

A SPATIAL ANALYSIS OF RAINFALL DAMAGE DATA USING C-BAND WEATHER RADAR IMAGES

Matthieu Spekkers^{1,*}, Matthijs Kok¹, Francois Clemens¹, Marie-claire ten Veldhuis¹

¹ Delft University of Technology, the Netherlands

* PO Box 5048, 2600 GA, Delft, the Netherlands; Tel: +31 15 2788578; E-mail: M.H.Spekkers@tudelft.nl

ABSTRACT

Intense rainfall may locally cause considerable damage in cities, e.g. because of sewer floods. A better understanding of relationships between rainfall characteristics and rainfall-related damage can help insurers and flood authorities to respond more adequately to rainfall extremes. This paper aims to study to what extent weather radars can be helpful to predict damage locations and characteristics. For the Netherlands, damage data were analysed based on daily insurance records. A database of C-band radar images was used to extract characteristics of rainfall events. These characteristics were linked to various damage statistics at district level and per day. Results are based on the 150 days with largest damage amounts nationwide. Rainfall and damage locations clearly showed similar spatial patterns when visualized on maps, particularly for the case maximum hourly rainfall intensity and rainfall volume. Through linear regression with log-transformed damage variables, highest correlation was found between claim ratio and maximum hourly rainfall intensity ($r=0.38$). The average damage per claim does not show any significant correlation with rainfall variables, except a weak relationship with maximum hourly rainfall intensity ($r=0.12$). This means that more intense rainfall mainly affects the number of households claiming damage, while it hardly influences the amount of damage per individual household.

KEYWORDS

Extreme rainfall; insurance damage data; pluvial flooding; regression analysis

1. INTRODUCTION

Intense rainfall may locally cause considerable damage in cities, for instance, as a result of flooding from urban drainage systems or rainwater intrusion through building roofs. It is of interest for many stakeholders to understand the process of how rainfall results in damage. In the case of building and building content damage, for instance, insurers are interested to know how characteristics of rainfall explain damage claim data and to what extent these relationships can be extrapolated to estimate damage under climate scenarios. Authorities responsible for the management of sewer floods may prioritize their investments by knowing the amount and location of flood damage. Meteorologists can improve the effectiveness of weather alarms when there is empirical evidence for rainfall thresholds that trigger high damage (Hurford et al., 2012).

In this context, it is useful to have regression models that link rainfall characteristics to damage data, which may reveal predominant rainfall characteristics that cause high damage. Damage data used for such analysis may potentially come from insurers, as insurance databases usually cover many records that are continuously collected in time; even though, strict privacy regulations often limit the amount of data that is available for research. Few studies have examined relationships between rainfall and insurance damage data, rainfall-related or water-related property and content damage data (Zhou et al., in press; Einfalt et al., 2012; Cheng, 2012; Spekkers et al., 2013). General conclusions cannot be drawn yet from these studies as they vary greatly in terms of temporal and spatial resolution, length and quality of the available damage and rainfall data.

This paper contributes to the knowledge of rainfall-damage relationships. It aims to study to what extent weather radar data can be helpful to predict damage locations and characteristics. Weather radars become more and more available and have improved considerably throughout the last decade, which makes them a potentially interesting source of "hazard" data. Damage data that are analysed are provided by the Dutch Associations of Insurers and cover property and content damage claims in the Netherlands in the period of 1998-2011. A database of corrected C-band radar images provided by

the Royal Netherlands Meteorological Institute (Overeem et al., 2009) was used from which various rainfall characteristics were extracted.

2. METHODS

2.1 Insurance databases

The insurance databases, provided by the Dutch Association of Insurers, cover water-related damages to private properties and contents in the Netherlands (Table 1). The insurance claims are related to rainfall-induced damages, such as rainwater intrusion through roofs and walls and floods from sewers or watercourses; and to other, non-rainfall-related causes, such as bursts of water supply pipes. They include the costs of cleaning, drying and replacing materials and objects and the costs of temporarily rehousing people. Daily records are available at the level of 4-position postal districts, i.e. neighbourhood level, which have typical surfaces areas of 1-5 km² (urban) to 10-50 km² (rural). The databases have been extensively checked on missing or incorrect values and inconsistencies as described in Spekkers et al. (2013).

2.2 C-band weather radar data

A database of adjusted C-band weather radar images was provided by the Royal Netherlands Meteorological Institute and cover the entire land surface of the Netherlands, see Table 1. The images are composites based on two C-band Doppler radars, which have been corrected for various biases using data from manual and automatic rain gauges (Overeem et al., 2009). The standard deviation in hourly rainfall volume per radar pixel is 3.8 mm for rainfall volumes larger than 5 mm and 6.6 mm for rainfall volumes larger than 10 mm, based on verification with data from automatic rain gauges (Overeem et al., 2009).

2.3 Data selection

Insurance and rainfall data are used for the period of 1998-2011. This paper discusses preliminary results based on a selection of days. The top 120 days with largest damage nationwide were selected and ranked according to their total number of claims per insured household for both databases. Table 2 lists the first 8 days for both property and content damage claims. The dates of both lists together made a list of 150 unique days. Due to missing radar images, in particular for the 2.5-km radar images, 16 out of 150 days (11%) were discarded from the analyses. Furthermore, first days of the month were excluded, because it is sometimes used by insurers as a default date when claim date was unknown or not entered correctly. Another eight days were removed because on these days (almost) no rainfall was observed, but nonetheless showing considerable claim numbers. Although not confirmed with precipitation data, claims on these days may be related to snowfall as most of the days happen to be in December or January.

2.4 Damage variables

For each day and district (around 4000 districts), the following damage statistics are available: number of claims, number of insured households and total amount of damage. From these, claim ratio, normalized total damage and average damage per claim are calculated; see Table 3 for definitions. Damage values before 2002 were converted from guilder to euro using the conversion ratio 1 guilder = 0.454 euro. All values are in 2011 euros. Every value associated with a year before 2011 was

Table 1. Summary of rainfall and insurance data. The availability of radar data is based on the fraction of available 5-minute composites, see Overeem et al. (2009, 2011).

Data source	Period	Temporal resolution	Spatial resolution	Availability
C-band weather radar data	1998-2008	5 minutes	2.5 km x 2.5 km pixels	83.5%
	2009-2011	5 minutes	1 km x 1 km pixels	≈100%
Property damage database	1986-2011	by day	district level	order 10 ⁵
Content damage database	1992-2011	by day	district level	order 10 ⁵

Table 2. Top 8 days with largest water-related damage nationwide in the period of 1998-2011. Days are sorted by the normalized number of claims per household. Data in column 4 and 5 are normalized by the maximum value recorded in the period of 1998-2011.

Insurance type	Rank	Date	Normalized number of claims per household [-]	Normalized amount of damage per household [-]	Radar images available	Plotted in Figure 1
Property	1	06-06-1998	1.00	1.00	no	no
	2	05-07-1999	0.69	0.90	yes	no
	3	14-09-1998	0.61	0.67	yes	no
	4	18-01-2007	0.54	0.51	yes	no
	5	26-08-2010	0.49	0.73	yes	yes
	6	30-06-2003	0.49	0.87	yes	no
	7	25-11-2005	0.46	0.61	yes	no
	8	14-07-2011	0.45	0.46	yes	yes
	1	06-06-1998	1.00	0.89	no	no
	2	14-09-1998	0.95	0.95	yes	no
	3	05-07-1999	0.83	1.00	yes	no
	4	10-07-2010	0.76	0.64	yes	no
	5	28-06-2011	0.73	0.63	yes	no
	6	14-07-2010	0.70	0.46	yes	no
	7	26-08-2010	0.67	0.97	yes	yes
	8	26-05-2009	0.65	0.67	yes	no

adjusted for inflation according to Statistics Netherlands (2012).

The distinction between rainfall-related and non-rainfall-related claims is not explicitly made in the data. Non-rainfall-related claims occur throughout the year, whereas rainfall-related claims are clustered on wet days. Consequently, a high number of claims in a district is more likely to be associated with rainfall. Spekkers et al. (2013) proposed a method to label districts that show high numbers of claims on a particular day compared to what is expected on dry days. This method is based on the distribution of the number of claims observed in districts on dry days. Given the statistical properties of this distribution, the probability of any observation to be drawn from the distribution was calculated. Observations with probabilities smaller than a significance level, which was set to 0.001 for this study, are likely to be related to rainfall. In the remainder of this paper, these observations are labelled as “damage observations”.

Districts with only a small number of insured households are more likely to show extreme values of damage variables and may distort results. Based on visual inspection of calculated confidence intervals of claim ratio, districts with less than 300 insured households were therefore left out. For the selected 150 days, a total number of 2514 damage observations were found that met the aforementioned criteria and were included in the analysis.

2.5 Rainfall characteristics

Damage observations were linked to various rainfall characteristics (Table 3). The procedure to link rainfall characteristics from radar images to damage observations is as follows.

Rainfall time series are processed on individual pixel level. Rainfall data were abstracted for all damage days and days prior to these. Then independent rainfall events were selected based on intermediate dry period of at least 12 hours, with “dry” being defined as < 0.083 mm for a 5-minute time step. Only rainfall events that coincide at least for one time step with the damage day are kept. This results in either zero, one or two independent rainfall events that can be associated with a damage day. In the case of zero events, all rainfall characteristics are assigned zero values, except the time of rainfall peak, which is marked as not available. In the case of two events, the maximum value out of the two events is taken. This way, maps can be plotted with the spatial distribution of rainfall characteristics as is done in Figure 1. The radar pixel value at the district's centroid is selected to be representative for the district.

Table 3. Definitions of rainfall and damage variables. Damage-related variables are aggregated by day and district and separately for content and property damage.

Variable abbr.	Description	Unit
Damage-related		
dratio	Claim ratio = number of claims in a district per day divided by number of insured households in a district per day	-
dtot	Normalized total damage = total damage in a district per day divided by number of insured households in a district per day	euro/day
dmean	Average damage per claim = total damage in a district per day divided by number of claims in a district per day	euro/day
Rainfall-related		
rmax	Maximum hourly rainfall intensity = maximum intensity of a rainfall event based on an 1-hour moving time window	mm/h
rvol	Rainfall volume (of event)	mm
rdur	Rainfall duration (of event)	h
rmean	Mean rainfall intensity = rainfall volume of event divided by rainfall duration of event	mm/h
rtime	Time of rainfall peak, ranging from -1 to 1, giving the relative time of the rainfall peak between 00:00 the day before and 24:00 the same day	-
rvolbp	Rainfall volume before rainfall peak	mm

2.6 Log-linear model

A linear regression model was applied using log-transformed values of damage variables. Distributions of damage variables encountered in insurance data are typically strongly non-normal (De Jong and Heller, 2008) which is also the case in this study. In case the distribution is log normal, the values of damage variables are log-transformed to approximate normality and linearity assumptions of a linear model. In this study, a log transformation works out well for average damage per claim, but in a lesser extent for claim ratio and normalized total damage. Nevertheless, because this study is an exploratory study, small deviations of the distribution from log normal were assumed acceptable.

Policyholders are not subject to a deductible, which, if it was the case, puts a lower limit to the amount of damage policyholders may claim. It is therefore assumed that the distribution of the damage per claim is not left truncated. Some left truncation of the data can be expected as people may choose not to take the trouble of claiming small damages; however, this factor is ignored in this study. The distribution of the average damage per claim is assumed not to be censored by the insured sum, as water-related damages are typically much smaller than the insured sum.

3. RESULTS AND DISCUSSION

3.1 Spatial patterns of rainfall and damage

To compare rainfall patterns and damage locations spatially, two days were selected for which the spatial variability of maximum hourly rainfall intensity and rainfall volume were plotted on a colour map (Figure 1). The two days were selected from the top 8 days, indicated with "yes" in the last column of Table 2. The damage observations are marked on the map with red dots (content-related) and black crosses (property-related).

By comparing the rainfall and damage data visually, it can be concluded that rainfall and damage show similar patterns. For example, on 26 August 2010 both rainfall extremes and damage locations are concentrated in a horizontal band across the centre of the Netherlands, with rainfall intensities of 20 mm/h or more, whereas in the rest of the Netherlands, with rainfall intensities less than 20 mm/h, no significant damage was reported. On 14 July 2011 rainfall volumes were highest along the west coast, with rainfall volumes of 70 mm or more, while most of the damage observations are clustered in the same region.

In Figure 2 the empirical cumulative distribution functions are given for the maximum hourly rainfall intensity (left) and rainfall volume (right) associated with the occurrence of damage observations. The

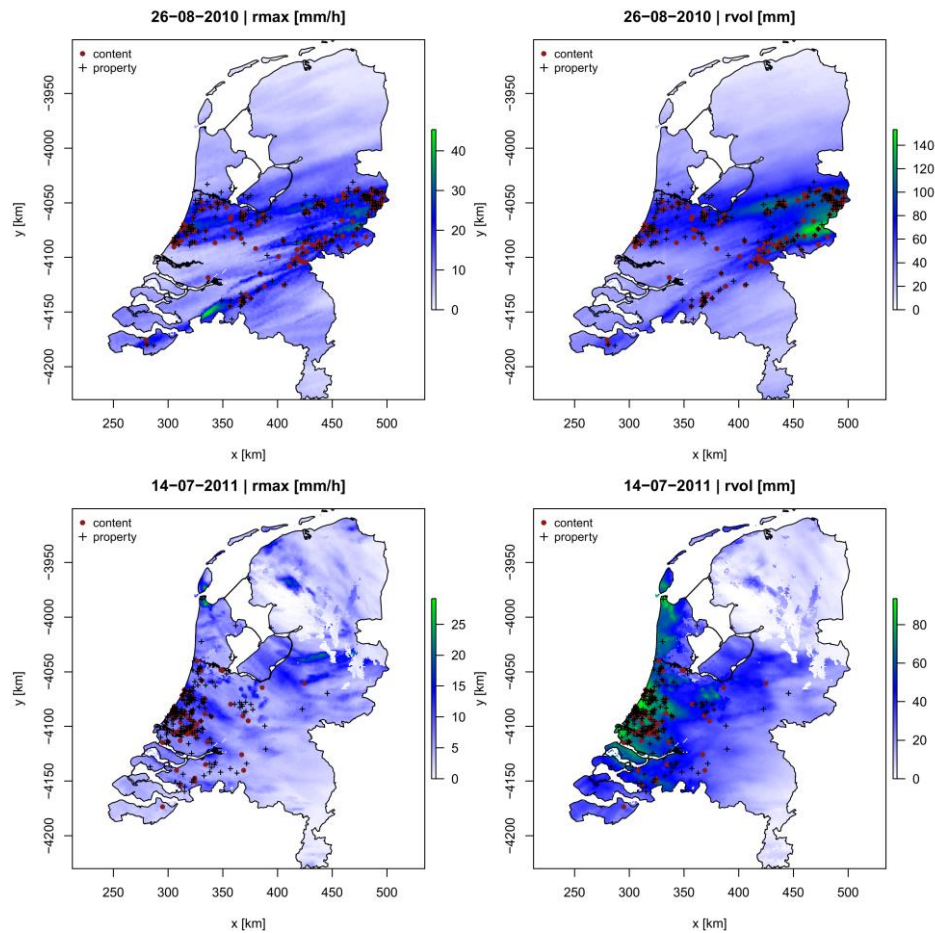


Figure 1. Maps with event maximum hourly rainfall intensity (left) and event total rainfall volume (right) for 26 August 2010 and 14 July 2011. These days are selected from the top 8 list in Table 2 and are days with large amounts of water-related damage nationwide in the insurance claim database. The red dots (content-related) and black crosses (property-related) mark damage observations. Significance level is set to 0.001. Note that colour bars have different value ranges.

curves represent the fraction of damage observations that is below a particular value of a rainfall characteristic. 6.8% of the damage observations is associated with no rainfall, which may be caused by errors in the data, such as rainfall-related observations for which selected rainfall pixel value is not representative. Another reason is that the significance level, used to label damage observations, was set too loose. If significance level is set to 1×10^{-5} , than 2.1% of the damage observations is unrelated to rainfall. Half of the damage observations are observed when rainfall intensity is 12 mm/h or less and rainfall volume is 32 mm or less ($\alpha=0.001$). The shape of curve for $\alpha=0.001$, having a steep slope near the left of the figure, indicates that no rainfall threshold exists for occurrence of damage.

3.2 Regression analysis

Figure 3 shows a correlogram of rainfall-related and damage-related variables. Highest correlation score is found between maximum hourly rainfall intensity and claim ratio ($r=0.38$); rainfall volume and mean rainfall intensity are the second and third best predictors for claim ratio ($r=0.26$ and $r=0.25$). Slightly lower correlation coefficients were found when normalized total damage was taken as dependent variable. Although these relationships are significant, the strength of correlations is moderate. The average damage per claim is only significant with respect to maximum hourly rainfall intensity, but the relationship is weak ($r=0.12$). Time of rainfall peak is insignificant with respect to any of the damage variables.

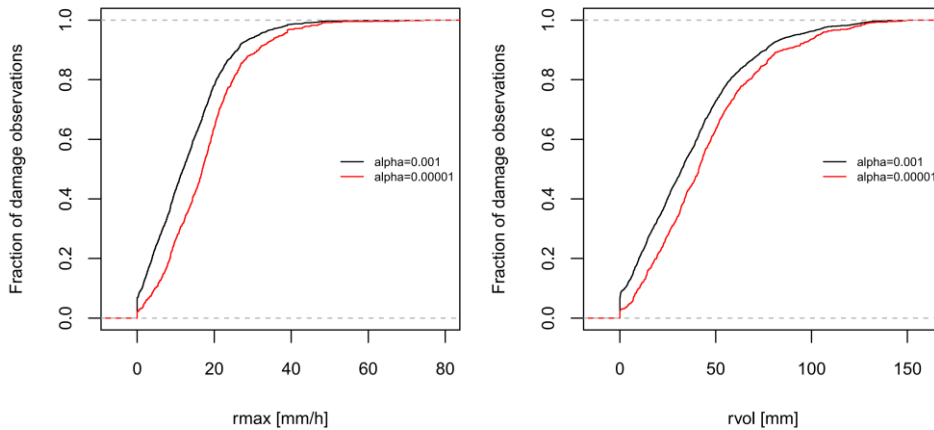


Figure 2. Empirical cumulative distribution functions of maximum hourly rainfall intensity (left) and rainfall volume (right) associated with the occurrence of damage observations. The curves represent the fraction of damage observations that is below a particular value of a rainfall characteristic. The black line is related to significance level of 0.001 (used in this study) and the red line shows the effect of setting a stricter significance level ($\alpha = 1 \times 10^{-5}$).

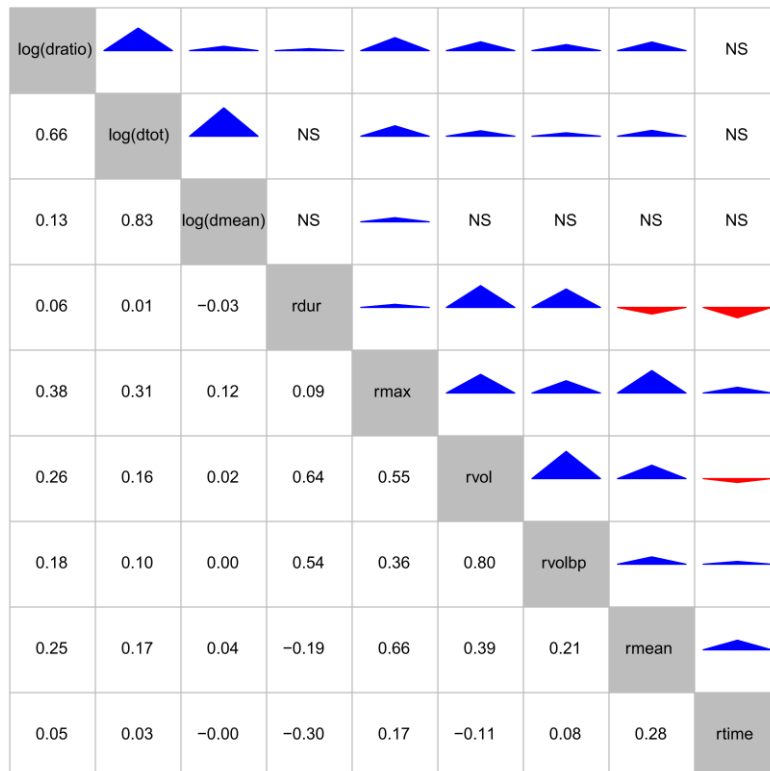


Figure 3. Correlogram of correlations among variables. The direction/colour and size of the triangle depicts the sign and magnitude respectively of the Pearson correlation coefficient between two variables. A blue upward pointing triangle indicates a positive correlation and a red downward pointing triangle a negative correlation. Not statistically significant relationships (1% significance level) are denoted with “NS”.

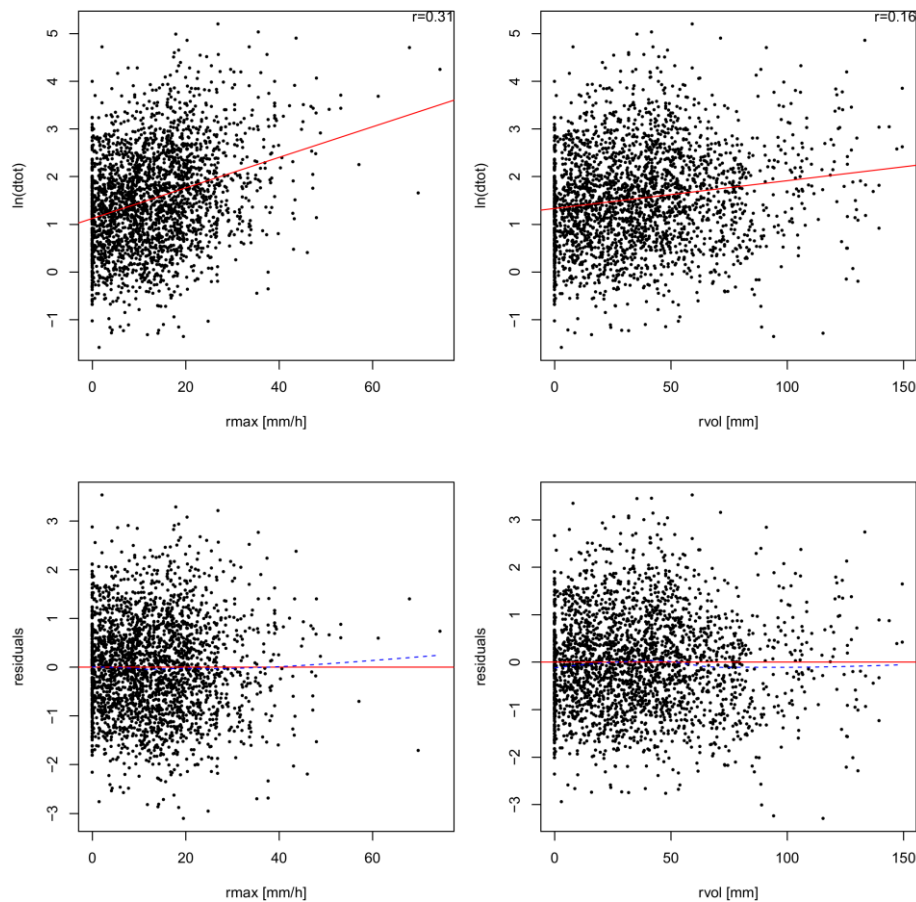


Figure 4. Scatter plots of log-transformed normalized total damage (dtot) against maximum hourly rainfall intensity (top left) and rainfall volume (top right). The red line is the linear regression model. The lower figures show the model residuals. The blue dashed line is the locally weighted regression curve (lowess function).

Scatter plots in Figure 4 of normalized total damage as a function of maximum hourly rainfall intensity (left) and rainfall volumes (right) confirm the moderate relationships, showing large spread of data around the linear fit. Nevertheless, the linear model with log-transformed dependent variable is an appropriate model choice, as the residuals are randomly dispersed around the horizontal axis (lower figures). Similar scatter plots can be made using the log-transformed claim ratio as dependent variable.

To summarize, more intense rainfall mainly affects the number of households claiming damage, while it hardly influences the amount of damage per individual household. This suggests that variations in the average damage per claim are probably related to a large extent to local characteristics other than rainfall, such as properties related to building and/or household.

The results are relevant, for instance, for the development of damage models. The results suggest that focus should be on rainfall thresholds related to rainfall intensity and to a lesser extent rainfall volume or mean rainfall intensity. The results also show that rainfall as single predictor is not good enough. Districts may respond differently to similar rainfall events, e.g. because of different urban drainage systems, and efforts should be made to collect other variables that describe these district-specific thresholds.

4. CONCLUSIONS

The aim of this study was to investigate the extent to which rainfall characteristics, extracted from C-band radar images, can explain claim statistics related to rainfall-related property and content damage data from insurance companies. The paper discusses preliminary results based on data from the 150 days with largest damage amounts in the Netherlands in the period of 1998-2011.

By comparing damage locations and spatial variability of rainfall visually, it can be concluded that rainfall and locations of reported damages show similar spatial patterns. No clear rainfall thresholds could be identified below which no damage occurs.

Using linear regression with log-transformed damage variables, highest correlation coefficient was found between claim ratio and maximum hourly rainfall intensity ($r=0.38$). Rainfall volume is a slightly less important predictor for damage compared to maximum hourly rainfall intensity. The average damage per claim does not show any significant correlation with the rainfall variables, except a weak relationship with maximum hourly rainfall intensity ($r=0.12$). This means that more intense rainfall mainly affects the number of households claiming damage, while it hardly influences the amount of damage per individual household.

A large part of the variance in damage variables is left unexplained. Future research will focus on linking damage data to district-specific characteristics, such as socio-economic characteristics of households and building properties.

5. 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.

6. REFERENCES

- Cheng, C. S. (2012). Climate Change and Heavy Rainfall-Related Water Damage Insurance Claims and Losses in Ontario, Canada. *Journal of Water Resource and Protection*, 04(02):49–62.
- De Jong, P. and Heller, G. Z. (2008). *Generalized Linear Models for Insurance Data*. Cambridge University Press, New York.
- Einfalt, T., Pfeifer, S., and Burghoff, O. (2012). Feasibility of deriving damage functions from radar measurements. In 9th International Workshop on Precipitation in Urban Areas, 245–249, St. Moritz (Switzerland).
- Hurford, A. P., Parker, D. J., Priest, S. J., and Lumbroso, D. (2012). Validating the return period of rainfall thresholds used for Extreme Rainfall Alerts by linking rainfall intensities with observed surface water flood events. *Journal of Flood Risk Management*, 5(2):134–142.
- Overeem, A., Holleman, I., and Buishand, A. (2009). Derivation of a 10-Year Radar-Based Climatology of Rainfall. *Journal of Applied Meteorology and Climatology*, 48(7):1448–1463.
- Overeem, A., Leijnse, H., and Uijlenhoet, R. (2011). Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resources Research*, 47(12):1–16.
- Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and Ten Veldhuis, J. A. E. (2013). A statistical analysis of insurance damage claims related to rainfall extremes. *Hydrology and Earth System Sciences*, 17(3):913–922.
- Statistics Netherlands (2012). StatLine online database: <http://statline.cbs.nl> (accessed on August 2012).
- Thieken, A. H. (2011). Methods for the Documentation and Estimation of Direct Flood Losses. In: Zenz, G. and Hornich, R., editors, Urban Flood Risk Management Conference, Graz (Austria).
- Zhou, Q., Panduro, T. E., Thorsen, B. J., and Arnbjerg-nielsen, K. (in press). Verification of flood damage modelling using insurance data. *Water Science & Technology*.