/products

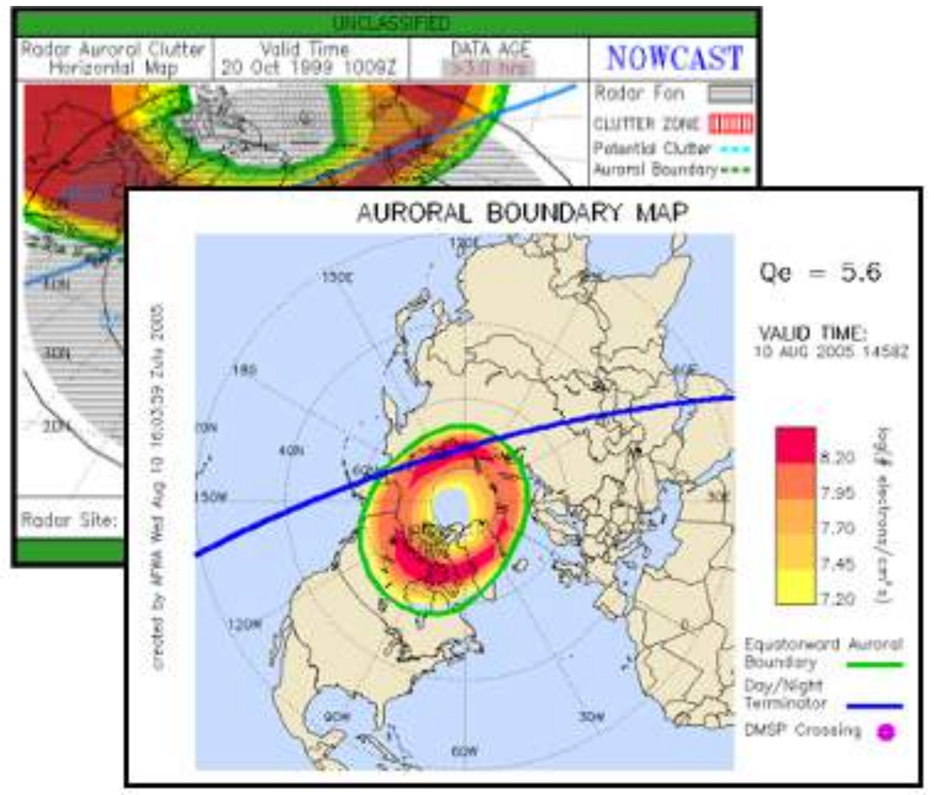
Concentration on communications and orbital systems



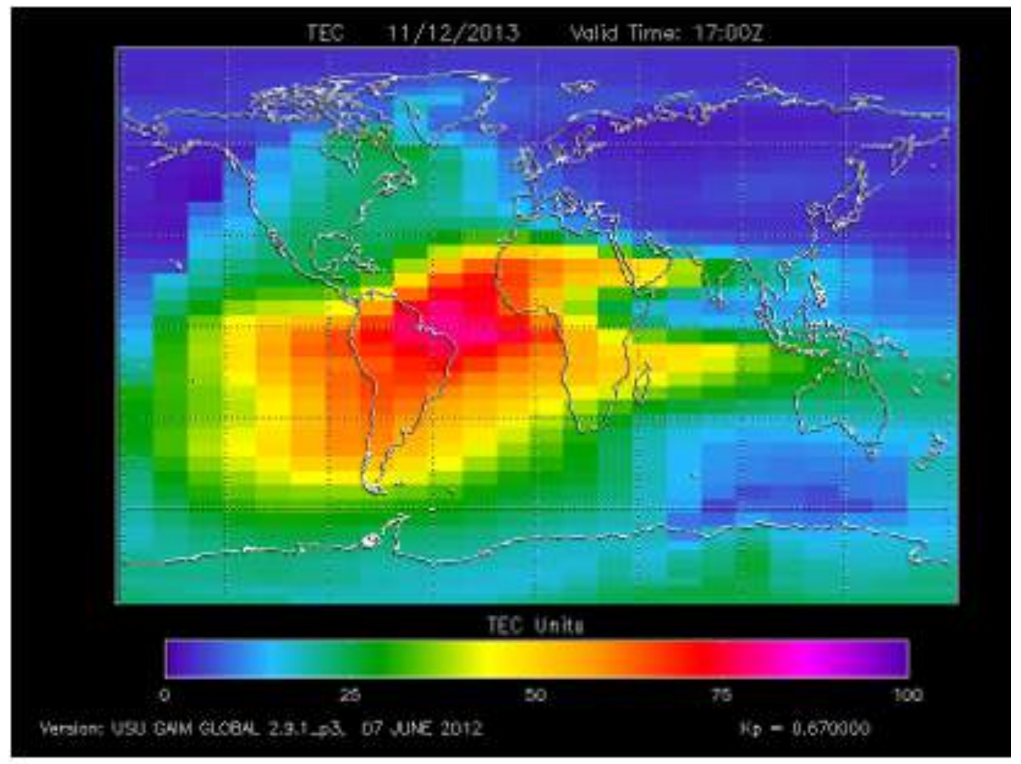
Magnetosphere – GEO radiation hazard



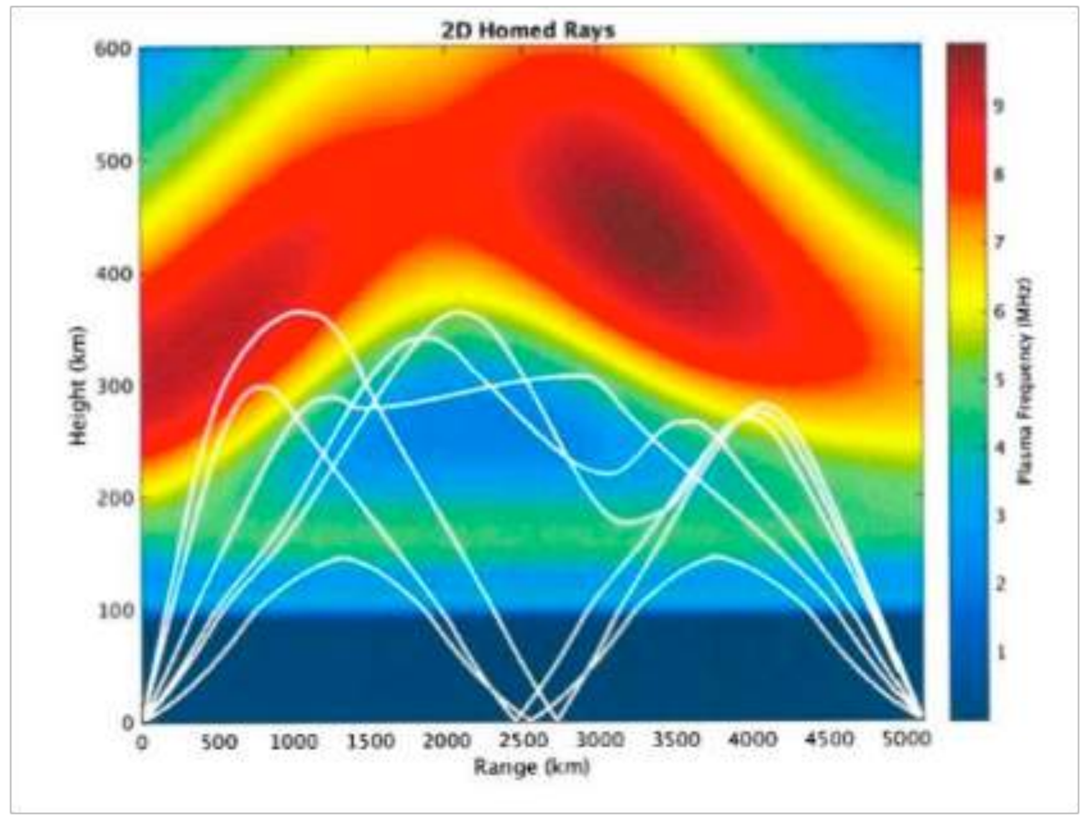
Magnetosphere – Wing Kp model



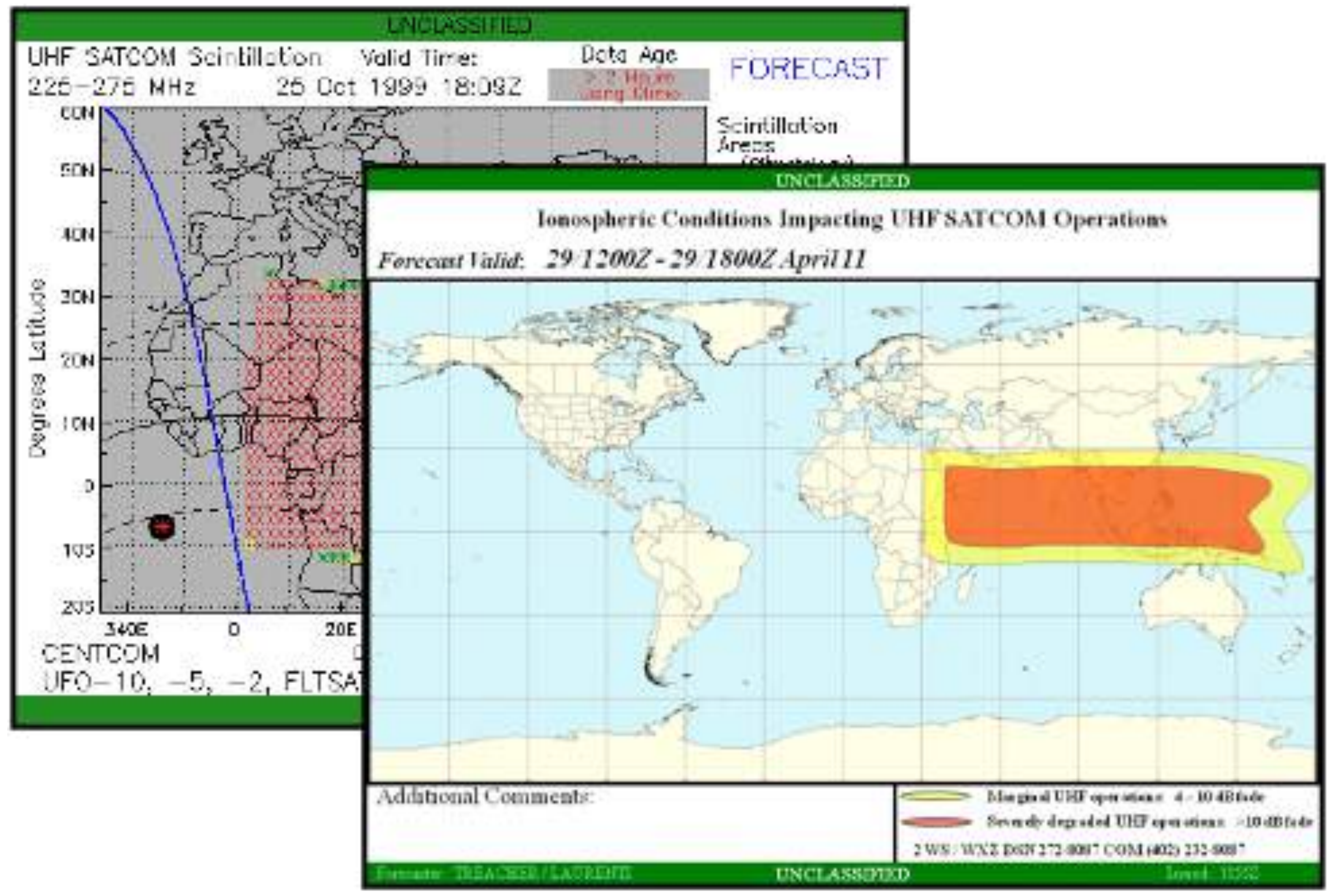
Magnetosphere – Auroral location



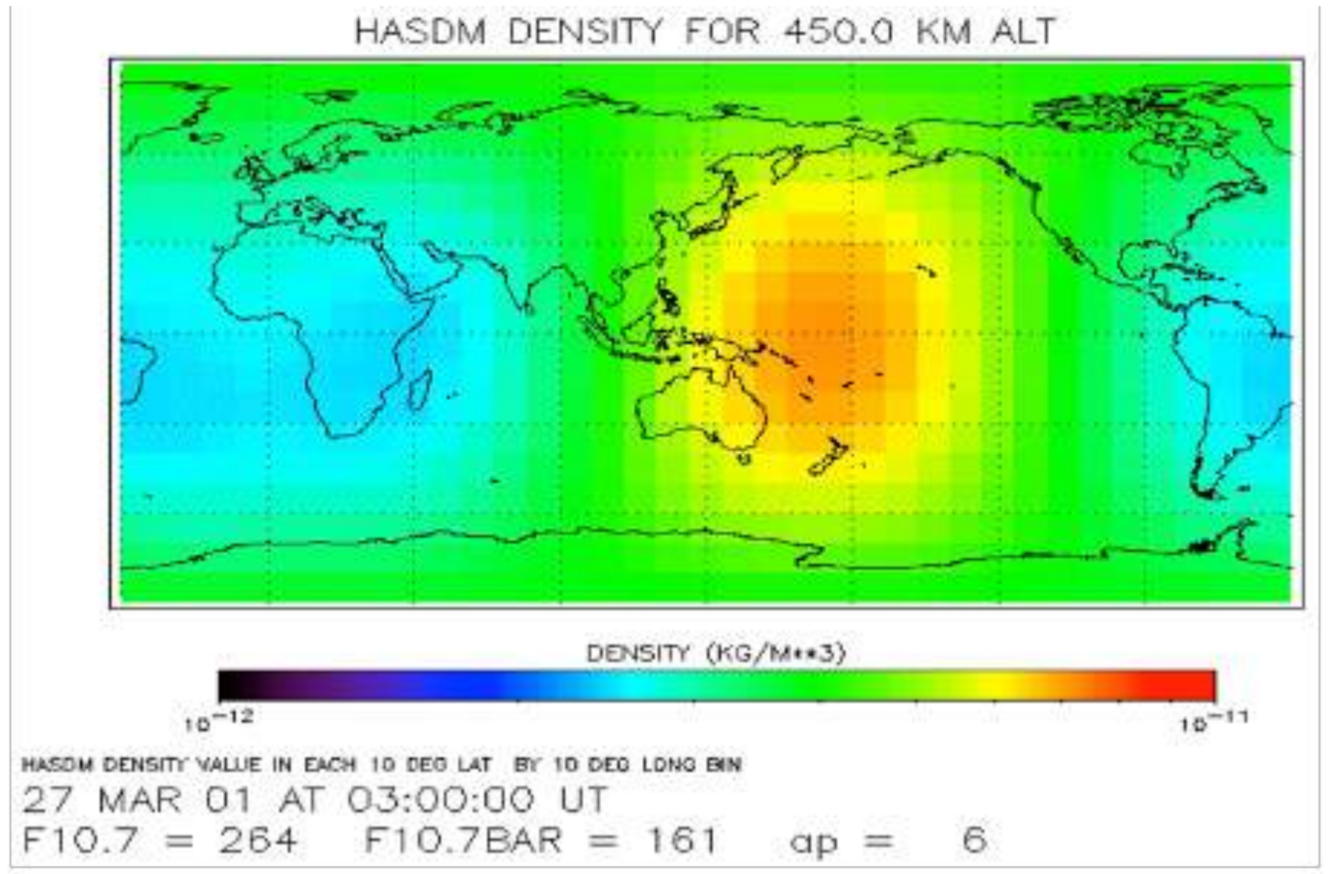
Ionosphere – GAIM model



Ionosphere – HF propagation tool



Ionosphere – scintillation alerts



Thermosphere – HASDM model

Official NOAA space weather forecasts & reports

Best summary product: Report & Forecast of Solar & Geophysical Activity

Forecasts

27-Day Outlook of 10.7 cm Radio Flux and Geomagnetic Indices
3-Day Forecast
3-Day Geomagnetic Forecast
Forecast Discussion
Predicted Sunspot Numbers and Radio Flux
Report and Forecast of Solar and Geophysical Activity
Solar Cycle Progression
Space Weather Advisory Outlook
USAF 45-Day Ap and F10.7cm Flux Forecast
Weekly Highlights and 27-Day Forecast

Reports

Forecast Verification
Geoalert - Alerts, Analysis and Forecast Codes
Geophysical Alert
Solar and Geophysical Event Reports
USAF Magnetometer Analysis Report

https://www.swpc.noaa.gov/products-and-data

Official NOAA space weather forecasts & reports

Best summary product: Report & Forecast of Solar & Geophysical Activity

```
:Product: 0830RSGA.txt
:Issued: 2021 Aug 30 2200 UTC
# Prepared jointly by the U.S. Dept. of Commerce, NOAA,
# Space Weather Prediction Center and the U.S. Air Force.
#
Joint USAF/NOAA Solar Geophysical Activity Report and Forecast
SDF Number 242 Issued at 2200Z on 30 Aug 2021
```

IA. Analysis of Solar Active Regions and Activity from 29/2100Z to 30/2100Z: Solar activity has been at low levels for the past 24 hours. The largest solar event of the period was a C1 event observed at 30/0154Z from Region 2860 (S29W36). There are currently 2 numbered sunspot regions on the disk.

IB. Solar Activity Forecast: Solar activity is likely to be low with a chance for M-class flares on day one (31 Aug) and likely to be low with a slight chance for an M-class flare on day two (01 Sep) and expected to be very low with a chance for a C-class flares and a slight chance for an M-class flare on day three (02 Sep).

Active region numbers have 10,000 subtracted (for completely unknown reasons…)

Other sources will show AR 12860

Activity over previous 24-hours
Starting 1 hour prior to issue

Forecast for next 72-hours

Official NOAA space weather forecasts & reports

Best summary product: Report & Forecast of Solar & Geophysical Activity

IIA. Geophysical Activity Summary 29/2100Z to 30/2100Z: The geomagnetic field has been at quiet to unsettled levels for the past 24 hours. Solar wind speed reached a peak of 438 km/s at 30/0000Z. Total IMF reached 9 nT at 29/2110Z. The maximum southward component of Bz reached -4 nT at 29/2127Z. Electrons greater than 2 MeV at geosynchronous orbit reached a peak level of 4481 pfu.

Activity over previous 24-hours
Starting 1 hour prior to issue

IIB. Geophysical Activity Forecast: The geomagnetic field is expected to be at quiet to unsettled levels on day one (31 Aug), quiet to minor storm levels on day two (01 Sep) and **unsettled to major storm levels** on day three (02 Sep).

Forecast for next 72-hours

Descriptive terminology for Kp levels

Adopted from terrestrial weather.

Not widely used.

Official NOAA space weather forecasts & reports

Best summary product: Report & Forecast of Solar & Geophysical Activity

```
III.  Event probabilities 31 Aug-02 Sep
Class M     30/20/10
Class X     05/01/01
Proton      05/05/05
PCAF        green

IV.  Penticton 10.7 cm Flux
Observed                30 Aug 091
Predicted    31 Aug-02 Sep 090/090/090
90 Day Mean             30 Aug 079

V.  Geomagnetic A Indices
Observed Afr/Ap 29 Aug  009/009
Estimated Afr/Ap 30 Aug  011/015
Predicted Afr/Ap 31 Aug-02 Sep  007/008-016/018-023/030
```

Note: customers want probabilistic forecasts

PCAF = Polar Cap Absorption Forecast

Red-yellow-green “stoplight” forecast of HF communication conditions in the Arctic

Current operational flare forecasting methodology

Dominated by human-in-the-loop processes

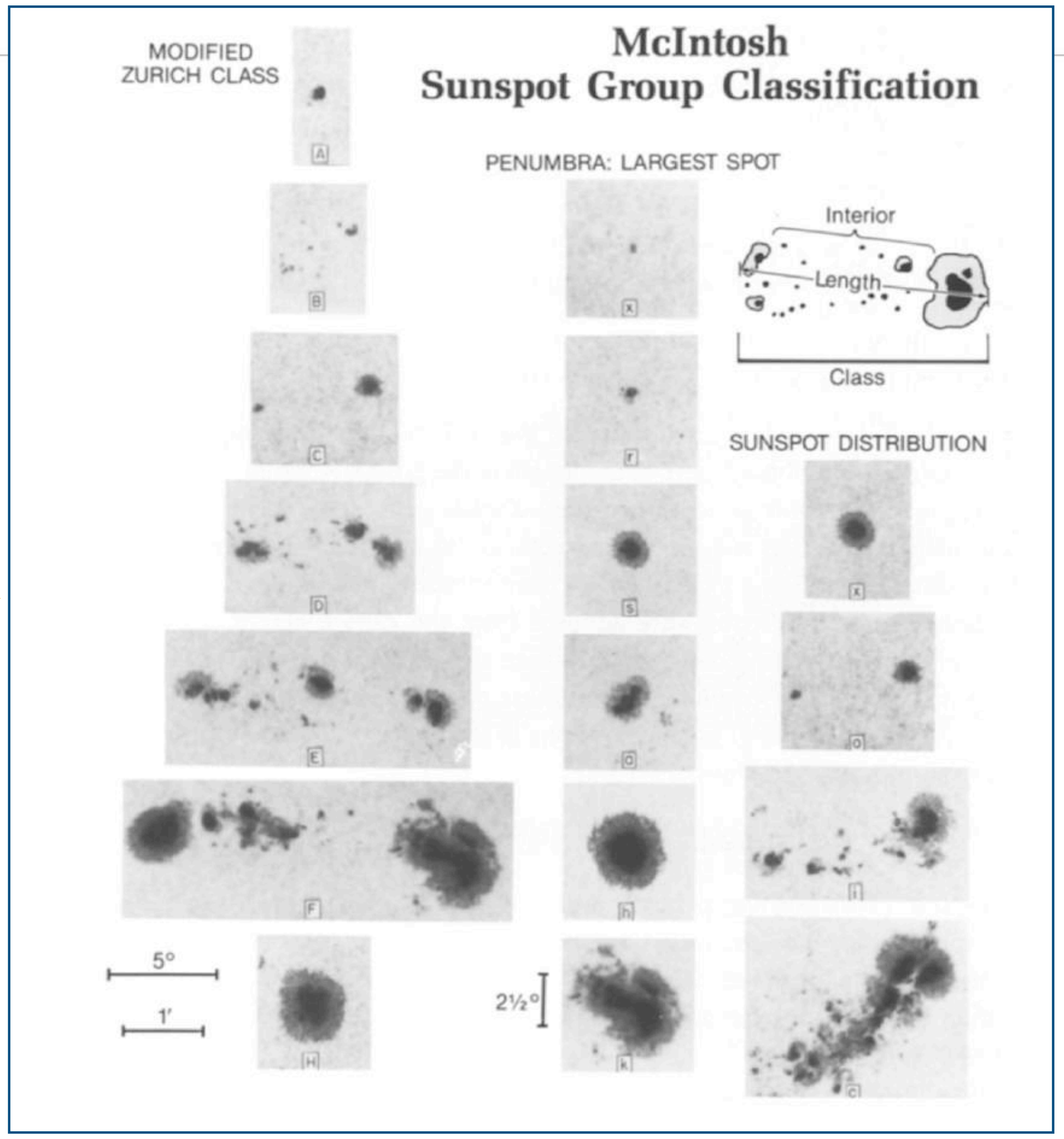
```
III.  Event probabilities 31 Aug-02 Sep
Class M      30/20/10
Class X      05/01/01
Proton       05/05/05
PCAF         green
```

GONG Magnetic Field

SOON* Sunspot imagery

* In practice, SDO/HMI is used as well

Magnetogram & Continuum Images

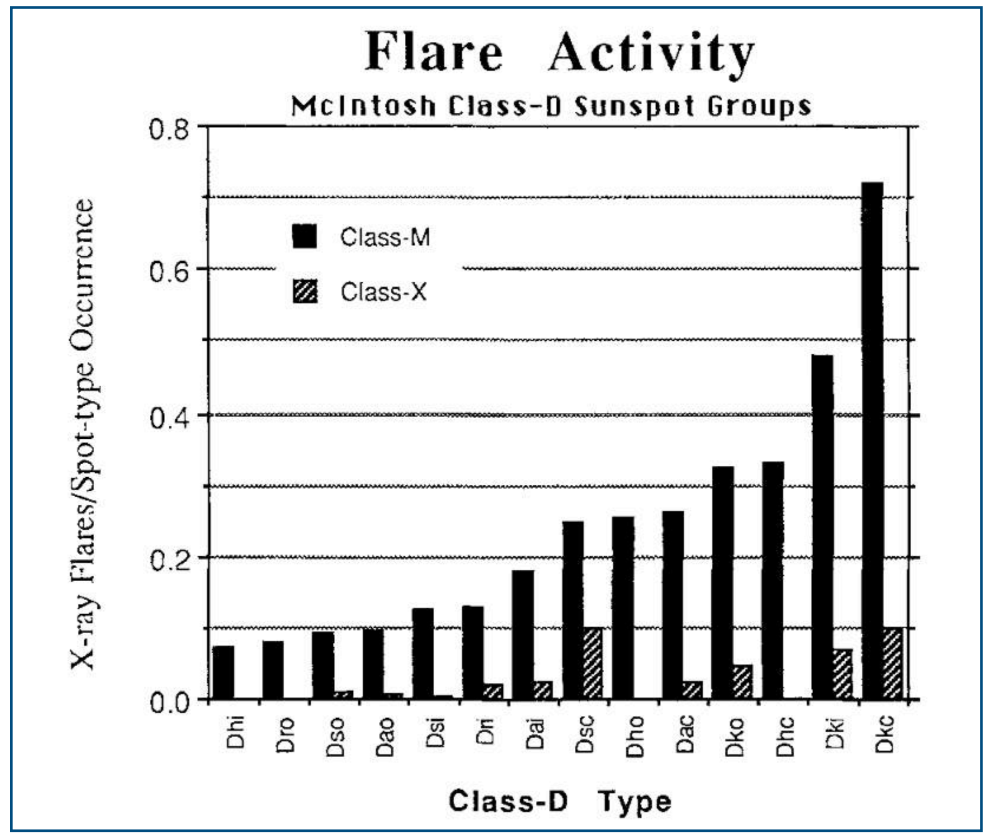


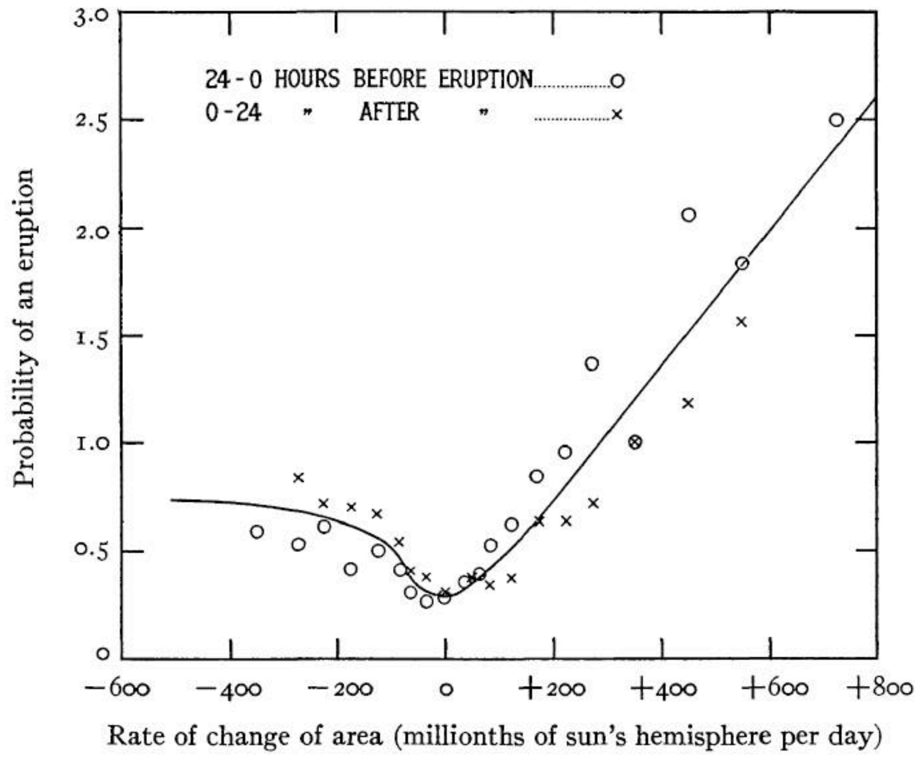
McIntosh, SolPhys, 125, 251, 1990

McIntosh Sunspot Classification

Forecaster processing

- Climatology look-up table
- Growth/decay of active region
- Total area of active region
- Flaring history
- Forecaster expertise





Giovanelli, R. G. (1939) *ApJ* 89(5), pp. 555–567.

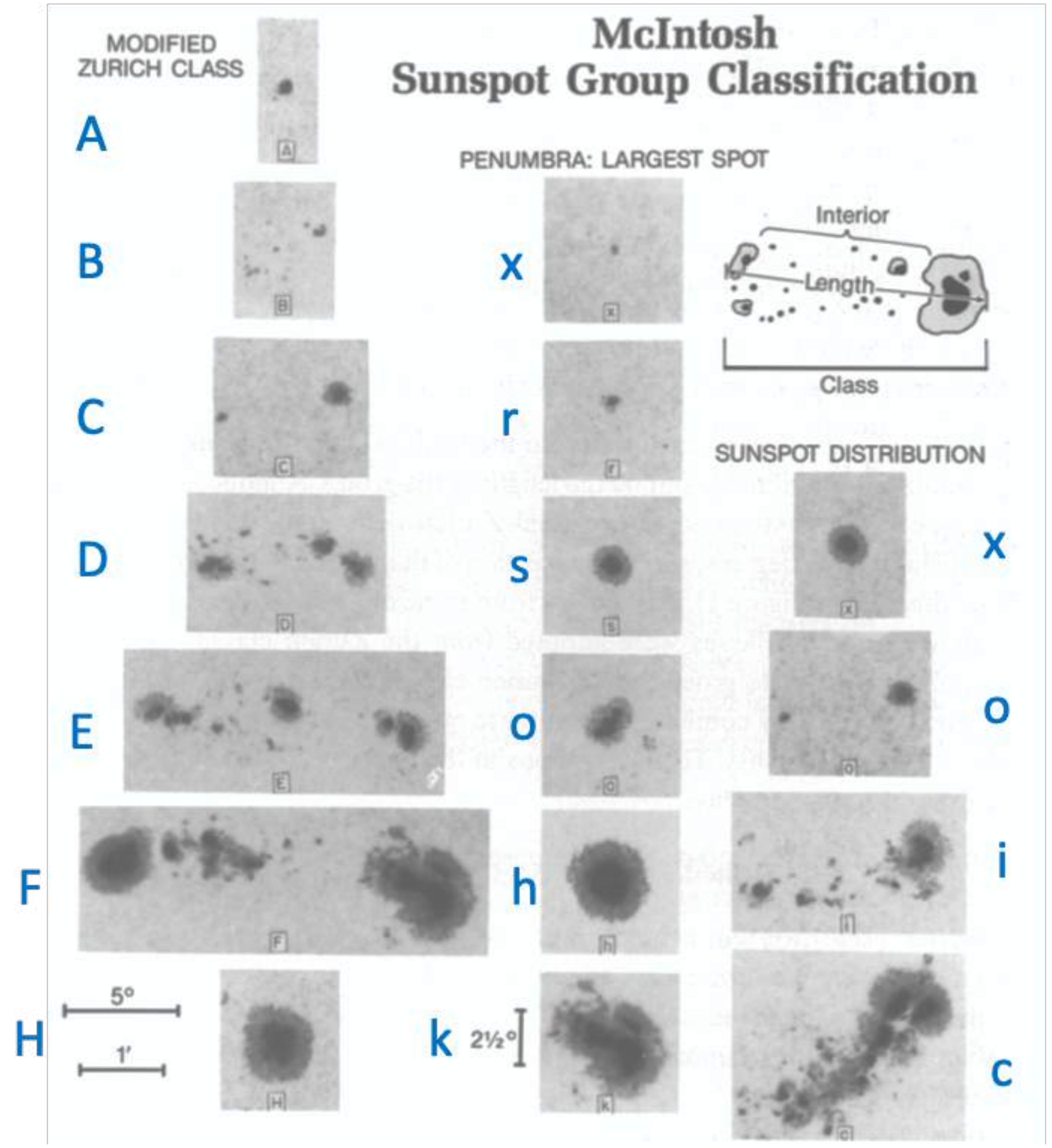
Probability of flaring in the next n hours

$P_f(n)$

n = 24, 48, 72 hours

Interlude: classification of active regions

McIntosh system: based on continuum images of sunspots



Three letter classification:

Modified Zurich Class (Length of AR): A, B, C, D, E, F, H

Size of Penumbra of Largest Spot: x, r, s, o, h, k

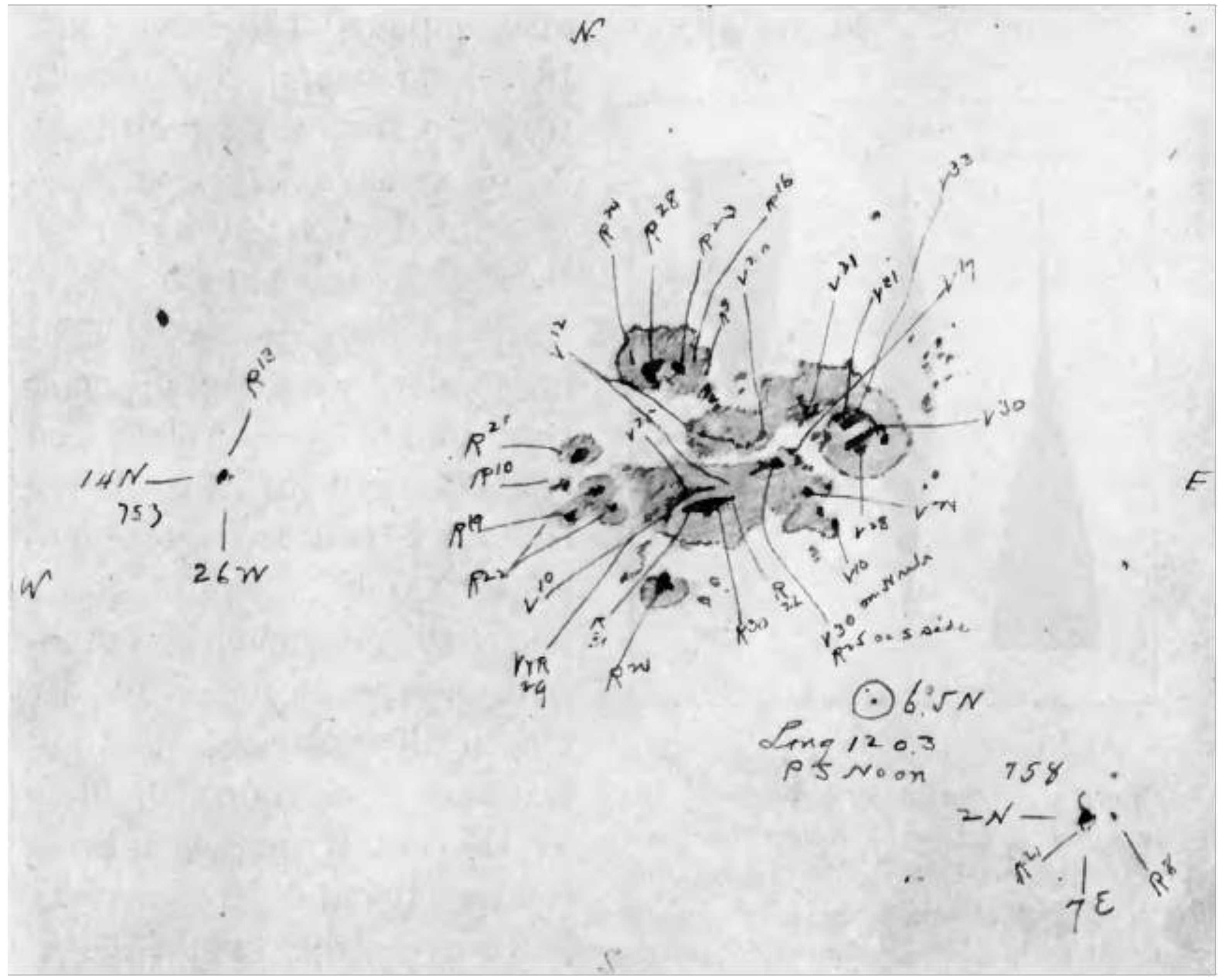
Spot Distribution in AR interior: x, o, i, c

The larger and "more complex" an AR, the more likely it is to flare:

Ekc and **Fkc** classes are the most flare productive according to ~50 years of flare data correlated to McIntosh classifications from NOAA.

Interlude: classification of active regions

Mt. Wilson system: based on magnetic field measurements



Mt. Wilson Sunspot Drawing
August 10, 1917

α: Unipolar sunspot (trailing polarity is plage/network).

β: Bipolar sunspots (leading and trailing spots of opposite polarity).

γ: Complex mix of polarities in same active region.

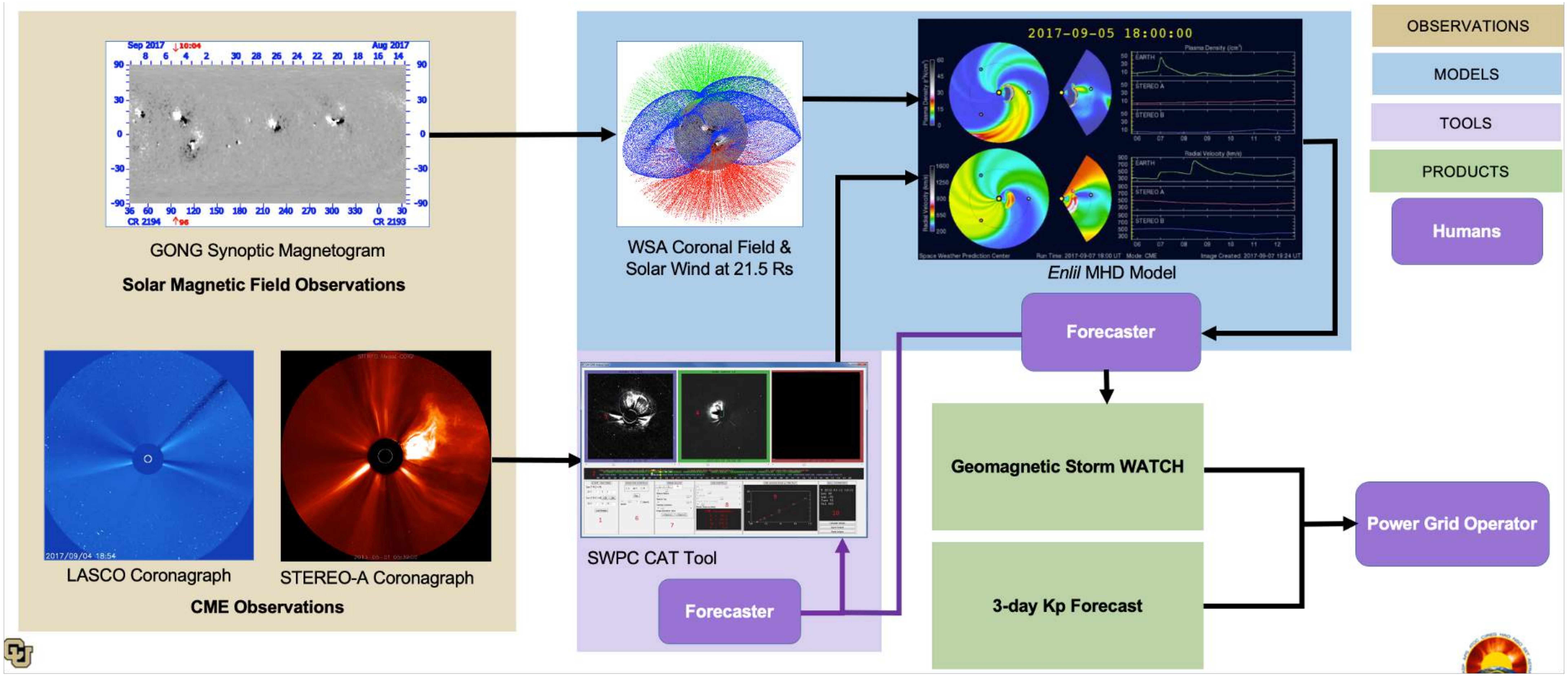
δ: Opposite polarity umbra in same penumbral area.

$\beta\gamma\delta$ is typically the most complex and flare productive active region classification in the Hale classification.

Operational solar wind & CME arrival time forecast

WSA/Enlil model: a complex chain of models

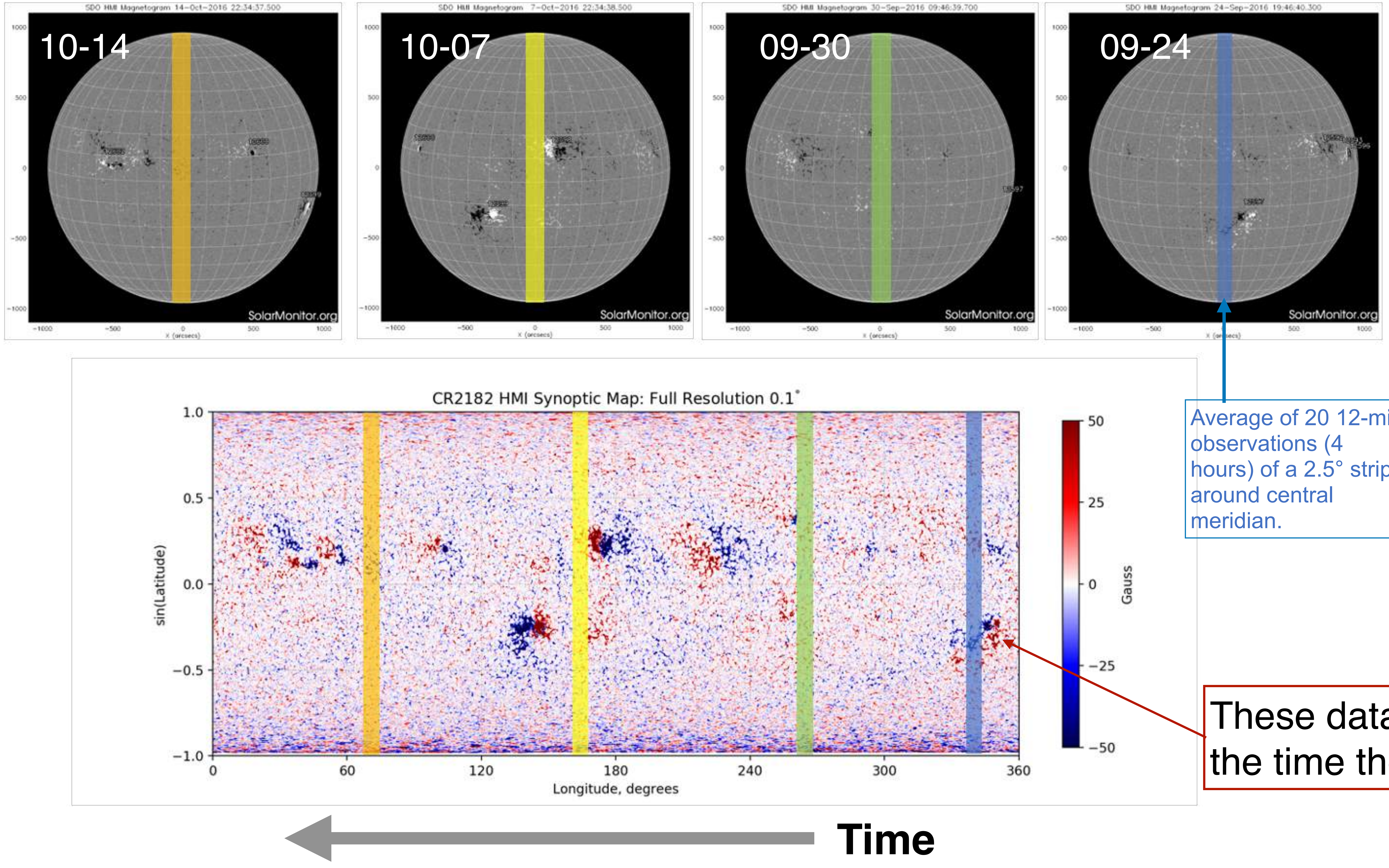
WSA: solar wind from corona to 21 Rsun.
Enlil: MHD model of heliosphere from 21 Rsun to 100 AU.



https://www.swpc.noaa.gov/products/wsa-enlil-solar-wind-prediction

Synoptic Magnetogram Construction

Classic "central meridian" method used by Stanford/HMI team



Average of 20 12-min observations (4 hours) of a 2.5° strip around central meridian.

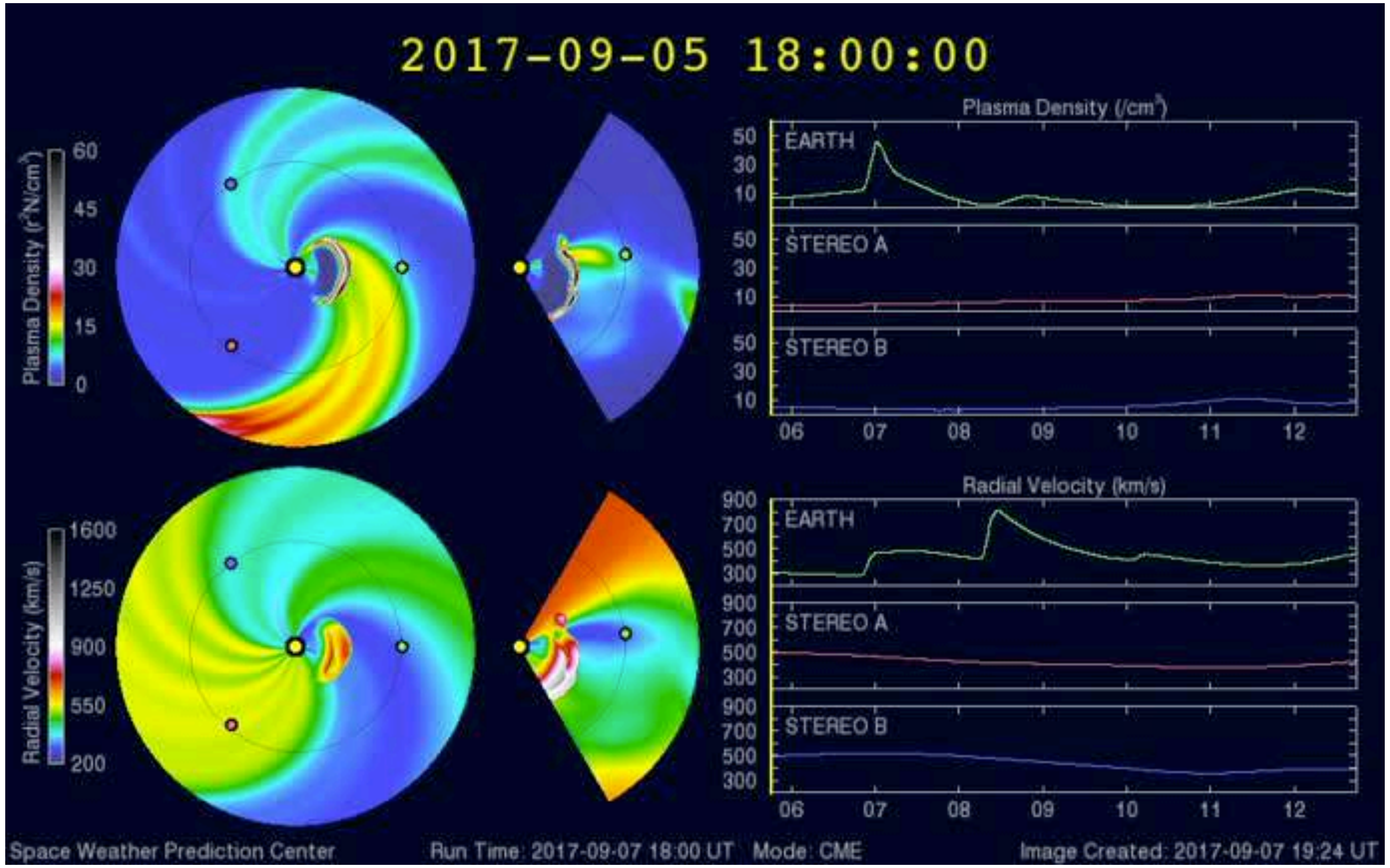
These data are 27 days old by the time the map is finished!

Synoptic magnetograms are *fictional* products that are more or less wrong, depending on what happens on the far side of the Sun at any time and at the poles over the course of the cycle.

What does this do to the coronal magnetic field models used to construct solar wind forecasts?

Operational solar wind & CME arrival time forecast

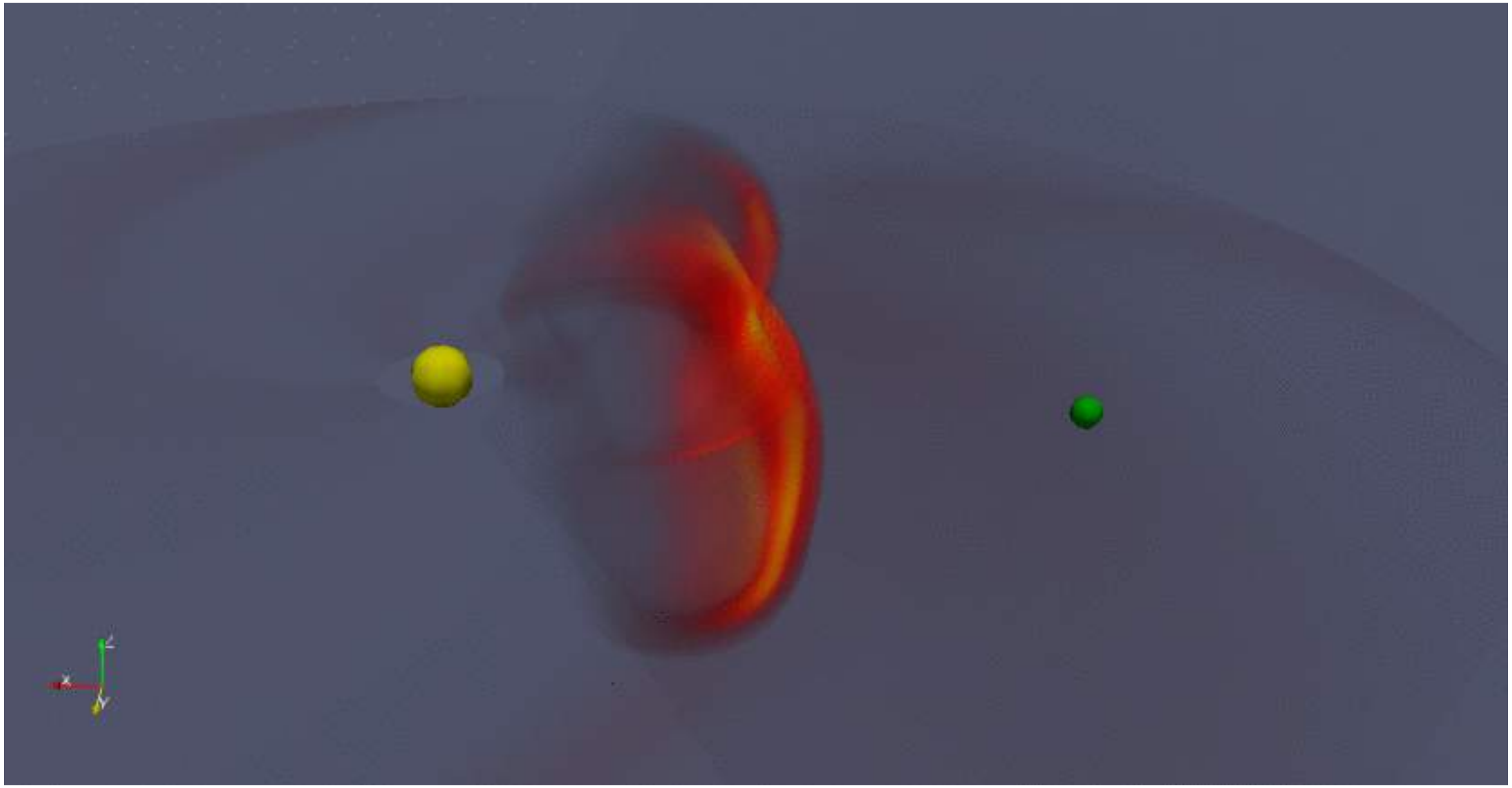
WSA/Enlil model: a complex chain of models



https://www.swpc.noaa.gov/products/wsa-enlil-solar-wind-prediction

Operational solar wind & CME arrival time forecast

WSA/Enlil model: 3D is always better!



https://enlil.swx-trec.com

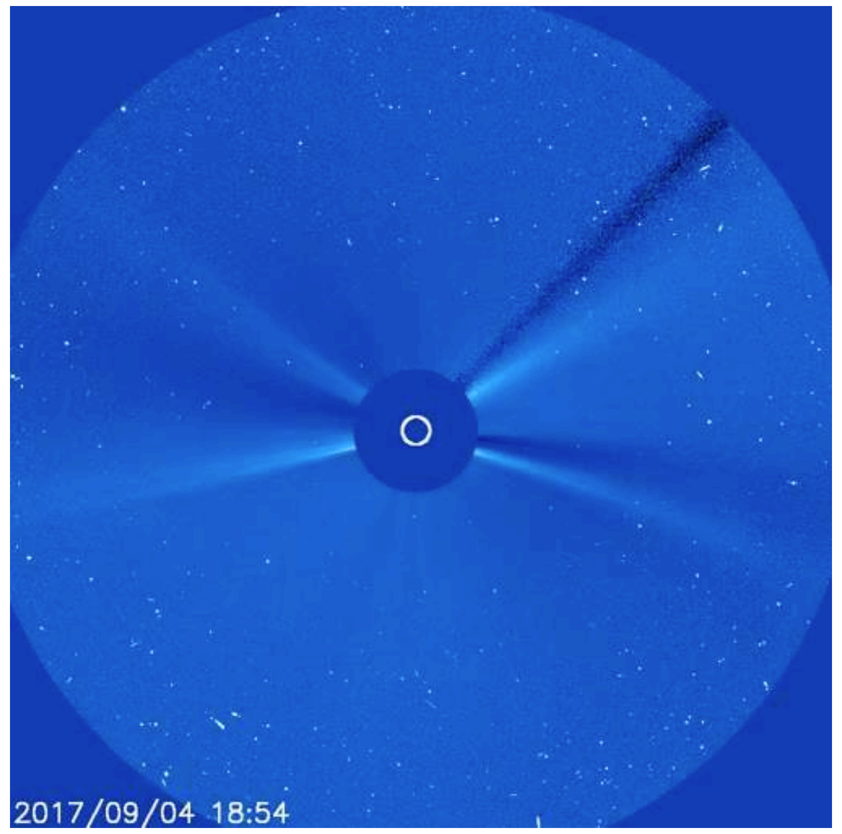
Operational solar wind & CME arrival time forecast

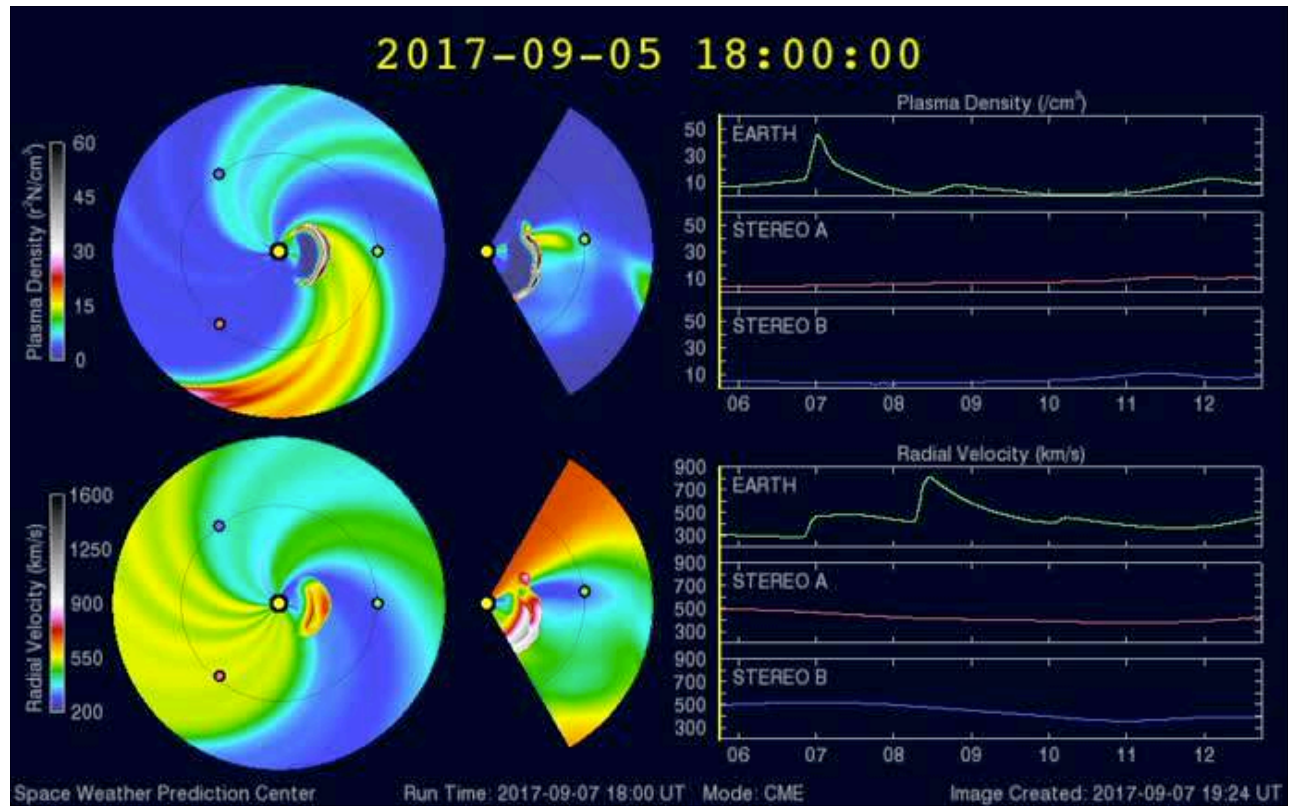
Timing of products relative to observations and model runs

1. Geomagnetic Storm WATCH

Issued upon coronagraph detection of Earth-directed CME and WSA-Enlil model run.

- 15—60 hours before Earth impact, if at all…

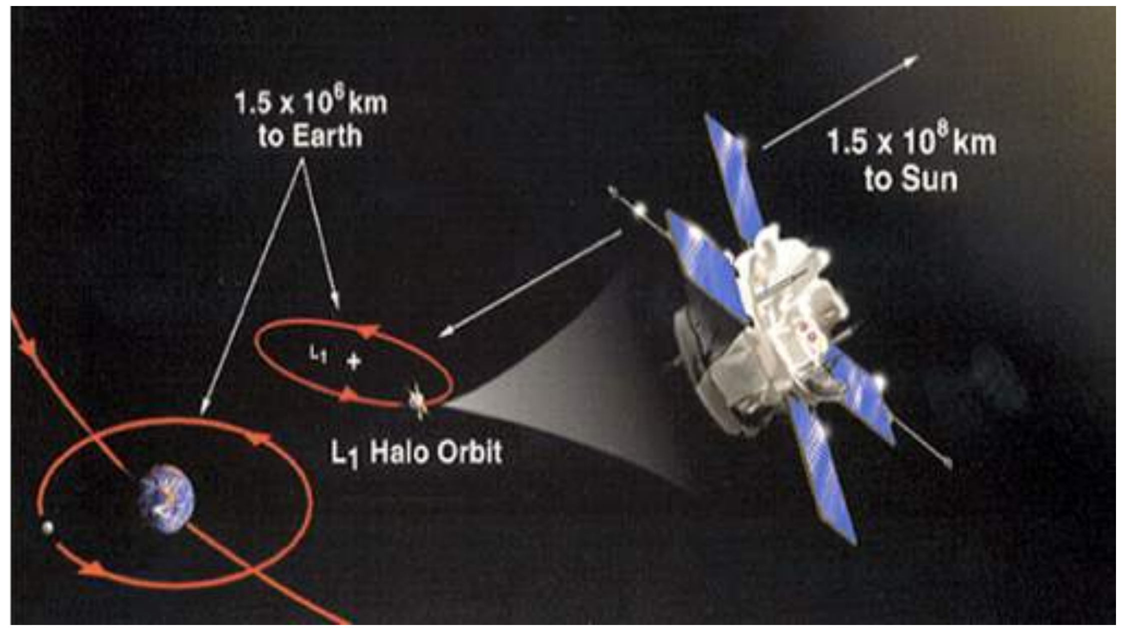


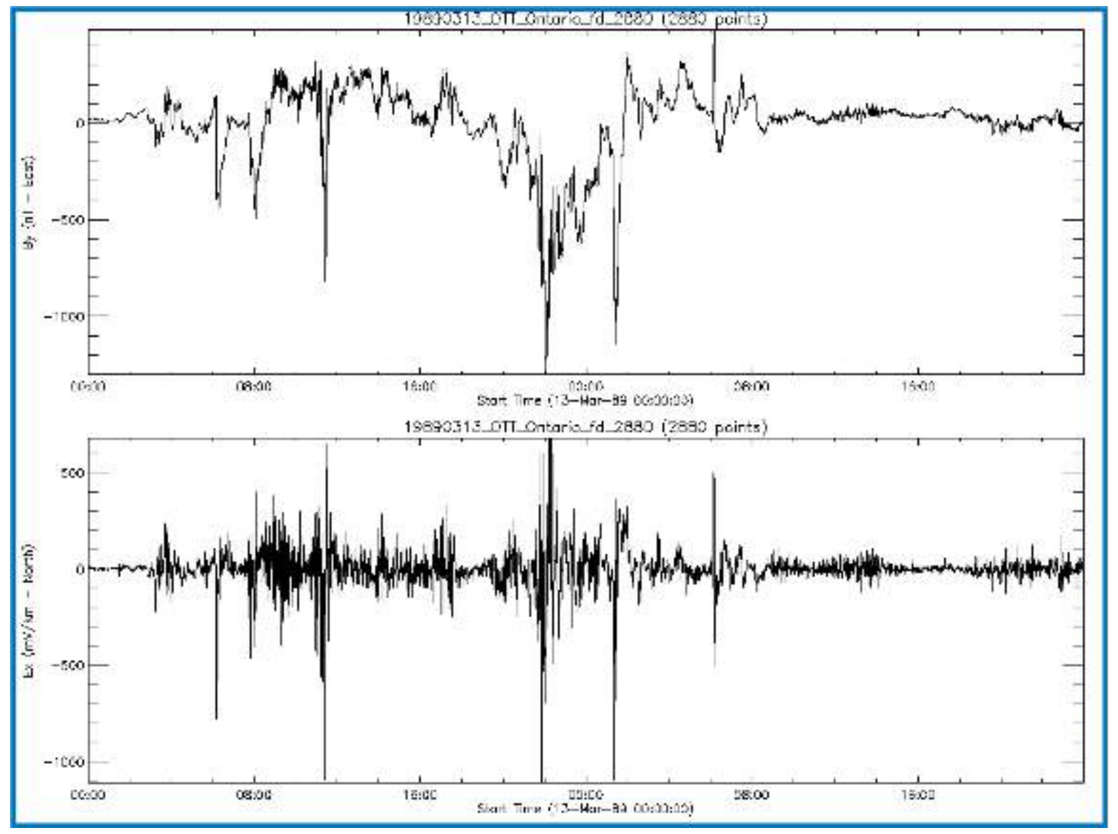


2. Geomagnetic Storm WARNING

Issued upon detection of CME shock wave at DSCOVR or ACE satellites at L1 Lagrangian point.

- 15—60 minutes before impact





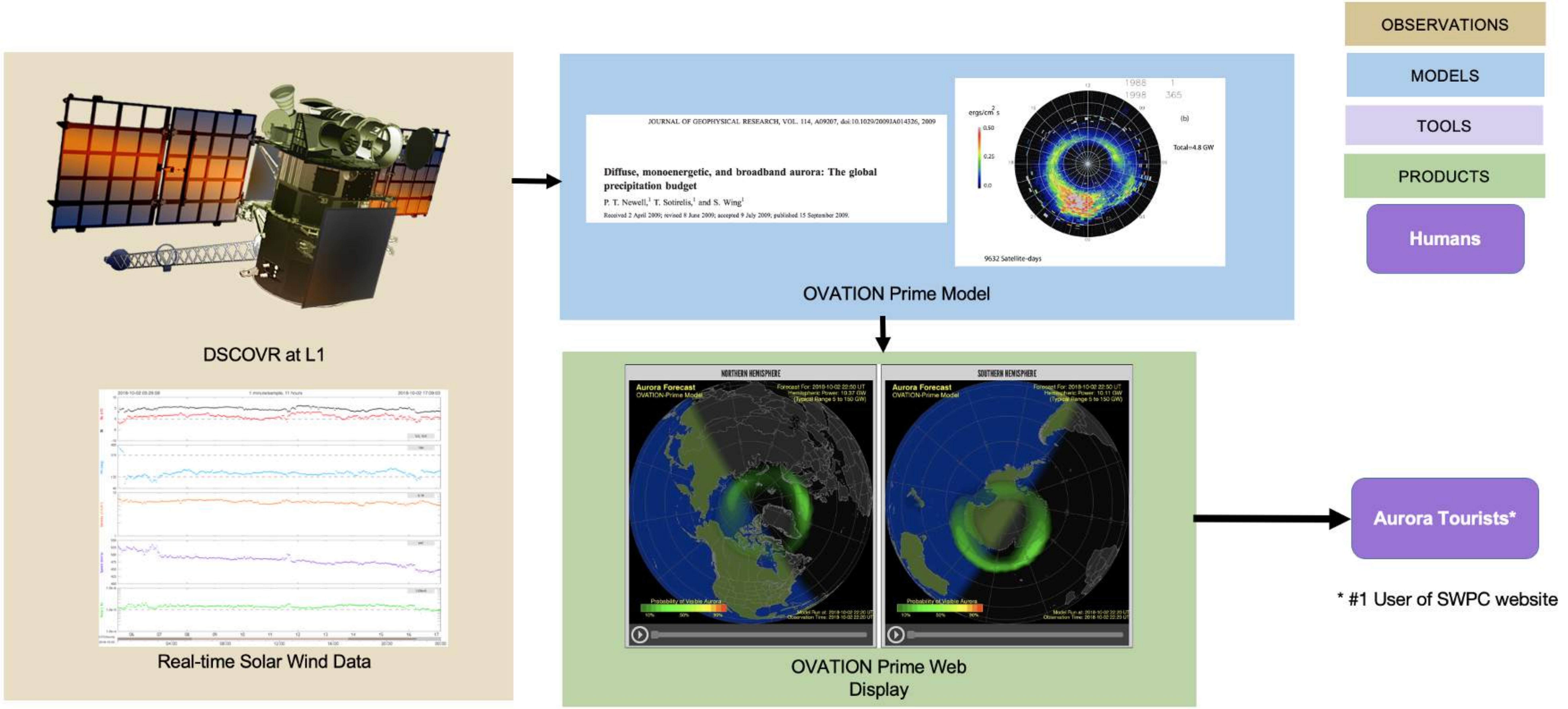
3. Geomagnetic Storm ALERT

Issued upon detection of USGS ground magnetometer disturbance.

- Current condition - nowcast

Operational aurora visibility forecast

OVATION prime model: 24-hour forecast window



https://www.swpc.noaa.gov/products/aurora-30-minute-forecast

Forecast evaluation metrics

How well do we forecast space weather?

Two main types of forecasts issued in weather and space weather:

1. Event-based “**binary categorical**” forecasts: will an event happen or won’t it?

2. **Probabilistic** forecasts: what is the likelihood that an event will occur?

Forecast evaluation metrics

Binary categorical forecast truth table or “contingency table”

		Forecast		
		Positive	Negative	
Observed	Positive	A (Hit, *TP*)	B (Miss, *FN*)	*Events*
	Negative	C (False Alarm, *FP*)	D (Null, *TN*)	*Non-events*

TP = true positive, ***FP*** = false positive

TN = true negative, ***FN*** = false negative

N = A + B + C + D = TP + FP + TN + FN

Descriptive Metrics	Formula
Base or Event rate, s	$\frac{A+B}{N}$
Forecast Rate, r	$\frac{A+C}{N}$
Frequency Bias, B	$\frac{A+C}{A+B}$
Ratio Test, R	$\frac{A+D}{N}$

Performance Metrics	Formula		
Precision	$\frac{A}{A+C}$		
Recall, Probability of Detection, Sensitivity	$\frac{A}{A+B}$	TPR	True Positive Rate
Probability of Missed Detection	$\frac{B}{A+B}$	1 - TPR	
Accuracy	$\frac{A+D}{N}$		
False Positive Rate[1]	$\frac{C}{C+D}$	FPR	False Positive Rate
Specificity	$\frac{D}{C+D}$	1 - FPR	
False Alarm Ratio[1]	$\frac{C}{A+C}$		
True Skill Statistic, TSS	$\frac{A}{A+B}-\frac{C}{C+D}$	TPR - FPR	
Critical Success Index	$\frac{A}{A+B+C}$		
F1 score	$\frac{2A}{2A+B+C}$		

Woodcock, F. (1976) *Monthly Weather Review*, 104(10), pp. 1209–1214.

Forecast evaluation metrics

Skill scores based on contingency table and reference forecasts

Skill = S – S_2 / S_1 – S_2

where S is the number of correctly partitioned occurrences (positive and negative) by a model in any trial, and S1 and S2 are standard predictor results over the same trial.

S = TP + TN = A + D

Standard Predictors[2]:

- Perfect: Sp = N
- False: Sf = 0
- Random: Sr
- Unskilled: Su

Skill Scores	Definition	Contingency Formula	
Skill Test, Sk	$\frac{2(S-Sr)}{N}$	$\frac{4(AD-BC)}{N^2}$	
Heidke Skill Score, HSS	$\frac{S-Sr}{Sp-Sr}$	$\frac{2(AD-BC)}{(A+B)(B+D)+(A+C)(C+D)}$	$[-\infty,1]$
Appleman's Discriminant, U	$\frac{S-Su}{Sp-Su}$	$\frac{D-B}{C+D}$	
Hansen & Kuipers Skill Score, HKSS = True Skill Statistic (TSS)	$R_{ev}-R_{non-ev}-1$	$\frac{A}{A+B}+\frac{D}{C+D}-1=\frac{AD-BC}{(A+B)(C+D)}$	$[0,1]$
Schrank's Discriminant, W	$\frac{R+S-1}{2}$	$\frac{(A+D)^2+(B+C)(A+D-1)}{2N}$	
Correlation Coefficient, r		$\frac{AD-BC}{[(A+B)(A+C)(C+D)(B+D)]^{1/2}}$	

Ratio Test, R = S – Sf / Sp – Sf

R = (A + D) / N

2. Woodcock, F. (1976) *Monthly Weather Review*, 104(10), pp. 1209–1214.

Forecast evaluation metrics

Common reference forecasts: S_2 in previous slide

- **Climatology Forecast:** the probability of an event occurring is the average of the probability over the relevant period.

For example, a climatology flare forecast would calculate the probability of an active flaring based on the probability of flaring for all recorded active regions over, say, the past and current solar cycle.

If you can't do better than climatology, your method should be abandoned.

- **Persistence Forecast**: things will stay just as they are right now, i.e., no flare is occurring now so that's the way it will stay.

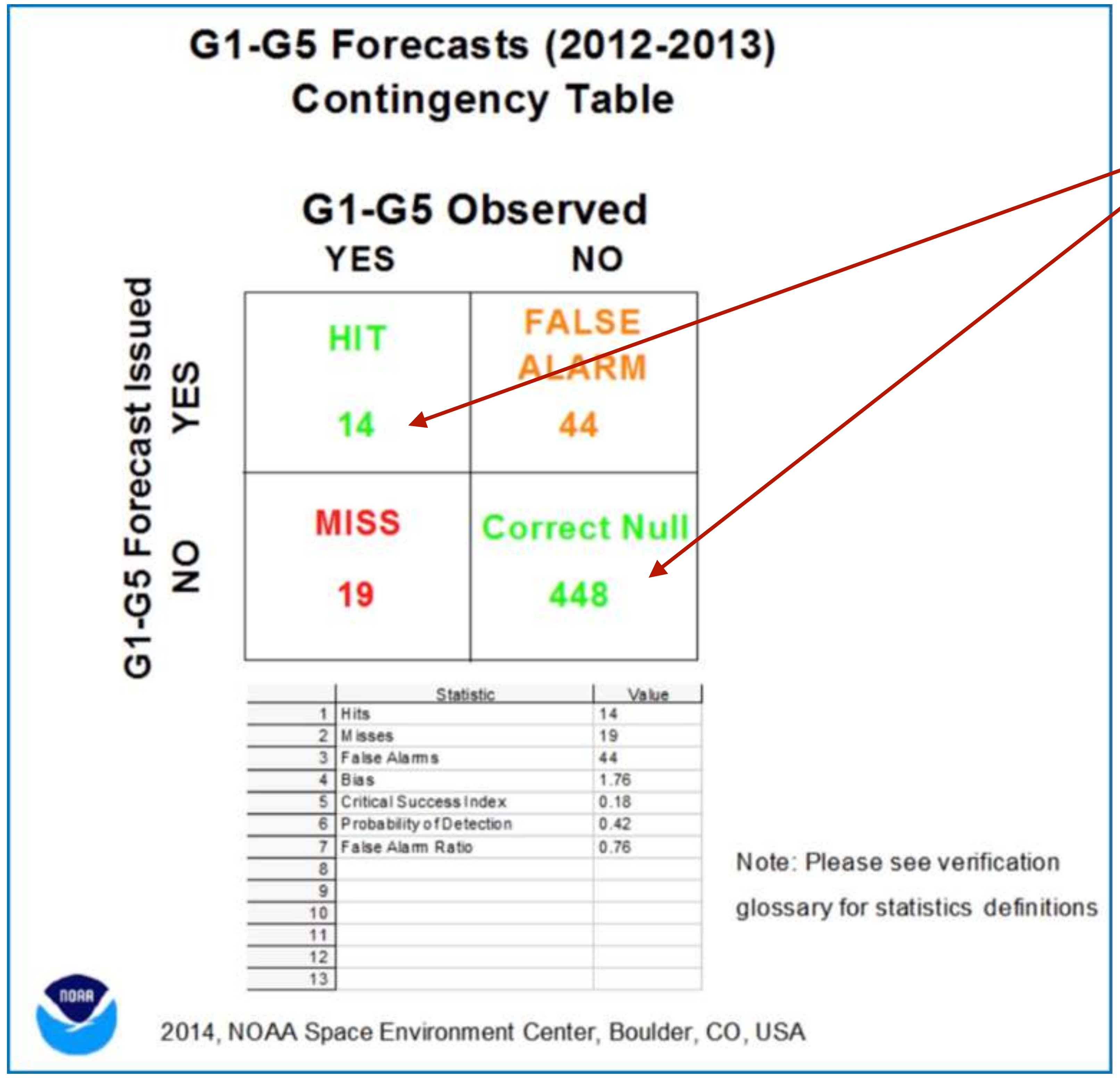
Note that this is a *very accurate* forecast 90+% of the time for episodic events. But it is also *useless* for high-impact episodic events like solar eruptions.

- **Recurrence Forecast**: the probability of an event occurring is the based on the probability of conditions returning.

This is the current operational method for forecasting coronal hole high-speed streams: the Sun rotates every 27 days so HSS events are predicted to return every 27 days.

Forecast evaluation: NOAA geomagnetic storm forecasting

Binary categorical forecast: G1 storm or greater vs. no storm



Note imbalance in events: many more negatives than positives.

P = 33
N = 492

This is typical of many space weather phenomena

Geomagnetic Storm Forecast Accuracy is measured as the percentage of times that the 24 hour geomagnetic storm forecast is correct for the 60 most recent geomagnetic storms.

The 24 hour geomagnetic storm forecast is considered accurate if a G1 or greater storm event was predicted. This measure is based on the next-day forecast of maximum Kp, where Kp=5 (NOAA Scale G1) or greater constitutes a storm, and is verified against the NOAA Kp estimated from ground-based magnetometer observations.

$$A_{forecast} = \frac{14 + 448}{525} = 0.88$$

$$A_{climatology} = \frac{FP + TN}{N} = 0.94$$

Just guess "No" every time when positives are rare and you can achieve high accuracy. But is that useful?

TSS = POD - FAR = 0.42 - 0.09 = 0.33

Forecast evaluation metrics

Probabilistic forecast: compare error in predicted rate to observed rate

- **Brier Skill Score (BSS): [0, 1]**

$$\text{BSS} = \frac{1}{n}\Sigma_{i=1}^{n}(P_i - O_i)^2$$

O_i = observed event frequency = 1 if event occurred, 0 if not.
P_i = forecast probability for event i.

NOAA Solar eruption probability skill:

BSS ~ 0.132 for 24 hrs, 0.077 for 48 hrs, and 0.040 for 72 hours
Slightly better than climatology forecast.

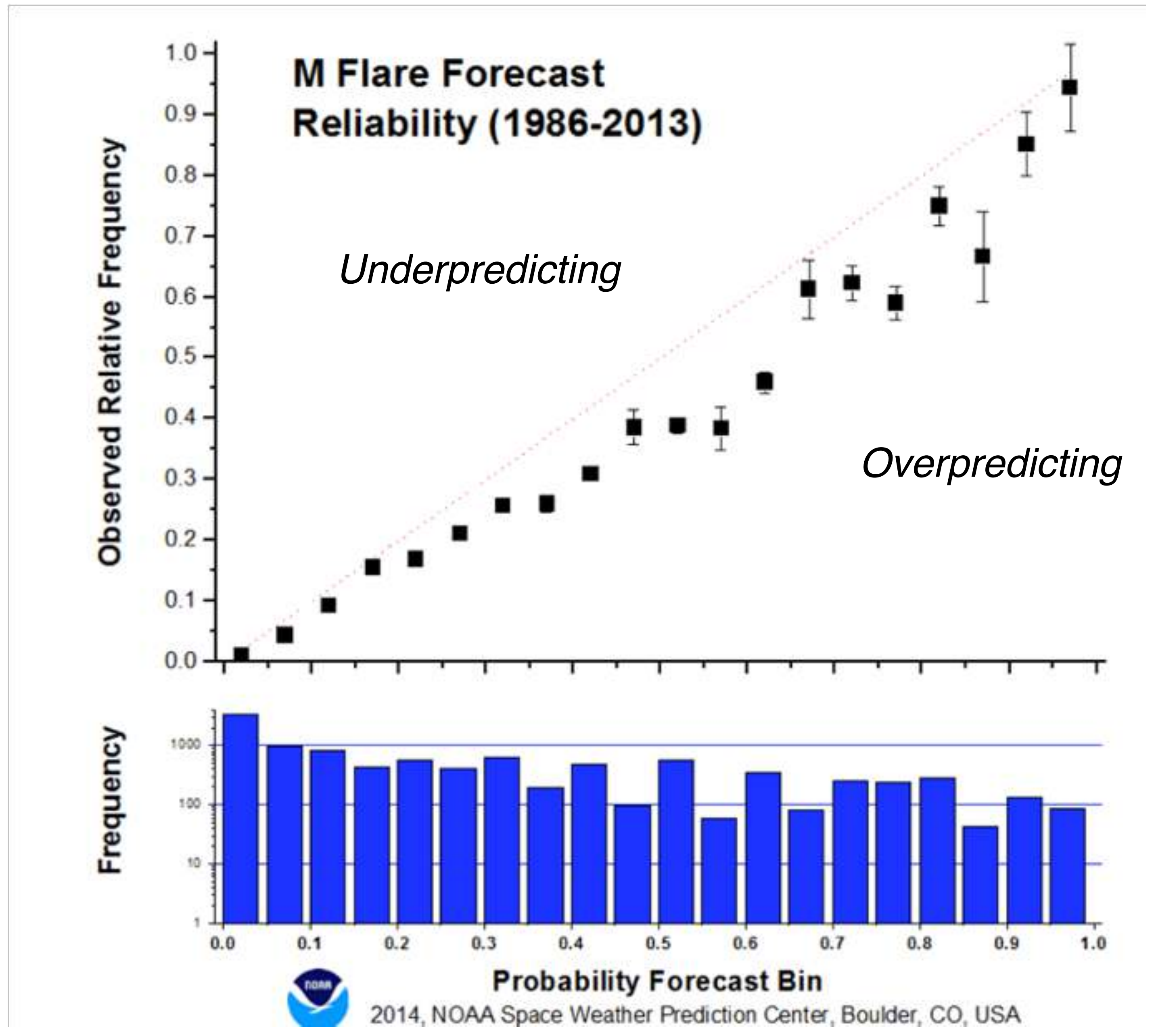
Not feasible to predict probability of eruption over time spans greater than major magnetic evolution time in sunspot ARs (30 hours, Schrijver, 2016) so 3-day forecast skill is very low.

See Leka et al., 2019 for comparison of current NOAA method to developing automated methods.

https://www.swpc.noaa.gov/content/solar-activity-forecast-verification

Reliability Curve

Plot of O_i vs. P_i binned by P_i



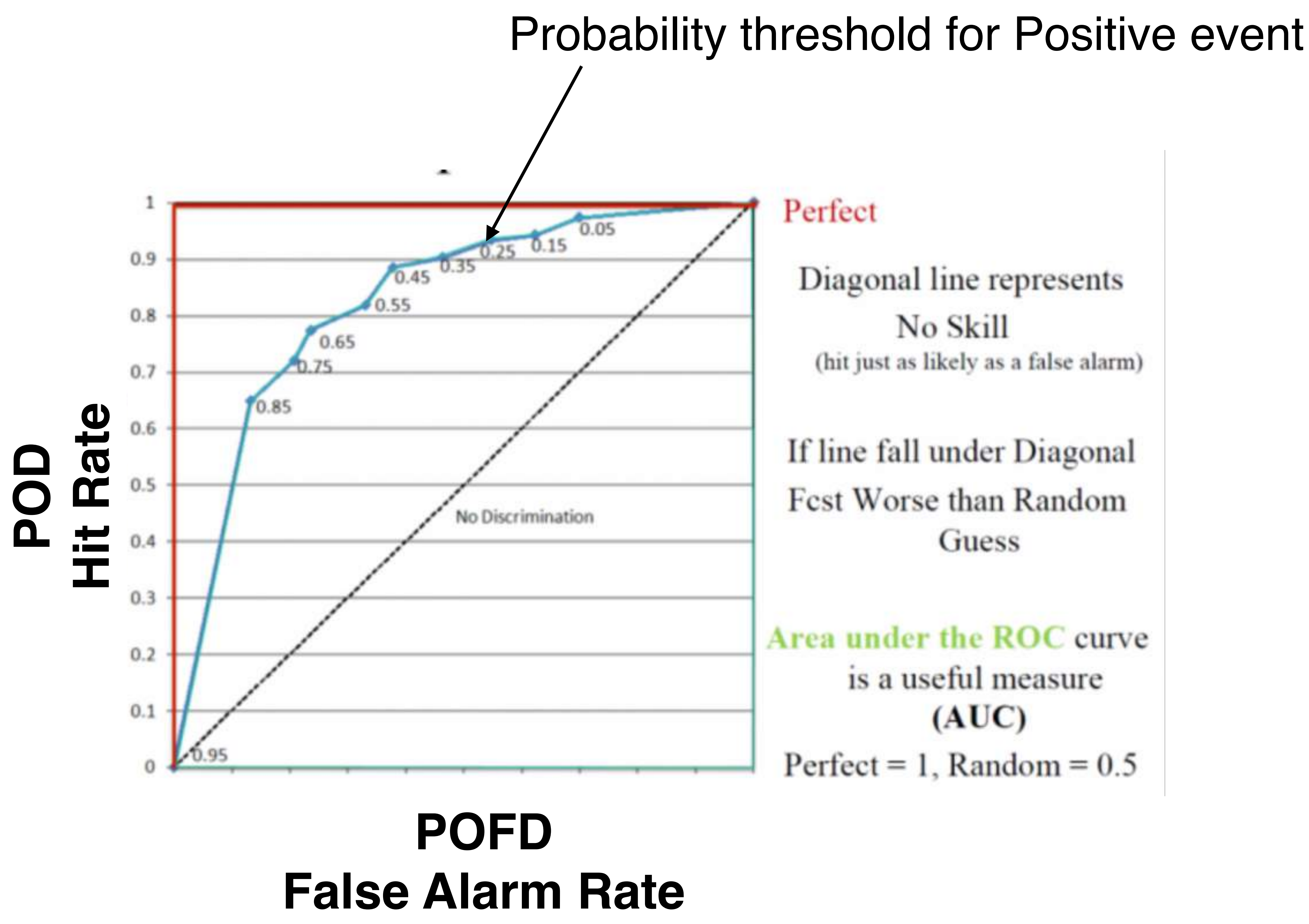
Forecast evaluation metrics

Threshold-dependent evaluation of probabilistic forecasts

- **ROC = Receiver Operating Characteristic**
  Plot of Probability of Detection vs.
  Probability of False Detection POD vs. POFD.

  Curve is a parametric function of probability threshold for event, *p*.

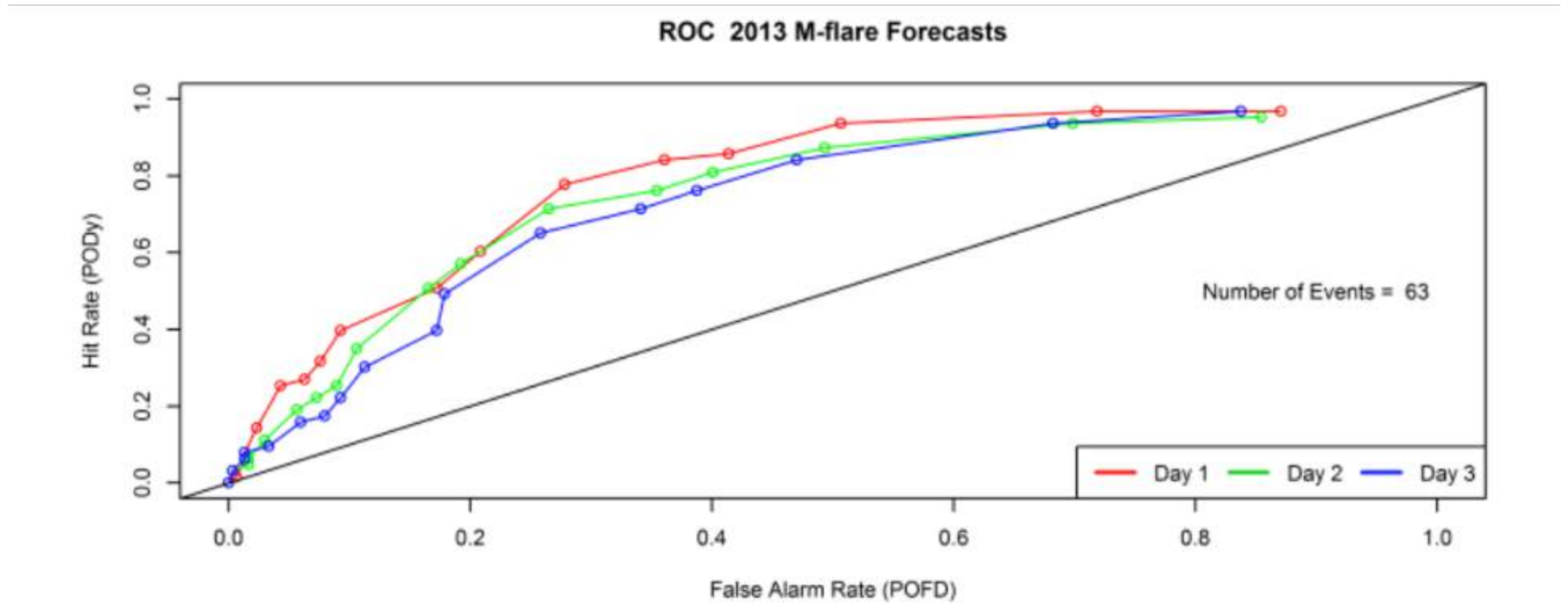
  p=1 means TP=FP=0: (0,0) on plot
  p=0 means FN=TN=0: (1,1) on plot

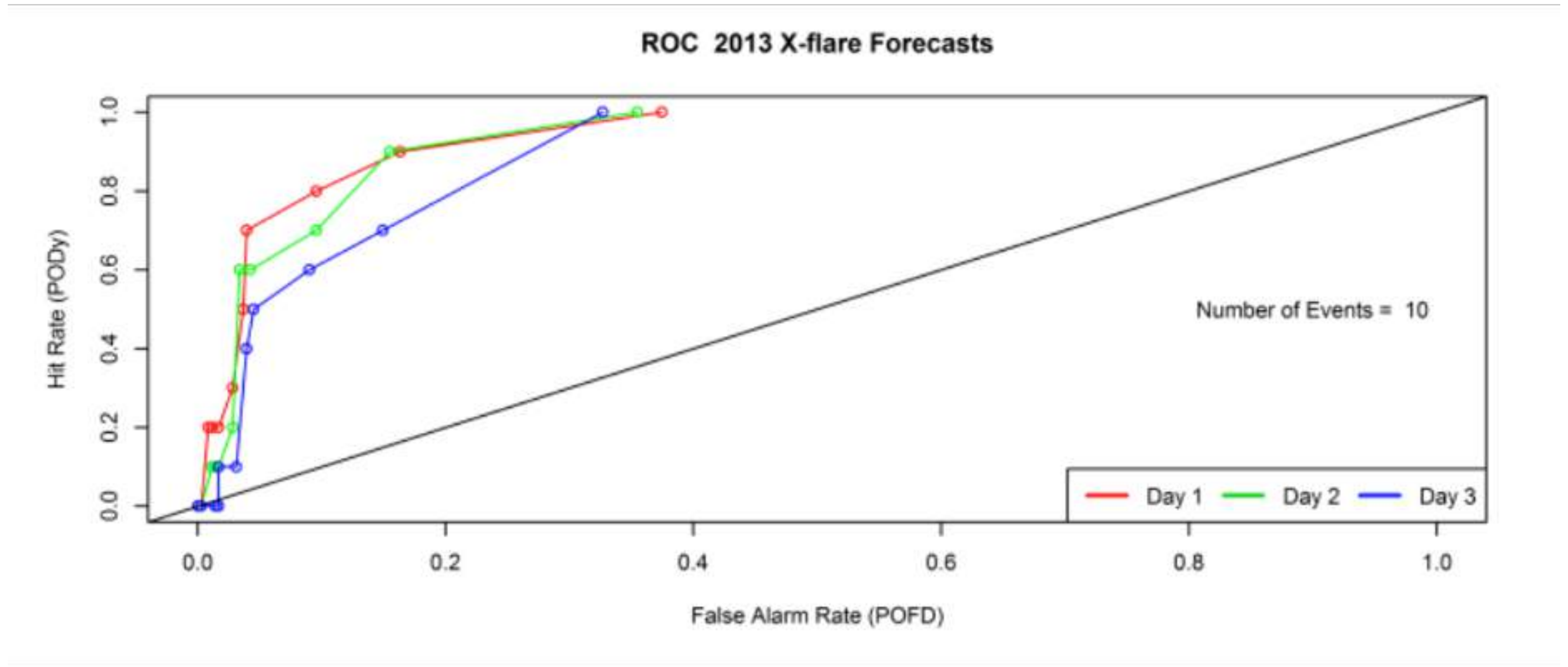


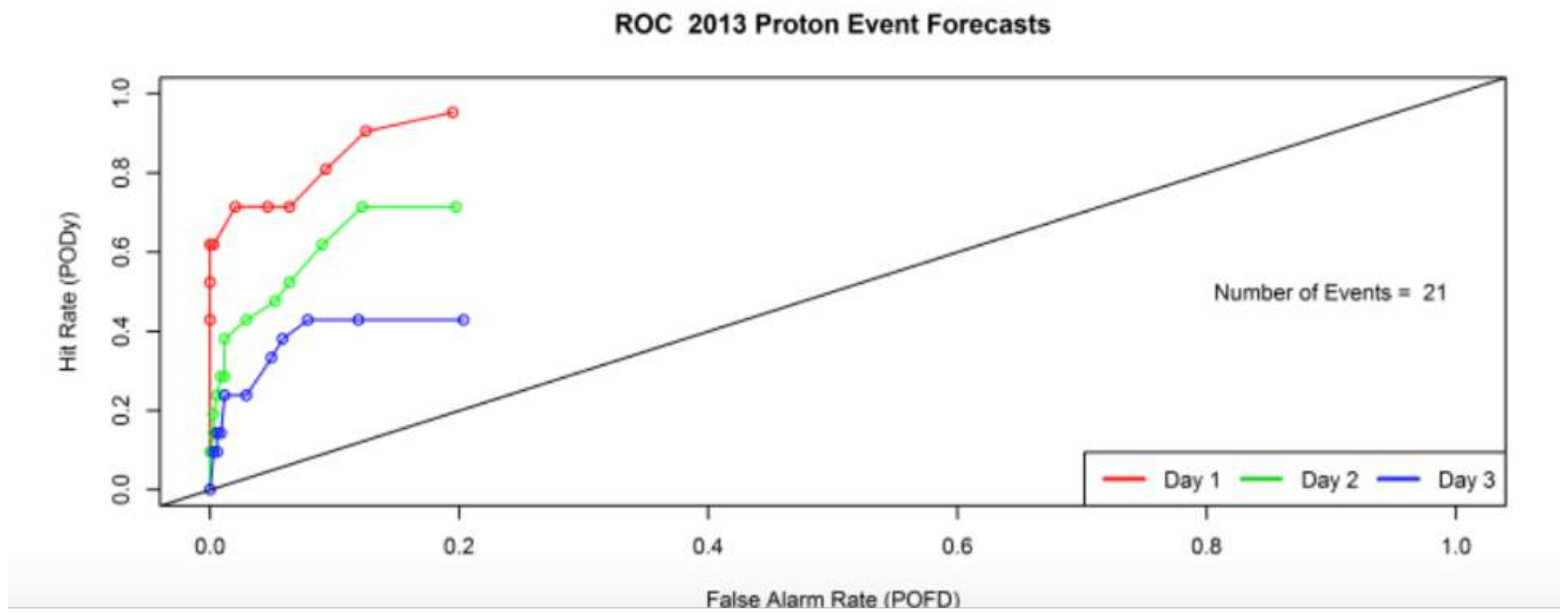
Note:
"Default" threshold = 50% for event occurrence
"Likely" = 60%
"Very Likely" = 70%

Forecast evaluation: solar flare & SEP prediction skill

NOAA/SWPC official reports







Very low number of events for 2013 and Solar Cycle 24 in general.

Data not posted beyond 2013.

https://www.swpc.noaa.gov/sites/default/files/images/u2/ROC_2013.png

Forecast evaluation: solar flare prediction skill

Binary categorical forecasts

Contingency tables are a must have for any fair comparison of systems!

Table 6. Contingency Tables for Look-up Table Probabilities by X-Ray Flare Class

Yes Forecast ≥ 0.25		C-Class Observed Yes	No	Yes Forecast ≥ 0.25		M-Class Observed Yes	No	Yes Forecast ≥ 0.15		X-Class Observed Yes	No
C-Class Forecast	Yes	2141	2042	M-Class Forecast	Yes	362	1049	X-Class Forecast	Yes	25	195
	No	1665	15786		No	543	19680		No	74	21340

Table 4. Contingency Tables for Subjective Forecast Probabilities by X-Ray Flare Class

Yes Forecast ≥ 0.50		C-Class Observed Yes	No	Yes Forecast ≥ 0.35		M-Class Observed Yes	No	Yes Forecast ≥ 0.25		X-Class Observed Yes	No
C-Class Forecast	Yes	2476	1630	M-Class Forecast	Yes	511	685	X-Class Forecast	Yes	50	67
	No	1458	25920		No	406	29882		No	52	31315

Table 5. Contingency Table Statistics by X-Ray Flare Class

	Bias	CSI	POD	POFD	PC	FAR	ETS	HSS	Records
Perfect score	1	1	1	0	1	0	1	1	
C-class flare subjective forecasts	1.043	0.445	0.629	0.059	0.902	0.397	0.389	0.560	31484
C-class flare predictions look-up table	1.099	0.366	0.563	0.115	0.829	0.488	—	0.431	21634
M-class flare subjective forecasts	1.304	0.319	0.557	0.022	0.965	0.573	0.304	0.466	31484
M-class flare predictions look-up table	1.559	0.185	0.400	0.051	0.926	0.743	—	0.276	21634
X-class flare subjective forecasts	1.147	0.296	0.490	0.002	0.996	0.573	0.294	0.455	31484
X-class flare predictions look-up table	2.222	0.085	0.253	0.009	0.988	0.886	—	0.151	21634

CSI = Critical Success Index
POD = Probability of Detection = TPR
POFD = Probability of False Detection = FPR
PC = Percent Correct = Accuracy

FAR = False Alarm Ratio
ETS = Equitable Threat Score: see Crown (2012) for definition
HSS = Heidke Skill Score

Crown, M.D. (2012), *Space Weather*, 10(6), p. n/a-n/a. doi:10.1029/2011SW000760.

Forecast evaluation: solar flare prediction skill

Probabilistic forecasts

Table 2. Brier Skill Scores [0,1] Closer to 0 is better

	All Region Types	Less complex regions Beta, Beta-Gamma, and Gamma Region Types	More complex regions Beta-Delta, Gamma-Delta, and Beta-Gamma-Delta Region Types
C-class flares observed subjective forecasts	0.100	0.123	0.183
C-class flares observed look-up table	0.111	0.139	0.193
M-class flares observed subjective forecasts	0.031	0.034	0.190
M-class flares observed look-up table	0.037	0.042	0.229
X-class flares observed subjective forecasts	0.004	0.002	0.067
X-class flares observed look-up table	0.005	0.003	0.080
Combined flaring and non- flaring subjective forecasts	0.045	0.053	0.147
Combined flaring and non- flaring look-up table	0.051	0.061	0.167

Bottom line: for M- and X-class flares (the only ones that matter), forecast skill decreases with increasing magnetic complexity of the region., likely due to over-forecasting bias (see previous slide).

Crown, M.D. (2012), *Space Weather*, 10(6), p. n/a-n/a. doi:10.1029/2011SW000760.

Reading for next week

Katsova, M.M. (2020) ‘The evolution of the solar–stellar activity’, *Journal of Atmospheric and Solar-Terrestrial Physics*, 211, p. 105456. doi:10.1016/j.jastp.2020.105456.