International Conference on Hydroinformatics

2014

Decision Tree Analysis Of Processes Generating Water-Related Building Damage: A Case Study In Rotterdam, The Netherlands

Matthieu Spekkers

Francois Clemens

Marie-Claire Ten Veldhuis

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

DECISION TREE ANALYSIS OF PROCESSES GENERATING WATER-RELATED BUILDING DAMAGE: A CASE STUDY IN ROTTERDAM, THE NETHERLANDS

SPEKKERS, M. H. (1), CLEMENS, F. H. L. R. (1), TEN VELDHUIS, J. A. E. (1)
(1): Delft University of Technology, P.O. Box 5048, 2600 GA, Delft, the Netherlands

The objective of this study was to identify the main failure mechanisms behind water-related building damage and to investigate to what extent these processes are related to characteristics of buildings and rainfall events. Results are based on the mining of property level insurance damage data, for a case study in Rotterdam, the Netherlands. This study has found that most frequent causes of water-related damage relate to roof leakages (28%), bursts of household water supply pipes (19%) and blocked household wastewater systems (18%). Cases of sewer flooding or depression filling were less present (2.4% and 0.6%), but showed stronger correlations with heavy rainfall events than any other failure mechanism. Classification tree analysis revealed that water discharges from neighbours is the main damage cause for high-rise buildings on days with no or minor rainfall (< 7.5 mm/h). Moreover, damage due to blocked household wastewater systems is associated with low-rise buildings younger than 50 years.

INTRODUCTION

A proper understanding of the damaging processes associated with weather events is an essential step in the development of weather-related damage prediction models. In this context, damage claim data from insurance companies can be used for data mining to study explanatory factors for damage [1–5]; however, claim data containing reliable, high-resolution information on the actual causes of damage are scarce in scientific literature, which is limiting data analyses both in terms of quality and level of detail.

This paper presents an application of data-driven analysis in water sciences. The objective of this study was to identify the main failure mechanisms behind water-related building damage and to investigate to what extent these processes are related to characteristics of buildings and rainfall events. For this, a property level damage database of around 2900 water-related records was mined, for a case study in Rotterdam, the Netherlands. Records include comprehensive transcripts of communication between insurer, insured and damage experts, which allowed detailed classification of claims according to the actual cause of damage. Preliminary results of a classification tree analysis are presented, where claims are classified into cause classes based on characteristics of buildings and rainfall events.

METHODOLOGY

The study focuses on Rotterdam, which is, with a population of around 620000, the second largest city of the Netherlands [6]. Because the city is relatively flat (maximum ground level variations of 10–15 meter), floods from heavy rainfall are typically characterized by small flood depths (up to a few decimetres) and limited surface run-off. Rotterdam's sewers are mainly combined systems (≈ 1800 km), some parts of the city have separated systems for wastewater and stormwater (≈ 700 km) [7]. Stormwater is transported to a branch of the river Rhine, which meanders through the city centre, or to wastewater treatment plants. In Rotterdam, water storage facilities have been built to temporarily store rainwater during extreme rainfall events, i.e., an underground storage facility (capacity of 10 million litres) and water retention squares. The majority of the buildings in Rotterdam was constructed in the 20th century, with the oldest buildings dating back to late 19th century. Rotterdam's urban fabric is characterised by a combination of terraced houses and high-rise residential and commercial buildings [8].

Insurance damage data were provided by a Dutch insurance company that is part of the Achmea insurance group. Records are available for the period of January 2007–June 2013, containing around 2900 water-related claims about damaged residential property and content. Claims are related to the individual property level. For each case, the following information is available: a full address, a date on which damage occurred, the claim size paid out by the insurer (not used in present study) and detailed transcripts of communication (e.g., calls, reports) between insurer, insured and damage experts. The transcripts typically describe the cause of damage, the goods and materials that were damaged and the costs related to cleaning, drying, repairing or replacing goods and materials. On average, the database contains information of around 16500 policyholders, which is 6% of the total number of households in Rotterdam. This number relates to one insurance company of the Achmea insurance group and does not reflect the market share in Rotterdam of the Achmea insurance group as a whole.

Cases were manually classified according to the cause of damage using the available information in the communication transcripts. For this purpose, an easy-to-use web interface was built based on the classification scheme listed in Table 1. Two main groups of damage causes can be distinguished: a group containing precipitation-related causes and a non-precipitation-related group. A set of explanatory variables was added to each case (Table 2): rainfall volume and maximum rainfall intensity were extracted from weather radar data from the Royal Netherlands Meteorological Institute, following an approach discussed in Spekkers *et al.* [5]. Building type and building age were derived from National Building Register [8]. Two terrain indices, the topographic position index and the steepness, were calculated using a 5 m x 5 m digital terrain model (DTM) of the Netherlands [9]. More background on the data sources used can be found in Spekkers *et al.* [5]

A classification tree model was used to investigate relationships between damage causes and explanatory variables. The philosophy of this approach is to learn a tree by recursively splitting data into two groups, such that the proportion of misclassified cases is minimized [10]. It should be noted that results of tree analysis were not cross-validated because of the relative small sample size. As a consequence, the tree describes structures in the data; the predictive power of the tree approach has still to be researched.

Table 1. Classification scheme for water-related damage causes applicable to residential buildings.

Id	Short name	Description
Precipitation-related		
1	Blocked roof gutters	Overflowing of roof gutters due to blockages (e.g., by leafs or ice)
2	Inflow route interruption	Flood water entering buildings as a result of inflow route interruption (e.g., blocked sewer inlet)
3	Depression filling	Flood water entering buildings as a result of depression filling, i.e., rainwater filling up depressions if no drainage facilities are available
4	Sewer flooding	Flood water entering buildings as a result of sewer flooding
5	Groundwater flooding	Groundwater flooding due to persistent rainfall
6	Melting snow	Intrusion of melting snow and ice
7	Rainwater through open window	Rainwater intrusion through open windows, open doors
8	Roof leakages	Rainwater intrusion through roofs, facades, walls, wall-window interfaces
9	Precipitation-related	Precipitation-related, but actual cause not specified
Non-precipitation-related		
10	Blocked public wastewater system	Flooding of wastewater due to blockage in public sewer system
11	External water discharges	External water discharges (e.g., extracted groundwater from construction site, fire extinguishing water)
12	Bursts of public water supply pipes	Bursts of public water supply pipes
13	Blocked household wastewater system	Flooding of wastewater due to blockage in household wastewater system
14	Leakages of central heating systems	Leakages of central heating systems
15	Bursts of household water supply pipes	Bursts of household water supply pipes, including attached facilities
16	Leakages of household appliances	Leakages of household appliances (e.g., washing machines, aquaria, waterbeds, coffee machines, sprinkler systems)
17	Non-precipitation-related	Non-precipitation-related, but actual cause not specified
Others		
18	Water-related	Water-related, but actual cause not specified
19	Water discharge from neighbours	Water discharge from neighbours, but actual cause not specified

Table 2. Definitions of explanatory variables.

Variable name	Definition	Min – Max
Rainfall volume (vol)	Volume of rainfall event at the radar pixel intersecting the building's centroid (mm). See for more details: Spekkers et al. [5].	0 – 73
Maximum rainfall intensity (max)	Maximum intensity of rainfall event at the radar pixel intersecting the building's centroid, using a 1 h moving time window (mm/h). See for more details: Spekkers et al. [5].	0 – 33
Building type (type)	1) Low-rise = a terraced or (semi-)detached house or 2) high-rise = a house that occupies only part of a building and that has its entrance at first floor or higher	NA
Building age (age)	Building age in years	0 – 153
Position index, 25 m (tpi1)	The elevation of a cell of the DTM at the building's centroid relative to the mean elevation of a 25 m x 25 m window around that cell (m)	-2.3 – 1.4
Position index, 255 m (tpi2)	The elevation of a cell of the DTM at the building's centroid relative to the mean elevation of a 255 m x 255 m window around that cell (m)	-4.2 – 3.8
Position index, 1005 m (tpi3)	The elevation of a cell of the DTM at the building's centroid relative to the mean elevation of a 1005 m x 1005 m window around that cell (m)	-3.8 – 5.6
Slope (slo)	The maximum rate of change in value from a cell of the DTM at the building's centroid to its neighbouring cells (◦)	0 – 27
Season (seas)	Season of the year: winter = December–February, spring = March–May, summer = June–August, autumn = September-November	NA

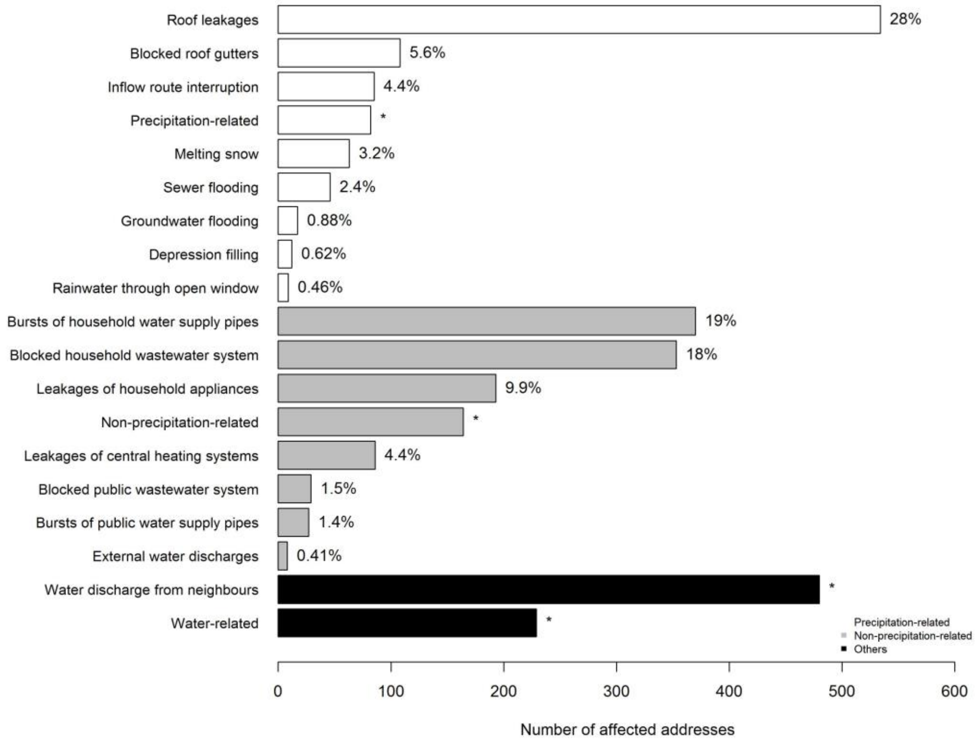


Figure 1. Causes of water-related building damage based on around 2900 insurance claims, for the period of January 2007–June 2013, for a case study in Rotterdam, the Netherlands. Percentages next to the bars are relative to the total number of claims for which the actual cause of damage were specified. Groups denoted with an asterisk next to the bar contain claims for which the actual cause of damage could not be specified.

RESULTS AND DISCUSSION

Explorative analyses show that the most common causes of water-related property damage are roof leakages (28% of the cases that were successfully classified, see Fig. 1), followed by bursts of household water supply pipes (19%) and blocked household wastewater systems (18%). Other frequent damage causes related to precipitation include blocked roof gutters (5.6%) and inflow route interruption (4.4%). Cases of sewer flooding or depression filling were less present (2.4% and 0.6%), but showed, compared to any other damage cause, stronger correlations with heavy rainfall events (Fig. 2). In contrast, cases related to roof leakages already occur for relative small maximum rainfall intensities and rainfall volumes (Fig. 2).

The classification tree in Fig. 3 relates the damage cause to rainfall-related and building-related characteristics. A tree size of five (i.e., number of terminal nodes) was selected. Larger trees resulted in only little improvement of the overall model performance. The tree uses four explanatory to classify cases: building type, building age, rainfall volume and maximum rainfall intensity. The two values at the terminal nodes are the fraction of correctly classified cases in that group and the percentage of cases that were classified into that group by the tree respectively. With regard to overall performance, the tree classifies 26% of the cases correctly, which is 8% better than a classification that is based on the “go with the majority” rule (i.e., all cases are classified as “roof leakage”). Although the overall classification performance is low,

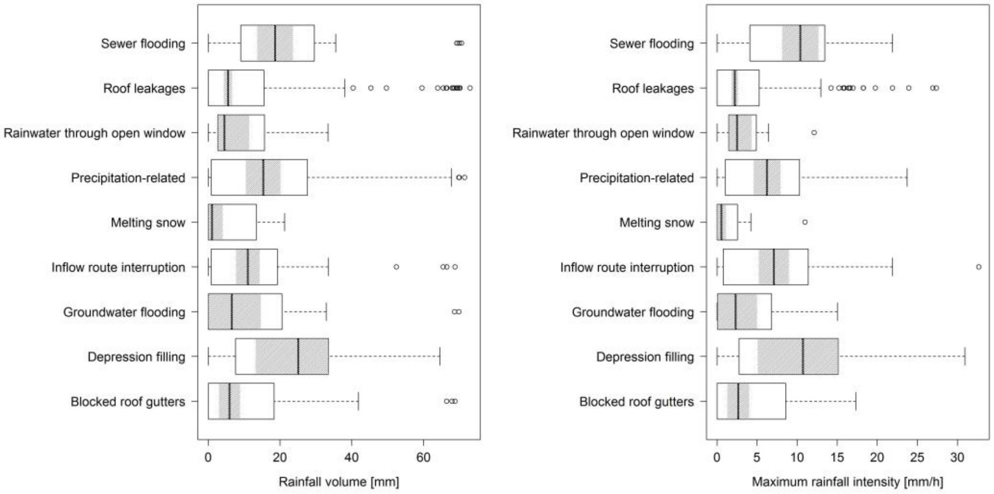
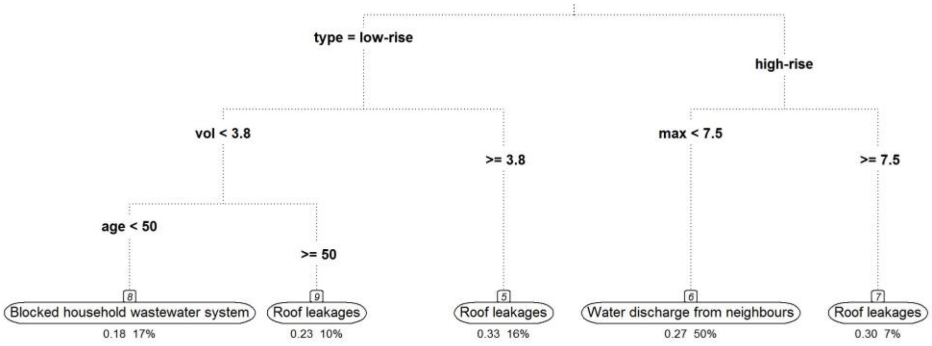


Figure 2. Boxplots of rainfall volume (left) and maximum rainfall intensity (right) associated with precipitation-related claims, per damage class. The grey rectangles display the 95%-confidence interval around the median. If the grey rectangles of two boxplots do not overlap, then there is a strong indication that the medians are statistically different.



Correctly classified by tree: 26%
Correctly classified by 'go with the majority rule': 18%
Correctly classified by 'blind guessing': 5%

Figure 3. Classification tree relating the damage cause to building type (type), building age (age), rainfall volume (vol) and maximum rainfall intensity (max). The two values at the terminal nodes are the fraction of correctly classified cases in that group and the percentage of cases that were classified into that group by the tree respectively.

the tree reveals that water discharges from neighbours is a dominant damaging process related to high-rise buildings, when there is no or minor rainfall (< 7.5 mm/h). This can be explained by the fact that high-rise buildings contain many interfaces between neighbours. Moreover, cases of blocked household wastewater system are relatively often found in buildings younger than 50 years. On rainy days (vol $\geq$ 3.8 mm or max $\geq$ 7.5 mm/h), roof leakages are the prevailing cause for water-related damage.

Results of this study are sensitive to a number of aspects. For instance, the study period is relatively short (6.5 years). As a consequence, the fraction of rainfall-related cases in the data set are sensitive to the magnitude and number of events in the rainfall time series. Another aspect that requires more study is the fact that the data are unbalanced. For example, the group "roof leakages" is a significant class in the data set. As a result, tree splits may be biased towards this class.

CONCLUSIONS AND FUTURE WORK

The objective of this study was to identify the main failure mechanisms behind water-related building damage and their relationships with building-related and rainfall-related variables. Based on a case study in Rotterdam, the Netherlands, analysis of property insurance data identified the three main causes of water-related damage to be roof leakages, bursts of household water supply pipes and blocked household wastewater systems. Classification tree analysis revealed that claims related to blocked household wastewater systems often occur for low-rise buildings younger than 50 years and that water discharge from neighbours is the main damage cause for high-rise buildings when there is no or minor rainfall. On rainy days, roof leakages are the prevailing cause for water-related damage. Still, tree analysis results are preliminary and need to be validated against independent data in a future study, for example, another case study.

Future steps for analysis will include an investigation of factors influencing the claim size and occurrence probability of failure mechanisms. Moreover, the database will be updated with claim data from a severe rainfall event (65–75 mm/day, associated with a return period of more than 30 years) that triggered relative many claims (100+). This event occurred outside the study period of the present paper and will be analysed in more detail in a future study.

ACKNOWLEDGEMENTS

The authors would like to thank the Achmea insurance group, Royal Netherlands Meteorological Institute and TU Delft Maps Room for their support and making available the data. Martijn Koole and Emiel Verstegen, MSc students at the TU Delft, are acknowledged for their help with the classification of claim data.

REFERENCES

[1] Botzen W. J. W., Bouwer L. M. and Van den Bergh J. C. J. M., "Climate change and hailstorm damage: Empirical evidence and implications for agriculture and insurance", *Resource and Energy Economics*, 32(3):341–362, (2010).

[2] Zhou Q., Panduro T. E., Thorsen B. J. and Arnbjerg-Nielsen K., "Verification of flood damage modelling using insurance data", *Water Science and Technology*, 68(2):425–32, (2013).

[3] Einfalt T., Pfeifer S. and Burghoff O., “Feasibility of deriving damage functions from radar measurements”, In *9th International Workshop on Precipitation in Urban Areas*, pp 245–249, St. Moritz (Switzerland), (2012).

[4] Schuster S. S., Blong R. J., Leigh R. J. and McAneney K. J., “Characteristics of the 14 April 1999 Sydney hailstorm based on ground observations, weather radar, insurance data and emergency calls”, *Natural Hazards and Earth System Science*, 5(5):613–620, (2005).

[5] 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”, *Natural Hazards and Earth System Sciences Discussions*, 2(4):2263–2305, (2014).

[6] Statistics Netherlands (2014). StatLine online database: http://statline.cbs.nl (viewed on April 2014).

[7] Municipality of Rotterdam, Municipal Sewer Plan 2011–2015, Technical report (in Dutch), http://www.rotterdam.nl/GW/Document/Waterloket/GRP%20rapport%202011-2015%20juni2011.pdf , (2011).

[8] Kadaster, Online viewer of the National Building Register, http://bagviewer.pdok.nl/ , (2014).

[9] AHN, Online viewer of the height map of the Netherlands, http://ahn.geodan.nl/ahn/ , (2014).

[10] Breiman L., Friedman J., Olshen R. and Stone C, “*Classification and regression trees*”, ISBN: 0-412-04841-8, Wadsworth, Belmont, California, (1984).