Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-205-RC2, 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "Predicting power outages caused by extratropical storms" *by* Roope Tervo et al.

Tim Kruschke (Referee)

tim.kruschke@smhi.se

Received and published: 30 September 2020

The manuscript "Predicting power outages caused by extratropical storms" by Tervo et al. presents a novel method to predict the danger of extra-tropical storms to cause power outages over Finland, which is mainly due to windthrow in forest landscapes. Based on meteorological data taken from the ERA5-reanalysis as well as forest inventory data and power outage information from two local power network companies and the national responsible authority, they developed and tested classification schemes potentially suitable for warning purposes by distinguishing between severe damage events, small damage events, and no damage events. This is certainly a very interesting an relevant topic and deserves publication in NHESS. However, I consider a number of modifications necessary before publishing.

General comments:

a) A general shortcoming I notice in the prediction and its evaluation is the lack of any geographical assignment. In principal the predicted event is just "severe damage", "small damage", or "no damage" for Finland as a whole, just complemented by the polygon(s) of the storm objects. From a user-perspective (electric power network providers etc.) the question is if such a prediction is really useful facing the potential consequences, that is the alert of manpower to fix potential damages to power lines which will be rather concentrated in specific regions for most events. Of course it is better than nothing but I am sure that the method could be easily advanced to provide more regionalized information. The least thing that could have been done is to provide information on the detail level of the (power network) input data. This would mean something like "severe damage in local network 1, small damage in local network 2 and region 3 of the national network".

b) I consider the explanations of the tested classification algorithms as too short. Maybe these different methods are self-explaining for members familiar with a variaty of sophisticated classification schemes and machine-learning but I think for the majority of the NHESS-readership which I assume to be with geoscientific background these methods are hard to assess. I would like the authors to provide a little more information about the general functionality, pros & cons, and existing studies in the context of weather and climate having made use of these approaches. For some approaches like the SVC or the GP, some of this information are already given, for others this is hardly the case.

c) Especially for readers with a geoscientific background (as said, probably the majority of NHESS-readership) it would be interesting to read something about the relative importance of the various factors listed in Tab. 1. I understand that this may be quite different for the different classification schemes. But at least for those schemes eventually assessed to yield the best performance a qualitative summary could be listed, may mentioning the five most important factors in order of relevance.

d) As far as I understand, the evaluation metrics in equations (4)-(7) are standard metrics used in the field of machine-learning based classification. However, what I am missing in these scores is any consideration of the distance between predicted and observed class. Clearly a prediction of "severe damage" in cases of no observed damage and vice versa is worse than predicting "small damage" in these cases. But this is not reflected in any penalty for the given scores. Maybe this is a wise solution given that the classes are very different in population. Otherwise an "algorithm" always predicting no damage might be superior with respect to a score taking this distance into account. I would ask the authors at least to comment on this matter and explain why they do not penalyze larger distance between prediction and observation.

e) I wonder why the authors decided to provide deterministic category predictions. This is to some degree a philosophical discussion but given the nature of the task to make a prediction and further supported by i) the rather arbitrary distinction between event classes and ii) the large number of influential factors (some of them considered in the categorization schemes but many more existing in the real world), I wonder why the authors didn't design a scheme that provides probabilities for the distinct event categories. It is often argued that end-users prefer deterministic predictions but it is clearly a fact that predictions such as produced in this study are subject to significant uncertainty. So, why not making this uncertainty transparent by providing related estimates in the form of probabilities? I do not ask the authors to re-design their whole approach but please comment on this issue. Maybe it is worth considering this as a future extension or advancement of the presented approach.

f) A very general issue is that the authors use the term prediction (and so do I in this review) but in fact the presented approach is based on atmospheric REanalysis data, i.e. it relies on data retroactively produced from observations. I would ask the authors to rephrase respective introductory and conclusive remarks in a way that it becomes clear that this study serves as a general introduction of this approach and a proof of concept while a quasi-operational implementation at weather services or

power network providers would have to be based on actual weather predictions which will introduce additional uncertainty to the final product.

g) Some of the figures need optimization. Please see my respective specific comments.

h) I am not a native English speaker myself, so I usually refrain from judging the language used in manuscripts written by others. However, in this example I have the strong impression that the language should be revised. A particular example are frequently missing definite and indefinite articles ("a" and "the"). Other examples can be found in my specific comments.

Specific comments:

1) line 14: Please revise your citation. This is certainly no person with family name "Re" who is cited here but a institutional citation referring to a publication by the Munich Re.

2) lines 20-21: "...up to 69% compared to previous years". What is meant here? Is it an increase of 69% compared to previous years or a total of 69% of outages in 2011/2013v which are associated with windstorms. If the latter is the case, then please delete "compared to previous years".

3) line 27: "Ulbrich et al. (2009)", not "Ulbrich et al. (Ulbrich et al., 2009)"

4) lines 31-33: Please rephrase to make clear that this sentence contains references to studies contradicting the fore-mentioned studies and their results.

5) line 46: Delete "large-scale storms" and "small-scale storms" and just name the meteorological phenomena themselves as they are now listed in brackets. It is misleading to call hurricanes large-scale if then coming to the extra-tropical storms which are even larger in spatial scale.

6) lines 52-55: The purpose of this sentence is a little unclear to me, especially the

reference to the IPCC-SREX-report. It's fine citing this report but not as one of several/many examples supporting this statement. It is basically the probably most comprehensive summary/review of studies indicating this.

7) line 64: Maybe replace "features" by "storm object features" or "storm object characteristics"

8) lines 68-73: Please indicate the purpose of each dataset in this study, e.g. "the ERA5 atmospheric reanalysis (Hersbach et al., 2019) provides the primary meteorological input data for this study...".

9) lines 74-80: Please indicate explicitely which level you use regarding the ERA5 wind data. I guess it is the 10m-winds but this is not said here. Additionally, you may comment on the issue regarding ERA5 surface winds which is described at https://confluence.ecmwf.int/display/CKB/ERA5%3A+large+10m+winds . As far as I can see this does not affect this study as all problematic occasions of unrealistic high wind speed happened at geographical locations far off the study domain. Still I consider this worth mentioning as some readers may not be aware of this issue in general (so the authors could contribute to a more widespread awareness of this problem) and others may be aware of the problem but not its location and related irrelevance for this study.

10) lines 84-88: What is the specific benefit of using the two local datasets on top of the national dataset for this study? Please comment.

11) lines 97-99: The reason given for using a threshold of 15m/s is valid as long as observed winds are considered. However, ERA5-winds are not observed winds, especially regarding gusts. It's basically model results. It should be noted here already, that at least a little bit of sensitivity tests have been performed yielding 15m/s to be the "best" choice. However, the motivation behind this study to develop a scheme which is applicable for quasi-operational forecasts would imply a transfer to a different source of meteorological data, basically weather predictions. Weather prediction models feature

quite different distributions of surface wind speeds. Hence, for such an application a thorough test of the use of this threshold would be necessary. I would like to point out that there are approaches existing in the published literature on wind damage that make use of thresholds which are tied to the specific wind climatology of respective datasets, e.g. by making use of specific quantiles rather than absolute values.

12) line 103: Do you mean "...connected to objects in preceding timesteps"?

13) line 103: Why do you call this "Algorithm 1" if there is no "Algorithm 2"? Why not simply calling it "Storm tracking algorithm"?

14) line 103-104: Maybe I missed something but it seems to me you are not providing any information about the criterion to define/identify a "pressure object".

15) line 103-104: You mention "the threshold" but such a threshold has not been introduced yet. This is done a few lines below. Please rephrase.

16) line 111: "That means that wind objects are not assumed to move..."

17) line 111: "45km" instead of "45km/h"; and please add "from one hourly timestep to another" to the end of this sentence.

18) line 115: "The first group is a number of object characteristics ... which are calculated ..."

19) line 117-118: Please provide more details how you aggregate. Are the minimum/maximum/average values calculated over all grid boxes identified to belong to the storm object (i.e. exceeding 15m/s)?

20) line 118: Replace "over" by "on"

21) line 119: Replace "features" by "characteristics"

22) line 120: Replace "in the damages" by "to the damages", "support" by "complement", and "with weather parameters" by "for weather parameters"

23) line 121: Replace "aggregated from" by "aggregated over".

24) line 124: Here you mention the samples for class 1 and 2 but the class definition has not yet been introduced. This happens in the next section. Please refer to this section and include a very brief definition of the two classes in this sentence, e.g. "severe damage" and "small damage"

25) line 131: Now you introduce the general class definition (no damage, low damage, high damage) but again the exact definition is found at the very end of section 3.3. Additionally, the thresholds used to distinguish between the classes, especially between the two classes containing damage, seems to be completely arbitrary. AT least there is no reason given why the respective number of outages is considered to be low-damage or high damage.

26) Fig. 1: Looking at the red lines in Fig. 1a & b I get the impression that only the lines for the northern local dataset illustrate actual power lines. The lines for the southern local dataset rather seem to be boundaries of sub-regions or so just as Fig. 1b contains region boundaries. I suggest to use different colors for different types of information. The spatial distribution of outages in Fig. 1c & d seems to having been smoothed. If so, please indicate this and the reason for doing so.

27) lines 143-149 (and especially when reading lines 145-146): The reader immediately wonders why the authors stay with the 15m/s-threshold and why this is not analyzed in terms of quantitative measures. A simple example might be hit rates and false alarm rates or so. It is only in Sec. 4 (lines 248-250) that the authors write that storm identification with 15m/s yields a bet\ ter basis for the following classification. Please refer to this later explanation here.

28) lines 155-158: Eventually the class definitions seem to be set arbitrarily. If there is a reason behind the particular thresholds, please name these.

29) lines 160-161: Why is centering and normalization necessary? Probably for some

classification algorithms but not for all of them, right?

30) lines 162-163: Please describe briefly what the application of SMOTE means and why this is beneficial/necessary.

31) lines 204-206: Why did you choose this specific topology? Did you test others? How is the sensitivity of the results to the networks topology?

32) line 236: Please explain the content of the confusion matrices briefly. Again this is probably clear to people profound in machine-learning based classification but not necessarily to the general readership in geosciences. If I understand correctly, it is simply the ratio of cases for each observed class that is show in the cells for the respective predicted classes, right?

33) Section 4.1: This whole section is where my major comment a) becomes visible. If I understand correctly, it is just the event as a whole which is assigned with the respective category, complemented by the polygon of the storm object(s). Is it possible that different objects of one specific event are assigned with different classes? Fig. 6a seems as if this is possible. On the other hand the northeastern object is outside of Finland, so it is clear that there is no damage (to Finnish power lines) observed. In this context it becomes also visible that intra-object refinement of the classification would be desirable. It makes hardly any sense for a prediction of potential damage to power lines (due to windthrow in forests) that the storm objects extend over the Baltic Sea. I understand that this is due to the primary identification being solely based on the exceedence of the wind speed threshold. However, I ask the authors to thin and comment on my general comment a). Additionally, this case study validation refers to observed wind gusts when qualitatively assessing the credibility of these specific predictions. But the authors made it very clear that the potential damage due to a windstorm depends on many more factors, partly non-meteorological but related to the forests themselves. This raises again the question of relative importance of the various factors.

34) lines 306-307: This sentence ignores the fact that the actual study was based on reanalysis data. Using actual weather predictions - which would be necessary for this prospect mentioned here - would introduce additional uncertainty and very likely lead to worse results than derived in the current study. This does not lower the value of the current study but it is worth mentioning when writing about such potential quasi-operational applicability.

35) line 309: Start the sentence with "Including data on..."

36) line 309: I agree that including data about forest soil and leaf index would probably be beneficial but it is questionable if such data is available in sufficient spatial and temporal resolution and coverage.

37) Appendix A: All text elements and axis labels in figures A1 and A2 are hardly readable.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-205, 2020.