Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2017-278-AC2, 2017

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "The relationship between precipitation and insurance data for flood damages in a region of the Mediterranean (Northeast Spain)" *by* Maria Cortès et al.

Maria Cortès et al.

mcortes@meteo.ub.edu

Received and published: 24 November 2017

Manuscript: nhess-2017-278 "The relationship between precipitation and insurance data for surface water floods in a Mediterranean region (Northeast Spain)".

Responses to reviewer #2:

Reviewer #2 (Summary of the article): The manuscript is analysing the link between the causes and impacts of floods by means of precipitation measurements and insurance claims. The main objective of this study is to identify the best indicators for describing this relationship. The topic is of great interest. However, the manuscripts is weak due

to a few but important points.

Response: We would like to thank reviewer for her/his very constructive comments.

Referee's Comment: After this summary lets jump into the study itself. I will try to describe the work with my own words, stopping here and there to add my toughts and concerns. 2.1 Data Three principal sources of information are used by the authors. First the Inungama database from which basic data about flash floods in Catalonia are drawn. These are: âĂć affected municipalities âĂć affected basins âĂć start and end date of the event. An event in the context of this article is therefore, as I understood it, an entry in the Inungama database. At this point the authors know when a (flash) flood happend and which municpialies in which basin were affected. Now the second source of data is entering the stage: the flood damage data from the Spanish Insurance Compensation Consortium (CCS). The event data is at the municipality level and therefore the CCS is aggregated also at that level (which is the finest grain of spatial resolution for this study). By performing a join based on the smallest temporal distance between the event and the date of the insurance claim every event should now also have a variable called Compensations. The last data source is meteorological data from the Spanish State Meteorological Agency. This should add another collumn to the data set with the accumulated 24h precipitation on the event day. Suggestion 1: Because the focus of the article is on flash floods the authors should only include flash floods in their analysis. The difference between flash and non-flash floods must then be stated clearly maybe a (working) definition of flash floods could be based on the length of the event (less than 24 hours). Suggestion 2: Figure 4 is showing the number of floods and the amount of compensation per municipality. Some of the municipalities which have no flood event have compensation payments suggesting a flaw in the homogenisation procedure or simply a graphical one because the legend of figure 4b starts at 0 with a light pink tone.

Response: We wish to thank the anonymous reviewer for the description of the data, which allowed us to improve the explanation of our pre-processing of the data. In the region of study (Mediterranean area) most floods are due to in situ precipitation (surface water floods). For this reason, our hypothesis is that precipitation is the main cause of damaging floods. Most of them are flash floods, events that last less than 24 hours, however this is not true in every case, and for this reason we worked on all the flood events recorded in the INUNGAMA database. In the manuscript we clarified this in the following sentence: “Most floods that have affected the region of study, Northeast Spain, are surface water floods. This type of floods can be regarded as coming under the most general definition of rainfall-related floods (Bernet et al., 2017), including pluvial floods but also flooding from sewer systems, small open channels, diverted watercourses or flooding from groundwater springs (Falconer et al., 2009). River floods that affect great distances are very rare in the region, and are only related to catastrophic and extended floods (for the analysed period only the October 2000 floods were of this type). Nevertheless, these are usually absorbed by reservoirs. It is therefore expected that flood insurance data will correlate strongly with precipitation and surface water floods.” In the revised manuscript we work on a basin level. This domain is large enough to have a fairly large sample size for analysis (we select a total of 221 ”cases”), but small enough that the causes of flood damages are likely to be similar across the area. We also focus on the MAB area, where higher resolution precipitation data are available. In addition, working at a higher level of aggregation allows us not only to reduce possible heterogeneities between the databases, but also to ensure more robust data for each unit.

Referee’s Comment: 2.2 Aggregation. If I got it straight the dataset should consit of entries with the following structure: a flash flood event i affected ni municipalities in mi basins. From the ni affected municipalities ni − ki received compensations of Yi. The anticipated cause of the event is the 24h precipitation Xi,j recorded at the day of the event at station j. Next the auhtors try to find the pair (Yi, Xi,j) which yields the highest correlation in the log-log plane. Let us, for the moment, assume that the hypothesis: payed compensation is a linear function of precipitation, log(Y) = a + b âĹŮ log(X) is true. How could this be physically possible? First the compensation payed to cover

the damage is caused by a flood event. The flood is produced by a stream (may it ephemeral or not) and this stream has a basin. Finnally the precipitation collected by the basin is the fuel for the catastrophic machinery producing the flood. It follows that only the amount of precipitation in the basin of the damage causing-stream should be related to the amount of compensation. Sounds logical tome. The authors find that the maximum precipitation over all affected basins has the highest correlation with the sum of compensations in the affected basins. This is a minor contradiction with the flow of reasoning presented which I assume also the authors used. But further problems may emerge like that the damage itself depends on the number of damageable objects (exposure) in the basin aka at the time of the event. Let us assume that a rainfall of Px is causing the total damage of a building in a basins of size Ax for all buildings with a distance to the stream of say dx then only changing the number of building in the buffer dx will result in considerable difference in the amount of compensation. Suggestion 3: The exposure should be taking into account in other words a relative compensation should be formulated as the response variable in the analysis. Suggestion 4: Adding a scatterplot of precipitation versus compensations for all used aggregation procedures would strongly enhance the understanding of the results.

Response: We would like to thank the reviewer again for the useful suggestion to improve our study. Taking this into account, we have completely reformulated our study. First of all, as the reviewer proposes, we completely agree with the need to include exposure in the data, and, for this reason, we have tested our model using not only absolute damage (D in the manuscript), but also damage per capita (DPC) and damage per unit of gross domestic product (DPW). This means the relative impacts of socio-economic factors on damage can be estimated while taking into account population and wealth responses. Taking into account suggestion 4, we have changed our figures, adding scatter plots for both levels of aggregation (basin and MAB) in order to show the relationship between precipitation and insurance data. In addition, we have added more graphs in the supplementary material with the different thresholds of precipitation used and also considering the Spanish State Meteorological Agency (AEMET) warning areas.

Referee's Comment: 2.3 Results The authors present with figure 5 the key results of the regional analysis. Only guessing from the figure a linear model should be seriously influenced by the obersavtion at x = -1. If the log is the logarithm to the base 10 than this is a precipitation value of 0.1 mm which also seems unrealsitic. The authors also state a precipitation threshhold (100 mm) at which significant damages are observed suggesting that the probability of having a damage above 30.000 is maximized if the precipitation is above 100 mm no further explanation nor quantifcation is given. Suggestion 5: Look at the observation with the low precipitation in more detail. Is it a measurement error? Maybe their is a wrong decimal sign? Is it really a flash flood and is it caused by precipitation? Generally the definition of the analysed data should be made more precise aka the obersavtions should be checked if they belong to the set of interest aka not comparing apples with oranges The analysis on the basin scale is focusing on a black and white example: a basin showing high correlation and therefore supporting the hypothesis of the authors and on the other hand a basin with low correlation contradicting the hypothesis (the mean correlation for all basins is 0.47 (se +/- 0.4) which is rather low). To resolve the low correlation in the black basin the authors split the data set according to a population by maximizing the correlation coefficient turning the black into a white one. Suggestion 6: Using the population as a basis for classifing rural and urban regions reminds me of using a dummy variable in regression from their it is only a slight jump to use population as variable in conjunction with preciptation. Using a ANOVA (or testing against a 0 slope of population or precipitation) would do the trick to see which one of the two is more important. But following suggestion 3 the influence of the population should vanish if and only if the hypothesis of a linear model in only influenced by precipitation is correct. The last subset of observation is the MAB (metropolian area of barcelona) suggesting that a finer temporal grain (30 min) of the precipitation is enhancing the predtiction of compensation payments. Then the precipitation is correlated with the precipitation in 24h which results in a low correlation. Now the whole other data analysis is based on the 24h precipitation but the 30 min seems

to be better suited. What are the implications for the 24h precipitation used for the other data sets? Suggestion 7: Presenting scatter plots are much better suited than maps in my humble opinion. The whole point of the study is the assumption of linearity between precipitation and compensation and simple plot could demonstrade this with elegant ease.

Response: First, the precipitation data went through a quality control process, only taking into account those stations with operations higher than 90% for the period of the study. In addition, different precipitation thresholds (for 24 h and 30 minutes in the case of the MAB) were tested in the model, and their results are shown (in the manuscript and in the supplementary material). As mentioned before, we considered the relative impacts of socio-economic factors on the damage in our models. That is, we consider three damage categories: total damage (D), damage per capita (DPC) and damage per unit of gross domestic product (DPW). For the MAB region we tested the model skill using two different time resolutions for precipitation data: 30 minutes and 24 hours. As shown in Figure 6 of the revised manuscript, the insurance data is more correlated with 30 minute precipitation. For this reason, we used this data in the logistic regression. Unfortunately this data is not available for all of Catalonia. Finally, following suggestions 4 and 7, we have added scatter plots to the manuscript (Figures 2 and 6) and the supplementary material (Figures 4, 5, 6, 7).

Referee's Comment: 3 Final Statement I hope the review was not unpolite and has in any way offented the auhtors which was not at all my purpose. I think the study needs a major overhaul regarding the data preprocessing as well as the techniques used to draw conclusions

Response: We want to show our sincere gratitude for all the comments and suggestions made by the reviewer. They have been very useful and constructive to make substantial improvements to the article.

References:

Bernet, D. B., Prasuhn, V. and Weingartner, R.: Surface water floods in Switzerland: What insurance claim records tell us about the damage in space and time, Nat. Hazards Earth Syst. Sci., 17(9), 1659–1682, doi:10.5194/nhess-17-1659-2017, 2017.

Falconer, R. H., Cobby, D., Smyth, P., Astle, G., Dent, J. and Golding, B.: Pluvial flooding: New approaches in flood warning, mapping and risk management, J. Flood Risk Manag., 2(3), 198–208, doi:10.1111/j.1753-318X.2009.01034.x, 2009.

Please also note the supplement to this comment:
https://www.nat-hazards-earth-syst-sci-discuss.net/nhess-2017-278/nhess-2017-278-AC2-supplement.zip

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2017-278, 2017.

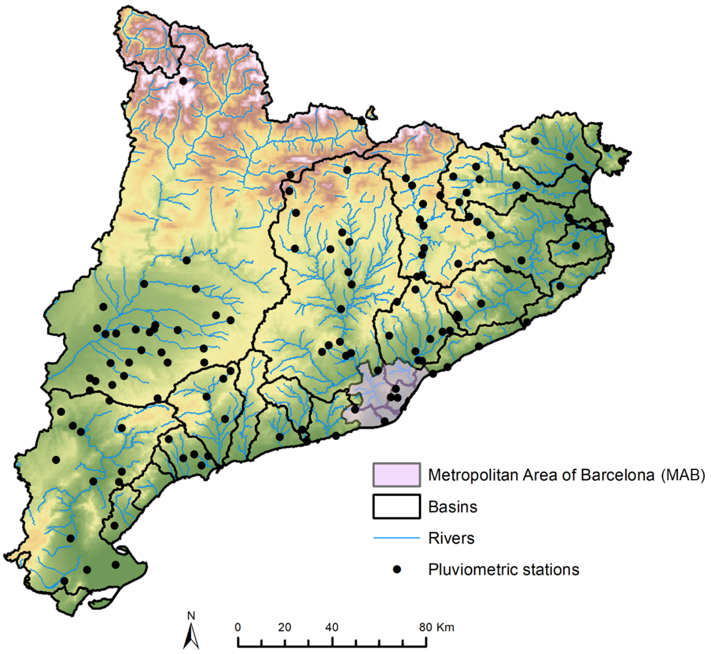

Fig. 1. Map of Catalonia showing the aggregated basins, the Metropolitan Area of Barcelona (MAB), the main rivers and the pluviometric stations used.

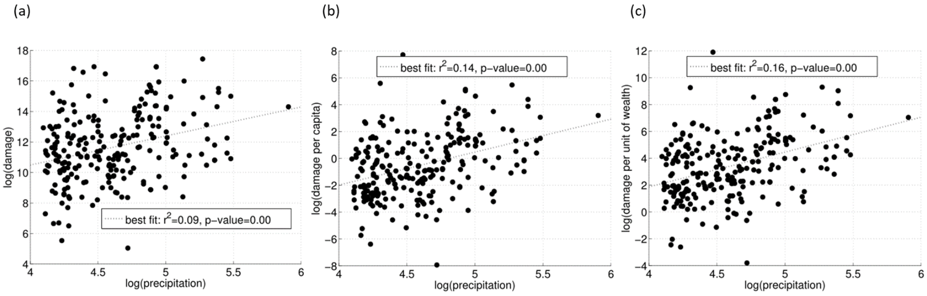

Fig. 2. Scatter plot between basin-aggregated maximum precipitation in 24 h and (a) total damages (D); (b) damage per capita (DPC); and (c) damage per unit of wealth (DPW), for flood events recorded in Catalo

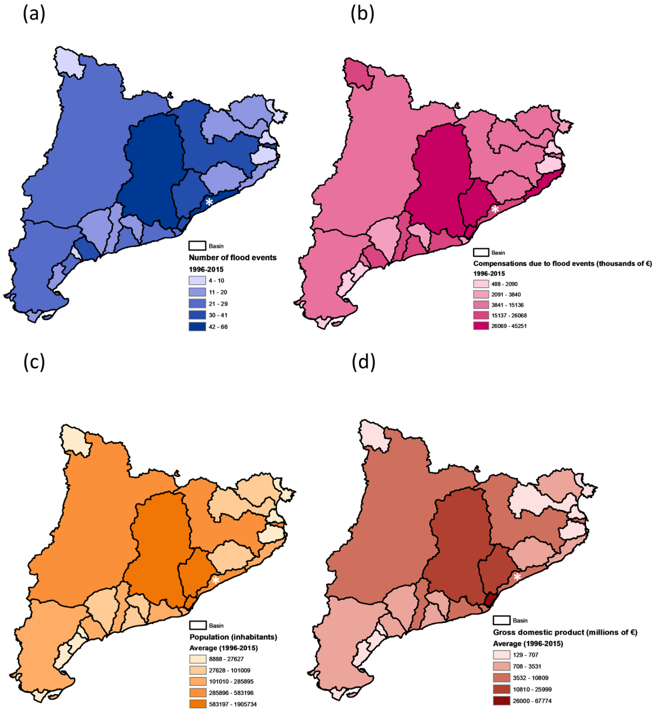

Fig. 3. Basin distribution of (a) flood events (1996-2015); (b) total insurance compensations for floods made by CCS (1996-2015); (c) average total population; and (d) average gross domestic product. Asterisk

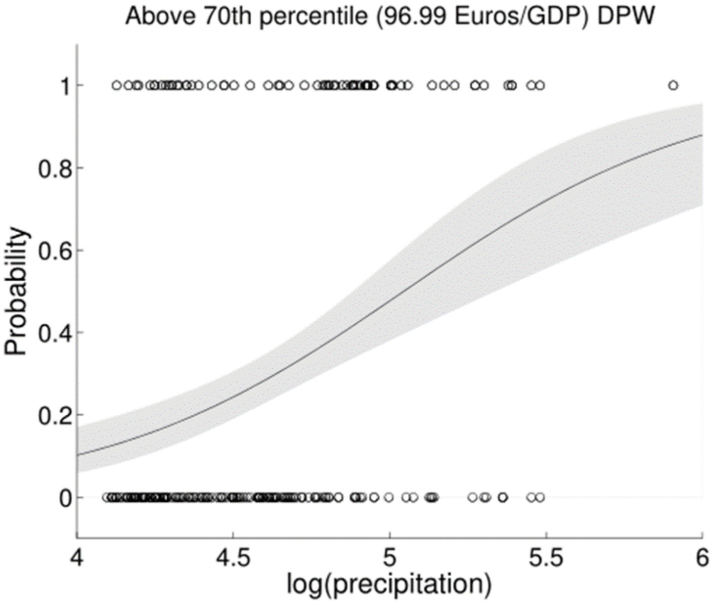

Fig. 4. Example of logistic regression result to model DPW damages above the 70th percentile as a function of precipitation (log-transformed). The solid line indicates the best estimate while the shaded band

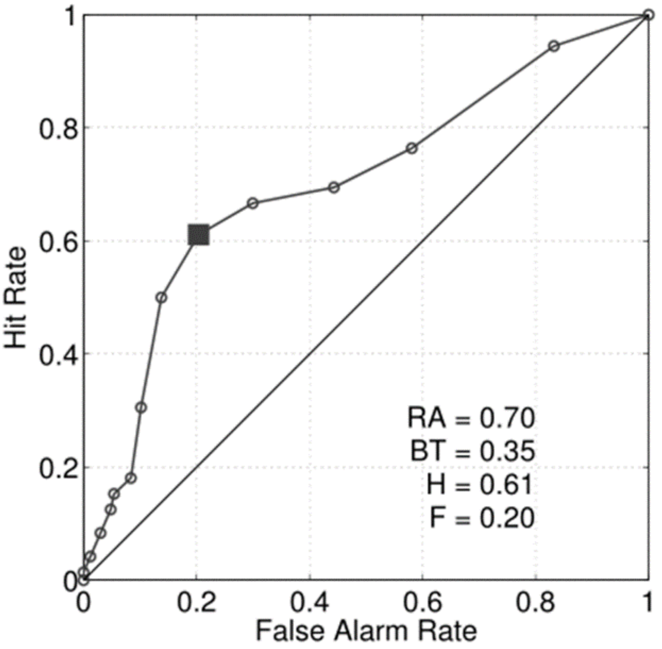

Fig. 5. Relative operating characteristic (ROC) diagram for above 70th DPW predictions using the logistic regression of Eq. (1). The open dots indicate a set of probability forecasts by stepping a decision th

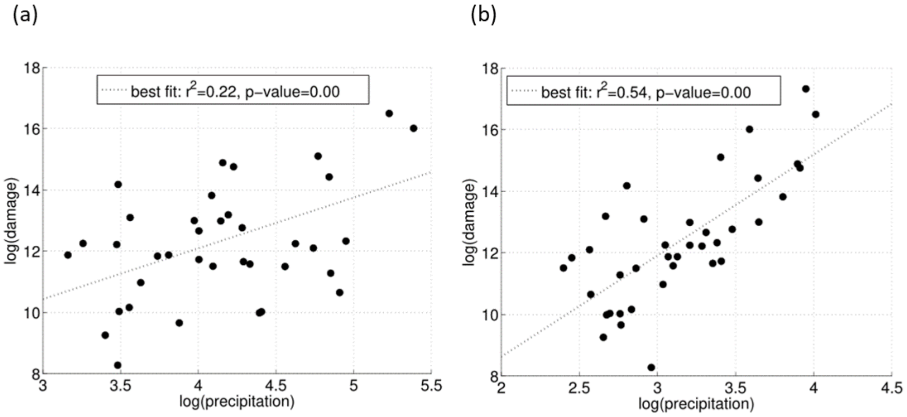

Fig. 6. Scatter plot (a) damages (D) versus 24 h precipitation and (b) damages (D) versus 30 minute precipitation.

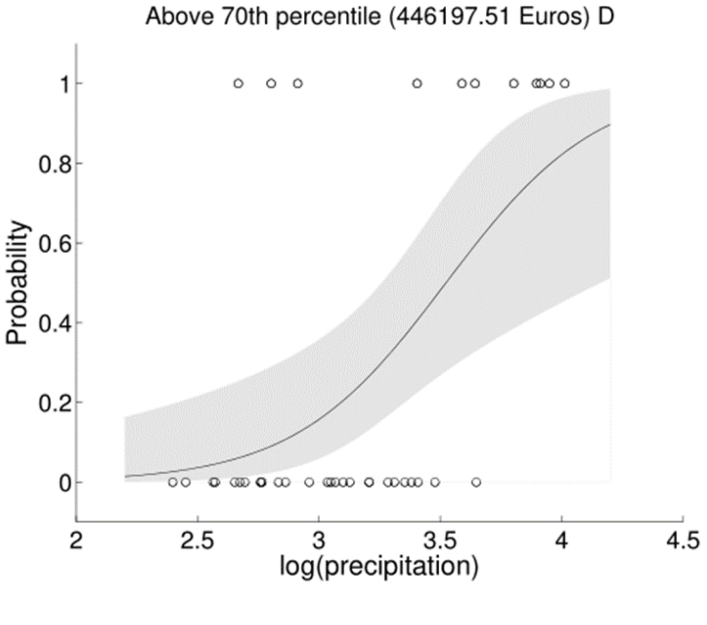

Fig. 7. Example of a logistic regression result to model damages (D) above the 70th percentile as a function of 30 minute precipitation for the MAB. The solid line indicates the best estimate while the shaded

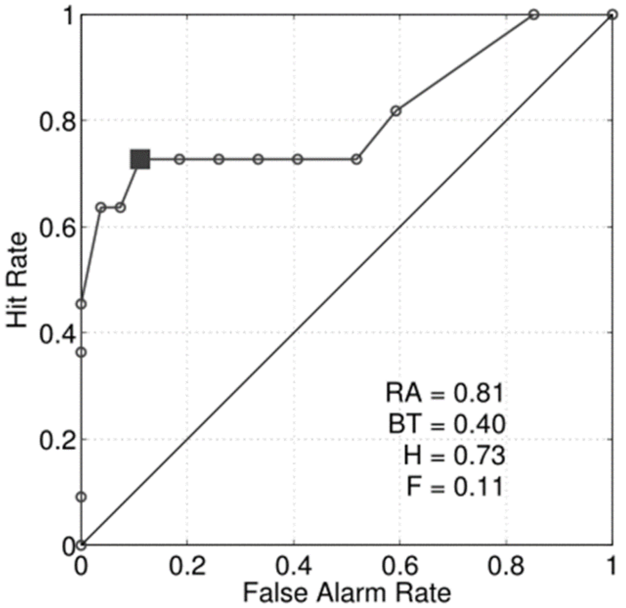

Fig. 8. Relative operating characteristic (ROC) diagram for predictions for damage indicator D above the 70th percentile for the MAB using the logistic regression of Eq. (1). The open dots indicate a set of p