Review of Probability Estimation of March 1989-like event .. Overview: The authors atempted to assess the likelihood of a reoccurrence of the March 1989 storm using the Dst index as a proxy.

Comments: The authors have used the pyextremes Python package to model the occurrence of large storms based on threshold levels of the Dst and provided a lot of explanation of the methodology but without any new insights.

There is a lot or repetition about the importance for the insurance industry and I was expecting to see some sort of economic analysis, but there is litle or no novel results in this paper. The paper is very poorly referenced, shows litle understanding of what the use of extreme value statistics is about and has not provided any new information on either the geomagnetic or the economic impact.

It seems unclear what the purpose of the paper is. For example, a return period of a known size is usually the rank order appearance of the already measured event. If the March 89 storm is the ranking event in the Dst index, which Figure 1 indicates, then it is a one-in-70 year event (as Dst starts from 1957) by definition.

You don't need to use EVT to provide an estimated return period of 72 years – that is the length of the input record! EVT tries to provide answers about events are that not in the record. There also has to be some level of understanding of the limitations of the method. Figure 6 shows the 1000 year return period but this is not really to be believed given there's only 70 years of data and the Sun's activity levels are clearly not stationary on 1000 year timescales - something which is not mentioned in the paper.

There have been a number of recent papers such as Rogers et al (2021) and Chapman et al (2020) which do a far beter job at detailing the extreme events and likelihoods. This is the sort of level of research that is now required in this area. The references are incomplete.

There are several cited that are not in the reference list and the reference to Thomson et al (2011) is completely incorrect. Also note that Thomson et al. and Love (and Rogers et al.) were looking at ground magnetic fields rather than a global proxy like Dst.

The tables are not correctly formated and the Figures show 'datetime' which is not actually a word but a Python package label.

Finally, the 2013 RAE report by Cannon et al will soon be updated and should provide fresh insight into the evolving risks and mitigations that are being researched in relation to space weather impacts on technology.

Response to Referee's Comments

Dear Referee,

We appreciate the time and effort you have invested in reviewing our manuscript. We have carefully considered each point and have prepared our replies accordingly. We'll make revisions to our paper to address these concerns based on your feedback on our answers.

1. Lack of New Insights and Economic Analysis

We acknowledge the referee's point about the paper seeming repetitive, particularly with regard to its importance for the insurance industry. However, it is very important to note that this research is not focused on speculation about current economic impacts. Our aim is limited to using the concrete scientific data available, applying statistical data science to this dataset, and meeting the scientific requirements necessary for the scientific interpretation of the results, and to propose a methodology for geomagnetic storms, risks that are not yet modelled and mitigated in the insurance sector. The new insights mentioned in the paper relate to the insurance industry and its ability to physically influence the mitigation of such risks. The insurance industry is not yet modelling such risks.

2. Purpose of Using EVT

We apologize for not making the purpose of using Extreme Value Theory (EVT) explicit. EVT is applied not merely to provide a return period for known events but to predict the likelihood of unrecorded extremes. Therefore, the probability of an event occurring above an event of 589 nT was investigated. Events such as Carrington, 1909, 1921 and 1946 are not included in the data set. However, this does not mean that the model is not capable of explaining these events.

Thank you for pointing out the concept of a return period based on rank-order appearance. Indeed, if the Dst index starts from 1957 and the March 1989 storm ranks highest, then by definition, it could be considered a one-in-66-year event based solely on the length of the dataset. However, this approach would only provide insights into what has already occurred, not on what could happen in the future, especially for events that have not yet been observed.

We used Extreme Value Theory (EVT) to extrapolate beyond the limitations of our 66-year record, aiming to provide a statistically rigorous estimate for the occurrence of extreme events that are potentially more severe than any previously recorded. While EVT can produce estimates for the return periods of extreme events not in the current record, we acknowledge that there are limitations to these projections.

3. Timescale and Stationarity

We appreciate the reminder that caution should be exercised when interpreting results, especially for very long return periods like 1000 years, given the non-stationary nature of solar activities. This limitation will be explicitly stated in the revised manuscript. We intend to focus on understanding and explaining these limitations in future studies, to provide a more comprehensive outlook on extreme geomagnetic storms.

5.4 Limitations and future research & 5.1 Comparison with previous studies have been updated.

4. Comparative Research Quality

The works of Rogers et al. (2021) and Chapman et al. (2020) were consulted, prior to creation of this manuscript. As well as all the other available pioneering works.

The focus of our research is on events that are statistically more likely to occur within our lifetimes, rather than the more extreme but less frequent Carrington-like events often examined in space weather science. Our approach aims to provide actionable insights that are immediately relevant for today's society and technological infrastructure. In line with this, we opted to use the Dst index as our measure because it is a more standardized and continuous metric, suitable for assessing global-scale trends.

For those interested in a more in-depth analysis of extreme space weather events, we recommend studies like Rogers et al. (2021) and Chapman et al. (2020), as well as the comprehensive paper on Probability estimation of a Carrington-like geomagnetic storm: https://pubmed.ncbi.nlm.nih.gov/30787360/

We believe that our focus on more probable, albeit less extreme, events complements existing research and offers valuable information for a variety of stakeholders, from policymakers to industries affected by space weather. We will clarify this rationale, along with the reasons for our choice of the Dst index, in the revised version of our paper.

This paper might be worth referencing as well:

Bergin, A., Chapman, S. C., Watkins, N. W., Moloney, N. R., & Gjerloev, J. W. (2023). Extreme event statistics in Dst, SYM-H, and SMR geomagnetic indices. Space Weather, 21, e2022SW003304. https://doi.org/10.1029/2022SW003304

The reference above has been added to the references and our commentary on the pioneer findings has been added to section 5.1 Comparison with previous studies and 5.4 Limitations and future research

5. Incomplete and Incorrect References

"There are several cited that are not in the reference list and the reference to Thomson et al (2011) is completely incorrect."

We regret the oversight concerning the incomplete and incorrect references, and we will rectify these errors upon further clarification. Please let us know regarding any other missing references other than "*Thomson et al (2011)".*

We apologize for any confusion our initial citation of Thomson et al. may have caused and will clarify the specific sections and arguments from Thomson et al. that our work refers to.

The comment on Thomson et al.'s study in the paper was made addressing his study and the specific section shown as below.

[Line: 408] *Accurate risk assessment and management are essential for a variety of industries, as shown by Thomson et al.'s study on the probabilistic assessment of extreme geomagnetic storms using statistical methods, and additional research in this area may result in more potent mitigation techniques.*

https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2011SW000696

[28] The non-stationary nature of geomagnetic data implies that our analysis probably needs refinement. There is a solar cycle dependent variability seen in residuals, and perhaps de-trending by smoothed sunspot number or by monthly mean numbers of coronal mass ejections is required. We may return to this in future work. However, in our results we look at likely trends over many solar cycles and the significance of such "short" period variations may be less. One may also consider the appropriateness of the "block-averaging" versus "point-over-threshold" methods and perhaps re-examine the threshold and de-clustering choices we have made. These may all have some impact on the robustness of our results. However, the results already given here should find application in hazard assessment and in magnetic navigation applications, not least in helping to assess the risk to power systems and to magnetic navigation activities within Europe and beyond.

Pg4 lines 94 and 103 (Dahen et al. and Rootzen and Tajvidi) were added in the references section.

Pg5 lines 133 and 135 (Schrijver et al., 2014; Eastwood et al., 2017 & Oughton et al., 2017; Kappenman, 2001) were corrected and references were added.

6. Tables and Figures

We acknowledge that the presentation of tables and figures should adhere to journal standards for readability and academic rigor. The use of 'datetime' was an oversight; it is indeed a Python package label and not an appropriate term to use in the context of the figures. In the revised version of the paper, we will replace 'datetime' with the more appropriate 'Date and Time,' provided this change aligns with the referee's expectations.

All Figures are updated

Crameri, Fabio. (2023). Scientific colour maps (8.0.1). Zenodo. https://doi.org/10.5281/zenodo.8409685

We are committed to rectifying these formatting issues to ensure that the paper meets the highest standards of academic publishing. Thank you for bringing these concerns to our attention.

Tables are updated as per the guidelines:

https://www.natural-hazards-and-earth-system-sciences.net/submission.html#figurestables

7. Upcoming 2013 RAE Report by Cannon et al.

We are eagerly awaiting the updated RAE report by Cannon et al., which promises to shed new light on the evolving landscape of risks and mitigations in relation to space weather impacts on technology. We concur that such emerging research could offer valuable insights that may enhance the context and relevance of our own work. We appreciate the referee's recommendation to consider this forthcoming update. Once the report is available, we aim to incorporate its findings into our future research efforts to keep our work at the forefront of this dynamic field. Thank you for making us aware of this impending publication.

Best regards,

D. Güney Akkor