

Machine-Learning Research in the Space Weather Journal: Prospects, Scope and Limitations

Noé Lugaz, Huixin Liu, Mike Hapgood, Steven Morley

Abstract: *Manuscripts based on machine-learning techniques have significantly increased in Space Weather over the past few years. We discuss which manuscripts are within the journal's scope and emphasize that manuscripts focusing purely on a forecasting technique (rather than on understanding and forecasting a phenomenon) must correspond to a substantial improvement over the current state-of-the-art techniques and present this comparison. All manuscripts shall include information about data preparation, including splitting of data between training, validation and testing sets. The software and/or algorithms used for to develop the machine-learning technique should be included in a repository at the time of submission. Comparison with published results using other methods must be presented, and uncertainties of the forecast results must be discussed.*

While machine-learning techniques in space weather have been around since the first years of the field (e.g., see O'Brien and McPherron, 2003; Anderson et al., 2004), there has been a clear growth in published articles using machine-learning techniques in the past four years. Since 2018, there has been at least 10 articles published in *Space Weather* per year using machine-learning techniques, whereas no year before 2018 had more than three such articles. As a result, about 15% of articles published in *Space Weather* in 2021 have the term “machine learning”, “deep learning” or “neural network” in their abstract (18 so far). Overall, this growth covers all aspects of space weather, about equally distributed among work focusing on forecasting the radiation belts, geomagnetic indices (Dst, Kp), the ionosphere Total Electron Content (TEC) and solar flares. *Journal of Geophysical Research: Space Physics* has witnessed a similar growth with about as many articles published using machine-learning (ML) techniques but as that journal publishes about six times more articles than *Space Weather*, this is still a very small portion of total number of published articles.

In light of this growth, the *Space Weather*'s editorial team has been discussing the place that ML-based forecasts of space weather phenomena should have within space weather research. Note that Camporeale (2019) in a Grand Challenges review discussed the prospect and technical challenges of applying machine learning to space weather research. Readers interested in the specific applications and recommendations in moving towards probabilistic forecast, and assessing uncertainties, for example, are referred to this article. Hereafter, we focus on the scope of articles focusing or relying on ML techniques to be published in *Space Weather*. At the core is the current scope of the journal (“to understand and forecast space weather”) as well as the readership composed of a mix of researchers, forecasters, end-users, and policy makers.

Pure ML manuscripts, *i.e.* manuscripts presenting a new technique are not within scope, as there are specialized journals more appropriate for such articles. Manuscripts presenting a ML-technique to better **understand** the drivers of space weather are in scope but must, as any other manuscript, bring significant new insight to specific space weather phenomena to be published. This assessment is typically done by reviewers but is sometimes undertaken directly by the editor. AGU's *Earth and Space Science* is a journal where manuscripts presenting a new technique which confirms existing knowledge are within scope. Manuscripts presenting a new **forecasting** scheme

based on an existing ML technique are only in scope if they present a substantial improvement over state-of-the-art existing forecasting models. Comparison of the results of the new model to those from state-of-the-art models needs to be presented within the manuscript. The comparison cannot be limited to the most simplistic models (climatology, persistence, etc.) unless there aren't more appropriate models. AGU open data/software policy results in models published in the past few years being accessible for this comparison.

The lack of sufficient comparison with published results using other methods has been one major obstacle to judging the advancement/improvements of a new ML method. To facilitate the proper use and further development of ML in the field of space weather, straightforward inter-comparison between different studies must be made available. Towards this end and in line with AGU software policy, the model/code used in any new study submitted to *Space Weather* should be included in a repository at the time of submission. If this is a standard machine-learning code available in existing libraries, the set-up, version numbers, input parameters/coefficients need to be provided in supplementary information. The parameter space for which the model can be used and was validated should be provided in the manuscript. The data used to train, validate, and test the model should be presented. Issues with splitting the data, and the cyclical dependency (diurnal, seasonal, solar cycle) of many space weather phenomena, should be clearly described in the main text of the manuscript. This also includes the treatment of data gaps, extreme events, and outliers, which could directly affect the ML model outcome. Furthermore, the uncertainty of the prediction should be discussed. Finally, the improvement over existing ML methods and new physical insights should be discussed.

The presentation of metrics to quantify the goodness of the technique should go beyond correlation, RMSE and MAE. Specific examples of metrics and community best practices were provided in the Space Weather Capabilities Assessment topical issue of *Space Weather* and should be followed if possible. This includes the forecast of geomagnetic indices (Liemohn et al., 2018), thermospheric neutral densities (Bruinsma et al., 2018), radiation and plasma environment (Zheng et al., 2019), and arrival time of coronal mass ejections (Verbeke et al., 2019), among others. When appropriate, specific thresholds should be defined in order to develop a binary classification. Deterministic forecasting metrics should then be included, including skill scores. Studies using probabilistic forecasts (Camporeale, 2019) are also highly encouraged. For both approaches (deterministic and probabilistic) authors must cite references that describe their metrics and, if using metrics developed by the ML community, show how those are related to metrics used by the wider forecasting and research community. We note that these two communities often use different names for the same metrics. For example, the metrics of precision and recall used in the ML are identical to metrics of success ratio and probability of detection as used by the forecasting community (see section 6 of Morley et al., 2020 for more examples). We consider it important to develop a joint understanding of these two sets of metrics as ML becomes more widely used for space weather purposes.

Bibliography

Anderson et al., 2004 10.1029/2004SW000095

Bruinsma et al, 2018 [kws-v-22gr-ltruj243435<2534;VZ 33535:](https://doi.org/10.1029/2018SW002061)

Camporeale, 2019 <https://doi.org/10.1029/2018SW002061>

Liemohn et al., 2018 [kwsv=22grlruj2431435<2534;VZ 33539;](#)
Morley, 2020 [kwsv=22grlruj2431435<2534;VZ 33543;](#)
O'Brien and McPherron, 2003, 10.1029/2002JA009324
Verbeke et al., 2019 [kwsv=22grlruj2431435<2534;VZ 335379](#)
Zheng et al., 2019 [kwsv=22grlruj2431435<2534;VZ 335375](#)