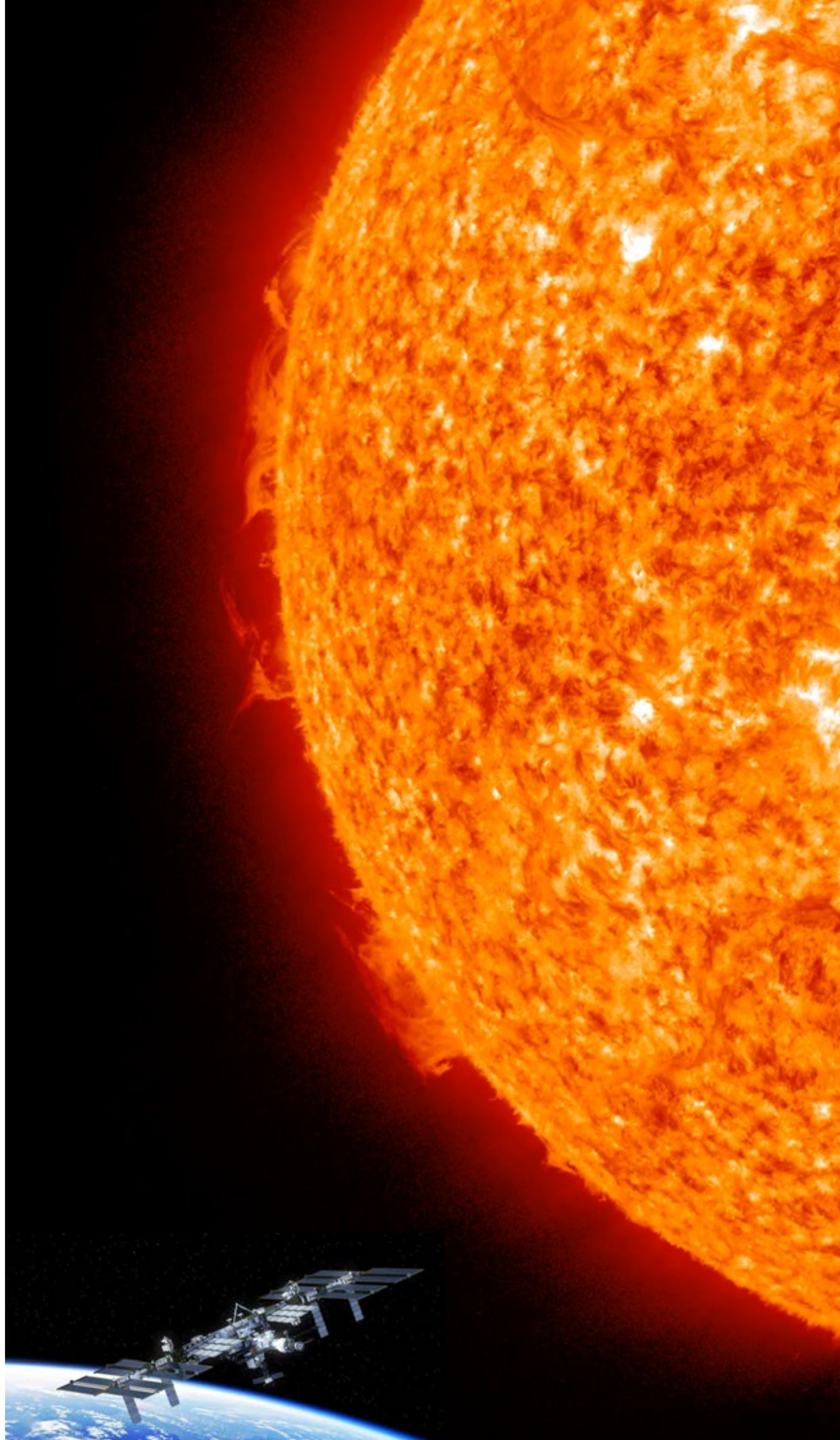
Introduction to Space Weather

COLLAGE Program
Spring 2023
Wednesday 1:00—2:00
Aerospace N250

Lecture 5

Space weather forecasting – how, and how well, can we predict space weather?

Instructor: Dr. Thomas Berger
thomas.berger@colorado.edu



Outline

1. **What is a Forecast?**
   - Forecasts vs. Simulations
   - Quantifying space weather events & impacts
   - The ART of forecasting

2. **What is a Nowcast?**

3. **Products: Watches, Warnings, and Alerts**

4. **Operational space weather capabilities in the US (civil only)**

5. **How well can we forecast space weather?**
   - Probabilistic vs. binary classification forecasts
   - Solar eruption forecasting skill
   - CME arrival time forecasting skill
   - Geomagnetic storm forecasting skill

- **Future developments**

What is a Forecast?

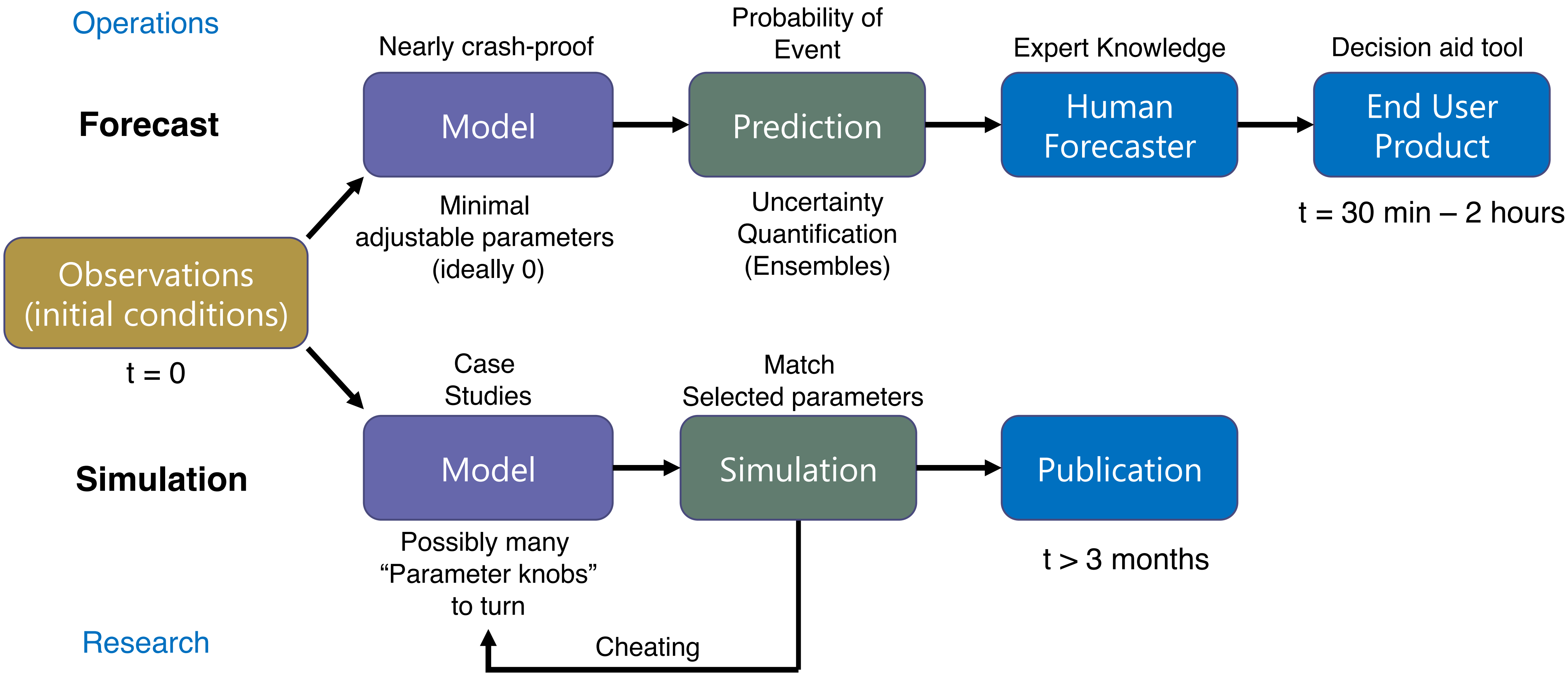
- A well-characterized, consistent (i.e., reliable), quantitative, thoroughly tested, estimate of future conditions that is used to make valuable (i.e., *actionable*) decisions.

- Contrast to **Simulation:** usually a large physics-based numerical model output whose primary use is as a *representation* of a complex system for the purpose of research.

An ***Operational* Forecast** is produced with models and/or observations that:

1. Have undergone documented statistical **Validation & Verification** ("V&V"), i.e., tested over many "re-forecast" runs on historical events before being used in operations.
2. Are available **continuously** ("24/7/365") and with **fail-safe backup** capabilities.
3. Have associated **products** with standardized formats and distribution channels.
4. Are processed, checked, and disseminated by trained specialists who are often **not scientists**.
5. Are continuously **monitored** for performance relative to established metrics. "How well has this forecast been doing (e.g., in terms of prediction of onset time) over the past N geomagnetic storms?"
6. Are **used** by people who make critical decisions.

What is a “forecast”?



Quantification of Space Weather Events

1. Solar Flares

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 positioning for several hours on the sunlit side of Earth, which may spread into the night side.	X20 (2×10^{-3})	Less than 1 per cycle
R 4	Severe	**HF Radio:** HF radio communication blackout on most of the sunlit side of Earth for one to two hours. HF radio contact lost during this time. **Navigation:** Outages of low-frequency navigation signals cause increased error in positioning for one to two hours. Minor disruptions of satellite navigation possible on the sunlit side of Earth.	X10 (10^{-3})	8 per cycle (8 days per cycle)
R 3	Strong	**HF Radio:** Wide area blackout of HF radio communication, loss of radio contact for about an hour on sunlit side of Earth. **Navigation:** Low-frequency navigation signals degraded for about an hour.	X1 (10^{-4})	175 per cycle (140 days per cycle)
R 2	Moderate	**HF Radio:** Limited blackout of HF radio communication on sunlit side, loss of radio contact for tens of minutes. **Navigation:** Degradation of low-frequency navigation signals for tens of minutes.	M5 (5×10^{-5})	350 per cycle (300 days per cycle)
R 1	Minor	**HF Radio:** Weak or minor degradation of HF radio communication on sunlit side, occasional loss of radio contact. **Navigation:** Low-frequency navigation signals degraded for brief intervals.	M1 (10^{-5})	2000 per cycle (950 days per cycle)

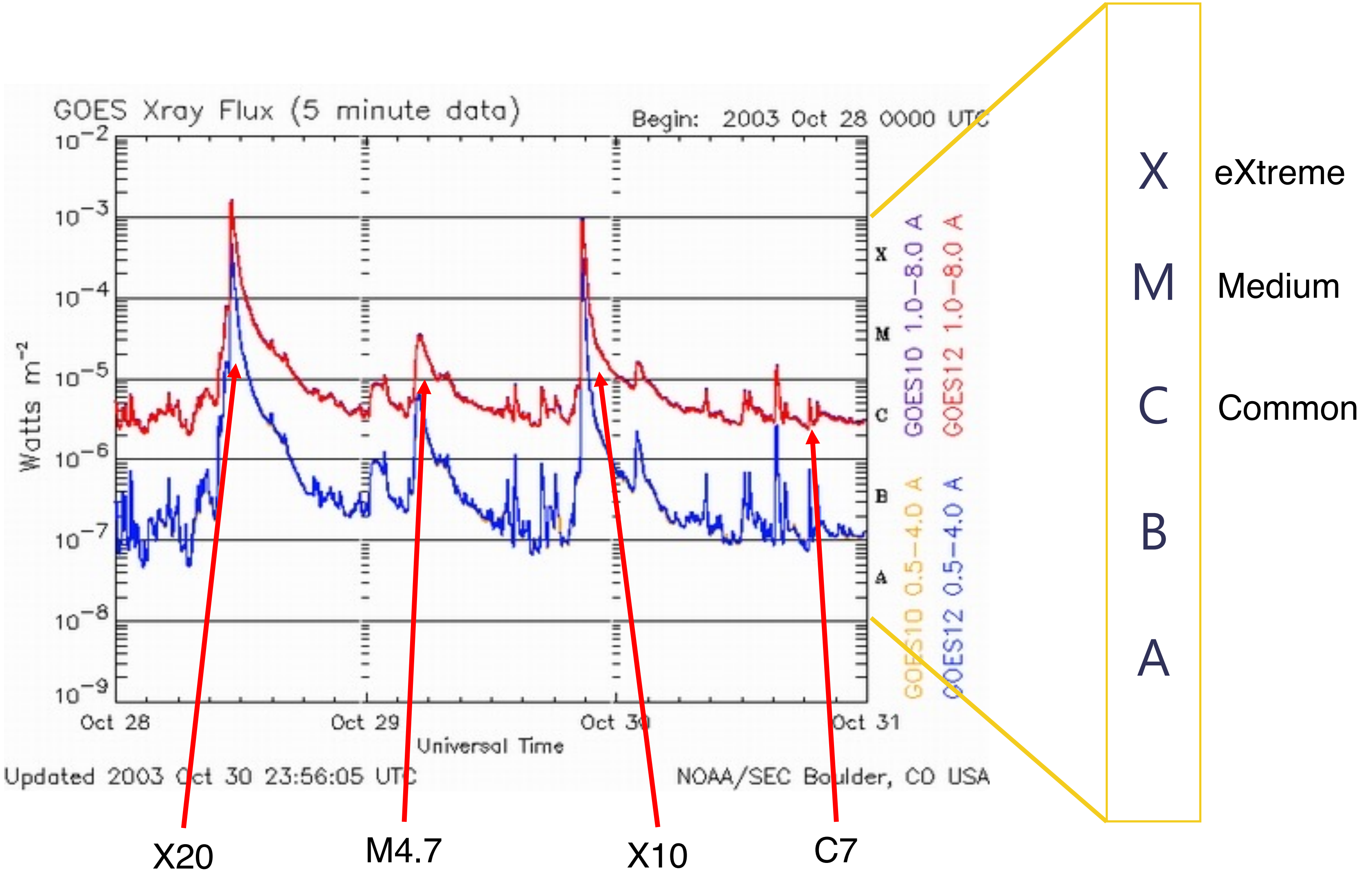
https://www.swpc.noaa.gov/noaa-scales-explanation

Based on GOES XRS X-ray irradiance measurements.

Impacted end users:
- Over-the-horizon (OTH) radio and radar operations (Airlines, shipping, search & rescue, national security)
- GPS

Aside: solar flare classification

Flares are *quantified* by peak intensity measured in NOAA GOES X-ray 1.0—8.0 Å full Sun irradiance (red curve).



Quantification of Space Weather Events

2. Solar Energetic Particle (SEP) events. “Radiation Storm”, “Solar Proton Event”

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:** Complete blackout of HF (high frequency) communications possible through the polar regions, and position errors make navigation operations extremely difficult.	10^5	Fewer than 1 per cycle
S 4	Severe	**Biological:** Unavoidable radiation hazard to astronauts on EVA; passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** May experience memory device problems and noise on imaging systems; star-tracker problems may cause orientation problems, and solar panel efficiency can be degraded. **Other systems:** Blackout of HF radio communications through the polar regions and increased navigation errors over several days are likely.	10^4	3 per cycle
S 3	Strong	**Biological:** Radiation hazard avoidance recommended for astronauts on EVA; passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** Single-event upsets, noise in imaging systems, and slight reduction of efficiency in solar panel are likely. **Other systems:** Degraded HF radio propagation through the polar regions and navigation position errors likely.	10^3	10 per cycle
S 2	Moderate	**Biological:** Passengers and crew in high-flying aircraft at high latitudes may be exposed to elevated radiation risk. **Satellite operations:** Infrequent single-event upsets possible. **Other systems:** Small effects on HF propagation through the polar regions and navigation at polar cap locations possibly affected.	10^2	25 per cycle
S 1	Minor	**Biological:** None. **Satellite operations:** None. **Other systems:** Minor impacts on HF radio in the polar regions.	10	50 per cycle

https://www.swpc.noaa.gov/noaa-scales-explanation

Based on GOES SEISS proton flux measurements.

Impacted end users:

- Airlines
- Satellite operations
- Astronaut operations
- Polar radio, radar, and GPS (Polar Cap Absorption)

Quantification of Space Weather Events

3. Geomagnetic Storms

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)

https://www.swpc.noaa.gov/noaa-scales-explanation

Based on Kp:
a 3-hour average of measurements from magnetometer stations between 44—60 degrees representing “planetary” magnitude of magnetic variation.

This is a logarithmic index scale, not a quantitative scalar measure of intensity.

Problem: Kp and G-scale saturate at 9 and 5

- Historical storms have saturated at Kp = 9+ and demonstrated the need to have an unbounded measure

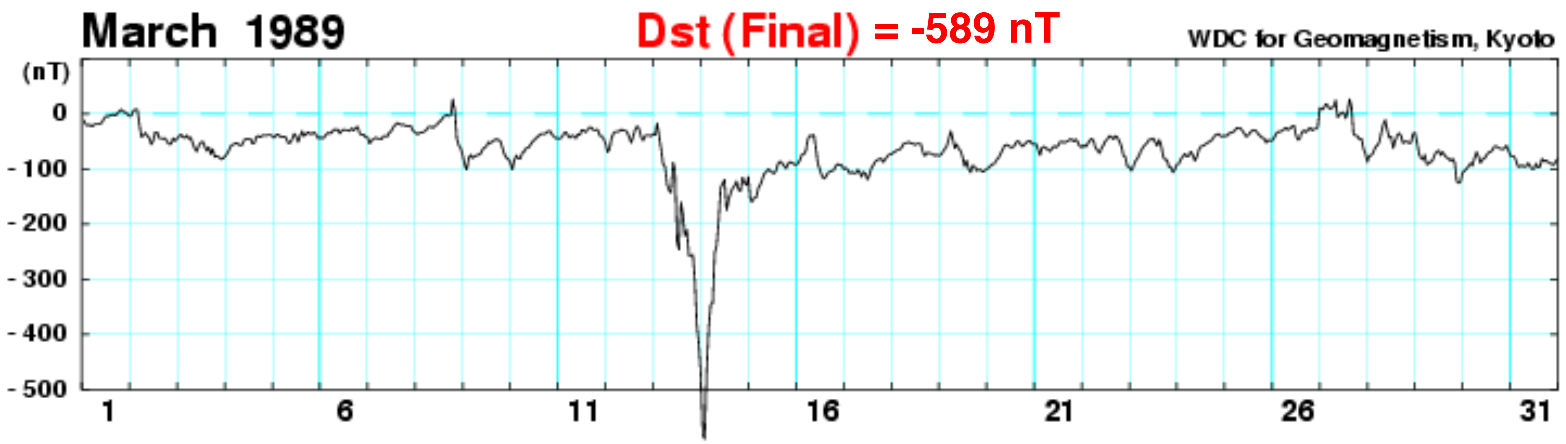
Classification	NOAA G-scale	Kp Threshold	Ap Threshold
Minor	G1	5	40
Moderate	G2	6	65
Strong	G3	7	110
Severe	G4	8 to 9-	180
Extreme	G5	9 to 9+	400

Solution: Disturbance Storm Time (Dst) “index”

Dst

- Derived from average decrease of H component at 4 “equatorial” stations
- Units of **nT:** a **scalar quantity** – not an index scale like Kp.
- A proxy for **Ring Current** intensity.
- Cadence = 1 hour.
- Definitive data: http://wdc.kugi.kyoto-u.ac.jp/dstdir/

March 1989 Superstorm: Kp = 9+



https://wdc.kugi.kyoto-u.ac.jp/dst_final/198903/index.html

Solution: Dst allows classification of “SuperStorms”

- Unlimited range to Dst values

Classification	NOAA G-scale	Kp Threshold	Ap Threshold
Minor	G1	5	40
Moderate	G2	6	65
Strong	G3	7	110
Severe	G4	8 to 9-	180
Extreme	G5	9 to 9+	400

***Suggested* classification based on Dst Index**

Follows NOAA terminology up to “extreme” class and extends it to “SuperStorm” class following Lakhina & Tsurutani (2016).

The ART of Forecasting

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

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

To be useful to anyone, a forecast must be

Accurate / Actionable

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 / Relevant

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.

Contrast: What is a Nowcast?

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 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 downlink to NOAA/SWPC forecast office.

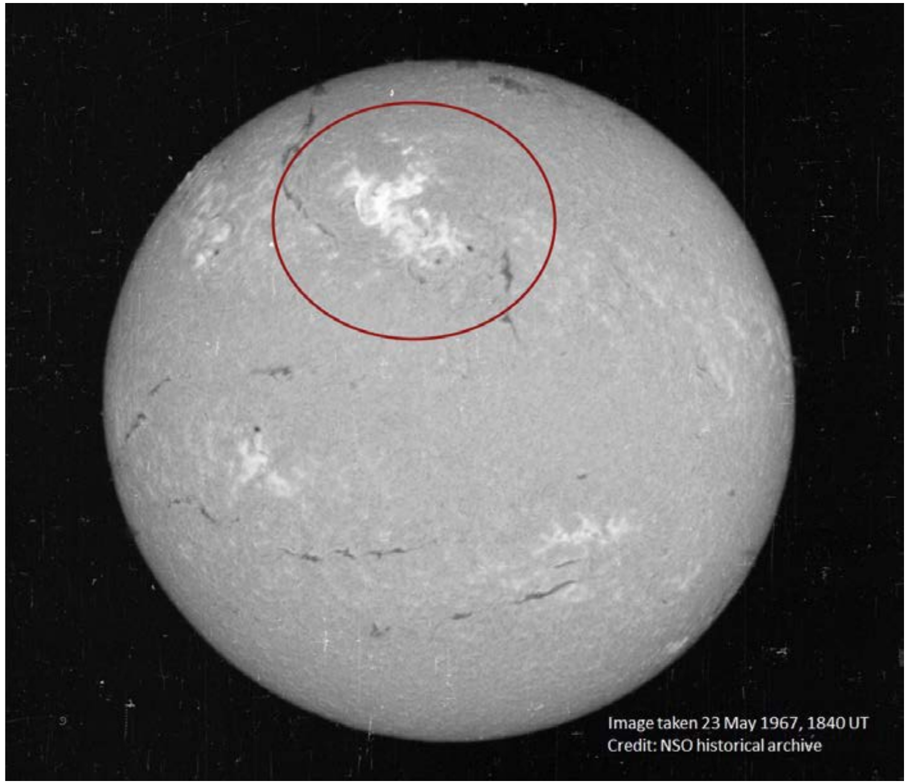
Related: **All Clear** announcement = official statement of event termination and return of safe conditions.

Does not currently exist in operational civil space weather forecasting. Liability is the issue.

However, DOD and NASA are pursuing this capability.

The Value of Nowcasting

- A solar radio burst nearly caused WWIII



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 in Alaska
Sun was low in the Eastern sky at time of flare

Knipp et al., 2016

Products: Watches, Warnings, and Alerts

Watch: “Something has been detected 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 SWx Watch, Warning, and Alert Capability

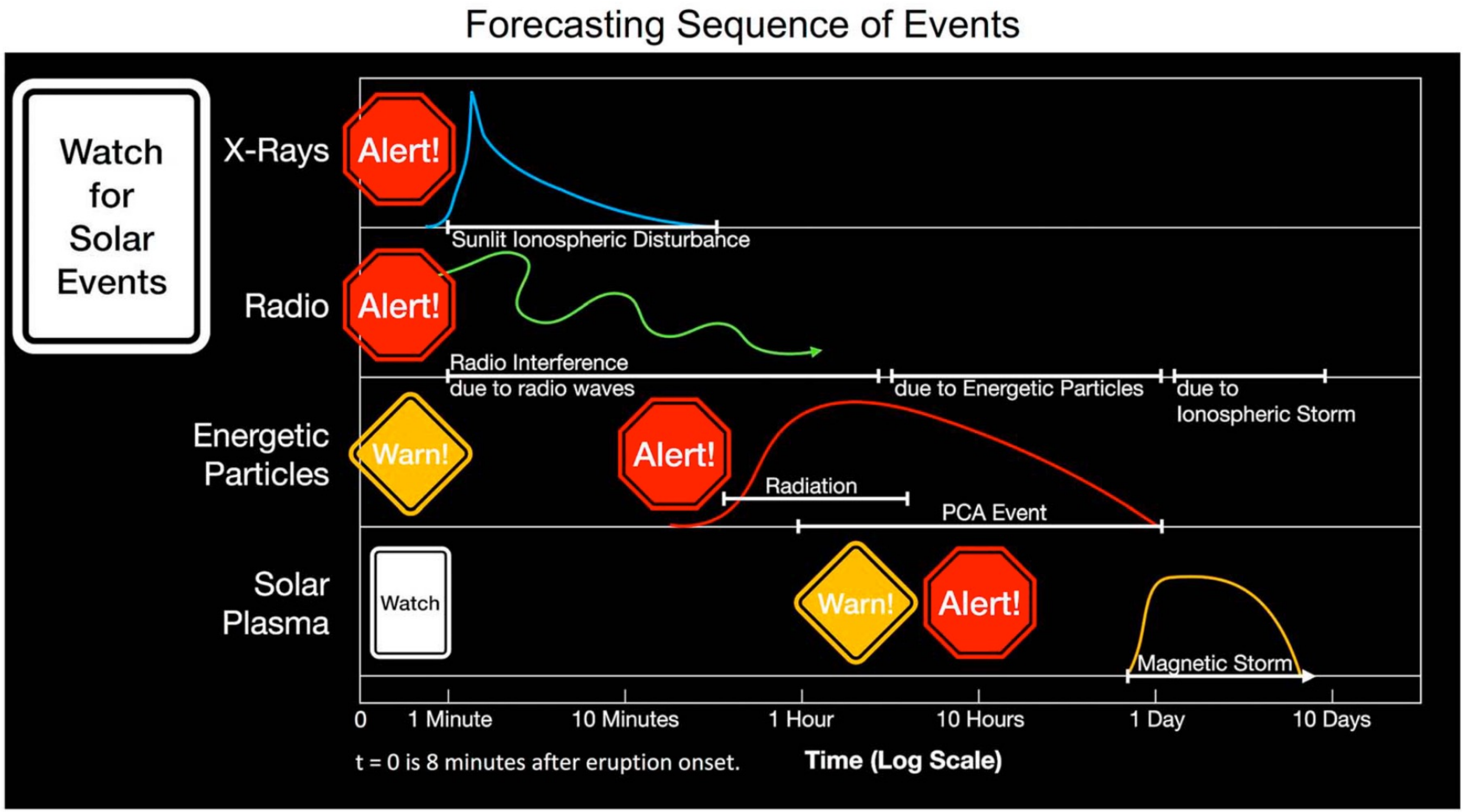
- Obvious lack of solar magnetic eruption forecast capability

Event	Watch	Warning	Alert
Eruption ("flare")			$*_1$
Radiation Storm		$*_2$	$*_3$
Geomagnetic Storm	$*_4$	$*_5$	$*_6$

1. Issued when GOES/XRS X-ray intensity crosses $5x10^{-5}$ W/m^2 = M5.0 = R2 on SWPC scale.
2. Issued depending on flare location and magnitude – up to 15—30 minutes
3. Issued on crossing 10 particles cm^{-2} s^{-1} sr^{-1} for 10 MeV protons = S1 on SWPC scale.
4. Based on analysis of coronagraph imagery and WSA/Enlil model runs to determine if CME will hit Earth.
5. Issued once CME is detected by instruments at L1 Lagrangian point, $1.5x10^6$ km upstream of Earth. Typically gives 30—45 minutes warning before impact.
6. Issued once geomagnetic disturbance is detected in ground-based magnetometer networks.

Current SWx Watch, Warning, and Alert Capability

- Obvious lack of solar magnetic eruption forecast capability



Redmon et al., 2018

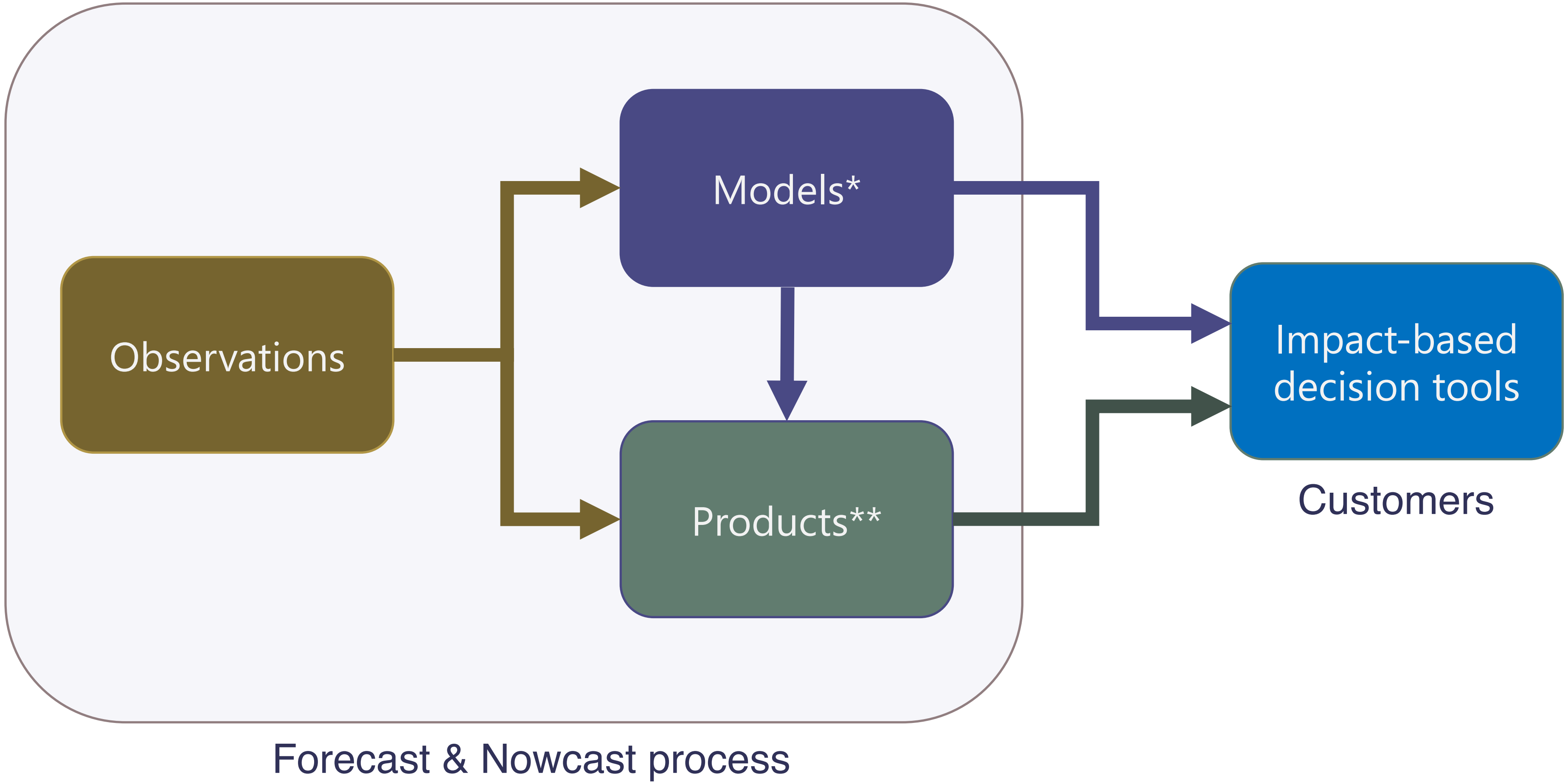
Solar Magnetic Eruptions: Earthquake or Volcano?

Adapted from USGS: https://volcanoes.usgs.gov/vhp/forecast.html

Solar Magnetic Eruptions: Earthquake or Volcano?

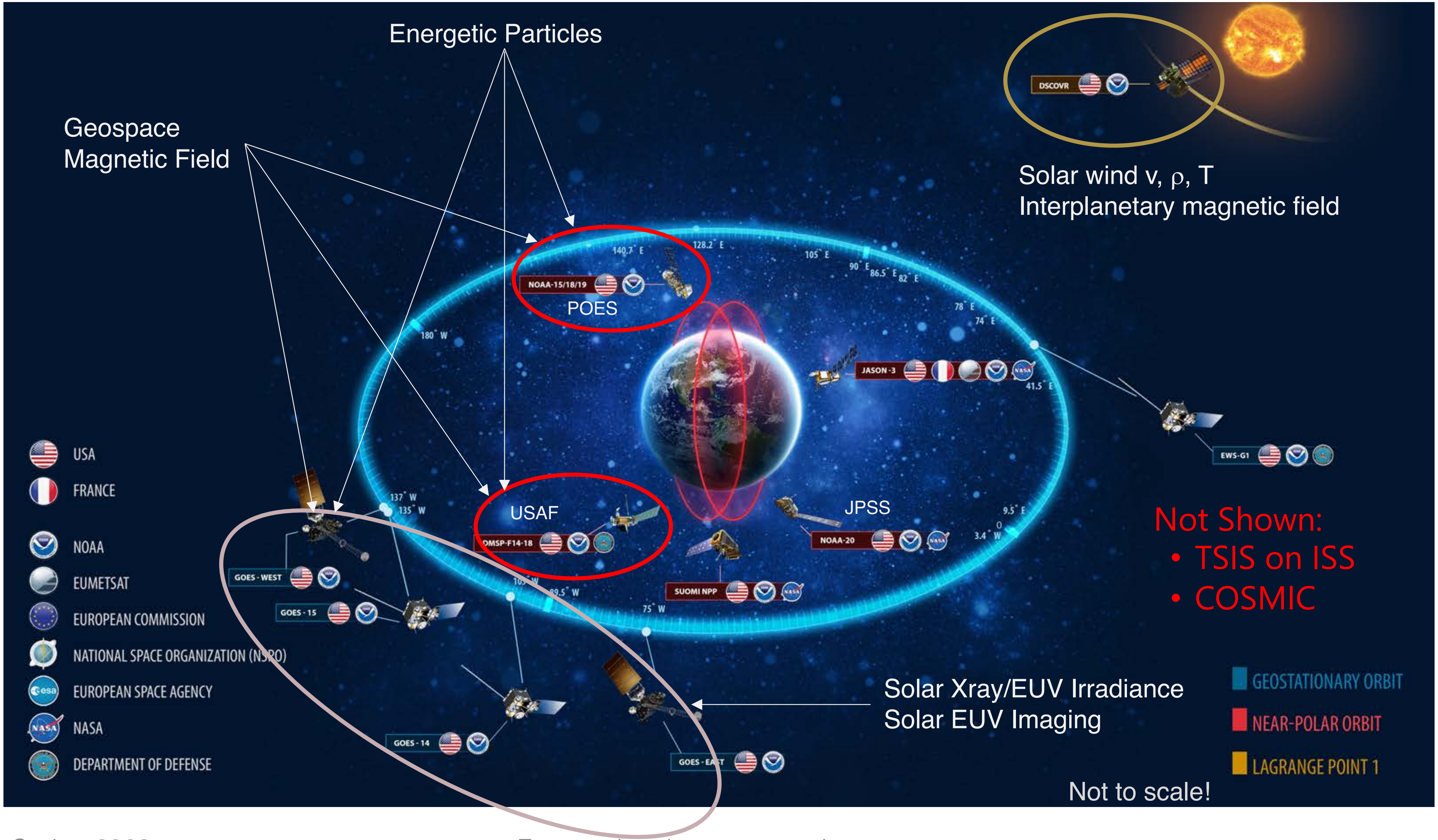
Adapted from USGS: https://volcanoes.usgs.gov/vhp/forecast.html

SWx Forecasting Information flow

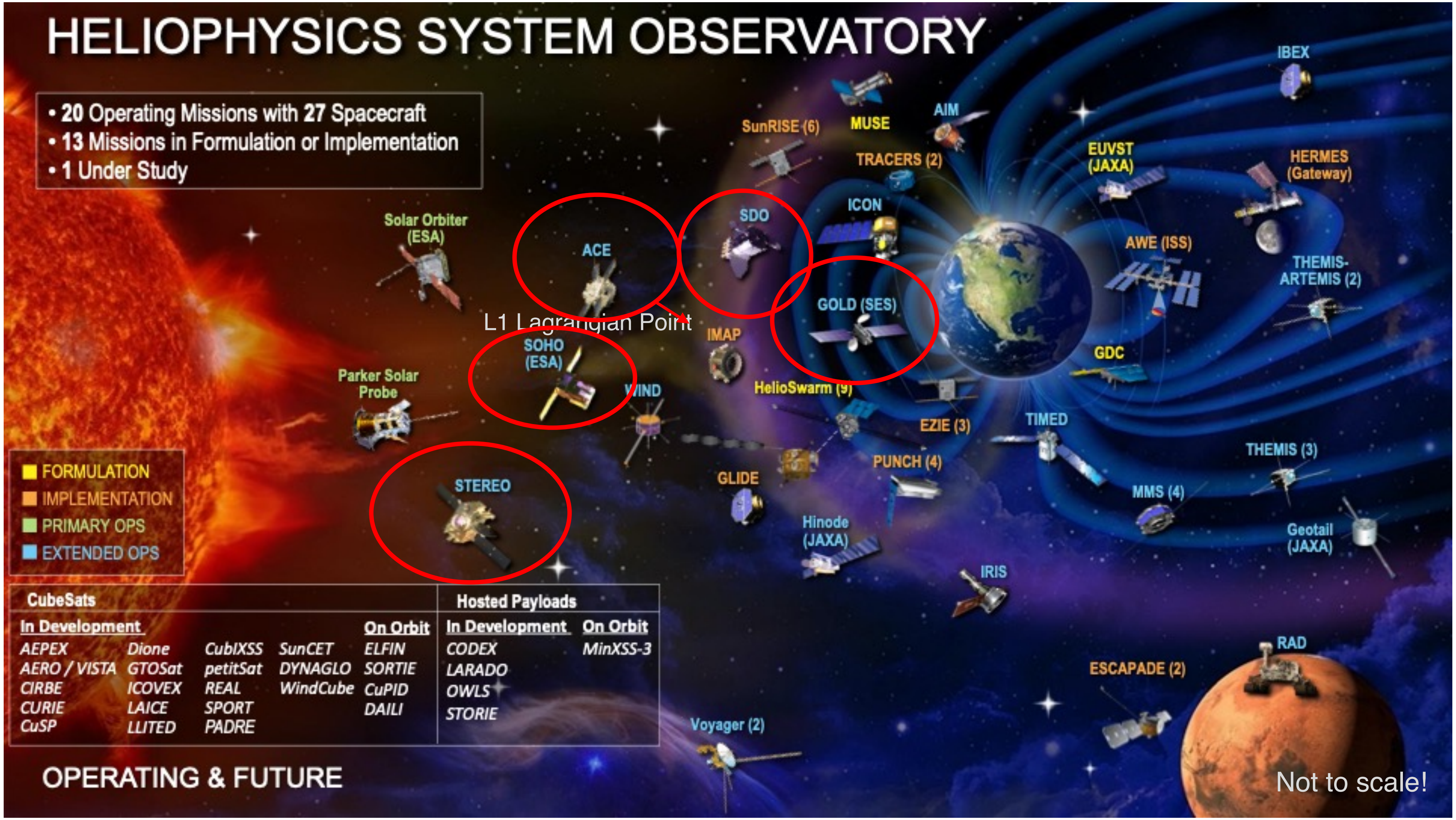


* Complex, multi-input, empirical or physics-based environment models.
** Watches, Warnings, Alerts, nowcasting products like real-time solar radio spectrum, etc.

NOAA SWx Operational Observations



NASA SWx Measurements in the Heliosphere



Used in operational forecasting.

SOHO: (L1 orbit)
- LASCO C2 & C3 coronagraphs

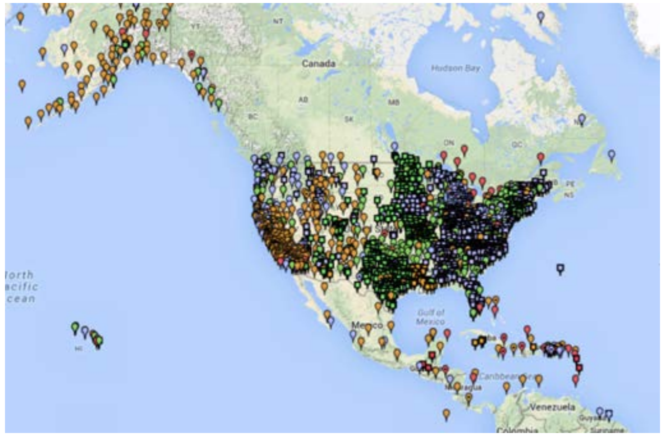
ACE: (L1 orbit)
- backup solar wind and magnetic field data to NOAA/ DSCOVR satellite

SDO: (GEO orbit)
- HMI photospheric magnetic field and continuum movies
- AIA chromospheric and coronal movies

STEREO-A: (Helio orbit)
- COR-1 & 2 coronagraphs
- Energetic particle data
- Magnetic field

Civil Space Weather Measurements on the Ground

USGS Magnetometer Network
Geomagnetic Data



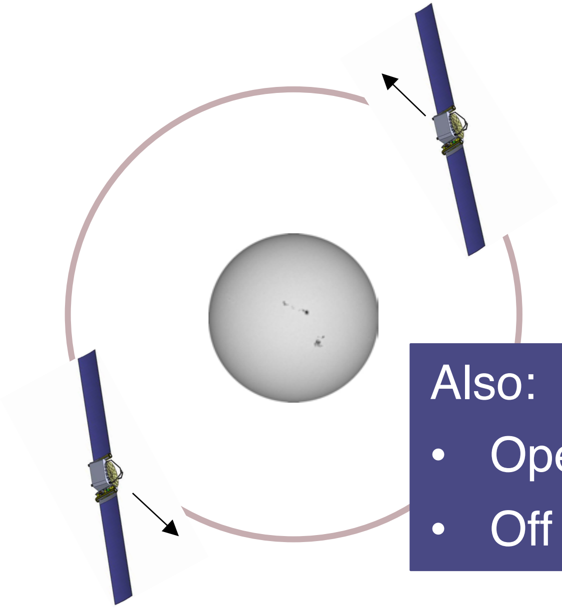
USCG CORS Network
Ionospheric TEC Data

IGS RTIG Network
Ionospheric TEC Data

NSF GONG Network
Solar Magnetogram and Hα Images

Neutron Monitor Network
Ground-level Radiation Event Detection

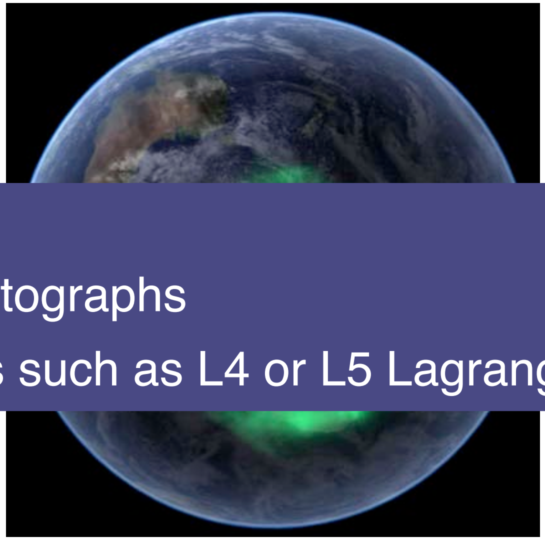
Operational Space Weather Observations: what’s missing?



Solar Polar Monitors
Accurate coronal magnetic field models
CMEs at *all* ecliptic angles (Earth, Mars, etc.)
Long-term Helioseismology: solar cycle

Radiation Belt Monitors
Real-time radiation belt monitoring

Thermospheric Sentinels
Real-time density & composition.
Storm-time IR cooling response.
...orecast model
...s.

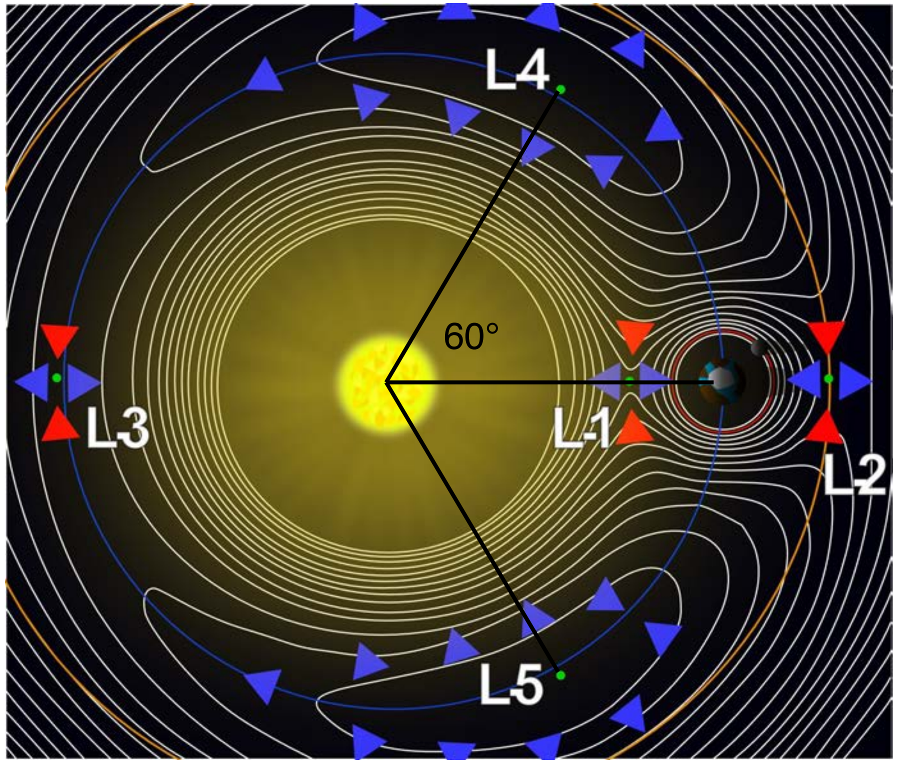


Real-time Auroral Imaging
Ovation model has no sub-storm structure.
More detailed monitoring for
polar GNSS, communications, and Auroral Tourism

Also:

- Operational space-based solar magnetographs
- Off Sun-Earth Line (SEL) observations such as L4 or L5 Lagrangian point stations

Aside: Off Sun-Earth line observations: Lagrangian points, etc.



By Lagrange_points.jpg: created by NASAderivative work: Xander89 (talk) - Lagrange_points.jpg, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=7547312

L1

- ESA/NASA SOHO (coronagraph)
- NASA ACE (solar wind)
- NOAA DSCOVR (solar wind)
- NASA IMAP (solar wind, 2025)

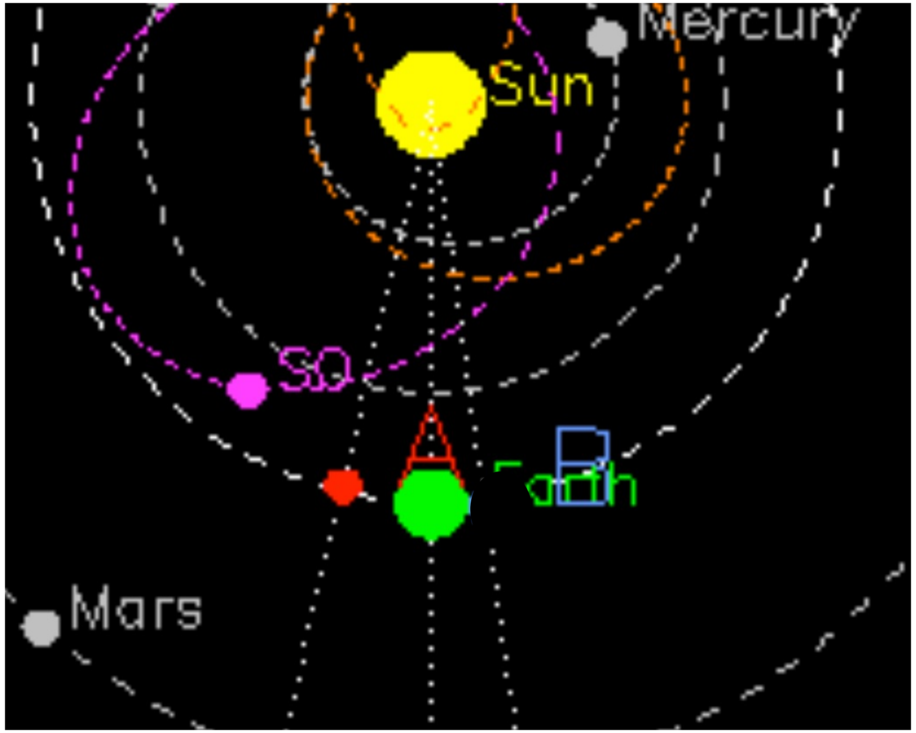
L2

- JWST

L5

- ESA Virgil (2030)

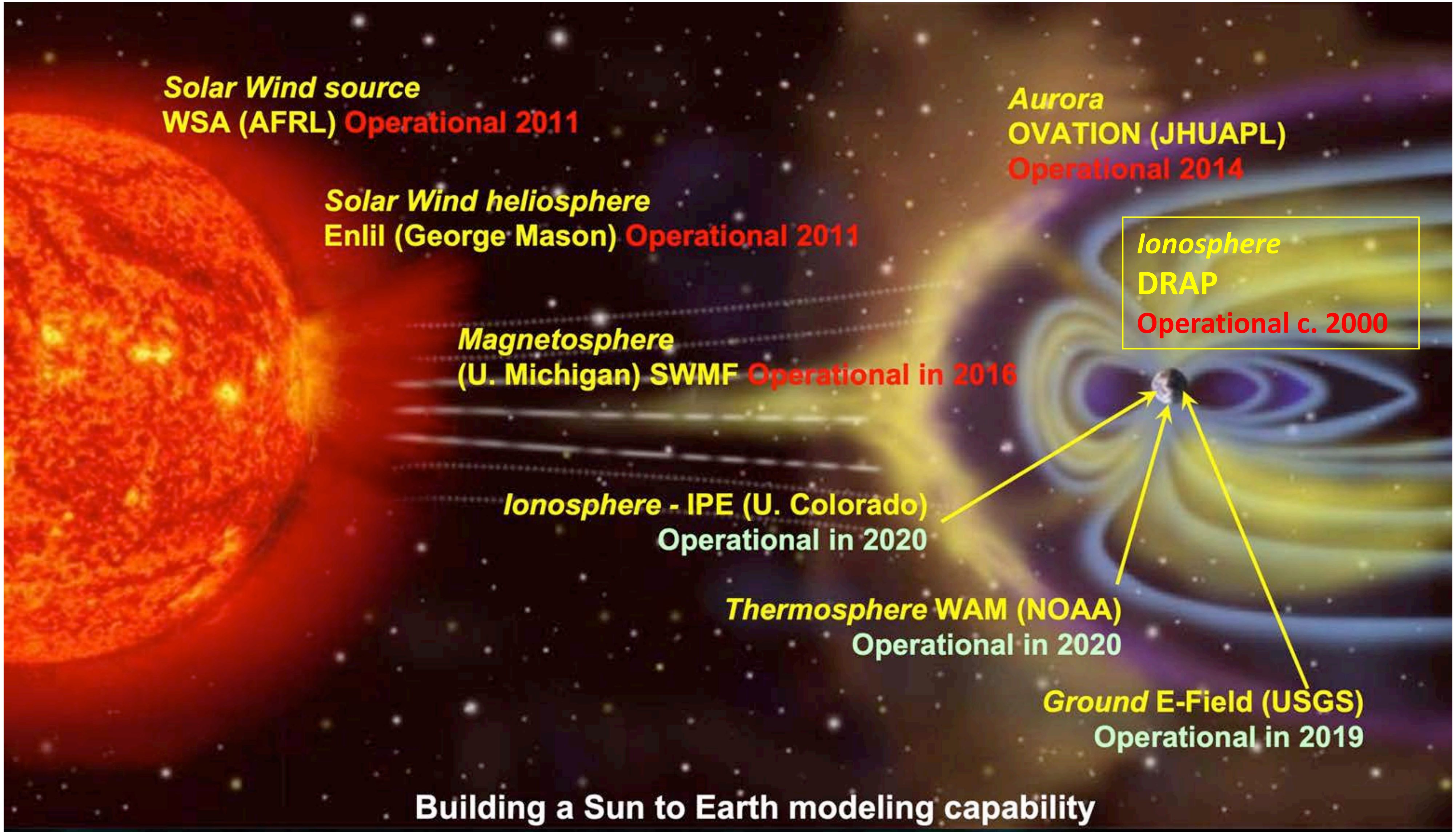
What about STEREO?



STEREO orbit relative to Earth ~18 years

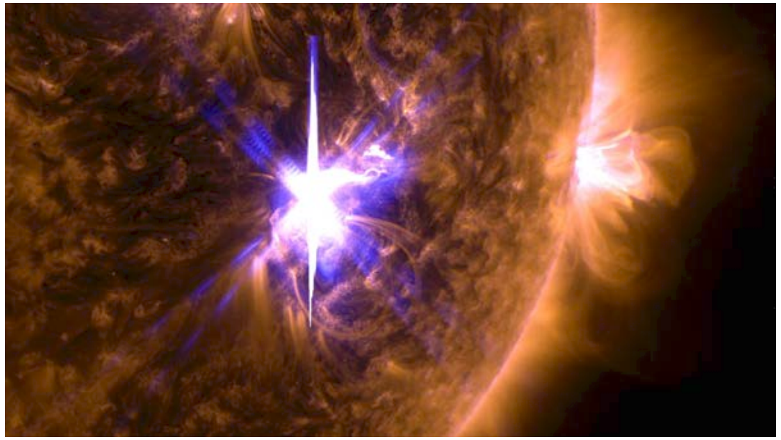
Orbital mechanics fun: you can park a S/C anywhere at 1 AU (i.e., you don't need to orbit only at L4 or L5)

NOAA Operational Forecasting Models

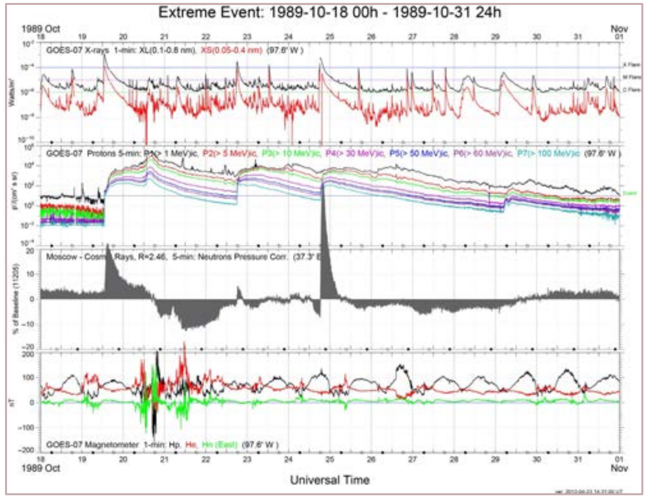


Adapted from E. Talaat, NAS SWx Infrastructure Workshop I, 16-June-2020

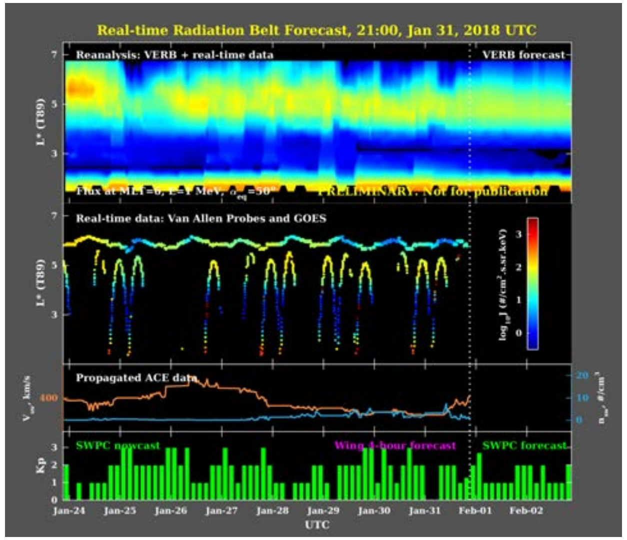
Missing Operational Space Weather Models/Products



Solar Eruption Forecasting Model



SEP Event Forecasting Model



VERB model, courtesy A. Kellerman (UCLA)

Radiation Belt Forecasting Model

Energetic particle flux at any orbit of interest

Whole Atmosphere + Satellite Drag Model

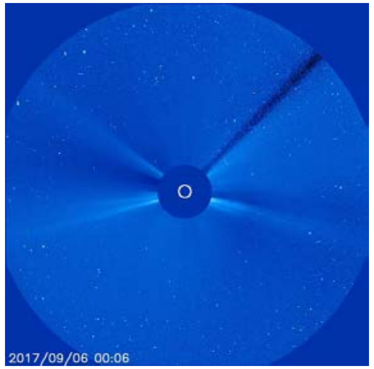
Geomagnetic storm satellite positioning prediction

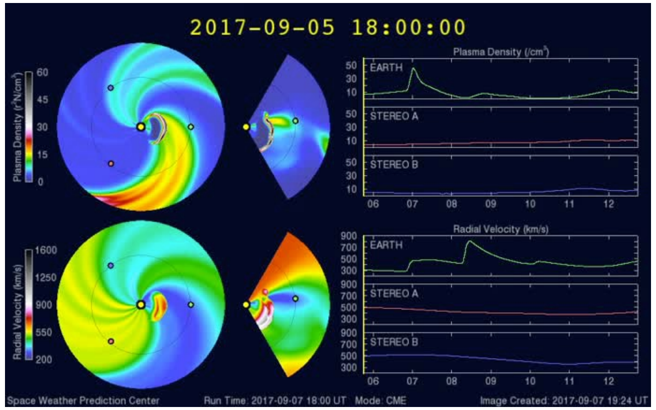
ALSO

- Data assimilation into full-physics operational forecasting models
- Ensemble forecasting systems

Operational Solar Wind & CME Arrival Products

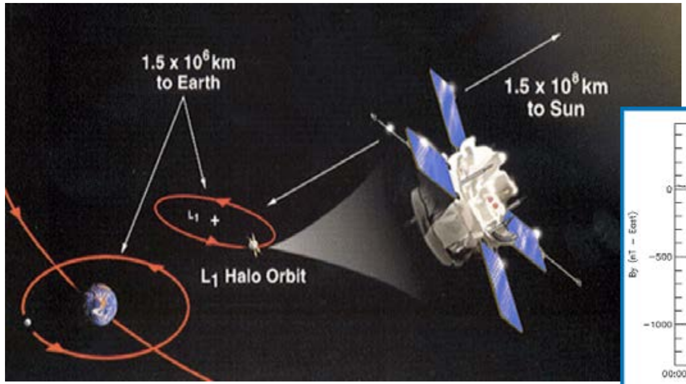
- Timing of products in relation to observations.

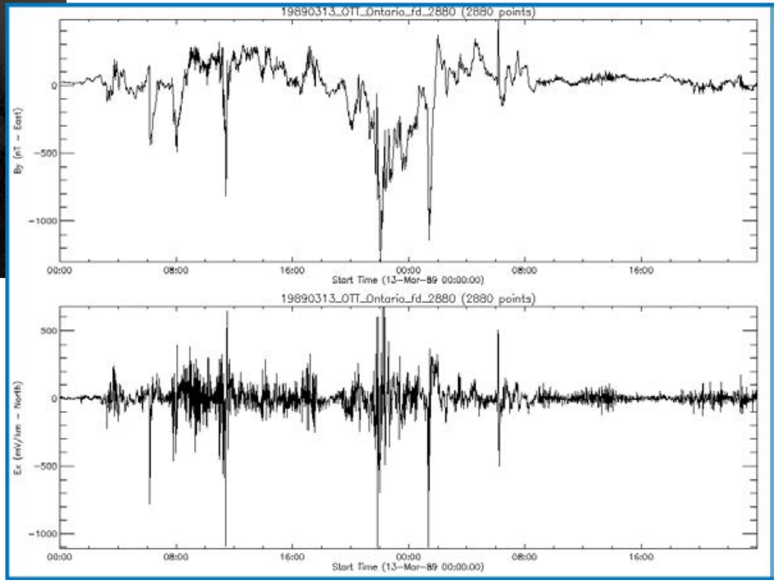




1. Geomagnetic Storm Watch
Issued upon coronagraph detection of Earth-directed CME and WSA-Enlil model run.

- 15 – 60 hours before impacting Earth (if at all)





2. Geomagnetic Storm Warning
Issued upon detection at DSCOVR or ACE spacecraft at the L1 Lagrange point.

- 15-60 minutes before impacting Earth

3. Geomagnetic Storm Alert
Issued when geomagnetic storm is detected on USGS ground-based magnetometers

- Current condition

NOAA Forecasts and Reports

- Best Summary Product: Report and Forecast of Solar and 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

- Report and Forecast of Solar and Geophysical Activity: 30 Aug 2021

```
: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
```

AR numbers have 10,000 subtracted

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).

Official NOAA Space Weather Forecasts & Reports

- Report and Forecast of Solar and Geophysical Activity: 30 Aug 2021 (cont.)

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.

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).

Official NOAA Space Weather Forecasts & Reports

- Report and Forecast of Solar and Geophysical Activity: 30 Aug 2021 (cont.)

```
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
```

```
VI.  Geomagnetic Activity Probabilities 31 Aug-02 Sep
A.  Middle Latitudes
Active                 20/40/35
Minor Storm            05/25/30
Major-severe storm     01/05/10

B.  High Latitudes
Active                 15/10/10
Minor Storm            30/25/25
Major-severe storm     30/65/65
```

PCAF = Polar Cap Absorption Forecast. Stoplight quantification: Red = bad, Yellow = not so bad, Green = no issue.

Forecast Evaluation

- How do you judge a forecast?
- How well do current space weather forecasts perform?

Two main types of forecasts in space weather:

1. Event-based: “**binary classification**” – will something happen or won’t it?
2. **Probabilistic**: what is the probability that some event will happen?

Which is more actionable and reliable? What does your weather app give you?

Forecast Evaluation: binary classification

- **Contingency Table**

- **Binary categorical tasks produce either "True/Positive" or "False/Negative" results.**
  "AR 10973 will flare in the next 24 hours" (P) or "AR 10973 will not flare in the next 24 hours" (N).

- **Contingency Tables are used to create Skill Scores based on the relative numbers of**
  - True Positive (TP)
  - False Positive (FP)
  - True Negative (TN)
  - False Negative (FN)

Forecast \ Observed	P	N
P	"Hit" **TP**	"False Alarm" **FP**
N	"Miss" **FN**	"Nothing happened" **TN**

Woodcock, 1976

Forecast Evaluation Metrics

- Some of the many binary classification **Skill Scores**

- **Accuracy: [0,1]** $A_{\text{forecast}} = \frac{TP + TN}{N}$ *N* here is total number of events: *N* = TP+TN+FP+FN
- **Probability of Detection (POD) or "Hit Rate": [0,1]** $\text{POD} = \frac{TP}{TP + FN}$
- **False Alarm Rate (Prob of False Detection): [0, 1]** $\text{FAR} = \frac{FP}{FP + TN}$
- **True Skill Statistic (TSS): [-1, 1]** $\text{TSS} = \frac{TP \cdot TN - FP \cdot FN}{(TP + FN)(FP + TN)}$
  Also called the Hanssen & Kuipers Skill Score (H&KSS). Not sensitive to the differences in size between event and non-event populations. POD - FAR.

For any Skill Score, the final measure of Skill $= \frac{SS_{\text{forecast}} - SS_{\text{reference}}}{SS_{\text{Perfect}} - SS_{\text{reference}}}$

$SS_{\text{reference}}$ = Score from Reference forecast

SS_{Perfect} = Perfect value for given SS

Forecast Evaluation Metrics

- Common **Reference Forecasts** against which other methods are compared

  - **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.

Current Operational Flare Forecasting Process (in US)

- Human-in-the-loop forecasting is the current state-of-the-art.

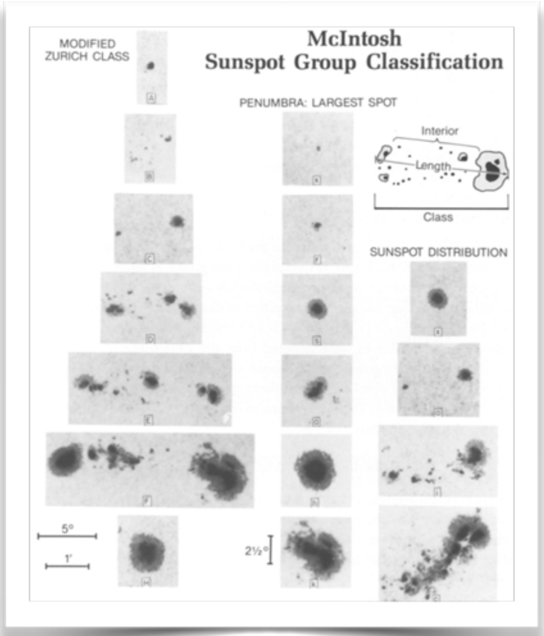
Input data

GONG

SDO/HMI

Continuum & Magnetogram Images

AR Classification System



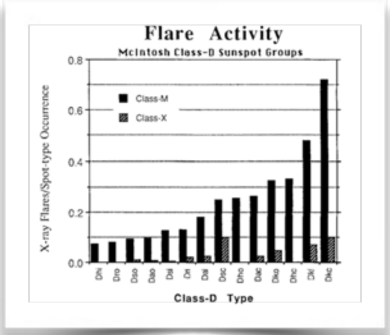
McIntosh, SolPhys, **125**, 251, 1990

McIntosh and/or Mt. Wilson Classification

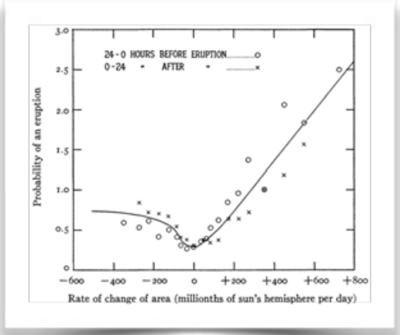
Analysis

Human Forecaster Processes

- Climatology look-up table: P_f for given class
- Growth/decay in spot & total AR area
- Flaring History
- Forecaster expertise



McIntosh, SolPhys, **125**, 251, 1990



e.g., Giovanelli, ApJ, **89**, 555, 1939

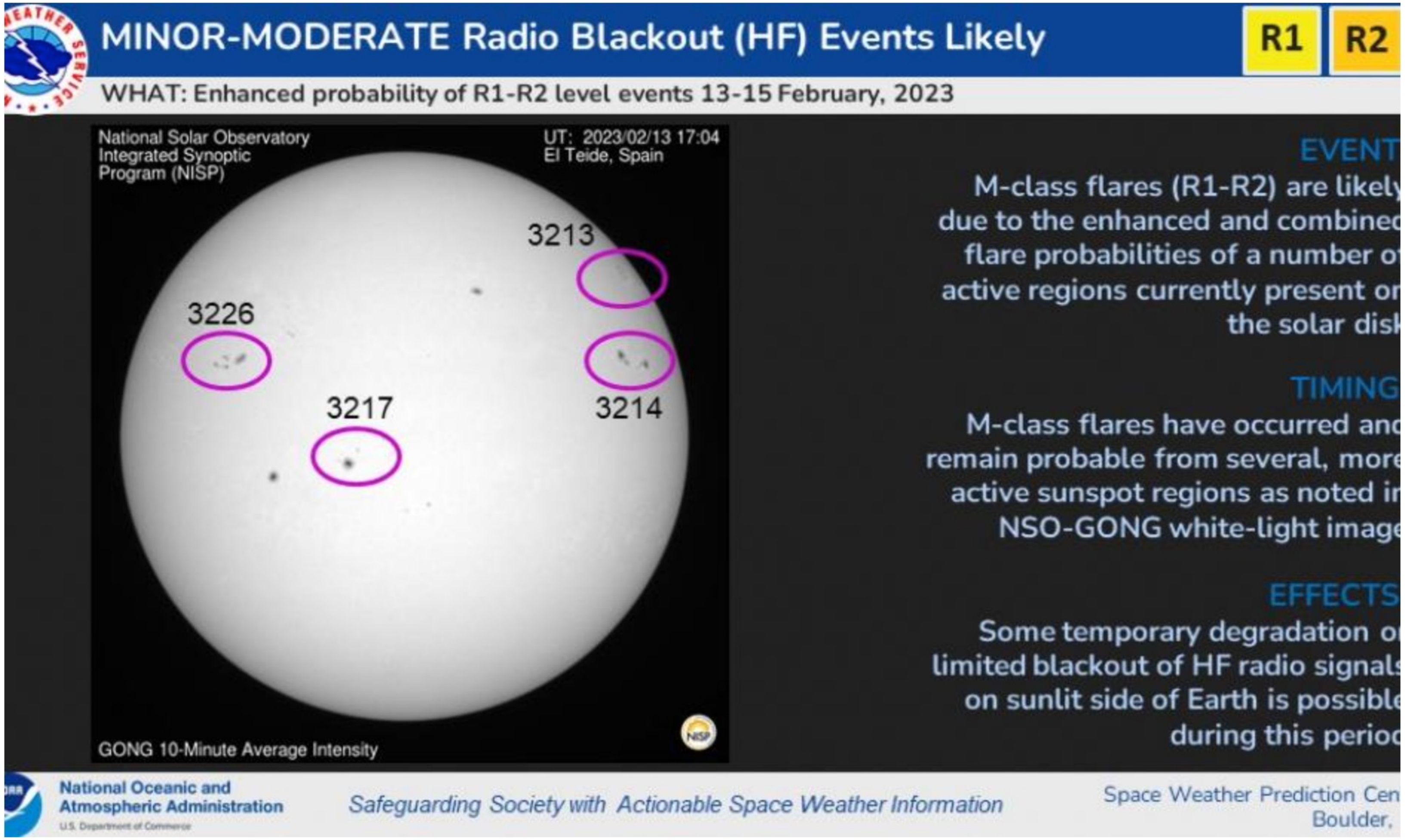
Output

Probability of X-ray flare of magnitude C,M,X

P_f (n)

n = 24, 48, 72 hours

From probabilities to language: likely?



From the RFSGA:

III. Event probabilities 15 Feb-17 Feb

Class M 55/50/50
Class X 10/10/10
Proton 10/01/01
PCAF green

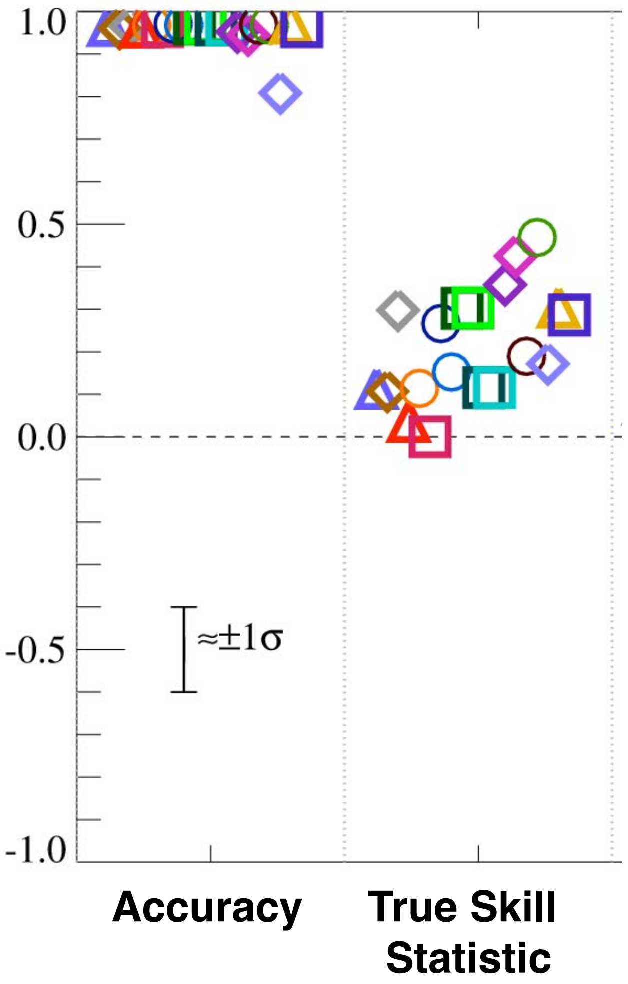
Apparently, "likely" ≥ 50% prob.

Flare Forecasting: stuck in neutral

- Comparison study on binary event classification – easier than doing probabilistic validation.
- Will there be an M1 or greater flare in the next 24 hours?

High Accuracy!

(because
TN ~ N >> TP)



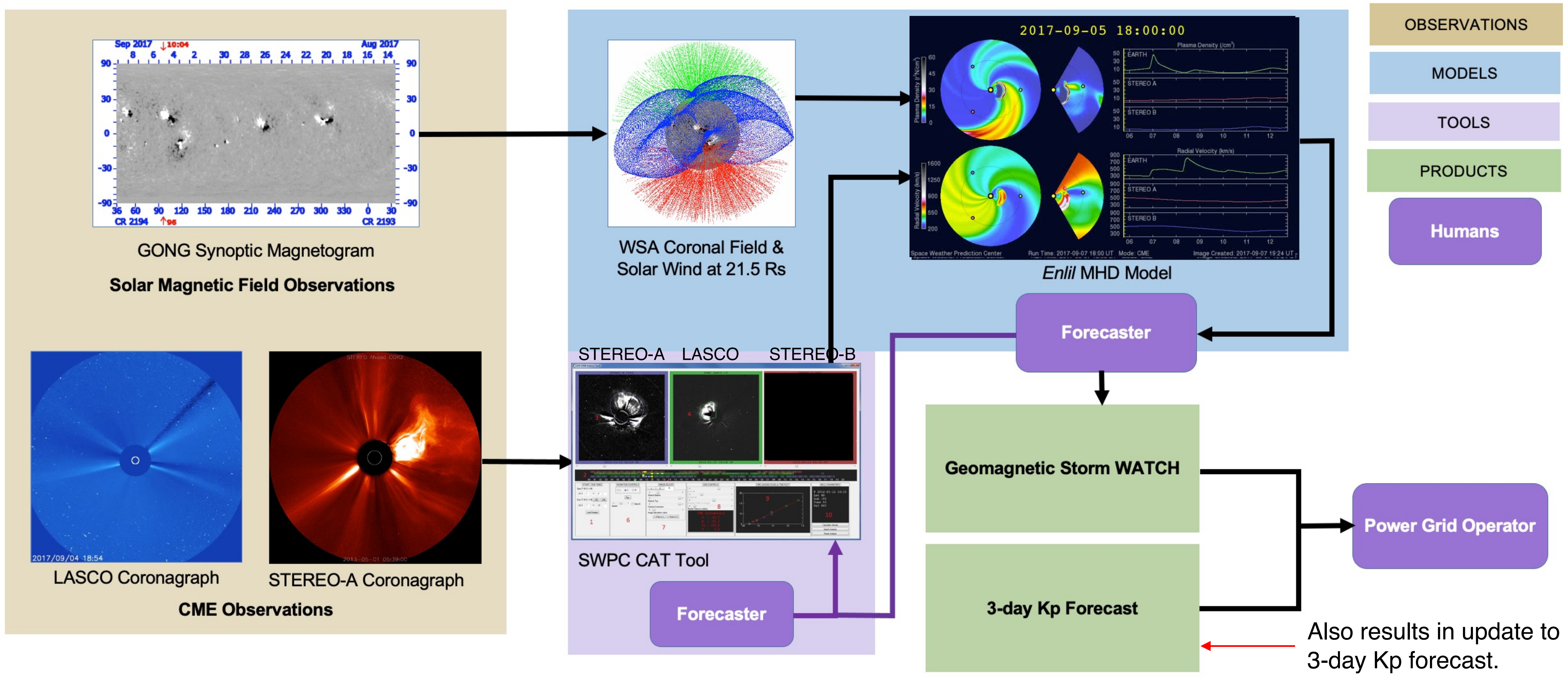
Huge False Alarm Rate

Leka et al., ApJSS (2019)

Label	Symbol	Reference(s)
A-EFFORT	△	Georgoulis & Rust (2007)
AMOS	◇	Lee et al. (2012)
ASAP	◇	Colak & Qahwaji (2008, 2009)
ASSA	△	Hong et al. (2014), Lee et al. (2013)
BOM	○	Steward et al. (2011, 2017)
CLIM120	□	Sharpe & Murray (2017)
DAFFS	○	Leka et al. (2018)
DAFFS-G	○	" "
MAG4W	□	Falconer et al. (2011);
MAG4WF	□	also see Appendix A
MAG4VW	□	
MAG4VWF	□	
MCSTAT	◇	Gallagher et al. (2002), Bloomfield et al. (2012)
MCEVOL	◇	McCloskey et al. (2018)
MOSWOC	○	Murray et al. (2017)
NICT	○	Kubo et al. (2017)
NJIT	◇	Park et al. (2010)
NOAA	△	Crown (2012)
SIDC	□	Berghmans et al. (2005), Devos et al. (2014)

Operational Solar Wind & CME Arrival Forecasting

- Primary geomagnetic Storm forecast process. Generates 3-day forecast and watch product.

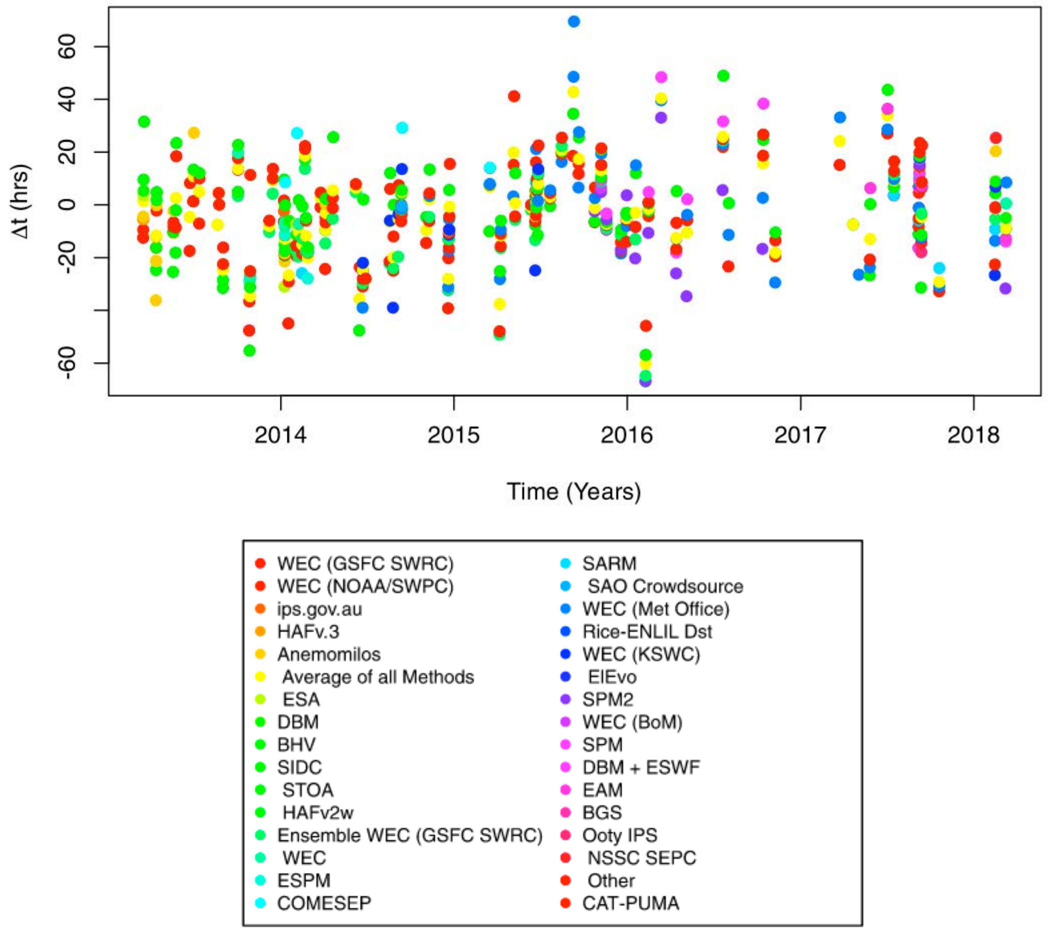


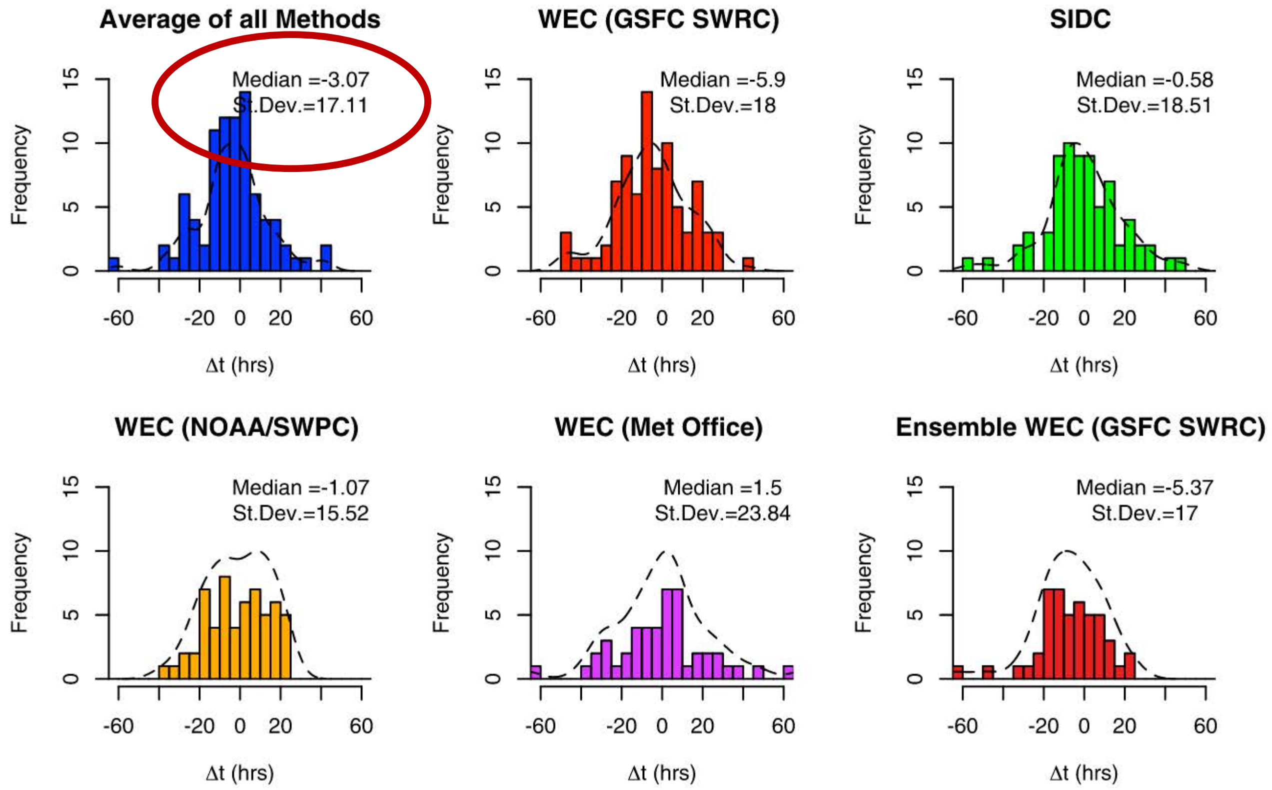
CME arrival time forecasting

- Note: binary event forecast. Will the CME hit Earth at time *t* ?

Aug 4, 1972 CME arrived in 14 hours

±17 hrs: not a reliable forecast

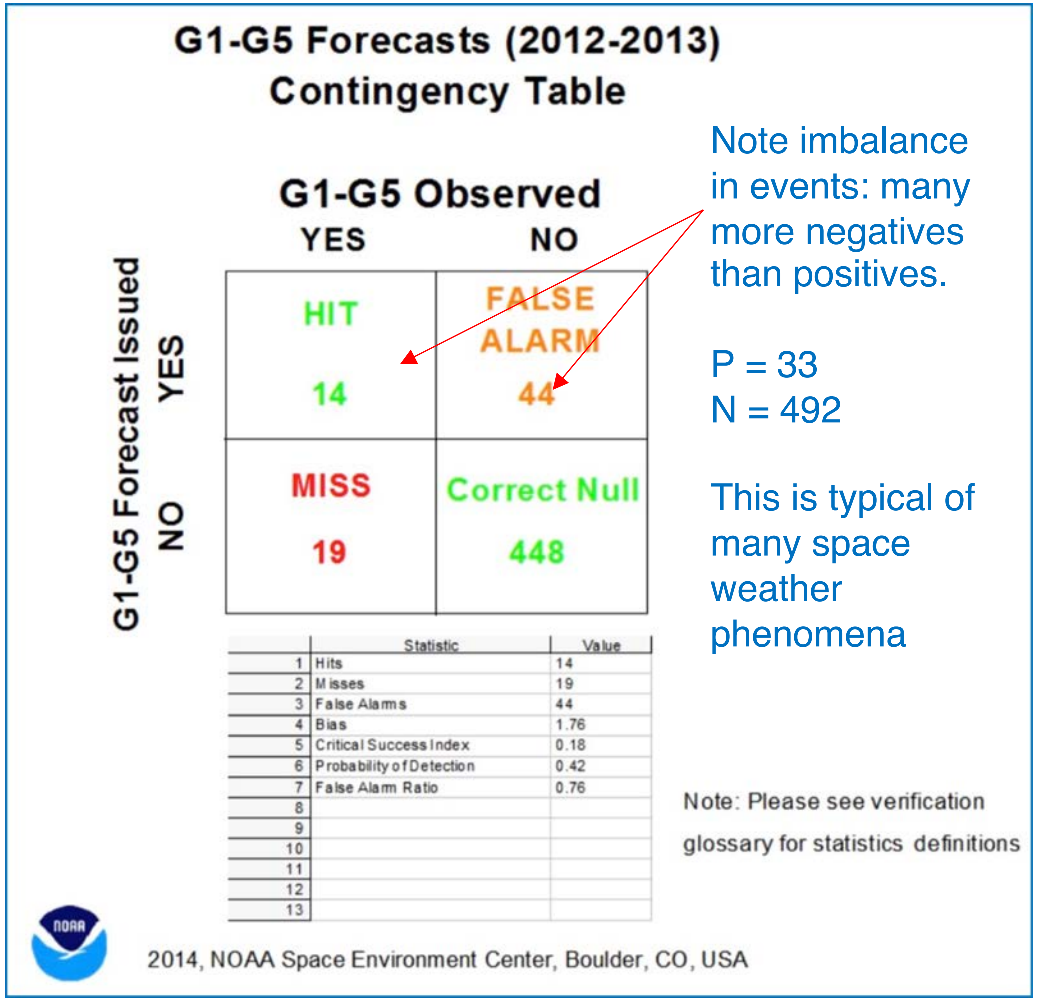




Riley et al., 2018

Forecast Evaluation: NOAA Geomagnetic Storm Forecasting

- Binary classification: G1 or greater geomagnetic storm or no storm.



	Statistic	Value
1	Hits	14
2	Misses	19
3	False Alarms	44
4	Bias	1.76
5	Critical Success Index	0.18
6	Probability of Detection	0.42
7	False Alarm Ratio	0.76
8		
9		
10		
11		
12		
13		

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:

- Percentage of times that the 24-hour geomagnetic storm forecast is correct for the 60 most recent geomagnetic storms.

Definition of a "Hit" for geomagnetic storm forecasting:

- G1 or greater was forecast and *G1 or greater* occurred. i.e., they are *not accounting for the magnitude of the storm*. Essentially a binary "storm/no-storm" forecast.

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

$$A_{\text{climatology}} = \frac{\text{FN} + \text{TN}}{\text{N}} = 0.95$$

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**

Future developments

Machine learning (data-driven) models: 50,000 times faster than physics-based numerical simulation models.

Equally skillful?

FOURCASTNET: A GLOBAL DATA-DRIVEN HIGH-RESOLUTION WEATHER MODEL USING ADAPTIVE FOURIER NEURAL OPERATORS

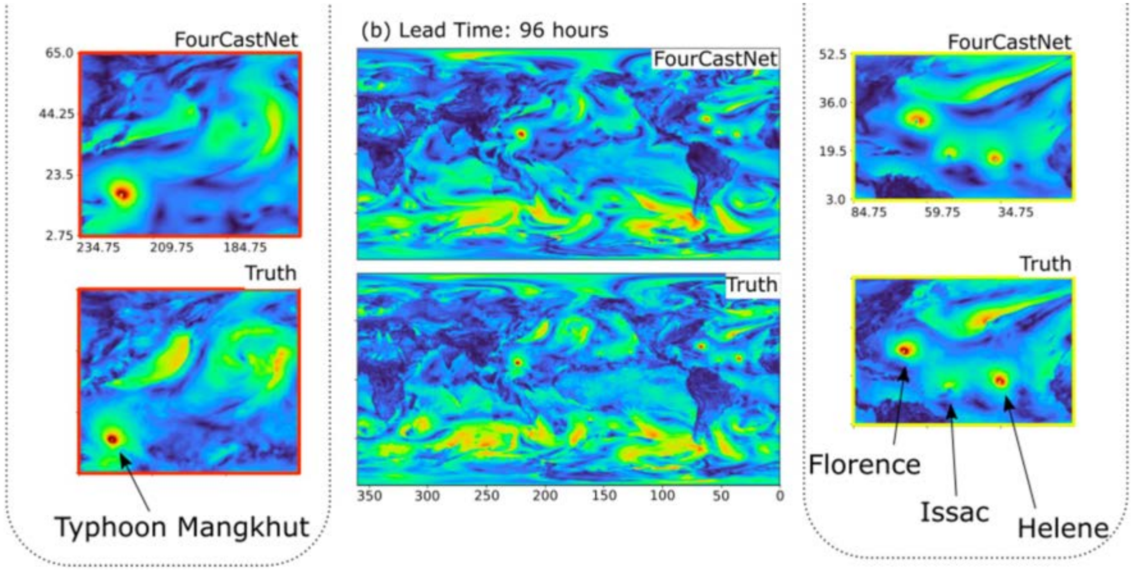
A PREPRINT

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109



global near-surface wind forecast generated by FourCastNet over the entire globe at a resolution of 0.25° compared to the ECMWF NWP model (labeled as “Truth”).

Questions? Comments?

Homework:

1. Show that the True Skill Statistic (TSS) is equal to the difference between the Probability of Detection (POD) and the False Alarm Rate (FAR).
2. Monitor the SWPC website for the 10 days and develop a rudimentary Contingency Tables based on the number of flares above M1 that occurred each day and the number of G1 or greater geomagnetic storms that occurred each day. For the forecast quantities, assume that a probability ≥ 50% is a predicted “True” event. How well did the forecasts do?

Bibliography

Giovanelli, R. G. (1939) ‘Eruptions and Sunspots’, *Astrophysical Journal*, 89(5), pp. 555–567.

Knipp, D.J. *et al.* (2016) ‘The May 1967 great storm and radio disruption event: Extreme space weather and extraordinary responses: MAY 1967 SOLAR AND GEOMAGNETIC STORM’, *Space Weather*, 14(9), pp. 614–633. Available at: https://doi.org/10.1002/2016SW001423.

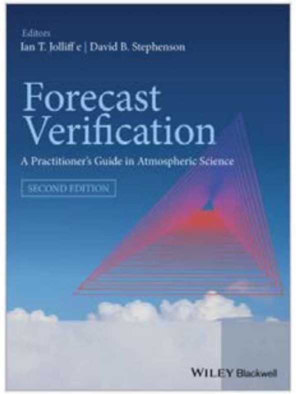
McIntosh, P.S. (1990) ‘The classification of sunspot groups’, *Solar Physics*, 125(2), pp. 251–267. Available at: https://doi.org/10.1007/BF00158405.

Redmon, R.J. *et al.* (2018) ‘September 2017’s Geoeffective Space Weather and Impacts to Caribbean Radio Communications During Hurricane Response’, *Space Weather*, 16(9), pp. 1190–1201. Available at: https://doi.org/10.1029/2018SW001897.

Riley, P. *et al.* (2018) ‘Forecasting the Arrival Time of Coronal Mass Ejections: Analysis of the CCMC CME Scoreboard’, *Space Weather*, 16(9), pp. 1245–1260. Available at: https://doi.org/10.1029/2018SW001962.

Woodcock, F. (1976) ‘The Evaluation of Yes/No Forecasts for Scientific and Administrative Purposes’, *Monthly Weather Review*, 104(10), pp. 1209–1214.

The forecast verification bible:



Forecast Verification: A Practitioner's Guide in Atmospheric Science, 2nd Edition

Ian T. Jolliffe (Editor), David B. Stephenson (Editor)

ISBN: 978-0-470-66071-3 | December 2011 | 296 Pages

		Forecast	
		Positive	Negative
Observed	Positive	A Hit *TP*	B Miss *FN*
	Negative	C False Alarm *FP*	D Null *TN*

Events (row Positive); *Non-events* (row Negative)

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
Probability of Missed Detection	$\frac{B}{A+B}$	1 - TPR
Accuracy	$\frac{A+D}{N}$	
False Positive Rate[1]	$\frac{C}{C+D}$	FPR
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	$\frac{2A}{2A+B+C}$	

T. Berger, 2022

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

Ratio Test, R = S – Sf / Sp – Sf

R = (A + D) / N

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