**Natural Hazards
and Earth System
Sciences**

Open Access

Discussions

Interactive comment on "The potential of big data and machine learning for weather index insurance" *by* Luigi Cesarini et al.

Anonymous Referee #1

Received and published: 24 August 2020

The manuscript presents an assessment of two machine learning methods for weather index insurance. My overall impression is that this is a methodologically sound albeit not overly innovative study. It avoids many pitfalls that are sometimes overlooked even in peer-reviewed publications. The train - validation - test split is performed accurately, the problem of class imbalance is tackled adequately (using suitable performance metrics), and presentation quality in terms of figures is good. My recommendation is that this paper can be accepted subject to minor revisions.

I have some minor remarks on the contents and some technical remarks, which I have noted below.

1 General Remarks

- That the authors have put some effort into making the study is easily understandable for people not intimately familiar with machine learning methods. While this is commendable in a journal with a core focus on natural hazards, I feel that the manuscript is a bit lengthy at times. Some parts of sections 1 (Introduction) and particularly 2 (Methodology) could be shortened in order to make them more concise. Some parts read like an introductory book on machine learning. I suggest to go over section 2 again and streamline some of the rather basic parts.

- Please also double check the language throughout the text, especially syntax (e.g. line 289: *Hereinafter is proposed a procedure (...)*).

- I would refrain from using the term 'big data' in this context and adjust the title accordingly. Simply because the authors use larger data sets, this is not a novel big data problem per se.

2 Specific remaks

- Introduction

  - Line 15: I recommend to avoid the term 'significant' in a methodology-oriented paper. This might lead to confusion with respect to statistical significance.

  - Line 65 ff: This section states the core aim of the paper. Please add information on the input data source that is used. Currently, this essential statement is missing. In addition, I suggest to more precisely refer to flood and drought in this statement: '(...) is capable of objectively identifying and classifying

extreme *flood and drought* events *from satellite and gauge data products* in near-real time (...)'.

- Methodology

  – Line 129ff: These 5 criteria are important. I would welcome a reference of the actual values for these five criteria in the text. Maybe add a table featuring spatial and temporal metadata of the datasets used?

  – Line 325: Missing year in Mueller and Massaron

  – It is not ultimately clear which method for tackling class imbalance was used. I realized this when reading the results section, but the authors might want to add a sentence that this was also tuned as a model parameter in the methods section.

  – Since different methods for approaching class imbalance were used: Is there a reasons why procedures for undersampling were not considered? Or combinations such as SMOTE + undersampling?

  – Line 366f: Please check reference (M. and M.N., 2015)

  – Similar to the class imbalance method, it is unclear in the methods section which performance metrics have been used to compare the performance of the model. Was one specific metric used, or was the decision reached using all metrics presented in Tab. 5 by comparing all of them somehow? This is mentioned in the results section, but it is not clear when reading the methods section. I would argue that this is a methodological decision, not a result of the analysis.

  – The reference model on logistic regression is not ultimately clear. Did the authors use simple logisitic regression? Which link function was used? Did the authors include interaction effects? Did the authors use nonlinear effects? Simple logistic regression is fine as a reference model, but I think this could be stated more clearly.

– Is there any particular justification why these two methods were selected specifically? My guess would be that a simple random forest with default parameters would probably perform equally well.

• Results and discussion

– I think more focus on the discussion would be beneficial. Results are described in this section, and findings are briefly commented. However, I am under the impression that there is some imbalance between the first half of the manuscript, which is quite extensive, and the discussion of the results, which is quite sparse. What have we learned from this study? Which novel aspects does this analysis show? What do the results mean for the Dominican Republic? Which impacts do the findings have on the study area?