Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-205-SC1, 2020

Natural Hazards and Earth System Sciences
Discussions

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

Peter Watson

peter.watson@uconn.edu

Received and published: 5 October 2020

A thoroughly interesting paper. The methodology for identifying storm is especially interesting. However, there may be a few ways to improve the work presented. More specifically:

1) In lines 46 to 48, the authors claim that modeling power outages caused by extratropical events is an understudied problem. However there are actually several papers that describe a power outage prediction system designed specifically for modeling power outages from extratropical storms that are not cited: Yang et al, https://www.mdpi.com/2071-1050/12/4/1525; and Cerrai et al, https://ieeexplore.ieee.org/abstract/document/8656482

2) In figure 4b, it's unclear why the data contains prominent examples where there are very few or no outages, but have a large number of customers affected. Is this trend real, or is it an artifact of noise in the data?

3) By using week as a predictor variable the authors may be over-fitting. For example, to my knowledge, there's no specific mechanism of why a storm on the 42nd week of the year would be particularly strong. But if you had several examples of strong storms on that week, the model would learn that trend and begin to predict strong outages just because of the week, independent of the actual meteorological characteristics of the storms. There are probably other, less problematic ways to describe seasonal aspects of storms to the model.

4) I would recommend a more rigorous and comprehensive method for validating the model. As discussed in the paper, the k-fold cross-validation approach may not sufficiently isolate temporally or spatially correlated information from the model, and thus inflate the model's performance. The 2010 to 2011 holdout approach is presented as alternative to this approach, but the types of storm events that occur often vary widely from year to year. A leave-one-day/week/month/year-out cross validation (where for each day, week, month, or year in the database you hold out that data, train the model on the remaining data, and predict on the withheld data. Then evaluate the model on all of those results) would provide more compelling results.

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