Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-347-RC1, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Quantification of Continuous Flood Hazard using Random Forrest Classification and Flood Insurance Claims at Large Spatial Scales: A Pilot Study in Southeast Texas" *by* William Mobley et al.

Anonymous Referee #1

Received and published: 23 November 2020

General Comments:

This manuscript develops a random forest model using landscape (hydrology, topography) predictors, training the model on flood insurance claims, to develop flood hazard maps. The authors state this is an improvement over previous efforts by including historic records of flood damage to create a probabilistic floodplain. The paper is well written, and the methods are sound. The year-by-year analysis is unique and useful, in that it demonstrates that the model can be updated with new claims (as the authors note in the Conclusion section). It expands on previous applications of floodplain de-

Printer-friendly version

Discussion paper

lineation via machine learning, a worthy addition to the field. The case study uses Houston, TX, USA, an area often impacted by flooding.

Specific Comments:

L40: although not ML, physics-based models using the shallow water equation should be mentioned briefly, such as Wing et al. (2017). This was completed for the entire CONUS. Or alternatively, this could be included somewhere in the Background section. Wing, O. E. J., P. D. Bates, C. C. Sampson, A. M. Smith, K. A. Johnson, and T. A. Erickson (2017), Validation of a 30 m resolution flood hazard model of the conterminous United States, Water Resour. Res., 53, 7968–7986, doi:10.1002/ 2017WR020917.

L46: this is a statement that needs more elaboration, why do previous sampling methods undermine model reliability? Why is your proposed method superior (and what is it)?

L118-120: a clarification needed here. The sampling to generate pseudo-absences was based on building footprints that did not have an NFIP claim?

L120-125: How are these one-to-one matched on a parcel that has two buildings, i.e. any combination of a house, shed, and detached garage? How were these situations handled, since the claims were parcel-level, not structure-level?

L205-210: error curve vs. number of trees and/or tree depth would be useful for the appendix.

L210-212: Worth stating the features determined important here.

An additional figure that would be useful to the reader would be to compare a location with the modeled flood for that year versus the parcel-level claims for a single year (or multiple panels for two very different years). Perhaps show a coastal area and an inland area, or contrasting coastal, fluvial, and pluvial? This would complement both fig 5 and fig 2.

L230-235: are water areas masked out from the accuracy assessment? That is, there is no prediction on permanent water bodies? Please clarify.

L267: does this mean that a more nuanced model would perform better? That is, isolating areas based on their characteristics (inland vs. coast) and training/testing separate models to improve performance in both areas, and likely resulting in different important variables?

L296: statistical ML model?

Figure A1: increase font size, difficult to read. In addition, I think this would be worthwhile to include in the main text, given the amount of time devoted to variable discussion in the methods.

Technical Corrections: L39-40: spaces missing here.

L53-55: Are these lines necessary?

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-347, 2020.