Space Weather Service Network Preliminary Product Validation for the Period of Heightened Activity Observed in September 2017

Sophie Burley[1], Alexi Glover[1], Juha-Pekka Luntama[1], Jesse Andries[2], Claudia Borries[3], Lenka Zychova[4], Mark Dierckxsens[4], Norma Crosby[4], Manolis Georgoulis[5], Guram Kervalishvili[6], Ioanna Tsagouri[7], Peter Wintoft[8], Federico Da Dalt[9], Gabor Facsko[9], Ralf Keil[9]

[1]ESA, [2]ROB, [3]DLR, [4]BIRA-IASB, [5]RCAAM, [6]GFZ, [7]NOA/IAASARS, [8]IRF, [9]RHEA System GmbH for ESA

With around 600,000 monthly hits from over 1000 registered users, the European Space Agency's Space Situational Awareness (SSA) Space Weather (SWE) Portal[3] provides users with access to 29 pre-operational space weather services built upon a large variety of products, tools and alerts, together with expert user support and guidance. Providing reliable information to the end users is of the utmost importance, and to do this, analysing the performance of the different service elements under a range of space weather conditions is vital.

INTRODUCTION

As validation is a requirement across the portal, it is necessary to outline the essential procedures under which the products must be validated. As a result of this, an initial set of **'Guidelines for common validation in the SSA SWE Network'[1]** were developed by a working group consisting of representatives of the SWE Expert Service Centres aiming to provide a baseline for future validation campaigns.

These guidelines provide descriptions for 'categories' of products, and different validation techniques dependent on these categories.

This project tested these guidelines by applying them to five products on the portal: **SIDC CACTus**, **A-EFFort**, **COMESEP Alert Service**, **GFZ Kp Nowcast**, and **IRF Kp Forecast.**

Specifically this project differs from other validation projects in two ways. Firstly, the data tested was taken from a limited time period, the September 2017 event. It is important to note that this significantly limits the amount of data used, and so plots and results may not reflect similar correlation/validation results that would be expected from large datasets. Secondly, the product as a whole was validated, not just the algorithm behind the product. For example: if the product was offline for a day, this would count as a "miss" or failing to perform.

TIME LINE OF RELEVANT EVENTS – SEPTEMBER 2017

September 4th
CME1 ejected
C/M flares

September 5th
G1 geomagnetic Storm
S2 SEP Storm

September 6th
CME2 ejected
CME1 arrives
X2.2 + X9.3 flares
S2 SEP Storm

September 7th
X1.3 flare
G3 geomagnetic storm
S2 SEP Storm

September 8th
CME2 arrives
G4 geomagnetic storm
S2 SEP Storm

September 10th
CME3 ejected
X8.2 flare
S3 SEP Storm

September 11th
S3 SEP Storm

September 12th
CME3 arrives
G1 geomagnetic storm

Though firmly within the quiet period of solar cycle 24, between the 4th and 12th of September there was an increased amount of space weather activity giving rise to one of the most flare-active periods of solar cycle 24.

Effects were seen across the different domains of space weather and, as a result, were picked up and recorded by products on the portal[3].

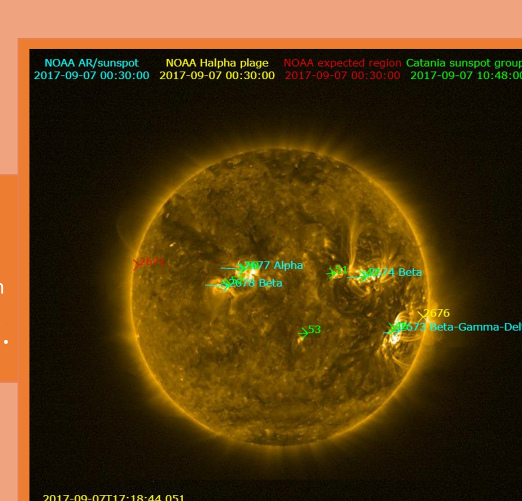


Fig.1 - PROBA-2 Image from SIDC/ROB for 7th September 2017.

This heightened period of activity was chosen, as it is the ability of the portal[3] to handle such events that is important to many users, and therefore understanding how it copes under increased conditions is essential.

COMESEP Alert Service

COMESEP is a notification system designed by the Space Physics Division of BIRA-IASB to provide alerts for imminent geomagnetic and Solar Energetic Particle (SEP) events (>10MeV and >60MeV) based on the strength and probability of occurrence.

- Initially product classified as "Probabilistic"
  →Applying techniques, Limiting data to one event prevents this method of validation.
- Product re-evaluated as a "Dichotomous" product (threshold = 0%) with secondary validation to consider the issued probabilities.
  →This approach was successful.
- As product provides event strength with alerts, occurrence of the event within the predicted timeframe was validated with and without associated strength.
- During event, an issue with input data prevented geomagnetic storm alerts being issued.
  →Issue since been resolved and so only SEP alerts are validated here.
- There only exists reference data for >10MeV SEP events, and so only these alerts are considered.
- This project defines the start of an event to be the first data point with flux ≥ 10pfu.
- The end is the last time the flux ≥ 10pfu.

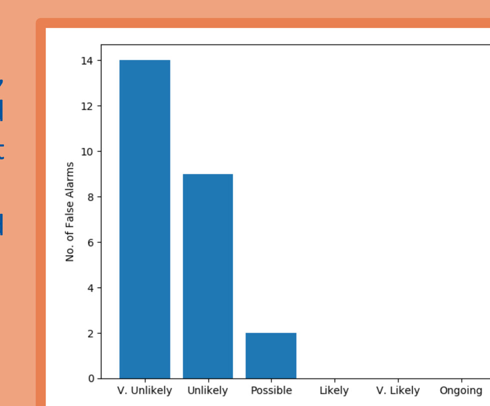


Fig.7 - Bar chart showing the distribution of False Alarms by their predicted occurrence probability.

Results:
From reference GOES data 9 events were detected.
→Validating both predicted timeframe and strength produced 6 Hits.
→Validating timeframe alone (ignoring the strength prediction) produced 8 Hits.
→The two methods also produced 26 and 25 False Alarms respectively.

The distribution of the False Alarms with respect to their associated probabilities was examined and displayed in Fig. 7. As can be seen, all occur when a low likelihood of an event was predicted, and thus do not seem unreasonable.

Conclusion:
Some products should be validated in this way, with primary and secondary steps to examine specific elements. These methods may not necessarily match their 'guideline category'.

GFZ Kp Nowcast

The German Research Centre for Geosciences (GFZ) currently provides the 3-hourly real-time nowcast Kp index (and related ap and Ap indices) to the SWE Portal.

- Product is "Continuous"
- In order to avoid logarithmic errors, the linear version of Kp index, ap, was used.
- All methods from the guidelines were successfully applied.

Results:
Figures 5 and 6 show the GFZ Nowcast correlates closely with the definitive reference data. Overall validation suggests the product is performing well.

Suggested technique	Applicability
Scatter plot	✓
Box plot	✓
Mean Error	✓
Mean Absolute Error	✓
Root Mean Squared Error	✓
Mean Squared Error	✓
Mean Relative Error	✓

Table 3 - List of techniques from the guidelines for a 'continuous' product with indicated applicability.

Conclusion:
The continuous methods in the guidelines are sufficient. It should be stated that the linearly scaled data[4] should be used for validating where possible.

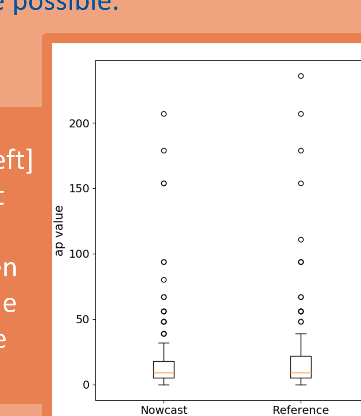


Fig.5 - Scatter plot [Left] and Fig.6 - Box Plot [Right] showing relationship between GFZ Nowcast and the definitive reference data.

A-EFFort

Athens Effective Solar Flare Forecasting (A-EFFort) from the Research Center for Astronomy and Applied Mathematics (AOA/RCAAM) is an online solar flare prediction service that monitors and evaluates active regions, providing advance warning of major solar flare activity.

The service provides probabilities for existing active regions in the earthward solar hemisphere and extrapolates them toward a full-disk flare probability. Predictions are provided with zero latency (effective immediately) and are refreshed every three hours. In this project, Active Region predictions were validated against the Hinode X-ray Telescope (XRT) Flare Catalogue.

- Predictions provided for M1+, M5+, X1+ and X5+ class solar flares.
- By guidelines this would make the product multi-categorical.
- Limiting the data sample size to a single event this was not possible.
- Still many M1+ class flares but the higher classes had significantly less events.
  →M1+ validated only.

Suggested technique	Applicability
Reliability diagrams	✓
ROC curves	✓
Discrimination diagrams	✓ ✗
Brier Score	✓
Brier Skill Score	✓ ✗
Multi-category methods	✗

Table 1 - List of techniques from the guidelines for a 'probabilistic' product with indicated applicability.

Table 1 shows two techniques with both a 'tick' and 'cross'. This is to indicate that the technique could be carried out, but due to lack of data, the results are not reliable.

Results:
Fig.2 shows that high probability predictions did not occur as frequently as expected (usually aligning closely with y=x)

→Likely due to the low number of events to begin with. With greater sample size, we could expect a better alignment.

Conclusion:
Different 'targets' or 'margins of error' should be established for validation of events and thus smaller sample sizes.

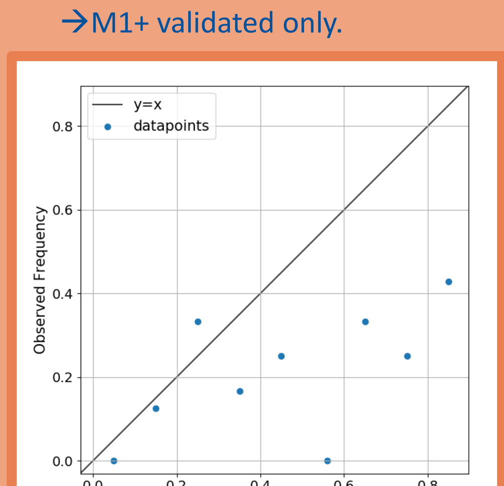


Fig.2 - A reliability diagram showing the correlation between the predicted probability and observed frequency for the 31 M1+ flares detected between 01/09/2017 and 30/09/2017.

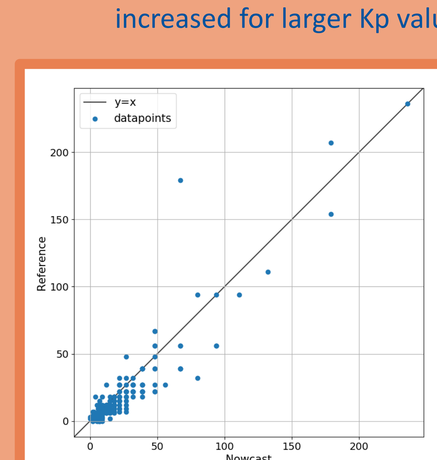
IRF Kp Forecast

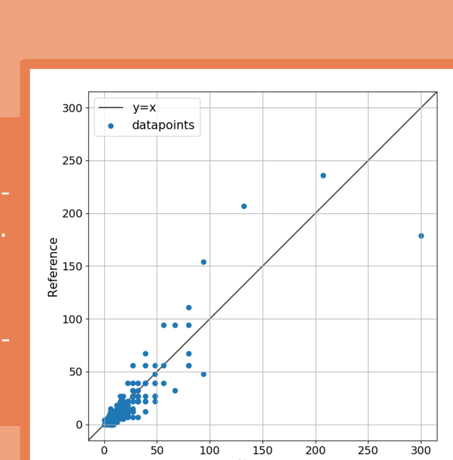
Driven by solar wind data from SWPC, the Swedish Institute of Space Physics (IRF) provide a forecast of Kp producing 4 predictions with a lead time of 0,1,2, and 3 hours.

- Product is "Continuous"
- Kp conversion to ap was not straight forward as the forecast provides a non-discrete near-Kp value.
- This was converted, after discussion with IRF, by rounding to nearest Kp value and then converting.
- All continuous methods from guidelines shown in Table. 3 could be successfully applied.

Results:
As can be seen in Figures 3 and 4, there is positive correlation between the Kp Forecast and the definitive reference data. The 3-hour advanced forecast showed greater error, and less correlation than the 1-hour forecast, which could be expected. The error also increased for larger Kp values.

Conclusion:
Continuous methods from the guidelines work. It should be stated that the linearly scaled data[4] should be used for valiation. Here, this required discussion with IRF.





[Left] Fig.3 - Scatter plot for 1-hour Kp forecast.

[Right] Fig.4 - Scatter plot for 3-hour forecast.

SIDC CACTus

CACTus (Computer Aided CME Tracking) is a software developed by the Royal Observatory of Belgium Solar Influences Data analysis Center (ROB/SIDC) to use image sequences from SOHO/LASCO to autonomously detect Coronal Mass Ejections (CMEs) without human involvement[2]. This results in a catalogue of CME events, with associated data including principle angle, angular width and velocity estimations for each of the listed CME.

Suggested technique	Applicability
Contingency Table	✓
Accuracy	✗
Bias score	✓
Probability of Detection	✓
False Alarm Ratio	✓
Success Ratio	✓
Threat Score	✓
All continuous methods!	✓

Table 2 - List of techniques from the guidelines for a 'dichotomous' product with indicated applicability.

- CACTus CME detection categorised as "Dichotomous"
- Product also provides secondary "Continuous" data associated with detected CMEs.
  →Product was therefore validated both ways.
- Product clears and re-issues list of CME detections every 3 hours.
  →Each run validated separately, significantly reducing data sample size.

Results:
With repeated runs, CACTus often produces different results with near-same input.

→Usually caused by miss-detecting one CME as two or more. This happens increasingly during times of high activity as there may be more than one actual CME to detect.

True Negative is difficult to define as there is no set time increment for 'no CME'.

→True Negative could not be determined, and thus Accuracy which relies on this value.

With 'continuous' validation methods, both Principle Angle and Velocity both showed low errors within the given product requirements. Angular Width error varied wildly between runs.

Two possible reasons:
→Very small data sample (1 run)
→Known Halo CME issue with CACTus.

Conclusion:
Currently there is no clear process in the guidelines for validating a product that repeats and rewrites its detections.
Two possible solutions:
→Validating the first detection of CME only.
(not ideal as this is not necessarily the detection users would see)
→Averaging the individual run validations.

SUMMARY OF CONCLUSIONS

ISSUE
Validating a limited period, limits the data you have to work with.

FINDINGS
- The methods for validating 'dichotomous' and 'continuous' products in the guidelines are described well and could all be applied.
- Limited data affected whether the suggested guideline techniques can be applied as described, and the kind of results they produce.
- Multi-categorical products suffer most, as the already smaller dataset is sifted into further smaller categories.
- Some products also span multiple categories for example CACTus, that both detects and provides continuous data about CMEs.

SOLUTIONS
- Products can be validated using methods from more than one category, using 'primary' and 'secondary techniques. (E.g.: A-EFFort + CACTus)
- Examples of "acceptable" results for validation on small data sets should be defined, as we would expect greater errors with less data.

NEXT...
These are suggestions for the guidelines, specific to the validation of single events.

REFERENCES

[1]Tsagouri, I. et al. (2019), *Guidelines for common validation in the SSA SWE Network, Issue 1*
[2]Robbrecht, E. & Berghmans, D. (2004), *Automated recognition of coronal mass ejections (CMEs) in near-real-time data*, AAP, 425, 1097
[3]*SSA Space Weather Service Network*, swe.ssa.esa.int
[4]GFZ German Research Centre for Geosciences, *Geomagnetic ap Index*, https://www.gfz-potsdam.de/en/kp-index/#gfz-collapse-c55295

CONTACT DETAILS
E-mail: Sophie.Burley@ESA.int
LinkedIn: sophie-burley-57b4ab10a
Address: ESA/ESOC
Robert-Bosch-Str. 5
64293 Darmstadt, Germany

esa