SpaceOps-2023, ID # 552

Deeply Understanding Space Weather

Matthieu Melcot[a], Stijn Calders[b], Mark Dierckxsens[c]

[a] Department of Control & Data Centres, Space Applications Services, Leuvensesteenweg 325, 1932 Sint-Stevens-Woluwe (Brussels Area), Belgium, matthieu.melcot@spaceapplications.com
[b] Space physics department, Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3 Avenue Circulaire, 1180 Brussels, Belgium, stijn.calders@aeronomie.be
[c] Space physics department, Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Ringlaan 3 Avenue Circulaire, 1180 Brussels, Belgium, Mark.Dierckxsens@aeronomie.be

Abstract

Space weather (SWx) is the physical and phenomenological state of natural space environments. The associated discipline aims, through observation, monitoring, analysis and modelling, at understanding and predicting the state of the Sun, the interplanetary and planetary environments, and the solar and non-solar driven perturbations that affect them, and also at forecasting the potential impacts on biological and technological systems. Under the European Space Agency's (ESA) Space Safety Programme (S2P), a system to monitor, predict and disseminate Space Weather information and alerts is being developed.

Machine learning (ML) approaches have been demonstrated in a broad range of applications. The study DENSER, Deeply uNderstanding Space Weather, has been conducted to review current practices in the field of ML and lessons learned from recent initiatives which have applied these techniques to space weather problems. The study has implemented a demonstration system utilizing state-of-the-art ML techniques to analyze the potential role of these approaches in further developing space weather prediction capabilities.

This paper reports the evaluation of the value of ML technologies in the SWx domain, using both traditional supervised learning techniques where experts make adjustments to the algorithms when needed, and Neural Network based Deep Learning, where the model itself was able to determine its performance and correct itself, provided that sufficient training and validation data is available. A literature study was performed by experts from the Royal Belgian Institute for Space Aeronomy (BIRA-IASB) and the Finnish Meteorological Institute (FMI).

The paper will describe two products developed in this study, both from a different domain, namely forecasting of solar proton flux (guided by Space Radiation experts at BIRA-IASB) and geomagnetic activity (guided by Geomagnetic Conditions experts at FMI). This served the purpose of not only validating that ML can provide added value to SWx applications, but also that developed algorithms can be reused across products. In addition, this two products approach allowed us to define more general "apply ML in SWx" workflows and immediately validate these workflows by putting them into practice. The developed products have been integrated into an application geared towards ESA's SWx Network and made available via the ESA SWx Portal. Finally, the paper will provide lessons learned from the study and development/pre-operation phases. It will provide recommendations and roadmap for further integration of ML technologies in the SWx domain.

Keywords: Space Weather, Machine Learning, Neural Network, LSTM

Acronyms/Abbreviations

ACC	Accuracy
ACE	Advanced Composition Explorer
AWT	Average Warning Time
BIRA-IASB	Royal Belgian Institute for Space Aeronomy
CME	Coronal Mass Ejection
CORR	Correlation

CSV	Comma-separated values
DENSER	DEeply uNderstanding Space weathER
EVA	Extra-vehicular activity
FAR	False Alarm Rate
FMI	Finnish Meteorological Institute
GEN	General Data Services
HSS	Hanssen-Kuipers Skill Score
IMAGE	International Monitor for Auroral Geomagnetic Effects
LAU	Launch Operations
LSTM	Long Short-Term Memory network
MLT	Magnetic Local Time
NRT	Near real-time
NSO	Non-space System Operation
OUJ	Oulujoki
PCA	Polar Cap Absorption
POD	Probability of (True) Detection
POFD	Probability of False Detection
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
S2P	Space Safety Programme
SCH	Human Space Flight
SCO	Spacecraft Operation
SEP	Solar Energetic Particle
SEPEM	Solar Energetic Particle Event Model
SR	Success Rate
SSA	Space Situational Awareness
SST	Space Surveillance & Tracking
SWx	Space Weather
TIO	Trans-Ionospheric Radio Link

TS	Threat Score
TSS	True Skill Statistic
UI	User Interface
UT	Universal Time

1. Introduction

In DENSER (DEeply uNderstanding Space weathER), we aimed to evaluate the potential value of Machine Learning (ML) technologies in the Space Weather (SWx) domain. The project started with an elaborate literature study that aims at creating an extensive overview of what ML activities have already been undertaken in the SWx domain. In a second phase of this study, a list of products for which ML might provide significant advantages -compared to the classical approaches of (physical) model design- was composed and elaborated upon.

In DENSER, two forecasting products have been specified and developed: ForMaL-SEP and ForMaL-Xrange. The models have been trained with data stemming from publicly available sources. The models' performances have then been evaluated using different validation techniques. This helped to compare the developed models against comparable pre-existing models.

In parallel, a web application that embeds the models and serves them as a web page has been developed.

2. What is Space Weather?

Space Weather describes the state of the environment in interplanetary space. The Sun is the main driver for the variability of this environment which includes changes in the electromagnetic and particle radiation levels, interplanetary and geomagnetic fields, and ambient plasmas. Galactic cosmic rays and micron-sized particulates, both not from solar origin, also contribute to the changes in the space environment. All these phenomena can potentially impact technological and biological systems. The effects are predominantly on spacecraft (electronics, solar panels, communications) and on human health during manned space missions. However, several effects can also be observed on or near ground such as increased radiation levels at aircraft altitudes, disturbances and damages to power grids and pipelines, and disruptions in radio communications.

The SWx discipline aims to monitor and measure the space environment so that the affected communities can take appropriate measures to limit the potentially damaging consequences. The observations of the environment and their effects are further analysed to build both physical and empirical models to describe the behaviour of the space environment. These models can be used during the mission planning and design phase to understand the environment that can be expected during the mission. They can also be employed to provide forecasts of the space environment allowing a longer lead time for the affected users to mitigate the consequences of Space Weather.

The Space Weather (SWx) segment of the ESA Space Situational Awareness (SSA) Program and its successor Space Safety Programme (S2P) is developing a system to monitor and predict the Space Weather environment and its effects and to disseminate this information to the end-user. Within the SWx network, the phenomena and their effects are divided into five domains: solar weather, heliospheric weather, space radiation, ionospheric weather and geomagnetic conditions.

3. Products

Two new products have been developed: FoRMaL-SEP, which predicts the expected maximum solar proton flux at geostationary orbit (GEO) with energies greater than 10 MeV, and FoRMaL-Xrange, which predicts the north component of the ground-based magnetic field.

3.1. ForMaL-SEP

The first product developed in DENSER is a prediction of the expected maximum flux levels at GEO for >10 MeV protons from solar origin for various time horizons. SEP (Solar Energetic Particle) events are sudden and large increases in the energetic particle fluxes observed in interplanetary space and can last for days or even weeks. These increased radiation levels are a consequence of particle acceleration during solar flares and Coronal Mass Ejections (CME).

The forecast model provides the expected maximum flux for protons with energies greater than 10 MeV (E >10 MeV) during the next hour, 3, 6 hours and 24 hours.

The input parameters are derived from time series of solar X-ray and differential proton flux measurements from 1986 to 2015 covering almost 3 solar cycles during which just over 200 SEP events were observed near Earth that were at least minor events (>10 MeV proton flux ≥ 10 s-1cm-2sr-1).

The parameters derived from the input data are for instance (but are not limited to) the actual flux value, a running average over various time ranges, the change and rate of change averaged over different time periods, etc. Characteristics of flares have been shown to be correlated with the expected proton peak flux, see e.g. [1]).

Protons accelerated during solar eruptive processes can reach energies from a few MeV up to a few GeV. Since these particles travel at sub-relativistic speeds, the highest energetic protons can be a good indicator for the start of a large event at lower energies. After the flux levels start rising, it typically takes 10s of minutes or even hours before the threshold for a minor event is reached. The early detection of an SEP event onset provides valuable input to increase the accuracy of short-term forecasts.

The ForMaL-SEP forecasts are presented to the user in the form of an image with the measured >10 MeV integral proton flux from the past days as well as the expected peak flux for the different forecast horizons. The different radiation storm levels according to the NOAA scale are also indicated.

On a separate page, the user is able to browse an archive of the forecasts where the predicted values are shown in comparison to the measured values, potentially also indicating the achieved performance over a certain period of time (e.g. maximum predicted flux in a time window versus maximum observed flux during that time window). The latest forecasts as well as the archived forecasts are accessible in numerical format.

3.1.1. Potential users

The forecasts provided by ForMaL-SEP are relevant for the User Domains Spacecraft Operation (SCO), Human Space Flight (SCH), Launch Operation (LAU) and Non-space System Operations - Service to Airlines (NSO/air). The product is relevant for all these user domains as energetic charged particles can endanger life and damage electronics in space.

More specifically, ForMaL-SEP is of interest in the following use cases:

- Forecast of SEP events are useful for:
    - Spacecraft Operation (SCO), to determine the likelihood of internal charging leading to discharge, single event effects and long-term radiation dose, and to take preventative measures and prepare recovery measures in case of disruption.
    - Human Space Flight (SCH), to estimate the exposure an astronaut might receive, put astronauts on alert so they are ready to shelter when a strong SEP occurs, and reduce or reschedule EVA activities if necessary.
    - Launch Operations (LAU), to estimate the risk of single event effects before and during launch.
    - Non-space System Operation (NSO)/air: to predict the potential occurrence of PCA events affecting communications at high latitudes.
    - General Data Services (GEN), to provide input for other SWx products estimating the potential effects of SEPs.
- Forecasts of all quiet conditions are useful for
    - Spacecraft Operation (SCO), when an operator has scheduled a critical manoeuvre.
    - Human Space Flight (SCH), when an astronaut is about to perform EVA.
    - Launch Operations (LAU), when a launch is scheduled

3.2. ForMaL-Xrange

The second product that has been developed in DENSER is a forecast within the 1, 3, 6 and 24 hours of the ground-based magnetic field north component (B-North-Range) which in turn is used to calculate the RX value, which is equal to Xmax-Xmin. RX is currently used in the AurorasNow! service of FMI as a proxy for enhanced auroral activity. The service forecasts the B-North-Range for the next hours from the basis of previous hour's averages of solar wind velocity and Interplanetary Magnetic Field (IMF) north-south component (Bz).

ForMaL-Xrange is a ML-based B-North-Range forecast for three Finnish stations: Kevo (magnetic latitude ~66°, standard auroral oval), Oulujarvi (magnetic latitude ~61°, expanded oval), and Nurmijarvi (magnetic latitude ~57°, sub-auroral region). The forecasts naturally have some trends depending on the Universal Time (UT), as they can capture activity only in a limited sector of Magnetic Local Time (MLT). In forecasts of lead times larger or equal to

~3 hours the system needs to model besides solar driven activity also variations due to rotation of the station from one MLT sector to another. This feature has been taken into account by including UT-time to the input parameters of the system.

Solar wind data (velocity, density, temperature and magnetic field) as measured by the Advanced Composition Explorer (ACE) [3] spacecraft at L1 point of Earth-Sun system are used as inputs to the for ForMaL-Xrange forecast

3.2.1. Potential users

The ForMaL-Xrange product is most likely relevant particularly in the non-space operations (NSO) user domain. Auroral tourism is an obvious application area due to the legacy of RX forecasts in the AurorasNow! Service. High confidence levels are required particularly in resource exploitation (e.g. in directional drilling and aeromagnetic surveys) where campaign costs are high.

Besides the most obvious users in NSO, ForMaL-Xrange is of interest in the following use cases:

- Forecasts of all quiet conditions are useful for
  - Spacecraft Operation (SCO) when critical orbital maneuvers and launch operations are planned.
  - Human Space Flight (SCH), when extravehicular activities are scheduled.
- Forecasts of geomagnetic storms are useful for
  - Trans-Ionospheric Radio Link (TIO) because geomagnetic storms often generate abnormal disturbances also in the ionospheric conditions controlling space-based navigation and communication. As the characteristics of ionospheric disturbance depend strongly on latitude, TIO gets benefits from forecasts tailored separately for global and regional activity.
  - SCO, where the information is used to take preventative measures and prepare recovery measures in case of disruption
- Information on the geomagnetic activity at different levels is useful for
  - Space Surveillance & Tracking (SST), to put staff on alert and predict risk of losing track of objects and to estimate high altitude neutral atmosphere density
  - General Data Services (GEN) because many models require characterization of geomagnetic activity as their input data.

Concerning the examples given above, it is good to recognize that in some cases ForMaL-Xrange cannot provide forecasts with the same reliability as a global activity index will do. However, when the ForMaL magnetometer stations are scanning the midnight sector, they are able to describe the activity there more accurately than global indices, like Kp or AE, can do. This is because the latter use only one station in each local time sector.

4. Models

Machine Learning techniques are a wide and dynamic field. In the past years, many techniques have emerged, and new ones appear regularly. There are different types and categories of Machine Learning techniques, and each of them can apply successfully or not to a specific problem.
In the frame of the time series predictions, the Recurrent Neural Network (RNN) models are a preferred choice as they generally provide good results.

A Recurrent Neural Network (RNN) is a type of deep learning neural network that is specifically designed to process sequential data, such as text, speech, or time series data for instance weather data. It is based on the idea that the output of a given time step can be used as input to the next time step, allowing the network to maintain an internal state that summarizes information from the past.

In DENSER, we chose to use Long Short-Term Memory (LSTM) neural network which is a specific type of RRN. A LSTM neural network is designed to remember information for longer periods of time. It is capable of learning long-term dependencies by using gated units to control the flow of information, allowing it to selectively remember or forget parts of its input. This allows an LSTM to learn more complex patterns of information and better capture temporal dependencies, making it effective at a variety of tasks such as machine translation, speech recognition, and time series prediction.

This means that the prediction at a N+1 timestamp is dependent on the series of inputs which is provided to the model, where each input modifies the internal state of the model.

The model implementation is made out of three layers (see Fig. 1):

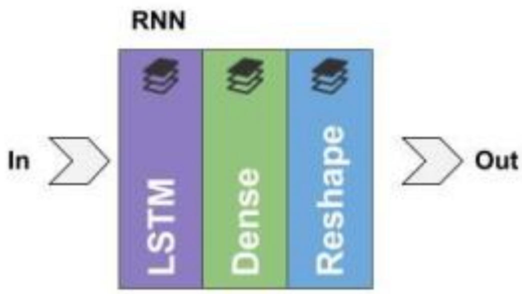

Fig. 1. LSTM model

- The LSTM layer contains the LSTM units that 'remember' the sequences that pass through them.
- The Dense layer with no activation (ie a Linear Layer) is used to reduce the dimensionality of the LSTM layer
- Finally, the Reshape layer separates the different predicted features in a new dimensionality. For instance, in the case of ForMaL-SEP, only the ">=10 MeV Proton Flux" feature will be considered (ie. other features are dropped).

5. Training and Execution Environment

Training a model needs significant processing power. It also needs multiple iterations so that one can adjust the hyper-parameters and model structure that will make the model performant.

In order to do so, a training environment is set up (see Fig. 2). This environment provides on one hand a set of tools for the developer to quickly iterate over the model and improve it, thanks to fast iterations with a REPL or an interactive notebook such as Jupyter Notebook.
On the other hand, this environment is backed by a graphical processing unit (GPU) powered infrastructure that can deliver strong processing capabilities.

Once the model has been trained, it is wrapped into a web server so that it can be used by a web application. In operations, the model uses the data input provided by the application backend and returns predictions that are stored in the application database.

Fig. 2 here below describes how the training and execution environments are structured.

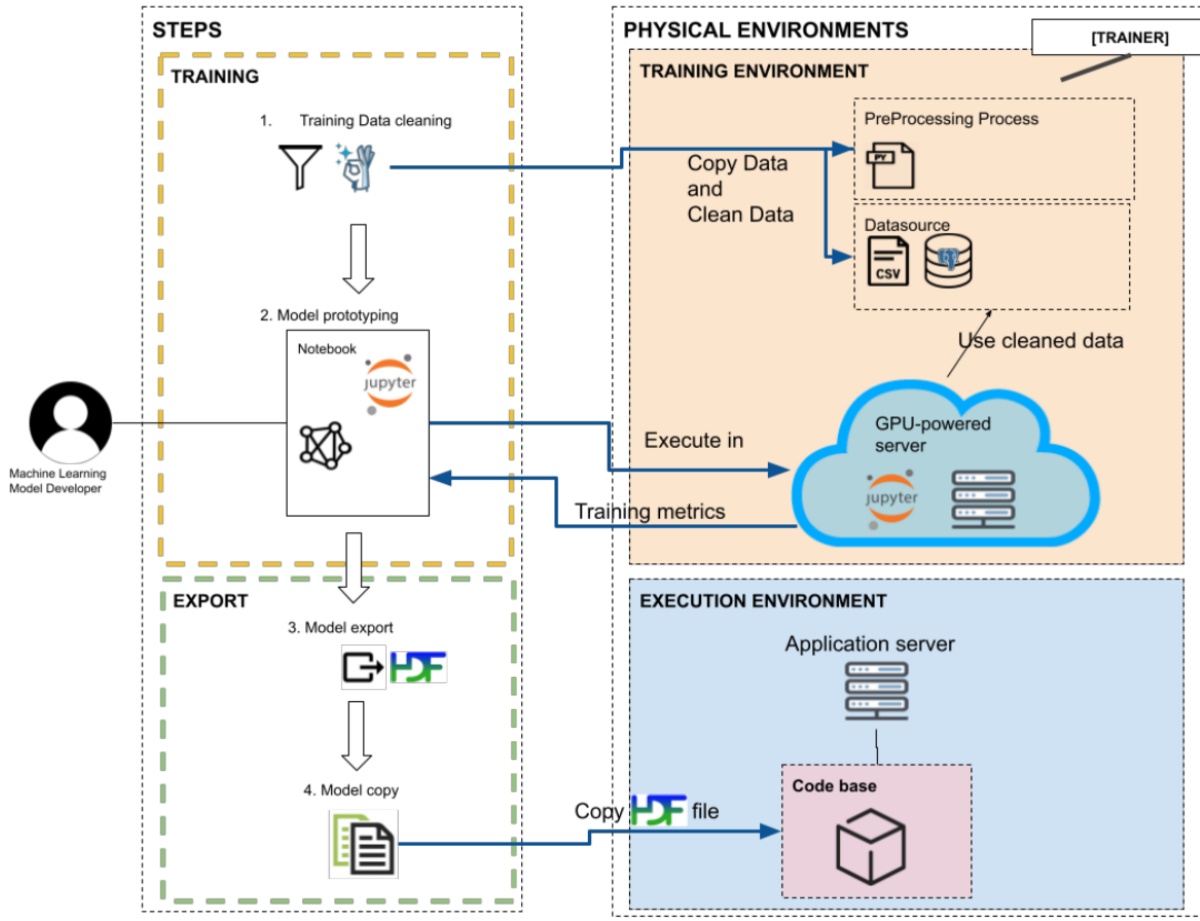

Fig. 2. Training and Execution environments

6. User interface

The ForMaL-SEP and ForMaL-Xrange products are served through a web application.

Both Products share a common UI structure (see Fig. 3): A menu and a calendar provide the possibility to browse the list of forecasts. Once a forecast is clicked, it is displayed in the central panel. The forecast view provides the time series prediction as a plot and a table. In addition, information and metrics about the forecast generation and metadata are provided. The view also allows the download of the forecast as a comma-separated values (CSV) file.

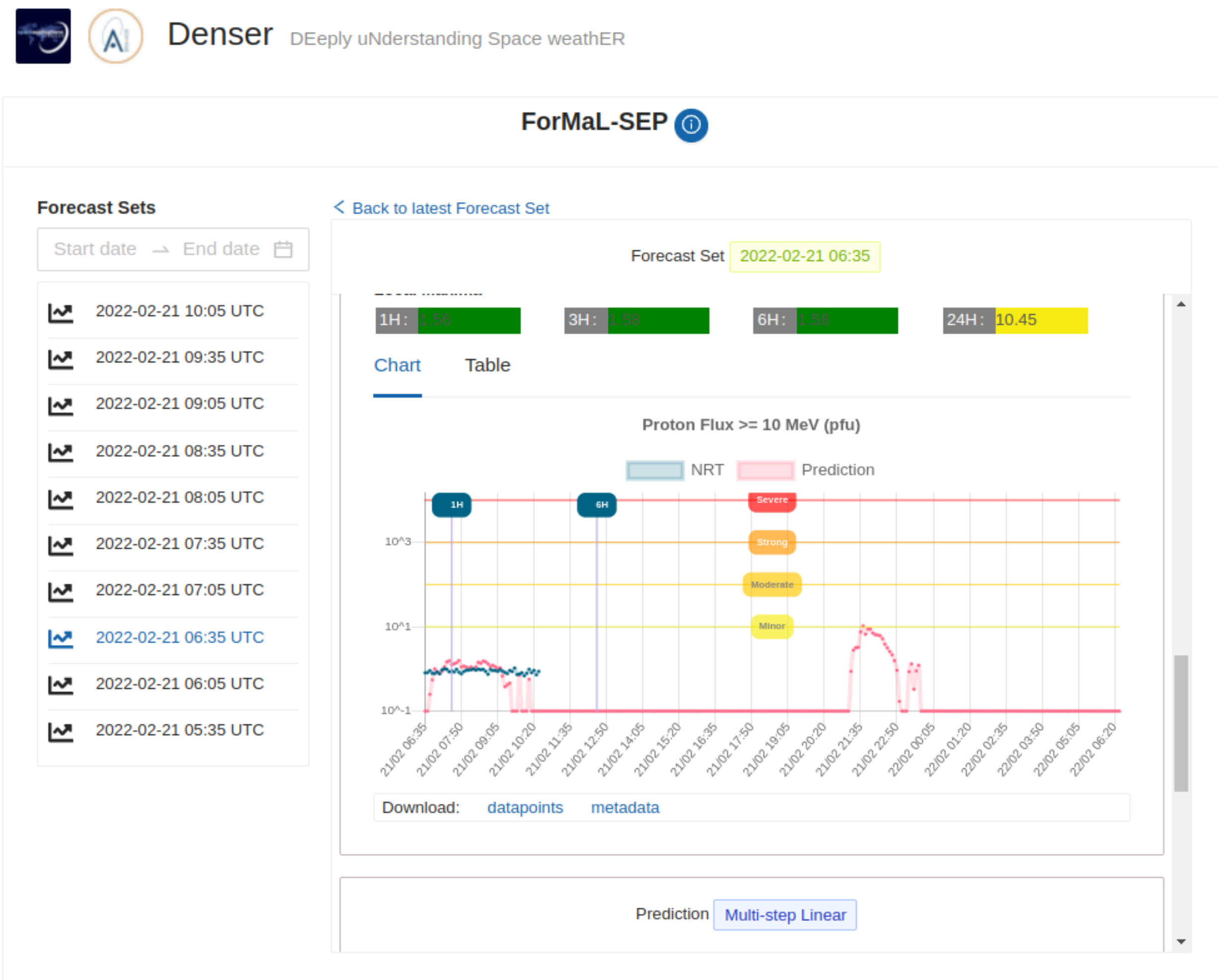

Fig. 3. DENSER UI for ForMaL-SEP product

7. Validation methodology

Both products have been validated using two different types of validation:

- Based on comparison with historical ground-truth data, deriving a set of metrics such as root-mean-square error (RMSE) and correlation (CORR) between the observed and predicted time series over different forecast horizons.
- Event-based validation on a list of identified past storm events, which aims at defining how well the models are related to the prediction of those events. In this method, we search to describe the model performances in terms of prediction lead time (i.e. time between the prediction and the actual event), but also the storm category matching.

Different validation tools such as contingency tables and derived quality measures such as Probability of Detection (POD), False Alarm Rate (FAR), and bias.

Finally, a list of existing reference forecast models is listed for both SEP event prediction and geomagnetic forecast, so that the performances of those models can serve as a reference when compared with DENSER models.

8. Validation results

8.1. ForMaL-SEP

ForMaL-SEP was validated following two approaches to derive the performance based on the comparison of all the predictions at each time step (i.e. at 5 minute intervals) against ground truth data as well as the capability to predict individual SEP events. The model is reacting clearly to the increases in both the flare and proton fluxes, but does not always respond correctly. However, one of the main issues is the very high level of false alarms, even after imposing extra requirements on the predicted flux profile. It is evident that further model development is desired by further enhancing the ML techniques employed and including more data features and input sources.

Table 1 here below presents the model performance metrics for the detection of SEP events, at 1 hours, 3 hours , 6 hours and 24 hours horizons:

Table 1. ForMaL-SEP performances at different horizons

	H1	**H3**	**H6**	**H24**
POD (Higher is better)	0.78	0.83	0.91	0.95
FAR (Lower is better)	0.62	0.63	0.66	0.86
HSS (Higher is better)	0.49	0.49	0.48	0.2
TSS (Higher is better)	0.75	0.8	0.87	0.77

8.1.1. Product Performances

The main conclusion from the validation activities is that ForMaL-SEP has a good Probability of Detection (POD) and Average Warning Time (AWT) compared with other models, but at the cost of a very high False Alarm Rate (FAR). The model also shows a tendency to underpredict the expected maximal value on shorter time scales and overpredict for longer prediction horizons. While looking at the behaviour of the predictions during individual events, it was observed that for gradual events (flare location not magnetically connected to Earth) first a positive prediction is issued, then the predicted event length decreases since the eruption is not immediately followed by an increase in proton flux, but once they start rising the predicted event length increases again.

The validation also highlights several other issues at the model level. One of the main issues is the large class imbalance between the active and quiet periods. After the initial validations, some changes were applied to the model training in order to try to fix issues with respect to the background subtraction of the input data taken with different instruments, the response to flat, rising and decaying inputs. However, none of the tried solutions has given satisfactory results.

8.2. ForMaL-Xrange

Correlations between ForMaL-Xrange forecasts in all lead times and measured Xrange values are lower than those reported in the literature for magnetic activity forecasts by machine learning methods. Therefore, ForMaL-Xrange 1-hour horizon forecasts cannot compete with those products. A potential improvement by ForMaL-Xrange to the already existing ESA SWx services would come from its 3-hour and 6-hours horizon forecasts. Based on the contingency tables, ForMaL-Xrange would be particularly good in forecasting magnetically quiet times. For example, for Oulujoki (OUJ) station the 3-hours horizon prediction the forecasts of quiet were correct in 16529 cases out of 16777 (96%) observed quiet. For the 6-hours horizon, the corresponding numbers are 15595 out of 16089 (97%). Of course, here it is important to understand that major parts for these good “all quiet” forecasts come from the times, when OUJ has been scanning magnetic local times where activity is typically low (i.e. outside the midnight sector). If moderate activity level is taken as a reference, then the success rates for OUJ are only 13% (50/388) for the 3-hours horizon and 10% (57/591) for the 6-hours horizon.

Comparisons of ForMaL-Xrange results with the Xrange forecasts by the AurorasNow! reveal that ForMaL-Range forecasts are in RMSE-sense systematically better than statistical forecasts based on 5, 50 and 95 percentiles used in AurorasNow!. Aurora enthusiasts today use most likely the 95 percentiles by AurorasNow! in their decision making. Based on the results from this validation study, ForMaL-Xrange would have better performance in guiding the AurorasNow! user community both during quiet and disturbed conditions.

Finally, our event-based validation with very strong activity (Kp>8) shows that ForMaL-Xrange 1-hour horizon forecasts cannot catch the highest peak values in observed Xrange values correctly, but in estimating the activity levels (from quiet to strong) the LSTM-based model behind ForMaL-Xrange and evaluated in this study works rather nicely.

9. Discussion

Several new space and ground based infrastructures are planned which will provide valuable space weather related measurements. However, there are several considerations common to many of them that should be taken into account when utilising these new datasets for data driven models. For many space weather impacts, it takes many years until a sufficiently large sample is recorded to be useful to build data driven models (whether they are based on ML algorithms or not). Initially, these new measurements will serve only to complement the input data in the implementations of the models to produce forecasts or as additional validation datasets, but not to train new models.

In order to use the measurements of these new instruments, it will be essential to perform some sort of cross calibration or other transformation to ensure that (derived) quantities are consistent with the existing observations. Furthermore, it will be important to gain understanding on the quality difference between NRT and post-processed versions of the new data set. Cross-calibration will even be needed if the instruments in different infrastructures are the same, as for instance the response functions will not be entirely the same. A good example of this within the DENSER project is the use of the Solar Energetic Particle Event Model (SEPEM) reference proton dataset [2] for ForMaL-SEP, which is a crosscalibrated sample of energetic proton measurements across several GOES missions using ever evolving technology of the same instrument. Furthermore, the newly designed particle monitor on GOES-16 is used as near-real time input data, and the energy channels need to be rescaled to the ones corresponding to the SEPEM dataset. Finally, in order to be useful as input to the forecasts, the data needs to be made available in near-real time.

Although the above considerations are particularly important for ML-based predictions, they are relevant for a broad range of space weather applications.

10. Conclusions

With DENSER, we studied how Machine Learning techniques can be applied to predict Space Weather. The models implemented in that study reveal that if in some aspect they perform quite well, but other (like the False Alarm Rate (FAR) still need improvement. This can be an issue when space weather alarms imply costly operations to prevent potential impacts on systems and infrastructures. Therefore the developed models can already be envisioned in certain situations such as aurora predictions, but are not yet mature enough to replace traditional, physics based models. The Machine Learning domain is in constant evolution. New time-series prediction techniques and algorithms are unveiled regularly, which means that performances will continue to increase.
It is also important to note that a big part of the models' performances is driven by the quality of training data, and the features derived from them. Thus, feeding the same models with additional inputs could sensibly improve the results.

Acknowledgments

This work was supported by Space Applications Services, BIRA-IASB and FMI. We thank Kirsti Kauristie (FMI) and Ilja Honkonen (FMI) for providing the content about the ForMaL-Xrange product. We thank Ralf Keil (Rhea system GmbH for ESA) and Alexi Glover (ESA Space Weather Office) for their support on the project.

References

[1] Dierckxsens, M. et al, Relationship between Solar Energetic Particles and Properties of Flares and CMEs: Statistical Analysis of Solar Cycle 23 Events, Solar Physics 290, 841, 2015, doi:10.1007/s11207-014-0641-4.
[2] N. Crosby et al., “SEPEM: A tool for statistical modeling the solar energetic particle environment,” Space Weather, 13, 7, 2015, doi: 10.1002/2013SW001008

[3] Smith, C., L'Heureux, J., Ness, N. *et al.* The ACE Magnetic Fields Experiment. *Space Science Reviews* 86, 613–632 (1998).