General Comments

The authors have collected a substantial database of insurance damage claims and linked those claims to precipitation measured in the close vicinity. Such data are rare to come by and are certainly of high interest for further study. The analysis performed by the authors is well structured and presented; no fundamental faults could be found in the statistical analysis. The statistical discrimination between "damage/no damage" is nicely done , the fitting of a logistic regression is logical (but see suggestion in specific comments 3) and discussed in considerable depth. The results are certainly worth publishing.

It seems to me though there are many other aspects of the data that could have been also studied (see specific comments 1 and 2). I can appreciate that the collection of the data and the presented analysis already took considerable effort, but I would recommend to state somehow in the conclusions section and/or point out in the introduction that further study could be made and that this is only a first analysis of the data focusing on a specific aspect.

Specific Comments

1. As far as I could see, the analysis focuses on a single aspect: establishing a relationship between the probability of pluvial flooding occurring (i.e. the number of claims is significantly higher than can be statistically expected for claims on dry days) with precipitation at the nearest gauge station (if within a certain distance, i.e. 10 km). The statement in the abstract that a relation is established between "high numbers of damage claims ... with high rainfall intensities" becomes clear later but is somewhat misleading at the outset, since the data are reduced to the binary outcome: no damage/some damage. A statement "probability of flood damage ... with the intensity of rainfall" would seem to me to cover better the real content of the study. If the authors agree, I would suggest to point out (for instance in the conclusions and recommendations) that further study of the data could be of interest: i.e. is there a significant statistical relationship between the fraction of insured buildings that have been damaged and the precipitation for events that are labeled as "damage events"; is there a significant relationship between the aggregrate claim of all damaged buildings and the precipitation. These relationships would be of considerable interest for simplified flood risk calculations.

2. The interpretation and discussion of the results appears to be focused on a 'damage forecasting' application and for such an application the results appear to be not very encouraging (as stated implicitly by the authors). As in 1. above, perhaps it should be pointed out that for other applications where no real-time forecasting is needed, but only the expected frequency of floods and a measure of the flood damage needs to be estimated (i.e. to assess the frequency of total losses for a given portfolio, to examine how changes in precipation frequency due to global climat warming would affect losses) the results in Figure 3 and the very high significance that is found are quite encouraging.

3. I would suggest to say that the logistic model in Equation (4) is only one of many possibilities, i.e. below the header of 2.5, one might say "The binary outcome ... can be linked ... using various types of logistic models. Here we choose to use ..."

My doubt is not so much about the choice of the logistic function (i.e. instead of the probit function), but on the use of the precipation value I instead of its logarithm ln(I). Using ln(I) is a more logical choice, because for I=0 the condition that teta=0 is then automatically satisfied. Relating the logit(teta) to I does not achieve this and puts some unnecessary constraint on the intercept and slope parameter. Note, that if one uses the model in ln(I), the model can be also rewritten in a parametric form that is more easily interpreted:

odds-ratio = (I/I_0.5)^beta

with I_0.5 representing one parameter (the precipitation at which the probability of damage is 50%) and the second parameter beta determining how quickly this probability increases as I becomes larger than I_0.5. It would be interesting for instance to see whether I_0.5 and/or beta when estimated for a single rain-gauge region shows significant variations between different regions and whether these differences can be related to some region-attribute (average landuse, area, ...).

Technical Corrections.

1. page 11620, line 15: "close to (not further defined)". I had to read this a couple of time to understand. Suggestion: "at a rain gauge station that is close to (without close being precisely defined) the location"

2. page 11623, line 9: "maximum rainfall intensity". "maximum" is previously defined, but for those with a short memory it might read easier " maximum rainfall intensity (maximum within 1 day for the chosen time-window Z)".

3. page 11625, line 9, and caption of Figure 3: "one standard deviation of uncertainty". Suggest to add "on the empirical proportion estimate"

4. page 11633, caption Table 2. "z= 60 mm h^-1". I don't understand this. Should it be "z= 60 min" and add "b_1 has units h/mm"?