Natural Hazards
and Earth System
Sciences

The relationship between precipitation and insurance data for floods in a Mediterranean region (northeast Spain)

Maria Cortès[1,2], **Marco Turco**[1], **Montserrat Llasat-Botija**[1,2], **and Maria Carmen Llasat**[1,2]

[1]Department of Applied Physics, University of Barcelona, Barcelona, 08028, Spain
[2]Water Research Institute (IdRA), University of Barcelona, Barcelona, 08001, Spain

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

Abstract. Floods in the Mediterranean region are often surface water floods, in which intense precipitation is usually the main driver. Determining the link between the causes and impacts of floods can make it easier to 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 could provide a proxy for flood impact, and this information 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. The relationship between precipitation and damage provides insights into flood risk in the Mediterranean and is also promising for supporting flood management strategies.

1 Introduction

Flooding is one of the largest natural risks in the world. Between 2005 and 2014, more than 85 000 000 people were directly affected by flood events annually, and around 6000 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 damage, and the vulnerability, that is, the characteristics and circumstances of a community/system

that make it susceptible to potential flood damage (UNISDR, 2009; Kundzewicz et al., 2014; Winsemius et al., 2015). Vulnerability can include factors such as exposure and other societal factors such as early warning systems, building capacity to cope with natural hazards and disaster recovery infrastructure (Jongman et al., 2014; Nakamura and Llasat, 2017).

A large number of researchers are making efforts to create 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 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 damage (Barredo et al., 2012). Several recent works have used this kind of data to explore the causes and impacts of floods. For instance, in several European regions researchers have noted that precipitation has a significant influence on flood insurance data (see, for instance, Spekkers et al., 2013, 2015, for the 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). These data are very valuable for establishing causal relationships between the costs of flood damage and precipitation extremes, for developing risk maps and to use 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.

Most floods that have affected the region of study, northeast Spain, are surface water floods that caused catastrophic damage (Llasat et al., 2014, 2016a). This type of flood 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 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. However, relatively few studies exist for the Mediterranean region, being mostly limited to urban flood damage assessment (Freni et al., 2010; Papagiannaki et al., 2015; Bihan et al., 2017), while analysis of the possible links between precipitation and economic flood damages are yet to be undertaken across Mediterranean regions. This may be due to limitations in insurance data records and difficulties in estimating how heavy precipitation could affect monetary damages. In the Mediterranean region of Catalonia, in northeast Spain, 20 years of flood-related insurance damage claims are available from the Spanish public reinsurer, the Insurance Compensation Consortium (Consorcio de Compensación de Seguros, CCS), a public institution that compensates homeowners for damage caused by floods, playing a role similar to that of a reinsurance company (Barredo et al., 2012). This means an assessment of the links between precipitation and impacts is now possible. This analysis would greatly help policy makers and civil protection agencies, improving early warning systems and allowing for more efficient management strategies. Furthermore, assessing the relationship between precipitation and flood damage would provide relevant information on the underlying mechanisms of how floods evolve, as well as those particular to Mediterranean regions.

The aim of this study is to develop and evaluate a methodology to estimate surface water flood damage from heavy precipitation in the Mediterranean region of study (henceforth, "flood" refers to surface water floods). The relationship between precipitation and insurance data has been assessed, using logistic regression models, to ascertain the probability of large monetary damages in relation to heavy precipitation events. Specifically, our main goal is to answer the following research questions:

Can we predict flood damage with parsimonious precipitation–damage models?

To what extent do exposure and vulnerability (through the commonly used proxies of gross regional domestic product, GDP) and population (Pielke and Downton, 2000; Choi and Fisher, 2003; Barredo, 2009) determine the damages corresponding to precipitation events?
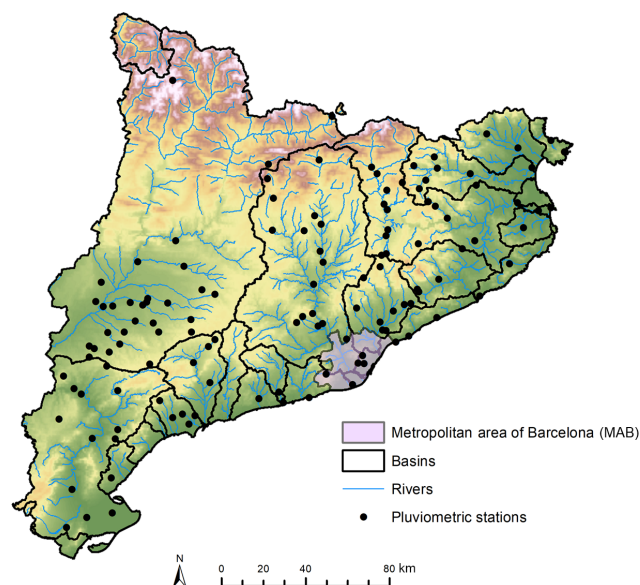
Figure 1. Map of Catalonia showing the aggregated basins, the metropolitan area of Barcelona (MAB), the main rivers and the pluviometric stations used.

Which thresholds used to define large flooding damages and heavy rainfall events determine the best applicability range?

To sum up, the results of this study can help better understand flood risk in Mediterranean areas by analysing flood causes and impacts, and can help more accurately estimate flood damage when high levels of rainfall are forecast.

The study is organised as follows. After the Introduction, the Methods section describes the study region, the observed data and the methodology used. Then, the Results section presents the regression models obtained. Finally, the Conclusions section summarises the main findings of this study.

2 Methods

2.1 Study region

The study area is Catalonia, a Spanish region of $32\,108\,\mathrm{km}^2$ in the northeast Iberian Peninsula. The region is characterised by three mountain ranges (Fig. 1): the Pyrenees in the north (maximum altitude above 3000 m a.s.l.) and parallel to the Mediterranean coast (SW–NE) between the Catalan Pre-Coastal Range (maximum altitude around 1800 m a.s.l.) and the Catalan Coastal Range (maximum altitude around 600 m a.s.l.). This marked orography is the key reason for the development of floods, both from a hydrological point of view (small torrential catchments) and due to meteorological factors (the orography forces water vapour to rise from the Mediterranean, triggering instability; Llasat et al., 2016a). The region is divided into 42 counties and 948 municipalities, with a total population of 7.5 million, most of

whom live along the coast, where more than 70 % of the flood events occur (Llasat et al., 2014), 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 located in very small municipalities for which some data needed are not available (i.e. GDP). For this reason we have aggregated some of the basins and worked with a total of 29 (see Supplement).

We also analyse the metropolitan area of Barcelona (MAB, 534.7 km^2; Fig. 1) 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, it contains 48 % of the population (IDESCAT, 2016). It is affected by an average of over three flood events per year, most of which are floods due to very convective local precipitation (Cortès et al., 2017). The city of Barcelona is crossed by 20 streams that have their source in the Serra de Collserola (Catalan Coastal Range), and which are covered by 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).

2.2 Data

The flood damage data were obtained from the insurance compensation for floods paid by the Spanish Insurance Compensation Consortium (CCS). The CCS compensates for damage 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 on a postal code level for the 1996–2015 period (no previous information is available with this level of detail). 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 events that have occurred in Catalonia on a municipal, district and basin level. 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 GDP 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, AEMET). To ensure temporal homogeneity, we have only considered the stations located in Catalonia with more than 90 % valid data over the 1996–2015 period (Fig. 1). For the MAB we also considered 30 min weather data obtained from the network of automatic meteorological stations belonging to the Meteorological Service of Catalonia (Servei Meteorológic de Catalunya, SMC). Table 1 summarises the data used.

Compensation 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 7-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 7 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

Table 1. Summary of the data used. Precipitation refers to the number of meteorological stations considered; the number of flood events is the total sum for the period 1996–2015; the average population is the total number of inhabitants; the average GDP is in millions of euros; the damages refer to the compensation paid by the CCS for the 1996–2015 period in millions of euros.

1996–2015	Catalonia	MAB	Source
Precipitation 24 h	127	26	AEMET
Precipitation 30 min	–	14	SMC
Number of flood events	166	61	INUNGAMA/ PRESSGAMA
Population	6 854 302	3 141 703	IDESCAT
Gross domestic product	164 162.30	95 438.57	IDESCAT
No. of municipalities	948	36	IDESCAT
Damages	436.40	86.30	CCS

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 recorded, 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 damage are likely to be similar across the area.

The same methodology was applied for another spatial aggregation based on the AEMET warning areas (included in the Supplement), 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 (14 to 15 warning areas).

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

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^{-1} is better for urban areas. In the main text we consider the threshold on 60 mm 24 h^{-1}, while results obtained using the lower threshold are available in the Supplement.

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^{-1}, 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^{-1} to study flood events in this

Mediterranean region. Since the sample size is still small, a 10 mm threshold was also used (results for the 20 mm threshold are, however, available in the Supplement).

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.

The large spread of Fig. 2 indicates that modelling insurance compensation 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 insurance compensation and precipitation suggests that rainfall data can be used to extract information on damage in Catalonia. To do this, we applied a logistic regression model to gauge the probability of large damaging events occurring 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 compensation exceeding the 50th, 60th, 70th, 80th and 90th percentile of the total sample. This methodology is repeated for both thresholds (40 and 60 mm) and for the three damage indicators (D, DPC, DPW) for the basins and warning areas, meaning we made a total of 60 models.

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

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

where π is 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 generalised linear models (GLMs). The Wald χ^2 statistic is used to assess the statistical significance of individual regression coefficients (Harrel, 2015).

2.4 Verification method

We plotted the relative operating characteristic (ROC) diagram, a commonly used logistic prediction diagnostic, show-
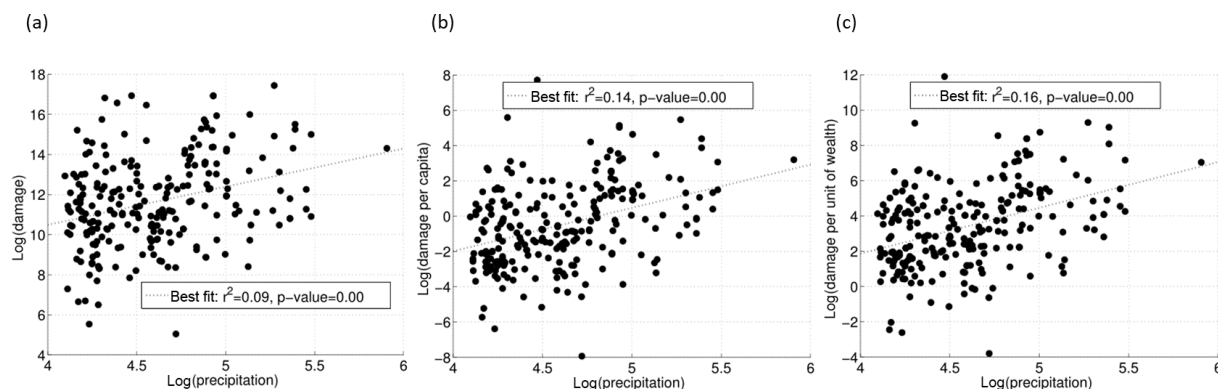
Figure 2. Scatter plot showing maximum precipitation in 24 h (mm) and **(a)** total damages (D), **(b)** damage per capita (DPC), and **(c)** damage per unit of wealth (DPW), for flood events recorded in Catalonia between 1996 and 2015 (log-transformed values; damages are given in euros). Each point represents the insurance compensation series (D, DPC or DPW) and the maximum 24 h precipitation for each basin. The dashed line indicates the fit based on a linear regression model.

Table 2. Contingency table to support Eqs. (2) and (3).

		Observed	
		Yes	No
Forecasted	Yes	a	b
	No	c	d

ing the hit rate (i.e. the relative number of times a forecasted event actually occurred) against the false alarm rate (i.e. the relative number of times an event had been forecasted but did not actually happen) for different potential decision thresholds (Mason and Graham, 2002). Thus, for each insurance compensation percentile and for each precipitation threshold, we first calculated the forecast probabilities for that event, and then grouped the probability forecasts into batches (here 20 with a width of 0.05) to count the observed occurrences/non-occurrences. That is, we converted the observed and forecasted series, expressed as continuous amounts, into "exceedance" categories (yes–no statements indicating whether the data equal or exceed a selected probability). We then plotted the resulting elements on a standard contingency table (see Table 2).

The ROC diagram shows the hit rate (H) against the false alarm rate (F). These indices are defined as

$$H = \frac{a}{a+c}; \quad 0 \le H \le 1, \tag{2}$$

$$F = \frac{b}{b+d}; \quad 0 \le F \le 1. \tag{3}$$

3 Results

3.1 Damaging events and precipitation in Catalonia

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^{-1} precipitation threshold and 81 went over the 60 mm 24 h^{-1} 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^{-1} and 239 for 60 mm 24 h^{-1}). 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 buildings and infrastructure are on the coast), exposure (population and tourism are concentrated in the coastal regions) and hazards (floods associated with 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 these events taking place between September and November (Llasat et al., 2016a). The compensation paid by the CCS for floods during this period in Catalonia amounted to EUR 436.40 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 good correlation among the four variables, as expected. The basins with more recorded flood events are those that received more insurance compensation for flood damage, with a higher population and GDP. The Maresme Basin was affected by 41 % of the recorded events (Fig. 3a) with damages that add up to EUR 24.60 million between 1996 and 2015 (Fig. 3b).
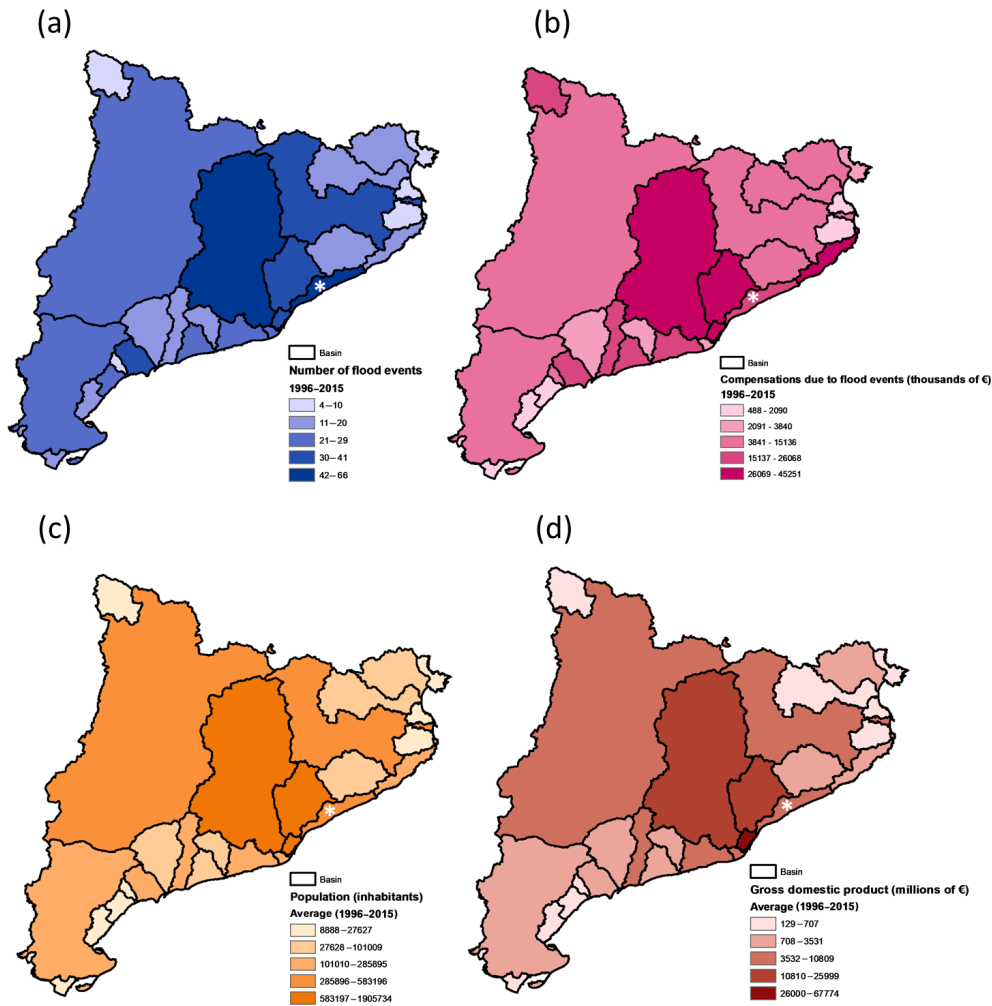
Figure 3. Basin distribution of **(a)** flood events (1996–2015), **(b)** total insurance compensation for floods made by CCS (1996–2015), **(c)** average total population, and **(d)** average GDP. Asterisk indicates Maresme Basin.

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 Eq. (4) is

$$\log\left(\frac{\pi}{1-\pi}\right) = -10.5 + 2.08 \cdot P. \tag{4}$$

Table 3 shows the values of β_0 and β_1, considering cases with a threshold of 60 mm for the different combinations of damage indicators and percentiles.

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 to occur 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 ROC diagram (see Fig. 5).

The area under the ROC curve (RA) is a useful measure for summarising the skill of a model. RA ranges from 0, for
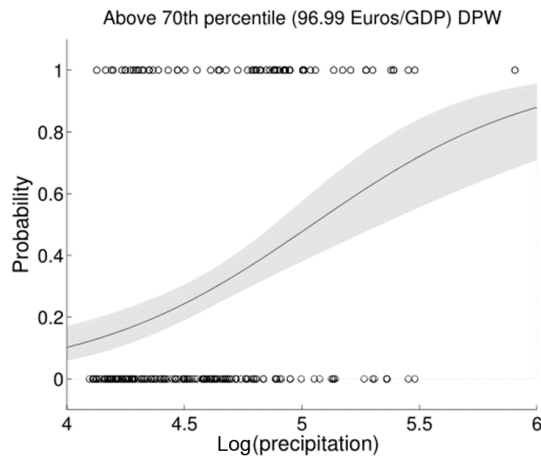
Figure 4. Example of logistic regression result used to model DPW damages above the 70th percentile as a function of precipitation (log-transformed precipitation given in millimetres). The solid line indicates the best estimate while the shaded band indicates the 95 % confidence interval. Open circles along the horizontal axis show the events that are above (top) and below (bottom) the 70th percentile.

Table 3. Parameters of the logistic model and RA values for the basin level with $60\,\text{mm}\,24\,\text{h}^{-1}$ maximum 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

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
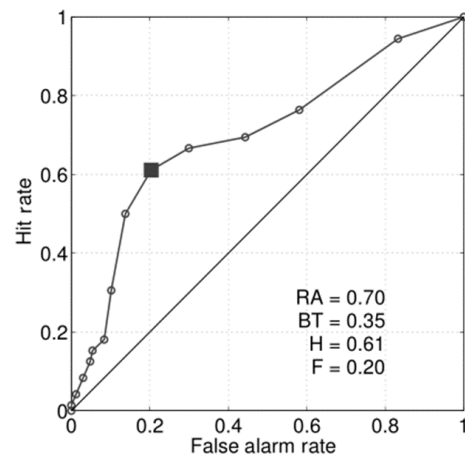

Figure 5. Relative operating characteristic (ROC) diagram for above-70th DPW predictions using the logistic regression of Eq. (1). The open dots indicate a set of probability forecasts by stepping a decision threshold with 5 % probability through the modelling results. The numbers inside the plots are 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).

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 (EUR 97 / GDP).

Table 3 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. Finally, a number of sensitivity tests were carried out. We repeated the analysis considering (i) a precipitation threshold of 40 mm instead of 60 mm, (ii) the AEMET warning areas, (iii) only coastal regions and (iv) the basin-averaged precipitation instead of the maximum values, obtaining similar results (see Supplement).

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 three 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 compensation paid by the CCS for floods

(a)

(b)
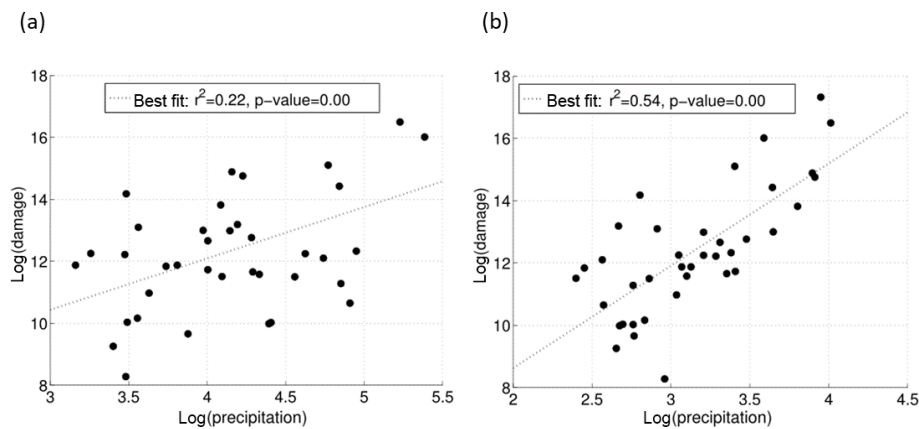
Figure 6. Scatter plot **(a)** damages (D) versus 24 h precipitation and **(b)** damages (D) versus 30 min precipitation in units of log(mm).
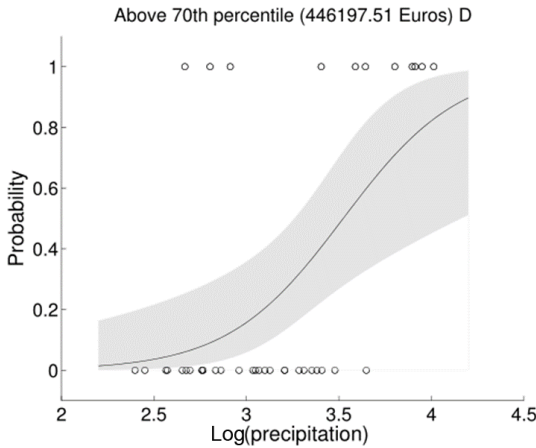
Table 4. Parameters of the logistic model and RA values for the MAB level with $10\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$ maximum precipitation threshold. All the results are significant (p value < 0.05). Number of flood cases: 38.

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

Figure 7. Example of a logistic regression result used to model damages (D) above the 70th percentile as a function of 30 min precipitation in units of log (mm) 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.

amounted to EUR 86.30 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 EUR 19 million).

As can be seen in Fig. 6, the total damages (D) correlate more with 30 min precipitation than with 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 4).

We then repeated the logistic modelling exercise using 30 min 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 min precipitation in this case. As could be expected, this probability increases with precipitation. The same methodology was applied using a precipitation threshold of $20\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$ (see Supplement) and using the 50th, 60th, 80th and 90th percentiles (Table 4). For this example, the regression equation is

$$\log\left(\frac{\pi}{1-\pi}\right) = -11.3 + 3.21 \cdot P. \tag{5}$$

Figure 8 shows the ROC diagram for predictions of total damages (D) above the 70th percentile for the MAB, using a precipitation threshold of $10\,\mathrm{mm}\,30\,\mathrm{min}^{-1}$. The total RA (0.81) shows that our model for the MAB has skill. In
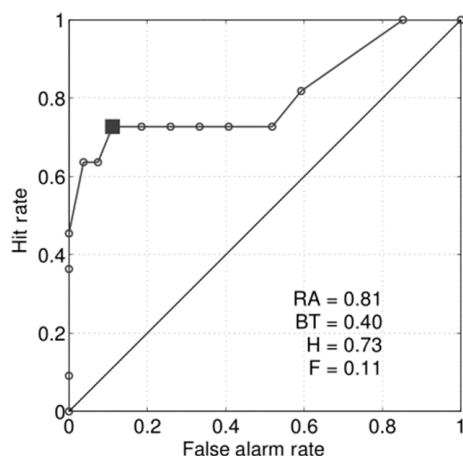
Figure 8. Relative operating characteristic (ROC) diagram for predictions for damage indicator D above the 70th percentile for the MAB using the logistic regression of Eq. (1). The open dots indicate a set of probability forecasts by stepping a decision threshold with 5 % probability through the modelling 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).

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 % are false alarms events ($F = 0.11$). In this example, a precipitation amount higher than $30\,\text{mm}\,30\,\text{min}^{-1}$ is needed to expect a damaging event above the 70th percentile for damage indicator D (EUR 0.45 million).

Table 4 summarises the model parameters and performance considering all the percentiles and the three damage categories used for a precipitation threshold of $10\,\text{mm}\,30\,\text{min}^{-1}$ (see Supplement for results using $20\,\text{mm}\,30\,\text{min}^{-1}$ for the MAB). Similar results in terms of RA have been obtained for damage categories, whether using a 10 mm (Table 4) or a 20 mm threshold (Supplement).

4 Discussion and conclusions

The Mediterranean is an area frequently affected by flood events that produce significant socioeconomic damage. Catalonia, located in the west of the Mediterranean, is affected by an average of more than eight events per year. The majority of the damage caused by these events is due to local events, with intense and short-lived rainfall rather than river overflow (Llasat et al., 2014). Therefore, it is assumed that precipitation is the main contributing factor for damage caused by this type of event. To corroborate this hypothesis, the relationship between precipitation and compensation paid by insurance companies was studied. To take into account the differences in vulnerability and exposure in the ter-

ritory, 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 compensation, 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. 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. That is, our model is able to simulate the probability of a damaging event as a function of precipitation.

The methodology was also been applied for the metropolitan area of Barcelona (MAB) region, an urban area affected by more than three flood events per year. In this case precipitation data at subdaily scale is available. Linear regression has shown that 30 min precipitation is linked more closely with damages than 24 h precipitation, and also the logistic regression models present better results in terms of RA for the urban area considering higher-resolution data, with values higher than 0.8 in all cases. Therefore, we have been able to confirm that 30 min rainfall is a better predictor of the probability of large damages than daily rainfall in urban areas, and this result confirms previous studies such as that of Torgersen et al. (2015), who found a significant relationship between insurance data and short-lasting rainfall when studying urban floods in Norway. In addition, Spekkers et al. (2013) showed that high claim numbers associated with private property and content damage were significantly related to maximum rainfall intensity, based on a logistic regression, with rainfall intensity for 10 min to 4 h time windows.

Overall, our results confirm the hypothesis that precipitation is a key factor in explaining the damage caused by flood events in regions in which water surface floods are the main type of flood, as is the case in the Mediterranean region of study. Also our findings align with the results of previous studies (Spekkers et al., 2013; Zhou et al., 2013; Wobus et al., 2014; Torgersen et al., 2015) and further indicate that insurance databases are a promising source for flood damage assessment at local (Garrote et al., 2016; Le Bihan et al., 2017; Zischg et al., 2018; Zhou et al., 2013) and at regional

scale (Barredo et al., 2012; Kim et al., 2012; Wobus et al., 2014; Zhou et al., 2017).

To summarise, we have developed a new model that allows us to predict the probability that a flood event causing large damage (where the meaning of "large" depends on the user) will occur, based on precipitation, and taking into account the exposure and vulnerability of the region in the model. That is, the parsimonious empirical models linking flood damages to heavy precipitation developed in this study make a substantial contribution towards developing a warning forecast system with flood management strategies. For instance, from the relationship shown between precipitation and insurance compensation it is possible to predict when damaging events will occur as a result of a certain precipitation threshold.

These results were obtained by following a simple and transparent statistical methodology that can also be applied to other areas. These links could also provide a basis to predict flood damage in future climate change scenarios as done for instance by Wobus et al. (2014) that estimated monetary damages from flooding in the United States under a "business as usual" climate change scenario. As a word of caution it is worth noting that the complex relationships between climate variability, human activities and flood damage may limit the applicability of these findings to conditions that are very different from current ones. In addition, more complex analyses including more sophisticated empirical methods, and other factors such as soil physical characteristics (e.g. slope, soil characteristics, vegetation) could provide additional understanding on flood drivers and impacts. For instance, in Garrote et al. (2016) different simulation scenarios were defined considering the modifications to the terrain due to construction of fluvial defence structures in the area.

Despite these limitations, this work has provided the first assessment of the link between precipitation and flood damage in a Mediterranean region, and our results suggest that by exploiting the relationship between precipitation and flood damage, the model could provide satisfactory prediction of monetary compensation.

Data availability. The data are freely available for research purposes by contacting the corresponding author.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/nhess-18-857-2018-supplement.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "Damage of natural hazards: assessment and mitigation". It is not associated with a conference.

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, 2643–2652, https://doi.org/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, 1843–1852, https://doi.org/10.5194/nhess-14-1843-2014, 2014.

Barnolas, M. and Llasat, M. C.: A flood geodatabase and its climatological applications: the case of Catalonia for the last century, Nat. Hazards Earth Syst. Sci., 7, 271–281, https://doi.org/10.5194/nhess-7-271-2007, 2007.

Barredo, J. I.: Normalised flood losses in Europe: 1970–2006, Nat. Hazards Earth Syst. Sci., 9, 97–104, https://doi.org/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, 1723–1729, https://doi.org/10.5194/nhess-12-1723-2012, 2012.

Barrera, 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, 505–518, https://doi.org/10.5194/nhess-6-505-2006, 2006.

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, https://doi.org/10.5194/hess-19-465-2015, 2015.

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, 1659–1682, https://doi.org/10.5194/nhess-17-1659-2017, 2017.

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, https://doi.org/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, 2145–2159, https://doi.org/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, 198–208, https://doi.org/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, 2979–2993, https://doi.org/10.2166/wst.2010.177, 2010.

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

Garrote, J., Alvarenga, F. M., and Díez-Herrero, A.: Quantification of flash flood economic risk using ultra-detailed stage-damage functions and 2-D hydraulic models, J. Hydrol., 541, 611–625, https://doi.org/10.1016/j.jhydrol.2016.02.006, 2016.

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), Anuari estadístic de Catalunya (Statistical Yearbook of Catalonia), available at: http://www.idescat.cat/, last access: 1 June 2017, 2016.

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, 1965–1984, https://doi.org/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. Change, 4, 1–5, https://doi.org/10.1038/NCLIMATE2124, 2014.

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

Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C. J. H., Apel, H., Aronica, G. T., Arnbjerg-Nielsen, K., Bouwer, L. M., Bubeck, P., Caloiero, T., Chinh, D. T., Cortès, M., Gain, A. K., Giampá, V., Kuhlicke, C., Kundzewicz, Z. W., Llasat, M. C., Mård, J., Matczak, P., Mazzoleni, M., Molinari, D., Dung, N. V., Petrucci, O., Schröter, K., Slager, K., Thieken, A. H., Ward, P. J., and Merz, B.: Adaptation to flood risk: Results of international paired flood event studies, Earth's Future, 5, 953–965, https://doi.org/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, 1–28, https://doi.org/10.1080/02626667.2013.857411, 2014.

Le Bihan, G., 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, Hydrol. Earth Syst. Sci., 21, 5911–5928, https://doi.org/10.5194/hess-21-5911-2017, 2017.

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, https://doi.org/10.5194/nhess-9-2049-2009, 2009.

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, https://doi.org/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, 24–37, https://doi.org/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., and 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, 2145–2166, https://doi.org/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, 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, 2469–2485, https://doi.org/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, 919–943, https://doi.org/10.1007/s11069-017-2802-x, 2017.

Papagianaki, 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., 15, 1859–1871, https://doi.org/10.5194/nhess-15-1859-2015, 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, 7, 137–143, 2013.

Pielke Jr., R. A. and Downton, M. W.: Precipitation and damaging floods: Trends in the United States, 1932–97, J. Climate, 13, 3625–3637, https://doi.org/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. Info., 27, 116–130, https://doi.org/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, 2305–2324, https://doi.org/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, https://doi.org/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, 261–272, https://doi.org/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., 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, 127–138, https://doi.org/10.4090/juee.2015.v9n2.127138, 2015.

UNISDR, International Strategy for Disaster Reduction (ISDR): Terminology on Disaster Risk Reduction, Geneva, Switzerland, last access: 10 May 2017, available at: https://www.unisdr.org/we/inform/publications/7817, 2009.

UNISDR, International Strategy for Disaster Reduction (ISDR): Making Development Sustainable: The Future of Disaster Risk Management, Global Assessment Report on Disaster Risk Reduction, Geneva, Switzerland, last access: 10 May 2017, available at: https://www.unisdr.org/we/inform/publications/42809, 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. Change, 1–5, https://doi.org/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, 217–229, https://doi.org/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, 425–432, https://doi.org/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, 5354, https://doi.org/10.1038/s41598-017-05773-4, 2017.

Zischg, A. P., Mosimann, M., Bernet, D. B., and Röthlisberger, V.: Validation of 2D flood models with insurance claims, J. Hydrol., 557, 350–361, https://doi.org/10.1016/j.jhydrol.2017.12.042, 2018.