Manuscript: nhess-2017-278
OLD TITLE "The relationship between precipitation and insurance data for flood damages in a region of the Mediterranean (Northeast Spain)".
NEW TITLE "The relationship between precipitation and insurance data for surface water floods in a Mediterranean region (Northeast Spain)".

Dear Dr. Thomas Thaler,

In accordance with your instructions, we submit the revised version of the manuscript nhess-2017-278 for consideration for publication in Natural Hazards and Earth System Sciences (NHESS), taking into account the reviewers' comments. We thank the reviewers and the editor for their thoughtful suggestions which have helped us to significantly improve how the analysis is presented.

The most evident changes applied to the manuscript are summarised below:
i) We have applied a logistic regression model to gauge the probability of large damaging events given a certain precipitation amount.
ii) We have included exposure in the model, taking into account the population and the gross domestic product in the response variable of the model (insurance data).

We are confident that these major changes have improved the statistical significance of the analyses, and have improved the clarity of the presentation. Most importantly, the changes have strengthened the main message of the previous version and make the key findings more robust.

We enclose the revised version of the paper, a letter with detailed responses to all the editors and reviewers' comments, and a version of the article highlighting the differences between the original and the revised versions.

We hope that the revised manuscript now meets the criteria for publication in NHESS.

Sincerely,

Maria Cortès, on behalf of the co-authors

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

Responses to reviewer #1:

Reviewer #1 (Highlight): *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 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.

Referee's Comment*: The use of the Spearman rank test on the data is to test the correlation between two indicators. However, this test is not providing any information for drawing conclusion. The correlation between precipitation and flood losses is to be expected and reported anywhere. Indeed, other literature goes far behind correlation only and analyses the form of the relationship resp. provides e.g. minimal rainfall intensity thresholds for losses. The added value of this study is therefore disputable.*
Response: We would like to thank you for this very important comment. To address it, we applied a logistic regression model to gauge the probability of large damaging events given a certain precipitation amount, an approach that is frequently used for this kind of modelling study (Kim et al., 2012; Wobus et al., 2014).

Referee's Comment*: A detailed definition of the terms "floods", "pluvial floods", "flash floods" and "urban floods" is missing. It is not clearly defined which processes are relevant for the respective insurance claims. This is important because of the chosen approach in defining the aggregation units. Depending on the size of the aggregation unit, the spatial distances between rainfall (in the catchment) and flood impacts (in the floodplain) may be very different for riverine floods and pluvial floods (in situ precipitation).*
Response: The insurance company (CCS) offers insurance compensations for all the claims related to flood damages, regardless of the type of flood. Nevertheless, the floods that most frequently affect the region of study, the Mediterranean, are caused by in situ precipitation (pluvial floods or, in broader terms, "surface water floods") (Llasat et al., 2014). Surface water floods can be thought of as the most general form of rainfall-related (pluvial) floods (Bernet et al., 2017).
For this reason, the hypothesis of the study is that precipitation is the main cause of damaging floods in Catalonia, and it is expected that flood insurance data will show a strong correlation (Cortès et al., 2017).
Taking into account the limitations of the data, we decided to work on a basin scale. For each basin and event, the maximum 24 h precipitation and total flood compensation was estimated. In the case of the MAB, the availability of precipitation data for lower time intervals (30 minutes) allowed us to make a comparison for both results.
Finally, we added this sentence in the manuscript:

> *"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."*

Referee's Comment*: The authors chose three different spatial aggregation units: regional, basin, and local scale. With this, the study ignores the Modifiable Areal Unit Problem (MAUP) as described in Openshaw (1984). The size and shape of the aggregation units may influence the results of the test. This arises because the authors do not explicitly differentiate between losses to houses due to pluvial floods and riverine floods as exemplarily shown by Bernet et al. in this special issue (https://doi.org/10.5194/nhess-2017-136).*
Response: As mentioned before, we decided to work by aggregating the data on a basin level and for the MAB region, because the available data are too sparse to support our statistical assessment on a municipal scale. On the other hand, the floods in question, as we have explained before, are pluvial floods. Furthermore, we considered another spatial aggregation based on the Spanish State Meteorological Agency (AEMET) warning areas, obtaining similar results (included in the supplementary material).

Referee's Comment*: The interpolation of the rainfall measurements of the single stations and the aggregation method to the different spatial units is not described.*
Response: Due to the existence of few meteorological stations, for precipitation data we used the maximum over 24 h recorded for each basin and event. There was no interpolation process. In order to better clarify this point, we add:

> *"Because the available data are too sparse to support our statistical assessment on a municipal scale, we assessed the precipitation-compensation link for Catalonia as a whole. That is, we sampled pairs of the response variable (i.e. the compensation series) and the maximum 24 h precipitation for each basin, and pooled them into one sample for the entire region (Catalonia) to correlate them. For each event there can be more than one pair of values, depending on the number of affected catchments. From now on we will use the expression "flood case" for each pair of values corresponding to a basin affected by a flood event. This area is large enough to have a fairly large sample size for analysis, but small enough that the causes of flood damages are likely to be similar across the area."*

Referee's Comment*: The authors analyse insurance claims in the period 1996-2015. While the compensations are adjusted with the consumer price index, the increase in the total stock resp. changes in the overall exposure to potentially flooded areas in this period are not considered. In the study period, a relatively high increase in the total building stock should be expected due to the construction activities before the financial crisis in 2008. Thus, the losses in the insurance dataset may be supposed to a remarkable instationarity.*
Response: We agree with the reviewer. Modelling insurance compensations is a complex issue due to the limitations in observational data and the concurrence of a variety of factors that affects them. For instance, in order to statistically link rainfall and insurance compensation, both

precipitation and monetary data would need to be compiled precisely and consistently over time and across the region. While precipitation data follows a formal quality control, the data for insurance compensations are no standardised. For instance, the value of assets exposed and insurance coverage may change over time (Barredo et al., 2012). Unfortunately, precise data on the value and location of assets exposed are not available. However, we show that rainfall data can be used to extract information on damages in Catalonia. To do so, we applied a logistic regression model to gauge the probability of large damaging events given a certain precipitation amount. That is, our aim is not to estimate the precise magnitude of the monetary compensation, but to estimate when a "large" damaging event will occur given a certain precipitation amount. In addition, we take into account the relative impacts of socio-economic factors on damage, considering not only the total damage, but also the damage per capita (DPC) and damage per unit of gross domestic product (DPW).

Referee's Comment*: The analysis of the correlation between precipitation and compensation paid was made on the basis of the recorded flood events (or flood episodes). The definition of a flood event (above 75th percentile) is not made transparent resp. not clearly enough described.*
Response: We have changed this part of the study. In order to define a flood event we used the INUNGAMA database (Barnolas and Llasat 2007; Llasat et al., 2016a), from which we took the event dates and duration. This database records all the flood events that have affected Catalonia, most of them caused by in situ precipitation (surface water floods).

Referee's Comment*: The paper has to be reworked fundamentally.*
Response: The revised manuscript has been thoroughly rewritten. One of the most important points is that we now model the compensations with a logistic regression strategy, testing the sensitivity of our results to the different threshold used to define the events. We have carried out a major overhaul of the data processing and the methods and techniques applied. We are confident that these major changes have improved the statistical significance of the analyses, and improved the clarity of the results presented.

References:

Barnolas, M. and Llasat, M. C.: System Sciences A flood geodatabase and its climatological applications: the case of Catalonia for the last century, (2005), 271–281, 2007.

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.

Cortès, M., Llasat, M. C., Gilabert, J., Llasat-Botija, M., Turco, M., Marcos, R., Martín Vide, J. P. and Falcón, L.: Towards a better understanding of the evolution of the flood risk in Mediterranean urban areas: the case of Barcelona, Nat. Hazards, 1–22, doi:10.1007/s11069-017-3014-0, 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.

Kim, Y. O., Seo, S. B., & Jang, O. J. (2012). Flood risk assessment using regional regression analysis. Natural hazards, 63(2), 1203-1217.

Llasat, M. C., Marcos, R., Llasat-Botija, M., Gilabert, J., Turco, M. and Quintana-Seguí, P.: Flash flood evolution in North-Western Mediterranean, Atmos. Res., 149, 230–243, doi:10.1016/j.atmosres.2014.05.024, 2014.

Llasat, M. C., Marcos, R., Turco, M., Gilabert, J. and Llasat-Botija, M.: Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia, J. Hydrol., 541 (September 2002), 24–37, doi:10.1016/j.jhydrol.2016.05.040, 2016a.

Openshaw, Stan, 1984. The Modifiable Areal Unit Problem. Geo Abstracts Univ. of East Anglia, Norwich (Catmog: concepts and techniques in modern geography).

Wobus, C., Lawson, M., Jones, R., Smith, J., & Martinich, J. (2014). Estimating monetary damages from flooding in the United States under a changing climate. Journal of Flood Risk Management, 7(3), 217-229.

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:
- *affected municipalities*
- *affected basins*
- *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,
$log(Y) = a + b * log(X)$
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.

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

Responses to reviewer #3:

Reviewer #3 (Article summary): *The article analyses the correlation between extreme rainfall events and compensation costs triggered by flash floods, which are drawn from insurance records. The correlation coefficient is used to draw conclusions on the causal effect between precipitation and damage magnitude, using different scales of aggregation as tests.*
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 below.

Referee's Comment*: The topic is of great importance and the use of empirical data is a plus, however the thinking behind the paper is a bit too much straightforward. Indeed precipitation is a major driver of flash flood damage; but it is not the only factor. The paper do not take in account other factors influencing damage (slopes, land cover and soil sealing, vegetation), and explains the effect (damage) by stating the cause (heavy rainfall); the conclusion uses correlation values to confirm the hypotesis.*
Response: We would like to thank you for this very important comment. To address it, we applied a logistic regression model to gauge the probability of large damaging events given a certain precipitation amount, an approach that is frequently used for this kind of modelling study (Kim et al., 2012; Wobus et al., 2014). Since most of floods that affect this region are caused by in situ precipitation (surface water floods), our hypothesis is that precipitation is the main cause. We agree with the reviewer that other factors can influence the insurance compensations for floods. For this reason we now consider three damages categories: (i) total damages (D), (ii) damage per capita (DPC) and (iii) damage per unit of gross domestic product (DPW). In this way the relative impacts of socio-economic factors on damage can be estimated, while taking into account population and wealth (Zhou et al., 2017).

Referee's Comment*: The statistical analysis needs to go deeper and to add more insights in relation to the distribution of damage along different typologies of exposure. The analysis uses 4 different aggregation scales based on admistrative units; most commonly in these kind of studies the scale would be smaller than the municipality. A projection of the data over built-up areas from land cover, building units or a regular grid cells would improve the analysis by linking the variables at a more detailed and homogeneus unit compared to the administrative boundaries. I would suggest then to present only the results relative to the better performing aggregation method, as the comparison on administrative units do not produce added value for the conclusions.*
Response: We have taken into account these useful comments on the methodology of our study. Our current model includes exposure in the damage data, as mentioned before. Following the suggestion of the reviewer, we have also aggregated the data on a basin level. In the supplementary material we also include the results based on the warning areas used by the Spanish State Meteorological Agency (AEMET). Our aim is not to estimate the precise amount of insurance payments made, but to estimate when a damaging event will occur given a certain precipitation amount. For this reason, we have applied a logistic model for different precipitation thresholds and types of damage in order to know when a damaging event occurs.

Referee's Comment: *The difference between different kind of floods and to which kind exactly the compensatory records refer is not clearly stated in the paper. Overall, both the record data and the spatial data needs to be presented more precisely and clearly.*

Response: In our model we have used all the flood events recorded in the INUNGAMA database. This database records the flood events that have affected Catalonia, most caused by in situ precipitation (surface water floods). For this reason, our hypothesis is that precipitation is the main cause of damaging floods. However, bearing in mind the possibility of having insurance data related with river floods, we used different precipitation thresholds in the model.

In order to clarify this point, we added this sentence to the revised manuscript:

> *"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."*

Referee's Comment: *Maps can be easily reduced in numbers and made more readable: figure 1 "a" and "b" can be combined by showing only the necessary information (river, basins, population, scores). Same goes for figure 3, it could be combined into 1 or 2 showing the information (dots) in different shapes/colors.*

Response: We have followed the suggestion made by the reviewer and have changed the figures in the manuscript in order to make the paper more readable.

Referee's Comment: *Finally, I agree with the insightful comments by reviewer 1 and 2 and suggest to majorly revise the paper by rethinking its objectives and methods.*

Response: Taking into account all the comments and suggestions made by the reviewers, we have completely rewritten the manuscript, and we are confident that these major changes have improved the statistical significance of the analyses.

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.

Kim, Y. O., Seo, S. B., & Jang, O. J. (2012). Flood risk assessment using regional regression analysis. Natural hazards, 63(2), 1203-1217.

Wobus, C., Lawson, M., Jones, R., Smith, J., & Martinich, J. (2014). Estimating monetary damages from flooding in the United States under a changing climate. Journal of Flood Risk Management, 7(3), 217-229.

Zhou, Q., Leng, G. and Feng, L.: Predictability of state-level flood damage in the conterminous United States: the role of hazard, exposure and vulnerability, Sci. Rep., 7(1), 5354, doi:10.1038/s41598-017-05773-4, 2017.

The relationship between precipitation and insurance data for surface water floods ~~damages~~ in a ~~region of the~~ Mediterranean region (Northeast Spain)

Maria Cortès, Marco Turco, Montserrat Llasat-Botija, M. Carmen Llasat

Department of Applied Physics, University of Barcelona, Barcelona, 08028, Spain

***Corresponding author*:** Maria Cortès, Department of Applied Physics, University of Barcelona, Martí I Franqués 1, 08028 Barcelona, Spain,mcortes@meteo.ub.edu

The relationship between precipitation and insurance data for surface water floods ~~damages~~ in a ~~region of the~~ Mediterranean region (Northeast Spain)

Maria Cortès, Marco Turco, Montserrat Llasat-Botija, M. Carmen Llasat

Department of Applied Physics, University of Barcelona, Barcelona, 08028, Spain

Correspondence to: Maria Cortès (mcortes@meteo.ub.edu)

Abstract. Floods in the Mediterranean region are often flash floods, ~~where~~ in which short and intense precipitation is usually the main driver behind the events. Determining the link between the causes and impacts of floods can ~~help better~~make it easier to ~~characterise~~ calculate the level of flood risk. However, up until now, the limitations in quantitative observations for flood-related damages have been a major obstacle when attempting to analyse flood risk in the Mediterranean. Flood-related insurance damage claims for the last 20 years~~,~~ which could provide a proxy for flood impact, and this information~~cover the last 20 years, are~~ is now available in the Mediterranean region of Catalonia, in northeast Spain. This means a comprehensive analysis of the links between flood drivers and impacts is now possible. The objective of this paper is to develop and evaluate a methodology to estimate flood damages from heavy precipitation in a Mediterranean region. Results show that our model is able to simulate the probability of a damaging event as a function of precipitation.~~to analyse the possible relationship between precipitation and flood damage compensation for the period of 1996-2015. Results show high correlation values between daily precipitation and insurance data on a regional, basin and local scale. These results confirm the hypothesis that precipitation is the main contributing factor to damages caused by flash flood events~~. The relationship~~s~~ between precipitation and damage ~~shown~~ provides insights into ~~the~~ flood risk in the Mediterranean and ~~are~~ is also promising for supporting flood management strategies.

1 Introduction

Flooding is the main natural risk in the world. Between 2005 and 2014, more than 85,000,000 people were directly affected by flood events annually, and around 6,000 people were killed on average each year due to floods (UNISDR, 2015). The main factors involved in flood risk analysis are the hazard, or the likelihood of a natural phenomenon causing damages, and the vulnerability, that is, the characteristics and circumstances of a community/system that make it susceptible to potential flood damage (UNISDR, 2009; Kundzewiczet al., 2014; Winsemiuset al., 2015). Vulnerability can include factors such as exposure and other societal factors such as early warning systems, the construction capacity to cope with natural hazards, and disaster recovery ~~capabilities~~ infrastructure (Jongman et al., 2014; Nakamura and Llasat, 2017).

A large number of authors are making efforts to create ~~appropriate~~ methodologies that are able to analyse the impacts of floods, due to the significant consequences of this phenomenon (Messner and Meyer, 2006; García et al., 2014). Indeed, progress is being made on incorporating ~~the~~ impact and vulnerability analysis in flood risk assessment, although the limitations of the impact data (availability and quality) make it difficult to carry out these studies (Elmer et al., 2010; Petrucci and Llasat, 2013; Jongman et al., 2014; Papagiannaki, et al., 2015; Thieken et al., 2016; Kreibich et al., 2017).

Insurance data may provide a good proxy for describing flood damages (Barredo et al., 2012). Several recent works have used this kind of data to explore ~~flood drivers~~the causes and impacts of floods. For instance, in several European regions researchers have noted ~~a significant influence of~~that precipitation has a significant influence on flood insurance data (see for instance Spekkers et al., 2013, 2015 for Netherlands; Zhou et al., 2013 for Denmark; Sampson et al., 2014 for Ireland; Moncoulon et al., 2014 for France; Torgersen et al., 2015 for Norway). This data is very valuable for establishing causal relationships between the costs of flood damage and precipitation extremes, for developing risk maps, and ~~for being~~to use~~d~~ as a validation tool for damage models (Zhou et al., 2013). These studies agree on the potential of insurance data to assess the damage caused by pluvial and urban floods.

The Mediterranean region is prone to flash floods, where torrential rain concentrated in small catchments can ~~turn~~ result in extraordinary runoffs and cause catastrophic damage (Llasat et al., 2014, 2016a). 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 floods (October 2000 was the only catastrophic river flood event that affected Catalonia for the period of study). Nevertheless, these are usually absorbed by reservoirs. It is therefore expected that flood insurance data will ~~be strongly related to~~ correlate strongly with precipitation and surface water floods~~heavy rainfall~~. However, relatively few studies exist for this region (Freni et al., 2010;~~,~~ Papagiannaki et al., 2015; Bihan et al., 2017). This may be due to limitations in insurance data records and difficulties in estimating how heavy precipitation could affect monetary damages~~assessing the reliability of the spatial scale of the data~~.

In the Mediterranean region of Catalonia, in Northeast Spain, 20 years of flood-related insurance damage claims are available~~In Spain, insurance data can be only collected~~ from the Spanish public reinsurer, the "Insurance Compensation Consortium" (*Consorcio de Compensación de Seguros*, or CCS), a public institution that compensates homeowners for ~~the~~ damage ~~produced~~ caused by floods, which plays a role similar to that of a private insurance company (Barredo et al., 2012). This means an assessment of the links between flood causes and impacts is now possible. The aim of this study is to develop and evaluate a methodology to estimate surface water flood damages from heavy precipitation in this Mediterranean region (from now we will use the expression “flood” to refer to surface water floods). The relationship between precipitation and insurance data has been assessed using logistic regression models for the probability of large monetary damages conditioned to heavy precipitation events. In order to determine the best applicability range, we consider different thresholds to define large flooding damages and heavy rainfall events. Also, to analyse flood damages, we consider the role played by exposure and vulnerability (through the commonly used proxies of Gross Domestic Product) and population (Pielke and Downton, 2000; Choi and Fisher, 2003; Barredo, 2009) to determine the damages corresponding to precipitation events. ~~However, they only provide aggregated~~

~~data and it is not possible to carry out analysis as detailed as for central Europe, or to access information collected during post-event surveys (Elmer et al., 2010). Floods account for more than 70 % of total insurance compensation paid out for extraordinary natural risks in Spain for the period of 1971-2015, amounting to a total of €5,564.3 million (CCS, 2016).~~

~~In this study we analyse the possible relationship between precipitation and damages caused by flood events in an area affected mainly by flash floods: Catalonia, in northeast Spain (Llasat et al., 2014). The starting hypothesis is that in this region precipitation is the main factor responsible for damage, since the majority are episodes produced by in situ precipitation and flash floods. The relationship between precipitation and insurance data is assessed considering different spatial aggregation of the data, determining the best range of applicability. Specifically, this study considers three difference spatial scales belonging to what is called a meso-scale within the scales described in Messner and Meyer (2006): (i) regional (Catalonia as a whole), (ii) basin (catchments in Catalonia) and (iii) local (the Metropolitan Area of Barcelona).~~ The results of this study can help to better understand flood risk in Mediterranean areas by analysing ~~drivers~~ causes and impacts, and ~~specifically~~ can help to more accurately estimate~~ing~~ flood damage when high rainfall ~~amounts are~~is forecast.

The study is organised as follows. After the Introduction, the section on “Methods” describes the study region, the observed data and the methodology used. Then, the “Results” section presents the regression models obtained. ~~the results obtained for the three different scales.~~ Finally, the “Conclusions” section summarises the main findings of this study.

2 Methods

~~This study has been carried out on different spatial scales, from the entire region (Catalonia) to basins and local scales (Barcelona) (Fig. 1). The period of the study is 1996-2015, when insurance data is available for damages.~~

2.1 Study region

The study area is Catalonia, a Spanish region of 32,108 km^2 in the northeast of Iberian Peninsula. The region ~~First of all we considered Catalonia as a whole. This region is located in the northeast of the Iberian Peninsula. Its surface is 32,108 km^2 and it~~ is characterised by three mountain ranges (Fig. 1~~a~~): the Pyrenees in the north (maximum altitude above 3,000 MASL) and parallel to the Mediterranean coast (SE-NE) between the Pre-Littoral mountain range (maximum altitude around 1,800 MASL) and the Littoral mountain range (maximum altitude around 600 MASL). This marked orography is the key reason for the development of flash floods, both from a hydrological point of view (small torrential catchments) and due to meteorological factors (~~for example,~~ the orography forces water vapour to rise from the Mediterranean, triggering instability; Llasat et al., 2016a). The region is divided into 42 districts and 948 municipalities, with a total population of 7.5 million, most of them living along the coast (Fig. 1b), where more than 70 % of the flood events occur (Llasat et al., 2014), ~~which makes~~making it a very vulnerable area. From a hydrological point of view the region is divided into 31 basins, most of them with surface areas of less than 500 km^2. Some of them are formed by very small municipalities for which some data needed is not available (i.e.

Gross Domestic Product, GDP). For this reason we have aggregated some of the basins and worked with a total of 29 (see supplementary material).

~~At a basin scale, we analysed the Catalan basins that have recorded the highest number of flood events, which are: Maresme, Llobregat, Besòs, Ter, streams in the MAB, Tarragona Sud and Segre. Next, we studied the Maresme basin in detail, where the greatest number of flood episodes was registered (68 for the period of study) and the Ter basin, where a total of 38 flood events were recorded between 1996 and 2015.~~

We also analyse ~~Finally, we considered~~ the Metropolitan Area of Barcelona (MAB, 534.7 km^2) (Fig. 1~~a~~) in detail, which consists of the city of Barcelona (1,608,746 inhabitants in 101.3 km^2) and 35 municipalities. Although it represents less than 2 % of the surface area of Catalonia, ~~the area~~it contains 48 % of the population (IDESCAT, 2016). It is affected by an average of ~~more than~~ over 3 flood ~~episodes~~ events per year, most of which are flash floods due to very convective local precipitation (Cortès et al., 2017~~Llasat et al., 2014~~). The city of Barcelona is crossed by 20 streams ~~with~~ that have their source in the Serra de Collserola (Littoral mountain range), and ~~they~~ which are covered as part of the Barcelona drainage system, managed by the Barcelona Water Cycle (*Barcelona Cicle de l'Aigua* or BCASA). The United Nations International Strategy for Disaster Reduction (UNISDR) marked Barcelona as a resilient city and a model city for dealing with floods (Nakamura and Llasat, 2017), as it has a permanent surveillance and warning system running on hydraulic modelling that includes 15 rainwater tanks (13 underground and 2 open) that allow for better flood prevention. As a result, flood damages have decreased over time (Barrera-Escoda et al., 2006) while the daily rainfall threshold associated with damaging floods has increased (Barrera-Escoda and Llasat, 2015).

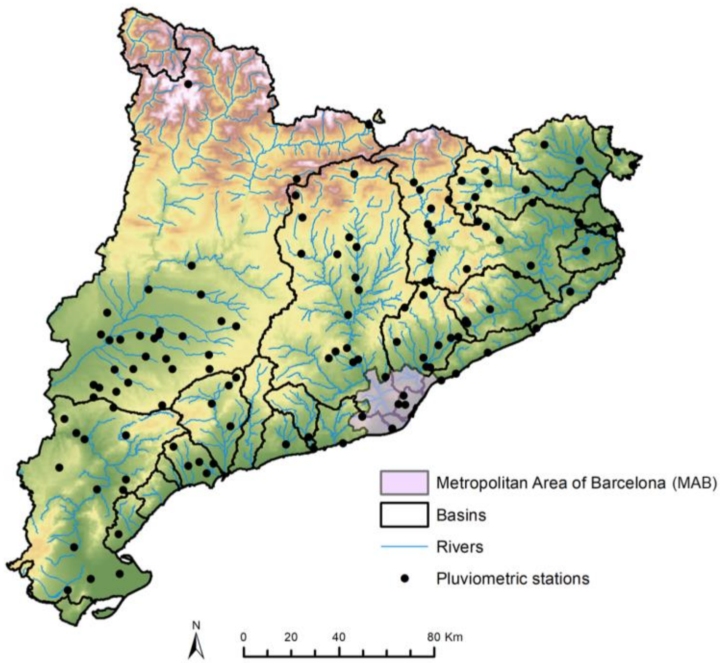
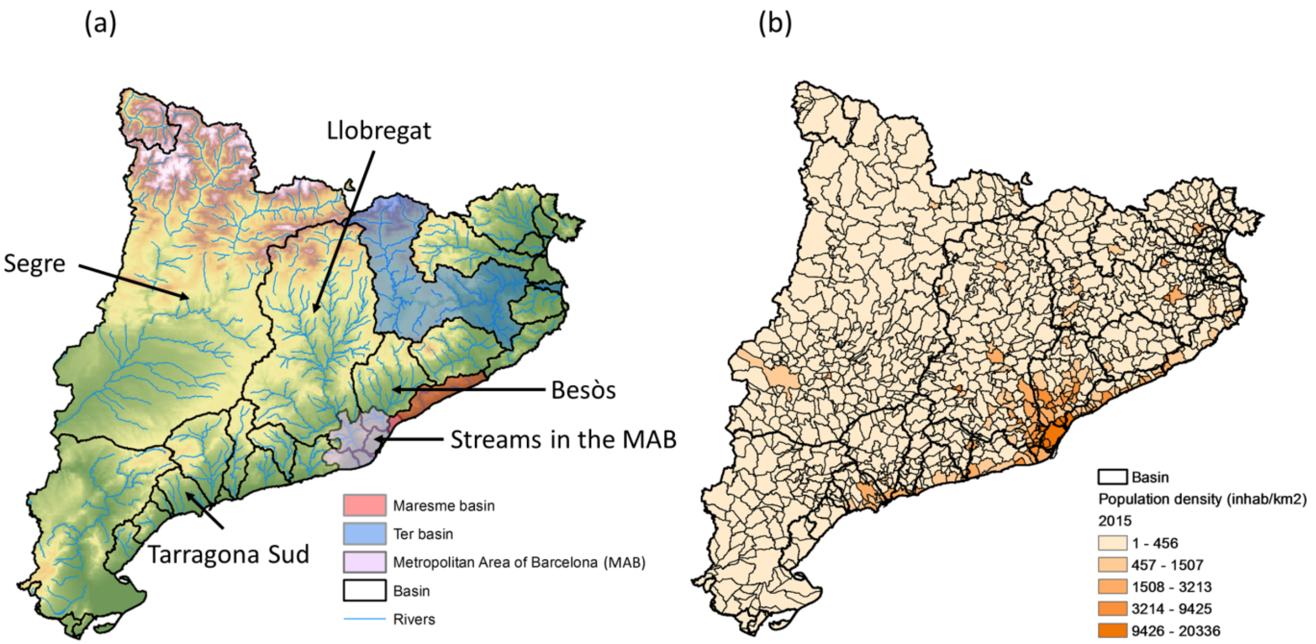


Figure 1: Map of Catalonia showing the aggregated basins, the Metropolitan Area of Barcelona (MAB), the main rivers and the pluviometric stations used.



~~**Figure 1: (a) Map of Catalonia showing the basins cited in the text and the Metropolitan Area of Barcelona (MAB); (b) distribution of the municipal population density in 2015.**~~

2.2 Data

The flood damage data were obtained ~~comes~~ from the compensation for floods paid by the Spanish Insurance Compensation Consortium (CCS). The CCS compensates for damages caused to people and property by floods and other adverse weather events covered by an insurance policy. The CCS database includes more than 58,000 records of claims paid for floods in Catalonia provided ~~at~~ on a postal code level for the 1996-2015 period (no previous information is available with this level of detail). ~~The data is available for the 1996-2015 period.~~ For flood events we use the INUNGAMA (Barnolas and Llasat, 2007;~~,~~ Llasat et al., 2016a) and PRESSGAMA (Llasat et al., 2009) databases, which report the flood ~~episodes~~ events that have occurred in Catalonia on a municipal, district and basin level ~~(Table 1)~~. Basic data on damaging events (i.e. event dates, duration and some hydrometeorological data) are identified using the INUNGAMA database. The PRESSGAMA database was used for the events and the description of their impacts, and to identify the worst-affected places. Population and Gross Domestic Product data were obtained from the Statistical Institute of Catalonia (*Institut d'Estadística de Catalunya*, IDESCAT). The population and GDP used correspond to the year when the flood event took place. We use daily precipitation data provided by the meteorological station network run by the Spanish State Meteorological Agency (*Agencia Estatal de*

Meteorología, or AEMET). To ensure temporal homogeneity, we have only considered the stations located in Catalonia with more than 90 % of valid data over the 1996-2015 period (Fig.1). For the MAB we also considered 30-minute weather data obtained from the network of automatic meteorological stations belonging to the Meteorological Service of Catalonia (*Servei Meteorològic de Catalunya*, or SMC). Table 1 summarises the data used.

Compensations paid by the CCS were adjusted to the value of the euro in 2015, following the methodology defined by the Spanish National Institute of Statistics (INE, 2007). This consists of using the exchange rate in the Consumer Price Index (CPI) between the two years to adjust the values shown in euros. To compare this data with other variables, we first aggregated them at a municipal level. This task was made more difficult by the fact that a municipality can include different postcodes and one postcode can correspond to two municipalities. These difficulties were solved by aggregating the municipal postcodes and looking at press information. Finally, to calculate the total damages per event, we took the payments made on the day the event occurred, and the following seven days. We used this seven-day window as this is the period of time that the CCS allows insurance claims to be made. When the time difference between two events is less than seven days, damages are associated with the first event, if the date of the claim was before the first day of the second event.

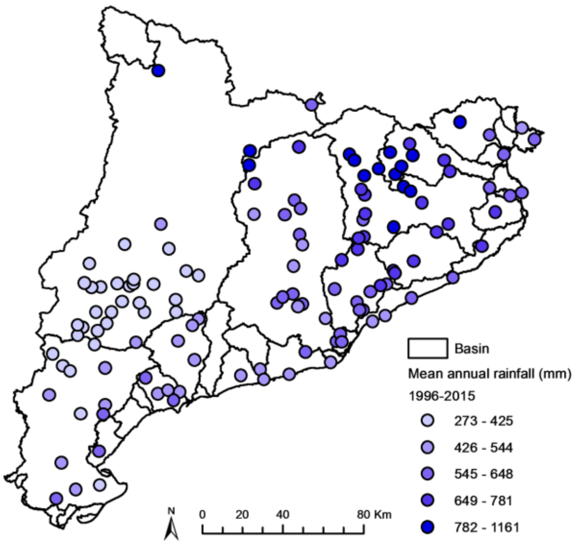
Because the available data are too sparse to support our statistical assessment on a municipal scale, we assessed the precipitation-compensation link for Catalonia as a whole. That is, we sampled pairs of the response variable (i.e. the compensation series) and the maximum 24 h precipitation for each basin, and pooled them into one sample for the entire region (Catalonia) to correlate them. For each event there can be more than one pair of values, depending on the number of affected catchments. From now on we will use the expression “flood case” for each pair of values corresponding to a basin affected by a flood event. This area is large enough to have a fairly large sample size for analysis, but small enough that the causes of flood damages are likely to be similar across the area.

The same methodology was applied for another spatial aggregation based on the Spanish State Meteorological Agency (AEMET) warning areas (included in the supplementary material), and which has also been used in other studies like Quintana-Seguí et al. (2016). Similarly as for the basins, an aggregation process was carried out (15 to 14 warning areas).

Finally, we considered three categories of damages: (i) total damages (D), (ii) damage per capita (DPC) and (iii) damage per unit of gross domestic product (DPW). This meant the relative impacts of socio-economic factors on damage could be estimated, while taking into account population and wealth (Zhou et al., 2017).

~~Population data was obtained from the Statistical Institute of Catalonia (*Institut d’Estadística de Catalunya*, IDESCAT).~~

~~We use daily precipitation data provided by the meteorological station network run by the Spanish State Meteorological Agency (*Agencia Estatal de Meteorología*, or AEMET). To ensure temporal homogeneity, we have only considered the stations located in Catalonia with more than 90 % of valid data over the 1996-2015 period (Fig. 2). For the MAB we also considered 30-minute weather data obtained from the network of automatic meteorological stations belonging to the Meteorological Service of Catalonia (*Servei Meteorològic de Catalunya*, or SMC).~~



~~**Figure 2: Pluviometric stations and average annual precipitation (1996-2015).**~~

~~Table 1 shows the data used for the different spatial scales studied: regional (Catalonia), basin (specifically Maresme and Ter basins) and local (MAB) areas.~~

~~2.3 Methodology~~

~~The CCS database includes more than 58,000 records of claims paid for floods in Catalonia provided at a postal code level for the 1996-2015 period (no previous information is available with this level of detail). To compare this data with the other variables we aggregated them at a municipal level. This task was made more difficult by the fact that a municipality can include different postcodes and one postcode can correspond to two municipalities. These difficulties were solved by aggregating the municipal postcodes and looking at press information. The compensations were adjusted to the value of the euro in 2015, following the methodology defined by the Spanish National Institute of Statistics (INE, 2007). This consists of using the exchange rate in the Consumer Price Index (CPI) between two periods to adjust the values shown in euros. Finally, to calculate the total damages per episode, we took the payments made during the period for the episode and the following 7 days. We used this 7-day window since this is the period of time that the CCS allows insurance claims to be made. When the time lag between two episodes is less than 7 days, damages are associated with the first event, if the date of the claim was before the first day of the second episode. The event duration is identified in the INUNGAMA database. The PRESSGAMA database was used for a description of the event and the affected places.~~

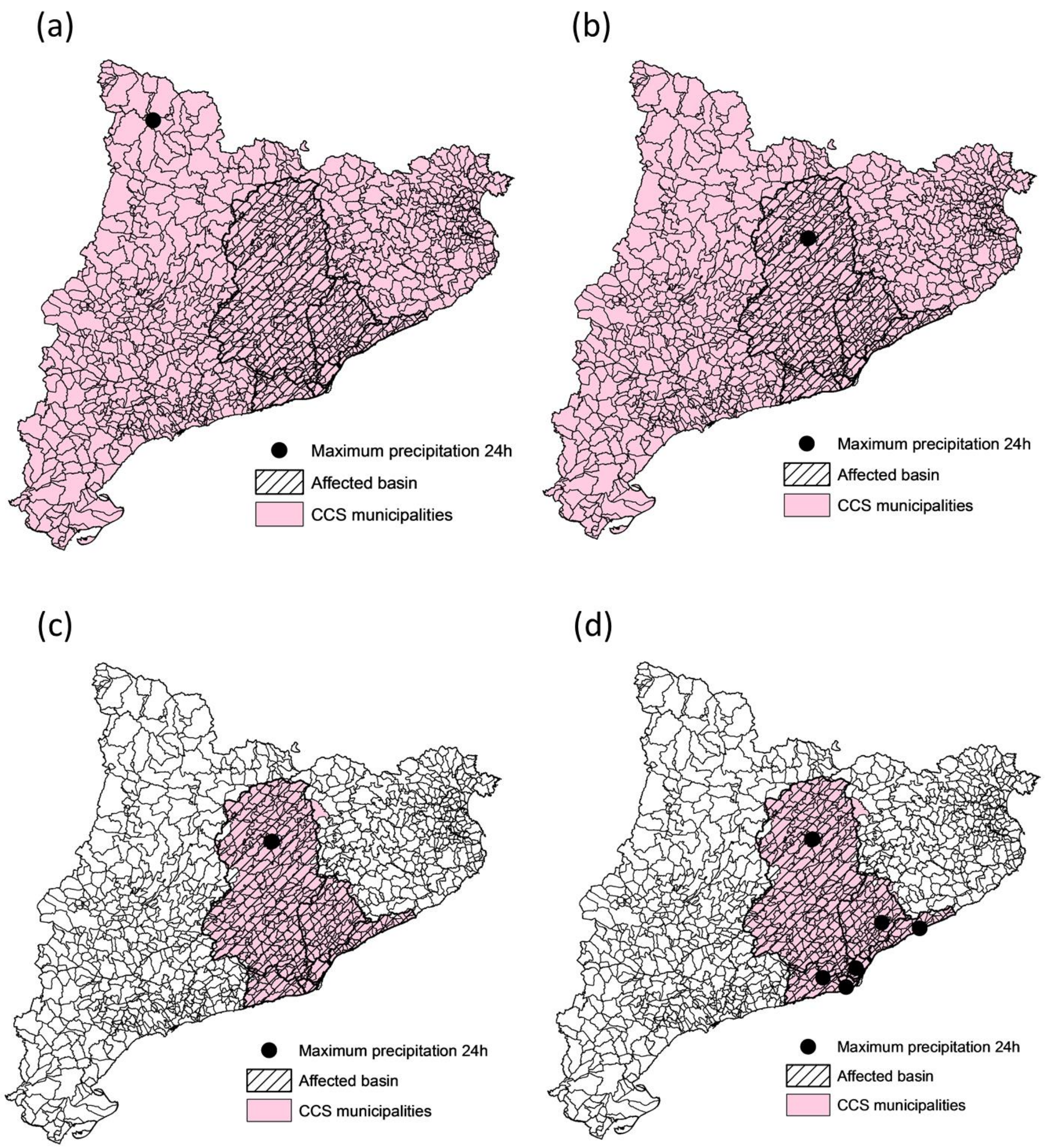
~~In order to analyse the potential links between precipitation and flood damages throughout Catalonia as whole, we explored four different configurations. We decided to do this as floods could happen in areas that are far away from the regions where the highest precipitation occurred. More specifically, the scheme used to correlate precipitation and damages is illustrated in Fig. 3 and is described below. We calculated:~~

~~The correlation between the maximum 24 h precipitation recorded in Catalonia during the event and the total amount paid by the CCS for the damages that occurred in the region;~~

~~The correlation between the maximum 24 h precipitation recorded in the basins where river or pluvial floods occurred and the total amount paid by the CCS for damages that occurred in the entire region;~~

~~The correlation between the maximum 24 h precipitation recorded in the basins where river or pluvial floods occurred and the total amount paid by the CCS for damages that occurred in the affected basins;~~

~~The correlation between the maximum 24 h precipitation recorded and the total amount paid by the CCS in each basin affected by river or pluvial floods.~~



~~**Figure 3: A schematic view of the analysis carried out to find the scale with the highest correlation between precipitation and flood damage. The pink shaded area shows the municipalities considered when aggregating damage. The dashed area shows the catchment affected by floods. The black circles show the areas with maximum precipitation in 24 h in: (a) the region as a whole; (b), (c) all the affected catchments; (d) each affected catchment.**~~

~~When looking at the basin scale, we analysed the correlation between precipitation and compensation paid for 7 basins. Those that have recorded a number of flood events above the 75^th^ percentile (26 flood events). These basins are (from more to less~~

~~flood events recorded): Maresme, Llobregat, Besòs, Ter, streams in the MAB, Tarragona Sud and Segre. Moreover, for the Maresme and Ter basins, we carried out a more in-depth analysis (Sect. 3.2.2), looking for correlations between compensations, flood episodes and the population at a municipal level. The Maresme basin is a torrential basin affected for by flash flood events and with a high permanent population density that increases in the summer (Llasat et al., 2010). The Ter basin is important both in terms of water resources and for agriculture and tourism, including both rural and urban areas, as well as hydraulic systems for flood prevention (dams and retention dams). For informative purposes, the city of Girona, the biggest city in the basin, was affected by 22 catastrophic flood episodes between 1301 and 2012 (Barrera-Escoda and Llasat, 2015). In order to better understand the relationship between flood events, compensation and population in a heterogeneous basin like the Ter basin, it has been divided in different sub-basins based on the population density of each municipality.~~

~~Finally, in the MAB the relationship with compensation was analysed for three periods of rainfall accumulation: 30 minutes, 24 hours and for the whole event.~~

~~All correlation calculations were carried out using the Spearman test.~~

2.3 Modelling damage probabilities

After gathering together a list of all the floods that affected Catalonia between 1996 and 2015, we filtered them based on specific rainfall thresholds. The Social Impact Research Group, created within the framework of the MEDEX project (MEDiterranean EXperiment on cyclones that produce high-impact weather in the Mediterranean; http: //medex.aemet.uib.es) has established a threshold – when a maximum rainfall of over 60 mm in 24 h was recorded – to indicate the expected social impact for rain events in Catalonia (Amaro et al., 2010; Jansà et al., 2014). Barbería et al. (2014) suggest that the threshold of 40 mm/24 h is better for urban areas. In the main text we consider the threshold on 60 mm/24 h, while results obtained using the lower threshold are available in the supplementary material.

In the case of the MAB, the minimal unit of study is the entire MAB region, which means each flood event corresponds to a single flood case. Taking into account that applying the precipitation thresholds of 40 and 60 mm for the MAB will result in samples that are too small (36 and 23 flood cases, respectively), and that the analysis would not be robust enough, we have used lower precipitation thresholds. It is worth noting that in this case we also used 30 min precipitation, which means a lower threshold might still have significant consequences. For instance, a previous study shows that with precipitation over 20 mm/30 min, extraordinary and catastrophic flood events can occur (Cortès et al., 2017) in the region. In addition, other studies (Barrera-Escoda and Llasat, 2015) have used 20 mm/24 h to study flood events in this Mediterranean region. Since the sample size is still small, a 10 mm threshold was also used (but results for the 20 mm threshold are available as supplementary material).

Figure 2 shows the relationship between the three categories of damages considered (D, DPC and DPW) and precipitation (log-transformed) in Catalonia. Even if a linear regression indicates a significant link (p-value<0.01), the explanatory power of the model for D is rather low (r^2=0.09). Marginally better results are obtained for the damage indicators DPC and DPW (r^2=0.14 and r^2=0.16 respectively), underlying the importance of considering the impacts of population and wealth on damage.

That is, this analysis corroborates the common experience that, given the same level of heavy precipitation, the total damage is larger where the level of wealth is higher.

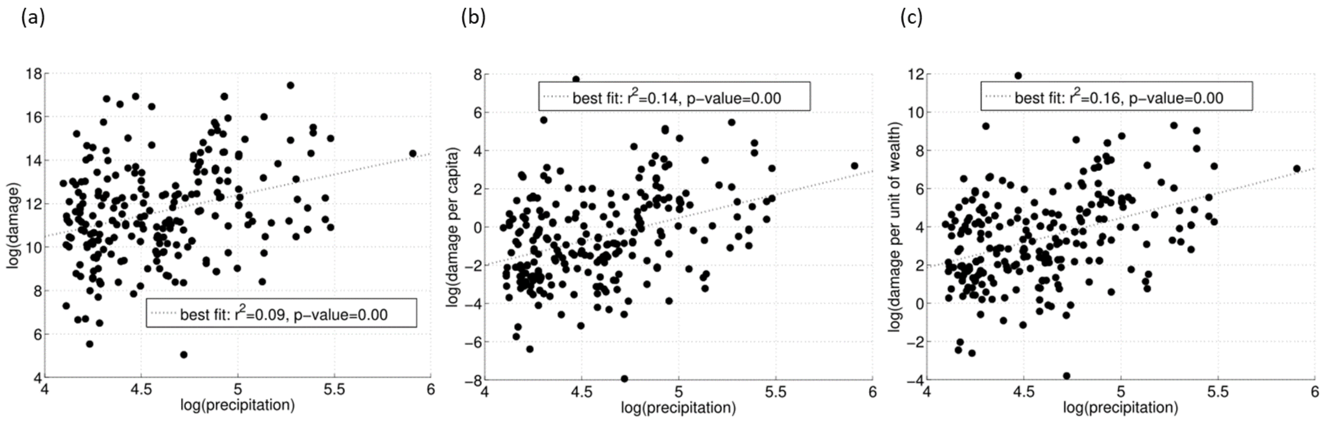


Figure 2:Scatter plot between basin-aggregated maximum precipitation in 24 h and (a) total damages (D); (b) damage per capita (DPC); and (c) damage per unit of wealth (DPW), for flood events recorded in Catalonia between 1996 and 2015 (log-transformed values). The dashed line indicates the fit based on a linear regression model.

The large spread of Figure 2 indicates that modelling insurance compensations is a complex issue due to the limitations in observational data and the concurrence of a variety of relevant factors. For instance, monetary data could be affected by limitations, as the value of the assets exposed and insurance coverage may change over time (Barredo et al., 2012). Unfortunately, exact data on the value and location of assets exposed are not available.

However, the significant correlation between compensations and precipitation suggests that rainfall data can be used to extract information on damages in Catalonia. To do this, we applied a logistic regression model to gauge the probability of large damaging events given a certain precipitation amount (an approach that is frequently used for this kind of modelling study: Kim et al., 2012; Wobus et al., 2014). That is, our aim is not to estimate the precise amount of the monetary compensation, but to estimate when a "large" damaging event will occur given a certain precipitation amount. Since there is not a standard definition of a large damaging event, we tested several cases: insurance compensations exceeding the 50th, the 60th, the 70th, the 80th and the 90th percentile of the total sample. This methodology is repeated for both thresholds (40 mm and 60 mm) and for the three damage indicators (D, DPC, DPW) for the basins and warning areas. It means we made a total of 60 models.

Finally, the logistic model is calculated following the Eq. (1):

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \cdot P, \qquad (1)$$

where, π, are the response variable (i.e. the probability above a certain percentile) and P is the predictor (precipitation in our case). The value of the β coefficient is determined using Generalized Linear Models (GLM). The Wald chi-square statistic is used to assess the statistical significance of individual regression coefficients (Harrel, 2015).

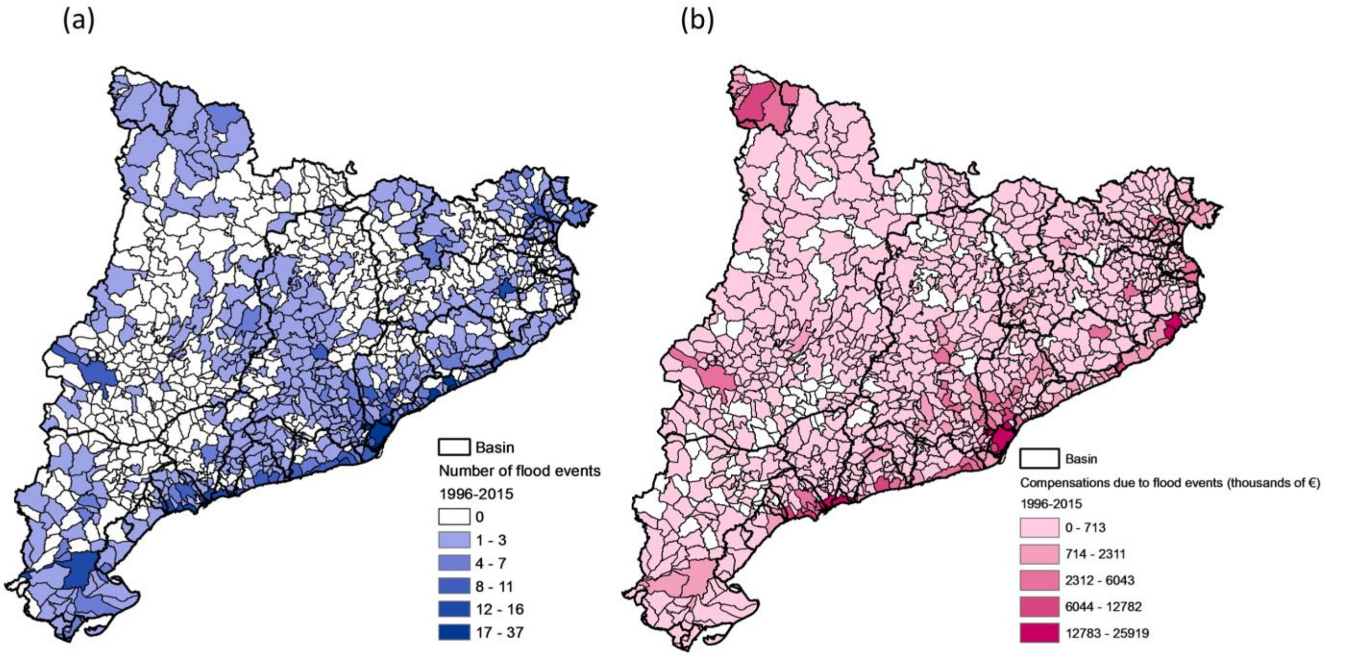
3 Results

3.1 Damaging events and precipitation in Catalonia

~~3.1 Analysis for the entire region~~

~~3.1.1 Flood episodes~~

~~The total number of flood episodes recorded in Catalonia for the 1996-2015 period was 166. Around 49 % of the episodes occurred during the months of July, August and September, with the latter month having the highest percentage of episodes (22 %). The most severe or catastrophic episodes occurred in autumn, with 77 % of the events between September and November (Llasat et al, 2016). The compensation paid by the CCS for floods for this period in Catalonia was €436.4 million.~~



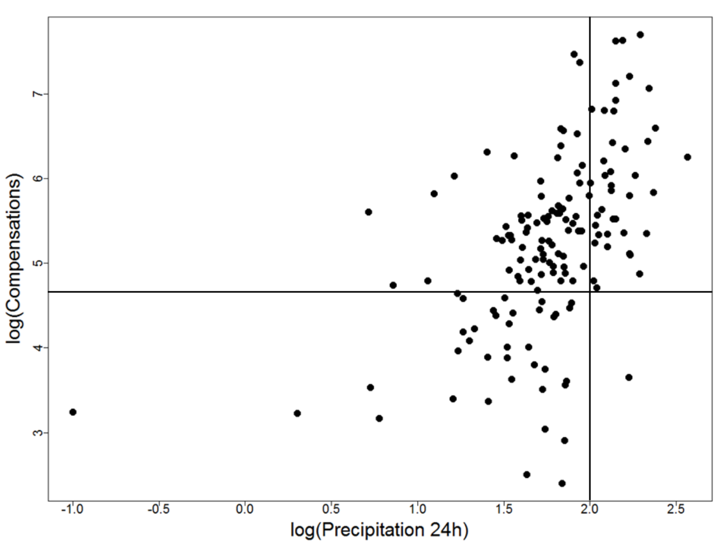
~~Figure 4: (a) Municipal distribution of flood events; (b) municipal distribution of total compensation for floods paid by CCS. Period: 1996-2015.~~

~~Figure 4 shows the number of flood episodes and the total number compensation paid for floods by the CCS on a municipal level. Coastal municipalities were the most affected by flood events (Fig. 4a) and where there was the most damage (Fig. 4b). This is the region where most of the population of Catalonia and the most tourism are concentrated, which makes it a very exposed and vulnerable area.~~

~~3.1.2 The relationship between precipitation and flood damages~~

~~Table 4 shows the result of the correlations between the accumulated precipitation in 24 h and the compensation paid by the CCS, applying the 4 methodologies described in section 2. In every case, the results show a positive and significant correlation. This result is consistent with the fact that most floods are caused by high intensity rainfall in basins where the time concentration is very small. This means 24 h precipitation could be considered a good indicator for flood risk. The best results are obtained using criteria number 3, only considering the basins affected by floods (Fig. 5). The fact that the correlation decreases considering the basin individually (criteria 4), suggests that there are not necessarily more damages where more precipitation occurs, since differences in vulnerability and exposure between one basin and the next may play an important role.~~

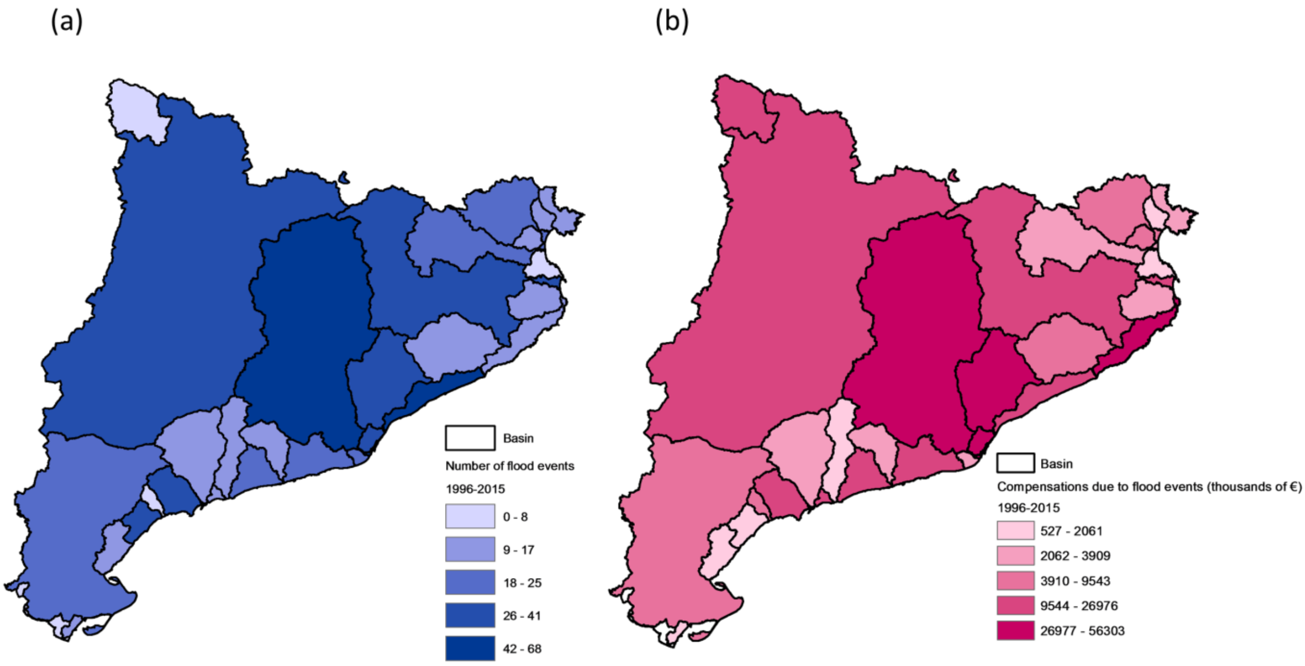
~~Figure 5 shows the correlation between 24 h precipitation and compensation using criteria 3. This graph makes it possible to discern a precipitation threshold from which significant damage is observed. For instance, the lines show that for the events where the 24 h precipitation was above 100 mm, damages exceeded €30,000, except in one case where the compensation did not surpass € 5,000, because it mainly affected a rural region with a low population.~~



~~Figure 5: Scatter plot (in log scale) between precipitation in 24 h and compensations paid by CCS for flood episodes recorded in Catalonia between 1996 and 2015, using the criteria number 3.~~

~~3.2 Analysis at a basin scale~~

~~3.2.1 Flood episodes~~



~~Figure 6: (a) Number of flood events per basin; (b) total compensation paid by the CCS per basin. Period: 1996-2015.~~

~~Figure 6 shows the total number of flood events recorded (Fig. 6a) and the total compensation paid by CCS for flooding (Fig. 6b) in each basin. In general, there is a good correlation between the flood events recorded and the compensation paid, as expected. The Maresme basin was affected by 41 % of the recorded episodes with damages that add up to €26,976,181.34 between 1996 and 2015 (Fig. 6b), especially for damages to individuals (Table 3).~~

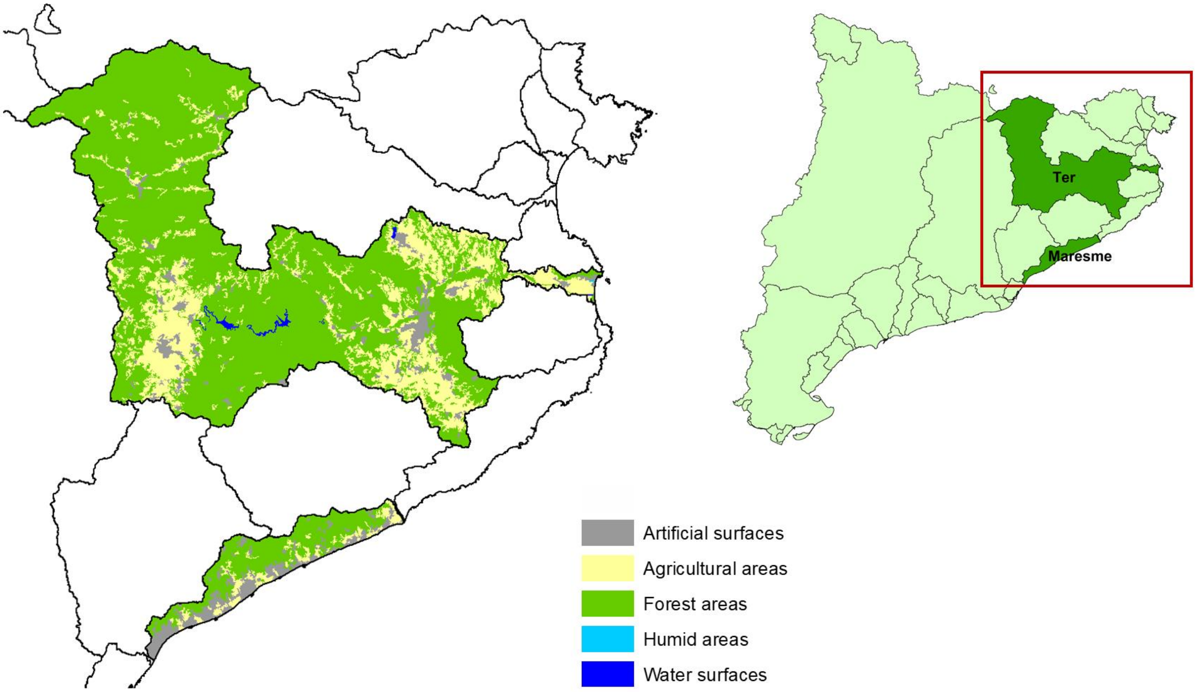
~~3.2.2 The relationship between precipitation and flood damages~~

~~Table 3 shows that most of the correlations are positive and statistically significant for the selected basins. The best results were obtained for the Maresme basin, with correlation of 0.7. This basin is made up of 30 municipalities and is characterised by a succession of villages crossed by torrential streams with their sources in the Littoral Range (Fig. 7), and where flash floods occur every year due to local convective precipitation events (Llasat et al., 2016a). These factors provide a possible explanation of the high correlation between precipitation and the damage caused.~~

~~On the other hand, the worst results are found in the case of the Ter river basin. Although this basin is affected by both flash flood and flood events due to continuous rains, the latter are less frequent (Barriendos et al., 2003). In addition, the use of reservoirs to control overflows means the floods are mainly caused by intense local rains in the sub-basins. Contrary to the Maresme basin, the Ter is a large basin where there are many differences in land use: the lower part of the basin is much more urbanised than upstream (Fig. 7). This non-homogeneous behaviour can also be observed on the map of compensation paid by CCS (Fig. 4b).~~

~~Next, we analyse the basins that present the best and the worst correlations in more details, for the results for 24 h precipitation and compensation in the previous section (Table 3), for the Maresme and Ter basins respectively. In order to analyse these differences in more depth, a correlation analysis was carried out for each basin at a municipal level for the flood events recorded for the 1996-2015 period, the total compensations received for the same period and the population of the municipality in the year that the episode occurred. Table 4 shows these results for the two basins studied. As might be expected, in both cases the compensation is better explained by the population, since the compensations paid are higher where there are more people exposed to floods. Furthermore, in the case of the Maresme basin, the correlation result between flood events and compensation paid are quite high and significant (0.53) while in the Ter basin this value, despite being significant, is low.~~

~~For this reason, and taking into account the greater differences between municipalities in the latter basin, analysis was carried out while separating the Ter river basin into two groups, according to the population density of the municipalities, using a threshold of 85 % (263.7 inhabitants/km^2): we defined "urban" areas as those that surpasses this threshold, and "rural" areas as all the others. This population density threshold was chosen for presenting a better correlation result between the number of flood events and compensation. It is worth noting that similar results were obtained considering different percentiles, as reported below.~~



~~**Figure 7: Land use (Corine Land Cover 2012) for the Ter and Maresme basins. The 5 land use categories of the Corine Land Cover Map are shown with the corresponding colour code.**~~

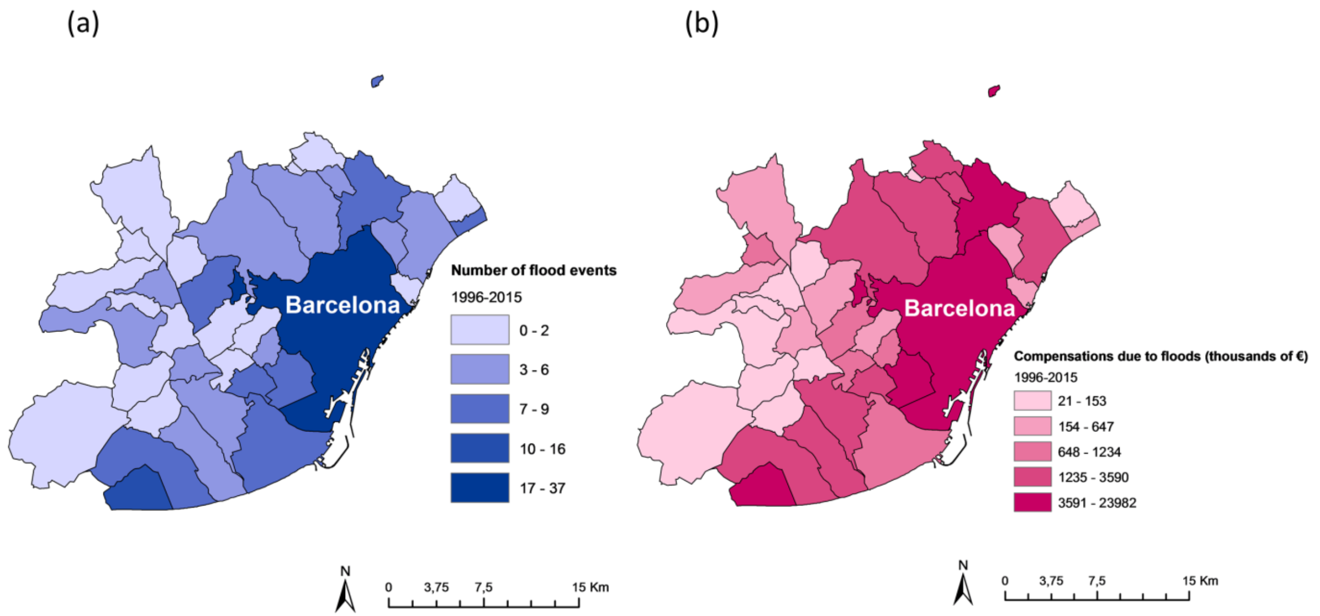
~~Table 5 shows that the correlations results are quite high considering the "urban" areas: the correlation value between flood events and damages is 0.60. If we consider a lower threshold to define "urban" areas, the 75^{th} percentile (163.1inhabitants/km^2), then the correlation is still quite high (0.53). In addition, the correlation values between compensations and population are high as expected (0.67 for the 75^{th} percentile and 0.81 for 85 %), as there are more claims where more people are exposed to risk. On the other hand, when considering the "rural" areas, the correlations are lower. These results can be explained by the poor availability of information in these regions, where it is probable that not all alluvial events were recorded and agricultural assets were not insured by the CCS~~

~~3.3 Metropolitan Area of Barcelona~~

~~3.3.1 Flood episodes~~

~~A total of 61 episodes of floods were recorded in the Metropolitan Area of Barcelona (Fig. 7a), which means an average of more than 3 episodes per year. The summer and autumn months were the ones with the highest number of flood episodes, with~~

~~September being the first (31 %), followed by October (16 %). The compensation paid by the CCS for floods amounted to € 86.3 million, which represents 20 % of the total compensation paid by the CCS in Catalonia (Fig. 7b).~~



~~**Figure 8: (a) Municipal distribution of the number of flood episodes in the MAB; (b) municipal distribution of the total compensation paid by the CCS in the MAB. Period 1996-2015.**~~

~~Figure 8 suggests a relationship between the MAB municipal distribution for the number of flood episodes and the distribution of the total compensation paid by the CCS. The municipality of Barcelona recorded a total of 37 episodes between 1996 and 2015, all due to in situ precipitation and drainage problems in the city (Llasat et al., 2016b). The city of Barcelona is also notably the most compensated for floods (around € 19 million).~~

~~**3.3.2 The relationship between precipitation and flood damages**~~

~~Table 6 shows the results of the analysis of correlation the between precipitation variables and the damage for all flood episodes that affected the MAB between 1996 and 2015. The best-correlated variable with compensation paid by CCS is the maximum precipitation in 30 minutes, with a value of 0.64. This result, together with the strong correlation between precipitation in 24 h and accumulated precipitation, corroborates the hypothesis that most frequent type of flood in the study region is flash floods, that is to say, episodes caused by intense precipitation of a short duration (one day or less).~~

The total number of flood events recorded in Catalonia for the 1996-2015 period was 166 (109 of them went beyond the 40 mm/24 h precipitation threshold and 81 went over the 60 mm/24 h threshold) resulting in a total number of flood cases (i.e. pair of precipitation-damage values at a basin scale) of 642 (331 for 40 mm/24 h and 239 for 60 mm/24 h). Coastal municipalities are the most affected by flood events and where there is the most damage. This is a consequence of high vulnerability (the most vulnerable structures and infrastructures are on the coast), exposure (population and tourism are concentrated in the coastal regions) and hazards (flash floods associated to local heavy rain events are frequent) (Llasat et al., 2014, 2016a). Around 49 % of the events occurred during the months of July, August and September, with the latter month having the highest percentage of events (22 %). The most severe or catastrophic events occurred in the autumn, with 77 % of the events taking place between September and November (Llasat et al., 2016a). The compensations paid by the CCS for floods during this period in Catalonia amounted to €436.4 million.

Figure 3 shows the number of flood events recorded between 1996 and 2015 (Fig. 3a), the total insurance losses paid by CCS for flooding (Fig. 3b) during this period, the average population (Fig. 3c) and the GDP (Fig. 3d) in each basin. In general, there is a good correlation between the four variables, as expected. The basins with more recorded flood events are those that received more insurance compensations for flood damages, with a higher population and gross domestic product. The Maresme basin was affected by 41 % of the recorded events (Fig. 3a) with damages that add up to €24,561,762.4 between 1996 and 2015 (Fig. 3b).

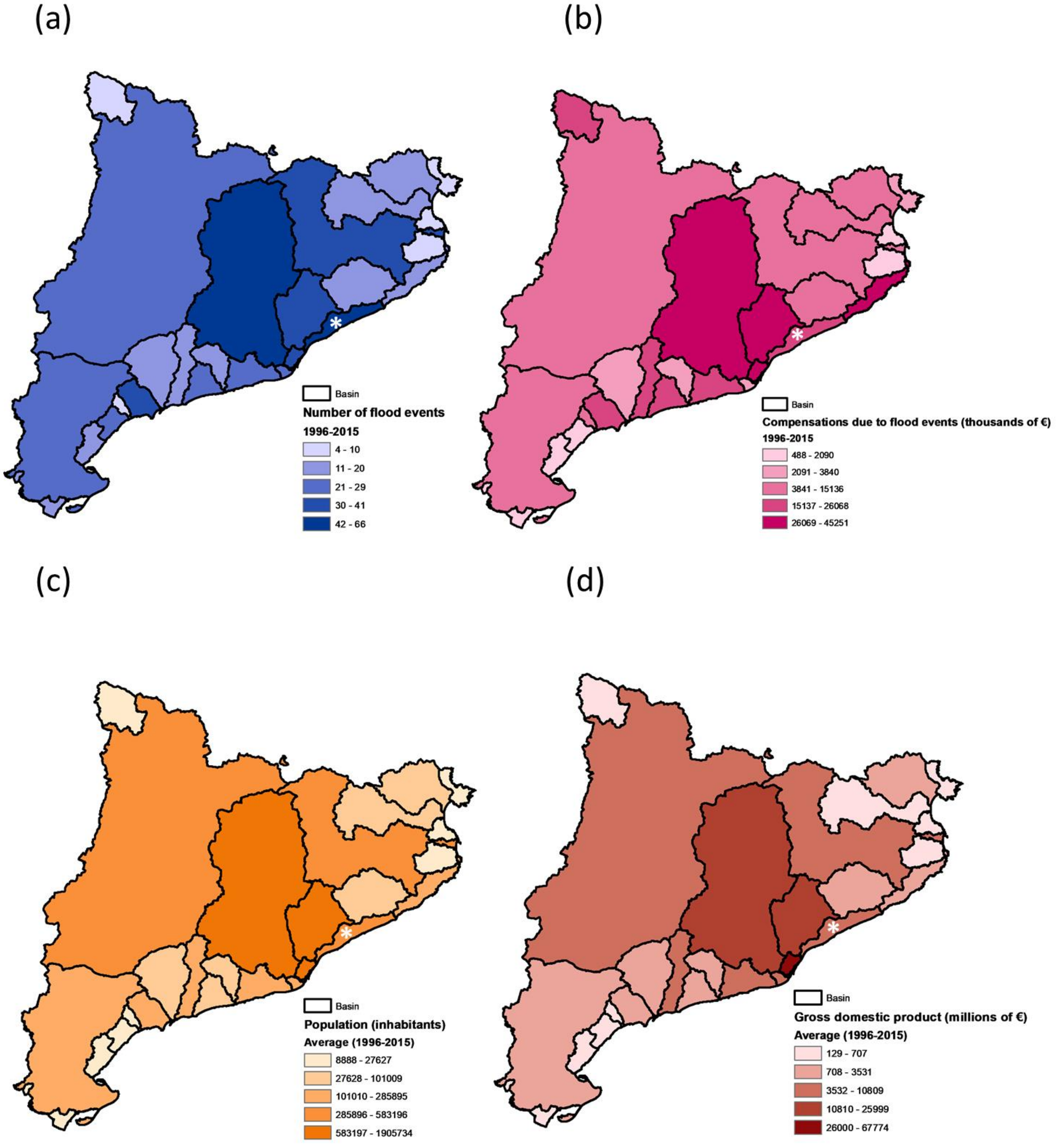


Figure 3: Basin distribution of (a) flood events (1996-2015); (b) total insurance compensations for floods made by CCS (1996-2015); (c) average total population; and (d) average gross domestic product. Asterisk indicates Maresme basin.

In order to estimate when a “large” damaging event will occur with a given precipitation amount, a logistic regression was used. Figure 4 shows a logistic regression example that indicates the model is able to simulate the probability of DPW above and below the 70th percentile as a function of precipitation. This figure illustrates that the probability of reaching above the 70th percentile for DPW increases when there is a large amount of rain. This result is consistent with the hypothesis that 24 h precipitation could be considered a good indicator for flood risk. For this example the regression equation [Eq. (2)] would be:

$$log\left(\frac{\pi}{1-\pi}\right) = -10.5 + 2.08 \cdot P, \quad (2)$$

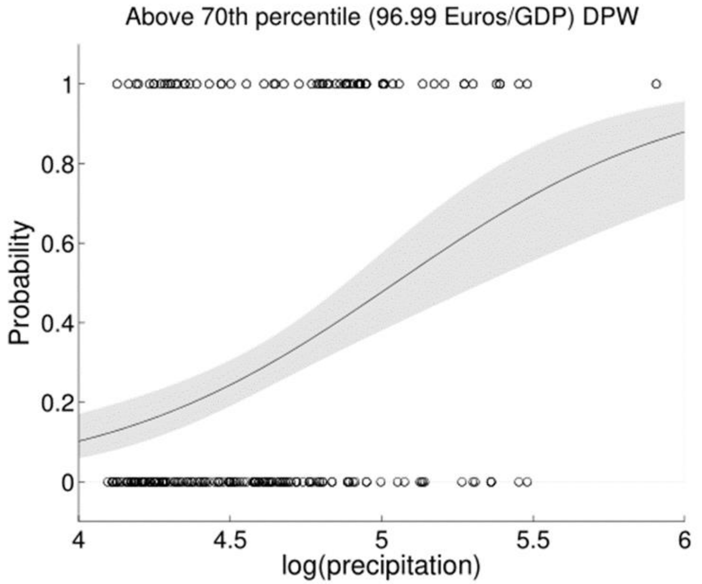


Figure 4: Example of logistic regression result to model DPW damages above the 70th percentile as a function of precipitation (log-transformed). The solid line indicates the best estimate while the shaded band indicates the 95 % confidence interval. Open circles along the horizontal axis show the events that are above (top) and below (bottom) the 70th percentile.

Table 2 shows the values of β0 and β1, considering cases with a threshold of 60 mm for the different combinations of damage indicators and percentiles (see supplementary material for 40 mm and the distribution in warning areas; the results are very similar).

It is important to assess whether this model can be used to separate positive and negative anomalies. Our models are not deterministic and users need to take into account the uncertainty of the forecast expressed by these probabilities. For example, users could decide to take action when a 10 % probability of an above-70th percentile event is forecast. In this case most of the observed events are forecasted, that is, the hit rate (i.e. the relative number of times a simulation event actually occurred) is close to 1, but this also implies a higher false alarm rate (i.e. the relative number of times an event had been simulated but did not actually happen). On the other hand, if a higher threshold is used, we can reduce the number of false alarms, but at the expense of a greater number of missed events. The choice of the decision threshold is a function both of the skill of the forecast and the cost/loss ratio of the user. In any case, in a forecasting system affected by uncertainties, missed events can be reduced only by increasing false alarms and vice versa. In order to validate the model, we considered the relative operating characteristic (ROC) diagram that shows the hit rate (H) against the false alarm rate (F) for different potential decision thresholds (Mason and Graham, 2002) (see Figure 5).

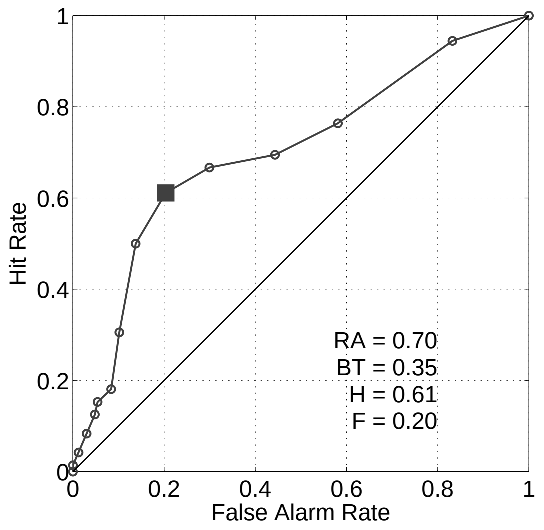


Figure 5: Relative operating characteristic (ROC) diagram for above 70th DPW predictions using the logistic regression of Eq. (1). The open dots indicate a set of probability forecasts by stepping a decision threshold with 5 % probability through the modelling results. The numbers inside the plots are the ROC Area (RA) and the Best Threshold (BT), here defined as the threshold that maximises the difference between the hit rate (H) and the false alarm rate (F).

The area under the ROC curve (RA) is a useful measure to summarise the skill of a model. RA ranges from 0, for a forecast with no hit and only false alarms, to 1, indicating a perfect forecast. Models with an RA above 0.5 have more skill than random forecasts. Figure 5 shows that our model has skill: the ROC curve is well above the identity line, with an RA of 0.7. The “best threshold” in this illustrative example is 0.35. This means that if we want to maximise the H-F difference (but please note that users could define other best thresholds according to their cost/loss ratio), an above 70th percentile damaging event is to be expected when our model predicts a probability higher then 0.35, resulting in H=0.61(this means that 61 out of 100 events are correctly modelled) and F=0.20 (this means that 20 out of 100 events were modelled as an “event” when it did not actually happen). For example, in this case (BT=0.35) a precipitation amount higher than 115 mm is needed to expect a damaging event above the 70th percentile for the damage indicator DPW (97 euros/GDP).

Table 2 summarises the model parameters and performance considering all the percentiles and the three categories of damage used. In each case, precipitation is a significant predictor (p-value<0.05) and the models have skill and significant RA values (the significance is estimated using a Mann-Whitney U-test; Mason and Graham, 2002). Similar results were obtained for the damage categories, with slightly larger RA considering DPW.

3.2 Damaging events and precipitation in the Metropolitan Area of Barcelona

A total of 61 flood events were recorded in the Metropolitan Area of Barcelona (Fig. 1), which means an average of more than 3 events per year. The summer and autumn months had the highest number of flood events, with September having the most (31 %), followed by October (16 %). The insurance compensations paid by the CCS for floods amounted to € 86.3 million, which represents 20 % of the total compensation paid by the CCS in Catalonia. The municipality of Barcelona recorded a total of 37 events between 1996 and 2015, all due to in situ precipitation and drainage problems in the city (Llasat et al., 2016b). The city of Barcelona also receives the most compensation for floods (around € 19 million).

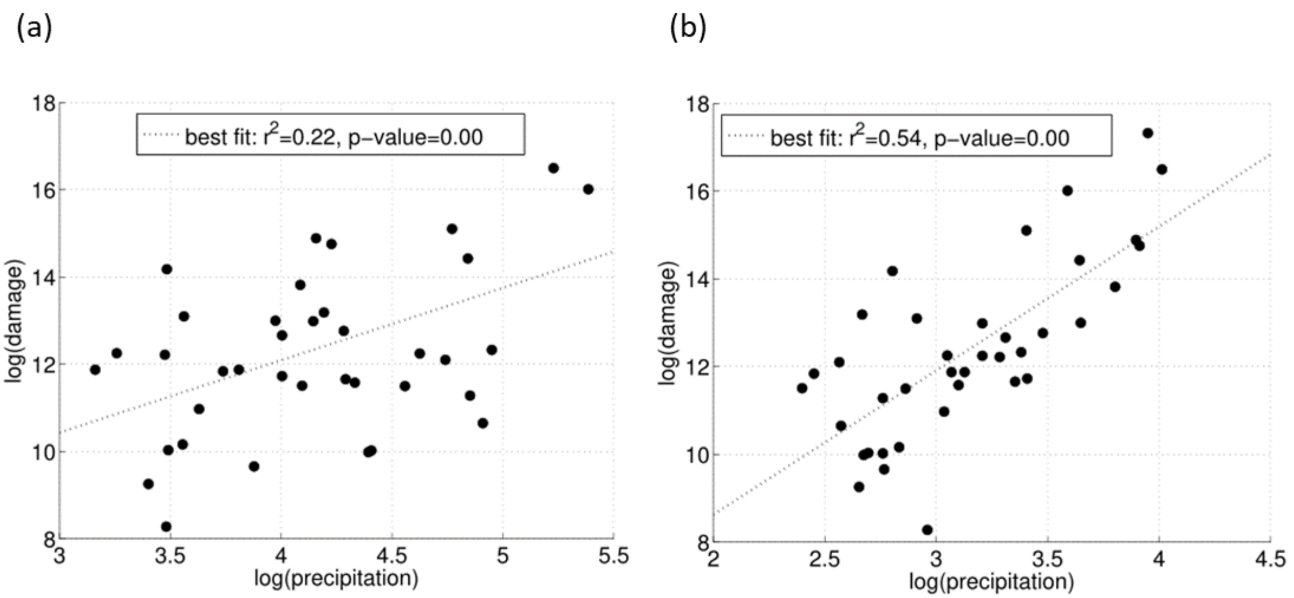


Figure 6: Scatter plot (a) damages (D) versus 24 h precipitation and (b) damages (D) versus 30 minute precipitation.

As it can be seen in Figure 6, the total damages (D) relate more to 30 minute precipitation than to 24 h precipitation, with significant results in both cases. In this particular case, similar results are obtained for the other damage categories (DPC and DPW, see Table 3).

We then repeated the logistic modeling exercise using 30 minute precipitation. Figure 7 shows a logistic regression for the events that affected the MAB. As in the basin level aggregation, the model is capable of simulating the probability of total damage (D) above and below the 70th percentile as a function of 30 minute precipitation in this case. As could be expected, this probability increases with precipitation. The same methodology was applied using a precipitation threshold of 20 mm/30 min (see supplementary material) and using the 50th, 60th, 80th and 90th percentiles (Table 3). For this example, the regression equation would be:

$$log\left(\frac{\pi}{1-\pi}\right) = -11.3 + 3.21 \cdot P, \qquad (3)$$

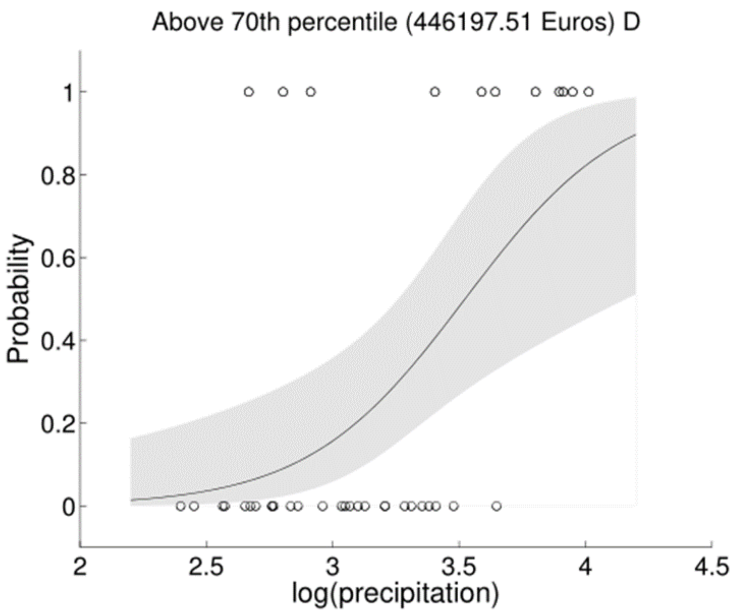


Figure 7: Example of a logistic regression result to model damages (D) above the 70th percentile as a function of 30 minute precipitation for the MAB. The solid line indicates the best estimate while the shaded band indicates the 95 % confidence interval. Open circles along the horizontal axis show the events that are above (top) and below (bottom) the percentile 70th.

Figure 8 shows the ROC diagram for predictions of total damages (D) above the 70th percentile for the MAB, using a precipitation threshold of 10 mm/30 min. The total RA (0.81) shows that our model for the MAB has skill. In this case, we would obtain the biggest difference between the hit and false rates when our model predicts a probability higher than 0.4. That is, the best threshold is 0.40, with 73 % of the events well-predicted (H=0.73) and only 11 % of false alarms events (F=0.11). In this example, a precipitation amount higher than 30 mm/30 min is needed to expect a damaging event above the 70th percentile for damage indicator D (0.45 million of euros).

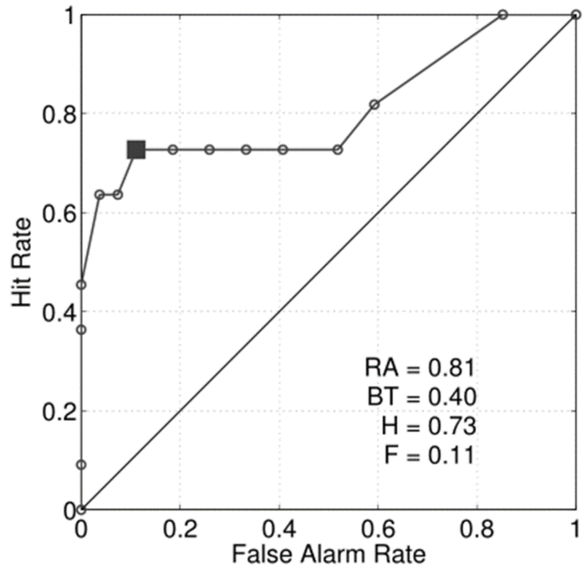


Figure 8: Relative operating characteristic (ROC) diagram for predictions for damage indicator D above the 70th percentile for the MAB using the logistic regression of Eq. (1). The open dots indicate a set of probability forecasts by stepping a decision threshold with 5 % probability through the modelling results. The numbers inside the plots are the ROC Area (RA) and the Best Threshold (BT), here defined as the threshold that maximises the difference between the hit rate (H) and the false alarm rate F).

Table 3 summarises the model parameters and performance considering all the percentiles and the three damage categories used for a precipitation threshold of 10 mm/30min (see supplementary material for results using 20 mm/30 min for the MAB). Similar results in terms of RA have been obtained for damage categories, whether using a 10 mm (Table 3) or a 20 mm threshold (supplementary material).

4 Conclusions

The Mediterranean is an area affected by flood events that produce significant socioeconomic damage. Catalonia, located to the west of the Mediterranean, is affected by an average of more than 8 ~~episode~~events per year. The majority of the damage caused by these ~~episode~~events is due to local events, with intense and short-lived rainfall ~~and not~~rather than river overflows (Llasat et al., 2014). Therefore, it is assumed that precipitation is the ~~maximum~~ main contributing factor for damages caused by this type of ~~episode~~event. To corroborate this hypothesis, the relationship between precipitation and compensation paid by insurance companies ~~at different spatial scales~~ was studied. To take into account the differences in vulnerability and exposure in the territory, we considered three types of damage: total damage, damage per capita (divided by the population) and damage per unit of GDP.

Although linear regression indicates a significant link (p-value<0.01), suggesting that rainfall data can be used to extract information on damages in Catalonia, the variance explained for the model is rather low (r^2=0.09 for D, r^2=0.14 for DPC and r^2=0.16 for DPW). For this reason, the relationship was assessed using logistic regression models in order to estimate the probability of large monetary damages occurring as a result of heavy precipitation events. That is, our aim is not to estimate the precise amount of insurance compensations, but to estimate when a "large" damaging event will occur given a particular precipitation amount. As could be expected, the logistic regression shows an increase in the probability of a damaging event occurring when precipitation increases. Our model is able to simulate the probability of a damaging event as a function of precipitation. In order to validate the model, we considered the Relative Operating Characteristic (ROC) diagram. The area under the ROC curve (RA) proved our model skill. The results show an RA above 0.6 in all percentiles of the three types of damages and thresholds of precipitation, most of them with values higher than 0.7.

The methodology was also been applied for the MAB region, an urban area affected by more than three flood events per year. Linear regression has shown that 30 minute precipitation is linked more closely with damages than 24 h precipitation. That is, we repeated the analysis for 30 minute precipitation and, as expected, the model presents better results in terms of RA for the urban area than for Catalonia as a whole, with values higher than 0.8 in all cases. Therefore, we have been able to confirm that 30 minute rainfall is a better predictor of the probability of large damages than daily rainfall in urban areas.

~~We observed that the best correlation results (up to 0.6) between the two variables are obtained when the affected basins are taken into account, rather than the entire study region. However, we also found that when we analysed large and heterogeneous basins (like the Ter basin) the correlation was only 0.37, and better results are achieved considering sub-areas defined in terms of population density, whether "urban" or "rural" areas. In the urbanised areas the correlation is 0.60, while lower and non-significant correlations were obtained for the rural areas. On the other hand, in small and homogeneous, very urbanized basins, such as the Maresme basin, precipitation explains most of the damage caused by flood events (with a correlation of 0.68). For the particular case of the Metropolitan Area of Barcelona, for which sub-daily rainfall data is available, a correlation above 0.60 was found for the maximum precipitation recorded in 30-min and damages, in spite of the lower correlation obtained for the precipitation accumulated in 24 h. These results suggest that prevention measures (such as rainwater retention tanks in the city of Barcelona) helped to mitigate the risk (for example, controlling water channels), but when precipitation is very intense and of short duration, these measures may not be sufficient.~~

These results confirm the hypothesis that precipitation is a key factor in explaining the damage caused by flood events in regions where ~~flash floods and urban~~water surface floods are the main type of floods, as is the case in this Mediterranean region. The strong relationships found in this study can be a useful tool for improving early warning systems and emergency management. For instance, from the relationship obtained between precipitation and compensations it is possible to predict when damaging events will occur as a result of a certain precipitation threshold. In other words, we have developed a new model that allows us to predict the probability that a flood event causing large damages (with its meaning depending on the user) will occur based on precipitation and taking into account the exposure and vulnerability of the territory in the model. ~~For instance, from the correlation results obtained between precipitation and compensations it is possible to predict the damages caused by a certain precipitation threshold on a different spatial scale.~~ ~~Also, t~~These links could also provide a basis to predict flood damage in future scenarios of climate change scenarios.

Competing interests

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the Spanish Project HOPE (CGL2014-52571-R) of the Ministry of Economy, Industry and Competitiveness, and the Metropolitan Area of Barcelona Project (no. 308321) (Flood evolution in the Metropolitan Area of Barcelona from a holistic perspective: past, present and future). It was ~~developed~~ conducted under~~in~~ the framework of the HyMeX Programme (HYdrological cycle in the Mediterranean EXperiment) and the Panta Rhei WG Changes in Flood Risk. We would like to thank AEMET and SMC for the meteorological and hydrological information provided for this study. Thanks also to BCASA for the detailed information about the system used to prevent and manage floods. Marco Turco was supported by the Spanish Juan de la Cierva Programme (IJCI-2015-26953). We would also like to acknowledge Hannah Bestow for ~~the correction of~~correcting the English language of this paper.

References

Amaro, J., Gayà, M., Aran, M. and Llasat, M. C.: Preliminary results of the Social Impact Research Group of MEDEX: The request database (2000-2002) of two Meteorological Services, Nat. Hazards Earth Syst. Sci., 10(12), 2643–2652, doi:10.5194/nhess-10-2643-2010, 2010.

Barbería, L., Amaro, J., Aran, M. and Llasat, M. C.: The role of different factors related to social impact of heavy rain events: considerations about the intensity thresholds in densely populated areas, Nat. Hazards Earth Syst. Sci., 14(7), 1843–1852, doi:10.5194/nhess-14-1843-2014, 2014.

Barnolas, M. and Llasat, M. C.: System Sciences A flood geodatabase and its climatological applications : the case of Catalonia for the last century, , (2005), 271–281, 2007.

Barredo, J. I.: Normalised flood losses in Europe: 1970–2006, Nat. Hazards Earth Syst. Sci., 9, 97–104, doi:10.5194/nhess-9-97-2009, 2009.

Barredo, J. I., Saurí, D. and Llasat, M. C.: Assessing trends in insured losses from floods in Spain 1971–2008, Nat. Hazards Earth Syst. Sci., 12(5), 1723–1729, doi:10.5194/nhess-12-1723-2012, 2012.

Barrera-Escoda, a. and Llasat, M. C.: Evolving flood patterns in a Mediterranean region (1301-2012) and climatic factors – the case of Catalonia, Hydrol. Earth Syst. Sci., 19, 465–483, doi:10.5194/hess-19-465-2015, 2015.

Barrera-Escoda, A., Llasat, M. C. and Barriendos, M.: Estimation of extreme flash flood evolution in Barcelona County from 1351 to 2005, Nat. Hazards Earth Syst. Sci., 6(4), 505–518, doi:10.5194/nhess-6-505-2006, 2006.

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.

Bihan, G. Le, Payrastre, O., Gaume, E., Moncoulon, D. and Pons, F.: The challenge of forecasting impacts of flash floods : test of a simplified hydraulic approach and validation based on insurance claim data, (June), 1–27, 2017. ~~(under review).~~

~~CCS: Estadística – Riesgos extraordinarios – Serie 1971-2015, Consorcio de Compensación de Seguros, Ministerio de Economía y Hacienda, Madrid, 143 pp, 2016.~~

Choi, O. and Fisher, A.: The impacts of socioeconomic development and climate change on severe weather catastrophe losses—Mid-Atlantic region (MAR) and the US, Clim. Change, 58, 149–170, 2003.

Cortès, M., Llasat, M. C., Gilabert, J., Llasat-Botija, M., Turco, M., Marcos, R., Martín Vide, J. P. and Falcón, L.: Towards a better understanding of the evolution of the flood risk in Mediterranean urban areas: the case of Barcelona, Nat. Hazards, 1–22, doi:10.1007/s11069-017-3014-0, 2017.

Elmer, F., Thieken, a. H., Pech, I. and Kreibich, H.: Influence of flood frequency on residential building losses, Nat. Hazards Earth Syst. Sci., 10(10), 2145–2159, doi:10.5194/nhess-10-2145-2010, 2010.

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.

Freni, G., La Loggia, G. and Notaro, V.: Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation, Water Sci. Technol., 61(12), 2979–2993, doi:10.2166/wst.2010.177, 2010.

García, L.E., J.H. Matthews, D.J. Rodriguez, M. Wijnen, K.N. DiFrancesco, P. Ray.: Beyond Downscaling: A Bottom-Up Approach to Climate Adaptation for Water Resources Management. AGWA Report 01. Washington, DC: World Bank Group, 2014.

Harrell Jr, F. E.: Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis, Springer International Publishing, 2015.

Institut d'Estadística de Catalunya (IDESCAT, Statistical Institute of Catalonia) (2016). Anuari estadístic de Catalunya (Statistical Yearbook of Catalonia). http://www.idescat.cat/. Accessed 01 June 2017.

Instituto Nacional de Estadística - INE: Índice de Precios de Consumo. Base 2006 - Metodología, Metodología, 1–74, 2007.

Jansà, A., Alpert, P., Arbogast, P., Buzzi, A., Ivancan-Picek, B., Kotroni, V., Llasat, M. C., Ramis, C., Richard, E., Romero, R. and Speranza, A.: MEDEX: A general overview, Nat. Hazards Earth Syst. Sci., 14(8), 1965–1984, doi:10.5194/nhess-14-1965-2014, 2014.

Jongman, B., Hochrainer-stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., Bouwer, L. M., Pflug, G., Rojas, R. and Ward, P. J.: Increasing stress on disaster-risk finance due to large floods, Nat. Clim. Chang., 4(4), 1–5, doi:10.1038/NCLIMATE2124, 2014.

Kim, Y.-O., Seo, S. B. and Jang, O.-J.: Flood risk assessment using regional regression analysis, Nat. Hazards, 63(2), 1203–1217, doi:10.1007/s11069-012-0221-6, 2012.

Kreibich, H. et al.: Adaptation to flood risk: Results of international paired flood event studies, Earth's Future, 5, doi:10.1002/2017EF000606, 2017.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L. M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G. R., Kron, W., Benito, G., Honda, Y., Takahashi, K. and Sherstyukov, B.: Flood risk and climate change: global and regional perspectives, Hydrol. Sci. J., 59(1), 1–28, doi:10.1080/02626667.2013.857411, 2014.

Llasat, M. C., Llasat-Botija, M. and López, L.: A press database on natural risks and its application in the study of floods in Northeastern Spain, Nat. Hazards Earth Syst. Sci., 9, 2049–2061, doi:10.5194/nhess-9-2049-2009, 2009.

~~Llasat, M. C., Llasat-Botija, M., Prat, M. a., Porcú, F., Price, C., Mugnai, a., Lagouvardos, K., Kotroni, V., Katsanos, D., Michaelides, S., Yair, Y., Savvidou, K. and Nicolaides, K.: High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database, Adv. Geosci., 23, 47–55, doi:10.5194/adgeo-23-47-2010, 2010.~~

Llasat, M. C., Marcos, R., Llasat-Botija, M., Gilabert, J., Turco, M. and Quintana-Seguí, P.: Flash flood evolution in North-Western Mediterranean, Atmos. Res., 149, 230–243, doi:10.1016/j.atmosres.2014.05.024, 2014.

Llasat, M. C., Marcos, R., Turco, M., Gilabert, J. and Llasat-Botija, M.: Trends in flash flood events versus convective precipitation in the Mediterranean region: The case of Catalonia, J. Hydrol., 541 (September 2002), 24–37, doi:10.1016/j.jhydrol.2016.05.040, 2016a.

Llasat, M.C.; Cortès, M.; Falcón, Ll.; Gilabert, J.; Llasat-Botja, M.; Marcos, R.; Martín-Vide, J.P.; Turco, M.: A multifactorial analysis of flood variability in the Metropolitan Area of Barcelona, ICUR2016 Proceedings, ISBN:978-989-95094-1-2, 2016b.

Mason, S. J. and Graham, N. E.: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation, Q. J. R. Meteorol. Soc., 128(584), 2145–2166, doi:10.1256/003590002320603584, 2002.

Messner, F. and Meyer, V.: Flood damage, vulnerability and risk perception - challenges for flood damage research, in Flood Risk Management Hazards Vulnerability and Mitigation Measures, vol. UFZ Discus, pp. 149–167., 2006.

Moncoulon, D., Labat, D., Ardon, J., Leblois, E., Onfroy, T., Poulard, C., Aji, S., Rémy, A. and Quantin, A.: Analysis of the French insurance market exposure to floods: A stochastic model combining river overflow and surface runoff, Nat. Hazards Earth Syst. Sci., 14(9), 2469–2485, doi:10.5194/nhess-14-2469-2014, 2014.

Nakamura, I. and Llasat, M. C.: Policy and systems of flood risk management: a comparative study between Japan and Spain, Nat. Hazards, 87(2), 919–943, doi:10.1007/s11069-017-2802-x, 2017.

Papagiannaki, K., Lagouvardos, K., Kotroni, V. and Bezes, a.: Flash flood occurrence and relation to the rainfall hazard in a highly urbanized area, Nat. Hazards Earth Syst. Sci. Discuss., 3(5), 3119–3149, doi:10.5194/nhessd-3-3119-2015, 2015.

Petrucci, O. and Llasat, M. C.: Impact of disasters in mediterranean regions: An overview in the framework of the HYMEX project, in Landslide Science and Practice: Social and Economic Impact and Policies, vol. 7, pp. 137–143., 2013.

Pielke R.A., J. and Downton, M. W.: Precipitation and damaging floods: Trends in the United States, 1932-97, J. Clim., 13(20), 3625–3637, doi:10.1175/1520-0442(2000)013<3625:PADFTI>2.0.CO;2, 2000.

Quintana-Seguí, P., Peral, C., Turco, M., Llasat, M. C. and Martin, E.: Meteorological analysis systems in North-East Spain: Validation of SAFRAN and SPAN, J. Environ. Informatics, 27(2), 116–130, doi:10.3808/jei.201600335, 2016.

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. B., Freer, J. E. and Cloke, H. L.: The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model, Hydrol. Earth Syst. Sci., 18(6), 2305–2324, doi:10.5194/hess-18-2305-2014, 2014.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R. and Ten Veldhuis, J. A. E.: A statistical analysis of insurance damage claims related to rainfall extremes, Hydrol. Earth Syst. Sci., 17(3), 913–922, doi:10.5194/hess-17-913-2013, 2013.

Spekkers, M. H., Clemens, F. H. L. R. and Ten Veldhuis, J. A. E.: On the occurrence of rainstorm damage based on home insurance and weather data, Nat. Hazards Earth Syst. Sci., 15(2), 261–272, doi:10.5194/nhess-15-261-2015, 2015.

~~Thieken, a. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S. and Schröter, K.: The flood of June 2013 in Germany: how much do we know about its impacts?, Nat. Hazards Earth Syst. Sci. Discuss., (January), 1–57, doi:10.5194/nhess-2015-324, 2016.~~

Thieken, A. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S., and Schröter, K.: The flood of June 2013 in Germany: how much do we know about its impacts?, Nat. Hazards Earth Syst. Sci., 16, 1519-1540, https://doi.org/10.5194/nhess-16-1519-2016, 2016.

Torgersen, G., Bjerkholt, J. T., Kvaal, K. and Lindholm, O. G.: Correlation between extreme rainfall and insurance claims due to urban flooding ??? Case study Fredrikstad, Norway, J. Urban Environ. Eng., 9(2), 127–138, doi:10.4090/juee.2015.v9n2.127138, 2015.

UNISDR. International Strategy for Disaster Reduction (ISDR): Terminology on Disaster Risk Reduction, 2009.

UNISDR. International Strategy for Disaster Reduction (ISDR) Making Development Sustainable: The Future of Disaster Risk Management, 2015.

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., van Vuuren, D. P. and Ward, P. J.: Global drivers of future river flood risk, Nat. Clim. Chang., (December), 1–5, doi:10.1038/nclimate2893, 2015.

Wobus, C., Lawson, M., Jones, R., Smith, J. and Martinich, J.: Estimating monetary damages from flooding in the United States under a changing climate, J. Flood Risk Manag., 7(3), 217–229, doi:10.1111/jfr3.12043, 2014.

Zhou, Q., Panduro, T. E., Thorsen, B. J. and Arnbjerg-Nielsen, K.: Verification of flood damage modelling using insurance data, Water Sci. Technol., 68(2), 425–432, doi:10.2166/wst.2013.268, 2013.

Zhou, Q., Leng, G. and Feng, L.: Predictability of state-level flood damage in the conterminous United States: the role of hazard, exposure and vulnerability, Sci. Rep., 7(1), 5354, doi:10.1038/s41598-017-05773-4, 2017.

Tables

Table 1: Summary of the data used. Precipitation refers to the number of meteorological stations considered; t~~T~~he number of flood ~~episode~~events is the total sum for the period 1996-2015; ~~The~~the average population is the total number of inhabitants~~ in 2015~~; the average Gross Domestic Product is in millions of euros; t~~T~~he damages refer to the compensations paid by the CCS for the 1996-2015 period in millions of euros.

~~1996-2015~~	~~CATALONIA~~	~~MARESME BASIN~~	~~TER BASIN~~	~~MAB~~	~~SOURCE~~
~~Precipitation 24 h~~	~~127~~	~~3~~	~~18~~	~~26~~	~~AEMET~~
~~Precipitation 30-min.~~	~~-~~	~~-~~	~~-~~	~~14~~	~~SMC~~
~~Number of flood events~~	~~166~~	~~68~~	~~38~~	~~61~~	~~INUNGAMA/ PRESSGAMA~~
~~Population~~	~~7,508,106~~	~~638,733~~	~~413,307~~	~~3,213,775~~	~~IDESCAT~~
~~N. of municipalities~~	~~948~~	~~30~~	~~105~~	~~36~~	~~IDESCAT~~
~~Damages~~	~~436.4~~	~~27~~	~~20.2~~	~~86.3~~	~~CCS~~

1996-2015	CATALONIA	MAB	SOURCE
Precipitation 24 h	127	26	AEMET
Precipitation 30 minute	-	14	SMC
Number of flood events	166	61	INUNGAMA/PRESSGAMA
Population	6,854,302	3,141,703	IDESCAT
Gross Domestic Product	164,162.3	95,438.57	IDESCAT
No. of municipalities	948	36	IDESCAT
Damages	436.4	86.3	CCS

Table 2: Parameters of the logistic model and RA values for the basin level with 60 mm/24 h precipitation threshold. All the results are significant (p-value<0.01). Number of flood cases: 239

PERCENTILE	DAMAGE	β_0	β_1	RA
50	D	-5.31	1.16	0.61
	DPC	-9.19	2.00	0.67
	DPW	-8.73	1.90	0.67
60	D	-6.89	1.41	0.64
	DPC	-8.90	1.84	0.67

	DPW	-9.58	1.99	0.68
70	D	-7.65	1.47	0.65
	DPC	-11.26	2.24	0.72
	DPW	-10.50	2.08	0.70
80	D	-10.19	1.89	0.70
	DPC	-10.44	1.94	0.70
	DPW	-11.84	2.24	0.73
90	D	-11.13	1.90	0.71
	DPC	-11.58	1.99	0.70
	DPW	-12.86	2.26	0.74

Table 3: Parameters of the logistic model and RA values for the MAB level with 10 mm/30 minute precipitation threshold. All the results are significant (p-value<0.05). Number of flood cases: 38

PERCENTILE	DAMAGE	β_0	β_1	RA
50	D	-14.61	4.7	0.88
	DPC	-14.61	4.7	0.88
	DPW	-10.02	3.21	0.81
60	D	-13.34	4.06	0.85
	DPC	-13.34	4.06	0.85
	DPW	-13.72	4.18	0.86
70	D	-11.30	3.21	0.81
	DPC	-11.30	3.21	0.81
	DPW	-15.05	4.33	0.87
80	D	-16.62	4.58	0.89
	DPC	-16.62	4.58	0.89
	DPW	-16.62	4.58	0.89
90	D	-17.72	4.53	0.91
	DPC	-17.72	4.53	0.91
	DPW	-17.72	4.53	0.91

~~Table 2: The results of correlations between 24 h precipitation and CCS compensation for the 4 different criteria applied.~~

~~METHODOLOGY~~	~~CORRELATION VALUE~~
~~1: PPT 24 h Catalonia vs total damages Catalonia~~	~~0.57**~~
~~2: PPT 24 h basins vs total damages Catalonia~~	~~0.53**~~

~~3: PPT 24 h maximum of all affected basins vs total damages of affected basins~~	~~0.60**~~
~~4: PPT 24 h vs damaged basins affected (one by one)~~	~~0.43**~~

~~* p-value<0.05; **p-value<0.01~~

~~Table 3: The results of correlations between 24 h precipitation and CCS compensation for the selected basins (1996-2015).~~

~~BASIN NAME~~	~~BASIN SURFACE (km^2)~~	~~FLOOD EVENTS~~	~~CORRELATION VALUE~~
~~Besòs~~	~~1,024~~	~~41~~	~~0.43*~~
~~Ter~~	~~2,960~~	~~38~~	~~0.37~~
~~Llobregat~~	~~4,925~~	~~54~~	~~0.49**~~
~~Segre~~	~~11,267~~	~~29~~	~~0.42**~~
~~Maresme~~	~~338~~	~~68~~	~~0.68**~~
~~Streams in the MAB~~	~~93~~	~~36~~	~~0.47**~~
~~Tarragona Sud~~	~~338~~	~~32~~	~~0.43*~~

~~* p-value<0.05; **p-value<0.01~~

~~Table 4: The correlation results for the Maresme/Ter basins.~~

	~~COMPENSATIONS~~	~~FLOOD EVENTS~~	~~POPULATION~~
~~COMPENSATIONS~~	~~-~~	~~-~~	~~-~~
~~FLOOD EVENTS~~	~~0.53**/0.20*~~	~~-~~	~~-~~
~~POPULATION~~	~~0.60**/0.42**~~	~~0.62**/0.34**~~	~~-~~

~~* p-value<0.05; **p-value<0.01~~

~~Table 5: The correlation results for the “urban”/“rural” areas of the Ter basin.~~

	~~COMPENSATIONS~~	~~FLOOD EVENTS~~	~~POPULATION~~
~~COMPENSATIONS~~	~~-~~	~~-~~	~~-~~
~~FLOOD EVENTS~~	~~0.60*/0.04~~	~~-~~	~~-~~
~~POPULATION~~	~~0.81**/0.35**~~	~~0.68**/0.25*~~	~~-~~

~~* p-value<0.05; **p-value<0.01~~

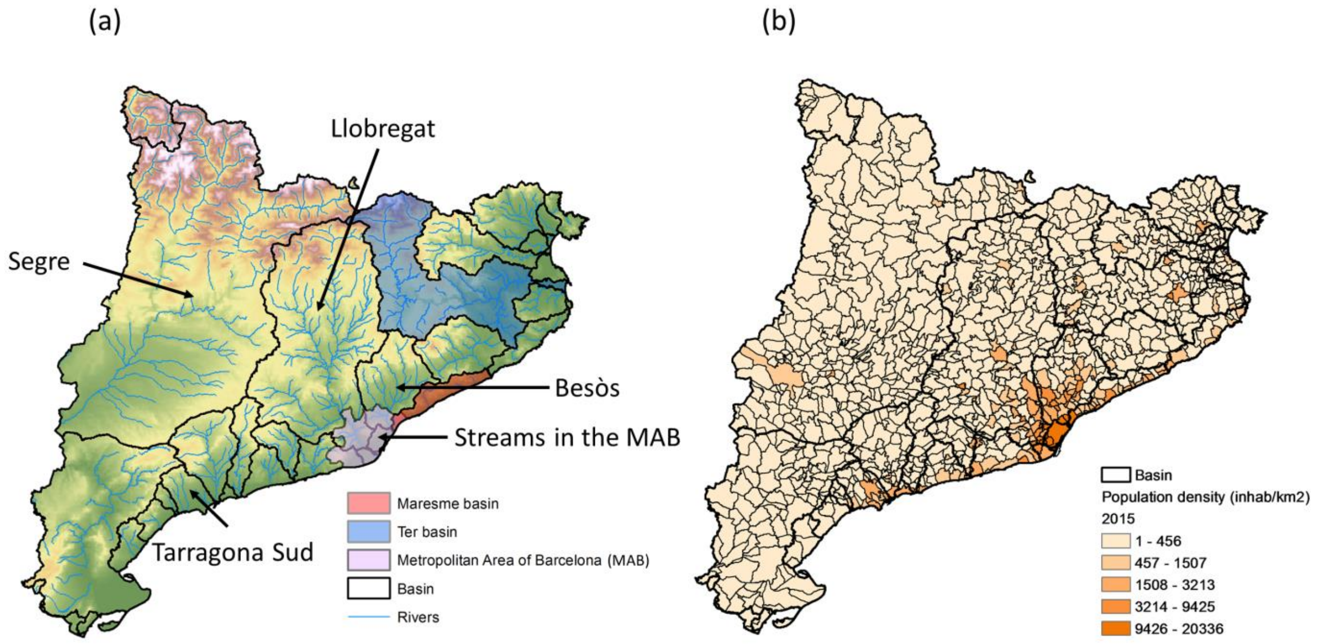
~~Table 6: The correlation results for compensation in the MAB area and different precipitation indices (period: 1996-2015).~~

	~~PRECIPITATION 24 h~~	~~30-MINUTE PRECIPITATION~~	~~ACCUMULATED PRECIPITATION~~	~~COMPENSATIONS~~

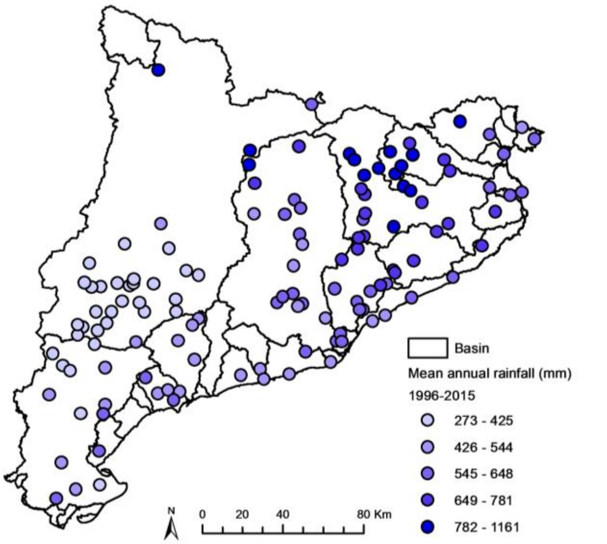
~~PRECIPITATION 24 h~~	-	-	-	-
~~30-MINUTE PRECIPITATION~~	~~0.39**~~	-	-	-
~~ACCUMULATED PRECIPITATION~~	~~0.97**~~	~~0.38**~~	-	-
~~COMPENSATIONS~~	~~0.31*~~	~~0.64**~~	~~0.31*~~	-

~~* p-value<0.05; **p-value<0.01~~

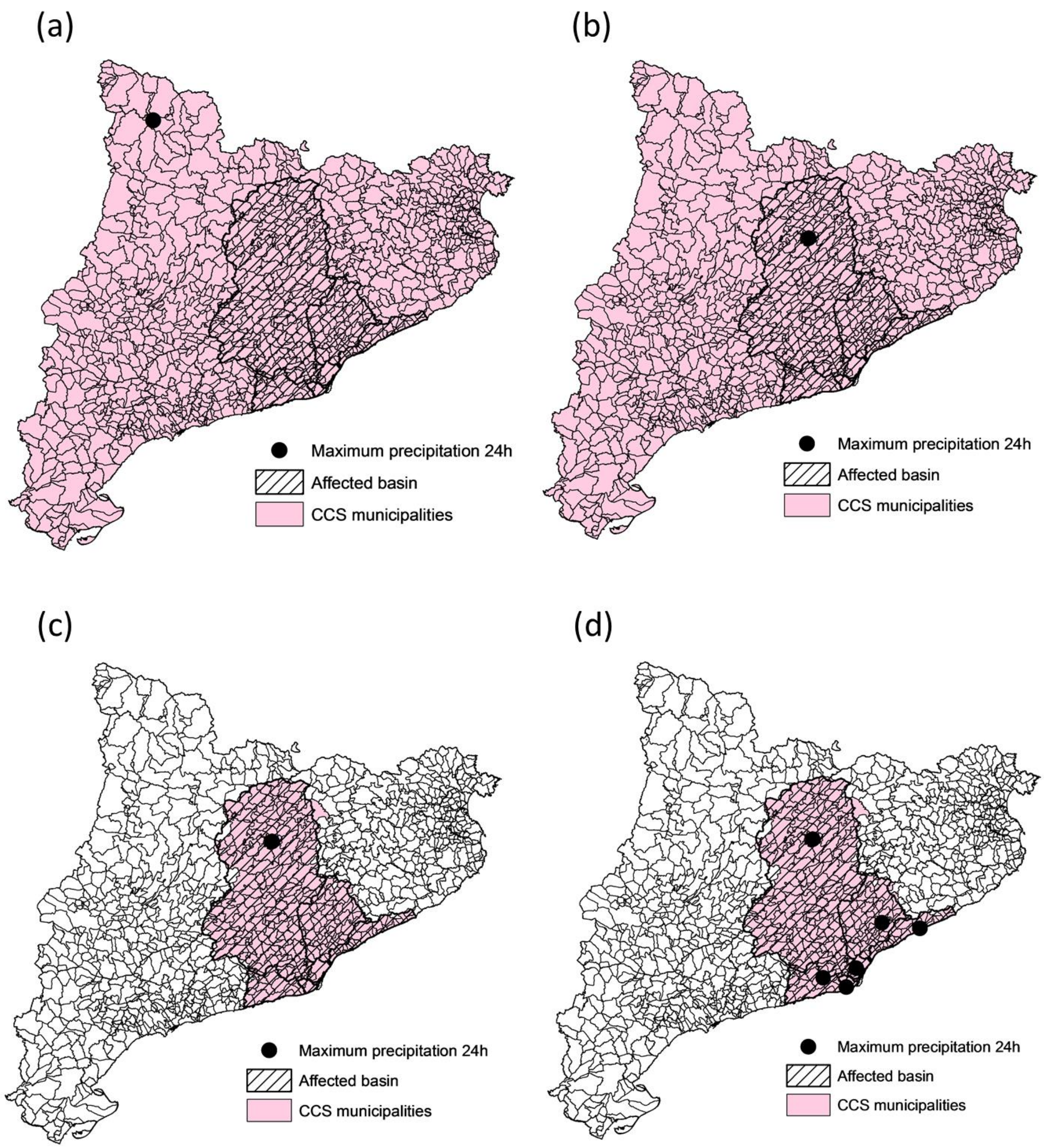
~~**Figures**~~



~~Figure 1: (a) Map of Catalonia showing the basins cited in the text and the Metropolitan Area of Barcelona (MAB); (b) distribution of the municipal population density in 2015.~~

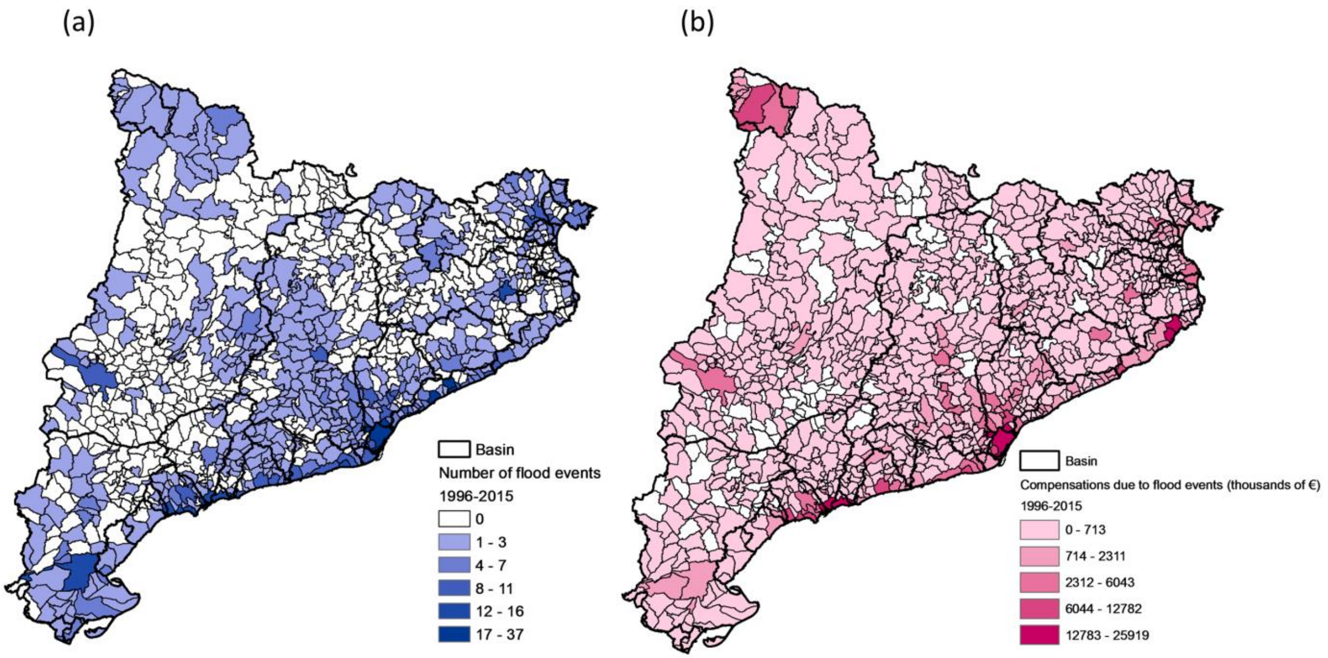


~~Figure 2: Pluviometric stations and average annual precipitation (1996-2015).~~

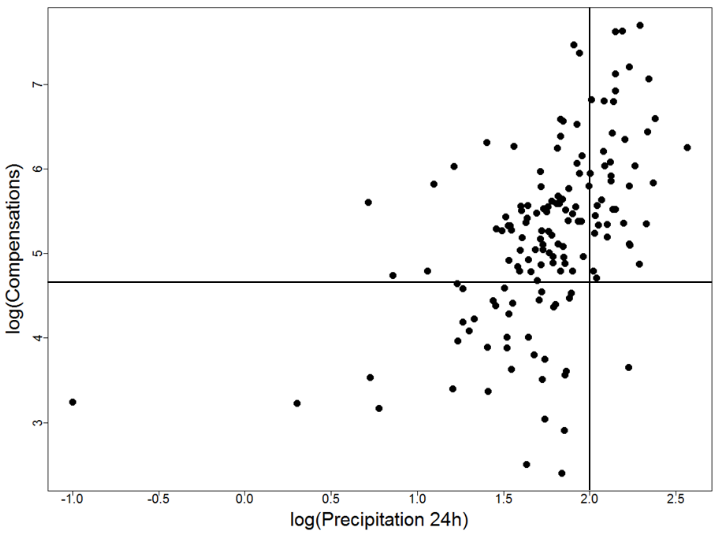


~~Figure 3: A schematic view of the analysis carried out to find the scale with the highest correlation between precipitation and flood damage. The pink shaded area shows the municipalities considered when aggregating damage. The dashed area shows the catchment affected by~~

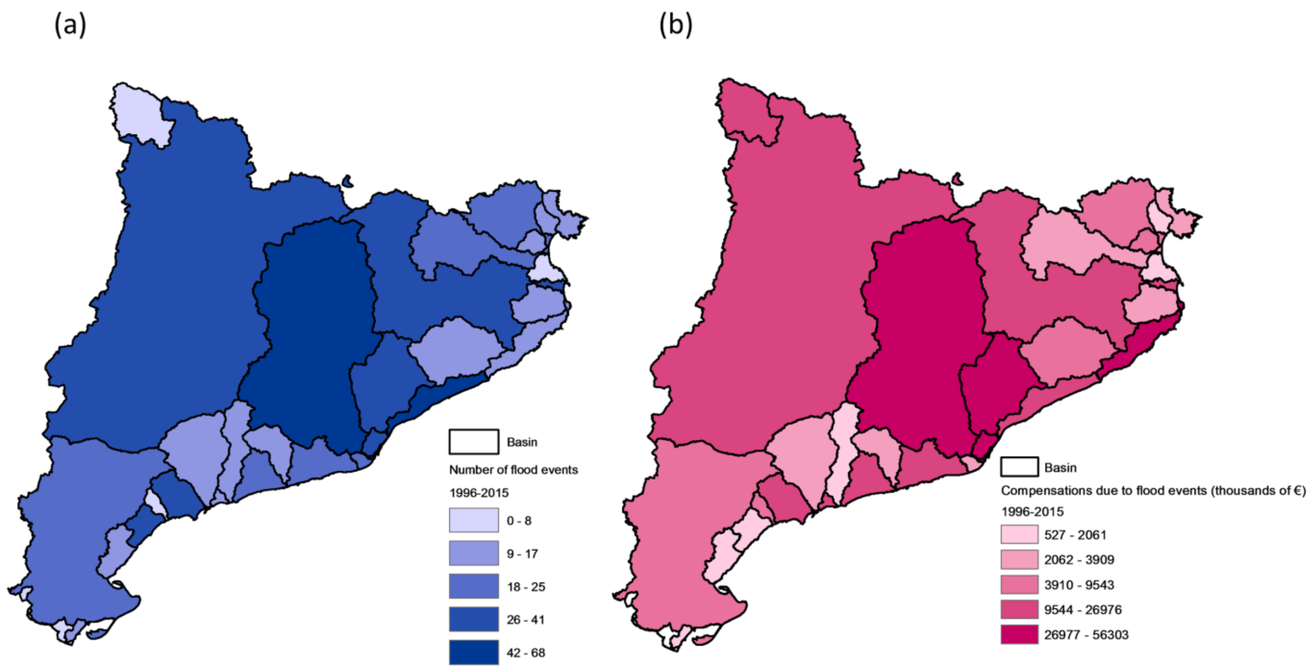
~~floods. The black circles show the areas with maximum precipitation in 24 h in: (a) the region as a whole; (b), (c) all the affected catchments; (d) each affected catchment.~~



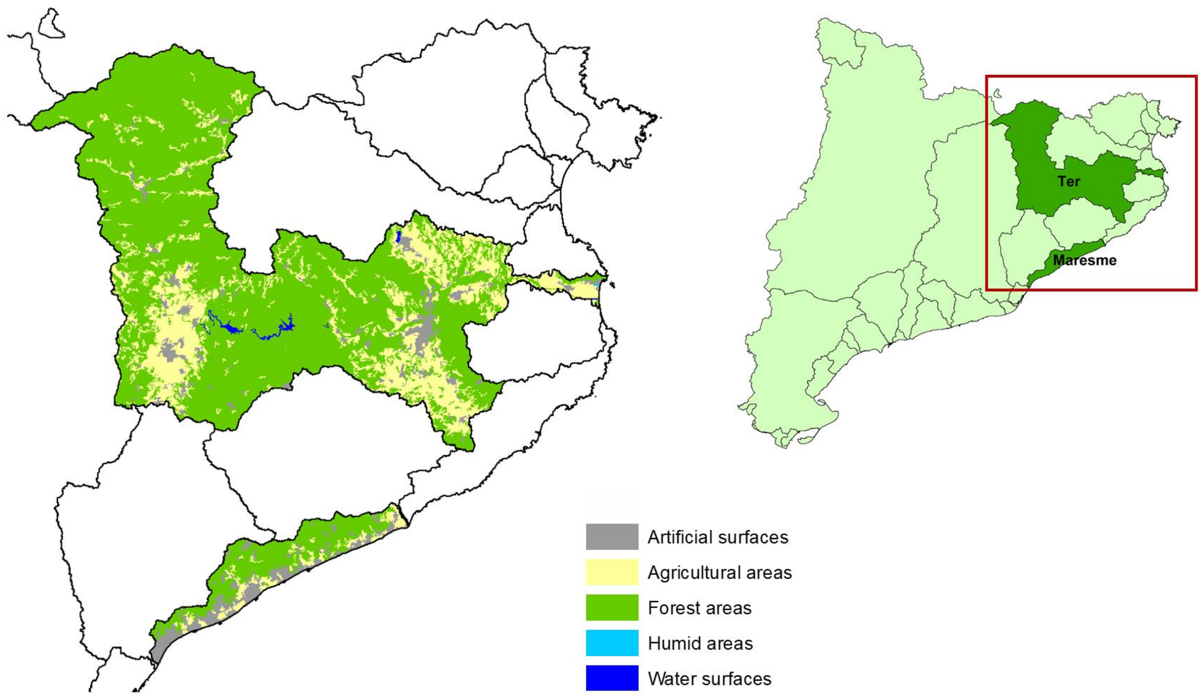
~~Figure 4: (a) Municipal distribution of flood events; (b) municipal distribution of total compensation for floods paid by CCS. Period: 1996-2015.~~



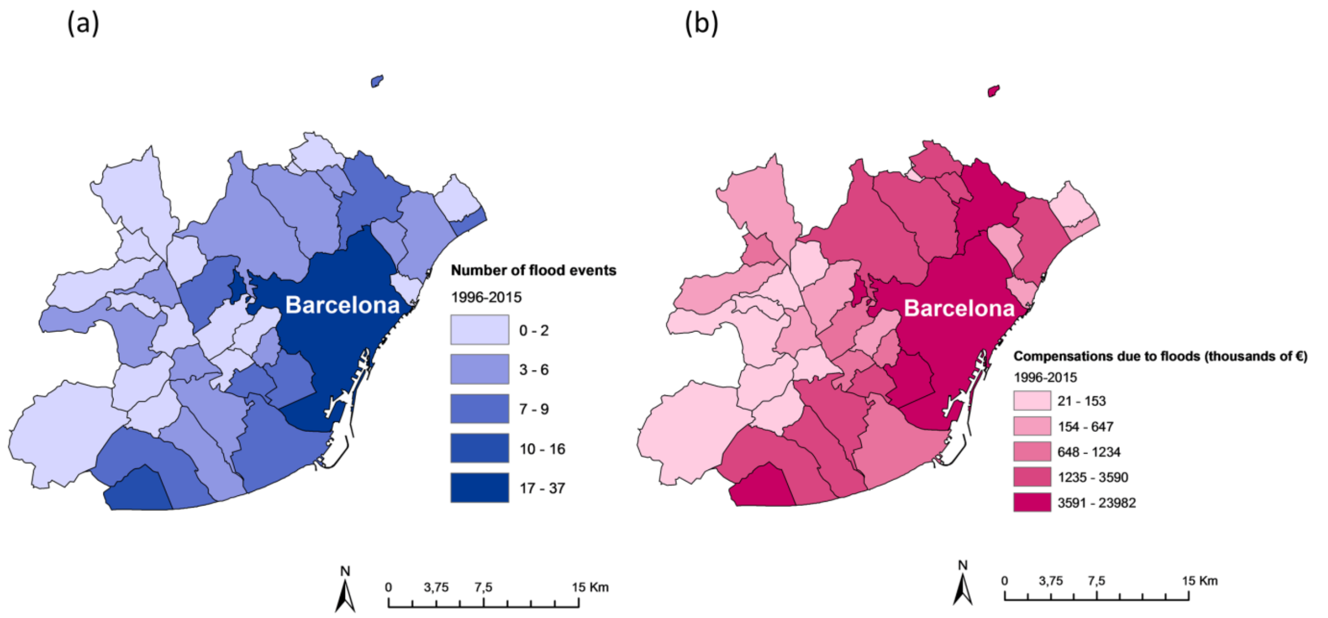
~~Figure 5: Scatter plot (in log scale) between precipitation in 24 h and compensations paid by CCS for flood episodes recorded in Catalonia between 1996 and 2015, using the criteria number 3.~~



~~Figure 6: (a) Number of flood events per basin; (b) total compensation paid by the CCS per basin. Period: 1996-2015.~~



~~Figure 7: Land use (Corine Land Cover 2012) for the Ter and Maresme basins. The 5 land use categories of the Corine Land Cover Map are shown with the corresponding colour code.~~



~~Figure 8: (a) Municipal distribution of the number of flood episodes in the MAB; (b) municipal distribution of the total compensation paid by the CCS in the MAB. Period 1996-2015.~~