Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2019-253-AC2, 2019

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Changes in flood damage with global warming in the east coast of Spain" *by* Maria Cortès et al.

Maria Cortès et al.

mcortes@meteo.ub.edu

Received and published: 20 November 2019

Reviewer #2. The paper presents an analysis of the expected change of flood damages under future climate scenarios for two regions on the North-East of the Iberian Peninsula. The authors present a model to estimate the probability of occurrence of damaging events based on daily precipitation and population. The model is calibrated for observed events in the two regions and then it is used to estimate changes in the probability of occurrence of damaging events with a global warming of 1.5, 2 and 3 C and population estimates consistent with SSPs. The topic is relevant because the scientific community is currently addressing flood risk changes under climate change incorporating exposure. In my opinion, the main contributions of the paper are the application of the Generalized Linear Model to obtain the probability of a damaging

event and the set of results obtained for the two regions under different scenarios. I also found interesting the discussion on the role of explanatory variables (precipitation and population) on determining the flood risk. These results are relevant for scientists working on climate risks and the methodology can be extended to other domains. Overall, the manuscript is correctly organized and well written, adequately illustrated with figures and tables. The topic fits well within the scope of Natural Hazards and Earth System Sciences, the objectives are clear and well identified and the conclusions are adequately supported by the results and discussion. Therefore, I think the paper deserves publication in NHESS. I am just offering a few suggestions for minor revisions that could improve the manuscript with a little further work.

Response: We wish to thank the anonymous referee for his/her useful and constructive comments. Each specific point has been addressed in the manuscript as explained in the following document.

Reviewer #2. In my opinion the presentation of the formulation and validation of the GLM model could be improved with some additional explanations. I found the problem formulation a bit confusing. The authors mention "the probability of large damaging events occurring given a certain precipitation amount". The "large damaging event" is related to a certain quantile of the sample of insurance compensations. I understood that the "certain precipitation amount" is 40 mm in 24 h. Therefore, the model estimates the probability of having damages exceeding the threshold when mean precipitation exceeds 40 mm in 24 h. However, as shown in the appendix, basin populations are very different and damages should be evaluated according to the population. How do you account for the fact that basins are heterogeneous in size and have different population densities? Do you assume that the entire basin population is affected by the flood?

Response: As the reviewer correctly mentioned, the model estimates the probability of a damaging event, defined as the exceedance of different damage thresholds, when mean precipitation exceeds 40 mm in 24 h. In the model, we take into account the

heterogeneities of the basin incorporating the population as explanatory variable. As the reviewer commented, we assume that the entire population of the basin is affected by the flood event, since the smallest unit of study is the basin. We also considered the area of the basin as explanatory variable, but it didn't provide information on the response variable. We have added a new sentence in the following paragraph of the article to clarify this point (line 31-32, page 7): "The minimum geographical unit of the study is the river-basin-scale and we only consider the flood cases that recorded a mean precipitation in the basin higher than 40 mm in 24 h. Therefore, the model estimates the probability of having economic damages exceeding different damage thresholds (percentiles) when mean precipitation in the basin exceeds 40 mm in 24 h."

Reviewer #2. The second question is related to model validation through the ROC diagram. I got the impression that the same sample was used for model fitting and for model validation. Could the authors please clarify this point?

Response: In this case, the out-of-sample validation has not been included in the study, only the in-sample validation. So the reviewer is right: we use the whole data to estimate the model and validate the predictions. When dealing with mixed models, the cross-validation can be a bit more complicated, because of the grouped structure of the data. In this study, there are two random effects (basin and flood event). The spatial effect is not significant, so this factor can be negligible for the cross-validation procedure. But the flood event is significant and the number of basins affected from a flood event is not balanced, ranging from 1 to 11, and the cross-validation procedure must take this configuration into account. One method to minimize the impact of such limitation is to use the leave-one-out procedure, where the model is fitted with all the observations except the one used for predicting. When the number of basins affected by one flood event is only 1 (in 32 cases) the random effect for this event is not well estimated, implying a worse prediction for these data. We have used the in-sample validation in the main text of the article since the main objective for the development of the climate present model is to use it for prediction. However, we have added the

results of the area under the ROC curve (RA) using the leave-one-out method in the tables and the ROC diagrams for the examples of the two regions in the Appendix section. We have clarified this in the methodology section adding a sentence in the Validation part (line 32, page 8): "The model validation has been carried out using both in-sample and out-of-sample (i.e. leaving-one-out cross-validation) methods."

Reviewer #2. The third question is related to model application in climate scenarios. The results shown correspond to changes in the probability of damaging events. How are those changes computed? The GLM produces the probability of having a "damaging" event, given a precipitation amount P and a basin population R. Therefore, it produces one probability per event. Since the number and nature of events are different in the control and in the future periods, how are the changes in probability computed using the GLM? I would appreciate if the authors could elaborate on this, since the Methodology section ends abruptly after the presentation of precipitation and population projections.

Response: We have used the same relation between variables obtained in the development of the present climate model for the future model. Therefore, we assume that the relation is stationary, however, the values of the variables change, because we incorporate the projections values for precipitation and population. Any other change in the model would involve collecting new data from future observations that we do not have. In the future, when we will have new data, this model can be estimated again by adding new observations to improve it and check its performance. Regarding the random effects, the model used for the future prediction (considering precipitation and population projections data) is the average model, thus, it is based on the fixed effects and does not consider the random effects, which are assigned to zero, their expected value. In order to clarify this point, we have followed the suggestions of the reviewer and add a sentence after the population projections in the Methodology section (line 15, page 11): "After both precipitation and population data for future scenarios have been corrected, the model developed for the present climate has been used to esti-

mate probability of a damaging event in the future. Since the number of flood events and their nature in the future are unknown, the average model has been applied for the prediction."

Reviewer #2. Regarding the climate projections, the authors mention that they selected 30-year periods of EURO-CORDEX simulations starting from the year when the 20-year running mean exceeds the temperature thresholds. These periods are shown on Table 1. However, the SSP population projections are time dependent, but not temperature dependent. Where the analyses made with a different population for each model? Is this a methodological inconsistency? Could the authors provide a brief discussion on this?

Response: The SSP population projections are certainly year dependent, therefore, we have considered the population, regarding to each SSP, of the year of the observation. As each model has its 30-year period window per each level of global warming, we have considered the precipitation observation of these years as well as the population (for the 5 SSP) of the year of each precipitation observation.

Reviewer #2. Regarding data, the authors mention several data sources to identify flood events in the two regions (INUNGAMA, PRESSGAMA and FLOODHYMEX). They seem to use the events identified in these datasets to obtain the damage data provided by the Insurance Compensation Consortium, with a continuous record 1996-2015. Did you check if there are events with relevant damage data in the ICC dataset not included in the other data sources?

Response: As the reviewer mentioned, we have only considered the compensation produced in the basins affected by a flood event that is registered in the flood databases. The CCS database contain more claims than that we have considered, however they are not always are related with a real flood event. For this reason, we have only taken into account the days when we know that a flood event took place.

Reviewer #2. Apart from the above points, there are a few practical details that could

improve the paper: Pag 3, line 14, "summarises"... should be "summarise"?

Response: We have changed it in the manuscript.

Reviewer #2. Pag 13, lines 2-3, "which was affected by 69 flood events between 1996 and 2015, resulting in 171 flood cases". Which is the difference between a "flood event" and a "flood case"?

Response: A flood case is each basin affected by a flood event. Therefore, each observation of the sample. In the page 8 (lines 5-7) of the manuscript is explained in these words: "For each event there can be more than one set of values, depending on the number of affected catchments. From now on we will use the expression "flood case" for each set of values corresponding to a basin affected by a flood event."

Reviewer #2. Pag 16, line 15, "showed"... should be "shown"?

Response: We have changed it in the manuscript.

Reviewer #2. Pag 22, line 17, "capture"... should be "captured"?

Response: We have changed it in the manuscript.

Reviewer #2. Pag 26, Basin 130 is missing from the list in Table A2.

Response: We have corrected and added a new row in the table. Thank you very much for your observation.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-253, 2019.