Hydrol. Earth Syst. Sci. Discuss., 9, 11615–11640, 2012
www.hydrol-earth-syst-sci-discuss.net/9/11615/2012/
doi:10.5194/hessd-9-11615-2012

Hydrology and Earth System Sciences Discussions

A statistical analysis of insurance damage claims related to rainfall extremes

M. H. Spekkers[1], M. Kok[2], F. H. L. R. Clemens[1], and J. A. E. ten Veldhuis[1]

[1]Delft University of Technology, Department of Water Management, Delft, The Netherlands
[2]Delft University of Technology, Department of Hydraulic Engineering, Delft, The Netherlands

Received: 24 September 2012 – Accepted: 27 September 2012 – Published: 11 October 2012

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In this paper, a database of water-related insurance damage claims related to private properties and content was analysed. The aim was to investigate whether high numbers of damage claims were associated with high rainfall intensities. Rainfall data were used for the period of 2003–2010 in the Netherlands based on a network of 33 automatic rain gauges operated by the Royal Netherlands Meteorological Institute. Insurance damage data were aggregated to areas within 10-km range of the rain gauges. Through a logistic regression model, high claim numbers were linked to maximum rainfall intensities, with rainfall intensity based on 10-min to 4-h time windows. Rainfall intensity proved to be a significant damage predictor; however, the explained variance, approximated by a pseudo-R^2 statistic, was at most 34 % for property damage and at most 30 % for content damage. When directly comparing predicted and observed values, the model was able to predict 5–17 % more cases correctly compared to a random prediction. No important differences were found between relations with property and content damage data. A considerable fraction of the variance is left unexplained, which emphasizes the need to study damage generating mechanisms and additional explanatory variables.

1 Introduction

In the autumn of 1998 extreme rainfall caused around 410 million Euros (1998 value) of direct damages to households, agriculture and industries in the Netherlands. Damage experts from the Dutch insurance sector identified a total number of 10 660 agricultural companies, 2470 buildings, 1220 other companies and 350 governmental agencies as being damaged by rainwater (Jak and Kok, 2000). The rainfall event with an associated return period of about 125 yr resulted in flooding of areas before rainwater was able to enter natural or engineered drainage systems. This type of floods is commonly known as pluvial flooding (e.g. Hurford et al., 2012; Blanc et al., 2012; Falconer et al., 2009).

HESSD

9, 11615–11640, 2012

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Other severe events that are well-documented are the summer floods of 2007 across the UK, for example in the City of Hull, that are believed to be for a great deal related to pluvial flooding (Pitt, 2008; Coulthard and Frostick, 2010), and the 2004 and 2006 floods in Heywood, Greater Manchester (Douglas et al., 2010). These events are just a few of the many examples that illustrate the serious consequences of high-intensity rainfall. But also minor events with relative small flood volumes and extensions can produce considerable damage in the long run due to its high frequency of occurrence (Freni et al., 2010; Ten Veldhuis, 2011). The aforementioned events have demonstrated that pluvial floods often occur at much smaller ranges of spatial and temporal scales than fluvial and coastal floods.

An increasing number of authors have acknowledged that a lack of data availability and quality have been important limitations in quantitative flood damage estimations (e.g. Freni et al., 2010; Merz et al., 2004; Hurford et al., 2011). In the absence of damage data, a common approach in flood damage estimation is to combine simulated flood depths and/or flow velocities and stage-damage curves (e.g. Ernst et al., 2008; Jonkman et al., 2008; Pistrika and Jonkman, 2009; De Moel and Aerts, 2011; Middelmann-Fernandes, 2010). The stage-damage curves are usually related to direct damages occurring in large catchments and are derived through synthetic and/or empirical approaches. Only few studies have focused on modelling damages of pluvial floods related to the dysfunctioning of urban drainage systems (e.g. Zhou et al., 2012a).

A promising source for flood damage data are insurance databases. These databases often contain many claim records that have been collected continuously in time. Disadvantages are the restricted access and the limited recordings of process information, such as flood depth and extent measurements, details on damage causes, and building information (Elmer et al., 2010; Thieken, 2011; Zhou et al., 2012b).

A few recent studies have analysed insurance data related to pluvial floods. Freni et al. (2010) conducted a damage assessment based on the outcomes of a simple and a detailed hydrodynamic model in combination with stage-damage functions derived

HESSD
9, 11615–11640, 2012

from around 600 insurance damage claims and water depth measurements for a case study in Palermo, Italy. They concluded that uncertainty in stage-damage function (40–50 % of average value) was higher than the accuracy gained by adopting a detailed hydrodynamic model. In another study, 1000 insurance damage claims related to sewer surcharging for the case of Aarhus, Denmark, showed that costs per claim were not explained by rainfall (Zhou et al., 2012b). They did find a significant relationship between rainfall and total costs per day. These studies confirmed the need to obtain accurate damage data to further investigate costs of pluvial floods.

In this study, data from an insurance database containing 20 yr of water-related claims for private properties and contents in the Netherlands, provided by the Dutch Association of Insurers, was analysed. The analysis built on earlier work by the Dutch Association of Insurers, where relationships between rainfall and claim data were studied at a regional scale (Ririassa and Hoen, 2010). The aim of this study was to investigate whether high numbers of damage claims are associated with high rainfall intensities, considering rainfall at scales most closely related to functioning of urban drainage systems. Separate relationships were analysed between rainfall data and property damage data as well as content damage data, through statistical analysis. A better understanding of relationships between rainfall extremes and floods is useful in the development of, for example, warning systems for pluvial floods (Hurford et al., 2012; Parker et al., 2011; Priest et al., 2011).

This paper is structured as follows. In Sect. 2 data sources as well as the statistical model to link rainfall and insurance damage data are described. Results of the statistical analysis are discussed in Sect. 3, as well as the significancy of predictor variables and the model performance. Conclusions and recommendation are summarized in Sect. 4.

2 Methodology

This section first describes the data sources. Next, the statistical approach that was used to link events with high claim numbers to rainfall characteristics is explained. All data-mining procedures and statistical computing are carried out within R 2.15.0 software environment.

2.1 Rainfall data

Rainfall data are based on two networks of rain gauges held by the Royal Netherlands Meteorological Institute (KNMI); a network of 300+ manual rain gauges (see Fig. 1, triangular markers) and a network of 33 automatic rain gauges (solid circles). The temporal resolution of the automatic network is 10 min and the spatial density is about 1 station every 1000 km^2 (see also Table 1), with most of the rain gauges located in rural areas or close to city boundaries. The manual network measures daily volumes based on 08:00 UTC–08:00 UTC intervals. The spatial density of the manual network is about 1 station every 100 km^2. All gauge data have been extensively validated by KNMI (KNMI, 2000).

2.2 Insurance data

The insurance databases cover water-related damages to private properties and content in the Netherlands and are summarized in Table 1. Data related to property and content damage is available from 1986 until 2010 and from 1992 until 2010 respectively. The database consists of data from a number of large insurance companies in the Netherlands, covering about 20–30 % of the Dutch market related to property and content policies.

House owners can insure both property and content; tenants can only insure content, while the rented property is considered a commercial building. Commercial buildings

HESSD
9, 11615–11640, 2012

are covered in a separate database that is not used in this study. More details and key characteristics of the databases can be found in Spekkers et al. (2011).

Water-related damages can be divided into two groups: (1) non-rainfall-related damages and (2) rainfall-related damages. Examples in the first group are bursts of water supply pipes and leakages of washing machines. Examples in the second group are leakages of roofs and flooding from sewer systems or regional watercourses. This distinction is not explicitly made in the data provided by insurance companies. Insurance companies use different systems to classify claims and the quality with which claims are assigned to groups varies between companies.

Damage due to pluvial flooding is included in most of the insurance policies after 2000 following an advice issued by the Dutch Association of Insurers (Ministry of Transport, Public Works and Water Management, 2003). Damage due to pluvial floods should be directly and solely related to local extreme rainfall for a claim to be accepted. Rainfall is considered "extreme" when rainfall intensity is higher than 40 mm in 24 h, 53 mm in 48 h or 67 mm in 72 h close to (not further defined) where damage occurred. The intensities are associated with occurrence frequencies of once every 3 to 7 yr in the Netherlands. It is unclear how and to what extent fulfilment of this requirement is examined by the insurance companies. Upon further inquiry, companies have indicated that detailed rainfall data to examine individual cases of local rainfall is usually lacking.

The insurance database consists of four sub databases: (1) a damage claim database with records related to property, (2) a damage claim database with records related to building content, (3) a database with policy holder information related to property insurances and (4) a database with policy holder information related to content insurances. The variables in the databases include, among others, the damage value that has been paid out by the insurance company, the date on which the damage occurred and the address of the insured household at 4-position district (i.e. neighbourhood) level. Typical surface areas of districts are 1–5 km^2 for urban areas and 10–50 km^2 for rural areas. Recorded damages include the costs of cleaning, drying and replacing materials and objects and the costs of temporarily rehousing of people.

HESSD
9, 11615–11640, 2012

For the analysis in this paper, it is assumed that the number of insurance policies is constant during one year. In case an insurance policy is only active for a part of the year, the insurance policy is counted proportionally for that year. Duplicate records were removed, as well as records with missing or incorrect date, location or damage value (around 6 % of the original database). Records with damage value equal to zero were also removed, as these are damage claims that did not meet the policy conditions. First and last day of the month were excluded as they, in a few cases, were showing unrealistically high claims numbers compared to other days. These days are probably due to software defaults when exact damage date was unknown or not entered by the insurer's employee.

2.3 Aggregating rainfall and insurance data

This study covers data from April 2003 to 2010. Insurance damage data were selected within a 10-km radius from the automatic rain gauges based on the distance between the district's centroid and its nearest automatic rain gauge (version shapefile of districts: March 2011). It is assumed that rainfall measured at the rain gauges are uniformly distributed in the rain gauge area. Rain gauge data are generally assumed to be representative within a range of several kilometers. Several ranges were tried and a 10-km range proved to be the best compromise between distance from rain gauges and number of data covered: within the 6-km range few data were present while from 10 km outward the effect of spatial rainfall variability was deemed too large to still consider the data spatially representative. Figure 2 shows two rain gauges and their neighbouring districts. Insurance data were converted to count data: the number of water-related claims k_i and number of insured households K_i were aggregated by day and by rain gauge area. The subscript i denotes the index of the observation. The number of insured households per rain gauge area ranges from around 300 to 55 000 for property insurance and from around 300 to 120 000 for content insurance. The higher number of content insurances is explained by the fact that property insurance only concerns house owners, whereas content insurance concerns both house owners and tenants.

HESSD
9, 11615–11640, 2012

Observations with less than 5000 households were filtered out as they were found very sensitive to errors in data. The maximum rainfall intensity $I_{i,z}$ is determined for each day and rain gauge area, where subscript z denotes the length in minutes of the moving time window, for z values 10 (original data), 20, 30, 40, 50, 60, 70, 80, 90, 120, 180 or 240 min.

2.4 Distinguishing rainfall-related and non-rainfall-related events

The distinction between non-rainfall-related and rainfall-related claims is not explicitly made in the data provided by insurance companies. Non-rainfall-related claims occur throughout the year, whereas rainfall-related claims are clustered on wet days. Consequently, high claim numbers are more likely to be associated with to rainfall. In the remainder of this paper, these observations are defined as "damage events".

The number of claims that can be expected on dry days was estimated based on claims recorded on dry days in 10-km ranges from the network of 300+ manual rain gauges, in order to obtain an independent estimate of the data associated with gauges in the automatic network. Observations were only selected in case of two subsequent dry days, because the daily volumes recorded by manual gauges are based on 08:00 UTC–08:00 UTC intervals. It was found that the number of non-rainfall-related claims is well described as a binomially distributed random variable:

$$k_i \sim B(K_i, \zeta), \tag{1}$$

where K_i is the number of insured households and ζ the probability that an individual insured household will have a non-rainfall-related claim on a day. It is assumed that ζ is constant in both time and space. Best fits with data were found for $\zeta = 3.2 \times 10^{-5}$ (property data) and $\zeta = 1.3 \times 10^{-5}$ (content data). The probability of obtaining y claims at least as extreme as k_i, the one observed, given the number of insured households K_i (i.e. p-value) is therefore:

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

HESSD

9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

$$\Pr(y \geq k_i \mid K_i) = 1 - \sum_{y=0}^{k_i-1} \binom{K_i}{y} \zeta^y (1-\zeta)^{K_i - y}. \quad (2)$$

Any p-value below a significance level α indicates occurrence of a damage event, as it is unlikely to be associated with non-rainfall-related claims. Different levels of significancy ($\alpha = 1 \times 10^{-2}$, 1×10^{-3}, 1×10^{-4} and 1×10^{-5}) are used to study its effect on the results. A binary variable Y_i is introduced to classify the observations that are considered a damage event $Y_i = 1$ and those that are not $Y_i = 0$:

$$Y_i = \begin{cases} 1 & \text{if } p\text{-value} < \alpha \\ 0 & \text{if } p\text{-value} \geq \alpha. \end{cases} \quad (3)$$

2.5 Linking binary outcome to maximum rainfall intensity

The binary outcome (damage event or not) is linked to the maximum rainfall intensity using a logistic regression model that is well-suited to analyse binary data (McCullagh and Nelder, 1989). The logistic function yields:

$$\operatorname{logit}(\theta_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \beta_1 I_{z,i}, \quad (4)$$

where θ_i is the probability of a damage event ($Y_i = 1$) and β_0, β_1 are regression coefficients. The regression coefficients are estimated using maximum likelihood estimation. The likelihood ratio (LR) test is used to test if β_1 is significantly different from zero; thus, if maximum rainfall intensity is a parameter that contributes to high numbers of damage claims. There is no universally-accepted goodness-of-fit measure in logistic regression that represents the proportion of variance explained by the predictors, such as R^2 in ordinary least squares regression. Several pseudo-R^2 statistics have been developed that mimic the R^2 in evaluating the variability explained, which is one of the

HESSD
9, 11615–11640, 2012

approaches used in this study. In this study McFadden's-R^2 is used, which compares the log-likelihood of the model without predictor and log-likelihood of the model with predictor (Long, 1997, p. 104). The other approach directly compares observed and predicted values from the fitted model using contingency tables, using a cutoff point of $\theta = 0.5$.

3 Results and discussion

3.1 Logistic regression results

In Table 2 the results of the logistic regression are summarized. Results are based on the 60-min rainfall intensity. The significance levels α, used for the dichotomization of damage data, range from 1×10^{-2} to 1×10^{-5}. Table 2 lists estimates for slope coefficient β_1, since this is the most important parameter for interpretation of logistic regression results. The standard error in β_1 is denoted as SE. The slope coefficient is expressed in exponential form, $\exp(\beta_1)$, which is the odds ratio. The odd ratio should be interpreted as the factor with which the odds (probability of a damage event divided by probability of no damage) change as an effect one unit change in the maximum rainfall intensity. For a large number of observations, $\mathrm{LR} \sim \chi^2$ with degrees of freedom equal to the number of parameters being estimated.

The slope coefficient is significantly different from zero in all cases (at $p < 0.05$ level), which means the maximum rainfall intensity is a significant predictor for damage. The odd ratios ($\exp(\beta_1)$) vary between 1.28–1.35 for property damage and 1.26–1.30 for content damage, indicating a 28–35 % (property) and 26–30 % (content) increase in odds of a damage event for each $\mathrm{mm\,h^{-1}}$ change in rainfall intensity. Different time windows between 10 min and 4 h have been investigated and produce similar results.

In Fig. 3 four examples of logistic functions are plotted as well as the data on which models were fitted. The plots are related to cases of property damage (with the dichotomization based on $\alpha = 1 \times 10^{-3}$) and 10-, 20-, 30- and 90-min rainfall intensities.

HESSD
9, 11615–11640, 2012

The function links the probability of a damage event θ on the y-axis to maximum rainfall intensity I_z on the x-axis. The steepness of the slope of the logistic function is determined by β_1 (see also Table 2); a large slope coefficient makes the transition between “low damage” and “damage event” more abrupt. The grey dots are the observations, either $Y = 0$ in case of “low damage” or $Y = 1$ in case of a “damage event”. A jitter function was applied to better visualize the density of the data points. The open circles are the calculated empirical proportions (number of observed $Y = 1$ in a bin divided by total number of observations in a bin n) for eight non-overlapping equally-sized bins. The error bars represent one standard deviation σ of uncertainty, where $\sigma = \sqrt{\theta(1-\theta)/n}$.

Most observations without damage ($Y = 0$) are associated with low-intensity rainfall, e.g. 99 % of the observations without damage are below 6.9 mm in 10 min. Few observations of low damage are associated with high-intensity rainfall. The $Y = 1$ observations are distributed over a larger range of rainfall intensities. The differences in the distributions of $Y = 0$ and $Y = 1$ are also reflected in the empirical proportions (open circles), with increasing values for higher rainfall intensities. Due to the low number of observations for high rainfall intensities, large uncertainty ranges occur for values of $\theta > 0.5$.

3.2 Goodness-of-fit using pseudo-R^2

McFadden’s-R^2 statistic was calculated using different time windows (z) and thresholding criteria (α). Results are listed in Table 3. The maximum rainfall intensity accounts for at most 34 % (for property damage) and at most 30 % (for content damage) of the variance explained, taking into account that these values are approximations and depend on the selected pseudo-R^2. There is a slight improvement in the model predictability if rainfall intensity is based on longer time windows, with an “optimum” between two and four hours. The differences are, however, rather small to be conclusive about what time window best predicts damage. An optimum, if true, may reflect the temporal scale at which failure mechanisms (e.g. floodings, leakages of roofs) have caused damage.

The results suggest that for this kind of analysis there is no need to collect rainfall data with temporal resolutions smaller than 10 min. Lowering the significance level α, and hence selecting higher damage observations, improves the predictability by high rainfall intensities. The results indicate that higher damage observations are more likely to be associated with rainfall data than lower damage observations. Property damage is better explained by rainfall than content damage, although the differences are marginal (1–4 % point).

3.3 Goodness-of-fit using contingency table

Another way to look at model performance is to directly compare observed and predicted values using contingency tables. The model is said to have predicted a significant damage event if the estimated θ is greater than or equal to 0.5 and no damage if θ is smaller than 0.5. The rainfall intensity for which the probability of success equals 0.5, is here defined as the rainfall threshold, although it does not necessarily imply a sudden transition from "no damage" to "damage". The rainfall thresholds are listed in Table 4 for different α and z. The thresholds are slightly higher for lower significance levels and higher for content damage compared to property damage; however, these differences are small compared to uncertainty introduced by assuming gauge measurement to be representative for the area in a 10-km range of the rain gauge.

In a 2×2 contingency table the observed Y (0 – no damage observed or 1 – damage observed) is compared with the predicted Y (0 – no damage predicted or 1 – damage predicted). Table 5 presents the contingency table for $\alpha = 1 \times 10^{-5}$ and $z = 60$ based on property damage data. The percentage correct predictions ($= \frac{a+d}{n} = 0.997$) is heavily skewed in this case due the high number of days without damage. An alternative performance index, less sensitive to skewness of observations, is the sum of fractions of correctly predicted observations ($= \frac{a}{a+b} + \frac{d}{c+d}$) (Kennedy, 2003). Using this approach, scores are presented in Table 6 for a range of z and α. The models score around 5–17 % better compared to random predictions. In most cases, property damage is better predicted by rainfall than content damage, although the differences are small

HESSD
9, 11615–11640, 2012

and for a few cases scores are equal. The scores do not improve when lowering the significance level from 1×10^{-4} to 1×10^{-5}. The highest scores are obtained for time windows between 30 and 50 min, which are smaller than the two to four hours found using McFadden's-R^2.

In yet another way, contingency tables can be used to address the fractions of Type 1 errors and Type 2 errors. Type 1 errors (b in Table 5) can be indicative of local rainfall that caused damage, while it was not recorded by the local rain gauge due to insufficient spatial density of the rain gauge network. They can also indicate that rainfall intensity does not sufficiently represent the damage generating mechanism and that other exploratory variables such as total rainfall volume, wind speeds or building characteristics need to be added to the model. Type 2 errors (c in Table 5) can be related to local rainfall that hit the rain gauge, but not the surrounding urban area. They can also be related to cases of overnight rainfall where people claim the day after. The time window approach used in this study allowed rainfall intensity to be based on rainfall prior to midnight, still rainfall that fell before the start of the time window was not analysed. Both types of errors could be reduced with a higher spatial resolution of rainfall data. Weather radar data are able to provide a better representation of spatial variability, although it is less accurate in determining the intensity than gauge measurements.

The need to reduce Type 1 and Type 2 errors can be different for different stakeholders. An example from the water manager's perspective: a decision to open or not to open a water storage facility may lead to unpreparedness in case of a Type 1 error or unnecessary costs in case of a Type 2 error.

4 Conclusions and recommendations

This study investigated relationships between water-related damage data from insurance companies and rainfall extremes for the period of 2003–2010 in the Netherlands. The results show that high claim numbers related to private property and content damages were significantly related to maximum rainfall intensity, based on a logistic

HESSD
9, 11615–11640, 2012

regression, with rainfall intensity for 10-min to 4-h time windows. The variance explained by rainfall intensity, approximated by a pseudo-R^2 statistic, was at most 34 % for property damage and at most 30 % for content damage, depending on the selected time window. When directly comparing predicted and observed values, the model was able to predict 5–17 % more cases correctly compared to a random prediction. No important differences were found between property and content damage data. A considerable fraction of the variance is left unexplained, which emphasizes the need to study damage generating mechanisms and other explanatory variables, such as total rainfall volume, wind speed or building characteristics. There is also a need for high-resolution rainfall data at the urban scale to have better spatial linkages between rainfall and claim data. A better documentation of exact damage causes in insurance databases is essential to detail relationships with damages caused by failure mechanisms of urban drainage systems.

Acknowledgements. This work has been funded by the EU 7th Framework Programme project Smart Resilience Technology, Systems and Tools (SMARTeST 2010–2012). The authors would like to thank the Dutch Association of Insurers and Royal Netherlands Meteorological Institute for their support and making available the data.

References

Blanc, J., Hall, J., Roche, N., Dawson, R., Cesses, Y., Burton, A., and Kilsby, C.: Enhanced efficiency of pluvial flood risk estimation in urban areas using spatial-temporal rainfall simulations, J. Flood Risk Manage., doi:10.1111/j.1753-318X.2012.01135.x, 2012. 11616

Coulthard, T. and Frostick, L.: The Hull floods of 2007: implications for the governance and management of urban drainage systems, J. Flood Risk Manage., 3, 223–231, doi:10.1111/j.1753-318X.2010.01072.x, 2010. 11617

De Moel, H. and Aerts, J. C. J. H.: Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates, Nat. Hazards, 58, 407–425, doi:10.1007/s11069-010-9675-6, 2011. 11617

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. 11617

Elmer, F., Seifert, I., Kreibich, H., and Thieken, A. H.: A delphi method expert survey to derive standards for flood damage data collection, Risk Analysis, 30, 107–124, doi:10.1111/j.1539-6924.2009.01325.x, 2010. 11617

Ernst, J., Dewals, B., Archambeau, P., Detrembleur, S., Erpicum, S., and Pirotton, M.: Integration of accurate 2D inundation modelling, vector land use database and economic damage evaluation, in: Flood Risk Management: Research and Practice, Proceedings of the European Conference on Flood Risk Management – Flood Risk 2008, Balkema, Rotterdam, 1643–1653, 2008. 11617

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 Manage., 2, 198–208, doi:10.1111/j.1753-318X.2009.01034.x, 2009. 11616

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, 2010. 11617

Hurford, A. P., Parker, D. J., Priest, S. J., and Lumbroso, D. M.: Validating the return period of rainfall thresholds used for Extreme rainfall alerts by linking rainfall intensities with observed surface water flood events, J. Flood Risk Manage., 5, 134–142, doi:10.1111/j.1753-318X.2012.01133.x, 2012. 11616, 11618

Hurford, A. P., Priest, S. J., Parker, D. J., and Lumbroso, D. M.: The effectiveness of extreme rainfall alerts in predicting surface water flooding in England and Wales, Int. J. Climatol., 32, 1768–1774, doi:10.1002/joc.2391, in press, 2011. 11617

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, The Netherlands, 139–146, 2000. 11616

Jonkman, S. N., Bockarjova, M., Kok, M., and Bernardini, P.: Integrated hydrodynamic and economic modelling of flood damage in the Netherlands, Ecol. Econ., 66, 77–90, doi:10.1016/j.ecolecon.2007.12.022, 2008. 11617

Kennedy, P.: A Guide to Econometrics, MPG Books, 5th Edn., Bodmin, Cornwall, 2003. 11626

KNMI: Handbook for the meteorological observation. Chapter 6: Precipitation, Tech. rep., KNMI, De Bilt, Netherlands, available at: http://www.knmi.nl/samenw/hawa/pdf/Handbook_H01_H06.pdf (last access: 1 August 2012), 2000. 11619

Long, J. S.: Regression Models for Categorical and Limited Dependent Variables, Sage Publications, Thousand Oaks, USA, 1997. 11624

McCullagh, P. and Nelder, J.: Generalized Linear Models, 2. Edn., Chapman and Hall, Chicago & London, 1989. 11623

Merz, B., Kreibich, H., Thieken, A., and Schmidtke, R.: Estimation uncertainty of direct monetary flood damage to buildings, Nat. Hazards Earth Syst. Sci., 4, 153–163, doi:10.5194/nhess-4-153-2004, 2004. 11617

Middelmann-Fernandes, M.: Flood damage estimation beyond stage-damage functions: an Australian example, J. Flood Risk Manage., 3, 88–96, doi:10.1111/j.1753-318X.2009.01058.x, 2010. 11617

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. 11620

Parker, D., Priest, S., and McCarthy, S.: Surface water flood warnings requirements and potential in England and Wales, Appl. Geogr., 31, 891–900, doi:10.1016/j.apgeog.2011.01.002, 2011. 11618

Pistrika, A. K. and Jonkman, S. N.: Damage to residential buildings due to flooding of New Orleans after hurricane Katrina, Nat. Hazards, 54, 413–434, doi:10.1007/s11069-009-9476-y, 2009. 11617

Pitt, M.: Learning lessons from the 2007 floods, and independent review by Sir Michael Pitt, The Pitt Review Cabinet Office, London, UK, 2008. 11617

Priest, S. J., Parker, D. J., Hurford, A. P., Walker, J., and Evans, K.: Assessing options for the development of surface water flood warning in England and Wales, J. Environ. Manage., 92, 3038–3048, doi:10.1016/j.jenvman.2011.06.041, 2011. 11618

Ririassa, H. and Hoen, A.: Rainfall and damage: a study on relationships between rainfall and claims in relation to climate change, Tech. rep., Dutch Association of Insurers, The Hague, The Netherlands, 2010. 11618

HESSD
9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

Spekkers, M. H., Ten Veldhuis, J. A. E., Kok, M., and Clemens, F. H. L. R.: Analysis of pluvial flood damage based on data from insurance companies in the Netherlands, in: Urban Flood Risk Management Conference, edited by: Zenz, G. and Hornich, R., Graz, Austria, 461–466, 2011. 11620

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. 11617

Thieken, A. H.: Methods for the documentation and estimation of direct flood losses, in: Urban Flood Risk Management Conference, edited by: Zenz, G. and Hornich, R., Graz, Austria, 2011. 11617

Zhou, Q., Mikkelsen, P., Halsnæs, K., and Arnbjerg-Nielsen, K.: Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits, J. Hydrol., 414–415, 539–549, doi:10.1016/j.jhydrol.2011.11.031, 2012a. 11617

Zhou, Q., Petersen, T., Thorsen, B., and Arnbjerg-Nielsen, K.: Verification of flood damage modelling using insurance data, in: Proceedings of the Ninth International Conference on Urban Drainage Modelling, edited by: Prodanović, D. and Plavšić, J., Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia, 2012b. 11617, 11618

HESSD

9, 11615–11640, 2012

Table 1. Summary of rainfall and insurance data sources.

Data source	Temporal resolution	Spatial resolution	Availability	Records
Manual rain gauge network	daily volumes	$\approx 1/100\,km^2$	1950–today	
Automatic rain gauge network	10-min volumes	$\approx 1/1000\,km^2$	2003–today	
Property damage database	by day	district level	1986–2010	≈ 300 000
Content damage database	by day	district level	1992–2010	≈ 270 000

HESSD
9, 11615–11640, 2012

Table 2. Logistic regression results for model fits on property and content data. The results are based on $z = 60\ \mathrm{mm\,h^{-1}}$ and a range of α levels.

data	α	β_1	SE	LR	d.f.	p	$\exp(\beta_1)$	95 % C.I. $\exp(\beta_1)$ Lower	Upper
property	0.01	0.265	0.0093	766	1	< 0.001	1.30	1.28	1.33
	0.001	0.309	0.0113	723	1	< 0.001	1.36	1.33	1.39
	0.0001	0.319	0.0126	626	1	< 0.001	1.38	1.34	1.41
	0.00001	0.325	0.0141	528	1	< 0.001	1.38	1.35	1.42
content	0.01	0.248	0.0081	882	1	< 0.001	1.28	1.26	1.30
	0.001	0.281	0.0097	782	1	< 0.001	1.32	1.30	1.35
	0.0001	0.276	0.0107	597	1	< 0.001	1.32	1.29	1.35
	0.00001	0.282	0.0118	516	1	< 0.001	1.33	1.30	1.36

HESSD
9, 11615–11640, 2012

Table 3. Evaluation of model performance using the McFadden's-R^2. Outcomes are given for ranges of z and α.

		$z = 10$	$z = 20$	$z = 30$	$z = 40$	$z = 50$	$z = 60$	$z = 90$	$z = 120$	$z = 180$	$z = 240$
property	$\alpha = 0.01$	0.102	0.111	0.114	0.117	0.118	0.120	0.123	0.124	0.126	0.127
	$\alpha = 0.001$	0.186	0.205	0.212	0.215	0.218	0.220	0.224	0.228	0.230	0.227
	$\alpha = 0.0001$	0.234	0.255	0.263	0.268	0.273	0.275	0.277	0.278	0.280	0.275
	$\alpha = 0.00001$	0.280	0.305	0.314	0.323	0.329	0.331	0.335	0.339	0.344	0.340
content	$\alpha = 0.01$	0.092	0.099	0.103	0.107	0.109	0.110	0.114	0.116	0.118	0.116
	$\alpha = 0.001$	0.167	0.177	0.183	0.189	0.192	0.195	0.202	0.207	0.212	0.210
	$\alpha = 0.0001$	0.190	0.201	0.209	0.217	0.223	0.227	0.237	0.244	0.250	0.248
	$\alpha = 0.00001$	0.232	0.244	0.256	0.266	0.272	0.277	0.285	0.292	0.298	0.294

HESSD

9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

Table 4. Rainfall thresholds: rainfall intensity in mm h^{-1} for time window z at which probability of a damage event $\theta = 0.5$.

		$z = 10$	$z = 20$	$z = 30$	$z = 40$	$z = 50$	$z = 60$	$z = 90$	$z = 120$	$z = 180$	$z = 240$
property	$\alpha = 0.01$	52.2	36.3	27.8	22.7	19.3	17.0	12.6	10.3	7.8	6.4
	$\alpha = 0.001$	56.2	39.1	29.8	24.4	20.8	18.2	13.5	10.9	8.2	6.8
	$\alpha = 0.0001$	60.1	42.0	32.1	26.2	22.2	19.4	14.5	11.8	8.8	7.3
	$\alpha = 0.00001$	64.5	45.2	34.6	28.2	23.9	20.9	15.6	12.5	9.3	7.7
content	$\alpha = 0.01$	56.3	39.4	30.1	24.5	20.8	18.2	13.5	10.9	8.2	6.8
	$\alpha = 0.001$	60.8	43.1	33.2	27.0	22.8	20.0	14.7	11.9	8.8	7.2
	$\alpha = 0.0001$	67.8	48.4	37.3	30.3	25.7	22.4	16.5	13.2	9.8	8.0
	$\alpha = 0.00001$	71.6	51.2	39.6	32.2	27.2	23.8	17.6	14.1	10.4	8.6

HESSD
9, 11615–11640, 2012

Table 5. Contingency table, cutoff point $\theta = 0.5$ ($\alpha = 1 \times 10^{-5}$, $z = 60$, property data).

	damage predicted $I_z \geq 20.9$	no damage predicted $I_z < 20.9$	Total
damage observed	$a = 19$	$b = 101$	120
no damage observed	$c = 13$	$d = 34\,056$	34 069
Total	32	34 157	n = 34 189

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
HESSD
9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

Table 6. Scores using alternative performance index $(= \frac{a}{a+b} + \frac{d}{c+d})$.

		$z = 10$	$z = 20$	$z = 30$	$z = 40$	$z = 50$	$z = 60$	$z = 90$	$z = 120$	$z = 180$	$z = 240$
property	$\alpha = 0.01$	1.05	1.07	1.07	1.07	1.07	1.08	1.07	1.07	1.07	1.07
	$\alpha = 0.001$	1.08	1.13	1.14	1.14	1.14	1.12	1.12	1.11	1.10	1.10
	$\alpha = 0.0001$	1.11	1.16	1.17	1.16	1.16	1.15	1.15	1.14	1.13	1.11
	$\alpha = 0.00001$	1.11	1.15	1.17	1.16	1.16	1.16	1.16	1.16	1.13	1.14
content	$\alpha = 0.01$	1.04	1.05	1.06	1.06	1.07	1.07	1.06	1.06	1.07	1.06
	$\alpha = 0.001$	1.07	1.09	1.11	1.10	1.10	1.10	1.11	1.11	1.11	1.10
	$\alpha = 0.0001$	1.06	1.08	1.10	1.12	1.12	1.12	1.14	1.12	1.13	1.12
	$\alpha = 0.00001$	1.07	1.07	1.09	1.11	1.13	1.12	1.12	1.14	1.14	1.12

HESSD
9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

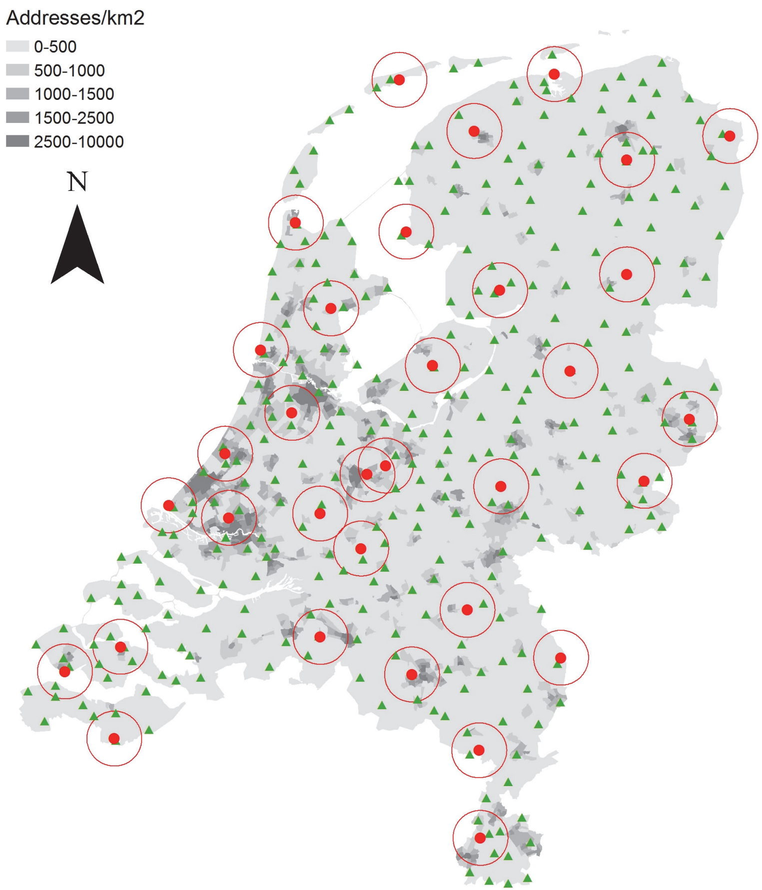

Fig. 1. Locations of 33 automatic rain gauges (solid circles) and 300+ manual rain gauges (triangular markers) and the area within a 10-km radius of automatic rain gauges (open circles). Urban density (addresses/km^2) is presented in grey scales.

HESSD
9, 11615–11640, 2012

Insurance damage data and rainfall extremes

M. H. Spekkers et al.

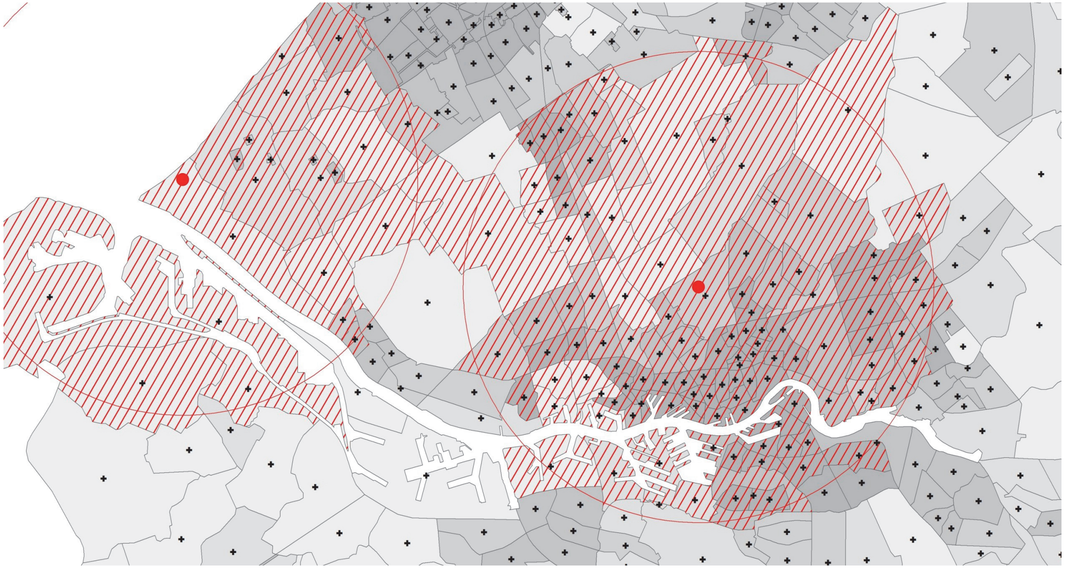

Fig. 2. Example to illustrate the subsetting of insurance data. The two red dots are rain gauges and the open circles the rain gauge areas. The black crosses are the centroids of the districts. The shaded areas are the districts that have been subsetted.

HESSD
9, 11615–11640, 2012

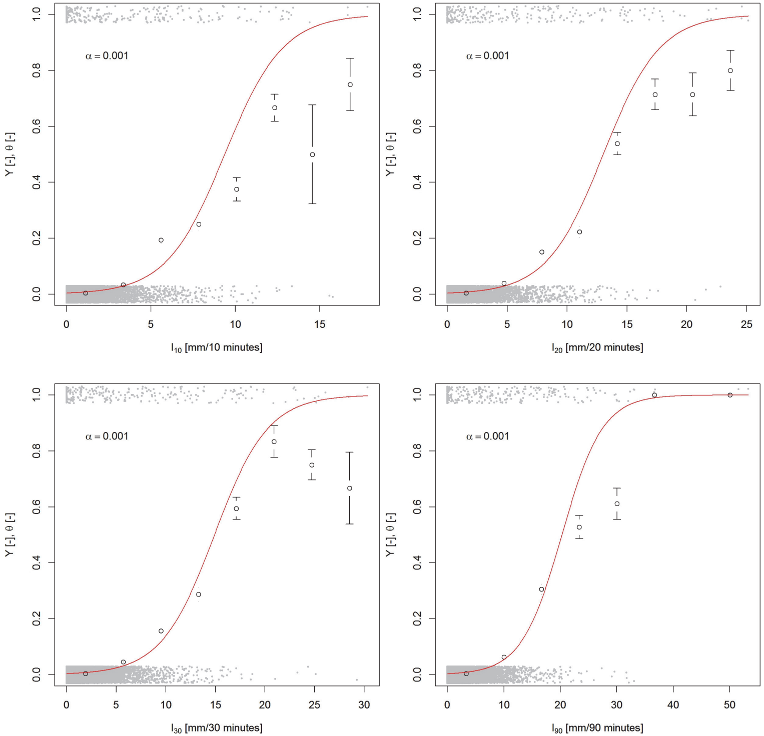

Fig. 3. Logistic functions (solid lines) fitted on property damage data. Plots are related to the cases of $z = 10$, 20, 30 and 90, using $\alpha = 1 \times 10^{-3}$. The grey small dots are the binary observations, either $Y = 0$ or $Y = 1$. A jitter function was applied on the binary observations to better visualize the density of the data points. The open circles are the calculated empirical proportions for eight non-overlapping equally sized bins. The error bars represent one standard deviation of uncertainty.

HESSD
9, 11615–11640, 2012