Title: The use of insurance data in the analysis of Surface Water Flood events – a systematic review

Klodian Gradeci [a*1], Nathalie Labonnote [a], Edvard Sivertsen [a], Berit Time [a]

[a] *SINTEF Building and Infrastructure, Trondheim, Norway*

Abstract

This study employs a systematic literature review to investigate how insurance data can be applied in the analysis of Surface Water Flood events. The study firstly identifies the variables expressing insurance data and those explaining them, together with their interrelationships. Damage variables may be expressed as either monetary-based or number of claims-based. Explaining variables may be subdivided into four categories: meteorological, geographic, demographic and property/building-based. Most of the common and under-researched combinations of these variables and their expression are discussed. Secondly, a comparative analysis is presented of current models, highlighting their differences and similarities. The study demonstrates that the scope and approach of the models varies in relation to scale, the coverage and period of incorporated insurance claims, and the methods used for model development and validation. Thirdly, the study proposes a generic and adaptable framework, constructed from an aggregation of information contained in relevant literature, to define a workflow for model development and future deployment. The study concludes with a discussion of the challenges facing model development and opportunities for deployment.

Keywords: Surface Water Floods; pluvial; insurance data; insurance claims; flood damage; flood risk

1 Introduction

Flooding is a common environmental hazard that endangers the physical, economic and social environment (J. I. Barredo, 2009; Falconer et al., 2009; Kron, 2005). Pluvial flooding is triggered by accumulated rainfall that results in overland water flow and ponding that cannot be drained away, either

[1] Corresponding author. Email: klodian.gradeci@sintef.no

by natural or artificial systems (Bernet, Prasuhn, & Weingartner, 2017; Falconer et al., 2009; Hurford, Parker, Priest, & Lumbroso, 2012). Surface water flooding (SWF) represents a combination of pluvial flooding, stormwater flooding, sewer flooding, flooding from small open-channel and culverted urban watercourses, and overland flows from groundwater springs (Bernet et al., 2017; Falconer et al., 2009; Hurford et al., 2012; Kaźmierczak & Cavan, 2011). The term SWF can be regarded as the optimal general definition of rainfall-related (pluvial) flooding events (Bernet et al., 2017). Economic loss resulting from SWFs, including both tangible and intangible consequences, has increased dramatically in recent decades, and is expected to do so in the future as reported for several countries in Europe and as well as USA and Canada (J. Barredo, Saurí, & Llasat, 2012; Bernet et al., 2017; L. M. Bouwer, 2013; Cheng, Li, Li, & Auld, 2012; Kousky & Michel-Kerjan, 2017; Kron, 2005; David Moncoulon et al., 2016; Wobus, Lawson, Jones, Smith, & Martinich, 2014; Zhou, Panduro, Thorsen, & Arnbjerg-Nielsen, 2013). On the one hand, patterns and intensities of rainfall events are expected to alter in response to climate change, leading to more frequent and severe flooding events (Cheng et al., 2012; Falconer et al., 2009). On the other, a large body of research currently points towards increasing concentration densities of valuable assets due to urbanization and an expanding population as the principle cause of the increasing cost of natural disasters (J. Barredo et al., 2012; J. I. Barredo, 2009; Bernet et al., 2017; Laurens M Bouwer, 2011; L. M. Bouwer, 2013; Kreibich & Thieken, 2008; Spekkers, Clemens, & ten Veldhuis, 2015). Consequently, risk mapping and risk assessment are applied as methodologies for the identification of risk-influencing factors and the evaluation of risk-mitigating measures.

The term risk in this context is commonly expressed as the multiplication of the factors hazard, vulnerability and exposure (Crichton, 1999; Field, Barros, Stocker, & Dahe, 2012; IPCC, 2012; Koks, Jongman, Husby, & Botzen, 2015; Kron, 2005). Hazard refers to threatening natural events, such as intense rainfall, expressed in terms of probability of occurrence. Vulnerability refers to the capacity, or inability, of a society to deal with the hazard. Exposure refers to that of the human population involved, combined with the value of the assets subject to the hazard (Crichton, 1999; Koks et al., 2015; Kron, 2005). An understanding of each component of this risk triangle is required as a basis for analysing how risk due to flooding can be reduced most effectively. Research over the past decades has mostly focused on improving our understanding of the hazard component (Grahn & Nyberg, 2017; Kaźmierczak & Cavan, 2011; Koks et al., 2015; Mechler & Bouwer, 2015; Mechler et al., 2014), while vulnerability and exposure have started to gain attention only during the past decade in the field of flood risk assessment (Cutter, Emrich, Morath, & Dunning, 2013; Koks et al., 2015; Lujala, Lein, & Rosvoldaune, 2014; Rød, Opach, & Neset, 2015). Hazard is a very uncertain phenomenon, which cannot be predicted. The ranges of levels of vulnerability and exposure are very wide and constantly changing. For this reason it is important to develop policies that are able to address a range of different outcomes (Falconer et al., 2009; Kron, 2005). To achieve this, it is important to understand the fundamentals of flood damage data and its possible causes or influences.

Insurance databases represent a potential source of flood damage data. Consequently, analytical research has been carried out in recent years to apply insurance data as a proxy for the analysis of the impact of flooding events (Bernet et al., 2017; Cortes, Turco, Llasat-Botija, & Llasat, 2018; Grahn & Nyberg, 2017; Sorensen & Mobini, 2017; Spekkers et al., 2015; Spekkers, Kok, Clemens, & ten Veldhuis, 2013; Torgersen, Bjerkholt, Kvaal, & Lindholm, 2015; Torgersen, Rød, Kvaal, Bjerkholt, & Lindholm, 2017; Zhou et al., 2013). The outcomes of these studies have included an understanding and ranking of the variables that can explain damage data, the development of models that can predict the likelihood of an SWF event, and the implementation of said models' flood risk assessment frameworks. These studies share a common objective – the development of models that explain insurance data in terms of other rainfall-related, geographic and socio-economic factors. However, the models differ in their identification and expression of the variables used, their interrelationships, the methods used to develop and validate the models, and their further implementation and deployment. The studies have concluded that such models can provide an insight into the relationship between insurance data and key explaining variables. However, much of the statistical variance is left unexplained, emphasising the need: to increase the availability, completeness and reliability of relevant data on one hand; and, to consider alternative ways of expressing selected variables, as well as the inclusion of other explaining variables and their interrelationships, and the methods used to develop the models on the other hand. In the light of this, an aggregation and synthesis of the relevant literature is required in order to compare the similarities and dissimilarities between these studies, and thereafter deliver recommendations for future application based on current best practice.

The aim of this study, carried out in the form of a systematic literature review, is to look into how insurance data can be used to analyse SWF events. It has the following objectives:

- to analyse the historical development of the use of insurance data for analysing SWF events (Section 3.1),
- to identify the variables that express insurance data, and those explaining them (Section 3.2),
- to identify current interrelationships between insurance data and other explaining variables (Section 3.3),
- to provide a current overview of existing models and analyse their differences and similarities (Section 3.4),
- to propose a generic framework based on an aggregation of current models and methods as a basis for a discussion of the challenges related to model development, as well as opportunities for their deployment (Section 3.5)

Section 1 describes the motivation, aims and scope of the study. The methodology is then presented in detail in Section 2. The results and discussions, addressing the aforementioned objectives, are presented in Section 3, and conclusions are set out in Section 4.

2 Methodology

The literature review presented in this study is based on an established research methodology (Booth, Papaioannou, & Sutton, 2011) that ensures a comprehensive search process and systematic review of the relevant literature. The methodology originates from the field of health and social sciences, but its principles are applicable to other fields of study. The approach provides a tool capable of providing a transparent and reproducible research synthesis, thus offering greater clarity, internal validity and audibility (Booth et al., 2011).

The first step in the review process is to define the scope of research that directs focus on the research question (Booth et al., 2011). In the present study, the research question opts to identify how can the insurance data be used to derive models explaining SWF events. In this study, the definition of an SWF, as set out in references (Bernet et al., 2017; Falconer et al., 2009; Hurford et al., 2012; Kaźmierczak & Cavan, 2011), is used because it covers the different types of floods described in the studied literature. The CIMO framework (Petticrew & Roberts, 2008) is used to define the key concepts of the research process (Table 1). The research question is identified as follows: "How (**O**) do we use insurance data (**I**) to analyse (**M**) Surface Water Flooding events (**C**)?"

Table 1.The CIMO framework

Context	Surface Water Flooding/pluvial floods/rainfall/precipitation/urban floods/surface water/storm water
Intervention	The use of insurance data to predict/analyse/explain/understand the occurrence of floods
Mechanisms	Analysis/derivations/relationships between insurance data and other explaining variables to model/predict the occurrence of pluvial floods
Outcomes	Models representing/explaining/associating flood occurrence and insurance data

The keywords, presented in Table 2, were identified based on the titles, abstracts and keywords provided in the literature (Bernet et al., 2017; Cortes et al., 2018; Grahn & Nyberg, 2017; Sorensen & Mobini, 2017; Spekkers et al., 2015; Spekkers et al., 2013; Torgersen et al., 2015; Torgersen et al., 2017; Zhou et al., 2013) following a preliminary screening (first step) using the electronic database Scopus, and Google Scholar. The search scheme and exclusion criteria are shown in Figure 1 and Table 3.

Table 2. Keywords and Boolean operators

What?		**Where?**		**How?**	
Intervention		**Context**		**Outcomes/Mechanisms**	
		storm*		analysis	model*
		pluvial		assessment	relation*
		precipitation		occurrence	statistic*
insurance	and	flood*	and	technique	verification
		rain*		correlation	regression
		urban flood*		risk	validation
		surface water			

Three electronic databases of peer-reviewed literature were used for the final screening (second step). Scopus, Web of Science and Engineering Village are all relevant sources of information in this research area (Aghaei Chadegani et al., 2013; Falagas, Pitsouni, Malietzis, & Pappas, 2008; Jacso, 2005). The keywords, operators and nesting combinations are presented in Table 2. The keywords were applied at title - abstract - keywords - topic level. The last search was performed on 25 April 2018. All years of publication were included in the search process.

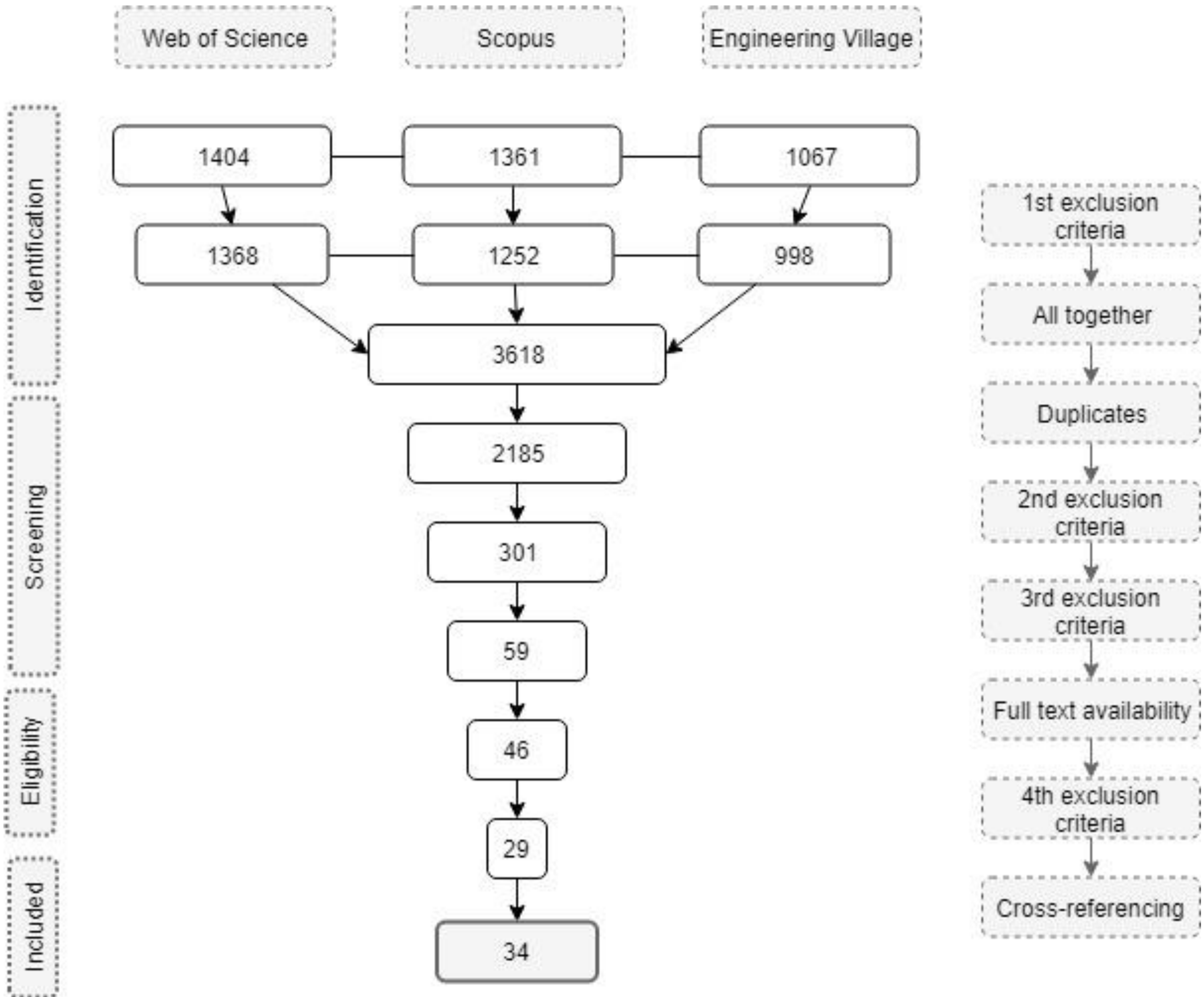

Figure 1. PRISMA framework (Moher, Liberati, Tetzlaff, Altman, & Group, 2009) showing the literature screening process

Table 3. Exclusion criteria

Exclusion	*1st exclusion criterion*	*2nd exclusion criterion*	*3rd exclusion criterion*	*4th exclusion criterion*
Reason for exclusion	Qualitative based on type of literature	Scientific based on keywords, titles	Scientific based on abstract	Scientific based on article and quality assessment

What is excluded	Literature other than Article; Review; Proceeding Paper; (Chapter) Books; English	Other types of floods or storms or drought; insurance programmes, schemes or policies; implementations and types of insurance programmes; description of insurance data, but not how to use it.

Literature screening based on a full content, cross-referencing methodology and author searching was used to check for additional sources. In cases of similar studies being included in different literature sources, priority was assigned to the most recent publication. The final number of selected publications was 34. Subsequently, a data extraction process (Booth et al., 2011) was developed to retrieve and code relevant variables and elements in order to facilitate comparison and identify patterns, themes or trends. Table 4 shows the subgroups of data extraction that help to structure the literature review results described in the following sections.

Table 4. Extraction of the reviewed literature

Data/ Variables	Damage variable	Expression
		Categorisation
	Explanatory variables	Combination with other variables
Models	Establishment/development	
	Validation	
	Implementation	

3 Results and Discussion

3.1 Historical development: graphical representation

Figure 2 presents an historical development of the selected literature that has addressed the relationship between insurance claims and SWFs. The literature is subdivided into publications that have modelled this relationship (solid line) and others that have simply provided relevant research results and discussions (dotted line). Despite the fact that the potential of applying this relationship has been discussed over many decades, results show that it has only been in recent years that an increasing number of publications have appeared that actively develop a model. Nevertheless, these studies cite a limited number of cities, and countries including Canada, Denmark, France, Germany, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the USA. In some cases, more than one study per country is identified, some investigating different regions, and others cited by either similar or different authors.

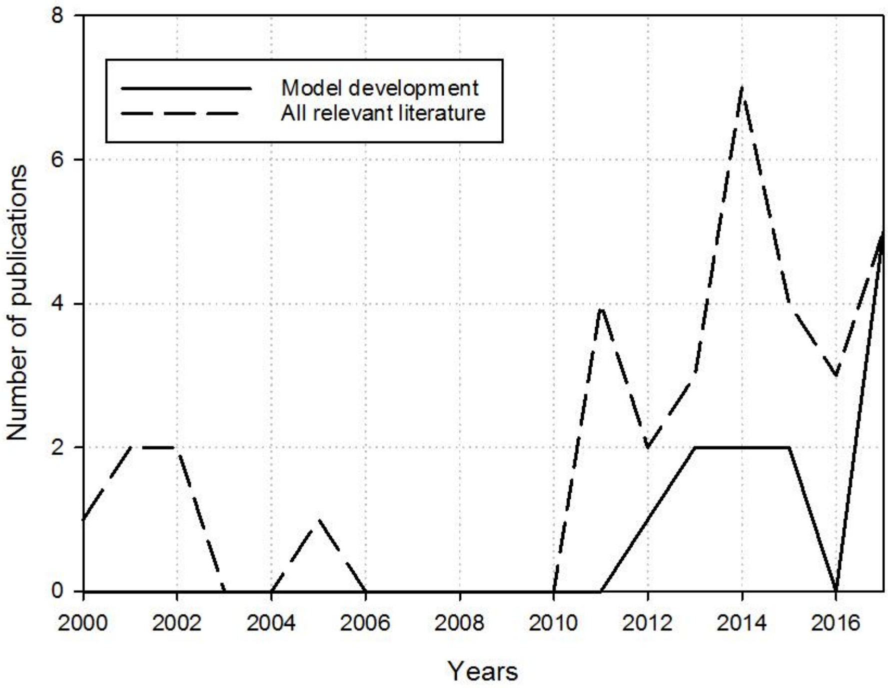

Figure 2. Historical development of selected literature divided between literature in which it was developed a model and all relevant literature

The increasing number of publications produced during the past decade suggests an increased interest in the use of insurance claim data to assess flood risk. Nevertheless, only very few countries have applied such methodologies, thus highlighting the potential for wider application of the approach. Consequently, an aggregation of relevant literature may provide the basis for further development and application of these models and, for this reason, a review of current models and their characteristics, including the variables used and their combinations, should be made available.

3.2 Identification and categorization of the variables used to develop the relationship between damage and explaining variables

A variety of different definitions in relation to the variables used to define insurance data and variables used to explain them are identified in the literature. The term *damage* is a dependent or response variable that expresses the nature of insurance data. The term *explaining* is a damage-influencing or independent variable used to account for or explain the damage variable. In the following, the terms *'damage variable'* and *'explaining variable'* are used.

Figure 3 presents a schematic overview, in the form of a bubble map, of the variables identified and used in these studies. The literature is indexed with numbers according to Table 6. The damage variables are grouped in two main categories based on their expression, as follows:

1- *Monetary-based*, which express the damage variable in terms of a currency value. They include values expressed both as a whole or as a fraction of, or relative to, other parameters or variables.

2- *Number of claims-based*, which express the damage variable in terms of the number of policies/claims that are dependent on other parameters or variables.

The explaining variables may be subdivided into four categories, based on their characteristics and their role in the risk triangle (hazard-exposure-vulnerability). The literature shows sometimes contrasting views when assigning different variables to one of the roles within the risk triangle. An example is the discussion of the variable 'density of built environment', which has been considered both as part of vulnerability and exposure (Koks et al., 2015).

The following categories are defined:

1- *Meteorology-based variables* (**M**) that describe physical atmospheric or natural extreme weather events such as intense rainfall. They may or do cause, influence or trigger the occurrence of flood events. This category belongs to the hazard risk component.

2- *Geographic-based variables* (**G**) that describe the spatial characteristics and parameters of the area under investigation. They may be expressed as single parameters or be combined in the form of a map describing terrain characteristics. These in turn may influence the degree of the hazard if an SWF event occurs, and the coping mechanisms of the system. This category belongs to both the exposure and vulnerability risk components.

3- *Demographic-based variables* (**D**) that state the inventory of elements and assets in a given area in which the SWF event may occur. Such variables may be used to aggregate damage variables. This category belongs to both the exposure and vulnerability risk components.

4- *Building/property-based variables* (**B**) that describe susceptible (at-risk) elements and the system's ability to cope with the hazard. They are directly related to the relevant damage variable. This category belongs to both the exposure and vulnerability risk components.

Damage variables: There are two ways of expressing the damage variables; 1) total number of claims, and 2) claim size. In both cases, these values can be aggregated with other parameters or explaining variables and be expressed as relative values. The review identifies a variety of ways of expression and conclusions in terms of their applications. Findings (Zhou et al., 2013) show that rainfall data cannot be used to explain variation in individual cost per claim. However, such data may be a suitable indicator of overall costs per day. In contrast, in Spekkers et al (2014) the cost per claim term was inadequate to express the damage variable, while claim frequency appeared to provide more satisfactory results. The latter sounds plausible, since cost per claim is related to real estate value, the cost of cleaning and the economic value of the insurance holders' belongings. Consequently, high-income neighbourhoods may appear to be more easily flooded, regardless of the real probability of SWF events in such areas (Sorensen & Mobini, 2017). Nevertheless, the total number of claims term may be biased if it is not aggregated or expressed in a relative manner. A neighbourhood containing a high building density or a high percentage of insured buildings will likely result in a larger total number of claims than an area that is less populated or less densely developed. This observation underlies the importance of using aggregated or relative values (Bernet et al., 2017; Spekkers, Kok, Clemens, & ten Veldhuis, 2014). However, relative values can also be misleading. For example, a neighbourhood containing only one-storey buildings may seem to be more easily flooded than a similar neighbourhood with the same number of multi-storey buildings. For this reason, the use of suitable parameters or variables that aggregate the damage variable may be more useful than using the 'cost of claims' or 'total number of claims' terms. Nevertheless, a combination of both claim size and total number of claims, aggregated by the use of different parameters or in terms of total values, is proposed in order to fully exploit the relationship.

Explaining variables: This review presents several variables that are used to explain the damage variables. Focus is directed mainly on the meteorological category since variables describing rainfall are considered to be the main causes of SWF events. However, a variable that has not been identified is ‘wind-driven rain’, which may damage certain parts of a building that are not accessible to vertical rainfall. Similarly, snow and hailstorms may also contribute to damages paid in response to insurance claims (Hanak & Korytarova, 2014). Moreover (and similar to the damage variable), results may be dependent on how the rainfall data are aggregated. For example, in (Grahn & Nyberg, 2017) the intensity variable, which takes both the duration of the rainfall and the aggregated volume of rain into account, exhibited a statistically significant effect on flood-related damages, while the aggregated volume of rain alone did not. This illustrates the importance of applying meteorological information that reflects the rainfall phenomenon in a temporal perspective. Despite the fact that rainfall may be the main cause of SWF events, previous research has concluded that use of this variable alone is not sufficient to explain observed variance (Cortes et al., 2018; Spekkers et al., 2015; Zhou et al., 2013), thus underlying the importance of considering the impacts of other categories. Different studies include different variables within the four identified categories of explaining variables. However, almost none of these studies include variables from each of these categories in the same analysis. The absence of key variables may explain the large unexplained variance.

The selection of variables also depends on the scale of the investigation (macro-, meso- or micro-, referring to city/country, neighbourhood and building scale, respectively). Different variables are associated with different scales, meaning that different variables and aggregations can be used to explain a given damage variable based on the scale of the latter. At microscales, detailed information regarding a given property may be very relevant (D. Moncoulon et al., 2014). On the other hand, the relationships between socio-economic variables and the damage occurred may be weaker at district level (compared to that of individual households), especially where such districts are heterogeneous. An example of this is in situations where there is a large variance in household incomes (Spekkers et al., 2014). Similarly, the type of insurance database plays an important role in the selection of variables. Different socio-economic variables may play different roles when insurance claims are subdivided between property and movable assets.

Other variables: This review indicates that many variables have been used and screened as important when explaining the damage variable. In addition to the common variables used for similar purposes in different studies, special variables are also noted. The latter include a binary variable depending on whether the event occurred during the day shift or night shift (Grahn & Nyberg, 2017), urban exposure (Grahn & Nyberg, 2017) or the permeability of surfaces (Torgersen et al., 2017); property value (Spekkers et al., 2014); or socio-economic variables such as household income, age and education of breadwinner or fraction of homeowner (Spekkers et al., 2014). Others include urban drainage system properties (drainage capacity, age of infrastructure, percentage of surface water), level of urbanization, socio-economic indices (household income and property value), and district-related parameters (percentages of low-rise and high-rise buildings, percentage impervious surface) (Spekkers et al., 2013), as well as the weather conditions prevailing during preceding days (Torgersen et al., 2015). Other variables have been identified as influential from other studies although they are not used in any of these publications. They include green spaces (Koks et al., 2015), self-protective behaviour (Grothmann & Reusswig, 2006), precautions, external response and early warning (Merz, Kreibich, Schwarze, & Thieken, 2010), as well as building condition (Yazdani, Dowgul, & Manzur, 2010). A systematic map of all the variables that may affect flood occurrence may be useful for the future application of similar research. Moreover, damage variables are also influenced by a complexity of factors associated with the social vulnerability of residents and communities to surface water flooding such as age of residents, willingness to pay for insurance, presence during occurrence of the event, and so on. Vulnerability may be a complex phenomenon to quantify, since it is represented as a composite of other economic, social, cultural and psychological factors that are themselves difficult to describe quantitatively (Holand, Lujala, & Rød, 2011; Shirley, Boruff, & Cutter, 2012).

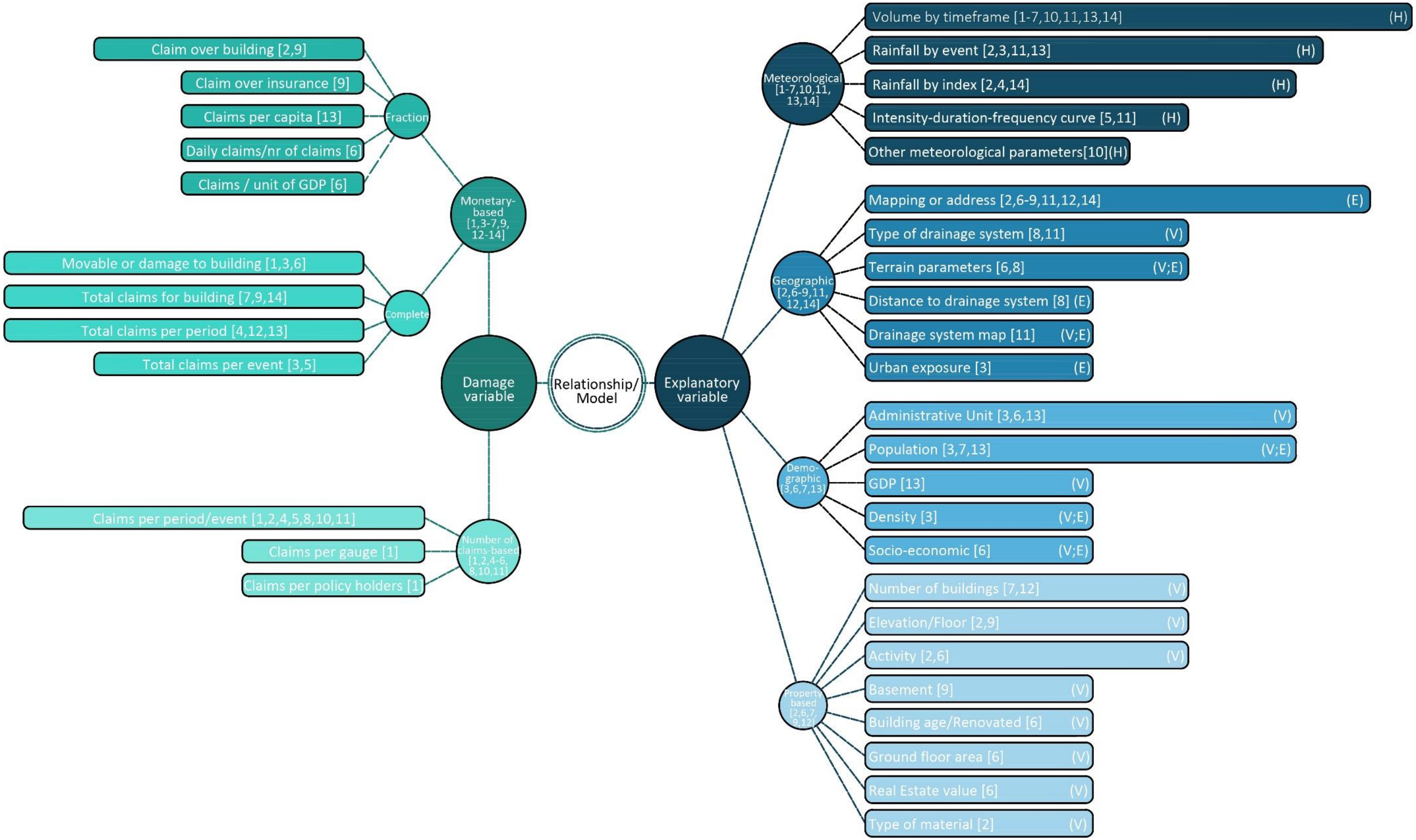

Figure 3. Map of identified variables. [Numbers] refer to the indexed literature in Table 6.Categories of damage (left) and explanatory (right) variables are grouped by colour shades (H-hazard; E-Exposure; V – vulnerability). The increasing area of each cell represents the increased frequency of variables/categories.

3.3 Interrelation between variables

Subsequent to the identification of the various influential variables, this section addresses the interrelationships between these variables. A quilt plot showing the frequency of all potential combinations between the variables used is presented in Table 5. Results from the quilt plot include the following:

- Within the categories expressing the damage variable, the occurrence of the ‘number of claims’ variable is greater than the ‘monetary terms’ variable.
- Within the categories expressing the explaining variables, the meteorological variable is the most commonly used. This makes sense since it is directly related to the cause of the floods. Thereafter, geographic, demographic and building/property-based variables occur, in that order.
- The most frequent combination of two categories (one from damage, and one from the explaining variable groups) is 'number of claims' combined with 'meteorological'.
- The most frequent combination of two variables (one from damage, and one from the explaining variable groups) is 'number of claims per period' combined with the 'rainfall by intensity'.
- Among the categories, the two most common variables are meteorological and monetary-based. However, among variables, the most common combinations involve one from meteorological and one from number of claims-based groups.
- The monetary-based variables are quite widespread in terms of frequency. The reason for this may be the different ways in which the databases are structured, while the number of claims variables are mostly focused on the number of claims per period. This may be explained by the fact that it is possible to retrieve the total number of claims from the databases during a specific timeframe.
- The most used damage variable is ‘number of claims over a specified timeframe’.
- The most used explaining variable is ‘rainfall by intensity’.

- Many variables exhibit low frequency of occurrence. Those exhibiting a single frequency and expressing the damage variable include 'Building claim over building value or insurance coverage', 'Claims per capita or GDP', 'Total daily claims per number of claims', 'Number of claims per gauge' and 'Number of claims per number of policy holders'. Those exhibiting a single frequency and expressing the damage variable include 'Other meteorological parameters', 'Density' and 'GDP'.

The most under-researched areas include the building/property-related and demographic categories. Both of these categories include variables that are vulnerability-based. All green-coloured cells in the quilt plot suggest new combinations between variables that have already been used. The red-coloured cells may provide a useful insight into what should be accounted for at the initial stages. For example, the most exploited relationship is that between 'number of claims' and 'rainfall intensity'. This may imply that these variables exhibit the strongest correlation, and as such may provide a useful insight into which relationship should first be accounted for. While a few of the damage variables specifically belonging to one of the four categories may have been considered as insignificant among the different studies, in general terms, the four categories have all been shown to be important. As a result, it may be expedient to combine variables derived from each of categories as follows:

$$I_{combined} = \left\{ \begin{matrix} I_1(meteorological) \\ I_2(geographic) \\ I_3(demographic) \\ I_4(building) \end{matrix} \right\} \qquad (1)$$

where I represent an index value.

The inclusion of four categories does not necessarily imply that the variance will be better explained here than in situations that include only two or three categories. However, part of the variance will always remain unexplained if no account is taken of variables from any of the identified categories. The results are highly dependent on the selection of both the variables within the given category and their combinations. Similarly, the choice of model used to develop this relationship significantly influences the results.

Table 5. Quilt plot showing the frequency and combination of identified variables, where from red to green colour means high to low frequency. [Numbers] refer to the indexed literature presented in Table 6.

			Damage variable											
			Monetary-based							Number of claims-based				
			Claim per building value or insurance coverage [2; 9]	Claims per capita or GDP [13]	Daily claims per number of claims [6]	Movable or damage to buildings [1;3;6]	Total claims for building [7;9;14]	Total claims per event [3;5]	Total claims per period [4;12;13]	Claims per period/event [1;2;4;5;8; 10;11]	Claims per gauge [1]	Claims per policy holders [6]		
Explaining variable	Meteorological	Rainfall by event [2;3;11;13]	1	1	0	1	0	1	1	2	0	0	4	11
		Rainfall by index [2;4;14]	1	0	0	0	1	0	1	2	0	0	3	
		Volume by timeframe [1-7;10;11;13;14]	1	1	1	3	2	2	2	6	1	1	11	
		Intensity-duration-frequency curve [5;11]	0	0	0	0	0	1	0	2	0	0	2	
		Other meteorological parameters [10]	0	0	0	0	0	0	0	1	0	0	1	
	Geographic	Drainage system [8;11]	0	0	0	0	0	0	0	2	0	0	2	8
		Mapping or address [2;6-9;11;12;14]	2	0	1	1	3	0	1	3	0	1	8	
		Terrain parameters [6;8]	0	0	1	1	0	0	0	1	0	1	2	
		Urban exposure [3]	0	0	0	1	0	1	0	0	0	0	1	
	Demographic	Administrative units [3;6;13]	0	1	1	2	0	1	1	0	0	1	3	4
		Density [3]	0	0	0	1	0	1	0	0	0	0	1	
		GDP [13]	0	1	0	0	0	0	0	0	0	0	1	
		Socio-economic [6]	0	0	1	1	0	0	0	0	0	1	1	
		Population [3;7;13]	0	1	0	1	0	1	1	0	0	0	3	
	Property-based	Building type and quality [2;6;9]	2	0	1	1	1	0	0	1	0	1	3	5
		Number of buildings [7;12]	0	0	0	0	1	0	1	0	0	0	2	
			2	1	1	3	3	2	3	7	1	1	Number of frequencies.	
			10							8				

High				Low

3.4 Modelling the relationship between the variables

Table 6 presents the characteristics of the case studies taken from the fourteen identified publications that have modelled the relationship between the damage and explaining variables. The applied case studies share the same scope, although they vary in terms of both extent and approach. This covers the scale involved, as well as the coverage and period of incorporated insurance claims. Similarly, the methods used to develop and validate the models are different. The methods applied in model development (see Table 6) include visual analysis techniques, a linear, logistic and Poisson regression model, decision trees, principal component analysis and partial least squares discriminant analysis. The percentage of the explained variance also varies. Lastly, the ways in which results are visualised and deployed vary from the provision of a variable correlation function to the development of probabilistic hazard maps.

The results show that regression is the most commonly applied method. This approach is widely used in the field of flood risk assessment (J. Barredo et al., 2012; Botzen & Bouwer, 2016; Changnon, Changnon, & Hewings, 2001; Donat, Pardowitz, Leckebusch, Ulbrich, & Burghoff, 2011; Haug, Dimakos, Vardal, Aldrin, & Meze-Hausken, 2011; Kim, Seo, & Jang, 2012; Lohmann & Yue, 2011; Peng et al., 2014; Wobus et al., 2014). Regressions are simple to apply and to visualise the results. However, the variation in validation techniques used indicate that the explained variance may be relatively low. One reason for this may be the choice of the modelling method. However, low variance may also be caused by 1) the poor availability of, or variation in, the aggregated data (Spekkers et al., 2014), 2) the assumptions regarding the variables included in the study (either by their absence or aggregation/expression), 3) the percentage of insured buildings as a ratio of all the buildings, or 4) alterations to insurance policies over the years.

The choice of modelling method is an important factor influencing outcomes. Different conclusions regarding model application and efficiency are drawn in different studies. According to Spekkers et al. (2014), decision-tree models perform better than global regression models in terms of the explained variance in damage data. Similar conclusions are drawn by Merz et al. (2010) in applications related to fluvial flooding. However, a satisfactory percentage of the variance may be explained using regression techniques (D. Moncoulon et al., 2014; Torgersen et al., 2017). Nevertheless, consideration should be given to the possibility of a non-linear relationship between the damage and explaining variables (Sorensen & Mobini, 2017; Spekkers et al., 2014; Zhou et al., 2013). Regression-based models may not be able to capture this variance. Furthermore, satisfactory results have been derived by applying principal component analysis even when account is taken only of variables within the meteorological category (Torgersen et al., 2015). Similarly, the partial least squares regression technique was also found to be suitable due to the high collinearity in the dataset (Torgersen et al., 2017), although this in turn may lead to poor results when using ordinary least squares regression (Tobias, 1995).

Many of the conclusions derived from the literature are contradictory and no specific modelling method has been proved to produce more satisfactory results than the others. However, the study does reveal that for a model to produce satisfactory results it is crucial to employ a combination of the variables and the methods used. Even if the choice and aggregation of variables corresponds to the specific characteristics of the case study in hand, explained variance and consequently outcomes may be improved by accounting simultaneously for the combination of variables derived from the main four categories. Sensitivity analysis and bootstrapping are additional techniques that can be used to verify and validate the models.

Table 6. Case study and model characteristics. SWF – Surface Water Flood; FV – Fluvial Flood; M – Meteorological, G – Geographical, D – Demographic, P – Property. Macro-, meso- and micro-scale refer to city/country, neighbourhood and building scale respectively.

Literature	Model characteristics					Case study description				
	Development: method	Validation: method	Validation: results	Visualisation	Outcome/Deployment	Coverage of insurance	Scale	Flood type	Categories[2]	Years covered
1 (Spekkers et al., 2013)	Logistic regression model	McFadden's R2/ Goodness-of-fit using contingency tables	34% (property damage) 30% (content damage) / 5-17% prediction accuracy	Logistic function/ tables/ graph	Function predicting damages based on rainfall intensities	20-30% of the market [Netherlands]	Macro/meso	SWF	M	2003-2009
2 (D. Moncoulon et al., 2014)	Logistic regression model/ square root function	Bootstrap method to determine confidence interval based on differences between simulations and extrapolations. Overlay of historical events with probabilistic maps.	74% of the flood claims are located inside the modelled areas	Hazard maps	Multi-peril exceedance probabilistic hazard maps	50% of claims for the market [France]	Macro	SWF + FV	M, G, B	1995-2010
3 (Grahn & Nyberg, 2017)	Logistic regression models	R2-value	3-57% of variance is explained by regression model and variables used	Function/ tables/graph	Aggregated flood damage graph showing relationship between damage and rain intensity	35% of the market [Kristianstad, Sweden]	Micro and meso	SWF	M, D	2000-2013
4 (Cheng et al., 2012)	Visual analysis/ relationship	-	-	Graph	Graph showing relationship between number of claims and monthly rainfall	20000 claims [Ontario, Canada]	Meso and macro	SWF + FV	M	1992-2002
5 (Torgersen et al., 2015)	Principal Component Analysis (PCA)	Correlation loading plot	Up to 99% of the variance is described by the model	Graphic analysis	Graph showing importance of each variable	90% of the market [Fredrikstad, Norway]	Meso	SWF	M	2006-2012
6 (Spekkers et al., 2014)	Decision trees; Poisson and linear	Cross-validation results/R2-value	22-26% of the variance is explained compared to 11-18% when global	Table/ decision tree graph	Ranking of importance of the explaining variables and how	22% of all households [Netherlands]	Macro and meso	SWF	M, G, D	1998-2011

[2] See section 3.2.

	regression model		multiple regression models are used		they affect the damage variable					
7 (Leal, Ramos, & Pereira, 2018)	Relationship defined by correlation coefficient	-	-	Map	Spatial distribution of claim per type of flooding	60% of the market [Lisbon, Portugal]	Macro and meso	SWF + FV	M, G, D, B	2000-2010
8 (Torgersen et al., 2017)	Partial Least Square-Discriminant Analysis (PLS-DA)	Correlation loading plot/cross validation	Up to 65% of the variance is described by the model	Graphic analysis	Graph showing importance of each variable	90% of the market [Fredrikstad, Norway]	Micro and meso	SWF	G	2006-2012
9 (Kousky & Michel-Kerjan, 2017)	Using fixed effects regressions; fractional logit model	Robustness check/R2-value	Up to 36% of the variance is explained	Table	Table showing importance of each variable	1,000,000 claims [USA]	Macro	SWF + FV	G, D, B	1978-2012
10 (Spekkers et al., 2015)	Logistic regression model	McFadden's R2/Wald test	Up to 20% of the variance is explained	Graph	Graph showing empirical probability of precipitation-related claim occurrence as a function of rainfall intensity	6% of the total number of households [Rotterdam, Netherlands]	Micro and meso	SWF	M	2007 - 2013
11 (Sorensen & Mobini, 2017)	Visual analysis	-	-	Map	Flood hazard map	Up to 8% of the market [Malmo, Sweden]	Meso and macro	SWF	M, G	20 years
12 (Bernet et al., 2017)	Visual analysis	-	-	Map	Spatial distribution of claim per type of flooding	Up to 48% of buildings [Switzerland]	Meso and macro	SWF + FV	G, B	2004-2013
13 (Cortes et al., 2018)	Linear and logistic regression model	Relative operating characteristic (ROC) diagram	Relative area under ROC curve up to 0.81	Table/graph	Graph simulating the probability of damage as a function of precipitation	43,640 claims [Catalonia, Spain]	Meso and macro	SWF	M	1996-2015

14 (Zhou et al., 2013)	Linear regression model	Significance level/boxplot	-	Table/map	Flood hazard map	1000 claims [Aarhus, Denmark]	Meso and macro	SWF	M, G	2005-2011

3.5 Proposed generic framework for developing models for the analysis and assessment of SWFs

3.5.1 On the relevance and development of the framework

The results from this review indicate that the process of developing models that use insurance data to explain SFW event occurrence depends on the characteristics of the case study, data availability and assumptions regarding how to interrelate data. Due to the specificity of their applications, no overall conclusions can be drawn regarding the variables or methods that can be used, or the steps for developing the models and their further application. Consequently, a generic and adaptable framework has been developed, using the aggregated results from relevant literature, to define a workflow that may be implemented to develop a model of the relationship between the damage and explaining variables, and its further application and deployment. This framework is presented in Figure 4 together with notes accompanying several of the steps (Table 7). It incorporates the assumptions and decisions that may be adapted to any specific case study in hand. The framework should be regarded as a guide to the development and further deployment of models used in the analysis and assessment of SWF events.

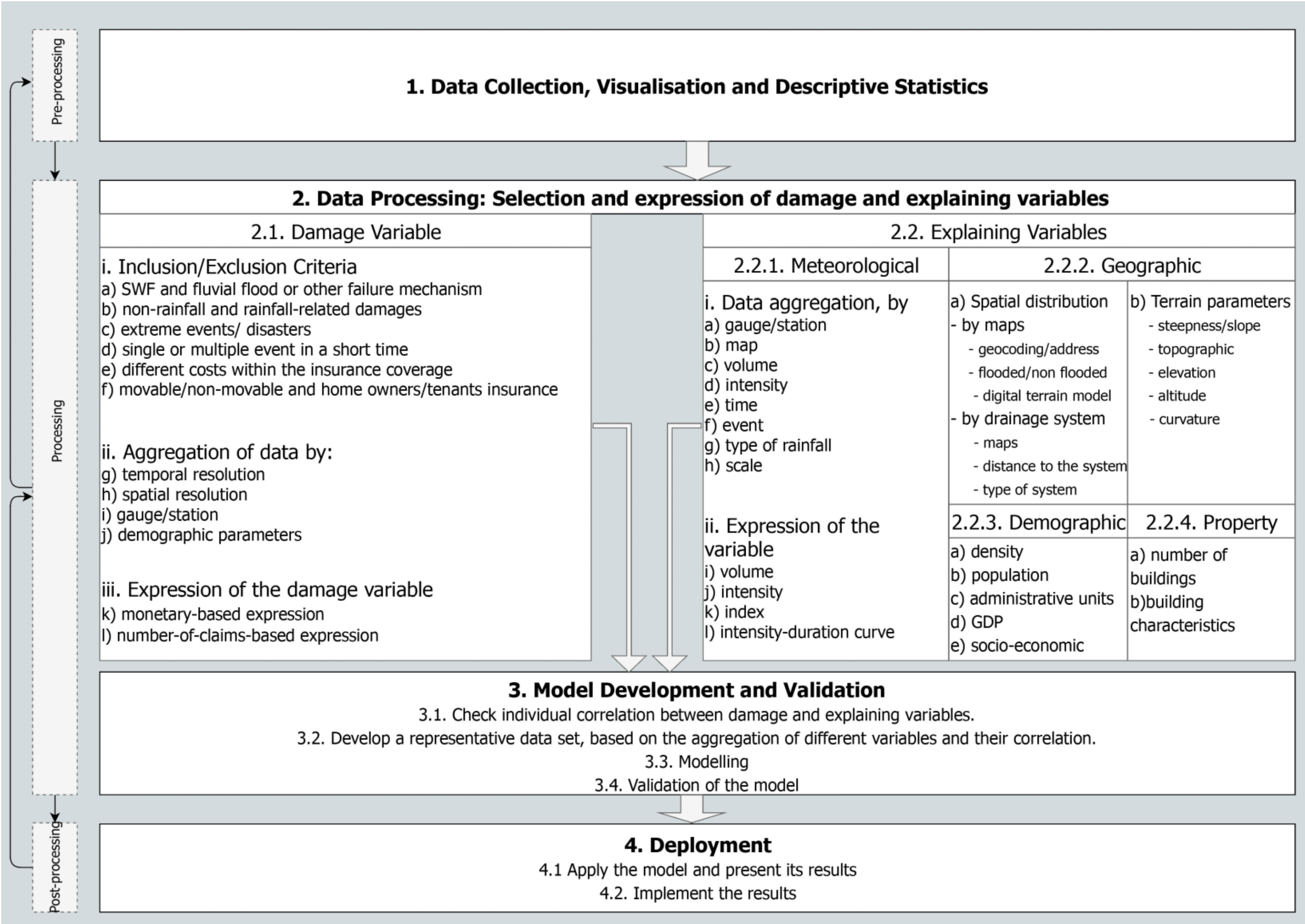

Figure 4. Proposed generic framework: guidelines for data collection, visualisation and descriptive statistics, the selection and expression of damage and explaining variables, model development and evaluation, and further deployment.

Table 7. Notes referring to the steps in the aggregated framework illustrated in Figure 4.

Step	Comments/notes
1	Coverage of the insured building within the studied area is of interest since this percentage may influence the variance of the data. A clear distinction should be made when data from different insurance companies are used within the same sample, since there may be differences in the policies they provide and their reporting procedures. Insurance data should be carefully checked for duplicates, missing data or outliers.
2.1. a)	A classification methodology for pluvial and fluvial flood events and other failure mechanisms (e.g. roof leakage), can be performed as in (Kaźmierczak & Cavan, 2011; Leal et al., 2018).
2.1. b)	The filter approach based either on (Spekkers et al., 2013) or press information (Cortes et al., 2018) can be applied.
2.1. c)	Event classification based on conditions covered by the insurance. The reader is referred to (Sorensen & Mobini, 2017).
2.1. d)	In cases of multiple event occurrence, the damage claim should be further investigated in order to find out whether it is a consequence of just one event or the sum of all events.
2.1. e)	In situations where costs of insurance coverage may be subdivided (costs for cleaning, replacement, etc.) – a study investigating both total and individual cost components is proposed.
2.1. f)	Division between these two factors may enable better differentiation between costs associated with structural damage and those associated with the residents. This may increase variance from one claim to another. In (Grahn & Nyberg, 2017), no difference was observed in the explained variance from property and movable components. However, in (Spekkers et al., 2013), which addressed only property damage, the variance was better explained than in the case of content damage.
2.1. g)	Different timeframe windows and intensities have been proposed by different studies. These range from 7-8 minutes to up to 12 days (Cortes et al., 2018; Sorensen & Mobini, 2017; Spekkers et al., 2015; Spekkers et al., 2013). This may enable a differentiation of claims that may be related to other failure mechanisms in the private domain.
2.1. h)	Data can be sorted according to location, number of buildings, address, district or neighbourhood level.
2.1. i)	A range of 10 kilometres from the rainfall gauge is proposed in (Spekkers et al., 2013), while 15 kilometres is suggested in (Berne, Delrieu, Creutin, & Obled, 2004). The range value may be influenced by several characteristics. For this reason, it is proposed that a study be carried out that defines the decorrelation distance used in the case study in hand.
2.1. j)	Insurance claims can be sorted using one of the demographic variables (see step 3.2.3.).
2.1. k)	When the damage variable is monetary-based, the value should be adjusted for inflation during the year in question. A transformation into normality can be performed by using the natural logarithm as applied in (Grahn & Nyberg, 2017). In addition, account should be taken of any insurance policy that states a minimum reimbursement amount as part of its terms and conditions (Grahn & Nyberg, 2017). Careful consideration should be made if the damage variable as monetary-based, since the cost of reimbursement may be highly dependent on the value of the real estate.
2.1. l)	Variable expression based on number of claims may reduce the influence of disproportionalities between areas with different property values.
2.2.	A qualitative analysis of the characteristics of the area is proposed as a means of identifying the kind of variables that can be used to explain, or relate to, the damage variable.
2.2.1. g)	Differentiation between precipitation types may be applied since some, such as snow, will not generate an immediate flood event response (Torgersen et al., 2015).

2.2.2.	Different techniques are available for the calculation of topographic variables (Wilson, O'Connell, Brown, Guinan, & Grehan, 2007) and the development of multi-dimensional terrain models (Yang, Grönlund, & Tanzilli, 2002; Zischg, Mosimann, Bernet, & Rothlisberger, 2018).
2.2.3.	Median instead of mean values can be used for variables that exhibit strong variance. This will reduce the influence of outliers (Spekkers et al., 2014).
3.1.	A visual correlation may be helpful as a means of identifying the linearity or monotonicity of the relationships. This in turn will provide a useful insight into subsequent steps and model selection. It will also provide an overview of what should be included in the explaining variables categories.
3.2.	The development of a representative dataset is based on the aggregation of different variables and their correlation.
3.3.	Parameters may be defined using the maximum likelihood (Kousky & Michel-Kerjan, 2017; Spekkers et al., 2015; Spekkers et al., 2014) or ordinary least squares method (Spekkers et al., 2015; Spekkers et al., 2013, 2014; Zhou et al., 2013). A sensitivity analysis may be carried out in order to obtain greater insight into the parameters involved and their influence on the output.
3.4.	Validation of the model depends on the method used in step 4.3. The reader is referred to specific literature examples presented in Table 7. Bootstrapping is also recommended.
4.1.	The results can be presented in the form of: 1. Graphical representations of the probability of damage occurrence due to SWF events as a function of one of the explaining variables (e.g. rainfall). 2. Flood hazard maps in the form of: - a spreadsheet of the claims reported based on location (point) and surface (degree of the damage). Visualisation of at-risk zones in the studied area based on rainfall intensity. - a visualisation of the spread of risk calculated according to a risk triangle. Future weather scenarios may be used to express the probability of hazard occurrence.

3.5.2 On the challenges facing model development

The outcome of applying an aggregated framework, which includes both the model and its further deployment, is dependent on the availability of the data (willingness to share) and their quality or reliability. These parameters are the result mainly of the degree of systematic recording (classification) of the data, combined with the uncertainties involved. Any scarcity, inhomogeneity, or lack of availability of data hinders the spatial and temporal correlation between the damage and explaining variables, and in turn constrains the quality of the model outcome, which may be represented by the explained variance.

A crucial issue that limits the potential for such model development is the availability of damage data, which is derived from either insurance companies or individual data collectors. It is observed a reluctance within the insurance industry to share detailed information (such as the exact location of the source of flood damage compensation (Grahn & Nyberg, 2017)). There may be many reasons for this reluctance, such as competition for market share, reputational issues, loyalty towards customers, or anxiety about the impact disclosure may have on residential property markets. A recent study, which carried out interviews and analysed the results from eight largest insurance companies in Norway in regard to their willingness and demands to share damage data, concludes that the largest insurance companies (representing 90 % of the market) are willing to share their data with municipalities and governmental agencies (Hauge et al., 2018). However, in order to share their data, several demands were identified: an arrangement that ensures restricted manageable admission of their data, especially to other (inter-)national companies; the availability of a data administrator and/or intelligent infrastructure that guarantees security and confidence in data protection; and, compliance and adaptation to new implemented regulations regarding protections of privacy (Hauge et al., 2018). The availability of such information would facilitate a better understanding of the vulnerability component of the risk triangle. Currently, there exists several databases in Norway and worldwide that have collected damage or other relevant data regarding SWF events, and a review can be accessed in (Labonnote, 2017; Labonnote, Hauge, & Siversten, 2018); however, data are spread around a heterogeneous community of stakeholders concerned with different motivations, needs, and levels of data processing. It is concluded that digitization and its opportunities can improve the workflow of data collection and analysis and increase the quality of data. The recent availability of Internet of Things, Big Data analytics and Artificial Intelligence can enable fast, systematic and sustainable (digital) data analytics, which can subsequently trigger a global data-driven evaluation system regarding the SWF event occurrence and their impact on society.

A commonly acknowledged issue that arises during the comparison of different studies is the lack of a consistent classification system for damage claims. Several schemes have been developed involving the classification of flood events by type (fluvial or pluvial), degree (event extremity), damage to assets (movable or non-movable), or origin/consequences (non-rainfall and rainfall-related damages) based on spatial resolution, temporal resolution, costs or degree of wetness. However, all these approaches have their shortcomings (Bernet et al., 2017), which in turn may decrease the explained variance derived from the model.

Even when data are accessible, they may be characterised by levels of uncertainty associated with both damage and explanatory variables. The temporal and spatial distribution of rainfall may not be correctly accounted for due to non-uniform distribution or a non-representative number of measurement gauges/stations. The spatial resolution of radar images may be too coarse to capture the spatial variability of rainfall at the subpixel scale, causing an underestimation of rainfall peaks of convective cells (Spekkers et al., 2014). Data variation in a spatial context is another source of uncertainty that may be attributed to a lack of specified addresses, the availability of which may enable the parametrization of geographical information at the level of other damage, demographic and meteorological variables (Spekkers et al., 2013, 2014; Zhou et al., 2013). It should be noted that an absence of recorded damage in a given area does not necessarily mean that the area has not been affected by a flood event (Bernet et al., 2017). It may simply indicate that no buildings were in the vicinity of the flooded area, or that the buildings were properly protected against the flood event, or the occurred damages were not properly registered. Lastly, the scale of a given area may increase the variability of the outcome because different scales of district (neighbourhoods/cities/countries) may be associated with different parameters linked to climatic conditions, insurance policies or the percentage of insured buildings. Another source of uncertainty resides in variables associated with the buildings themselves and the socio-economic status of their residents, which is related in turn to self-protective behaviour. For example, building refurbishment may not have been recorded. Moreover, tenants or owners may share different responsibilities, and consequently different levels of vulnerability.

It is widely acknowledged that risk assessment should provide an indication of the degree of reliability of risk quantification (Merz & Thieken, 2009), although such reliability may be constrained if data are scarce, missing or associated to uncertainties. Consequently, a strategy involving the implementation of a systematic and homogenous recording process that includes information from different explaining variable categories at both local and national scales is suggested even if a comprehensive harmonization at international scale has been deemed as unlikely to be effective (Surminski et al., 2015). Policies that regulate and digitise the claims process can better facilitate both grounds for claims and more accurate inputs as a means of improving current models. Subsequently, the application of risk assessment can be more useful to higher implementation schemes such as policies or programmes.

3.5.3 On the opportunities for model application presentation

The model relating the damage and explaining variables can be further applied within a risk assessment framework or sensitivity analysis. These applications can be useful for stakeholders such as insurance companies, government agencies and meteorological institutes. Figure 6 provides a schematic presentation of the interrelations between model application, implementation and involved stakeholders.

Risk assessment enables the graphical representation of risk distributed in a spatial and temporal context. One outcome is the production of probabilistic maps of metropolitan areas showing the likelihood of occurrence and degree of damage based on meteorological events, similar to that illustrated in Moncoulon et al. (2014). A graphical display such that the one used in (Brevik, Aall, & Rød, 2014) might then be employed (see Figure 5). Such a framework can be used to evaluate potential increases in damage resulting from flooding that may be caused by climate change. This may be achieved by incorporating a global climate model(Cheng et al., 2012). The likelihood of both SWF and fluvial flood events may be included as part of the overall hazard scenario. Furthermore, insurance data can be collected from different sources for different purposes related to residential, business or agricultural properties, as well as state-owned public buildings and transport infrastructure. Lastly, multi-dimensional models for flood events and specified terrains can be incorporated into risk assessment frameworks.

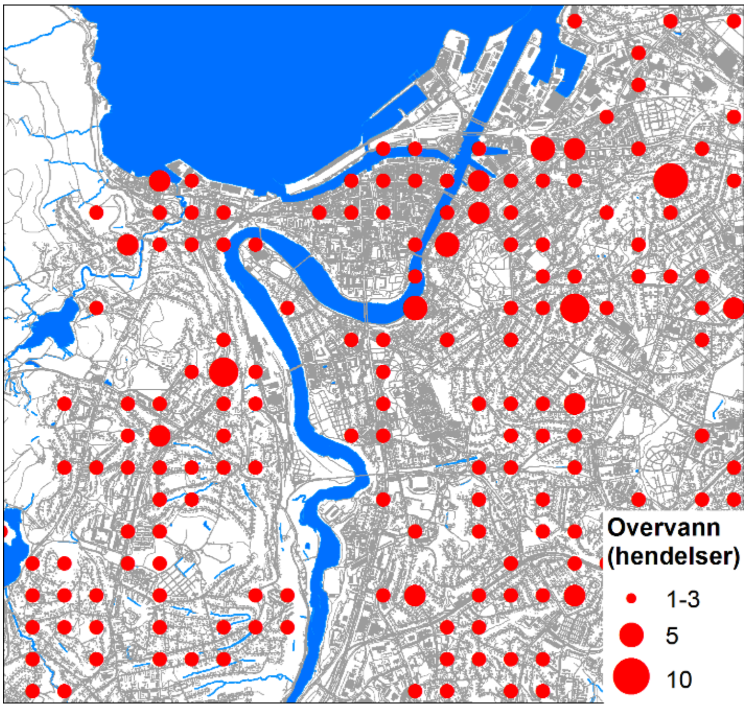

Figure 5. Example of graphical display showing the distribution of insurance claims in Trondheim, Norway. This figure has been reproduced with permission from Brevik et al.(2014). [Translation from Norwegian language: Overvann – Stormwater; Hendelser – Events],

Application of global sensitivity analysis (Saltelli et al., 2008), which is identified as a research gap, enables the understanding and quantification of a given system. As such, it is able to provide estimates of the influence of the inputs on the outputs. The relationship between the damage and explaining variables is replete with uncertainties. As a consequence, the application of global sensitivity analysis enables a ranking of the importance of given parameters and/or their uncertainties. Such rankings can support decision-making processes by means of facilitating comparisons of relative performance, and by optimizing design selection and the implementation of a policy or mitigation action.

3.5.4 On the opportunities for model deployment

The following is a summary of implementations of the models discussed in the foregoing:

- *Policy writing and the execution of mitigation measures*.

- The importance of socio-economic variables, as identified from sensitivity analysis, may provide an insight into their relative contributions in the vulnerability and exposure components of the risk triangle. The temporal distribution of damage claims may provide insights into what values of rainfall intensity and seasons of the year most closely correlate with damage claims. This facilitates a clearer identification of parameters that can reduce risk effectively. The latter can be applied during policy writing and the prioritization of mitigation measures.
- The implementation, based on risk assessment, of more proactive, cost-effective and politically achievable investments in infrastructure adaptation at local, regional, and national scales (Kousky & Michel-Kerjan, 2017).
- An understanding of trends in damage claims and causality based on sensitivity analysis. For example, several authors have concluded that the causes of increases within the vulnerability component of the risk tringle are associated with socio-economic factors, such as population growth and increased wealth among policy holders (and thus, the damaged products may be more valuable), rather than the hazard component associated with climate change (J. I. Barredo, 2009; Bernet et al., 2017; Laurens M Bouwer, 2011; L. M. Bouwer, 2013; Spekkers et al., 2015).
- Improvements in insurance policy writing, such as the inclusion of specific clauses related to rainfall intensity criteria (Spekkers et al., 2013).

\- *Writing and updating of design codes*.

- Current design criteria related to urban drainage system capacity or the return period of design storms (Spekkers et al., 2015) can be implemented or updated.
- The development or validation of damage models.

\- *Improving customer service*.

- Once a clear association between hazard intensity and its consequences is established, direct weather alerts or warnings can be communicated to residents. This will boost the emergency preparedness of residents, which may in turn may limit damage and levels of vulnerability.
- Better management of call centres during flood events. Many companies have indicated that there is a sudden increase in communication demand from clients during extreme events (Spekkers et al., 2013).

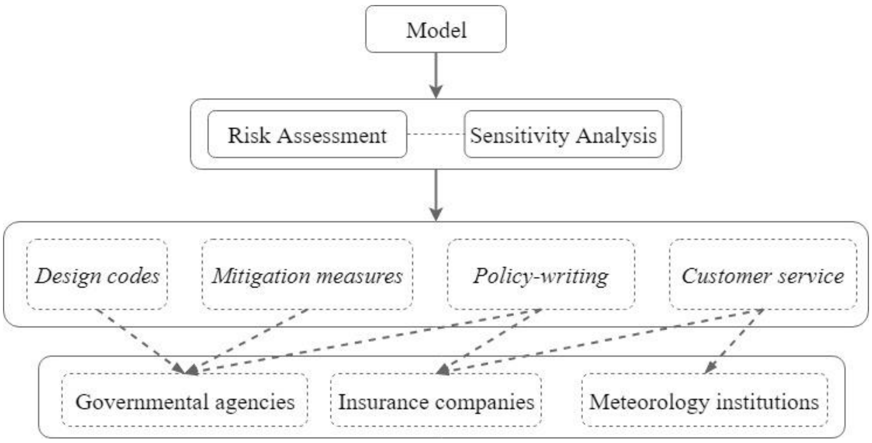

Figure 6. Opportunities for applying and implementing models

4 Conclusion

This study has carried out a systematic literature review to investigate how insurance data can be applied in the analysis of SWF event occurrence. The review concludes that models that identify the relationships between insurance data and explaining variables may provide an insight into the occurrence of surface water flood events. The study has identified four main categories of explaining variables (meteorological, geographical, demographic and property/building-related). Potential ways of expressing both damage and explaining variables, as well as their combinations, have been discussed, and recommendations for future applications proposed.

A generic framework providing guidelines for the development of models of similar scope and their further deployment has been aggregated on the basis of previous applications. The review shows that the outcome of such models is sensitive to factors such as the selected variables and their expression/aggregation, the combination of variables, the methodologies used to establish the model in question, data availability and quality. The study emphasises the importance of the systematic recording and public disclosure of insurance data as a means of improving the implementation of, and outcomes from, these models.

Such models can enable sensitivity analysis and risk assessment frameworks that can be further incorporated into decision-making processes, policy writing and implementations. The review demonstrates an increase in interest worldwide in the development of such models at local and national scales. However, their application is mostly geographically focused, which emphasises the potential for wider application.

Author Contribution Statement

Conceptualization (KG, NL, ES, BT); Data curation (KG, NL, ES, BT); Formal analysis (KG); Funding acquisition (NL, ES, BT); Investigation (KG); Methodology (KG, NL); Project administration (NL, ES); Software (KG, NL); Validation (KG, NL, ES, BT); Visualization (KG); Roles/Writing - original draft (KG); Writing - review & editing (KG, NL, ES, BT)

Acknowledgments

This study was funded by the Norwegian Centre for Research-based Innovation '*Klima 2050'* (www.klima2050.no) with the intention of promoting the development of a global, data-driven, evaluation system that will provide policymakers with know-how linked to societal risk associated with climate change, and to strengthen the innovation capacity of national agencies and private companies to address the consequences of climate change.

References

Aghaei Chadegani, A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., & Ale Ebrahim, N. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science, 9*(5), 18-26.

Barredo, J., Saurí, D., & Llasat, M. (2012). Assessing trends in insured losses from floods in Spain 1971–2008. *Natural Hazards and Earth System Sciences, 12*(5), 1723-1729.

Barredo, J. I. (2009). Normalised flood losses in Europe: 1970–2006. *Natural Hazards and Earth System Sciences, 9*(1), 97-104.

Berne, A., Delrieu, G., Creutin, J.-D., & Obled, C. (2004). Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology, 299*(3), 166-179. doi:https://doi.org/10.1016/j.jhydrol.2004.08.002

Bernet, D. B., Prasuhn, V., & Weingartner, R. (2017). Surface water floods in Switzerland: what insurance claim records tell us about the damage in space and time. *Natural Hazards and Earth System Sciences, 17*(9), 1659-1682. doi:10.5194/nhess-17-1659-2017

Booth, A., Papaioannou, D., & Sutton, A. (2011). *Systematic approaches to a successful literature review*: Sage.

Botzen, W. J. W., & Bouwer, L. M. (2016). Weather Indicators for Insured Hailstorm Damage to Motor Vehicles and Potential Climate Change Impacts. *Geneva Papers on Risk and Insurance-Issues and Practice, 41*(3), 512-527. doi:10.1057/gpp.2015.16

Bouwer, L. M. (2011). Have disaster losses increased due to anthropogenic climate change? *Bulletin of the American Meteorological Society, 92*(1), 39-46.

Bouwer, L. M. (2013). Projections of Future Extreme Weather Losses Under Changes in Climate and Exposure. *Risk analysis, 33*(5), 915-930. doi:10.1111/j.1539-6924.2012.01880.x

Brevik, R., Aall, C., & Rød, J. K. (2014). Pilotprosjekt om testing av skadedata fra forsikringsbransjen for vurdering av klimasårbarhet og forebygging av klimarelatert naturskade i utvalgte kommuner. In: Vestlandsforksningsrapport.

Changnon, S. A., Changnon, J. M., & Hewings, G. D. (2001). Losses caused by weather and climate extremes: A national index for the United States. *Physical Geography, 22*(1), 1-27.

Cheng, C. S., Li, Q., Li, G., & Auld, H. (2012). Climate change and heavy rainfall-related water damage insurance claims and losses in Ontario, Canada. *Journal of Water Resource and Protection, 4*(02), 49.

Cortes, M., Turco, M., Llasat-Botija, M., & Llasat, M. C. (2018). The relationship between precipitation and insurance data for floods in a Mediterranean region (northeast Spain). *Natural Hazards and Earth System Sciences, 18*(3), 857-868. doi:10.5194/nhess-18-857-2018

Crichton, D. (1999). The risk triangle. *Natural disaster management*, 102-103.

Cutter, S., Emrich, C., Morath, D., & Dunning, C. (2013). Integrating social vulnerability into federal flood risk management planning. *Journal of Flood Risk Management, 6*(4), 332-344.

Donat, M. G., Pardowitz, T., Leckebusch, G. C., Ulbrich, U., & Burghoff, O. (2011). High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Natural Hazards and Earth System Sciences, 11*(10), 2821-2833. doi:10.5194/nhess-11-2821-2011

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB journal, 22*(2), 338-342.

Falconer, R., Cobby, D., Smyth, P., Astle, G., Dent, J., & Golding, B. (2009). Pluvial flooding: new approaches in flood warning, mapping and risk management. *Journal of Flood Risk Management, 2*(3), 198-208.

Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*: Cambridge University Press.

Grahn, T., & Nyberg, L. (2017). Assessment of pluvial flood exposure and vulnerability of residential areas. *International Journal of Disaster Risk Reduction, 21*, 367-375. doi:10.1016/j.ijdrr.2017.01.016

Grothmann, T., & Reusswig, F. (2006). People at risk of flooding: why some residents take precautionary action while others do not. *Natural Hazards, 38*(1-2), 101-120.

Hanak, T., & Korytarova, J. (2014). Risk zoning in the context of insurance: Comparison of flood, snow load, windstorm and hailstorm. *Journal of Applied Engineering Science, 12*(2), 137-144. doi:10.5937/jaes12-6098

Haug, O., Dimakos, X. K., Vardal, J. F., Aldrin, M., & Meze-Hausken, E. (2011). Future building water loss projections posed by climate change. *Scandinavian Actuarial Journal*(1), 1-20. doi:10.1080/03461230903266533

Hauge, Å., Flyen, C., Venås, C., Aall, C., Kokkonen, A., & Ebeltoft, M. (2018). *Attitudes in Norwegian insurance companies towards sharing loss data* (11). Retrieved from

Holand, I. S., Lujala, P., & Rød, J. K. (2011). Social vulnerability assessment for Norway: a quantitative approach. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 65*(1), 1-17.

Hurford, A., Parker, D., Priest, S., & 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.

IPCC. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. *Sciences, 10*, 97-104.

Jacso, P. (2005). As we may search-Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *CURRENT SCIENCE-BANGALORE-, 89*(9), 1537.

Kaźmierczak, A., & Cavan, G. (2011). Surface water flooding risk to urban communities: Analysis of vulnerability, hazard and exposure. *Landscape and Urban Planning, 103*(2), 185-197. doi:https://doi.org/10.1016/j.landurbplan.2011.07.008

Kim, Y.-O., Seo, S. B., & Jang, O.-J. (2012). Flood risk assessment using regional regression analysis. *Natural Hazards, 63*(2), 1203-1217.

Koks, E. E., Jongman, B., Husby, T. G., & Botzen, W. J. W. (2015). Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. *Environmental Science & Policy, 47*, 42-52. doi:https://doi.org/10.1016/j.envsci.2014.10.013

Kousky, C., & Michel-Kerjan, E. (2017). Examining Flood Insurance Claims in the United States: Six Key Findings. *Journal of Risk and Insurance, 84*(3), 819-850. doi:10.1111/jori.12106

Kreibich, H., & Thieken, A. H. (2008). Assessment of damage caused by high groundwater inundation. *Water Resources Research, 44*(9).

Kron, W. (2005). Flood risk= hazard• values• vulnerability. *Water International, 30*(1), 58-68.

Labonnote, N. (2017). Stormwater-related databases–Review and Recommendations. *Klima 2050 Report*.

Labonnote, N., Hauge, Å., & Siversten, E. (2018). A climate services perspective on Norwegian stormwater-related databases. *Climate Services (submitted manuscript)*.

Leal, M., Ramos, C., & Pereira, S. (2018). Different types of flooding lead to different human and material damages: the case of the Lisbon Metropolitan Area. *Natural Hazards, 91*(2), 735-758. doi:10.1007/s11069-017-3153-3

Lohmann, D., & Yue, F. (2011). CORRELATION, SIMULATION AND UNCERTAINTY IN CATASTROPHE MODELING. In S. Jain, R. Creasey, & J. Himmelspach (Eds.), *Proceedings of the 2011 Winter Simulation Conference* (pp. 133-145).

Lujala, P., Lein, H., & Rosvoldaune, R. (2014). Quantifying vulnerability to flooding induced by climate change: The case of Verdal, Norway. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 68*(1), 34-49.

Mechler, R., & Bouwer, L. M. (2015). Understanding trends and projections of disaster losses and climate change: is vulnerability the missing link? *Climatic Change, 133*(1), 23-35.

Mechler, R., Bouwer, L. M., Linnerooth-Bayer, J., Hochrainer-Stigler, S., Aerts, J. C., Surminski, S., & Williges, K. (2014). Managing unnatural disaster risk from climate extremes. *Nature Climate Change, 4*(4), 235.

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article" Assessment of economic flood damage". *Natural Hazards and Earth System Sciences, 10*(8), 1697.

Merz, B., & Thieken, A. H. (2009). Flood risk curves and uncertainty bounds. *Natural Hazards, 51*(3), 437-458.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.

Moncoulon, D., Labat, D., Ardon, J., Leblois, E., Onfroy, T., Poulard, C., . . . Quantin, A. (2014). Analysis of the French insurance market exposure to floods: a stochastic model combining river overflow and surface runoff. *Natural Hazards and Earth System Sciences, 14*(9), 2469-2485. doi:10.5194/nhess-14-2469-2014

Moncoulon, D., Veysseire, M., Naulin, J.-P., Wang, Z.-X., Tinard, P., Desarthe, J., . . . Deque, M. (2016). Modelling the evolution of the financial impacts of flood and storm surge between 2015 and 2050 in France. *International Journal of Safety and Security Engineering, 6*(2), 141-149. doi:10.2495/SAFE-V6-N2-141-149

Peng, J., Kesete, Y., Gao, Y., Shan, X., Davidson, R. A., Nozick, L. K., & Kruse, J. (2014). *Math modeling to support regional natural disaster risk management.* Paper presented at the 10th

U.S. National Conference on Earthquake Engineering: Frontiers of Earthquake Engineering, NCEE 2014, July 21, 2014 - July 25, 2014, Anchorage, AK, United states.

Petticrew, M., & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*: John Wiley & Sons.

Rød, J. K., Opach, T., & Neset, T.-S. (2015). Three core activities toward a relevant integrated vulnerability assessment: validate, visualize, and negotiate. *Journal of Risk Research, 18*(7), 877-895.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). *Global sensitivity analysis: the primer*: John Wiley & Sons.

Shirley, W. L., Boruff, B. J., & Cutter, S. L. (2012). Social vulnerability to environmental hazards. In *Hazards Vulnerability and Environmental Justice* (pp. 143-160): Routledge.

Sorensen, J., & Mobini, S. (2017). Pluvial, urban flood mechanisms and characteristics Assessment based on insurance claims. *Journal of Hydrology, 555*, 51-67. doi:10.1016/j.jhydrol.2017.09.039

Spekkers, M. H., Clemens, F., & ten Veldhuis, J. A. E. (2015). On the occurrence of rainstorm damage based on home insurance and weather data. *Natural Hazards and Earth System Sciences, 15*(2), 261-272. doi:10.5194/nhess-15-261-2015

Spekkers, M. H., Kok, M., Clemens, F., & 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. doi:10.5194/hess-17-913-2013

Spekkers, M. H., Kok, M., Clemens, F., & ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences, 14*(9), 2531-2547. doi:10.5194/nhess-14-2531-2014

Surminski, S., Aerts, J., Botzen, W. J. W., Hudson, P., Mysiak, J., & Perez-Blanco, C. D. (2015). Reflections on the current debate on how to link flood insurance and disaster risk reduction in the European Union. *Natural Hazards, 79*(3), 1451-1479. doi:10.1007/s11069-015-1832-5

Tobias, R. D. (1995). *An introduction to partial least squares regression.* Paper presented at the Proceedings of the twentieth annual SAS users group international conference.

Torgersen, G., Bjerkholt, J. T., Kvaal, K., & Lindholm, O. G. (2015). Correlation between extreme rainfall and insurance claims due to urban flooding – Case study fredrikstad, Norway. *Journal of Urban and Environmental Engineering, 9*(2), 127-138. doi:10.4090/juee.2015.v9n2.127138

Torgersen, G., Rød, J. K., Kvaal, K., Bjerkholt, J. T., & Lindholm, O. G. (2017). Evaluating Flood Exposure for Properties in Urban Areas Using a Multivariate Modelling Technique. *Water, 9*(5), 318. doi:ARTN 318

10.3390/w9050318

Wilson, M. F., O'Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy, 30*(1-2), 3-35.

Wobus, C., Lawson, M., Jones, R., Smith, J., & Martinich, J. (2014). Estimating monetary damages from flooding in the United States under a changing climate. *Journal of Flood Risk Management, 7*(3), 217-229.

Yang, X., Grönlund, A., & Tanzilli, S. (2002). Predicting Flood Inundation and Risk Using Geographic Information System and Hydrodynamic Model. *Geographic Information Sciences, 8*(1), 48-57. doi:10.1080/10824000209480573

Yazdani, N., Dowgul, R. W., & Manzur, T. (2010). Deficiency analysis of coastal buildings toward storm damage reduction. *Journal of Performance of Constructed Facilities, 24*(2), 128-137.

Zhou, Q., Panduro, T. E., Thorsen, B. J., & Arnbjerg-Nielsen, K. (2013). Verification of flood damage modelling using insurance data. *Water Science and Technology, 68*(2), 425-432. doi:10.2166/wst.2013.268

Zischg, A. P., Mosimann, M., Bernet, D. B., & Rothlisberger, V. (2018). Validation of 2D flood models with insurance claims. *Journal of Hydrology, 557*, 350-361. doi:10.1016/j.jhydrol.2017.12.042