Natural Hazards
and Earth System
Sciences

# On the occurrence of rainstorm damage based on home insurance and weather data

**M. H. Spekkers, F. H. L. R. Clemens, and J. A. E. ten Veldhuis**

Delft University of Technology, Department of Water Management, Delft, the Netherlands

*Correspondence to:* M. H. Spekkers (m.h.spekkers@tudelft.nl)

**Abstract.** Rainstorm damage caused by the malfunction of urban drainage systems and water intrusion due to defects in the building envelope can be considerable. Little research on this topic focused on the collection of damage data, the understanding of damage mechanisms and the deepening of data analysis methods. In this paper, the relative contribution of different failure mechanisms to the occurrence of rainstorm damage is investigated, as well as the extent to which these mechanisms relate to weather variables. For a case study in Rotterdam, the Netherlands, a property level home insurance database of around 3100 water-related damage claims was analysed. The records include comprehensive transcripts of communication between insurer, insured and damage assessment experts, which allowed claims to be classified according to their actual damage cause. The results show that roof and wall leakage is the most frequent failure mechanism causing precipitation-related claims, followed by blocked roof gutters, melting snow and sewer flooding. Claims related to sewer flooding were less present in the data, but are associated with significantly larger claim sizes than claims in the majority class, i.e. roof and wall leakages. Rare events logistic regression analysis revealed that maximum rainfall intensity and rainfall volume are significant predictors for the occurrence probability of precipitation-related claims. Moreover, it was found that claims associated with rainfall intensities smaller than 7–8 mm in a 60-min window are mainly related to failure processes in the private domain, such as roof and wall leakages. For rainfall events that exceed the 7–8 mm h$^{-1}$ threshold, the failure of systems in the public domain, such as sewer systems, start to contribute considerably to the overall occurrence probability of claims. The communication transcripts, however, lacked information
to be conclusive about to which extent sewer-related claims were caused by overloading of sewer systems or failure of system components.

## 1 Introduction

Heavy rainfall causes considerable damage to building structure and content all over the world. Research on this topic has mainly concentrated on the adverse consequences of river flooding (Douglas et al., 2010; Jongman et al., 2012). Little research focused on damage caused by the malfunction of urban drainage systems and direct water intrusion due to defects in the building envelope. Severe rainstorms have demonstrated that the impact of local high-intensity rainfall to cities can be large. On July 2011, Copenhagen was hit by 150 mm of rainfall in 3 h, which resulted in surcharging sewer systems, flooded houses, shops, roads and railways. Danish insurers received more than 90 000 claims and paid out more than EUR 800 million (2011 value) in compensation (Garne et al., 2013). Another example is the heavy rainfall event of autumn 1998 in the Netherlands, which was associated with a return period of about 125 years and caused around EUR 410 million (1998 value) worth of damage to private buildings and agriculture (Jak and Kok, 2000). The cumulative damage of minor rainfall events can also be considerable in the long run due to their high frequency of occurrence (Ten Veldhuis, 2011).

Many authors, from fields related to different kinds of weather-related risks (e.g. hailstorms, landslides, river flooding, coastal flooding), have recognised that damage data are lacking or biased and that this is limiting the development

of reliable damage models (e.g. Pielke and Downton, 2000; Hohl et al., 2002; Elmer et al., 2010; Gall et al., 2009; André et al., 2013). The same is true for rainstorms; little research has focused on the collection of rainstorm damage data, the understanding of mechanisms causing damage and the deepening of statistical methods to analyse damage data. Among the exceptions are studies by Busch (2008), Smith and Lawson (2011), Einfalt et al. (2012), Cheng (2012), Zhou et al. (2013), Climate Service Center (2013), and Spekkers et al. (2013, 2014), who analysed damage data sources (i.e. from insurance industry, local media, rescue service reports) and their relationships to rainfall data. As a result, there is no strong foundation for the development and validation of prediction models for rainstorm damage. Such models could help homeowners and water authorities to make better decisions on measures to prevent or reduce damage (e.g. retrofitting of buildings and early warning systems).

A potential source of damage data are insurance damage databases. They often contain claims collected over many years and from a large number of insured. A difficulty of insurance databases is that information on the mechanisms that cause damage and building-related, weather and socioeconomic variables are not or only limitedly available in claim data, or cannot easily be retrieved from insurers' data archives (André et al., 2013; Spekkers et al., 2014).

This study aims to quantify the relative contribution of different failure mechanisms to the occurrence of building structure and content damage induced by rainstorms, and to investigate to what extent the probability of occurrence of these processes is related to weather variables. For this purpose, a property level database of around 3100 water-related damage claims was analysed, for a case study in Rotterdam, the Netherlands. An interesting feature of this database is that it includes comprehensive transcripts of communication between insurer, insured and damage assessment experts, which allowed the classification of claims based on the failure mechanisms causing damage. This information is, however, stored in an unstructured way that required substantial data classification efforts before data could be used for the analysis in this present paper.

The outline of this paper is as follows. In Sect. 2 insurance damage data and classification of claims are described, as well as the statistical method used to model the probability of claim occurrence as a function of weather variables. Results of data analyses and regressions are presented in Sect. 3, followed by discussion in Sect. 4. In Sect. 5, conclusions and recommendations are summarised.

## 2 Data and methods

### 2.1 Case study description

This study focuses on Rotterdam, which is, with a population of around 620 000, the second largest city of the

**Table 1.** Inflation adjustment according to the online database of Statistics Netherlands (2014). The average inflation per year for the Netherlands is used, based on the consumer price index. Every damage value associated with a year before 2013 was multiplied with a correction index.

| Year | Inflation [%] | Correction |
|------|---------------|------------|
| 2007 | 1.6 | 1.12 |
| 2008 | 2.5 | 1.10 |
| 2009 | 1.2 | 1.08 |
| 2010 | 1.3 | 1.07 |
| 2011 | 2.3 | 1.05 |
| 2012 | 2.5 | 1.02 |
| 2013 | 2.5 | 1.00 |

Netherlands (Statistics Netherlands, 2014). Because the city is relatively flat (maximum ground level variations of 10–15 m), floods from heavy rainfall are typically characterised by flood depths up to a few decimetres and limited surface run-off. Rotterdam's sewers are mainly combined systems ($\approx 1800$ km), although some parts of the city have separate systems for wastewater and stormwater ($\approx 500$ km) (Gemeente Rotterdam, 2011). The average density of sewer pipes in the city centre is $15.6$ km km$^{-2}$ and $13\%$ of the area is surface water (i.e. city canals and ponds, not rivers) (Statistics Netherlands, 2013; Gemeente Rotterdam, 2014). The majority of the buildings in Rotterdam were constructed in the 20th century. Rotterdam's urban fabric is characterised by a combination of terraced houses and high-rise residential and commercial buildings (Kadaster, 2013). It is assumed that within the study period no changes have been made to the sewer infrastructure or the building portfolio of Rotterdam that have significantly affected the results of the present paper.

### 2.2 Insurance data

Insurance damage data were provided by a Dutch insurance company that is part of the Achmea insurance group[1]. Data are available at property level for the period of January 2007–October 2013 (data collected on: February 2014), containing around 3100 water-related claims. A claim relates to building structure or content damage or a combination of the two, depending on the available insurance policies at the risk address.

From each claim, the following information was available: risk address; type of insurance coverage; damage date; amount of compensation; and detailed transcripts of communication between insurer, insured and damage assessment experts (e.g. calls, abstracts from reports). The database was checked on missing and incorrect values, such as duplicated records, inconsistencies in date formats and claim coding. Every value associated with a year before 2013 was adjusted

---

[1]Website of Achmea insurance group: http://www.achmea.nl.

**Table 2.** Key features of the home insurance policy related to the damage database used in present study.

| | Content insurance | Property insurance |
|---|---|---|
| For whom? | Homeowners and tenants | Homeowners, landlords, housing cooperatives, homeowners associations |
| Covers physical damage to | – Portable goods<br>– Semi-permanent objects (e.g. curtains, laminate, carpet, window blinds, shutters)<br>– Additions or refurbishments to the property which enhance the property value that have been made by a tenant ("tenants improvements") | – Building<br>– Building foundation<br>– Garden, garden sheds<br>– Permanent floors (e.g. floor tiles, glued wooden floors)<br>– Kitchens |
| Damage assessment is based on | Replacement value or current value if replacement value is less than 40 % of current value | Costs to repair or rebuild (part of) the building, deprecation costs |
| Other compensations | Temporary housing, costs of damage experts, costs to clean and dry goods and materials and costs to detect and repair leakages | |
| Damage assessment by means of | "Small" claims → proof of payment<br>"Large" claims → independent damage assessment expert | |
| Grounds for rejection | – Negligence by insured (e.g. windows or doors that were left open during rainfall events, valves of the central heating system that were not closed properly after refilling the system, no leaf basket installed in rain gutter)<br>– Lack of maintenance (e.g. poor quality sealant joints between walls and floors, rain gutter clogged with leaves)<br>– Damage caused by "slow" processes (e.g. rotting, moisture intrusion through walls)<br>– Construction errors (i.e. liability of building contractor or water company)<br>– Costs not covered (e.g. costs to repair leakage are in some cases not compensated)<br>– Floods from rivers or sea<br>– Groundwater flooding | |
| Others | – In case of underinsurance (i.e. insured sum is less than asset value), compensation is proportional to the level of underinsurance<br>– The insurance policies do not have deductibles | |

for inflation according to the correction indices in Table 1. On average, the data set contained information of around 16 000 risk addresses, which is 6 % of the total number of households in Rotterdam. These risk addresses constitute a total number of around 21 000 insurance policies, of which around 6000 insurance policies relate to building structure insurance and around 15 000 to building content insurance. These numbers relate to data from one insurance company of the Achmea insurance group and do not reflect the market share in Rotterdam of the Achmea insurance group as a whole. Table 2 summarises the key features of the home insurance policy related to the present database.

The general rule for a claim to be accepted is that damage should be unforeseen and to have occurred suddenly. Damage due to river flooding is not covered. Damage due to pluvial flooding is covered, provided that the damage is directly and solely related to localised heavy rainfall (Ministry of Transport, Public Works and Water Management, 2003).

## 2.3 Classification of claims

For the purpose of this study, claims were manually classified according to the actual cause of damage using the information in the communication transcripts. Transcripts contain telegram style summaries of calls and abstracts from reports and typically vary in length, from a few lines to a few thousand words. When a claim is first reported at the insurer's call centre, the client is asked a few basic questions to verify if the client was indeed insured at the time of the damaging event and to make a quick assessment on the severity of the damage (e.g. "Is the risk address still habitable?"). Follow-up calls typically describe the actual cause of damage, an inventory of damaged goods and materials and the costs related to cleaning, drying, repairing or replacing goods and materials.

An easy-to-use web interface and SQL database was built based on the classification scheme listed in Table 3. Failure mechanisms described in Table 3 are also shown graphically in Fig. 1. Only one cause class per claim could be selected. Labels were given to each cause class to indicate whether the class relates to precipitation or not and whether the class relates to failures of systems in the public domain (i.e. responsibility of water authorities) or private domain (i.e. responsibility of homeowner, landlord or housing cooperative). Next to the classification scheme, a number of checkboxes was available to specify if (1) building or content was underinsured, (2) insured has not responded to a request for a long time, (3) claiming process is still ongoing (a typical processing time is a few months), (4) claim was rejected because of

**Table 3.** Classification scheme of water-related failure mechanisms applicable to residential buildings. The column "Precipitation?" indicated if the claim is related to precipitation. The column "Domain" indicates whether damage prevention mainly concerns homeowners (private) or water authorities (public).

| Id | Short name | Description | Precipitation? | Domain |
|----|-----------|-------------|----------------|--------|
| 1 | Roof and wall leakages | Rainwater intrusion through roofs, facades, walls, wall-window interfaces and closed doors, which includes rainwater intrusion as a result of overloaded rain gutters | Yes | Private |
| 2 | Rainwater through open window | Rainwater intrusion through open windows, open doors | Yes | Private |
| 3 | Hail impacting roofs | Hail impacting roofs or windows | Yes | Private |
| 4 | Precipitation-related in private domain[1] | Precipitation-related in private domain, but other or unknown actual cause | Yes | Private |
| 5 | Melting snow | Intrusion of melting snow and ice, in particular snow blowing up under roof tiles | Yes | Private |
| 6 | Blocked roof gutters | Overflowing of roof gutters due to blockages in gutter or downpipe (e.g. by leaves or ice) | Yes | Private |
| 7 | Sewer flooding | Flood water entering buildings through doors or openings as a result of overloaded public sewer systems, including sewer backups | Yes | Public |
| 8 | Depression filling | Flood water entering buildings through doors or openings as a result of depression filling, i.e. rainwater filling up depressions if no drainage facilities are available | Yes | Public |
| 9 | Blocked sewer inlets | Flood water entering buildings through doors or openings as a result of blocked sewer inlets | Yes | Public |
| 10 | Flooding from local watercourses | Flood water entering buildings through doors or openings as a result of flooding from local watercourses (e.g. city canal, pond) | Yes | Public |
| 11 | River flooding | Flood water entering buildings though doors or openings as a result of flooding from river systems | Yes | Public |
| 12 | Precipitation-related in public domain[1] | Precipitation-related in public domain, but other or unknown actual cause | Yes | Public |
| 13 | Precipitation-related[1] | Precipitation-related, but other or unknown actual cause | Yes | Unknown |
| 14 | Leakages of household appliances | Leakages of household appliances (e.g. washing machines, dishwashers, aquaria, waterbeds) | No | Private |
| 15 | Burst household water supply pipes | Burst household water supply pipes, including attached facilities | No | Private |
| 16 | Leakages of central heating systems | Leakages of central heating systems, which includes boilers, radiators and pipes | No | Private |
| 17 | Blocked or leaking household wastewater systems | Flooding of wastewater due to blockage in or leakage of wastewater system located inside the building | No | Private |
| 18 | Non-precipitation-related in private domain[1] | Non-precipitation-related in private domain, but other or unknown actual cause | No | Private |
| 19 | Burst public water supply pipes | Burst public water supply pipes owned by water supply company | No | Public |
| 20 | External water discharges | External water discharges (e.g. extracted groundwater from a construction site, fire extinguishing water) | No | Public |
| 21 | Blocked public wastewater system | Flooding of wastewater due to blockage in sewer lateral or sewer main, not related to rainfall events | No | Public |
| 22 | Non-precipitation-related in public domain[1] | Non-precipitation-related in public domain, but other or unknown actual cause | No | Public |
| 23 | Non-precipitation-related[1] | Non-precipitation-related, but other or unknown actual cause | No | Unknown |
| 24 | Water discharge from neighbours[1] | Water discharge from neighbours, but other or unknown actual cause | Unknown | Private |
| 25 | Groundwater flooding | Groundwater flooding due to persistent rainfall or sudden wall failure | Unknown | Unknown |
| 26 | Water-related[1] | Water-related, but other or unknown actual cause | Unknown | Unknown |

[1] Residual group; a group of claims for which exact failure mechanisms could not be derived from communication transcripts.

a lack of building maintenance and (5) damage was (partly) not insured.

Classification was done by three people, by dividing the data set into three independent subsets containing 60, 36 and 4 % of the claims. On average, classification took 4 min per claim. The entire text was read first while making preliminary classification choices. A second reading was used to verify and finalise selections. If the available information was unclear or multi-interpretable, the claims were flagged for investigation by one of the other two persons. This happened to be the case for 7 % of the claims.

### 2.4 Weather variables

A set of weather variables was derived for each combination of risk address and day (i.e. a case) to investigate explanations for claim occurrence (Table 4). Rainfall volume and maximum rainfall intensity were extracted from weather radar data, provided by the Royal Netherlands Meteorological Institute (KNMI), according to a method described in Spekkers et al. (2014). Maximum rainfall intensity was calculated using a 15-min, 30-min and 60-min moving time window to study typical time scales of failure processes. Rainfall duration was not considered, because previous stud-

ies based on a similar insurance database for the Netherlands show that rainfall duration has no significant or weak effect to rainfall-related damage (Spekkers et al., 2013, 2014). Maximum temperature, daily-averaged wind speed, maximum hourly wind speed and wind gust were obtained from an automatic weather station operated by the KNMI, located in the north of the city, around 10 km from the city centre. The season of the year was included to account, for instance, for the occurrence of snow, hail and blockages of rain gutters due to autumn leaf fall.

### 2.5 Modelling the probability of claim occurrence

The modelling objective was to test the significance of weather variables in explaining the occurrence of precipitation-related claims. For each case, a unique combination of risk address and day, the outcome ($Y_i$) can be a reported claim (1) or not (0). The binary outcome can be linked to a set of weather variables ($x_1, \ldots, x_n$) using various types of models for binary data (McCullagh and Nelder, 1989). In this study a logistic regression model was used:

$$\log\left(\frac{\theta_i}{1 - \theta_i}\right) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_n x_{ni}, \tag{1}$$
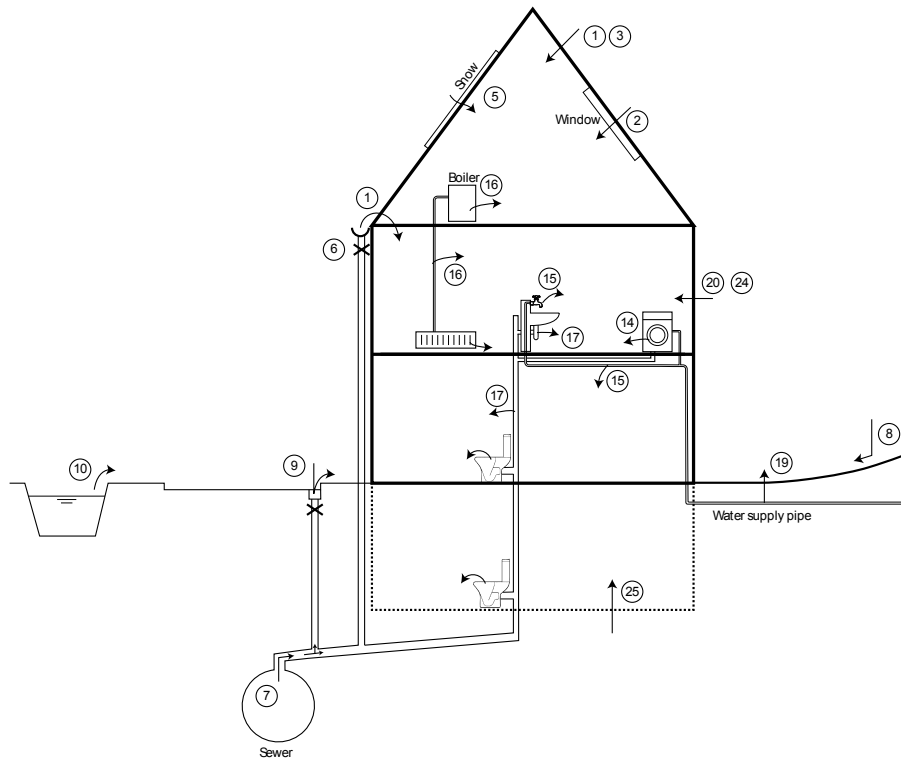
**Figure 1.** Water-related failure mechanisms applicable to residential buildings. A situation with a combined sewer system is depicted. Descriptions of the numbers are listed in Table 3.

**Table 4.** Definitions of explanatory variables and variable value ranges.

| Variable name | Definition | Min – Median – Max |
|---|---|---|
| Rainfall volume (vol) | Volume of rainfall event at the radar pixel intersecting the building's centroid (mm) | $0 - 4.9^1 - 86.2$ |
| Maximum rainfall intensity ($\mathrm{max}_{15}$) | Maximum intensity of rainfall event at the radar pixel intersecting the building's centroid, using a 15-min moving time window ($\mathrm{mm\,h}^{-1}$) | $0 - 3.8^1 - 102.3$ |
| Maximum rainfall intensity ($\mathrm{max}_{30}$) | Maximum intensity of rainfall event at the radar pixel intersecting the building's centroid, using a 30-min moving time window ($\mathrm{mm\,h}^{-1}$) | $0 - 2.8^1 - 62.3$ |
| Maximum rainfall intensity ($\mathrm{max}_{60}$) | Maximum intensity of rainfall event at the radar pixel intersecting the building's centroid, using a 60-min moving time window ($\mathrm{mm\,h}^{-1}$) | $0 - 2.0^1 - 34.2$ |
| Maximum temperature (temp) | Maximum temperature measured at the KNMI Rotterdam weather station (°) | $-6 - 14.8 - 35$ |
| Daily-averaged wind speed ($\mathrm{wind}_d$) | Daily-averaged wind speed measured at the KNMI Rotterdam weather station ($\mathrm{m\,s}^{-1}$) | $0.7 - 4 - 14.3$ |
| Maximum hourly wind speed ($\mathrm{wind}_h$) | Maximum hourly-averaged wind speed measured at the KNMI Rotterdam weather station ($\mathrm{m\,s}^{-1}$) | $2 - 6 - 16$ |
| Wind gust ($\mathrm{wind}_g$) | Wind gust measured at the KNMI Rotterdam weather station ($\mathrm{m\,s}^{-1}$) | $3 - 11 - 28$ |
| Season (seas) | Season of the year: winter = Dec–Feb, spring = Mar–May, summer = Jun–Aug, autumn[2] = Sep–Nov | NA |

[1] Median based on non-zero values only, [2] The level "autumn" was dropped to avoid multicollinearity.

where $\theta_i$ is the probability of claim occurrence ($Y_i = 1$) and $\beta_0, \ldots, \beta_n$ are regression coefficients. The regression coefficients were estimated using maximum likelihood estimation. The significance of the regression coefficients is tested using the Wald test. Logistic regression is known to generate biased estimates for rare events data, i.e. data series in which

only a low percentage of events occur, resulting in an underestimation of the probability of rare events (King and Zeng, 2001). In present database, only 1031 precipitation-related claims were recorded in the period of 2007–October 2013, which is on average $2.67 \times 10^{-5}$ claims per day per risk address. King and Zeng (2001) proposed a method, called
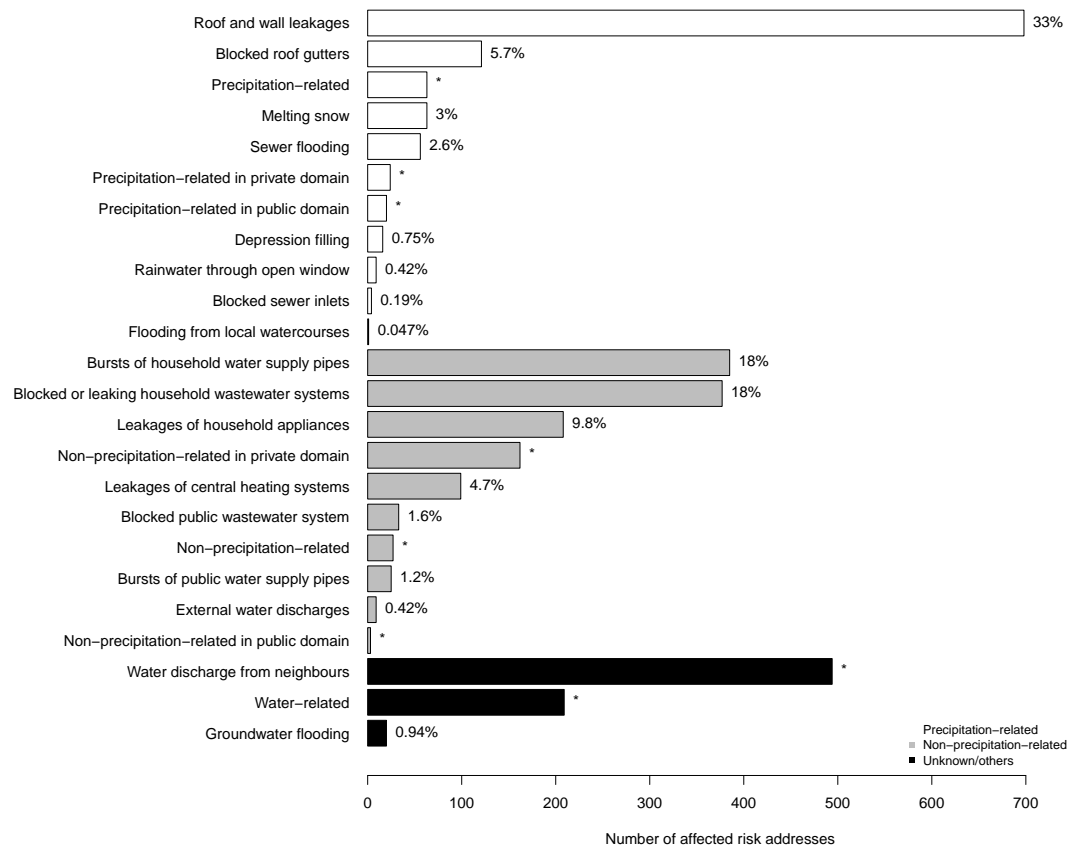
**Figure 2.** Occurrence rates and relative occurrence frequencies of failure mechanisms causing water-related claims ($n = 3126$). An asterisk next to a bar indicates a residual group: a group of claims for which exact failure mechanisms could not be derived from communication transcripts. Percentages are based on the number of claims in the non-residual groups.

rare events logistic regression, to deal with rare events data. This method encompasses a case-control design where ten times more non-events (i.e. no claim from an insured) are selected than events (i.e. a claim from an insured). The method first estimates regression coefficient using an ordinary logistic regression model (Eq. 1), then correcting regression coefficients for finite sample and rare events bias. For this purpose, the Rare Events Logistic Regression (relogit) routine from the Zelig package (Imai et al., 2007) for $R$ was used. Collinearity among explanatory variables was tested by calculating the Pearson's correlation coefficient between each pair of explanatory variables. None of the correlation coefficients yielded values > 0.7, which means that collinearity effects can be neglected (Dormann et al., 2013).

The likelihood ratio and a pseudo-$R^2$ statistic were used to evaluate goodness-of-fit of a model. The likelihood ratio compares the likelihood of a model with predictors to the likelihood of a model without predictors (i.e. intercept-only model), which tests if adding explanatory variables to a model significantly improves its fit. For logistic regression, there is no universally accepted measure that represents the proportion of variance explained by the predictors, such as

$R^2$ for ordinary least-squares regression. Several pseudo-$R^2$ statistics exist; however, these statistics generally score much lower than their equivalent in ordinary least-square regression and are therefore found less informative. They can be used, nevertheless, to compare predictability of nested models. In this study, McFadden's $R^2$ is used (e.g. Long, 1997).

## 2.6 Discarded data

During the validation process of the insurance data, it was found that on three extremely stormy days (i.e. storm Kyrill on 18 January 2007 and storms on 27 July 2013 and 28 October 2013), despite occurrence of rainfall, no or hardly any precipitation-related claims were recorded. Upon further inquiry, the insurer has indicated that on extremely stormy days, precipitation-related claims are often inaccurately recorded as storm-related claims. These 3 days were therefore excluded from the logistic regression analysis.

# 3 Results

## 3.1 Relative occurrence frequencies and costs of claims

Analyses of the relative occurrence frequencies of damage causes show that leakage of roofs and walls is the most frequent failure mechanism generating water-related claims, followed by burst household water supply pipes, blockage or leakage of household wastewater systems and leakage of household appliances (Fig. 2). Besides roof and wall leakages, other common precipitation-related failure mechanisms are blocked roof gutters, snow melting under roof tiles and sewer flooding.

In general, 34 % of the claims were related to precipitation and 43 % to non-precipitation causes. For the remaining 23 %, it was unknown if the claim was related to precipitation or not. These unknowns include claims caused by water discharges from neighbouring properties and groundwater flooding. For groundwater flooding particularly, there was insufficient information to be able to distinguish between floods as a result of persistent rainfall or because of sudden wall failures not related to rainfall. For insurers, there is no strong need to collect information on the actual cause of groundwater flooding as they do not compensate for this type of flood (see also Table 2).

The top of Fig. 3 shows the yearly distribution of precipitation-related claims (white), non-precipitation-related claims (grey) and claims for which the cause was unknown (black), for the years 2007–2013. There is an increase in the number of claims through the years. This increase is most apparent between 2008 and 2010 for precipitation-related claims and between 2009 and 2012 for non-precipitation-related claims. Possible explanations of these trends are discussed in Sect. 4. Most precipitation-related claims are recorded in July–August and December–January (bottom of Fig. 3); the December–January claims can partly be explained by the damage due to melting snow. Claims related to sewer flooding mainly occur in June–August.

Although claims related to sewer flooding were less present in the data, they are associated with significantly larger claim sizes (EUR 1150–3160, based on the 95 % confidence interval around the median in Fig. 4) than claims generated by roof and wall leakages (EUR 680–840), the majority class. Sewer floods are costly because of the required (chemical) cleaning of sewage spills and replacement of goods that cannot be cleaned properly. In contrast, costs related to roof and wall leakages, which usually do not involve large water volumes, are relatively low and are limited to the repair of the leak and the painting of walls and ceiling.

Based on a qualitative analysis of outliers, it was found that exceptionally large claim sizes are related to cases where water leakage could not be stopped easily (e.g. a burst water supply pipe just outside the property), flooding occurred
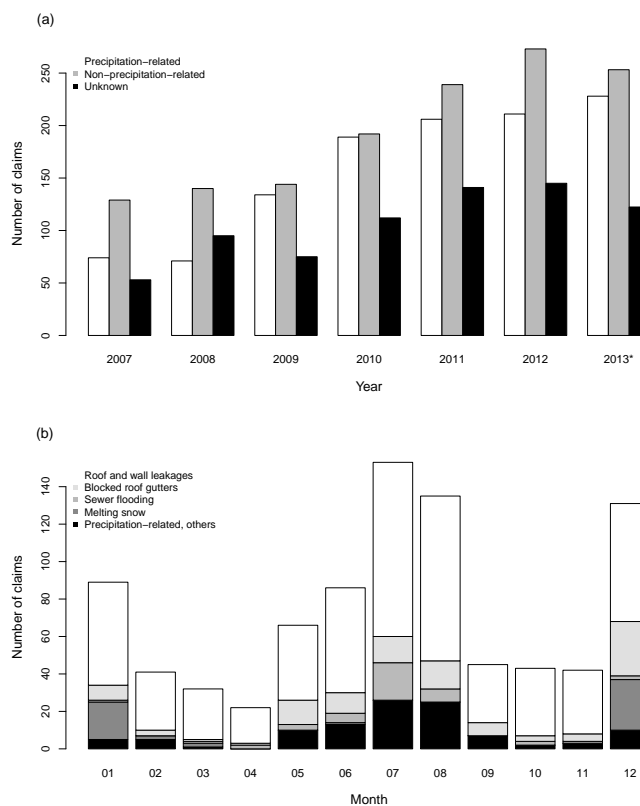


**Figure 3.** Yearly and monthly distribution of the number of claims: **(a)** the yearly distribution of precipitation-related claims (white), non-precipitation-related claims (grey) and claims for which the cause was unknown (black), for the years 2007–2013. The values of the year 2013 (denoted with an asterisk) are estimated, because no data was available for the months November and December; **(b)** The monthly distribution of precipitation-related claims for the years 2007–2012, per cause class. The year 2013 is excluded because data was not available for the entire year.

while no one was at home or temporary housing was required.

## 3.2 Effects of rainfall intensity on claim occurrence probability

In Fig. 5, the empirical probability of precipitation-related claim occurrence per day per risk address is shown, as a function of the rainfall intensity (black dots, based on 1031 claims). Within the subset of precipitation-related claims, a further distinction was made between the occurrence probability of claims caused by failure of systems in the private domain (grey dots, 876 claims) and the public domain (light grey dots, 89 claims), according to column 5 ("Domain") of Table 3. The empirical probability is calculated as follows: within a bin, with a size of 5 % of the range of $x$ values, the number of successes (i.e. claims) are divided by the sample size (i.e. number of combinations of days and risk addresses
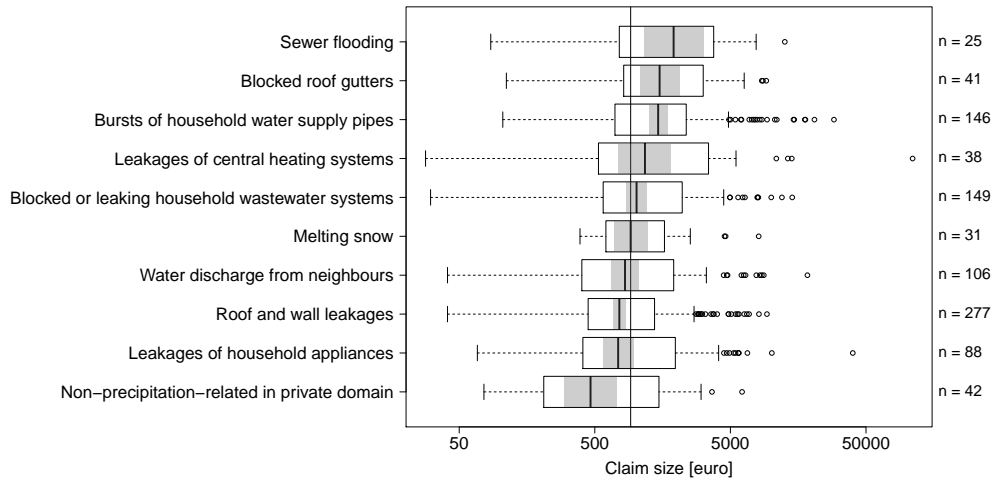
**Figure 4.** Distribution of claim sizes associated with various failure mechanisms. Claim size is the sum of property and content damage. Only risk addresses are included for which both property and content insurance were available. Results are only shown for failure mechanisms with at least 20 claim records. The grey rectangles display the 95 % confidence interval around the median. If the grey rectangles of two boxplots do not overlap, there is a strong indication that the median are statistically different. The vertical solid line represent the median claim size. The number of claims ($n$) within each class are given next to the boxplots.
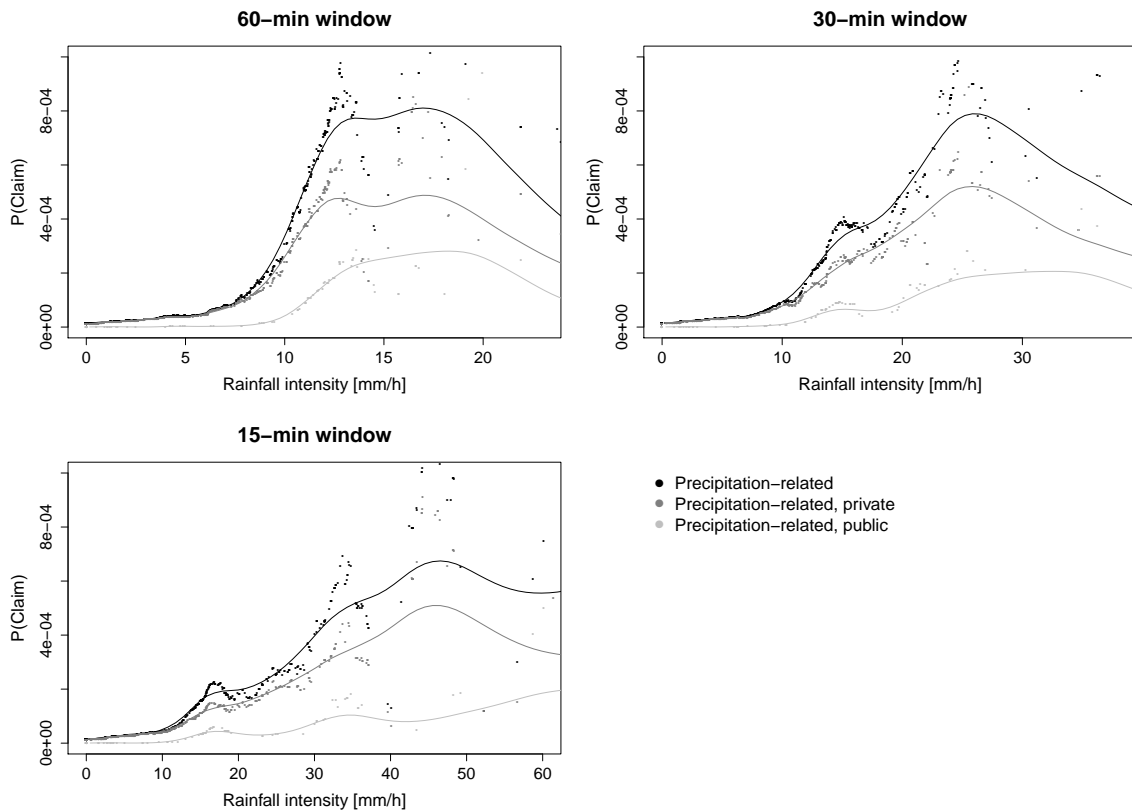


**Figure 5.** The empirical probability of precipitation-related claim occurrence per day per risk address as a function of rainfall intensity, using a 60-min (top left panel), 30-min (top right panel) and 15-min moving time window (bottom left panel). Results are related to precipitation-related claims (black dots), broken down to those classified as "private" (grey dots) and "public" (light grey dots). The range of $x$ values is from $0\,mm\,h^{-1}$ to the rainfall intensity associated with a return period of 5 years. Locally-weighted regression lines are based on penalised B-splines.

that can generate claims). Empirical probabilities are evaluated at each $x$ value that corresponds with a claim.

The top-left plot of Fig. 5, based on a 60-min window, shows that the occurrence probability of precipitation-related claims increases with increasing rainfall intensity and that it increases considerably when rainfall intensity exceeds 7–8 mm h$^{-1}$. For events with rainfall intensities smaller than 7–8 mm h$^{-1}$, the occurrence probability of precipitation-related claims is mainly determined by failure processes in the private domain, which are primarily roof and wall leakages. Thus, roof damage and wall leakage already occur at small rainfall intensities, which suggest that leaks may be latent before first observed during a rainfall event. For rainfall events that exceed the 7–8 mm h$^{-1}$ threshold, failure processes in the public domain start to contribute substantially to the overall occurrence probability. Similar conclusions can be drawn from the other two plots related to a 30-min (top-right) and a 15-min window (bottom-left), with the difference that rainfall threshold shift to 9–10 and 12 mm h$^{-1}$ respectively. The locally-weighted regression lines reveal that relationships using the 30-min and 15-min window have a less linear nature than the ones based on a 60-min window.

## 3.3 Logistic regression results

Logistic regression analyses were performed to test the significance of various combinations of explanatory variables in explaining the occurrence probability of precipitation-related claims. Separate analyses were made for the occurrence of claims caused by failures of systems in the public and private domain (according to column 5 in Table 3). From the three variants of maximum rainfall intensity, the one based on a 60-min window was used for modelling. Regression coefficients were estimated based on the data in the rainfall intensity range of 5 to 12 mm h$^{-1}$ (60-min window). Data associated with 12 mm h$^{-1}$ or larger are scarce and are, therefore, likely biased towards single rainfall events. In a first attempt to fit a logistic regression model to data in the range of 0–12 mm h$^{-1}$, it was found that much weight was given to the data in the range of 0–5 mm h$^{-1}$, resulting in a poor fit to data in the higher rainfall intensity range. Possibly, claims associated with rainfall intensities of 0–5 mm h$^{-1}$ are generated by a different process than the claims associated with rainfall intensities larger than 5 mm h$^{-1}$. More on this can be read in Sect. 4.

The goodness-of-fit measures of the various models, including a comparison of the likelihood ratio statistics between models, are summarised in Table 5. The models that combine maximum rainfall intensity and rainfall volume result in better fits compared to the intercept-only models. Maximum temperature, wind parameters and season significantly improve the model fit for claims caused by failures of private systems, but not for claims caused by failures or public systems. Most of the explanatory power derives from

maximum rainfall intensity and rainfall volume. Of all wind parameters, wind gust has best explanatory power.

Table 6 lists the estimates of the regression coefficients for the two models that include all explanatory variables (using wind gust as wind parameter), further referred to as the "private model" and the "public model". The categorical variable "season" was modelled as four separate binary variables, where one level was dropped to avoid multicollinearity. The summer season was found to positively correlate with the occurrence of claims related to failure of private systems. Moreover, regression analysis revealed that the regression coefficient of the maximum rainfall intensity is larger for the public model than for the private model, which means that rainfall intensity more strongly affects the claim occurrence by failures of public systems than private systems. The odds ratio ($\exp(\beta)$) related to maximum rainfall intensity varies between 1.16–1.27 for the private model and 1.33–1.95 for the public model, which means a 16–27 % and a 33–95 % increase in odds for each mm h$^{-1}$ change in rainfall intensity, for private and public models respectively.

## 4 Discussion

Based on the insurance data for the case study in Rotterdam, a distinct rainfall intensity threshold could be defined, above which failures of public systems start to contribute considerably to the occurrence of damage claims (Fig. 5). Interestingly, this threshold of 7–8 mm h$^{-1}$ (based on a 60-min window) is not in line with the design standards of sewers in the Netherlands. Dutch sewers are designed to cope with rainfall intensities of 20 mm h$^{-1}$, which is associated with an event return period of approximately 2 years (Van Mameren, 1997; Van Luijtelaar, 1997; Koot, 1977). This suggests that the threshold relates to some other damaging process than the simple overloading of sewer systems, for example, blockages in sewer pipes or malfunctioning of non-return valves in sewer laterals. On closer inspection of the communication transcripts of claims labelled as "sewer flooding", it was found that most cases relate to sewer backups from toilets or floor drains, and to a lesser extent to run-off entering buildings at ground level. Still, communication transcripts were inconclusive about the extent to which these claims were related to overloaded sewer systems or failure of system components.

Findings from the current body of work have implications for pluvial flood risk management. The return period of design storms as currently being used to design sewer systems in the Netherlands is largely based on political consensus. Potentially the results presented here can be used to obtain an objective design criterion based on risk assessment. Furthermore, this study provides insights into contributions of urban drainage systems to flood damage at city level. The results will support urban water managers in the evaluation of urban drainage system capacity and their decision-making

**Table 5.** Goodness-of-fit measures of logistic regression models.

| | | Claims related to failure of private systems | | | Claims related to failure of public systems | | |
|---|---|---|---|---|---|---|---|
| | Model | Likelihood ratio | d.f. | Pseudo-$R^2$ | Likelihood ratio | d.f. | Pseudo-$R^2$ |
| 1 | $Max_{60}$ | 134.97*** | 1 | 0.033 | 57.72*** | 1 | 0.094 |
| 2 | $Max_{60}$ + vol | 253.66*** | 2 | 0.063 | 112.68*** | 2 | 0.183 |
| 3 | $Max_{60}$ + vol + temp | 261.95*** | 3 | 0.065 | 114.24*** | 3 | 0.186 |
| 4 | $Max_{60}$ + vol + temp + seas + $wind_d$ | 276.11*** | 7 | 0.069 | 120.4*** | 7 | 0.196 |
| 5 | $Max_{60}$ + vol + temp + seas + $wind_h$ | 284.49*** | 7 | 0.071 | 121.01*** | 7 | 0.197 |
| 6 | $Max_{60}$ + vol + temp + seas + $wind_g$ | 302.71*** | 7 | 0.075 | 122.16*** | 7 | 0.199 |
| | Comparison models 2–1 | 118.69*** | 1 | | 54.96*** | 1 | |
| | Comparison models 3–2 | 8.29** | 1 | | 1.56 | 1 | |
| | Comparison models 4–3 | 14.16** | 4 | | 6.16 | 4 | |
| | Comparison models 5–3 | 22.54*** | 4 | | 6.77 | 4 | |
| | Comparison models 6–3 | 40.76*** | 4 | | 7.92 | 4 | |

\* $p$ value $< 0.05$, \*\* $p$ value $< 0.01$, \*\*\* $p$ value $< 0.001$
Note: because the *relogit* routine does not report goodness-of-fit statistics, statistics are based on the ordinary logistic regressions.

**Table 6.** Estimates of regression coefficients of the rare event logistic regression models.

| | Model related to failure of private systems | | Model related to failure of public systems | |
|---|---|---|---|---|
| | $\beta$ (SE) | $\exp(\beta)$ (95 % C.I.) | $\beta$ (SE) | $\exp(\beta)$ (95 % C.I.) |
| (Intercept) | −14.841*** (0.605) | | −22.263*** (3.504) | |
| Maximum rainfall intensity (60-min) | 0.192*** (0.046) | 1.21 (1.16–1.27) | 0.479* (0.191) | 1.61 (1.33–1.95) |
| Rainfall volume | 0.048*** (0.006) | 1.05 (1.04–1.06) | 0.053** (0.017) | 1.05 (1.04–1.07) |
| Maximum wind gust | 0.082*** (0.019) | 1.09 (1.06–1.11) | 0.154 (0.084) | 1.17 (1.07–1.27) |
| Maximum temperature | 0.083*** (0.019) | 1.09 (1.07–1.11) | 0.156 (0.080) | 1.17 (1.08–1.27) |
| Season: spring | 0.095 (0.265) | 1.10 (0.84–1.43) | 0.211 (0.890) | 1.23 (0.51–3.00) |
| Season: summer | 0.521* (0.242) | 1.68 (1.32–2.15) | 0.319 (0.825) | 1.38 (0.60–3.14) |
| Season: winter | 0.324 (0.251) | 1.38 (1.08–1.78) | 0.054 (0.976) | 1.05 (0.40–2.80) |
| Likelihood ratio $\chi^2$ | 302.71 | | 122.16 | |
| d.f. | 7 | | 7 | |
| $p$ value | 0.000 | | 0.000 | |
| Pseudo-$R^2$ | 0.075 | | 0.199 | |

\* $p$ value $< 0.05$, \*\* $p$ value $< 0.01$, \*\*\* $p$ value $< 0.001$
Note: standard error (SE) of estimate is given between brackets; the upper and lower bound of the 95 % confidence interval (C.I.) are $\exp(\beta \pm 1.96\,SE)$, assuming normality on the log odds scale.

about the need for and prioritisation of investment to increase drainage capacity. Further research is needed to explain why damage related drainage capacity occurs below the level of design capacity; this will help water managers to focus efforts on ensuring that their systems reach design capacity.

The results of this research have practical relevance for insurers. From this present case study, it became evident that the majority of the water-related claims are caused by roof and wall leakages. Thus, damage prevention programmes focusing on these causes may be helpful. When it is raining heavily ($> 7$–8 mm in a 60-min window) insurers can expect more claims related to sewer flooding that require special services for the cleaning of sewer spills.

In the higher range of rainfall intensities in Fig. 5, relationships between rainfall intensity and claim occurrence probability become less distinct, which can partly be ex-

plained by the limited amount of claim data associated with extreme rainfall events. The present insurance database covers almost 7 years of claim data (2007–October 2013), where around 80 % of the precipitation-related claims relate to rainfall events with return periods less than 2 years. Around 10 % of the claims can be attributed to two exceptional rainfall events with a return period of 14–18 years. As a consequence, empirical probabilities in the higher range of rainfall intensities are unreliable and biased towards single rainfall events.

Claims associated with rainfall intensities of 0–5 mm h$^{-1}$ in Fig. 5 (60-min window) are possibly generated by a different process than the claims related to rainfall intensities larger than 5 mm h$^{-1}$. It maybe the case that more specific damage processes can be distinguished within the existing cause classes. For example, the class "roof leakages" may contain two processes; one related to the presence of latent

leaks that are first observed when it is raining and another one related to the exceedance of the "hydraulic capacity" of roofs. The hypothesis could not be tested based on the present database, because it lacked information to distinguish between the sub-processes.

In the top panel of Fig. 3, an increasing trend is observed in the number of water-related claims in the period 2007–2013. There are a number of possible explanations for this trend. To start with, the number of policyholders may have increased in time. This could not be verified, because in the present study only policyholder data were available for a single snapshot in time. Another explanation may be related to burst household water supply pipes, which is the most frequent cause of non-precipitation-related claims. Based on an unpublished report, the insurance company has observed a substantial increase in defects in water supply systems in the recent years, mainly because of incorrectly installed compression fittings. Other explanations that may be worthwhile to investigate are differences in climate variables between years and the effect of the 2007–2008 financial crisis on people's claiming behaviour.

There are a number of aspects with regard to uncertainty in insurance data. The occurrence of claims that relate to causes that are not covered by insurers (e.g. groundwater flooding) are probably underestimated by the data, simply because people may be aware of the fact that damage is not covered and, thus, not make a claim. Moreover, the reported claim date may not always be the date on which the damage occurred, for example, because the exact damage data is unknown, which may be the case when people are on holidays. Furthermore, addresses of the insured are based on static policyholder information, i.e. a situation on a snapshot in time (reference date: 31 July 2013). Errors in addresses may occur if policyholder information has changed (e.g. policyholders moving to another address).

Failure of public systems (e.g. sewer system) will probably mostly affect buildings that occupy ground floor. In the present study, no distinction was made between terraced or detached houses and high-rise buildings (i.e. houses that occupy first floor or higher). As a consequence, claim occurrence probabilities related to the failure of public systems is likely to be higher than the probabilities estimated in the present study, which is based on all building types.

## 5 Conclusions and recommendations

The main goal of the current study was to investigate the relative contributions of different failure mechanisms to the occurrence of rainstorm damage to building structure and content, as well as the extent to which the probability of occurrence of these failure mechanisms relate to weather variables. For this purpose, a property level home insurance database of around 3100 water-related damage claims was analysed, for a case study in Rotterdam, the Netherlands.

The results of this investigation show that the leakage of roofs and walls is the most frequent failure mechanism causing precipitation-related claims, followed by blocked roof gutters, snow melting under roof tiles and sewer flooding. Although claims related to sewer flooding were less present in the data, they are associated with significantly larger claim sizes (EUR 1150–3160, 95 % confidence interval around the median) than claims generated by roof and wall leakages (EUR 680–840), the majority class. Rare events logistic regression analysis revealed that maximum rainfall intensity and rainfall volume are significant predictors for the occurrence probability of precipitation-related claims. Moreover, it was found that claims associated with rainfall intensities smaller than 7–8 mm in a 60-min window are mainly caused by failures of systems in the private domain, such as roof leakages and blocked roof gutters. For rainfall events that exceed the 7–8 mm h$^{-1}$ threshold, the failure of systems in the public domain, such as sewer systems, start to contribute considerably to the overall occurrence probability of claims. The communication transcripts, however, lacked conclusive information about the extent to which sewer-related claims were caused by the overloading of sewer systems or failure of system components.

It is worthwhile investigating the spatial distributions of water-related claim data in a future study, considering local conditions, such as building age and type and percentage of impervious area. An important limitation of this study is that the number of claims associated with extreme rainfall events was relatively small. Given the fact that manual classification took considerable amount of work, it is recommended to explore methods to automate and standardise classification of claim data, with the aim to process more data in future analyses.

Edited by: M.-C. Llasat
Reviewed by: two anonymous referees

## References

André, C., Monfort, D., Bouzit, M., and Vinchon, C.: Contribution of insurance data to cost assessment of coastal flood damage to residential buildings: insights gained from Johanna (2008) and Xynthia (2010) storm events, Nat. Hazards Earth Syst. Sci., 13, 2003–2012, doi:10.5194/nhess-13-2003-2013, 2013.

Busch, S.: Quantifying the risk of heavy rain: its contribution to damage in urban areas, in: Proceedings of the 11th International

Conference on Urban Drainage, Edinburgh, Scotland, UK, 31 August–5 September, 2008.

Cheng, C. S.: Climate change and heavy rainfall-related water damage insurance claims and losses in Ontario, Canada, J. Water Resour. Protect., 4, 49–62, doi:10.4236/jwarp.2012.42007, 2012.

Climate Service Center: Machbarkeitsstudie "Starkregenrisiko 2050", Tech. rep., available at: http://www.climate-service-center.de/ (last access: 1 May 2014), 2013.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography, 36, 27–46, doi:10.1111/j.1600-0587.2012.07348.x, 2013.

Douglas, I., Garvin, S., Lawson, N., Richards, J., Tippett, J., and White, I.: Urban pluvial flooding: a qualitative case study of cause, effect and nonstructural mitigation, J. Flood Risk Manage., 3, 112–125, doi:10.1111/j.1753-318X.2010.01061.x, 2010.

Einfalt, T., Pfeifer, S., and Burghoff, O.: Feasibility of deriving damage functions from radar measurements, in: 9th International Workshop on Precipitation in Urban Areas, St. Moritz, Switzerland, 6–9 December 2012, 245–249, 2012.

Elmer, F., Seifert, I., Kreibich, H., and Thieken, A. H.: A delphi method expert survey to derive standards for flood damage data collection, Risk Anal., 30, 107–124, doi:10.1111/j.1539-6924.2009.01325.x, 2010.

Gall, M., Borden, K. A., and Cutter, S. L.: When do losses count? Six fallacies of natural hazards loss data, B. Am. Meteorol. Soc., 90, 799–809, doi:10.1175/2008BAMS2721.1, 2009.

Garne, T. W., Ebeltoft, M., Kivisaari, E., and Moberg, S.: Weather related damage in the Nordic countries, Tech. rep., available at: http://www.fkl.fi/materiaalipankki/tutkimukset/Dokumentit/Weather_related_damage_in_the_Nordic_countries.pdf (last access: 1 May 2014), 2013.

Gemeente Rotterdam: Sewer Plan Rotterdam 2011–2015, Tech. rep., available at: http://www.rotterdam.nl/GW/Document/Waterloket/GRPrapport2011-2015juni2011.pdf (last access: 1 May 2014), 2011.

Gemeente Rotterdam: Open Data Centre Rotterdam: Shape file of Rotterdam's sewer sytem, available at: http://www.rotterdamopendata.nl/dataset/het-rioolsysteem-van-rotterdam (last access: 1 May 2014), 2014.

Hohl, R., Schiesser, H.-H., and Aller, D.: Hailfall: the relationship between radar-derived hail kinetic energy and hail damage to buildings, Atmos. Res., 63, 177–207, doi:10.1016/S0169-8095(02)00059-5, 2002.

Imai, K., King, G., and Lau, O.: relogit: rare events logistic regression for dichotomous dependent variables, in: Zelig: Everyone's Statistical Software, edited by: Imai, K., King, G., and Lau, O., available at: http://cran.r-project.org/web/packages/Zelig/index.html (last access: 1 May 2014), 2007.

Jak, M. and Kok, M.: A database of historical flood events in the Netherlands, in: Flood Issues in Contemporary Water Management, NATO Science Series 2, Environmental Security, Kluwer Academic Publisher, Delft, the Netherlands, 139–146, 2000.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., Gericke, A., Neal, J., Aerts, J. C. J. H., and Ward, P. J.: Comparative flood damage model assessment: towards a European approach, Nat. Hazards Earth Syst. Sci., 12, 3733–3752, doi:10.5194/nhess-12-3733-2012, 2012.

Kadaster: Online viewer of the National Building Register held by Kadaster, available at: http://bagviewer.pdok.nl/ (last access: 1 May 2014), 2013.

King, G. and Zeng, L.: Logistic regression in rare events data, Polit. Anal., 9, 137–163, 2001.

Koot, A. C. J.: Inzameling en Transport van Rioolwater, Waltman, Delft, the Netherlands, 1977.

Long, J. S.: Regression Models for Categorical and Limited Dependent Variables, Sage Publications, Thousand Oaks, 1997.

McCullagh, P. and Nelder, J.: Generalized Linear Models, 2nd Edn., Chapman and Hall, Chicago, Londen, 1989.

Ministry of Transport, Public Works and Water Management: Insurability of damages related to extreme rainfall and pluvial flooding, Tech. rep., RIZA, Lelystad, the Netherlands, 2003.

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

Smith, C. and Lawson, N.: Identifying extreme event climate thresholds for greater Manchester, UK: examining the past to prepare for the future, Meteorol. Appl., 19, 26–35, doi:10.1002/met.252, 2011.

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, 913–922, doi:10.5194/hess-17-913-2013, 2013.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: A spatial analysis of rainfall damage data using C-band weather radar images, in: Proceedings of the International Conference on Flood Resilience: International Conference on Flood Resilience (ICFR): Experiences in Asia and Europe, Exeter, UK, 5–7 September, 2013.

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: Decision-tree analysis of factors influencing rainfall-related building damage, Nat. Hazards Earth Syst. Sci., 14, 2531–2547, doi:10.5194/nhess-14-2531-2014, 2014.

Statistics Netherlands: Demographic statistics by municipality 2013, available at: http://www.cbs.nl/nl-NL/menu/themas/bevolking/publicaties/publicaties/archief/2013/2013-b55-pub.htm (last access: 1 May 2014), 2013.

Statistics Netherlands: StatLine online database, available at: http://statline.cbs.nl, last access: July, 2014.

Ten Veldhuis, J. A. E.: How the choice of flood damage metrics influences urban flood risk assessment, J. Flood Risk Manage., 4, 281–287, doi:10.1111/j.1753-318X.2011.01112.x, 2011.

Van Luijtelaar, H. and Rebergen, E.: Guidelines for hydrodynamic calculations on urban drainage in the Netherlands: Backgrounds and examples, Water Sci. Technol., 36, 253–258, doi:10.1016/S0273-1223(97)00590-8, 1997.

Van Mameren, H. and Clemens, F. H. L. R.: Guidelines for hydrodynamic calculations on urban drainage in the Netherlands: Overview and principles, Water Sci. Technol., 36, 247–252, doi:10.1016/S0273-1223(97)00591-X, 1997.

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, doi:10.2166/wst.2013.268, 2013.