Understanding trans-basin floods in Germany

Data, information and knowledge

Steffi Uhlemann

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2013/6886/
URN urn:nbn:de:kobv:517-opus-68868
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-68868

Institute of Earth- and Environmental Sciences

Understanding trans-basin floods in Germany

Data, information and knowledge

Cumulative Dissertation

for the degree of
Doctor of Natural Sciences (Dr. rer. nat.)
in Geoecology

submitted to the
Faculty of Mathematics and Natural Sciences
at the University of Potsdam, Germany

by
Steffi Uhlemann

Submitted: April 2013

Contents

List of Figures

List of Tables

Abstract

Large Central European flood events of the past have demonstrated that flooding can affect several river basins at the same time leading to catastrophic economic and humanitarian losses that can stretch emergency resources beyond planned levels of service. For Germany, the spatial coherence of flooding, the contributing processes and the role of trans-basin floods for a national risk assessment is largely unknown and analysis is limited by a lack of systematic data, information and knowledge on past events.

This study investigates the frequency and intensity of trans-basin flood events in Germany. It evaluates the data and information basis on which knowledge about trans-basin floods can be generated in order to improve any future flood risk assessment. In particu-lar, the study assesses whether flood documentations and related reports can provide a valuable data source for understanding trans-basin floods.

An adaptive algorithm was developed that systematically captures trans-basin floods using series of mean daily discharge at a large number of sites of even time series length (1952-2002). It identifies the simultaneous occurrence of flood peaks based on the exceedance of an initial threshold of a 10 year flood at one location and consecutively pools all causally related, spatially and temporally lagged peak recordings at the other locations. A weighted cumulative index was developed that accounts for the spatial extent and the individual flood magnitudes within an event and allows quantifying the overall event severity. The parameters of the method were tested in a sensitivity analysis. An intensive study on sources and ways of information dissemination of flood-relevant publications in Germany was conducted. Based on the method of systematic reviews a strategic search approach was developed to identify relevant documentations for each of the 40 strongest trans-basin flood events. A novel framework for assessing the quality of event specific flood reports from a user's perspective was developed and validated by independent peers. The framework was designed to be generally applicable for any natural hazard type and assesses the quality of a document addressing accessibility as well as representational, contextual, and intrinsic dimensions of quality.

The analysis of time-series of mean daily discharge resulted in the identification of 80 trans-basin flood events within the period 1952-2002 in Germany. The set is dominated by events that were recorded in the hydrological winter (64%); 36% occurred during the summer months. The occurrence of floods is characterised by a distinct clustering in time. Dividing the study period into two sub-periods, we find an increase in the percentage of winter events from 58% in the first to 70.5% in the second sub-period. Accordingly, we find a significant increase in the number of extreme trans-basin floods in the second sub-period.

A large body of 186 flood relevant documentations was identified. For 87.5% of the 40 strongest trans-basin floods in Germany at least one report has been found and for the most severe floods a substantial amount of documentation could be obtained. 80% of the material can be considered grey literature (i.e. literature not controlled by commercial publishers).

The results of the quality assessment show that the majority of flood event specific reports are of a good quality, i.e. they are well enough drafted, largely accurate and objective, and contain a substantial amount of information on the sources, pathways and receptors/consequences of the floods. The inclusion of this information in the process of knowledge building for flood risk assessment is recommended.

Both the results as well as the data produced in this study are openly accessible and can be used for further research.

The results of this study can contribute to an improved spatial risk assessment in Germany. The identified set of trans-basin floods provides the basis for an assessment of the chance that flooding occurs simultaneously at a number of sites. The information obtained from flood event documentation can usefully supplement the analysis of the processes that govern flood risk.

Zusammenfassung

Große Hochwasser der Vergangenheit in Europa haben gezeigt, dass Überschwemmungen gleichzeitig in mehreren Flusseinzugsgebieten vorkommen und katastrophale ökonomische und humanitäre Schäden hervorrufen können, welche das Katastrophenmanagement über das vorgesehene Maß beanspruchen. Für Deutschland ist das räumlich gleichzeitige Auftreten, die zugrunde liegenden Pro-zesse und der Beitrag von flussgebietsübergreifenden Hochwasserereignissen für eine nationale Risikobe-wertung weitgehend unbekannt. Die Analyse wird durch das Fehlen einer systematischen Daten-, Infor-mations- und Wissensgrundlage zu vergangenen Ereignissen begrenzt.

Die vorliegende Studie untersucht die Häufigkeit und Intensität von flussgebietsübergreifenden Hochwasserereignissen in Deutschland. Die Studie untersucht, welche Daten- und Informationsgrundlage nötig ist, um die einer Hochwasserrisikobewertung zugrunde liegenden Prozesse besser zu verstehen. Im speziellen wird untersucht, inwieweit Hochwasserereignisdokumentationen und verwandte Berichte als eine weitere Datenquelle für ein verbessertes Prozessverständnis genutzt werden können.

Es wurde ein adaptiver Algorithmus entwickelt, welcher flussgebietsübergreifende Hochwasser systematisch auf der Basis von Zeitreihen täglichen mittleren Abflusses an vielen Stationen und mit gleicher Zeitreihenlänge (1952-2002) erfasst. Das gleichzeitige Auftreten von Hochwasserscheiteln wird anhand der Überschreitung eines 10-jährlichen Hochwassers an einer Stelle und der Zuordnung räumlich und zeitlich versetzter, aber kausal mit diesem in Verbindung stehender weiterer Scheitel erreicht. Es wurde ein gewichteter kumulativer Index entwickelt, welcher es erlaubt die Gesamtereignisstärke zu quantifizieren. Der Index berücksichtigt sowohl die räumliche Ausdehnung als auch die jeweilig erreichten Abflussspitzen im Hochwassergebiet. Die Parameter der Methode wurden einer Sensitivitätsanalyse unterzogen. Im Rahmen der Studie wurden die Quellen und Verbreitungswege von hochwasserrelevanten Veröffentlichungen in Deutschland tiefgründig erhoben. Basierend auf der Methode des systematischen Re-views wurde ein Ansatz für eine strategische Suche nach relevanten Veröffentlichungen zu den 40 stärks-ten flussgebietsübergreifenden Hochwasserereignissen entwickelt. Für die Bewertung der Qualität von ereignisspezifischen Hochwasserberichten wurde ein neuartiger Ansatz entwickelt und durch unabhängige Experten validiert. Der Ansatz wurde so entwickelt, dass er grundsätzlich auch auf andere Naturgefahren übertragbar ist. Die Qualität eines Dokuments wird anhand von zugänglichkeits-, formellen, kontextuellen und intrinsischen Qualitätskriterien bewertet.

Die Analyse der Zeitreihen mittleren täglichen Abflusses führte zur Identifizierung eines Sets von 80 flussgebietsübergreifenden Hochwasserereignissen in Deutschland (1952-2002). Das Set wird von Ereignissen dominiert, welche im hydrologischen Winterhalbjahr auftreten (64%). 36% aller Ereignisse treten im Sommerhalbjahr auf. Das Auftreten von Hochwasser ist durch eine Häufung von Ereignissen in bestimmten Perioden gekennzeichnet. Die Unterteilung der Untersuchungszeitraums in zwei Teilperioden zeigt, dass der Anteil von Winterereignissen signifikant von 58% in der ersten zu 70,5% in der zweiten Periode zunimmt. Daran geknüpft lässt sich eine signifikante Zunahme hin zu schwereren flussgebietsübergreifenden Ereignissen in der zweiten Periode feststellen.

Im Rahmen der Studie konnten 186 hochwasserrelevante Berichte identifiziert werden. Für 87,5% der 40 untersuchten Ereignisse konnte wenigstens ein Bericht und für die schwersten Ereignisse eine erhebliche Anzahl an Berichten identifiziert werden. 80% des Materials kann als Grauliteratur eingestuft werden, d.h. als Veröffentlichungen welche nicht durch kommerzielle Verleger publiziert wird.

Die Ergebnisse der Qualitätsbewertung zeigen, dass die Mehrheit der ereignisspezifischen Hochwasserberichte von guter Qualität ist, d.h. die Berichte sind in ausreichender Qualität verfasst, größtenteils korrekt und objektiv und beinhalten eine substantielle Menge an Informationen zu den Ursachen, Verläufen, betroffenen Objekten und Schäden eines Ereignisses. Es wird empfohlen diese Informationen in die Wissenssynthese für die Hochwasserrisikobewertung einfließen zu lassen.

Sowohl Ergebnisse als auch Daten dieser Studie sind so publiziert, dass sie öffentlich zugänglich sind und für weitere Forschungsfragen genutzt werden können.

Die Ergebnisse unserer Forschung tragen zu einer Verbesserung der großräumigen Hochwasserrisikoanalyse in Deutschland bei. Das erstellte Ereignisset stellt die Grundlage für die Bewertung der Wahrscheinlichkeit dar, dass Hochwasser an mehreren Orten gleichzeitig auftreten. Der Informationsgehalt von Hochwasserereignisberichten stellt eine wichtige Ergänzung für die Analyse der das Hochwasserrisiko bedingenden Faktoren dar.

Chapter 1:

Introduction

1.1 The scale of floods

The European Union (European Union, 2007) defines a flood as "the temporary covering by water of land not normally covered by water." This covering of water can emerge from many sources and can occur on various pathways, i.e. as pluvial (heavy rainfall), fluvial (floods along rivers), ground water or coastal flooding (storm surge), or as a result of technical failure (dam break) or cascading events (e.g., floods due to landslide; tsunami due to submarine earthquake). This work will be concerned with river or fluvial flooding.

River floods extend over a period of time, with durations lasting from a few hours to several weeks, and affect a certain space, ranging from single catchments to several basins. Depending on the space-time scales of the source the flood response can be either spatially confined and spontaneous leading to local or small regional floods, or widespread and long lasting leading to major regional floods (Fig. 1.1). In as much as spatially small floods can exhibit very high magnitudes and lead to high losses at the sites affected the sum of damages and the options for emergency response are mostly manageable. On the contrary, widespread flooding can lead to catastrophic economic and humanitarian losses that can stretch emergency resources beyond planned levels of service (Lamb et al., 2010). For insurance and re-insurance the accumulation of losses is critical and needs to be taken into account when assessing premiums. In Central Europe prominent extreme river floods such as the August 2002 flood in the Elbe and Danube, the Pentecost flood of May 1999 in the Danube, the Odra flood 1997 and the two consecutive Rhine floods of December 1993 and January 1995 have demonstrated that that flooding can at the same time affect more than one federal state and often more than one river basin. We term these floods trans-basin floods.

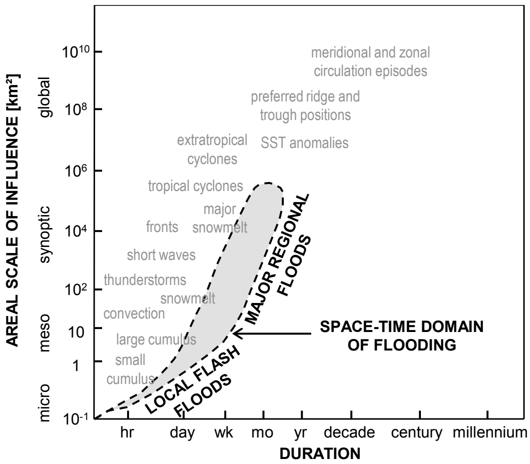

Figure 1.1: Characteristic space and time scales of floods and flood causing mechanisms (adapted from Hirschboeck, 1988).

1.2 Understanding floods

When we look at the chain of causalities that leads to river flooding (Fig. 1.2) we find that the system is complex requiring knowledge on a number of subsystems, their feedbacks and their space and time scales. This includes atmospheric processes like preconditioning factors that lead to certain moisture states in the catchments and initiating factors like the intensity and extent of the rainfall field or the amount of snowmelt. The catchment processes of runoff generation and concentration of excess water lead to the formation of a flood wave and its propagation in the river network. There the wave may be amplified by superposition of flood waves from tributaries or attenuated by processes of dispersion. Where bankfull flow conditions are exceeded the flood plain is inundated resulting in adverse consequences to the exposed receptors. Many of the governing processes are characterised by random behaviour and the flood event is the consequence of either one of the subsystems taking on an extreme state or due to an unfavourable combination of states in the subsystems.

Complexity is added by the dynamics of the coupled human-water system as institutions and governance processes interact with hydrological processes in floodplains to influence the frequency and severity of floods (Di Baldassarre et al., 2013a). Vice versa, structural defences such as flood retention basins, river weir cascades, or dike protection trigger changes in the exposure of receptors in the floodplain. For example, the process of building and raising levees on the one hand decreases the vulnerability of the urbanized hinterland, however, it also introduces a shift from frequent flooding to rare, but potentially catastrophic flooding (Di Baldassarre et al., 2013b). Ad-hoc flood defence measures and non-structural defences such as early warning or crisis management plans can help reducing the immediate flood impact. Other direct or indirect anthropogenic influences on the climate (changes in precipitation and temperature due to global warming) or catchments (e.g., land use change altering the runoff behaviour) propagate along the chain of processes and may also alter the flood

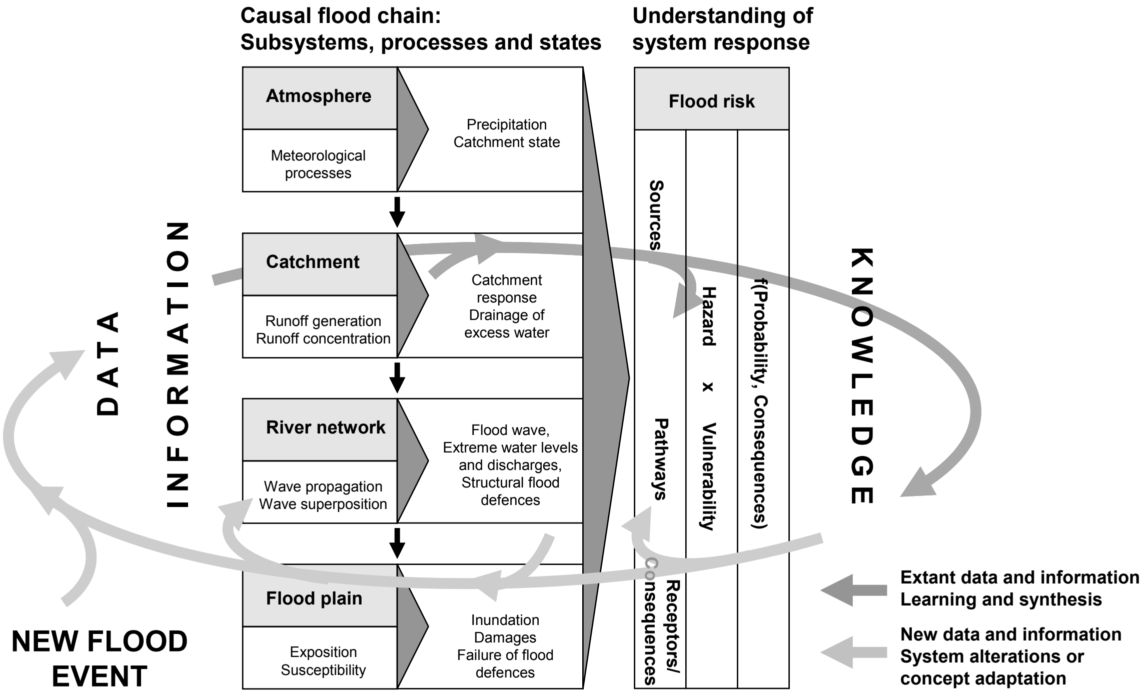

Figure 1.2: The iterative process of knowledge synthesis for understanding flood risk. Systems and processes adapted from Merz (2006); Flood risk conceptions as e.g. summarized in FloodSite glossary (Samuels and Gouldby, 2009).

hazard on a longer time scale and in consequence the flood risk.

When we review the documentations of some of the trans-basin flood events of the recent past in Central Europe, we can identify some particular characteristics of widespread floods: (1) The damages caused by the events are of a national or even international significance. (2) Within one event the pattern of flood magnitude and damages can vary enormously in space with some regions experiencing extreme flood flow conditions and others only moderate flows. (3) The regions affected by a trans-basin flood can vary from event to event. (4) Trans-basin floods can be caused by different weather systems and preconditions in the catchments and the flood impact can be significantly influenced by flood mitigation and adaptation measures.

However, when we move beyond the information on the most recent floods, we learn that trans-basin floods are little understood in the risk they pose on the national or larger scale compared to the tremendous damages they cause. In particular, current approaches do not address the question of the chance that many sites experience severe flooding at the same time. We also learn that little systematic knowledge is available to identify the causalities and consequences of these floods. This knowledge needs to account both for the commonalities of these floods as well as the individual factors that lead to certain patterns of flood impact. The question arises: Which data and information are needed in order to understand trans-basin floods and how can we assess the risk of trans-basin floods?

1.3 Spatial flood risk

Any flood risk assessment has to answer three questions (Merz and Thieken, 2004): (1) What are the possible flood scenarios? (2) How likely is their occurrence? (3) What are the consequences in case a scenario occurs? In principle, any approach is viable that allows determining all the relevant damage scenarios and their associated probability. Usually the probabilities of the hazard are used to infer the probability of the damage encountered during an event.

For an at-site risk analysis the probability of the hazard can be obtained from a frequency analysis of extremes in series of river flow or stage data which may be complemented by information from historical events. For long time series the probabilities can be estimated from the empirical frequencies. However, other combinations of the flood generating processes may result in an extreme event that has not been observed before, necessitating the extrapolation beyond the set of observations.

For trans-basin floods the analysis has to move from point estimates to many possible combinations of outcomes at a multitude of sites. An empirical estimate would require a database of past flood events that explicitly accounts for the spatial pattern of flood magnitude. Current collections of flood events fall short of a systematic approach in the inclusion of events and they do not provide the geographical referencing of the flood magnitudes encountered (Sturm et al., 2001). However, river flow series at a multitude of sites provide the opportunity of inferring the simultaneousness of peak flows in order to create a set of past events. Alternatively, the series of river flow data can be scrutinized directly for the statistical dependence between locations which in turn are used to create a synthetic set of events in a stochastic modelling framework. This leads to a very large and complex joint probability problem (Lamb et al., 2010) which is a recent field of research (e.g., Keef et al., 2009a, b). In both cases the probability estimate needs to be prescribed to the pattern of flood magnitude raising the need for an appropriate quantitative measure on which the frequency analysis can be performed.

A commonly used alternative approach is the simulation of the entire rainfall-runoff process (both event based or continuous). Either series of observed rainfall are used directly or a stochastic weather generator is used to produce patterns of rainfall that are input to a deterministic hydrological-hydraulic model chain. This approach is commonly used in the (re-)insurance industry (RMS, 2006; Reiche, 2012) and has found few national applications, e.g. the RASP-project (Hall et al., 2005) or FLORIS-project (FLORIS, 2005) in England/Wales and the Netherlands, respectively. Recently, numerous steps in the deterministic model chain have been started to be replaced by probabilistic approaches, e.g., Vorogushyn et al. (2010) developed a probabilistic dike breach model. In this way the spatial dependencies and temporal dynamics of the events are preserved. However, the approach involves the modelling of a number of complex processes. For the large scale the amount of data needed to drive this model chain is challenging and models on this scale need too simplify the physical processes and are therefore associated with high uncertainties.

Any assessment of flood risk and in particular of the probabilities of flood scenarios includes assumptions. In fitting an extreme value distribution to an empirical sample of events it is assumed that this sample is the result of a stationary process, i.e. that the events are independent and identically distributed and that the events belong to the same parent population. Considering the outlined complexities of the flood generating processes these assumptions have to be carefully checked. This concerns both the assessment of the dominant processes, i.e. whether the spatial dependence structure of the events is different for particular flood typologies, and whether the occurrence of the events is independent and stationary, i.e. whether any clustering of events in time or a trend can be observed.

1.4 Knowledge synthesis

The European Committee for Standardization's guide to good practice in knowledge management states (Centre for Evidence-Based Conservation, 2004): "Knowledge is the combination of data and information, to which is added expert opinion, skills and experience, to result in a valuable asset which can be used to aid decision making." The process of knowledge building is in turn an iterative cycle of learning and adapting (Fig. 1.2), e.g., improving models and concepts by means of new or better data and information that in turn lead to a better understanding of processes and possibly an improved knowledge of the system. In terms of flood risk we have already highlighted the complexities of the coupled human-water system where the knowledge of the risk leads to management actions that feedback on the single subsystems, either through direct alterations in the real world or through the improved conception of the modelling of the (sub)systems.

Considering these complexities, the role of past and current natural hazard events as learning examples for an improved risk management has been stressed at many instances (e.g., Hübl et al., 2002; IRDR, 2011). These events provide the specific information about the sources and pathways of the flood as well as the information about the receptors being affected and the resulting consequences. However, data, information and knowledge on past flood events are fragmented amongst a multitude of stakeholders (science, public administration, organisations, society) and disciplines (e.g., hydrology, social sciences, economics). The acquisition and analysis of hydro-meteorological data is a viable option for the German case as data is accessible for research purposes and for a dense network of stations. However, data on the flood impact, either inundation or damages, and information on the numerous individual factors that determine flood impact (like the operation of flood defences) cannot be obtained easily.

Since flood risk assessment is at foremost a subject of high societal relevance, it is inherently a subject of governmental action. A large body of authorities is concerned with the management of this risk and the planning of measures for flood loss reduction. Authorities are the primary body of (observational) data and information production and can claim to hold a high level of long-term technical experience. They are responsible for maintaining the national station network and are therefore equipped with first hand access to and control of the quality of the data. Consequently, authorities rather than the scientific community are involved in the production of reconnaissance reports in the aftermath of a flood event. These reports often not only address the hazard part but also provide a more holistic and possibly more detailed view on the event including sources, pathways, receptors and/or consequences. Therefore, beside observations and model outputs, flood documentations are another potential source of information for understanding trans-basin floods. However, flood event documentations are frequently excluded from the scientific knowledge synthesis as the mostly technical documents produced by authorities are commonly disseminated through other means than the scientific publication routes and as the credibility of these sources is perceived to be low. So far, the value and use of flood event documentation for flood risk research has remained largely unexplored.

1.5 Objective

Large Central European flood events of the past show that flooding can affect several river basins at the same time and that the spatial pattern of flood magnitude in one single event is highly heterogeneous. For Germany, the spatial dependence of flooding, the contributing processes and consequently their role for a national risk assessment is largely unknown and no systematic analysis has been presented so far. The objective of this study is to investigate flood events that at the same time affect more than one river basin and to provide the data and information basis on which knowledge about this type of floods can be generated in order to improve any future flood risk assessment. We address the objective by the following research questions:

1. The assessment of the accumulated risk over large spatial domains like that of entire countries requires not only studying the local extremes, but also extremes that affect many sites simultaneously. The question is: How can trans-basin floods be identified in Germany?
2. Classical flood frequency analysis allows extrapolating to probabilities beyond the observed sample of extremes (given by units of river discharges at a site of interest) by fitting extreme value distributions to the sample and under the assumption that the sample of extremes is homogeneous, and that the events are independent and identically distributed. For trans-basin floods the frequency estimate has to move from the point scale to a spatial pattern of flood magnitude. The question arises: How can we define a measure of severity for trans-basin floods that can be used for a frequency analysis and what are the frequency characteristics ob the observed sample of trans-basin floods?
3. The expression of the flood impact is the result of a complex interaction of inextricably linked processes, ranging from atmospheric to catchment processes to flood protection and ad-hoc mitigation measures. Understanding floods in their sources, pathways, receptors and consequences requires the availability of sufficient data on each aspect. However, barriers in access to, scarcity, or incompleteness of observations of the physical processes and damage influencing factors limit the options for a systematic analysis. In contrast, event specific documentations provide information on a range of flood aspects, however they are commonly perceived as being of low accessibility and credibility. The question arises: Can and should flood event documentations of the recent past be used as another source of data for understanding trans-basin flood events in their sources, pathways and consequences on particular receptors?

1.6 Tasks and structure

Formulating the objective as understanding trans-basin floods and as outlined by the three research questions, this work aims to contribute to fundamental aspects of data provision for a national flood risk assessment. We evaluate the objective for the specific case of trans-basin floods in Germany. The work has been split into three different parts each being addressed in a separate chapter (chapters 2-4):

First of all, it is necessary to identify trans-basin flood events in Germany in the past and to develop a complete and consistent set of events. The set of events is to provide the basis for any following assessment of the processes and event characteristics as well as any spatial risk assessment. For the purpose, a systematic and robust method needs to be developed that allows identifying causally related flood peaks at a large number of sites and that covers a long period. Further, in order to compare events, a measure needs to be developed that quantifies the overall event severity. Basic assessments of the nature of these floods as well as their frequency characteristics can be provided based on this set and the data used.

Second, it is necessary to understand these floods in terms of their sources, pathways, receptors, and consequences and the data and information available for this purpose needs to be identified. In this work the use of event documentations is explored as another set of data. For that purpose a systematic approach is to be used in order to quantify the size of the body of literature available on trans-basin floods in Germany and to identify the type of sources, their main characteristics in terms of accessibility, and hence to draw conclusions on their potential applicability in flood risk research.

Third, the issue of credibility of the previously identified sources needs to be addressed. In particular, for event documentations that are produced outside of the scholarly publication routes a framework needs to be developed that allows assessing the overall quality of the document with respect to the specific task of understanding trans-basin floods.

Finally, the main findings of the previous tasks will be summarized in a concluding chapter (chapter 5). The chapter will provide a synthesis over the achievements reached in understanding trans-basin floods and will provide an outlook into the next steps. In particular, the potential of the information obtained to expand the knowledge of and to build a knowledge base on trans-basin flood events in Germany.

The thesis is complemented by two data publications for the chapters 3 and 4.

Chapter 2:

A consistent set of trans-basin floods in Germany between 1952 – 2002

A consistent set of trans-basin floods in Germany between 1952 - 2002

Abstract

Floods that affect many sites simultaneously can pose great challenges in the co-ordination of flood disaster management actions, as well as for the insurance and re-insurance industry, since this type of flooding leads to an accumulation of losses and the risk assessment needs to be extended to a concept representing the spatial risk of flooding. The assessment of the accumulated risk, especially over large domains, requires an analysis of the spatial and temporal coherence of flooding. For Germany the extent of spatial dependence of flooding is largely unknown and no systematic analysis has been performed so far. In this paper, we present a methodology that is capable of capturing the simultaneous occurrence of flooding using multiple series of mean daily discharge. For the first time we present a complete and consistent set of trans-basin floods in Germany for the period between 1952 and 2002. Each flood is characterised by a specific value for the timing, the location and the magnitude of discharges within the entire river network. We propose a measure for quantifying the overall event severity considering both the heterogeneous spatial extent as well as the locally varying magnitudes of a trans-basin flood. In total, we identify 80 trans-basin floods in the entire time period. The set is dominated by events that were recorded in the hydrological winter (64%); 36% occurred during the summer months. 32 events affected more than one third of the entire river network. These most severe events are predominantly winter events. Dividing the study period into two sub-periods, we find an increase in the percentage of winter events from 58% in the first to 70.5% in the second sub-period. Accordingly, we find a significant increase in the number of extreme trans-basin floods in the second sub-period. A natural extension of this study is the quantification of the spatial and temporal dependencies in a multivariate framework. This framework needs to be supported by a flood typology based on the analysis of the physical processes relevant in the genesis of trans-basin floods.

Published as:

Uhlemann, S., Thieken, A.H., and Merz, B. (2010): A consistent set of trans-basin floods in Germany between 1952-2002, Hydrology and Earth System Sciences, 14, 1277-1295, doi:10.5194/hess-14-1277-2010.

2.1 Introduction

Flood events extend over a period of time, with durations lasting from a few hours to several weeks, and affect a certain space, ranging from single catchments to several basins. In this study we focus on river floods that, caused by the same hydrometeorological processes, affect multiple basins and that are consequently of durations exceeding several days. We term them trans-basin floods. Extreme river floods such as the well-known events of August 2002, May 1999, January 1995 and December 1993 have demonstrated that flooding in Germany can at the same time affect communities in more than one federal state and often in more than one river basin. This poses problems in the development of flood disaster management strategies since the co-ordination of flood activities in case of emergencies is subject to the federal state ministries and inter-state agencies are largely endowed with advisory competences only. Further, in the insurance and re-insurance industry this type of flooding leads to an accumulation of losses. Therefore, the risk concept needs to be extended to a concept representing the spatial risk of flooding, i.e. scenarios have to be developed that capture a large number of individual risks during a single event.

Any flood risk assessment has to answer three questions (Merz and Thieken, 2004): (1) What are the possible flood scenarios? (2) How likely is their occurrence? (3) What are the consequences in case a scenario occurs? The risk to a particular site can be estimated by using a wide range of well-established methods. However, the assessment of the accumulated risk, especially over large domains like that of entire countries, requires not only studying the local extremes, but also their spatial and temporal coherence. Methods to derive the spatial dependence and the accumulated risk of flooding are still in their infancy, and for Germany the extent of spatial dependence of flooding is largely unknown. This paper starts filling this gap by providing answers to the first question of risk assessment - the possible flood scenarios, and by laying the foundation to answering the second question, namely through an analysis of the past occurrences of trans-basin floods and their overall severity.

Past floods provide a range of scenarios and through their systematic analysis inferences can be drawn on the inherent spatial and temporal dependencies. Rodda (2005) demonstrated that past floods can be employed to derive synthetic trans-basin flood scenarios for the main rivers in the Czech Republic. He assembled a set consisting of the 11 most severe historical floods from 1935 to 2002 using reports and local knowledge. For each flood, series of daily mean discharge were acquired within a temporal envelope around the days of the flood peaks for 25 gauges. The author identified three different patterns of flooding, based on which 30 synthetic flood events were derived through a qualitative analysis of the historical events and knowledge of the hydrology of the river basins.

The question arises though, whether a chronology of floods assembled from documented data sufficiently reflects the range of flooding needed to infer future flood scenarios that do not only replicate a limited set of possible combinations. Further, if the frequency of simultaneous flooding is to be assessed, the frequencies of past occurrences need to be known. In the Central European context numerous studies have presented collections of flood events, e.g. the works of Glaser (2001), Stanescu (2002), Glaser and Stangl (2003), Jacobeit et al. (2003), Pohl (2004), Barnolas and Llasat (2007), Barredo (2007), and Müller et al. (2009) to name a few. However, uncertainties remain about the completeness with which flood events are identified and especially the issue of geographical referencing and the overall magnitude of floods can be addressed qualitatively only (e.g. Sturm et al., 2001).

From a frequentist's point of view, a series of observed extreme events provides the basis for an extreme value analysis by choosing either block maxima or peaks-over-threshold to define the sample within the observation period. Then distribution functions can be fitted to these samples allowing an extrapolation beyond the range of the observed values, i.e. to more extreme events. The problem with trans-basin floods is that no straightforward measure can be used to quantify the overall magnitude of the event. The only regionally unbiased measure would be the total area of inundation or the total damage caused by a flood. But no consistent information of the actual inundation area or damages is available for historical events. The problem is that historical data on flood losses are neither comprehensive nor standardised throughout Europe (Mitchell 2003). Blong (2003) presents a list of damage related indices and recently Barredo (2009) presented an attempt to normalise flood losses in Europe, using events listed in either the Natural Hazards Assessment Network (NATHAN, http://mrnathan.munichre.com/) of the Munich Re or the Emergency Events Database (EM-DAT, http://www.emdat.be/), maintained at Université catholique de Louvain. These archives contain selections of floods that exceeded certain thresholds of impact (monetary damage and number of casualties). Damage in turn is a product of the hazard itself and the vulnerability of people and assets towards flooding. In the course of the centuries, changes in the vulnerability, and there especially in the exposure, are evident. This leads to a bias in the archives, as floods of the same intensity may have caused little damage in the past but led to severe damages (and therefore their inclusion in the archives) in later times.

Only recently a somewhat complete catalogue on floods has begun to be maintained (Dartmouth Flood Observatory, available at http://www.dartmouth.edu/~floods) dating back to 1985, which mainly uses satellite data.

It is the aim of this study to develop both a method that allows the derivation of a consistent set of past trans-basin floods and that provides an indicator to compare the severity of these floods based on their spatial pattern of flood magnitude. We deduce trans-basin floods by jointly analysing series of mean daily discharge at many sites for the simultaneous occurrence of peak discharges. Discharge measurements are integrals of the meteorological, catchment, and channel processes and are therefore suitable to capture the temporal and spatial evolution of flood events. The method and the severity indicator are developed based on data of catchments in Germany; however, they are transferable to other regions.

The first important issue to address is the identification of suitable events in the multiple series of discharge. To allow flood peaks at different locations and at particular time lags to be identified as being mutually related, the identification procedure requires an appropriate definition of thresholds for the flood magnitude and dynamic. So far, only few studies have investigated the spatial coherence of flooding, making use of different combinations of these thresholds, dependent on the aim of the study and the data available.

Besides the before mentioned study of Rodda (2005), Merz and Blöschl (2003) identified mutually dependent annual maxima in Austria for the purpose of developing a flood typology. A maximum time lag of one day and a maximum distance between adjacent catchments of 50 km are applied to determine those annual maxima that belong to the same flood event. The spatial spread of a flood is then expressed as an ellipsoid that is built around the centroids of the affected catchments. However, not all rivers that are affected during the event may have experienced their annual maximum flood and in many cases flooding below the annual maximum occurred at more sites. The method favours floods which propagate along one particular river or affect only neighbouring catchments. Flood occurrences which affect multiple basins at a time lag of more than one day or floods of spatially dispersed origins cannot be captured. This is especially the case when not the entire basin can be used for analysis due to data unavailability. For example, in the case of the severe summer floods of 1954 and 2002 the flood triggering Vb cyclone first led to severe flooding in parts of the Danube basin and thereafter to flooding in the Elbe basin. Both occurrences are evidently related; meanwhile for the German context the missing data for the upper Elbe basin (Czech Republic) prevent a causal correlation through particular distance measures.

Keef et al. (2009a) develop a measure to capture the spatial dependence in extreme river discharges using a dense network of time series of mean daily discharge in Great Britain. For a range of return levels T they estimate the expected proportion of sites at a range of distances d from any gauged site that exceed the pth quantile during an event. Even though a range of distances are tested, the overall extent of a flood is limited to the proposed radii d. Also, the choice of the quantile largely influences how many sites are identified as responding simultaneously. Naturally, for very high quantiles the spatial coherence is rather low. No indications can be given on whether the all causally related flood peaks have been captured.

Our study extends the previous approaches in various ways: First of all, we aim at identifying all trans-basin floods in a period between 1952 and 2002 using as many sites as possible. We take a holistic approach on each flood event, meaning that rather than applying a strict quantile approach we are more interested in the system response. Therefore, we aim at identifying all peaks at all sites which are mutually related. For that purpose we relax the stringent thresholds of magnitude by defining that only one site needs to exceed the pth quantile. All other sites are checked for the simultaneous occurrence of flood peaks irrespective of the quantile reached.

In our approach we do not impose any distance measure to infer the mutual relation of flood peaks or set any a priori restrictions on flood extent. Since the spatial spread of a flood is confined to the river network, the flood characteristics derived from discharge measurements are not a space-filling phenomenon (Gottschalk et al., 2006). For the German context the experience from recent floods shows that flooding can develop over long distances and in a spatially asymmetric manner. This is largely due to the complexity and form of the river network, which is very diverse for the German river basins, and due to spatially varying and often multiple origins of floods. Therefore, distance measures like the Euclidean distance between catchment centroids or gauging stations cannot be employed in a straightforward manner to infer the dependence of flood peaks on a trans-basin scale. Rather, we will make use of the time-space correlation in the evolution of a flood event. Therefore, in this study the spatial distances encountered in the study area are treated implicitly by considering the timing of the peaks as a function of space. This allows capturing also dependencies amongst peak recordings at spatially remote locations that are not directly connected by a river network and/or belong to catchments which are not adjacent.

As outlined earlier, floods evolve over a period in time and extend over a certain area. That means flood peaks which can be attributed to the same flood event will be recorded at a time lag τ at the various stations. This time lag corresponds to the drift velocity of the weather system, the concentration time in the catchments and also the spatial distances between sites expressed as length and complexity of the river network including phenomena such as superposition of flood waves at confluences or retention of the flood wave due to dike breaches. In this study, flood peaks are considered as simultaneous when they occur less than a predefined time apart. This time is based on the physical understanding of the processes during flood development and it implicitly considers the spatial evolution of a flood event within as well as across basins. Hence, floods will be defined strictly in terms of the timing of their peak discharges.

A second central aspect of this work is the development of a measure of event severity for trans-basin floods. The translation of discharges into inundation area on a large scale using hydraulic approaches still poses severe demands on data and computing power and was hence not an option for this study. Alternatively, the measure of Keef et al. (2009a) gives a statistical indication on the spatial dependence in extreme river discharges. Mapping this measure for all stations provides an overview on where flooding tends to occur spatially coherently at various levels of recurrence but it cannot be used for an event based assessment.

In this study we will lend on the concept of flood impact by assessing the length of the river network potentially inundated during the event. We thereby define the potential for inundation as the exceedance of bankfull flow at any particular site. The events identified in the first step are characterised by information on the flood peaks per site. To derive an indicator that captures the spatial pattern of heterogeneous flood magnitudes, both the spatial extent and the magnitude have to be taken into account. Further, to ensure the generic applicability this measure has to be unspecific to the set of gauged sites. We therefore introduce a simple scheme to regionalise the site specific discharge peaks to the entire river network and normalise the discharge values by a threshold indicating, whether a river stretch has actually been ‘in flood’, hence, has exceeded the bankfull discharge. Conditional on this criterion, the measure is formed as a cumulative weighted function of the dimensionless, normalised flood magnitude and spatial extent.

We have chosen not to express flood magnitude through return periods. During many events peak discharges of very high magnitude can be expected. Return periods are estimates of the flood magnitude and the quality of that estimate at any site depends on the quality of the fit to an extreme value distribution and on the length of the underlying annual maximum series, here 51 years. For large quantiles these estimates are associated with high uncertainties due to an extrapolation beyond the range of the data. Further, expressing the event severity as the mean over all return periods from the affected gauges (even if weighted) would too easily be interpreted as the ‘true’ return period of the entire event. This is certainly not the case. The estimation of the return period of a trans-basin flood must be based on a frequency analysis over the entire event set and even more – moving away from an empirical estimate – must be based on an assessment of many (a thousand or more) synthetically generated flood scenarios that take into account the spatial dependencies amongst the flood peaks within the events.

The paper is structured as follows: Section 2.2 gives an overview on the data available for this study. Section 2.3 provides a detailed description of the methods developed for the identification of trans-basin floods and the indicator for flood severity. The resulting event set is presented in Section 2.4 together with an analysis of the main characteristics of trans-basin floods in Germany. To prove the reasonability of the thresholds applied for event identification, and to allow an adaptation of the methodology to different objectives or data, a sensitivity analysis is presented in Section 2.5. Section 2.6 discusses the results and concludes on the main findings of this study.

2.2 Study area and data

Series of mean daily discharge were obtained for the German parts of the river basins Rhine, Ems, Weser, Danube, Elbe, and Odra from various water authorities in Germany and the Global Runoff Data Centre (GRDC). The basins are located in Central Europe and

Table 2.1: Basin characteristics.

Basin	Number of gauges	Length of river network (km)	(%)
Danube	46	3019.6	25.5
Rhine	48	4440.1	37.4
Elbe	35	2301.4	19.4
Weser	28	1580.8	13.6
Ems	4	274.4	2.4
Odra	2	176.4	1.7
Sum	162	11918	100

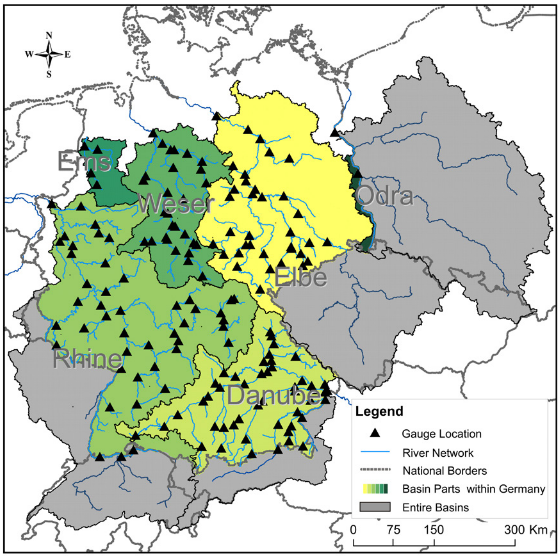

Figure 2.1: Basins, locations of gauges, and river network used in the study.

cover the territories of five states, which are Germany (with exception of the coastal zones), Switzerland, France, Austria (where the headwater catchments of the Danube and Rhine are located), Czech Republic and Poland which contain large parts of the Elbe and most parts of the Odra, respectively (see Fig. 2.1).

The gauges were selected based on catchment size whereby the catchment had to exceed a drainage area of 500 km². This threshold was chosen to exclude local floods from the study. The choice of gauges was further constraint by outweighing the best possible spatial coverage of the investigation area and the longest coherent time period. The period of the water years from 1952 to 2002 (a water year ranging from 1st November to 31st October) was chosen, since for this period a maximum of gauges with continuous measurements could be identified. All time series were checked for data errors and missing values. Series with more than two complete water years missing were excluded from the analysis.

162 gauges were selected with a mean catchment area of 16,880 km² and the maximum area comprising 159,300 km² (Rhine). A high percentage of nested catchments are included in the dataset. The stations are not evenly distributed across the basins, with less dense coverage in the Rhine basin and dense networks in the Danube and Weser basins (see Table 2.1). Fig. 2.1 illustrates the spatial distribution of all gauges and the relevant basins.

The river network used in this study is the pan-European River and Catchment database developed under the CCM2-activity (Catchment Characterisation and Modelling) of the Joint Research Centre (Vogt et al., 2007). The CCM2 dataset offers the stream network for Europe and explicitly allows the deduction of the river topology, also making reference to hierarchical structures like the Strahler system. It is therefore an ideal basis for any regionalisation of point discharge data to the entire river network. The resolution and quality of the data is sufficient for the scale of this study.

2.3 Method

Let F denote a trans-basin flood event and $Q_i(t)$ the discharge series at the sites i, with $i = 1,\ldots,N$ and N the total number of sites available, here 162, and t in daily time steps in the period of 51 years between 1952 to 2002. As outlined in the introduction, each event can be fully described by its spatial pattern of flood magnitude. We treat the spatial distances encountered in the study area implicitly by considering the timing of the peaks. This also allows capturing dependencies amongst peak recordings at distant locations that are not directly connected by a river network and/or belong to catchments which are not adjacent. We take a holistic view on each event, meaning that we are interested in the system response at each site given an identified extreme discharge at any other site. An extreme value is defined as at least one gauge exceeding the pth quantile of its annual maxima series during the event. The system response at any other site is given by a peak discharge that significantly deviates from the ordinary variance of $Q_i(t)$ around its running mean. Let Ω capture all sites i that have exhibited a significant peak within the time interval Δ. Then each flood F can be described by a set of time dependent discharge peaks (the peak value denoted by the superscript *P*) according to

$$\mathrm{F_e} = \left\{ Q_i^P (t^P) \right\} \quad \forall i \in \Omega, \quad t^P \in \Delta, e = 1, \ldots, E \tag{1}$$

with *E* the total number of identified floods. Δ expresses the overall duration of the event within which the maximum observed peaks at sites *i* are defined as mutually related or in other words, as having occurred simultaneously. The duration is supposed to last from the first identified peak to the last.

The identification procedure requires an appropriate definition of thresholds for the desired flood magnitude and for the flood dynamic allowing flood peaks at different locations and at particular time lags as being identified as mutually related. The procedure is comprised of the following four steps each being described in more detail in a subsection:

1. In each series of mean daily discharge those days are identified where a peak above or equal to the *p*th quantile (peaks over threshold, POT) has been observed. This step is introduced by an explanation of the general procedure for identifying peaks in series of daily mean discharge.
2. Subsequently, all peaks that belong to the same event need to be pooled. For each day during which a POT was recorded at any site the discharge series of all N stations are interrogated in a temporal envelope around that day to find significant peak discharges. The steps involved are:
   - The definition of an appropriate temporal envelope.
   - The evaluation of the significance of the peaks found.
   - The definition of an inter-event time criterion which ensures independence between consecutive trans-basin floods.
3. After pooling, by applying a simple regionalisation scheme, we translate the point values of discharge peaks into a spatial extent variable that allows us to identify those events that are of a trans-basin extent. The share of the network that is potentially affected by inundation is used as indicator for the spatial extent of each event.
4. Using the regionalisation scheme introduced in point 3, we can formulate a weight cumulative indicator that is a function of the spatial pattern of maximum observed discharges in the river network.

2.3.1 Peaks over threshold

As outlined, we are interested in the system response at any particular site towards a given extreme event at any other site. We adopt the POT approach to identify those days in the spatial series of discharge during which at least one gauge exceeded a discharge threshold *u*. Starting from these days we apply a pooling procedure to identify any peak discharge that can be mutually related to this event.

Discharge peaks for series of mean daily discharge *Q(t)* at any observation site *i* can be identified by evaluating the increments between the preceding and following discharge per day.

$$dQ(t) = Q(t) - Q(t-1) \tag{2}$$

Let z*(t)* depict the series of increments between the daily values of *Q(t)* simplifying it to positive, zero or negative differences.

$$z(t) = \begin{cases} +1, & \mathit{if}\ dQ(t) > 0 \\ 0, & \mathit{if}\ dQ(t) = 0 \\ -1, & \mathit{if}\ dQ(t) < 0 \end{cases} \tag{3}$$

Let *g* denote the index to the daily time series over the 51 year period (a total of 18,628 days). Then peaks can be identified according to three cases (see Eq. 4). Firstly, a clearly peaking hydrograph where in the course of a day the discharge reaches its maximum and after that immediately ceases (case 1, Eq. 4). These peaks are typical for fast-reacting catchments. For slowly reacting catchments or downstream observation sites these peaks might be prolonged and the flood crest may persist over a day or two until the water level falls again (cases 3 and 4, Eq. 4).

$$Q^P(t^P) = \begin{cases} & \mathit{if}\ \left[z(t_g) = 1 \wedge z(t_{g+1}) = -1 \right] \\ Q(t_g), & \mathit{if} \begin{bmatrix} z(t_g) = 1 \wedge z(t_{g+1}) = 0 \\ \wedge\, z(t_{g+2}) = -1 \end{bmatrix} \\ & \mathit{if} \begin{bmatrix} z(t_g) = 1 \wedge z(t_{g+1}) = 0 \\ \wedge\, z(t_{g+2}) = 0 \\ \wedge\, z(t_{g+3}) = -1 \end{bmatrix} \end{cases} \tag{4}$$

$Q^P(t^P)$ then contains the set of discharge peaks contained in the time series *Q(t)*. For cases 2 and 3 in Eq. 4, the first day of the sequence of increments is used as the day of the peak occurrence.

We choose the 10-year flood (Q10) as threshold u to define a minimum event severity above which a flood impact can be expected. The Q10 is commonly requested for risk maps as the first out of three or four zones in risk mapping (as for example proposed by the EU-Flood directive, or e.g. realised in the Rhine atlas (IKSR, 2001)), delineating areas of frequent flooding. In the insurance industry, objects which are situated within the exposure zone of Q10 are usually not considered as insurable (Kron 2005).

We estimate the 10-year flood using the 90th percentile of the series of annual maxima (AMS). The AMS are extracted from each series of $Q(t)$ choosing the annual maximum peak per water year. Then, the generalised extreme value distribution (GEV) is fitted by the method of L-moments to each of the AMS. From the fitted data the discharge threshold u of the 10-year flood is estimated. Considering the length of the time series used, this estimate is assumed to be reliable.

Using u, we identified all POT in each series of mean daily discharge. To ensure independence of the selected events, the minimum time lag between flood peaks are set according to Svensson et al. (2005). There, dependent on the catchment size, time lags are set to 5 days for catchments <45,000 km², to 10 days for catchments between 45,000 and 100,000 km², and to 20 days for catchments >100,000 km², respectively.

Evaluating the POT of all N gauging stations, those days in the time period during which at least one POT was recorded were compiled into a set of dates expressed as

$$\mathrm{D} = t^P \;\Big|\; \#\left\{Q_i^P(t^P) \geq u\right\} \geq 1 \tag{5}$$

Now, each of the days in D is used as a starting point for identifying trans-basin events. Let j denote the index to D, with j = 1,…, M. Based on the Q10 thresholds a pooling method has to be applied such that mutually dependent peaks at all sites i can be identified and grouped together.

2.3.2 Pooling of mutually dependent peaks

Temporal envelope W

Flood events that affect numerous catchments do not lead to peak discharges at the same day at all observation sites. For example, a low pressure system passes through the study area on a south-west to north-easterly track, leading to rainfall at is front and an influx of warm air. The movement may take a couple of days from its first appearance in the study area until it finally leaves the area or the precipitation field has rained out. This rain field meets specific catchment conditions, like e.g. the presence of a snow cover or saturated soils. The processes of runoff concentration and the respective concentration time determine the time lag from the initial precipitation and/or snowmelt to the recording of a flood peak at the respective gauges. The flood wave in turn propagates downstream at a particular speed, leading to lagged flood peaks downstream.

In this study the design of the temporal envelope W is intended to reflect the flood dynamic at a trans-basin scale. Several features in the flood generation of large events are in common to most types of floods and allow the definition of a general time window for flood peak detection. For example, Rodda (2005) uses a 10 day envelope around the date when the peak discharge of a historical flood event was recorded. This time window is then interrogated to find the maximum mean daily discharge for each station for each event. Keef et al. (2009a) test the bivariate temporal dependence for a selection of sites in Great Britain on a range of lags up to a maximum of 50 days. They find that 96% of all pairs have estimates of extreme dependence within a lag of $|\tau| \leq 3$ days. Larger lags occur if any of the pairs is a slow responding catchment.

Here, a more differentiated approach is applied. We infer the mutual dependence of peaks starting from any day D_j. Since this is likely not the point in time, when flooding has started, we check in both directions in time to find mutually related peaks. Therefore, the time window W around each day j will be composed of a pre-POT and a post-POT time lag. The pre-POT time lag reflects the drift velocity of a flood producing weather system over the study area and the time of concentration within the catchments, mostly resulting in time lags of a few days. The speed (which may include stationary conditions) and direction of the triggering weather system determine the point in time and the spatial order (succession) at which runoff generation is induced in the catchments. Runoff can thereby result from snow melt, rainfall, or both (rain on snow). Typically, frontal systems that are embedded in the westerlies pass over the study area on a west-easterly track in less than 24 hours. In case of quasi-stationary conditions a frontal system may persist and lasting precipitation (with varying intensities) over a couple of days can occur. Due to the wide spatial coverage of these systems, at the beginning of an event a number of spatially far apart catchments will react simultaneously or within a few days. The same accounts for sequences of disturbances which cross over the study area in short intervals. The time

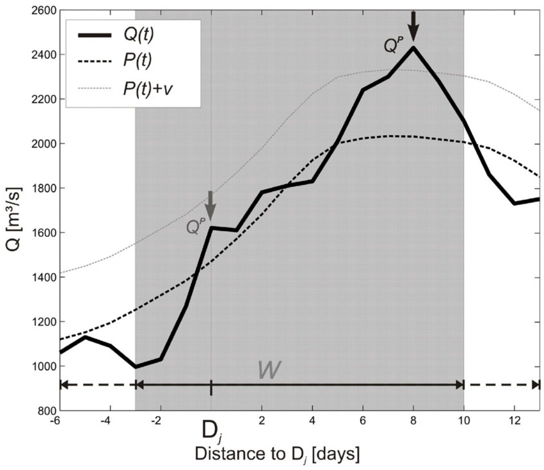

Figure 2.2: Procedure of identifying significant peak discharges. *Q(t)* depicts the hydrograph of an arbitrary gauge *i* in the interval *W*(grey shaded area) around a day D_j. The lowpass function *P(t)* is calculated using a moving average of 13 days. The significance of each peak is evaluated by calculating the 90th percentile *v* of the residuals between *Q(t)* and *P(t)*. From the identified peaks Q^P (arrow) only those are considered that exceed $P(t)+v$ and that are located within *W*. In the example this holds true for the second peak (black arrow).

of concentration, i.e. in this study the time until the first gauge reports a peak discharge, is mainly determined by the catchment size, the catchment characteristics, and the initial catchment state. Depending on the catchment size, various studies have used a time lag of one (catchments between 500-5000 km²) to three days (catchments >20 000 km²) to link the flood triggering circulation pattern with discharges (Duckstein et al., 1993; Frei et al., 2000; Bárdossy and Filiz, 2005; Petrow et al., 2007, 2009). We choose a pre-POT interval of 3 days.

The time for the propagation of a flood wave in the channel to the most downstream location leads to a much longer post-POT time lag to be considered. In the process of propagation, the flood wave can be either amplified at confluences due to simultaneous arrival of flood waves from tributaries, be maintained or dampened. For the first two cases, the flood peak can be monitored over long distances and the travel time of the flood wave can take several days, e.g. in the Elbe a flood wave recorded in Dresden (hence flooding originates in Czech Republic) reaches the outlet of the basin (Neu Darchau) approximately 8 to 10 days later. For the flood event in March 1988 the time lag between the flood crest at the lower Elbe and the preceding Q10 corresponded to exactly 10 days. Consequently, we set the post-POT interval of the temporal envelope to 10 days, and

W = 3, -2,…,0,…,+9,+10.

For any day D_j, all of the *N* discharge series are checked for the presence of distinct discharge peaks at any time lag $\tau \in W$, using the increment based approach as described in Eqs. (2) to (4). To also capture peaks at the very first or last day of the temporal envelope (i.e. in case 3 Eq. (4)), for computational accuracy it is necessary to extend the time window by three days at the beginning and end of the interval.

Nonetheless, peaks will only be considered if they fall within *W*.

Significance of peaks

The significance of each peak identified in *W* is evaluated by analysing whether it significantly deviates from the normal fluctuations of *Q(t)*. Rather than applying a global threshold based on quantiles of the annual maximum series per site, we chose to evaluate each peak detected in *W* locally by comparing it to the general behaviour of the hydrograph.

For that purpose we calculate the moving average *P(t)* (kernel width of 13 days) for the entire discharge series. The residuals between the observed runoff *Q(t)* and *P(t)* are then calculated at each time step producing a series of nearly normally distributed noise. We use the 90[th] percentile of this series as a threshold *v* that, if exceeded, reflects those periods in *Q(t)* during which the hydrograph significantly deviates from the normal fluctuations. We interpreted this as a reaction to a distinct surplus of water in the river network. Also it allows identifying the flood peak rather than any other minor peaks in the hydrograph. If more than one significant peak is detected in the interval *W*, the one of highest discharge is used for further analysis. The procedure can be expressed as

$$Q^P_{i\in\Omega}(t^P) = \max_{\tau\in W} \left\{ Q^P_i(D_j+\tau) \,\middle|\, Q^P_i(D_j+\tau) - P_i(D_j+\tau) \geq \upsilon \right\} \quad \forall j=1,\ldots,M \;\; \forall i=1,\ldots,N \tag{6}$$

If no significant peak is detected for site *i*, then $Q_i^P(t^P)$ is treated as missing value. The set Ω then comprises only those sites *i*, for which a significant peak discharge has been identified.

Figure 2.2 illustrates the procedure of peak identification for a typical flood hydrograph. Applying the interval *W* (grey shade) around any arbitrary day D_j,

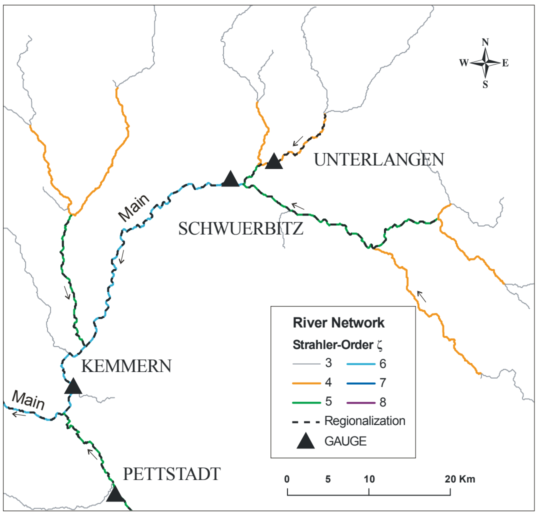

Figure 2.3: Scheme applied to regionalise the discharge peaks at each gauge to the river network of the upper or intermittent catchment, respectively. Those parts of the network which are used for the regionalisation are highlighted by the dashed lines. The entire river network used as a result of the regionalisation is shown in Fig. 2.1.

two peaks are identified. Clearly, the first peak is located on the rising limb of the hydrograph and should not be considered. The conditional of Eq. (6) allows this distinction and only the second peak is identified as significant and is chosen for the further analysis.

Independence of events

The last step in the identification of trans-basin floods is the definition of an inter-event time criterion δ that allows defining the independence of events. Often, when a flood is evolving, a number of gauges will exceed the threshold u within short succession. Then the time lag between some consecutive POT-dates can be of a few days only and peaks detected for day D_j may overlap with those identified for day D_{j+1}. These peaks are mutually dependent and need to be pooled into the same event group. The inter-event time criterion δ has to be defined such that no peaks are pooled which belong to separate events. The independence of significant peaks identified for each consecutive entry in D is determined by evaluating the time lag between the last peak identified for day D_j and the first peak identified for day D_{j+1}, with

$$\delta_j = \min_{i\in\Omega}(D_{j+1} + \tau_i) - \max_{i\in\Omega}(D_j + \tau_i)\,. \tag{7}$$

Unless δ_j does not exceed one day, all entries of D_j and D_{j+1} are pooled into the same event and the duration of the event grows with the additional dates. This step basically reflects the topological behaviour of the river network, once a flood is progressing downstream. In this way it is secured that also those peaks at downstream locations are picked that do not exceed the threshold of the 10-year flood and that are farthest apart from the place of the onset of an event. E.g. for a typical summer Vb type, it is possible to capture flood peaks at the basin outlet of the Elbe since they can be temporally linked to the last occurring POT in the study area which would be likely reported in the mountainous headwater catchments of the Ore-Mountains. If δ_j > 1day, the events are treated as independent.

Using the series of δ_j, an index set K can be created according to the following conditional

$$\mathrm{K} = j \big| \delta_j > 1 \tag{8}$$

where K defines the points in time at which peaks identified for any date $D_{j\in K}$ and those identified for the consecutive date D_{j+1} are treated as independent. e denotes the subscript to K, with $e = 1,\ldots,E$, and E gives the total number of flood events detected in the discharge series of N gauges within the period of 1952 – 2002. In case more than one significant peak at a particular site i is present within one event, the larger peak is used for further analysis.

Now, each event is fully described by the timing and the magnitude of the discharge peak at each site, as given by Eq. (1). The overall duration Δ of the pooled event is defined to last from the day of the first pooled peak to the day of the last pooled peak.

2.3.3 Exclusion of spatially small events

Since trans-basin floods are the subject of this work, all events E identified in the procedure described above need to be checked for their actual impact, i.e. their spatial extent. We regionalise the point observations of peak discharges per event to the entire river network and choose a truncation level of 10% potentially inundated rivers to limit the event set to reasonably large events. This threshold translates into at least 1200 river kilometres affected during the event. In most cases this results in flooding in more than one basin. For comparison, the overall length of the river network for each basin is given in Table 2.1.

Therefore, a discharge threshold κ has to be identified which reflects whether the discharge peak has led to inundation, hence has exceeded the bankfull discharge. Since no data on river morphology or water levels at bankfull conditions were available, the threshold is estimated using a quantile approach.

Several ranges of recurrence intervals for bankfull discharge have been proposed in literature for natural rivers. Using an annual series approach, Petit and Pauquet (1997) estimate the recurrence interval for the bankfull discharge of 30 gravel bed rivers in the north east of France. A linear relationship between catchment size and recurrence interval leads to an estimated range from about 1.8 to 2.5 years for catchments of 500 km² and larger. Using partial duration series, the results even indicate recurrence intervals in the range of 0.8 to 1.5 years. Although recent studies (Navratil et al., 2006; Wilkerson, 2008) show a strong relationship to the type of the river bed and also the methodology used to estimate bankfull discharge, an average of 2 years recurrence interval Q2 seems to be a reasonable threshold.

A simple regionalisation scheme is applied that encounters for both the stream length and the stream complexity. The hierarchical ordering of river networks as developed by Strahler (1964) provides a good measure to describe these features. Depending on the Strahler-order ζ at the gauge *i,* we regionalise the discharges to the river stretches upstream of a gauge. In case of nested catchments the length of the river network of the intermittent catchment is considered. For those gauges at lower parts of streams, like e.g. the Rhine ($\zeta \geq 6$), only those parts of the network are considered that are attributed with orders of $\zeta \geq 5$. Discharges at gauges with orders $5 > \zeta \geq 4$ are regionalised to upstream rivers of $\zeta \geq 4$ and those of $\zeta = 3$ only to their respective same order. Figure 2.3 shows the regionalisation for the example of the upper Main.

A few exceptions had to be considered: For the eastern tributary of the Elbe, the river Havel, time series were only available at downstream locations. The regionalisation of discharges at locations where large parts of the basin are located upstream with no further gauges is highly uncertain. Therefore, the most upstream gauge of the Havel (Ketzin) was assigned to only the same ordered river network ($\zeta = 6$). The basin of the Odra is almost completely located in Poland. Only the confluences of the two major tributaries are located on German territory and two gauges are situated there. Here, the discharge peaks of Hohensaaten and Eisenhüttenstadt were regionalised to the same ζ (8 and 6, respectively) river stretches both up- and downstream to allow an adequate consideration of the river length.

This kind of regionalisation can only be a very rough estimation of the true effect of each flood. Nonetheless, for the scale considered in this study and the dense network of gauges this is deemed reasonable.

Using simple GIS queries, the cumulative length l of the river network can be calculated for each gauge i according to the above mentioned procedure. Then the ratio of the catchment length to the total length of the entire river network provides the weights $\lambda_i = \frac{l_i}{\sum_i l_i}$.

The overall affected length of the river network L is conditional on the exceedance of the threshold level κ and is given in percent according to

$$L = \sum_{i \in \Omega} \lambda_i * 100 \Big| \, Q_i^P \geq \kappa \qquad (9)$$

We choose to truncate the event set at a level of $L <$ 10%. This level keeps only those events in the set that apply for a trans-basin analysis, i.e. in most cases more than one basin being affected. For an extreme value analysis the set can be chosen to be truncated at any higher level L which we leave to the practitioner to decide. We will present the results of this study differentiating between several extent classes to analyse possible differences in the processes that lead to trans-basin floods.

2.3.4 Event severity

In this study, the overall event severity is defined as a function of the spatial pattern of maximum observed discharges in the river network. Making the same impact-based assumptions as denoted in the previous section, we weight the normalised peak discharges by the median annual flood (Q2) and derive the weighted cumulative discharge indicator S according to

$$S = \sum_{i \in \Omega} \left\{ \lambda_i * \frac{Q_i^P}{\kappa} \right\} \Big| \, Q_i^P \geq \kappa \qquad (10)$$

The sum is formed only over those sites and their respective river length where the threshold for bankfull discharge κ had been exceeded. The normalisation to the inundation threshold or median flood $\kappa = Q2$ allows comparing the magnitude of a flood at each gauge, and the sum then serves as an indicator for both event magnitude and spatial extent.

Table 2.2: List of trans-basin flood events in the period 1952 – 2002. Classes of spatial extent are highlighted in orange (class 1, $L \geq 50\%$), yellow (class 2, $50\% > L \geq 33\%$), green (class 3, $33\% > L \geq 20\%$) and blue colours (class 4, $20\% > L \geq 10\%$). Winter events are displayed in black fonts, summer events in red fonts.

Rank	Start		End	Rank	Start		End	Rank	Start		End
1	15 3 1988	-	11 4 1988	29	28 2 1987	-	7 3 1987	57	30 5 1984	-	7 6 1984
2	22 2 1970	-	4 3 1970	30	5 12 1981	-	12 12 1981	58	17 3 1979	-	19 4 1979
3	23 1 1995	-	7 2 1995	31	11 5 1999	-	27 5 1999	59	19 5 1965	-	30 5 1965
4	31 12 1981	-	18 1 1982	32	20 3 2002	-	26 3 2002	60	28 5 1995	-	5 6 1995
5	29 10 1998	-	11 11 1998	33	26 1 2002	-	2 2 2002	61	12 1 1993	-	20 1 1993
6	9 3 1981	-	26 3 1981	34	7 2 1958	-	22 2 1958	62	8 8 1970	-	15 8 1970
7	3 3 1956	-	17 3 1956	35	18 12 1965	-	28 12 1965	63	21 7 1980	-	28 7 1980
8	20 12 1993	-	31 12 1993	36	28 6 1958	-	18 7 1958	64	5 7 1955	-	20 7 1955
9	4 2 1980	-	14 2 1980	37	22 2 1957	-	4 3 1957	65	8 2 1961	-	15 2 1961
10	30 12 1986	-	10 1 1987	38	11 4 1970	-	3 5 1970	66	5 3 1979	-	12 3 1979
11	15 1 1968	-	25 1 1968	39	19 7 1981	-	30 7 1981	67	28 8 1995	-	6 9 1995
12	2 7 1954	-	31 7 1954	40	25 5 1983	-	31 5 1983	68	17 3 1957	-	8 4 1957
13	10 4 1994	-	27 4 1994	41	29 12 1974	-	7 1 1975	69	4 6 1981	-	12 6 1981
14	9 8 2002	-	24 8 2002	42	13 7 1956	-	31 7 1956	70	5 12 1961	-	17 12 1961
15	1 6 1965	-	20 6 1965	43	25 3 1987	-	1 4 1987	71	24 1 1994	-	4 2 1994
16	25 2 2002	-	4 3 2002	44	25 12 1954	-	8 1 1955	72	17 6 1991	-	25 6 1991
17	23 1 1982	-	8 2 1982	45	11 5 1970	-	18 5 1970	73	30 7 1977	-	8 8 1977
18	8 12 1974	-	21 12 1974	46	19 1 1986	-	23 1 1986	74	22 8 1970	-	30 8 1970
19	31 12 1993	-	9 1 1994	47	4 12 1960	-	13 12 1960	75	25 6 1953	-	7 7 1953
20	24 2 1958	-	3 3 1958	48	9 2 1970	-	13 2 1970	76	8 8 1978	-	13 8 1978
21	24 12 1967	-	3 1 1968	49	18 3 1970	-	2 4 1970	77	30 4 1980	-	8 5 1980
22	6 2 1984	-	11 2 1984	50	16 3 1994	-	27 3 1994	78	21 12 1991	-	29 12 1991
23	10 1 1955	-	27 1 1955	51	22 7 1966	-	31 7 1966	79	28 6 1966	-	7 7 1966
24	9 4 1983	-	20 4 1983	52	13 2 1966	-	2 3 1966	80	22 9 1968	-	28 9 1968
25	20 2 1999	-	26 2 1999	53	17 6 1979	-	28 6 1979				
26	15 2 1990	-	20 2 1990	54	1 6 1961	-	20 6 1961				
27	2 3 1999	-	7 3 1999	55	22 5 1978	-	31 5 1978				
28	22 2 1997	-	3 3 1997	56	31 1 1961	-	5 2 1961				

2.4 Results

Applying the methodology with the chosen set of parameters, a total of 80 trans-basin events are detected within the years 1951 to 2002. Table 2.2 gives an overview on the events with ranks assigned in the order of event severity according to the indicator *S*. For each event the first and the last day with a significant peak discharge are given. Classifying the event set by the spatial extent, four event severity classes are further distinguished. Class 1 contains extreme events that affected more than 50% of the entire stream network. A total of 14 events belong to this class (highlighted in orange in Table 2.2). Class 2 (yellow) contains all events which affected between one third and 50% of the network, 18 in total, and another 21 events affected between one fifth and one third of the network (class 3 - green). The majority of events (27) exhibited a spatial extent just above the threshold level of 10% and up to a maximum of 20% (class4 - blue). The set is dominated by events (64%) that were recorded in the hydrological winter (1^{st} November to 30^{th} of April), 36% occurred during the summer months (1^{st} May to 31^{st} October), which are marked by red fonts in Table 2.2.

In the following, the main characteristics of the events are further analysed. Figure 2.4 gives an overview on the characteristic features of each event. The events in Fig. 2.4 are sorted in descending order according to the index *S,* and the according rank number is given on the x-axis. Using this rank number the event date can be obtained by cross checking in Table 2.2. For an easier overview, summer events are marked by red fonts and winter events by black fonts.

Focussing on the event severity first, it can be noted that *S* declines nearly exponentially with an initial sharp decline of event severity within the first two classes, that is, those events which affected $1/3^{rd}$ or more of the entire river network. The spatial extent of each event *L* (in %) is also displayed, highlighting the relative contribution of spatial extent and event magnitude to the indicator *S*. The farther both lines are apart,

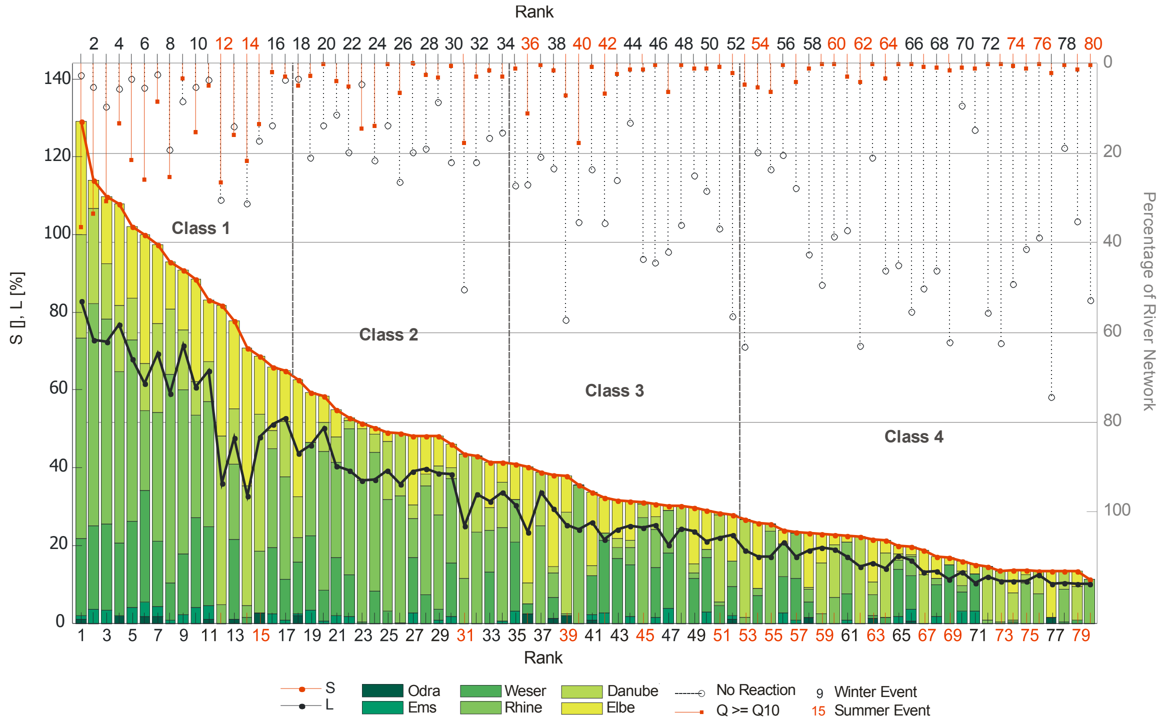

Figure 2.4: Characteristics of the identified trans-basin events, ranked according to the weighted cumulative discharge index *S* (solid red line). The share each major basin (Odra, Ems, Weser, Danube, Rhine, Elbe) takes in the formation of the index S is indicated by the green to yellow shading. The percentage of the river network affected is given by *L* (solid black line). On the secondary x and y-axis both the number of gauges is given at which the 10-year flood was exceeded (dashed red stems) and the number of those that did not show any significant reaction during the event (dashed black stems). The grey dashed vertical lines indicate the approximate division of the event set into the 4 classes of spatial extent as given in Table 2.2.

the higher is the share of river stretches that have been affected by severe flooding. For a better orientation the stems at the top of Fig. 2.4 indicate the share of the river network that has been either affected by discharges exceeding Q10 (red stems) or for that no significant peak discharges have been recorded during the event (black stems).

Figure 2.4 further gives an indication on the location of the floods, showing the relative contribution of each basin to the indicator *S* (colour shaded bars). During the most severe floods all major basins react (Rank 1: *L* = 83%). Consequently, the number of gauges which did not exhibit any reaction during the event is relatively small and often is exceeded by the number of gauges which recorded a 10-year flood or higher. This ratio changes for events of the extent classes 3 and 4. The less severe the events, the more often only one or two basins dominate the event.

2.4.1 Seasonality

Marked differences can be observed between winter and summer floods, both with respect to their region of occurrence as well as their magnitudes. Additionally to the information given in Fig. 2.4, Fig. 2.5 highlights which basins were actually affected by what level of flooding. Here the shares are differentiated in those parts, where no significant peak discharges could be detected (N-bar), those parts where a significant peak was observed but did not exceed the threshold of Q2, and those parts that contributed to the event severity by exceeding Q2. Two typical examples are given in Fig. 2.5 for (a) a winter flood (February 1970) and (b) a summer flood (July 1954). The maps in Fig. 2.6 illustrate the spatial extent and spatially heterogeneous magnitude of the two examples.

The examples of Figs. 2.5 and 2.6 show, that winter floods are characterised by moderate magnitudes that even for events of class 1 hardly reach those of sum-

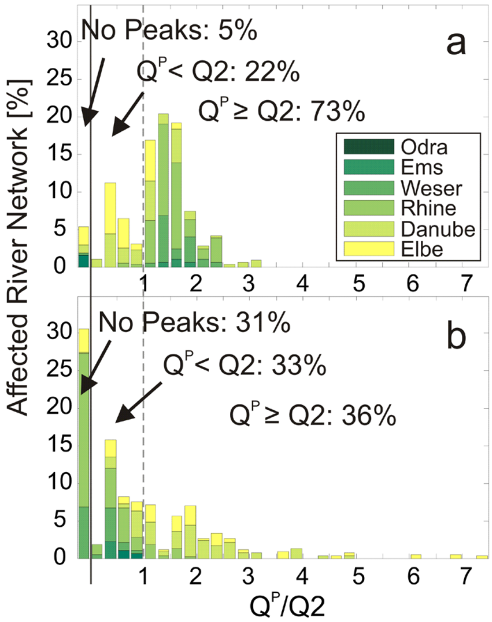

Figure 2.5: Histograms of the shares of the river network affected by peak discharges of a certain magnitude for (a) a typical winter flood (Rank 2, February 1970) and (b) a typical summer flood (Rank 12, July 1954). The colour shading is equivalent to that of Fig. 2.4, highlighting the contribution of the basins to each column of magnitude.

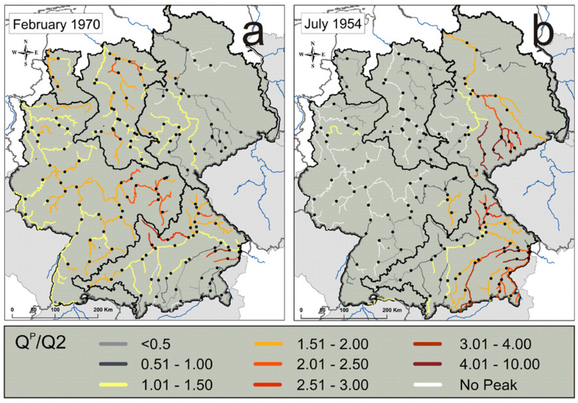

Figure 2.6: Characteristics of (a) the winter flood in February 1970 (22.2. – 4.3.1970, Duration: 11 days, Rank 2, L = 72.9%, S = 114.0) and (b) the summer flood of July 1954 (2.7. – 31.7.1954, Duration: 30 days, Rank 12, L = 36.1%, S = 81.7). The river network is coloured according to the regionalisation of the normalised peak discharges with colours from yellow to red indicating significant peaks above the threshold κ = Q2. Grey shaded river stretches still exhibited significant peaks but did not exceed κ = Q2. At river stretches coloured white no flow reaction could be observed during the event.

mer floods. It can be generalised that the events are characterised rather through their wide spatial extent (for the example of Fig. 2.5a: 73% of the river network exceeded the threshold Q2), rather than flooding with high magnitudes.

During winter, most of the affected river segments are usually located in the Rhine basin, often in combination with the Weser, covering most of west to central Germany (on a north-south extension). In the example of Fig. 2.5a, only 5% of the entire river network showed no reaction to the event at all. 22% reported significant peaks, though below the threshold Q2. Consequently, even though winter events are most common in the centre to west of Germany, the hydro-meteorological origins of the floods are also present in the east and south, leading to reactions in at least parts of the Elbe and Danube basin (mostly in the western and mountainous catchments). Clearly in Fig 2.5a, both basins make up nearly all of the significant peaks below Q2. Returning to the event overview given in Fig. 2.4, we can generalise, that no winter event is solely located in the south-eastern part of the study area. Also, the figure highlights the difference between the most severe events of class 1 and events of class 2. Events are only listed in the top-ranks if additionally to the Rhine and Weser also catchments in the Elbe and Danube are reacting. Events of class 2 are mostly only confined to the first two basins.

Contrastingly, during summer the north and west of Germany (Rhine, Weser) are hardly being affected. Fig. 2.6b illustrates for the July 1954 flood, that the most severe flood peaks were exclusively observed in Danube and Elbe. The remaining basins are often not reacting at all (in this case 40% without any reaction). The ranking of the event is dominated by the few extreme discharges. From the colour shading in Fig. 2.4 this can be generalised: Most summer floods almost exclusively affect the basins of the Danube and Elbe, and during nearly all these floods one third or even more of the river network does not respond.

A closer look on the monthly variability in the occurrence of trans-basin floods is taken in Fig. 2.7. It can be quickly captured that trans-basin floods occur predominantly during winter in the period between December and March. Only few events were detected in the transition months of spring and nearly non in autumn. Summer events occur predominantly between June and August. The differentiation into the event severity classes clearly shows that spatially large extents are almost exclusive to the winter months, with only 5 events in summer belonging to class 2.

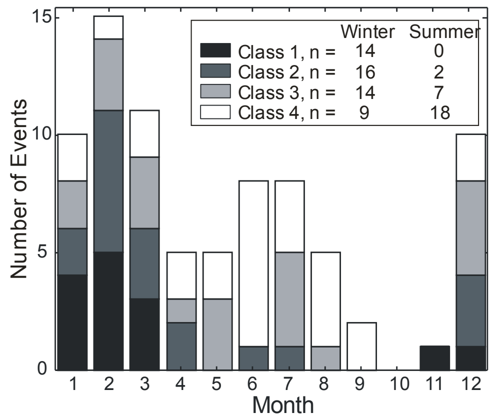

Figure 2.7: Number of events per month/season and severity class.

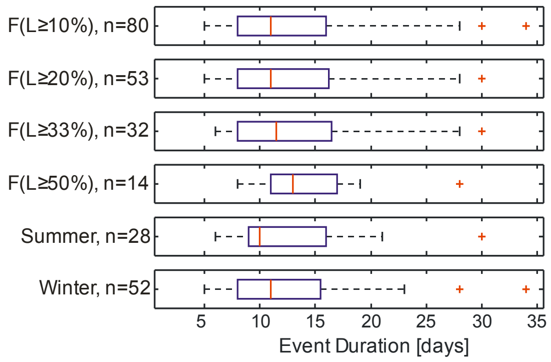

Figure 2.8: Event duration, stratified by spatial extent and season. n indicates the number of trans-basin floods F which fall into any particular class. In the box plots, the central mark denotes the median and the edges of the box the 25th and 75th percentiles. The whiskers extend to the most extreme data points which corresponds to approximately +/–2.7σ, and outliers are marked by the red crosses.

2.4.2 Event duration

The average event duration lies in the range of 10 to 15 days, with maximum durations of up to one month and shortest durations of 5 days. Fig. 2.8 shows the results differentiated by the spatial extent classes and by season. Extreme events (class 1, and therefore winter events) mostly take a longer course in their development with a median of 13 days. The outlying event of 28 days is the top-ranking event of the set, March 1988. During this event a succession of snowfall, snowmelt and rainfall led to a continuous increase in the water stages and the formation of several flood waves throughout the entire country. Due to the widespread nature of these floods longer event durations can be expected, since the flood waves propagate through all basins with varying onsets of the flood initiation. Many winter floods can be expected to be partially caused by snowmelt which often leads to delays in the concentration times due to initial storage of rainfall in the existing snow cover. Summer floods in turn show a faster reaction with an immediate rainfall-runoff transformation. Further, due to the limited area affected, also the outlet of the basin and, hence, last detectable flood peak are reached faster.

2.4.3 A note on stationarity

Figure 2.9 gives an overview of all events in the time period, indicating the number of events in each water year classified by the respective thresholds of spatial extent. It can be noted that the events tend to cluster in time, with periods of frequent, often even multiple floods per year and periods with few occurrences, if any (e.g. late 50s to mid 60s, early to mid 70s, end of 80s to beginning of 90s). This phenomenon has already been described in a number of studies (e.g. Shorthouse and Arnell, 1997; Mudelsee et al., 2004; Llasat et al., 2005; Sturm et al., 2001) and may be explained by distinct modes of inter-annual and inter-decadal oscillations in the climate. Aside from the clustering in flood occurrences, it is further interesting to analyse, whether an actual change in the frequency of the flood events can be observed and whether there are differences with respect to the event severity or season. A simple approach is adopted for this purpose (see Milly et al., 2002). The 51 year observation period is divided into two sub-periods, the first ranging from 1952-1977 (26 years) and the second from 1978 – 2002 (25 years). In Fig. 2.9 the frequencies per extent class and season, respectively, are given for each sub-period.

In total, 44 out of the 80 events occurred in the second half of the 51 year period. Differentiating the event frequencies by the classes of spatial extent reveals some interesting details. In the first half of the record, only 30.0% of all events belong to classes 1 and 2, in the second half 47.7% belong to those two groups. Comparing the total number of these extreme events, 11 were recorded during the first half and 21 during the second half. Assuming, that flood events were

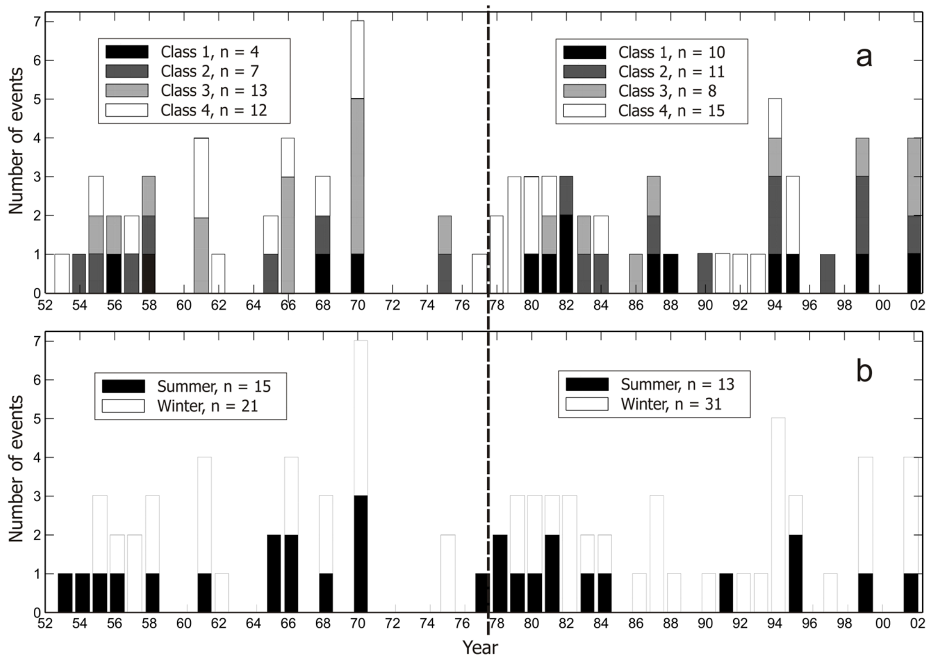

Figure 2.9: Time series of trans-basin floods in 1952 – 2002, stratified into (a) 4 classes of spatial extent and (b) hydrological summer (May-October) and winter (November-April) events. The dashed line indicates the separation into two sub periods (1952 – 1977; 1978 – 2002) and the legends reflect the respective numbers of event per period and for either the extent classes(a) or the seasons (b).

independent outcomes of a stationary process, these results can be compared to a binomial process. We determined a probability of 5.5% of having 21 or more extreme events (classes 1 and 2 together) out of 32 in the second half of the record what can be described as a significant deviation from a stationary process. For the overall event frequencies, considering all classes together and the frequency of events of classes 3 and 4 no significant changes can be observed.

Figure 2.9b further distinguishes the occurrence of flood events with respect to the season in which these occurred. As stated earlier (see Section 2.4.1), the most severe events are predominantly winter events. An increase in the percentage of winter events from 58.0% in the first to 70.5% in the second observation period can be noted. Even more, out of the 21 extreme events (classes 1 and 2 together) in the second half all were recorded during winter, in the first half 9 out of 11 events (so in total there are 30 extreme winter floods). Using the binomial theorem again, the probability of having 21 out of 30 winter events in the second half of the record is 2.1% and, hence, significantly different from the assumptions of a binomial process.

2.5 Sensitivity analysis

To verify the robustness of the resulting event set, it is interesting to revisit the assumptions made in the parameter settings of the methodology. Also, for different objectives in spatial risk assessment, different choices in the parameterisation of the method may be of interest. We test the sensitivity of the methodology for plausible thresholds u and κ, as well as for the time lags τ in the temporal envelope W and analyse the effects on the resulting event sets as compared to the results obtained in Section 2.4. One other important issue to test is the sensitivity of the resulting set of trans-basin floods to the data available, i.e. the number of gauges and therefore time series of daily discharge.

2.5.1 Thresholds u, κ, τ

The choice of the threshold u of the minimum desired flood magnitude, here Q10, is a key factor in the identification of flood events. An increase in the threshold to less frequent events will lead to a reduction in the number of trans-basin floods. Therefore, several dis-

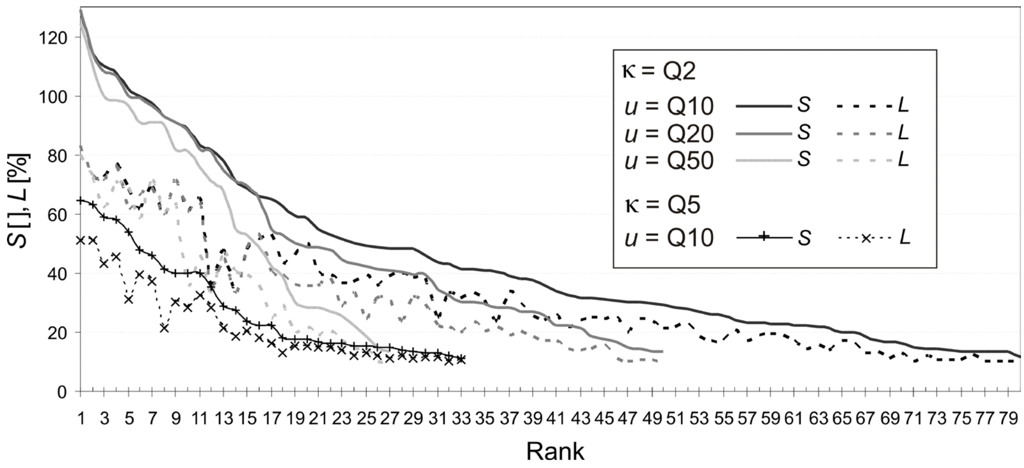

Figure 2.10: Comparison of event sets using various u (p = 0.9, 0.95, 0.98) and κ (p = 0.5, 0.8). The temporal envelope W = [3,10] is the same for all examples. The ranks indicate the number of events and their relative location in the set according to the severity indicator S, but between the sets the same rank may not refer to one and the same event.

charge thresholds u were tested using quantiles of p = 0.8, 0.95 and 0.98 of the annual maximum series, corresponding to T = 5, 20 and 50 year return period (Q5, Q20 and Q50).

Figure 2.10 shows the results for the indicators S and L for the original set u = Q10 and for u = Q20 and Q50. Generally, all resulting event sets are similar in the upper ranks. That is, the most severe events are detected irrespective of the choice of threshold. That is but for the exception of using Q50 for deriving u what severely reduces the number of identified events, also those of classes 1 and 2 as compared to results using the default settings. Certainly, only a limited number of the events even in classes 1 and 2 exhibit local magnitudes above the 50-year flood and, moreover, only very few of the winter events generally do, as has been already analysed in Section 2.4. E.g. the February flood in 1970 (see Section 2.4.1) is no longer detected. Also, caution has to be taken, since the overall length of the time series is just 51 years and, therefore, the thresholds derived for the 50-year floods have to be attributed with a much higher uncertainty than those of e.g. Q10. Decreasing u to *Q5* in turn doubles the total number of days M during which a POT at any gauge was observed (M(Q10) = 381 days; M(Q5) = 707 days). This poses problems in the separation of events even when decreasing τ in the temporal envelope W. Due to elevated flow conditions i.e. in winter a number of events become inseparable extending over 1 or 2 months.

To emphasise the magnitude of an event, the truncation level κ for defining bankfull discharge can be increased. As outlined in Section 2.3.1, Q2 is a rough approximation of bankfull discharge for natural rivers. Areas of high vulnerability are often embanked and bankfull discharge is increased to the level of dyke construction. Since no detailed information was available for the whole of Germany, we tested κ by increasing it to Q5 (keeping u and W of the original set). As can be seen from Fig. 2.10, this change mostly influences the values of the indicator S, reducing it to almost half of the original set (from 129.3 for κ = Q2 to 68.7 for κ = Q5, for the same event of March 1988). The increase promotes events with a generally high magnitude in discharge. Therefore, a number of winter events are largely reduced in their spatial extent since rivers often did not exceed the level of Q5 (L_{max} = 50% as compared to 82.9% for κ = Q2). In turn, the reduction in L is less present for the severe summer events ranging between 4% and 20% (median 9%), as opposed to 10 - 47.7% (median 20%) for winter events. Most of the originally detected severe events remain in the set, but the order of the events can change considerably, with a number of events previously ranking in class 2 now ranging below the extent threshold of 10%. The total number of events reduces from 80 to 33, with 9 out of 33 events recorded in summer. So, even though summer events tend to exhibit stronger flooding in the affected river, the overall spatial extent is also reduced to less than 10%. When adapting the extent threshold, considering e.g. L = 5% as minimum constraint, the number of events identified jumps to 63, bringing forth most of the events of classes 1 to 3 of the original set.

The temporal envelope W for flood peak detection was chosen, using process based assumptions on the ex-

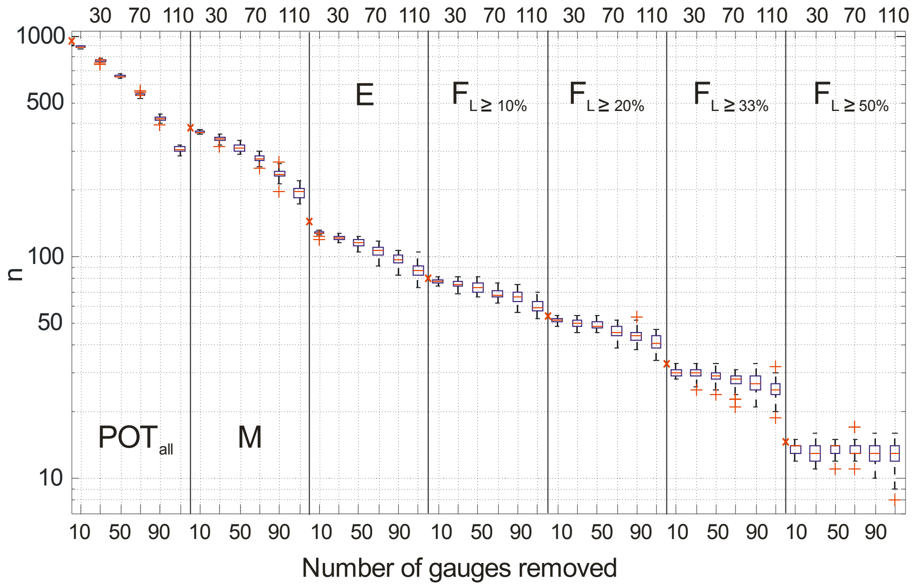

Figure 2.11: Bootstrap results for the various steps in event identification. Shown is the number of all identified peaks over threshold (POT_{all}), the redundancy free number of peaks over threshold (M), the total number of identified flood events E, and the number of trans-basin floods F that exceeded a particular spatial extent L, respectively, against the number of removed gauges. The notation of the box plots is the same as in Fig. 2.8.

pected time for flood evolution. The original parameter setting represents some maximum values for the context of the German wide assessment. Tests using shorter intervals of W = [-1,…,+5; -2,…,+7] days show no major changes in the event identification and ranking. Changes can be observed in the detection of peaks at the most downstream locations which results in a tendency to omit the last flood peak and, hence, leads to a slight reduction of the event duration.

2.5.2 Number of gauges

Finally, the influence of the number of data points on the resulting event set is analysed. For that purpose, a subset of gauges is randomly drawn from the original 162 gauges, removing 10, 30, 50, 70 and 90 gauges respectively (resulting in a sample size of N = 152, 132, 112, 92, 72, 52). The procedure for event identification was run 100 times for each subset, leaving the other parameters unchanged.

The results for the bootstrapping are summarised in Fig. 2.11 showing the changes at the various steps in the methodology, starting from the overall number of POTs, the redundancy free set D of dates with at least one POT observed at any of the N sites, to the resulting number of events E identified. F is then reduced to those events which are of a trans-basin character ($F_{L \geq 10\%}$) and that exceeded certain thresholds of spatial extent ($L \geq 20\%$, $L \geq 33\%$ and $L \geq 50\%$). The results for the original sample size of N = 162 are indicated on the solid vertical lines (expressing 0 number of gauges removed).

As outlined in Section 2.2.3, the POT of the spatial series tend to cluster in time with often many POT occurring on the same day. Using all available gauges (N = 162) during the 18,621 days of the 51 year observation period, a total of 950 POT are identified. These are distributed over only M = 381 days and finally a total of 130 events (E) are identified. From these only 80 events are considered as trans-basin floods ($F_{L \geq 10\%}$).

Applying the various levels of spatial extent, 53 events belong to $F_{L \geq 20\%}$, 32 to $F_{L \geq 33\%}$ and 14 to $F_{L \geq 50\%}$. Reducing N largely reduces the number of POT that can be used in the identification procedure. This reduction though becomes increasingly unimportant the more the peaks are aggregated due to their mutual dependence, highlighting the strong clustering of POTs in time. The most important result is that the most severe events are almost always being detected, even if the number of gauges used in the analysis is reduced by more than a half. For $F_{L \geq 50\%}$ the median of the 100 runs performed is nearly the same for all N

tested, between 12 and 13, as compared to 14 for the original set. In general, reducing the sample size by up to 50 gauges gives similar results for all levels of L. Reducing the set by more sites strongly increases the random component and those events that in the original set are detected due to only one gauge reporting a POT are often no longer detected. On the other hand, the extreme events remain in the set, since during severe flooding mostly a large number of sites report POT discharges (compare to red stems in Fig. 2.4). This leads to a promotion of summer events in the set, the fewer sites are used. As described, summer events are characterised by high magnitudes in most of the affected areas, even though the overall spatial extent is limited. Winter events in turn often report lower magnitudes in the affected areas with only a few sites exceeding u. Hence the likelihood of having a site with POT recording in the randomly chosen sample of sites is higher for summer events than for winter events of medium severity.

The other way around, it can be expected that an increase in the number of gauges eventually would not result in major changes to the event set due to the large redundancy in POT.

2.6 Discussion and conclusions

This study, for the first time, presents a complete and consistent set of trans-basin floods for Germany in the period between 1952 and 2002. We derive a methodology that is capable of capturing the simultaneous occurrence of flooding using multiple series of mean daily discharge. Based on physical reasoning, we assume thresholds for identifying the spatial and temporal dependencies amongst peak discharges, aiming at capturing the system response rather than using a strict quantile approach. Each flood is characterised by a specific value for the timing, the location and the magnitude of discharges within the entire river network.

The consistent and data-based approach allows formulating a cumulative indicator that considers both the heterogeneous spatial extent as well as the locally varying magnitudes of a flood and, hence, allows ranking the events with respect to their overall severity.

The results indicate that in Germany trans-basin floods are a frequent phenomenon, with 80 events detected in the entire 51-year period. Thereby, the western and central parts of the country are most frequently affected. During the most severe floods all major basins react and the number of gauges that do not exhibit any reaction is relatively small. The less severe the events, the more often only one or two basins dominate the event.

We find a distinct seasonal variation of the trans-basin event characteristics. Summer floods often exhibit very strong local magnitudes that are mostly confined to the basins of the Elbe and Danube and one third or even more of the river network does not respond. In turn, winter floods often can be detected in most basins of the entire study area, but the local magnitudes are less strong than during summer floods. The most severe and in this sense also the spatially largest events are predominantly winter events.

We analysed the frequencies per extent class and season, respectively. It can be noted that the events tend to cluster in time, with periods of frequent, often even multiple floods per year, and periods with few occurrences. By dividing the time period into two subsets we detected changes in the frequency. An increase in the percentage of winter events from 58% in the first to 70.5% in the second observation period can be noted. Coinciding we find a significant increase in the number of extreme trans-basin floods in the second period. This finding is in line with other studies that have detected a shift towards increased winter precipitation and the responsible circulation patterns in Central Europe (e.g. Caspary, 1995; Caspary, 2000; Jacobeit et al., 2003; Jacobeit et al., 2006; Belz et al., 2007; Pauling and Paeth, 2007; Petrow et al., 2007; Petrow et al., 2009).

An intrinsic parameter of the methodology is the spatial domain of the study area, here the national borders of Germany. As outlined earlier, summer events tend to be spatially rather limited. Nonetheless, all of the extreme summer floods which can be found in the event set were very prominent events that caused tremendous damages (i.e. July 1954, June 1965 and August 2002). Therefore, these events are well documented and analysed in their hydro-meteorological origins (Glaser, 2001; Christensen and Christensen, 2002; Jacobeit et al., 2003; Ulbrich et al., 2003a, b; Philipp and Jacobeit, 2003; Mudelsee et al., 2004; Pohl, 2004; Grünewald, 2006). These floods affected large parts of the basins of the Danube and Elbe that are located in Austria and Czech Republic and the spatial extent of the entire event by far exceeds that within the national borders of Germany. In contrast, the extreme winter floods in the set can be expected to have been captured more completely in their spatial extent. The rivers Rhine and Weser are located to their largest share or even entirely within the German territory and comprise over 50% of the entire river network used in this study. Both rivers can be categorised as belonging to a winter flood regime (Disse and Engel, 2001; Mudelsee et al., 2006; Belz et al., 2007; Beurton and Thieken, 2009; Petrow and Merz, 2009). Therefore, the dominance of winter events in the set is not surprising and it would be interesting to analyse, whether an extension of the study area to all catchments for each basin would considerably change this ratio. On the other hand, an effect of some winter floods also in the upper Elbe basin is not unlikely. It has to be emphasised, that the results presented here have to be interpreted solely within the national borders. For the purpose of national flood management and insurance issues this is certainly advantageous; but for an analysis of the physics behind these events the event characteristics will have to be analysed in the entire basins. Also, when extending the study area, the conclusions drawn for the changes in flood frequency will have to be revisited.

The method developed in this study has been parameterised based on the available data and in context of the spatial domain from which thresholds based on physical understanding of the flood genesis and on standard risk assessment techniques have been derived. When adapting the method to other regions and even more, when extending the event set to the entire basins (i.e. of Elbe, Danube and Odra) which are under study here, the choices for the time lag τ in the temporal envelop W, that define the spatial dependence amongst flood peaks, have to be adapted also. As in this study, this choice has to be made on the basis of physical reasoning (expected times of concentration and travel times in the channel network).

Depending on the desired aim of the analysis, i.e. the preferences towards spatial extent and/or magnitude, it is easy to adopt the method by choosing different percentiles for the thresholds u and κ. The choice of the POT threshold u influences the number of events that can be identified. Changing the threshold κ (bankfull discharge) alters the indicators L and S, since it raises the threshold for spatial extent. Any event set of trans-basin floods should contain events that are markedly connected with inundation and that are likely to have caused damages of considerable magnitude. Q2 is a rough approximation of bankfull discharge for natural rivers. Areas of high vulnerability are often embanked and bankfull discharge is increased to the level of dyke construction. For a good approximation of the inundation caused by a particular flood, the only solution is the definition of specific thresholds on bankfull discharge for each river reach and the routing of the flood wave through the network. Nonetheless, when changing κ the range of S changes but the intra-event comparison is still consistent. In this way, S can easily be adopted for applications in which more emphasise needs to be given to the event magnitude rather than the spatial extent of each event.

From the sensitivity analysis it can be concluded that the most sensitive parameter for the event identification is the number of days M in the time period during which at least one gauge recorded a discharge above the threshold u. M depends on the number of sites available but moreover on the choice of u. This threshold largely determines whether a flood event can be detected in the first place. Increasing the POT threshold u, first of all, the total number of events decreases. This is largely due to a decrease in the number of winter events, since the maximum recorded discharge during many winter events does not exceed high thresholds u. In turn, this promotes the relative share of summer events in the set. From the sensitivity analysis we conclude, that both Q5 and Q50 are inadequate thresholds u for the purpose of this study, because Q5 fails to separate between damaging flood events and periods of simply elevated discharges, and Q50 fails to detect some major events. Both Q10 and Q20 (or any thresholds in between) are recommended for an analysis of trans-basin flood events.

The method proved very robust to changes in the number of sites N with respect to the most severe events. We further conclude that up to a critical value of N = 110 the overall effects on the resulting event set in terms of number of events detected are insignificant. Nonetheless, when reducing the station network, care has to be taken in the regionalisation of point discharge values to the entire river network that, in turn, determines the quality at which the pattern of spatially heterogeneous flood magnitude can be captured and therefore determines the reliability of the indicator S. Certainly, for the regionalisation as many sites as available should be used to reduce the uncertainties. In addition to the consistent approach of this study, for example, time series that only partially cover the study period could be included and compared to the respective events presented in this study.

The robustness of the method to the number of sites also offers the possibility to extend the analysis further back in time. From the daily time series used in this study (n = 162) about 41 stations date back to 1922 or earlier, about 97 stations date back to 1932 or earlier. The series are more or less continuous with a major data gap for many stations during world war second. The spatial spread is not very even in the early 20th century, with many stations along the major rivers being established since long, but many (also large) tributaries only starting to be gauged in the 30s to 50s.

Also, until 1930 a strong regional bias can be observed with a dense network in the Danube but a poor coverage in Rhine and Weser. Since the sensitivity analysis of the resulting event set towards the number of available sites is performed by randomly removing stations from the set of time series, the spatial spread is more or less preserved. Now, for the real world situation a bias could be expected due to the location of the stations. If the event set is to be extended back in time this must carefully be taken into account. Using the series at hand we would be confident to extend the set by roughly 10 years back in time (with caution on data gaps in 1945). If the regional bias of the gauging station network and the before mentioned uncertainties in the regionalisation of point discharges to the river network can be taken into account the set may even be extended to the mid to late 1920s.

Besides the coherent occurrence of damages during trans-basin floods, for a concise analysis of accumulated risk, it is interesting to analyse the contribution of local floods to mean expected damage. These floods, even though restricted in their spatial extent and their probably uncorrelated occurrence over space and time may still lead to an accumulation of damages. To assess to which degree trans-basin floods and too which degree floods of smaller spatial extent contribute to mean expected damage on e.g. the national scale, an equally consistent approach as presented in this work for trans-basin floods would need to be developed to identify all relevant flood events of small spatial extent. To derive such a set, the pool of gauging stations used in this study would need to be extended by adding stations of smaller catchment sizes resulting in a denser network of stations. So far we have used only stations in catchments that exceed at least 500 km². In this way, we are able to reliably detect large scale flooding. For small floods, e.g. events resulting from convective storms, the uncertainty of the completeness of the event set increases, since a number of local flood events that occurred in ungauged basins will be missed out on. This issue needs to be carefully addressed before conclusions on accumulated risk are drawn.

A natural extension of this study is the quantification of the spatial and temporal dependencies between the peak discharges during the trans-basin floods in a multivariate framework as e.g. it has been proposed by Keef et al. (2009a). This framework needs to be supported by a thorough analysis of the responsible hydro-meteorological processes (atmospheric conditions, runoff generation in the catchment, and routing) and their quantification that allows developing a flood typology. In this way, also more understanding can be gained on the responsible mechanisms for flood genesis at the trans-basin scale. For a frequency analysis the conditions of stationarity and homogeneity in the time series of trans-basin floods have to be carefully evaluated, as already changes in the occurrence rates of winter and therefore the (spatially) most extreme floods were found in this study.

Acknowledgements. We thank the GFZ German Research Centre for Geosciences and the Helmholtz Association of National Research Centres for their financial support. The study was part of DISFLOOD (Disaster Information System for Large-Scale Flood Events Using Earth Observation), a joint project of the United Nations University Bonn, Institute for Environment and Human Security (UNU-EHS), the German Aerospace Centre (DLR), and the GFZ in cooperation with the Helmholtz research network "Integrated Earth Observing System" (Helmholtz-EOS). We dedicate our special thanks to the authorities that provided data. We thank the European Commission – Joint Research Centre (JRC) for providing the CCM River and Catchment Database.

Chapter 3:

The potential of grey literature for understanding floods

Data expansion: the potential of grey literature for understanding floods

Abstract

Sophisticated methods have been developed and become standard in analysing floods as well as for assessing the flood risk. However, increasingly critique of the current standards and scientific practice can be found both in the flood hydrology community as well as in the risk community who argue that the considerable amount of information already available on natural disasters has not been adequately deployed and brought to effective use. We describe this phenomenon as a failure to synthesize knowledge that results from barriers and ignorance in awareness, use and management of the entire spectrum of relevant content, that is, data, information and knowledge. In this paper we argue that the scientific community in flood risk research ignores event specific analysis and documentations as another source of data. We present results from a systematic search that includes an intensive study on sources and ways of information dissemination of flood relevant publications. We obtain 186 documents that contain information on the sources, pathways, receptors and/or consequences for any of the 40 strongest trans-basin floods in Germany in the period 1952-2002. This study therefore provides the most comprehensive meta-data collection of flood documentations for the considered geographical space and period. 87.5% of all events have been documented and especially the most severe floods have received extensive coverage. Only 30% of the material has been produced in the scientific/academic environment and the majority of all documents (about 80%) can be considered grey literature (i.e., literature not controlled by commercial publishers). Therefore, ignoring grey sources in flood research also means ignoring the largest part of knowledge available on single flood events (in Germany). Further, the results of this study underpin the rapid changes in information dissemination of flood event literature over the last decade. We discuss the options and obstacles of incorporating this data in the knowledge building process in the light of the current technological developments and international, interdisciplinary debates for data curation.

Published as:

Uhlemann, S., Bertelmann, R., and Merz, B. (2013): Data expansion: the potential of grey literature for understanding floods, Hydrology and Earth System Sciences, 17, 895-911, doi:10.5194/hess-17-895-2013.

The data used for this publication is freely available as:

Uhlemann, S. (2012): Supplement to: Data Expansion: the potential of grey literature for understanding floods. Deutsches GeoForschungsZentrum GFZ, doi:10.5880/GFZ.5.4.2012.001.

3.1 Introduction

Sophisticated methods have been developed and become standard in analysing extremes in time series, i.e. in estimating the frequency and magnitude of natural events. However, different process types hamper the assumptions of the classical frequency analysis. For the field of flood research, Merz and Blöschl (2008a, b) have called for “a shift away from solving the estimation problem to hydrological understanding”. They argue that the existing formal methods for flood frequency statistics need to be accompanied by hydrological reasoning, i.e. need to reflect the hydrological processes. They specifically argue that the hydrological knowledge gained in the past century is often unduly respected and highlight how the systematic combination of a maximum of relevant information from different complementary sources can help to adjust quantitative estimates from formal methods. Likewise, recently, several international and interdisciplinary groups (International Council for Science (ICSU), International Social Science Council (ISSC) and the UN International Strategy for Disaster Risk Reduction (UN-ISDR)) coined that the considerable amount of information already available on natural disasters has not been adequately deployed and brought to effective use (IRDR, 2011; White et al., 2001).

This phenomenon, which can be described as a failure to synthesize existing knowledge, is known in many disciplines and results from the complexities in the knowledge building process. This process requires an information structure that facilitates the distribution and accessibility of content, that is, data, information and knowledge, over time and distance (Luzi, 2010). Barriers in sharing and unsolved complexities in structuring and managing knowledge have been identified as the major bottlenecks in the knowledge building process (Borgman, 2011; Haendel et al., 2012; Whitlock et al., 2010). Numerous initiatives and international strategies have been set up recently to promote open science and the creation of structured knowledge, particularly aiming at (research) data curation and sharing. The basic motivation behind these initiatives (that ship under the frameworks of open access, open archives, or altogether open science) is the reflection that research results need to be reproducible and verifiable, that data needs to be available to the careful scrutiny of other scientists in order to ask new questions of extant data, to advance the state of research and innovation and finally to create public access to publicly funded results (Borgman, 2011; Wood et al., 2010; Evans and Reimer, 2009; Marx, 2012; The Scientist, 2012). The ‘Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities’[1], provided a first landmark in 2003 and has been signed by leading scientific institutions and organizations worldwide. It calls for the “… free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display …” of research results. Built on these principles recently the European Commission in 2012 released a communication, expecting all results – text and data – funded in Horizon 2020 to be open accessible (European Commission, 2012).

Debate and innovation in data sharing are led by data-rich disciplines such as biosciences or astrophysics and therefore largely focus on observational or experimental data. However, data can take many forms. The National Academy of Science (NAS) in the US uses the term broadly inclusive comprising digital manifestations of literature (incl. text, images and many more) as well as observational data and other forms of data either generated or compiled by humans or machine (Borgman, 2011). For the purpose of this paper we use the term accordingly. A number of applications have already emerged that consider the multifaceted life of data by building repositories that 1) include all kinds of data, from experimental data to publications (e.g., see the BigData movement on biocuration (Howe et al., 2008), BioSharing (Sansone et al., 2012), DRYAD (http://datadryad.org/), SIMBAD (http://simbad.u-strasbg.fr/simbad), see also Genova (2007)), and 2) that deliberately link all information, for example by showing all datasets on a particular genome and all publications that make references to this genome (text2genome project, http://text2genome.smith.man.ac.uk/, see also Haeussler et al. (2011)).

In flood research barriers in access and exchange of hydrometeorological data hamper the set up of central and openly accessible repositories of observational data (Viglione et al., 2010a; Hannah et al., 2011). Existing datasets often lack essential metadata to contextualize and interpret time-series (Hannah et al., 2011) often providing little detailed and no annotated information on location, station or catchment alterations, specifics on extremes etc. Beside observational data and model outputs, data on past events in the broad NAS definitions sense are another important source of information on which risk assessment and decision making needs to be based. In that sense, the role of past and current natural hazard events as learning examples has been stressed at many instances (e.g., Hübl et al., 2002; IRDR, 2011). Knowledge on

[1] http://oa.mpg.de/files/2010/04/berlin declaration.pdf.

flood specific occurrences is particularly important when interpreting extreme value statistics of long time series, particularly when attributing trends to some causal mechanism (see Merz et al. (2012) for a critical treatment on the current state of trend attribution) and for understanding differences in disaster consequences.

Data on past flood events and in particular event documentations are largely non-research data. Since flood risk assessment is at foremost a subject of high societal relevance, it is inherently a subject of governmental action. A large body of authorities is concerned with the management of this risk and the planning of measures for flood loss reduction. Authorities are the primary body of (observational) data and information production and can claim to hold a high level of long term (technical) experience. They are responsible for maintaining the national station network and are therefore equipped with first hand access to and control of the quality of the data, including data normally not available to the scientific community (different parameters, higher spatial and temporal resolution). Additionally, engineering knowledge, i.e. knowledge on (defence) structures, changes to these structures and their operation both in normal times and during events can mostly only be found at the responsible authorities or operators. Consequently, authorities rather than the scientific community are involved in the production of reconnaissance reports in the aftermath of a flood event. These reports often not only touch the hazard part but provide a more holistic and possibly more detailed view on the event including sources, pathways, receptors and consequences.

The mostly technical documents produced by these authorities are commonly disseminated through other means than the scientific publication routes. In the information science they are referred to as grey literature. Grey literature is defined by the Luxembourg Convention on Grey Literature (Farace and Schöpfel, 2010) as "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers" and, as opposed to 'white' or 'conventional' literature (books, scholarly journals etc.), "where publishing is not the primary activity of the producing body". Mackenzie Owen (1997) further highlights that the term grey does not imply any statement on the quality of the document. Rather, it is a characterisation of the distribution mode.

In the scientific community, grey literature seems to be largely ignored in the knowledge building process. Anecdotal evidence and everyday experience confirm two main reasons that have also been found by studies investigating the use and influence of grey in science and research synthesis (see MacDonald et al. (2010) and Rothstein and Hopewell (2009)). First, 'white' literature in the form of journal articles or text books is trusted since it is perceived as quality-labelled according to its production process that includes peer-reviewing and editorial control. Second, and probably most importantly, practical aspects of information retrieval hamper the use of grey in science. White literature is easily searched for and found since it is at the interest of the (commercial) publishers to make their products available, therefore ensuring both the bibliographic control of each item and provision of an information structure that facilitates the distribution and accessibility of the content (through e.g., Scopus or Web of Knowledge). On the contrary, accessing grey requires considerably more effort. Sources are dispersed amongst a multitude of producers or custodians that spend less effort in making the electronic meta-data or the full text of the document accessible (see e.g. Auger (1998), Farace and Schöpfel (2010) and Ranger (2004) for a detailed analysis). Grey documents are mostly produced in the national language of the producing body making it difficult to find and understand the content for a non-native speaker. However, grey can provide a significant added value to journal publishing through the considerably greater detail at which a topic can be treated in a report and the content of unique and significant scientific and technical information that is often not included in scientific journal articles or that otherwise is not published at all (Ranger, 2004; Weintraub, 2000; Farace and Schöpfel, 2010).

So far, no systematic approach has been presented that would allow to defer the size of the body of publications relevant for an event based assessment of floods and no reliable estimate can be made on its potential for combining existing knowledge like that contained in flood reports with a data based analysis. This paper's objective is to first of all identify the existing body of literature that is potentially useful for the specific task of understanding trans-basin floods. Trans-basin floods are extreme events occurring on a regional scale and across catchment boundaries. We aim at creating an openly accessible database of the meta-data of publications that contain information on the sources, pathways, receptors and/or consequences for any of the top 40 trans-basin flood events in Germany (as presented in Uhlemann et al. (2010a)) that can be used as another source of data for flood research.

In this study we develop a systematic search approach with a strong focus on grey publications, i.e. we review the administrative and information landscape in Germany to obtain an overview on the relevant institutions and tools available for the search. Based on the search results we will elucidate the accessibility and origin of the documents and want to capture their basic bibliographic characteristics. Further, we want to analyse the frequency at which trans-basin floods have been reported on and want to assess whether and what kind of changes have occurred in the production process during the study period. This will allow us to determine the potential applicability of flood event related publications both retrospectively as well as for future flood events. Further we will discuss the technical options on how this knowledge can be best deployed for future flood research.

3.2 Systematic search approach

When aiming to identify the existing body of literature available on trans-basin floods in Germany for the past 50 years the search effort has to include both scholarly sources as well as grey sources. In order to draw reliable conclusions from the search this requires a rigorous and transparent search strategy. We apply the analytical steps and rules developed for systematic reviews that provide the methodological rigour needed for the purpose of our study. In particular, we capitalize on the conception of search strategies as they are defined in systematic reviews.

Generally, a systematic review aims at accessing, appraising and synthesising scientific information (Centre for Evidence-Based Conservation, 2010), i.e. it is applied whenever there is need to synthesize the available evidence for a given question, to identify and assess consistent findings across diverse studies (i.e., statistical analysis of causal linkages, effectiveness of interventions) and to inform policy (Burton, 2010; Borenstein et al., 2009). It has originated from the medical and health service sector and become a recognised and standard method also in the field of environmental research and management (see e.g., Higgins and Green (2011), Centre for Evidence-Based Conservation (2010)). In the field of natural hazard research systematic reviews have recently been added to the set of research methodologies for Forensic Disaster Investigations of the IRDR (Burton, 2010; IRDR, 2011).

According to the guidelines for systematic reviews in environmental management (Centre for Evidence-Based Conservation, 2010) a search strategy is an a priori description of the methodology to be used to locate and identify studies pertinent to a systematic review. Based on a specific task at hand, it includes a list of search terms to be used when searching electronic databases, websites, reference lists and when engaging with personal contacts, and the formulation of a-priori inclusion criteria (eligibility criteria) that are applied to the search results. In order to ensure transparency and reproducibility of the results, the method requires the documentation of the entire search strategy. In this way it also entails the opportunity to extend and update the search. Our study differs from a full systematic review approach in that no meta-analysis is planned (Meta-analyses are quantitative procedures to statistically combine the results of studies (Cooper et al., 2009)). Therefore at this point we do not address the points of study characterisation and quality assessment as well as data extraction from the relevant studies. However, when aiming at synthesizing information on flood events from multiple sources as well as for comparing events, the quality of the sources of information needs to be carefully evaluated. We address this point in a separate paper and present a generic framework for quality assessment of natural hazard event documentations (Uhlemann et al., 2013c).

In order to develop our search strategy and in particular as we aim to include grey sources in the search we need to review the administrative and information landscape in Germany to obtain an overview on the tools available and relevant institutions for the search.

3.2.1 The German information landscape

Figure 3.1 depicts, in a generalized way, the various administrative levels that are concerned with flood risk and/or water resources management in Germany. Governmental agencies are organized into federal government agencies and agencies in the single federal states. Within each the organisational levels range from supreme (ministries) to lower administration (district offices). The institutional hierarchies vary within the federal states and Fig. 3.1 shows the example of the state of Saxony. Further, the relevant cross-federal and cross-national levels of organisations that are largely structured according to shares in river basin are depicted in the figure. It is important to note that Fig. 3.1 represents the current administrative landscape and relevant organizations and that numerous changes and reforms have altered this landscape over the past decades.

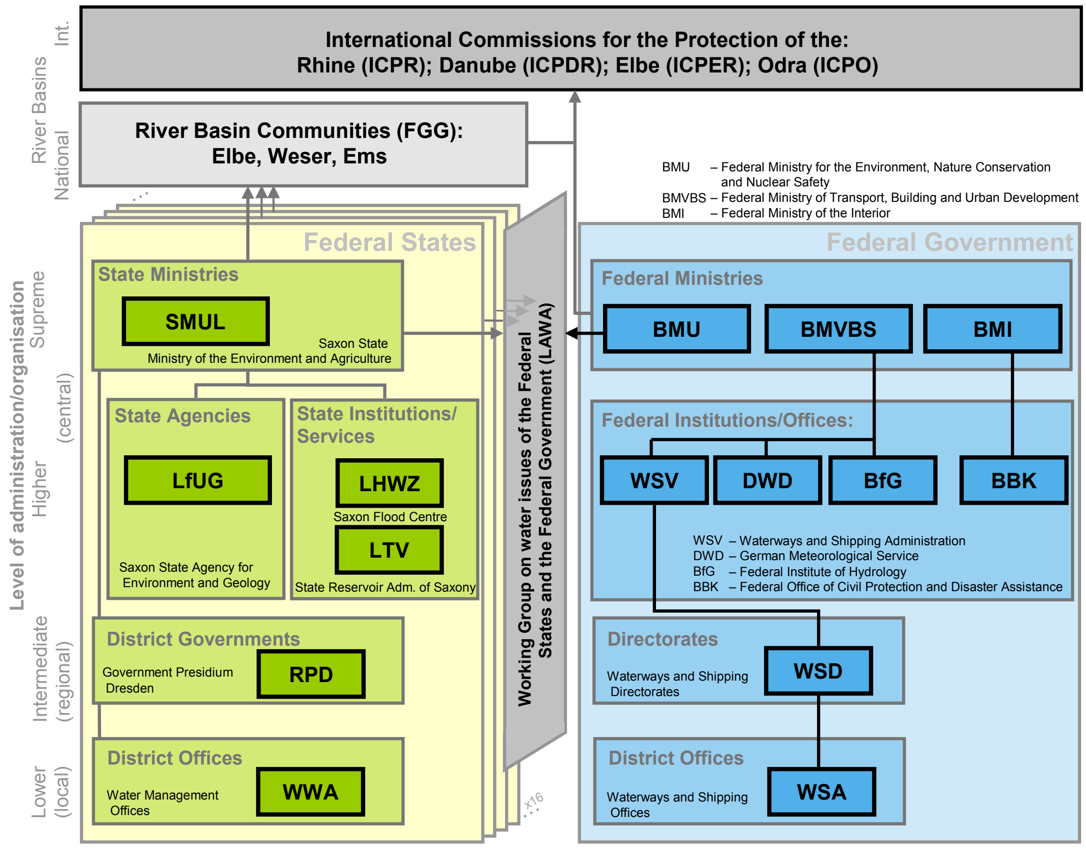

Figure 3.1: General organization plan of water management and related authorities in Germany in 2011, i.e. those relevant for floods. The number of levels at each federal state may vary. Also, numerous reforms have altered the organizational structure during the 20th century (ongoing process!). For the federal states the administrative bodies are given as examples for the state of Saxony (dating to 2002).

In Germany, any published document from public administration is obliged to enter as copy into the German National Library (Deutsche Nationalbibliothek, DNB) and into the respective state library in the federal state where it has been produced (statutory copy). However, the degree of acquisition varies from state to state. Catalogues of the supra-regional to national (DNB) libraries are generally publicly accessible. Further they are part of union catalogues which in turn are part of a widely used meta-search portal for Germany, the Karlsruhe Virtual Catalogue (Karlsruher Virtueller Katalog, KVK). For German publications and in particular for grey publications the KVK can be regarded as the standard national search gateway. It allows for simultaneous search in the union catalogues hence covering the entire German scientific libraries landscape.

Internationally, in an attempt to combat the threat of losing existing knowledge and to reduce the barriers for an effective use of grey literature, over the past two decades, major national institutions and libraries have started creating special ‘grey’ collections. The most important initiative in Europe is the information system OpenGrey (http://www.opengrey.eu), hosted by the Institute for Scientific and Technical Information (INIST-CNRS), France.

3.2.2 Search strategy

In the following subsections we present the single steps of the systematic search for this study according to the methodological design of systematic reviews.

Task at hand

The results of any systematic review are strictly related to the chosen question (task at hand) and the search strategy pursued. In our study we aim at identifying flood relevant literature for the particular purpose of understanding trans-basin floods. Therefore, instead of aiming to conduct a complete search for flood relevant publications we rather aim for consistency in the search approach. The task at hand for this study is phrased as: *Identify all studies that contain information on the sources, pathways, receptors and/or consequences (SPRC) for any of the top 40 flood events contained in the set of trans-basin floods (Uhlemann et al., 2010a) and concerning the territory of Germany.*

This task at hand and the expectation of consistency then impose several logic constraints to the choice of search terms, the search strategy and the formulation of the inclusion criteria. Consistency needs to be given in three particular aspects:

- Temporal and contextual consistency: The strongest constraint results from the limitation of the search to a selection of flood events (top 40 trans-basin floods) and to information on their respective sources, pathways, receptors and/or consequences.
- Scale and spatial consistency: The scale of the publications needs to match the spatial scale and extent of the flood event (large scale flooding, nationally confined to Germany). The scale consistency has implications on the choice of search tools, the languages for conducting the search and the types of references. In particular, the search can only be consistent (in Germany) at the level of white literature and publicly accessible grey literature of the higher governmental administration and national or international institutions as outlined in Fig. 3.1. At this level commonly used and publicly available tools for searching literature are available. Below higher administration the search volume inflates tremendously as the number of relevant administrative units that would need to be addressed inflates and the search mode would likely need to be extended into archival search. Under the limitations of the resources of our study this approach is not feasible as it could not be conducted nationally coherently.
- Accessibility consistency: With our study we address the scientific community. In order to be consistent with the daily scientific search routines the search tools for this study need to be readily available also to any other researcher and at adequate expense. Further, any material that has restrictions on public access per se (like confidential material…) will not be considered in our study.

Search terms and tools

Table 3.1 provides an overview of all search terms that are used in the English and German searches respectively. Any search is limited to the title field. Two sets of search terms are used (flood/inundation terms and defining terms), with individual terms separated by Boolean ‘OR’ operators and sets combined using ‘AND’. Wildcard symbols (indicated by an ‘*’) are used where appropriate. The particularities in German grammar and spelling (compound words and word conjugations) require forward and backward truncation and wildcard replacement of special characters to ensure full coverage of all titles. Generally, any search is the combination of one flood term with at least one of the defining terms.

In section 3.2.1 we provide an overview of the institutional landscape of governmental authorities on the various levels of administration (Fig. 3.1). As outlined, in order to conduct a consistent search we limit the systematic search to organisational levels at river basin scale and the supreme and higher governmental levels to reduce the degrees of freedom in the search and to match the spatial scale of trans-basin floods with that of the administrative levels. Regional and district levels are not approached during the internet search or expert contacts. Further, only scientific libraries at the level above local are included in the search. The Karlsruhe Virtual Catalogue (KVK) is limited to searching only national catalogues, WorldCat and Amazon books. No archive search and no media analyses (web news, print news etc.) are performed. Using the set of predefined search terms, the following search order and respective tools are pursued for the strategic search:

1. Scientific bibliographic databases: Web of Knowledge (In two versions: vs. (a) covering the entire period from 1896 to date but excluding conference proceedings (Science Citation Index, SCI), and vs. (b) limited to articles past 1990 but including conference proceedings (SCI plus Conference Proceedings Citation Index, CPCI)).
2. Meta-search of public open access library catalogues through KVK, and OpenGrey.
3. Search in catalogues not included in KVK: libraries and/or experts of federal or state agencies (e.g., the library of the Federal Institute of Hydrology (BfG)); personal contact; search in index lists of technical (German) journals.
4. Internet Search: Homepages of respective administrations and associations, unions, etc.; Google/Google Scholar searches are only applied to results from the previous search steps to check

Table 3.1: List of search terms.

Sets / Term Category	Flood (SPRC) terms	Defining terms: Year[a]	Place[b]	Month/Season
English	flood*, precipitat*, damage*, event	Tb-flood years between 1951 and 2002; historic*, century	Rhein/Rhine, Neckar, Main, Mosel/Moselle, Donau/Danube, Ems, Weser, Elbe, Mulde, Saale, Oder/Odra	April … December Spring, summer, autumn/fall, winter Christmas, Easter, “new year*”
German	*hochwasser*; *berschwemmung*, *flut*, *regen*, *niederschlag*, *schaden*, *ereignis*	Tb-flood years between 1951 and 2002; historisch*, jahrhundert*	Bayern/Bavaria, …, Sachsen/Saxony Deutsch*/German*, europ* Schwarzwald*, Alp*, „Bayrisch* wald“, Harz*, Erzgebirg*,	Januar* … Dezember*, Fr?hjahr*. Sommer*, Herbst*, Winter*, Weihnacht*, Oster*, Pfingst*, Neujahr*

[a] Where appropriate the year is searched for by using 19* or 20*. [b] For the English search both the English and the German words for places are used

for full text access to the document; if full text is not available the document is ordered via interlibrary loan.

This is a cumulative process and per iteration only those documents are added to the results list that had previously not been found. All relevant results from the strategic search are included in a reference database using appropriate reference management software.

A-priori inclusion criteria

Using the set of predefined search terms, the results are reduced to fit the task at hand. The inclusion criteria are applied to the title of each document and, where available, to the abstract provided. Abstracts are commonly only provided for documents listed in the SCI; for any of the meta-search tools for library catalogues only the bibliographic entries are available. Documents with indistinctive title are attempted to be retrieved and are then checked for inclusion.

According to the consistency criteria, only documents that report on any of the top 40 events of trans-basin floods observed between 1952 and 2002 on the territory of Germany are included in the results list. Included are event specific reports and reports that consider any of the contextual criteria of the source, pathway, receptor, consequence framework for any particular flood event. We deliberately exclude studies on water quality aspects and environmental effects such as soil contamination, effects on species or habitats, or sediment transport. Also (personal) experience reports or narratives are not included. As most river basins in Germany have significant upstream reaches in other countries it is useful to evaluate search hits also from Austria (Danube), Switzerland (Rhine) and Czech Republic (Elbe) and to a very limited degree from Poland (Odra) (none of the top 40 floods in the trans-basin flood event set exhibited major flooding at the Odra, compare to Uhlemann et al. (2010a)). If no additional information on source, pathways, receptors and consequences in Germany is obtained from these documents they are not included in the results list. Further, only reports with a regional scope or broader are included. Local studies that analyse or document the event at a district or city level are not considered. This is sometimes difficult to obtain from the title of a report and some reports on local aspects also account for the regional aspects of the flood, i.e. in the description of the hydro-meteorological causes. They are then also considered.

We include solely print material (both paper and e-prints). According to the strategic search only website contents of either scientific or agency origin are included. This excludes reports from Wikipedia, newspapers, internet news pages, broadcasting (videos, audios) or social networks. Also material in the form of presentations, mostly power points of meetings, classrooms, conferences etc. are excluded.

3.2.3 Document characteristics

By analysing the meta-data of each document that was retrieved through the systematic search and by classifying the document along event specific aspects, we aim to identify the key players in report production including a characterisation of the production process and want to characterise the material potentially useful to maximise the information per event (who produces what, when, how and why).

In order to identify the main producers of flood event related literature we associate the author(s) or issuing institutions to 7 classes according to their affiliation: (1) Specialised governmental agencies (any of the federal or states level shown in Fig 3.1 and commissioned with flood/water management tasks); (2) Non-specialised governmental agencies (governmental agencies not particularly commissioned with flood/water management tasks, mostly ministries); (3) non-governmental organisations or associations, (4) intergovernmental/ international commissions (e.g., ICPR, see Fig 3.1); (5) science/academia (research centres, universities), (6) business (e.g. insurances, associations for shipping etc.); and (7) other or unknown affiliation.

Further, to analyse what is being produced, we classify the reference type of each document. Table 3.2 lists all classes that are accounted for.

To analyse the accessibility of the material we analyse the results of the strategic search with respect to how the document was found (level of search: SCI, KVK, Open Grey, homepages, reference lists, etc.) and evaluate in how far the document's full text is openly accessible and which format (electronic or print) the documents have upon retrieval.

We provide a report typology that basically classifies the purpose of the document in terms of its specificity in being related to any particular flood event. Table 3.2 lists the classes and the definitions of each class.

Irrespective of the report typology assigned, each document contains information on one or many trans-basin flood events. For each document the full list of events is recorded, including the month and year of the flood and the rank given to the flood in the set of trans-basin floods allowing linking each document to the set of trans-basin floods and the existent characteristics per event. Also it allows identifying the number (and types of documents) available per event.

Table 3.2: Classification of reference type.

Reference Type	Definition
International Scientific Journal (CI)	Article in a peer-reviewed, SCI-listed journal
(National) Scientific/ Technical Journal	Article in a national journal, not necessarily peer-reviewed
Specialised Periodical (reg.)	Regularly published, mostly governmental continuous reports (e.g. hydrologic year books, monthly weather reports etc.)
Technical Report/Series (irreg.)	Technical Report or report published within a technical but irregularly appearing series
Monograph	Books, thesis
Edited Material	Article published in an edited book, a book section or as part of conference proceedings
Expert opinion	Commissioned work not produced directly by governmental agencies
Web page	Material that is published only online (i.e. on special portals)
Brochure	Material published for information of the public
Press Release	Releases from official authorities to the press
Other/Unknown	Document does not fit any of the above classes or type of reference cannot be identified

3.3 Results and discussion

The entire database of all references and the full evaluation table of the document characteristics are accessible via the data supplement accompanying this paper and can be addressed using the doi provided in Uhlemann (2012).

Table 3.3: Report typology.

Report Type	Definition
Special Report 1	Report on one, possibly two particular flood events aiming at documentation and analysis. If two events are treated they are described together due to their close temporal occurrences and/or related causes.
Special Report 2	Reports on two to five, rarely more events, sometimes with the aim of comparative analysis but generally aiming at an event description
Special Report 3	Reports or studies on certain aspects of flood analysis making reference to case studies (i.e. any trans-basin flood). Lessons Learned studies (any aspect).
Regional Report	Reports with a regional perspective (geographical region or particular river/basin) either presenting (extreme) flood event collections or studies on flood characteristics in that region that also contain useful information on a particular event.
Continuous Report	Official documents issued by governmental authorities for the purpose of data publication and continuous documentations of e.g. the state of rivers, water resources etc. In case of hydrologic yearbooks or monthly/ quarterly continuous reports flood events are naturally included. For meteorology also the effects of hydrometeorological events are listed (not consistently but frequently).
Other	Reports fitting none of the above classes.

3.3.1 Systematic search

Using the set of predefined search terms the systematic search resulted in the identification and acquisition of 186 documents that fulfilled the inclusion criteria.

In the cumulative process of the search the meta-data of an initial set of 26 documents was identified using the Web of Knowledge. The largest share of documents (114) was identified using the KVK (excluding documents that can be found through SCI). Using the OpenGrey platform resulted in generally very few relevant hits none of which was additional to the searches performed on KVK or SCI. Additional material was then found through searches at institutional homepages (13), through checking reference lists (15) and from tables of content provided for some technical non-SCI journals (4). 14 sources that were otherwise not found were obtained from the special library of the Federal Institute for Hydrology (BfG).

The effort in accessing the material varied considerably with the main differences originating from language specificities of the document and the technical capabilities of the search portal.

The SCI provides comfortable and standardized functionalities in the search options and output formats, including interfaces to referencing software. Publications included in SCI are provided with a link to the abstract of the document (if any is provided) allowing for application of a-priori inclusion criteria directly on the search results. This allows for an efficient search as the document does not need to be acquired before the inclusion criteria can be applied.

Search results conducted through the KVK cannot be saved or exported as no interfaces are provided. Results lists are provided separately for each of the included union catalogues leading to highly redundant search results. Also, keywords and abstracts are not provided along with the meta-data. Therefore, if the title of the document is inconclusive with respect to the a-priori criteria for inclusion of the document in the search results, the publication first needs to be acquired which is challenging with respect to financial and time resources. Further, to the time of this study the technical capabilities of the portal were partly limited with respect to the error free transmission of search terms to the embedded union catalogues. This meant that the particularities in German grammar and spelling (compound words and word conjugations) that require forward and backward truncation and wildcard replacement of special characters needed to be partially substituted by full-length word searches (see combinations of search terms of Table 3.1), therefore inflating the search.

Searches conducted directly at the producing body or their associated libraries (that are not included in KVK) also proved to be less straightforward as complete lists of all publications and the provision of central access points or search options to a data base of

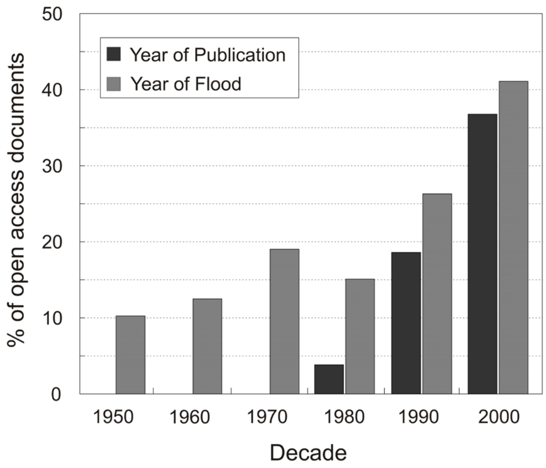

Figure 3.2: Percentage of documents with open access to the full text, aggregated into decades (2000 only accounting for 3 years, 2000-2002) considering (a) the year when the document was published and (b) the flood event years.

publications are not the rule. We find that except for the Federal Institute of Hydrology (BfG (Bundesanstalt für Gewässerkunde), 2012) no authority has provided an overview on its entire list of publications. The institutional library of BfG maintains a very large collection of flood relevant literature, however, the stock is largely confined to material concerning western Germany, the catalogue is not part of the meta-search portal KVK and old material is largely not searchable as it has not been included in the digital catalogue (the largest share of old publications (<1980) has not been entered). However, in the course of the study we obtained copies of paper records archived on microfiche that allowed us to also search the catalogue for documents published prior to 1980. Many authorities provide a publication list on their homepages, however, they are not complete, i.e. they mostly do not list publications before 1990 neither do they include scientific-technical articles submitted to journals by individual employees or as results of co-operation. Further, following up on the discussion on open access to digital works provided in the introduction of this study, a decade later we have to subscribe to the findings of Warnick (2001) that no agency has systematically digitized its legacy collection. Recent publications are frequently added as digital and downloadable documents at the authority's web pages; however, this has not resulted in the automatic indexing of the document's meta-data in an electronic database. In as much as this improves the access to full text, if the users' search strategy solely relies on searching electronic data bases, these documents will not be found.

Access to the electronic full text of any of the identified documents depends on the number of journal subscriptions and inter-library loan agreements that are provided by the hosting institution within which the search is conducted. Given the licenses and subscriptions at hand for our study, 49.2% of all documents were retrieved as print material. Electronic, machine readable text was obtained in 36.3% (33.5% as pdf, 2.8% as online material) of all cases. Electronic but not text processible scanned documents in pdf form comprise 14.5%. In sum, 22.7% of the documents identified for this study are fully open accessible, most being provided on agency web pages. Figure 3.2 shows the percentage of documents that are open accessible (OA) per decade. For one, the percentage is given for the number of documents produced on floods in a particular decade and, second, the percentage is given for the actual years in which the documents were published. The figure highlights that the share of publications with OA increases strongly with time. In particular it highlights that only reports published past 1980 are OA and that OA reports on events prior to 1980 have actually been published past 1980.

A note on completeness of the search

In our study we conduct a systematic search for publications relevant for the task at hand (trans-basin floods) and our search strategy builds on three criteria of consistency: Temporal and contextual consistency (documents that contain information on sources, pathways, and/or receptors and consequences on any of the top 40 trans-basin floods), scale and spatial consistency (conducting the search only at the organisational levels that correspond to the large scale type of the floods and the respective search tools available for that purpose), and accessibility consistency (limiting the search to material that is potentially accessible to any researcher). Our study therefore provides an estimate of the size of the body of literature that is specific for the task at hand. Completeness then has to be assessed with respect to whether the search strategy is comprehensive enough to be able to identify all material actually accessible under the given search scope.

At the level of the search that is independent from the documents producers our search is exhaustive as we deploy the entire spectrum of search tools available to retrieve scientific indexed publications as well as grey publications indexed in online public access catalogues (OPAC) in Germany. Misses at this level of the search may result from the technical limitations of the KVK, i.e. the partially dysfunctional transmission of search terms to the embedded union catalogues. However, we estimate that this concerns very few documents as most documents have been indexed in more than one library catalogue and can therefore still be identified. Misses may further occur due to lagging inclusion of older paper records into any OPAC. However, since we could access the full catalogue (paper and electronic catalogue) of the BfG library, which serves as a national collection for any kind of hydrologic publication, we also estimate that the number of missed documents is rather small.

At the level of the search that addresses the producing bodies directly we include the homepages of all higher and supreme agencies of the federal and state government including that of international organisations. This search addresses the present form of the administrative organisation and the search results depend on the degree to which documents of previous structural units have been included in the current homepages. In any case, the statutory copy obligation has been in place for the entire period investigated in our study (in West Germany). Therefore we estimate that the effect of missing documents from former governmental units is more a matter of the before mentioned lagging inclusion rates of file cards in OPACs.

Beside the accessible material it is likely that more grey literature has been produced that might contain information on the SPRC of trans-basin floods in Germany in the considered period but that has not entered the electronic databases so far or that falls outside of the scope of the search (i.e. does not meet the consistency criteria). The size of this body of literature cannot be inferred at this point of the study as it would require a different search strategy and a substantially larger effort. Based on the experience gained during the search in our study, we conclude that an (expert) survey amongst the producing bodies and custodians of knowledge would be necessary in order to obtain a best estimate of the material produced at (all) governmental institutions. Also, by extending the search to archives and non-public catalogues (governmental and institutional) a number of additional documents might be detected. This concerns particularly older documents (not included in an OPAC so far), documents produced on the lower organisational level of public administration and those of the former German Democratic Republic. For the latter it is however not unlikely that due to the restrictive and centralised publishing strategy in the GDR (Kühnert, S., personal communication, 2012) the number of detectable documents remains small. In any case, archival search requires adequate skills and is extremely time consuming and the cost-benefit would need to be carefully evaluated.

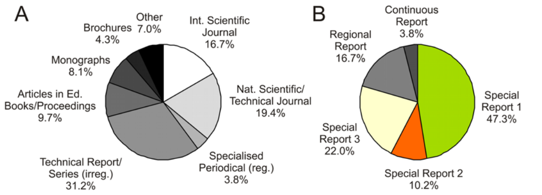

Figure 3.3: Typology of all documents by (A) reference class and (B) with respect to event specificity.

3.3.2 Basic characteristics of the material

We start analysing the material based on the meta-data characteristics described in section 3.2.3. Each document is referred to once; double counts due to multiple events described within one document are not considered at this stage. The study of the material revealed that two out of the first 40 events in the set of trans-basin floods are generally documented in pairs treating two consecutive events as mutually dependent. Both events are merged with the respective dependent events for the purpose of this study (December 1993/January 1994 flood, ranks 8 and 19; February/March 1999 flood, ranks 25 and 27). The strategic search is extended for two more ranks (to 42).

Figure 3.3A displays the reference types that can be ascribed to each document. Nearly one third of all documents belong to the group of technical reports or reports that are produced in irregular series. 17% of the relevant material has been published in international SCI-listed journals, 19% in technical journals that are published mostly for the national market, often specialised for particular regions or branches. 18% of the material is comprised of either monographs (including books and theses) or articles in edited books or conference proceedings. Specialised regular periodicals, i.e. yearbooks, monthly reports etc. contributed by 4%. The remaining material is evenly spread amongst brochures or other material (expert opinions, web pages, press releases…). Based on the definition of grey literature (Luxembourg Convention) the material is to about 80% comprised of grey items and only the SCI-listed journals as well as part of the monographs and edited books can be considered as fully white literature.

Using the typology of reports presented earlier we analyse the specificity of each document with respect to a particular flood event, see Fig. 3.3.B. Over half of the material is comprised of reports that were specifically produced to document or analyse one particular flood event (Special Reports Type 1, 47%) or several events (Special Reports Type 2, 10%). Reports of a mostly scientific nature, investigating certain aspects of floods and making reference to a case study (any trans-basin flood) or lessons learned studies contribute 22% to the material; regional studies make up 17%. For the moment, the share of continuous reports in the set of material only forms some 4%. We only count each type of continuous report not including each of the issues produced. The systematic search revealed that yearbooks and monthly reports on both hydrological and meteorological aspects have been produced for nearly the whole period and both on national and regional scales. However, in the course of this study only a limited number of issues from the continuous series could be retrieved. Acquiring all this material will significantly change the share of continuous reports and it will be interesting to systematically analyse the information content of this material in future.

Most of the material is produced within the national context and for a national auditorium. Only 7% of all documents have a cross border or European scope. A closer look reveals that European scale analysis can be found only for the most recent and also (probably) most damaging flood event of August 2002. This flood affected the central European space. The remaining material has in 27% of all cases a national scope and in 61% of the cases a regional (often related to a particular basin or river) or federal scope (due to the federal jurisdictions). According to the search strategy that largely excluded local searches, the share of material that has a very narrow spatial focus is small (5%). Further, as a consequence of the national task at hand, the main language (text body) of the retrieved documents is German (81% of all reports); 18% are published in English, and less than 1% in other languages.

Using the classification for the affiliation of authors and/or producing bodies, the analysis of the material reveals that the majority (54%) of documents retrieved was produced by governmental agencies (or their employees) that are specialised in the field of flood or water management. 30% of all relevant documents were produced in the scientific/academic environment and 3% by intergovernmental commissions. The remaining percentages are contributed to 5% by higher level, non-specialised governmental institutions (mostly ministries), non-governmental organisations (2%), and business (4%); 6% of the authors could not be associated to any particular institution. From the 54% of documents produced by specialised governmental agencies the largest share was produced on the states level (66%) and there nearly exclusively by the state agencies. At the national level, agencies associated to the Ministry for Transport, Building and Urban Development (BMVBS, see Fig. 3.1 for the overview) contributed to 25% in total, with 14% by the Federal Institute of Hydrology (BfG) and almost all publications main-authored by one person (H. Engel), 7% by the German Meteorological Service (DWD) and about 4% could be attributed to the Federal Waterways and Shipping Administration (WSV) (or its subdivisions). 7% of all documents were produced on a national level in the former GDR.

3.3.3 Event coverage

On an event basis, the analysis crosses over the entire set of documents per flood event and then summarizes over all events. The amount of reports for this analysis increases from 186 to 272 as documents that contain information on more than one event are listed several times.

The amount of documents that contain relevant information on any of the 40 trans-basin flood events varies considerably. Figure 3.4A shows in chronological order the total number of documents per event. For 5 out of the 40 events (12.5%) no relevant document could be retrieved. For the majority of events (60%) less than 5 reports were found. 8 events were documented by 5 to 10 reports and 20% of all events received extensive coverage of more than 10 reports; in three cases, of more than 20 reports. Figure 3.4A also denotes the season during which the flood occurred (stratified in summer (May to October) and winter half years (November to April)). Both floods that received the highest reporting frequency (7/1954, 8/2002) are summer floods. Generally, considering special reports of type 1 summer floods tend to be more intensively reported on. This effect can be largely attributed to the level of damage that was encountered during the flood. Floods with a large spatial extent but with less severe local magnitudes and consequently less damage often do not draw major attention. Many of the winter floods are characterised by this phenomenon. Trans-basin summer floods are characterised by high local magnitude, often with record breaking rainfall intensities, flash flood characteristics in head water catchments and high damages including fatalities.

Uhlemann et al. (2010a) provide an index of severity that allows comparing events according to their spatial extent and their patterns of magnitude. Using this index, Fig. 3.4B depicts the events in their order of

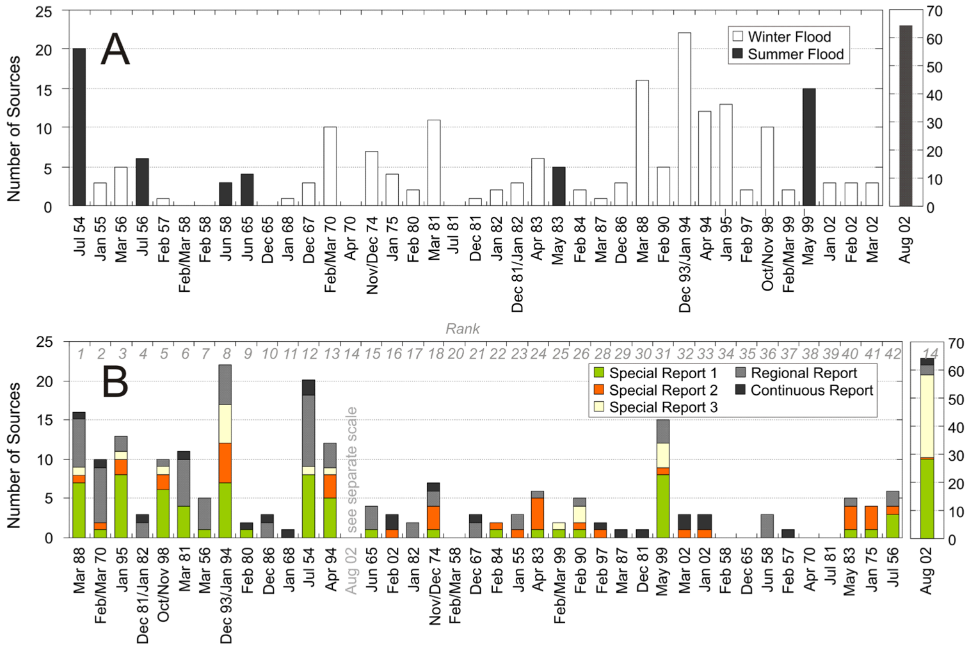

Figure 3.4: Number of documents per event for the top 40 trans-basin floods in Germany. (A) In chronological order; (B) in order of trans-basin flood severity and stratified by report typology. The August flood of 2002 is displayed at a separate scale.

severity; revealing that generally, the highest-ranking events are also those that have been reported on most. Considering the amount of special reports type 1 and 2 produced per event, this effect is even more pronounced. The seasonal effect described earlier leads to summer floods generally receiving more attention than winter floods of comparable rank.

The influence of flood magnitude on the number of publications per event is illustrated in Fig. 3.5, where magnitude is expressed as the exceedance of a certain return period at a certain number of stations (using the 162 gauges of Uhlemann et al. (2010a)). If return periods were encountered that equalled or exceeded the 50-year flood almost certainly at least one report was produced and made accessible and at a level of the 20-year flood the publication of a report is very likely. For both return periods a linear regression can be drawn (Fig. 3.5A) and the correlation coefficient is significant positive with r^2 of 0.75 for T=20a and r^2=0.60 for T=50a (excluding the flood of 2002 from the analysis). Generally, the more gauges exceeded peak dis-charges above T=20a the more reports are being produced. Clearly, damage is the unifying criteria both for report production and also for information dissemination. We find that publications on low frequency high-damaging events are more likely produced as special reports (SR 1-3) and therefore entail a meaningful title (mostly the flood event is mentioned in the title) and due to public and political interests more outreach activities are pursued by the producing bodies. In turn, high frequency but low damaging floods are more likely to be treated in con-text (regional reports, continuous reports) with less meaningful title and with less effort in creating access to the document.

However, the linear regression of Fig. 3.5A can only roughly explain the threshold behaviour of public interest in a flood event and holds only if the flood of August 2002 is excluded from the regression. Figure 3.5B shows the same analysis considering also the flood event of 2002. Taken the event magnitude and spatial extent of the flood the number of reports produced for this particular flood by far exceeds the reporting numbers of all previous events. Obviously some other mechanisms than the previous publication strategies were effective for this event. In the following we analyse the changes in publications frequencies over the investigated period.

As shown in Fig. 3.4a, within the body of material at hand for the analysis, the number of accessible publications and particularly the number of special reports increases with time. Starting from the late 80s this is very apparent as nearly each event starts to receive special documentation (SR1 or SR2). Figure 3.6 displays the time lag at which a report had been produced

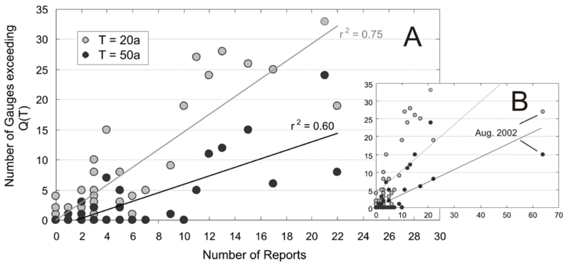

Figure 3.5: Correlation of number of documents per event and event magnitude (expressed as number of gauges that exceeded the return period T (20 years or 50 years) during the event). A) Excluding the August flood of 2002; B) including the August flood of 2002.

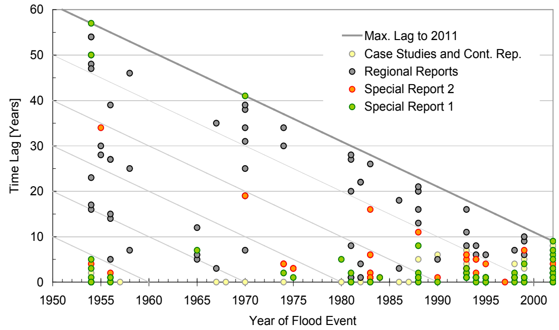

Figure 3.6: Time lag between the year in which the document was published and the year of the flood event. For better orientation the diagonals depict decadal isolines referring to the publication year (2011, dark line; 2000, 1990, 1980, 1970, 1960, light grey).

relative to the year in which the flood event that the document refers to had actually occurred. For a better orientation Fig. 6 includes a line indicating the maximum possible time lag that a document could have (referring to the year 2011 in which the systematic search was conducted for this paper). Considering the year of publication our results show that 35% of the material was published within the last decade and 25% in the 1990s. For the decades of the 60s to 80s this proportion is considerably smaller and it can be seen that only few documents were retrieved in the course of this study that refer back to earlier events. The figure also shows that, whenever special reports are produced, they are usually published immediately. Special reports of type 1 have to 57% of the time been published in the year of the flood and 87% within the first two years. For the flood event of 1954 two reports had been published more than 50 years later, invoked by a similar flood event (August 2002). Documents other than special reports, i.e. regional reports are published also a long time after the event. Often these documents implicitly contain information on a particular event of the past, for example in the analyses of the flood history of a region. The likelihood of a flood event to be included in a regional analysis naturally increases with time and consequently the share of regional reports in the number of publications produced per event is larger for floods of the past.

The increase in numbers of documents since the 80s coincides with three distinct social, political and environmental shifts: 1) the start of the digital period, 2) the reunification of the two German states in 1990 and 3) the onset of a flood rich period starting from the ‘80s (compare to Uhlemann et al. (2010a)) including many of the strongest trans-basin floods. Due to the clustering of floods in this period, the total number of documents in that period is also largest. However, the number of reports per event has also increased. It can be debated, whether this increase results from an increase in reporting frequencies in the last two decades or indicates that documents prior to 1990s are less accessible. We find that both aspects contribute to the effect. In Section 3.3.1 we highlight several reasons that may be responsible for a reduced accessibility to older documents. However, the strong increase in publications over time and in particular over the last two decades by far outweighs the more limited access to older documents. The clustering and the occurrence of some key events with particularly high damages together with the question of climate change impact on extremes has drawn marked media attention, increased awareness and fear in the society and subsequently has led to political action like the creation of (international) river commissions (ICP Rhine founded in 1950, ICP Danube River in 1998, ICP Elbe River in 1990, ICP Odra 1999), the initiation of flood management plans and political frameworks like EC Water Framework Directive (2000/60/EC) and the Flood Risk Management Directive (2007/60/EC) that in turn led to the creation of River Basin Communities (see Fig. 3.1.), the foundation of a Federal Office of Civil Protection and Disaster Assistance (BBK) for Germany in 2004, the initiation of research priorities and programs (United Nations: International Decade for Natural Disaster Reduction (IDNDR, 1990s), Ministry for Education and Research Germany: German Research Network Natural Disasters (DFNK, 2000-2004)). Further, authorities have realized that risk awareness can only be created through information dissemination. Beside event related reports this becomes clear as the number of publications intended for a general audience in form of brochures and pamphlets has significantly increased recently. Besides the increase in numbers of publishing bodies on the river scale it also has to be kept in mind that the number of federal authorities has increased by 6 in Germany as a result of the reunification of the two German states in 1990 and the re-creation of the federal structures in East Germany. Consequently the restrictive and centralised publishing strategy of East Germany (Kühnert, S., personal communication, 2012) has been replaced by federal authorities with responsibilities in free and impartial information of the public.

Several studies on flood damages (e.g., Petrow et al., 2007; Thieken et al., 2007) have shown that experience is a large factor in creating a sense of risk awareness, leading also to reductions in damages. The most prominent examples in Germany are the two consecutive flood events in the middle to lower Rhine in December 1993/January 1994 and in January 1995. Our strategic search results show an increase in special reports (Type 1) for the latter event which can be attributed to the high level of awareness still present both in public and administration. In total, the number of relevant publications is larger for the 1993 flood as comparisons are frequent in 1995 flood reports to 1993 hence covering both events. Another effect can be observed as a consequence of these two events, that is, the triggering of scientific event specific publishing (in the form of SCI-journal or proceedings articles, books, project reports). Previous to 1993, scientific event-based studies of floods were hardly published, but, with the series of flood events in the Rhine region (starting already before 1993 with floods in 1988 and 1990) the topic was put on the research agenda and remained there as further extreme floods occurred within critical periods of time. The community of researchers and projects has since increased leading to an increased publication output.

The most remarkable effect of scientific contributions has been the tremendous amount of scientific publications on the most recent flood (in the set of events) of August 2002. This event received an exceptional number of publications showing a distinct difference to all previous events (as shown in Fig. 3.4). For this event the share of scientific articles is almost equally as high as the number of all other publications. The 2002 event seems to have created an unparalleled type of case study for a wide array of research fields (social, health, economics, risk, engineering, hydrometeorology, ecology ...). Beside this scientific interest also the amount of publications from the agencies by far exceeds that of any previous flood and is remarkably larger than for floods of similar characteristics. For example, only half as many documents were found for the flood of 1954 although this event exhibited partially even higher magnitudes in the same region with higher losses. Further, more than 40% of the documents on this event are openly accessible (compare to Fig. 3.2); highlighting a distinct shift in the dissemination strategies pursued especially by the agencies.

3.4 Conclusions

In our study for the first time we present a systematic approach that allows deferring the size of the body of publications relevant for an event based assessment of floods. That is, this paper's objective was to first of all identify and characterise the existing body of material that is potentially useful for the task of understanding trans-basin floods in Germany from a scientific perspective. Based on the methodological steps developed for systematic reviews we present a search strategy that explicitly includes grey literature using the tools widely available to conduct a search in the German information landscape. The search results are reflective of the material that was actually accessible at the time of the search and under the search strategy pursued. The strength of the approach lies in the full documentation of the entire search process and in it allowing the search to be extended and updated.

We obtain 186 documents that contain information on the sources, pathways, receptors and/or consequences for any of the 40 strongest trans-basin floods in Germany in the period 1952-2002. Most of these 40 flood events have been documented (87.5%) and especially the most severe floods have received extensive coverage. Only 30% of the material has been produced in the scientific/academic environment and the majority of all documents (about 80%) can be considered grey literature. Therefore, our study reveals that ignoring grey sources in flood research also means to ignore the largest part of knowledge available on single flood events (in Germany).

We present the results of our study in an open accessible database (Uhlemann, 2012). This allows any potential user to circumference the tedious work of searching for material that is otherwise scattered amongst a multitude of producing bodies and information providers. The presented collection of material can be considered as a repository containing the digital, centralized meta-data of documents relevant for trans-basin flood event analysis. In that way, the results of this study are the first step into a structured deposition of content, therefore providing access to existing knowledge that would otherwise likely be ignored and the data base can be considered as another source of data for further flood research tasks.

With this study we not only want to provide a database of flood documentations but also create awareness in the research community that grey sources should be an integral part of the knowledge building process. The way in which this knowledge can effectively be combined with analyses based on observational data and modelling results has to be defined separately. However, a knowledge base that integrates all data available will largely facilitate this process and may be precondition to a successful combination of all sources.

So far, there is still a long way to go before a knowledge base on floods is created that could be scrutinized for any (new) scientific query or question. However, a number of critical conclusions can be drawn from the results of this study on the potential applicability of the current material and the next steps to be taken.

Two main barriers in deploying the identified material for research synthesis remain after completion of this study. This is for one the language barrier as 85% of the material is written in German and only a small share is accompanied by English title and/or captions. This certainly limits the widespread use of the material and to allow for interpretability these documents would have to undergo some processing, either in the form of direct translations or in the form of content tagging (keywording, header and caption translations, multilingual ontology that includes dictionaries). Second and most important: So far only a rather small share of the identified documents is openly accessible as full text (in total 22%). The remaining material is available only as digital meta-data set (see Data description) and full text needs to be acquired from the publishers (journal) or via interlibrary loan. Then still, upon retrieval a substantial share of documents is not digitized. Open Access to scientific results was discussed mostly for high ranking journal articles in the last years. In terms of open access publishing formerly grey literature could now become an integral part of scholarly publication. Quality assured reports, published by trusted institutions, typical for grey literature, could be fully accessible, leaving all distribution and accessibility problems behind them.

Digital full text documents are precondition to apply advanced text processing tools (semantic/ontology based text mining) that facilitate the efficiency of any search query and therefore improve largely the speed of information reception including that of machine information processing. For now, any information within the documents can be processed only intellectually. Providing open electronic access not only to current reports but digitizing the legacy of knowledge of flood related institutions would provide highly useful research data to the scientific community. Making such older research data electronically available is always expensive, including costs for annotating, adding relevant metadata to make it useful and, in the end, providing long-term accessibility (Houghton, 2011). Most institutions even don't see a priority to build such a resource. During our study we found no agency that has systematically digitized its legacy collection.

In order to ensure long term applicability extending and updating the knowledge base will be of high importance. This accounts both for past flood events (closing the gap to historic hydrology) as well as for future events. Based on our results we expect that adding material to the database on events after 2002 or any future event will likely become easier and the amount of material will be more abundant including coverage also of low-impact floods. We could show that the increase in numbers of publications per event over time results from a combination of factors, with accessibility barriers playing a subordinate however not unimportant role and changes in publication frequencies due to coincidence with three distinct societal (start of the digital age), political (reunification of the two German states) and environmental shifts (flood rich period) explaining most of the change.

These shifts led to an increase in the number of publishing bodies and to changes in publication philosophies, i.e. increased output from authorities as strong floods occurred within short intervals and flood risk management prominently entering the administrative and research agenda. The most distinct change occurred with the central European flood of August 2002. In as much as this flood was exceptional in many ways (large spatial, international extent, extreme damages) it also demonstrated the tremendous changes in recognition of floods at the administration due to information needs from public and due to the opportunities of internet publishing. The publishing strategies at the agencies have changed largely towards web-publishing of reports providing access to the full text. However, they are often far from fulfilling basic standards of electronic publications. Solutions have to be found on how to support agencies in their archiving and publishing strategies, i.e., it will be crucial to develop standards for information dissemination (i.e., guidelines to administration on how to assure persistent citability, or how to license content for scientific re-use as open as possible; Creative Commons licences, CC by). As research has started to contribute to event-based analysis largely in the form of case-studies since the 1990s and most prominently since 2002 the amount of digital publications (not necessarily open access) outweighs pure print publications since 2002. Currently, the high-level political pressure on open access to research results and on the value and access to data already alters the ways of scholarly communication. We expect that scientific publishing will continue to become more and more open and incorporating this (digital and largely annotated) knowledge in the knowledge building process will become increasingly faster and simpler.

Currently, particularly for grey, we see a gap between the rapidly evolving or already available technologies for information cropping (text mining, semantic search tools) and access options to digital content (see Marx, 2012, and Renear and Palmer, 2009 for future visions as well as critical discussions in Van Noorden, 2012, and Borgman, 2011) and an infrastructure that merely supplies the most basic functionalities for such a development. It can be expected that the fossil design of KVK will soon be replaced by more advanced search tools. Already, new search platforms have been established (e.g., 'Base' http://www.base-search.net). In Germany, the 'Wissenschaftsrat' recently published a paper on the perspectives of the German information infrastructure, paving the way for sustainable development of advanced tools (Wissenschaftsrat, 2012). However, in addition to the systematic search strategy presented in this paper good web search strategies will be needed.

For the future of knowledge management the roles of science and libraries in data and knowledge curation need to be critically reflected (as is currently done in an international debate, see, e.g., Smith (2011)). New standards in information and data literacy need to be developed for data and information curation and libraries are the natural experts in data management and can provide the institutional level of support to permeate these standards into the local cultures (Haendel et al., 2012); that is, both to the individual scientist's level as well as the institutional level including that of public administration. Already, technological developments and international projects in fields such as astrophysics and bioscience preordain the path for the future of data curation through the combined deposition and query of all data available on an object of interest. For hydrology, the development of such standards is currently being discussed for observational data archives (Hannah et al., 2011). In order to develop the full potential of the extant data and knowledge, these initiatives should be either complemented by or at least linked to publication databases and database developments should be thought together and not separately. As demonstrated by the biosciences (OBO Foundry, 2012; GO Consortium, 2012), semantic compatibility and therefore the setup of a (hydro-)ontology will be the basis to ensure interoperability of all data. Clearly, to provide these functionalities for and the access to databases will require a substantial investment (time and funds) as well as the coordination of many different parties and stakeholders. It will probably not be achieved easily. We believe that the results of our study can contribute and help to structure the current debate on knowledge management and curation for flood research.

Data description. The data used for this publication is freely available as data supplement under the creative commons license and can be permanently addressed following the doi given in Uhlemann (2012).

Acknowledgements. We thank the GFZ German Research Centre for Geosciences and the University of Potsdam for their financial support. The main author is grateful for partial funding through the excellence in teaching programme (Junior Teaching Professionals) of the Potsdam Graduate School and a scholarship by the equal opportunities office, both University of Potsdam. We want to thank the authorities and respective individuals who provided information and access to relevant publications for this study, assisted in their retrieval and for donations. We dedicate our special thanks to the library of the Federal Institute of Hydrology for their straightforward cooperation and to the staff of the library of the science park Albert Einstein in Potsdam. The numerous dedicated hours of work of the students Florian Betz, Sylvia Wesser, Annika Meiser, Tim Wahrenberg, and Fabian Gehrmann who assisted in the strategic search are gratefully acknowledged.

Chapter 4:

A quality assessment framework for natural hazard event documentations

A quality assessment framework for natural hazard event documentations: Application to trans-basin flood reports in Germany

Abstract

Written sources that aim at documenting and analysing a particular natural hazard event in the recent past are published at vast majority as grey literature (e.g., as technical reports) and therefore outside of the scholarly publication routes. In consequence, the application of event specific documentations in natural hazard research has been constrained by barriers in accessibility and concerns of credibility towards these sources and by limited awareness of their content and its usefulness for research questions. In this study we address the concerns of credibility for the first time and present a quality assessment framework for written sources from a user's perspective, i.e. we assess the documents' fitness for use to enhance the understanding of trans-basin floods in Germany in the period 1952-2002. The framework is designed to be generally applicable for any natural hazard event documentation and assesses the quality of a document addressing accessibility as well as representational, contextual, and intrinsic dimensions of quality. We introduce an ordinal scaling scheme to grade the quality in the single quality dimensions and the Pedigree score which serves as a measure for the overall document quality. We present results of an application of the framework to a set of 133 event specific documentations relevant for understanding trans-basin floods in Germany. Our results show that the majority of flood event specific reports are of a good quality, i.e. they are well enough drafted, largely accurate and objective, and contain a substantial amount of information on the sources, pathways and receptors/consequences of the floods. The validation of our results against assessments of two independent peers confirms the objectivity and transparency of the quality assessment framework. Using an example flood event that occurred in October/November 1998 we demonstrate how the information from multiple reports can be synthesised under consideration of their quality.

Published as:

Uhlemann, S., Thieken, A. H., and Merz, B. (2013): A quality assessment framework for natural hazard event documentations: Application to trans-basin flood reports in Germany, Natural Hazards and Earth System Sciences Discussions, 1, 143-192, doi:10.5194/nhessd-1-143-2013.

The data used for this publication is freely available as:

Uhlemann, S. (2013): Supplement to: A quality assessment framework for natural hazard event documentations: Application to trans-basin flood reports in Germany, Deutsches GeoForschungsZentrum GFZ, doi:10.5880/GFZ.5.4.2013.001.

4.1 Introduction

The role of past and present natural hazard events as learning examples for hazard assessment, risk prevention and disaster mitigation has been stressed at many instances (Hübl et al., 2002; IRDR, 2011). The underlying rationale is that any predictive method (model) and the effectiveness of disaster response likewise depend on observations and monitoring as well as on experience from real life situations (Hübl et al., 2002). Therefore, any systematic approach requires the evaluation of written documents on past events as well as the standardized documentation of any future event.

Written sources that aim at documenting and analysing a particular natural hazard event in the recent past are published at vast majority as grey literature (i.e., literature not controlled by commercial publishers) and therefore outside of the scholarly publication routes (Uhlemann et al., 2013a). Anecdotal evidence shows that many scientists are aware of the existence of grey literature (to a certain degree), i.e. they know about the publishing activities of state authorities or (inter-)governmental institutions in the course of natural disasters. However, the expectation towards the contextual depth of the documents seems to be rather unclear making it difficult to assess the potential applicability for the own research question.

Studies investigating the use and influence of grey literature in science and research synthesis (e.g., MacDonald et al. (2010), Rothstein and Hopewell (2009), Uhlemann et al. (2013a)) highlight two main reasons that hamper the effective use of the material in the scientific discourse. First, practical aspects that can be largely described as barriers in accessibility (both in terms of the languages used and the options for finding these documents) impose strong constraints on the applicability of these items. Second, the credibility of information published in grey literature is commonly perceived as low. The assumption is that grey literature is never or rarely impartially refereed (MacDonald et al., 2010) and therefore scientific quality standards (peer-reviewing and editorial control) are not ensured.

In contrast to the common perceptions on quality constraints, studies that have investigated the characteristics of grey literature (Ranger, 2004; Weintraub, 2000; Farace and Schöpfel, 2010) often highlight that these documents contain unique and significant scientific and technical information. And that the greatest advantage of grey documents is the considerably greater detail at which a topic can be treated. From their analysis of publications accessible on floods in Germany in the recent past Uhlemann et al. (2013a) conclude that ignoring grey sources in flood research also means ignoring the largest part of knowledge available on single flood events (in Germany).

However, in order to be used in scholarly argument, the issue of credibility of event documentation needs to be addressed, i.e. requires a quality assessment. Only then it will or should be included in the knowledge accumulation or in a decision making process.

Defining the quality of publications or any other documentary evidence is in itself a challenging task as no agreed standard measure has been proposed. Hjørland (2012) provide a concise list of methods for evaluating information sources as they are applied in the field of information sciences. In total they discuss 12 widely used concepts most of which originate from critical research assessment (and therefore primarily intended for ‘white’ literature) and that address quality through the classical peer review, author credentials, publisher reputation or journal impact factors. Inasmuch as these approaches are current gold standards in research assessment (Bailin and Grafstein, 2010) they are indirect methods for evaluating the quality of a document (or its academic worth) and as such have received substantial criticism (e.g., (Seglen, 1997; Simons, 2008)). Broader concepts that acknowledge the heterogeneity of information sources are checklist approaches (most common for evaluating websites, e.g., Hjørland (2011)), comparative studies (evaluating a study against ‘authoritative works’ in the field, e.g., Bragues (2009)) or evidence-based evaluations.

Evidence-based evaluations have found wide application and become standard in the health and medical sciences (Higgins and Green, 2011) and have also been transferred to environmental science and management (Centre for Evidence-Based Conservation, 2010; Norris et al., 2008; Osenberg et al., 1999). They are commonly applied in the course of systematic reviews and meta-analyses are used to synthesize the available evidence for a given question to identify and assess consistent findings across diverse studies (i.e., statistical analysis of causal linkages, effectiveness of interventions) and to inform policy (Burton, 2010; Borenstein et al., 2009). Beside the quantitative meta-analyses qualitative criteria-based methods of causal inference have been developed (see Weed (2000) for a comparison of both methods).

Documentation and analysis of natural disasters primarily aim at describing the event in its course and therefore they do not address classical evidence-based objectives like cause-effect relationships under particular experimental design. For this type of reporting evidence-based approaches are less suited to characterise the quality of the study. A special case of evidence-based evaluations are the techniques developed in the field of historic sciences for source criticism. When studying extreme events of the past, information about the climate or floods in the pre-instrumental period is commonly retrieved from documentary evidence that therefore constitutes the data basis of historic climatology or historic hydrology (Brázdil et al., 2006b). In order to make use of historic sources such as chronicles, annals etc. methods have been developed to critically analyse the evidence of the documents. Rather than defining quality measures of a document the sources are critically checked for the conditions under which the information of concern was produced. Glaser and Stangl (2003) as well as Glaser (2001) present a scheme for source criticism typically applied in historic climatology. In order to assess the reliability of the source intrinsic quality characteristics like the bibliographical level of the author, his mental perceptions and the environmental level (depicting the general Zeitgeist) are used as indicators and the content of the document is cross-checked for compliance with known historical and scientific facts of that time. Only if some basic agreement is reached the climate information is extracted and further used for quantitative analysis (Glaser and Stangl, 2003).

In order to assess the credibility and applicability of recent sources of event documentation and analyses the intrinsic quality assessment needs to be accompanied by a contextual assessment, i.e. an assessment of the quality of information provided on the geophysical processes that caused the hazard and the resulting consequences. Further, from a scientific perspective it is important that the information is accessible and of an intelligible nature. In summary, it is important to assess how much and how qualitatively good the information of the document is, and how likely it is that this information is actually used.

Therefore, a holistic assessment of the quality of written sources is needed in order to assess their applicability or, in other words their ‘fitness for use’. This user’s point of view is pursued in the field of information systems sciences and its linkages to organizations, management and consumers research where the quality of a product (data, information) plays a critical role. It is well accepted in this field of research that the quality of data and information cannot be assessed independent of the purpose for which it is being used (Strong et al., 1997). For more than two decades, the development of methods and standards for data and information quality assessment has received substantial interest (see Madnick et al. (2009) for an overview). Using a factor analysis on a large set of data quality attributes, Wang and Strong (1996) identified four distinct data quality categories all of which need to be addressed for a complete assessment of quality: accessibility, contextual, representational, and intrinsic quality. Each of the categories is further defined by certain quality dimensions (QD). Synonymous to Wang and Strong’s (1996) formulation of data quality, we can summarize the expectation towards a high-quality report on a natural hazard event as that it should be intrinsically good, contextually appropriate for the task at hand, clearly represented and accessible to the information consumer.

Considering the identified strengths and weaknesses of event specific documentations (that are largely characterised by grey literature features) and the apparent barriers for using these documents as another source of data and information in flood hydrology, the objective of this study is the development of a quality assessment framework for written contemporary sources. The aim is to provide a generic framework for information quality assessment and quality labelling of natural hazard event reports for an international research question. Using the example of trans-basin floods in Germany in the period 1952-2002 (Uhlemann et al., 2010a) we want to assess the quality of event specific reports that were produced for any of the top 40 trans-basin floods. Providing a holistic quality measure we want to address the concerns of credibility of sources and of their contextual depth. In that we aim to improve the use and awareness of the information contained in flood event specific reports.

The paper is structured as follows. In Section 4.2, first, we introduce the data used for this study and, second, present the quality assessment framework and the test for concordance to infer its validity for judging the quality of event reports. Results and discussion are presented in Section 4.3 starting with the concordance check and a discussion of the quality assessment framework. Following, we present the results of the overall quality of the entire material. We complete our results with an application to an example flood event. Finally, in Section 4.4 we conclude on the framework and the quality of the reports with respect to their applicability for flood research. The paper is accompanied by an open access data supplement (Uhlemann, 2013c).

4.2 Data and methods

4.2.1 Data

We use a subset of the literature compiled in Uhlemann et al. (2013a). The set comprises the results of a systematic search for publications that contain information on the sources, pathways, receptors and/or consequences (SPRC) for any of the top 40 trans-basin flood events in Germany in the period between 1952-2002 (Uhlemann et al., 2010a). The main criteria for the search can be summarized as follows: The spatial, temporal and contextual frame for the search is given by the task above. Only reports are included that treat any of the selected flood events within Germany and that are of a regional scope or broader. Local studies and studies that document the flood largely outside the German territory are only considered if they provide also information relevant for the scale of trans-basin floods in Germany. The search included solely print material (both paper and e-prints) and website contents of either scientific or agency origin. This excludes reports from other internet sources or media. Further, (personal) experience reports or narratives were not included. The main tools used for the search were 1) the Web of Knowledge, 2) the Karlsruhe Virtual Catalogue (KVK), which is the standard search tool for publications indexed in public open access library catalogues in Germany and 3) open catalogues of federal or state agencies libraries or their respective homepages. A detailed description of the search criteria applied for the systematic search is provided in Uhlemann et al. (2013a).

The search resulted in the identification of 186 relevant publications (see Uhlemann et al., 2013a and the respective data supplement provided at Uhlemann et al., 2012). For the purpose of this study we use only reports that explicitly aim at documenting one or a number of flood events. These types of reports are classified as ‘Special Report 1’ (reports on one, possibly two particular flood events aiming at documentation and analysis) or ‘Special Report 2’ (reports on two to five, rarely more events, sometimes with the aim of comparative analysis but generally aiming at an event description). In total 105 reports of this characteristic are listed in the set of documents. As some reports contain documentation on more than one trans-basin flood event the total number of event specific documentations sums up to 133.

4.2.2 Quality assessment framework (QAF)

In their analysis of the importance of quality attributes for consumers of data Wang and Strong (1996) identified four distinct quality categories (QC) all of which need to be addressed for a complete assessment of data quality: accessibility, contextual, representational, and intrinsic quality. They stratify each of the categories by a number of quality dimensions (QD) and consequently develop their framework for data quality from the perspective of the data user’s needs. This fitness for use is considered through the formulation of a task at hand. Synonymous to Wang and Strong’s (1996) formulation of data quality, we can summarize the expectation towards a high-quality report on a natural hazard event as that it should be intrinsically good, contextually appropriate for the task at hand, clearly represented and accessible to the information consumer.

The four quality categories of the conceptual hierarchical framework for data quality developed by Wang and Strong (1996) and Strong et al. (1997) provide the basis for the development of our quality assessment framework for natural hazard event documentations. We adapt the number and the definitions of the quality dimensions considered per quality category in the Wang and Strong (1996) concept to fit the purpose of assessing the quality of event documentations instead of data. The category of accessibility evaluates the extent to which the report is available, and easily and quickly retrievable. Representational quality includes aspects related to the format of the document and the meaning that a user can make from it. Our rational for defining representational quality is that a report that is written in a language that cannot be comprehended, or that is written clumsily, lacks structure and/or presentation skills, or does not relate to its objectives will not be easily understood and hence utilized for contextual information cropping. Contextual data quality as defined by Wang and Strong (1996) involves judgements on whether the data is value-added, relevant, complete, timely, and of an appropriate amount with respect to the task at hand. We adapt the definition of the contextual dimensions in such a way, that the variety of information needed to characterise any natural hazard event from its causes to the consequences can be best captured. Therefore we employ the commonly adopted Source-Pathway-Receptor-Consequence concept (SPRC) (Samuels and Gouldby, 2009). This concept is useful in that it can be universally applied to all natural hazards. It represents the systems and processes that lead to particular consequences of the hazardous event. We incorporate this

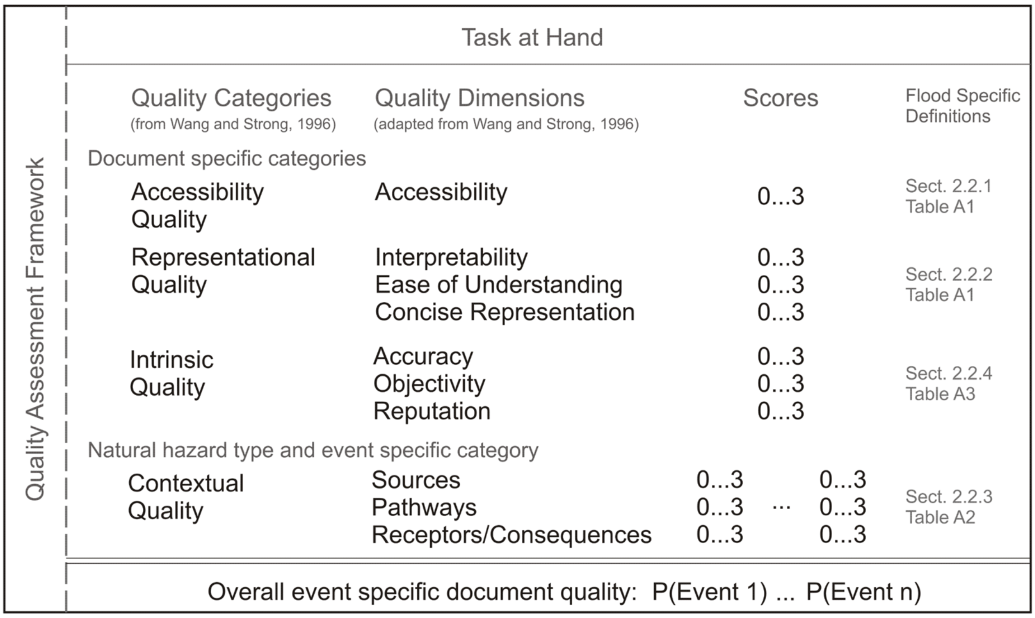

Figure 4.1: Overview of the quality assessment framework.

concept into the definition of the contextual quality criteria and create three quality dimensions: sources, pathways, receptors/consequences. Within each of these dimensions the original contextual dimensions of Wang and Strong (1996) are inherently considered. The intrinsic quality category reflects the overall ‘trustworthiness’ of the report. There are general rules for creating trust, like correct and sufficient data, replicable methods and results, discussion of results and uncertainties, unbiased and impartial conclusions that are supported by evidence, procedures for quality checking etc.

Figure 4.1 provides an overview of the framework and the quality attributes considered. The four quality categories can be differentiated into those that capture quality aspects specific to the document (accessibility, representational and intrinsic quality) and those that capture the quality of the document with respect to the information provided on the natural hazard event (contextual quality). Therefore, if a report covers more than one event it will be evaluated separately for each event in the contextual quality dimensions.

In order to quantify the level of information within each of the individual quality dimensions and to quantify the overall report quality we extend the concept of Wang and Strong (1996) by a pedigree scoring scheme. A pedigree scoring scheme was developed by Funtowicz and Ravetz (1990) and van der Sluijs et al. (2005) in order to assess the uncertainty and quality of environmental research that is relevant for policymakers. Selected data quality dimensions of Wang and Strong (1996) and a pedigree scoring scheme based on the work of van der Sluijs et al. (2005) were already combined to assess and illustrate the quality of flood damage data subsets that can be retrieved from the flood damage database HOWAS 21 (Thieken et al., 2009). In our study we apply a four level ordinal scaling scheme that grades the level of quality reached within each quality dimension in a range from 0 (no information/quality), 1 (low quality), 2 (medium quality) to 3 (high quality).

The ordinal scaling allows calculating an overall quality for each document (and each flood event). This so called pedigree score *P* (Funtowicz and Ravetz, 1990; Van Der Sluijs et al., 2001) is the sum over the scores QD reached in all *i* dimensions divided by the maximum score possible, in this case ten dimensions by three points (Eq. 1). We choose equal weights for all QD in Eq. 1, meaning that *P* is to 90% equally influenced by representative, contextual and intrinsic QC and to 10% by accessibility QC.

$$P = \frac{\sum_i QD_i}{\sum_i QD_{i,\max}} \qquad i = 1,\ldots,10 \qquad (1)$$

P can reach a value between 0 and 1. A pedigree of 1 would mean that, with respect to the task at hand, the document is complete in its description of the event, that the information contained can be trusted and that

Table 4.1: Quality label per document.

$\overline{QD_i}$	$\sum_i QD_i$	P	Quality label
0	0 – 4	0 – 0.13	No quality
1	5 – 14	0.17 – 0.47	Low quality
2	15 – 24	0.5 – 0.8	Medium quality
3	25 – 30	0.83 – 1.0	High quality

the document is easily found and comprehended. Assuming an average score $\overline{QD_i}$ of 0, 1, 2, or 3 over all dimensions (example: an average score of 2 would result in a score sum of 10 x 2 = 20 and P = 20/30 = 0.67), *P* can be interpreted according to the quality labels of no, low, medium and high quality. Table 4.1 provides an overview of the ranges applied for the interpretation of the overall quality of a document.

It is important to note that *P* is not meant to label a document as per se bad or good and any new task at hand will yield its own quality results. It provides a measure to assess the overall quality of a report and assists in creating an overview of the quality present in the material. At any instance, this overall score needs to be accompanied by an analysis of scores reached in the single dimensions or combinations of dimension in order to identify the contextual scope of the document and its strengths and limitations.

In our study we want to investigate the quality of reports that document trans-basin flood events in the period 1952 – 2002 in Germany. Therefore, we introduce the specific requirements of the task at hand (the user's perspective) in the formulation of the definitions. In the case of our study the task at hand is phrased as: *What were the governing hydrometeorological and hydraulic processes that have caused the trans-basin flood event of year x in Germany and what were its consequences?* This task at hand also reflects the scientific user's perspective that we take for the assessment of quality. In defining the quality dimensions we consider the spatial scope at which the report documents an event as reference for the quality expectation and assessment. In the document specific categories we introduce the convention that each report is assessed with respect to its objectives. The rationale for the choice and definition of quality dimensions is given in the subsections of this chapter (as indicated in Fig. 4.11). The Tables A1to A3 in the Appendix then provide the basis for the quality assessments of any of the 133 event specific documentations on trans-basin floods.

Accessibility quality

Accessibility assesses the bibliographic control of the document, that in turn determines the tools by which it can be found and retrieved, and its availability as full text. Further, a document is only fit for use if it can be found using standard search terms for the task at hand. In the case of our study the assessment is based on the systematic search of Uhlemann et al. (2013a) and the search terms and tools provided therein. We define the score classes of accessibility quality to range from inaccessible, e.g. documents not indexed in any public open accessible catalogue like archival material, to full access, i.e. indexed documents with an additionally openly accessible full text. The full definitions are provided in Table A1 in the Appendix.

We omit the quality dimension 'security' that is originally included in the concept for data quality of Wang and Strong (1996) as the level of protection is not relevant for the quality of a document, or, if understood as copyright or access restriction, it is reversely relevant as it limits the access to the document and therefore is already included in the quality dimension accessibility.

Representational quality

As natural hazards are confined to particular geographical regions the reporting on the event is largely endemic to this region and consequently the language(s) used. This is useful for communication amongst authorities; however, it creates barriers in deploying these documents in the context of an international research question if the native language is different to English. The dimension 'Interpretability' considers this aspect. It evaluates whether the document is drafted in appropriate languages and can therefore be comprehended. We define a document to be comprehensible by any individual if it is fully drafted in English. A report that in turn is neither written in the language of the flood affected region nor in English will be of little use for knowledge synthesize and therefore is rated as incomprehensible.

The dimension 'Ease of understanding' relates to formal aspects like the clarity of the writing, the appearance and style of text and figures and the clarity of the structure of the document. In particular, the report should sensibly structure the content that it presents, i.e. should use headings and subheadings in a logical order. This will allow the reader to easily reference into the document without having to scan the entire text. The dimension ‘Concise Representation’ evaluates whether the document is compactly presented without being overwhelming or too coarse and whether appropriate use of additional material (figures, maps etc.) is made. Generally this dimension assesses the amount of relevant and meaningful information provided by the standard of the documents objectives (either deferred from title of document or their direct definition in the text). For example, a report that is produced under the title of ‘The flood of 1983 in the river Rhine’ but that consists largely of photo documentation of single damage occurrences without providing a broader contextual frame is not concise in this regard.

The full definition of the score classes within each quality dimension is provided in Table A1 in the Appendix.

Additional to the three quality dimensions above, Wang and Strong (1996) consider ‘Consistent representation’ (defined as compatibility of the format of the data with that of previous data) for defining the representational quality of data. As highlighted, we do not define an a-priori standard of event documentation that could be used to infer compatibility of the evaluated document with this standard. We therefore exclude this dimension from the quality assessment.

Contextual quality

The grading for each of the three contextual quality dimensions is based on the amount of information presented with respect to a set of expected variables or attributes and the degree to which quantitative analysis are provided in favour of qualitative or descriptive information on the event. If no or an insignificant amount of information is provided the score is zero. A quantitative analysis of the expected list of attributes leads to a score of 3 and largely descriptive and rather incomplete information leads to score 1. In that sense the framework addresses also the added value that a flood report may provide as score 3 analyses contain attributes that are commonly difficult to obtain in the course of a flood risk assessment (or related flood research tasks). The full definition of the score classes within each quality dimension is provided in Table A2 in the Appendix.

Within the QD ‘Sources’ we assess the degree of information presented for the atmospheric processes, the catchment state and the runoff processes. In order to be complete the event documentation has to capture the processes on three temporal scales: preconditions, initiating and maintaining conditions. It should be spatially resolved at a level that allows identifying the main regions of flood origin (geographical and/or by stratified by elevation). Depending on the type of the flood (in particular the flood season) the report should include information on the following state variables or processes and highlight their role in the flood generation: circulation patterns, precipitation (snow/rainfall; intensity, duration, advection/convection/orographic enhancement), temperature, snow accumulation/ melt, catchment state: soil saturation (e.g. through precipitation indices, monthly anomalies, or runoff coefficients), frost.

The QD ‘Pathways’ grades the information provided on the flood propagation in the river, processes that influence the flood wave formation and the inundation effects encountered. Depending on the course of the flood the following variables or processes need to be addressed: affected river stretches, timing and duration of flood peak, effects of superposition of flood crest at confluences, flood volume; effects and types of flood retention due to defence failures (breach location and type) or operations of defence structures (polders, dams, ad-hoc defences by mobile flood protection); inundation extent (and duration).

We combine the evaluation of the amount of information provided for the affected elements of the flood and the damages encountered in the QD ‘Receptors/Consequences’. In a spatially explicit manner, the report needs to differentiate the documentation of the consequences and therein particularly of the (monetary) damages by the sectors that were affected (private households, business, infrastructure…) and by the type of damage (direct, indirect; insured, uninsured; intangibles like fatalities).

Intrinsic quality

‘Accuracy’ assesses to which degree the data used for the contextual aspects of the report are reliable, i.e. of sufficient amount and certified free of error, and whether adequate and documented methods were used. It therefore addresses the degree to which the results are reliable (error-free) and reproducible (amount and sources of data are clear). It has to be noted that this dimension provides a summary assessment over the entire report. It is likely that the accuracy of the presented data and methods varies within a report and particularly across the different aspects of the SPRC. The quality grading then integrates over all aspects providing a mean quality score, which is particularly relevant for reports that cover a number of the contextual aspects. Therefore the score should not be used independently of the quality framework, i.e. for labelling single information entities as in/accurate.

'Objectivity' assesses the validity of the results and conclusions reached, i.e. the degree to which they are supported by the evidence of the data and analysis. Any evidence for influence of interests (political, person, institutional…) leads to lower quality grades. As for accuracy, objectivity is assessed as a mean over the contextual aspects covered by the report.

The quality dimension 'Reputation' is reflective of the common concepts of assessing research by indirect measures like author credentials, publisher reputation or level of peer-reviewing (see e.g. Bailin and Grafstein (2010) and Hjørland (2012)). We consider the technical experience of the producing body as important for the quality of the report however acknowledge peer reviewing as the gold standard for quality control. Similarly we assume consortium work that included numerous publishing bodies of high technical experience as having undergone quality checks by multiple resorts. The quality grades decrease with the degree of generality of the producing body. If the author of the report or the publishing body cannot be affiliated or seems dubious the grade is set to zero.

The full definitions for the score classes of each of the quality dimensions within intrinsic quality are given in Table A3 in the Appendix.

4.2.3 Concordance check – Kappa test

To check for consistency in the interpretation of the definitions of the quality dimensions we test the framework using the kappa statistic κ, which is a widely used measure to express the degree of agreement between two or more independent judges (in the following referred to as peers) for categorical data. Kappa expresses the percentage agreement corrected for chance (Kraemer et al., 2002; Fleiss and Cohen, 1973), or in other words it evaluates the proportion of records for which there was agreement against the amount of agreement that would be expected by chance alone (Centre for Evidence-Based Conservation, 2010). For nominal categories Cohen's Kappa (Cohen, 1960) is given as

$$\kappa = \frac{P_o - P_c}{1 - P_c} \tag{2}$$

P_o – proportion of units in which peers agreed

P_c – proportion of units for which agreement is expected by chance

In an ideal case when there is perfect agreement or disagreement between the peers, Kappa is scaled to vary from -1 to +1, with zero indicating exactly chance agreement. However, in the likely case of encountering disagreements the marginals of the concordance matrix are unequal and the upper and lower limits of κ are slightly less than 1 or, respectively, slightly larger than -1. The exact upper boundary of the maximum possible kappa κ_M is defined as (Fleiss and Cohen, 1973)

$$\kappa_M = \frac{P_{oM} - P_c}{1 - P_c} \tag{3}$$

whereby P_{oM} is found by pairing the values of two peers such that the concordances are maximised. In any case, negative values always suggest agreement poorer than chance and positive values indicating better than chance agreement (Cohen, 1960). A kappa of unity (or the maximum possible kappa) means that both peers are in total agreement and, in the context of this study, that the quality attribution to the documents is not randomly done. In that case, the definitions of the quality dimensions proposed would be perfect in the sense that their interpretation is bijective.

The rating of the quality dimensions in this study is based on an ordinal scale. Variants of Cohen's Kappa, which is restricted in its use to strictly nominal (independent, mutually exclusive and exhaustive) categories, have been developed also for ordinal scales (Fleiss and Cohen, 1973; Lowry, 2012). In that case not only the absolute concordances are taken into consideration but also the relative concordances expressing the nearness of the assignments. We use a squared weighting scheme which is the reverse of the disagreement weights given in (Fleiss and Cohen, 1973) assuming equal imputed distances of 1 unit between the successive scores in each quality dimension (Lowry, 2012). Weights are then assigned to each combination of scores (score category i by peer A and score category j by peer B) according to

$$v_{ij} = 1 - \frac{(i-j)^2}{\max(i-j)^2} \tag{4}$$

According to Fleiss and Cohen (1973) the proportions of Eq. (2) are then calculated as weighted mean observed degree of agreement P_o

Table 4.2: Kappa statistics for the single quality dimensions (N=10) and pooled over all quality dimensions (N=100).

	Access-ibility	Inter-pret.	Ease of underst.	Concise Repres.	Sour-ce	Path-way	Recept. /Cons.	Accu-racy	Objec-tivity	Repu-tation	All QD
κ	0.91	1.00	0.66	0.29	0.84	0.76	1.00	1.00	0.78	0.90	0.89
κ_M	0.95	1.00	1.00	0.91	0.97	0.97	1.00	1.00	1.00	0.95	0.98
σ	0.27	0.00	0.72	0.72	0.27	0.32	0.00	0.00	0.47	0.30	0.07
CI 95%	0.53	0.00	1.42	1.42	0.54	0.62	0.00	0.00	0.91	0.60	0.14

$$P_o = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij} v_{ij} \qquad (5)$$

and mean degree of agreement as expected by chance P_c. P_c is given by the weight adjusted joint probabilities of the marginals,

$$P_c = \frac{1}{N^2}\sum_{i=1}^{m}\sum_{j=1}^{m} n_{i\bullet} n_{\bullet j} v_{ij} \qquad (6)$$

N – total number of subjects

m – number of categories

n_{ij} – number of subjects assigned to category i by peer A and category j by peer B

$n_{i\bullet}$ – total number of subjects assigned to category i by peer A

$n_{\bullet j}$ – total number of subjects assigned to category j by peer B

The sampling characteristics of κ can be captured by computing a confidence interval that expresses the chances that the population value of κ falls within the computed limits. The standard error of κ can be approximated by (Cohen, 1960)

$$\sigma_\kappa = \sqrt{\frac{P_o(1-P_o)}{N(1-P_c)^2}} \qquad (7)$$

With large N it can be assumed that the sampling distribution of κ approximates normality and the respective confidence limits at 95% are given by $\kappa = \pm 1.96\,\sigma_\kappa$.

4.3 Results and discussion

The paper is supplemented by the entire set of documents evaluated in the frame of this study including both the scores given in the single quality dimensions as well as the overall pedigree. Additionally, the results of the kappa test are presented. The data supplement is available through an open access data server and can be permanently addressed and referenced to using the doi provided in Uhlemann (2013c).

4.3.1 Kappa test results and method discussion

We test the quality assessment framework by drawing a random sample of 10 studies from the entire set of documents which are then evaluated by 2 scientists (5 studies each) experienced in the field of flood risk assessment but outside of the collective of authors of this study and who were not involved in the design of the quality assessment framework. In the following we will refer to them as peers. Results from both peers are pooled and checked for concordance with the scores given by the authors of this study.

Table 4.2 comprises the results of the weighted kappa κ for the single quality dimensions and over all QD. The κ measure is accompanied by its respective maximum value κ_M, its standard error σ and confidence interval (CI at 95% confidence). The summary κ over the entire framework is computed by pooling the assessments of all quality dimensions. In as much as the subjects covered in the quality dimensions are necessarily different the scores assigned to define the level of quality reached are the same (0…3). Given the assumption that the imputed distances between the scores are the same in all QD, the κ so computed is then reflective of concordance of the peers in the overall framework. As the number of cases N increases to 100 (10 quality dimensions, 10 studies) the

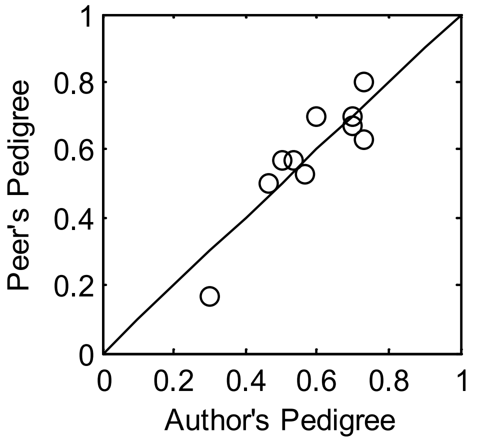

Figure 4.2: Comparison of the peer's P with the authors P (N=10).

measure is statistically more robust than for the single QD (*N*=10). For the entire framework we compute a weighted kappa of $\kappa = 0.89 \pm 0.14$ at 95% confidence interval. The maximum possible kappa for the given marginals is $\kappa_M = 0.98$ and the κ therefore reaches 91% of the maximum possible agreement. The result indicates a high degree of concordance between the peers and we can exclude pure chance operations in the application of the framework to an arbitrary document. Further, investigating the concordance matrices shows that all disagreements deviate by only one score.

When computing κ for the single QD the number of cases N is limited to 10 and the statistical significance of the kappa is limited entailing a high sensitivity to single changes in just one pair of evaluations, as can be inferred by the large standard error and confidence intervals. Keeping the small N-size in mind, in the following we will discuss the agreements reached in the single QD with respect to the strengths and weaknesses of the definitions of the dimensions and their respective grades. In order to substantiate the interpretation we requested a feedback from the peers upon completion of their assessments.

Some distinct differences in the concordances can be observed between the quality dimensions. For the dimensions accessibility, interpretability and reputation a nearly perfect agreement in the scoring between the peers can be observed. These QD can be assessed in a strictly objective manner as they entail no need for text interpretation by the peer.

A very high to high degree of concordance is given for the contextual quality dimensions verifying that the assignment of a document to any of the grades is not ambiguous and will be similar for any peer. Differences in the assignments can be attributed to the large amount of variables that are covered in any of the contextual QD which introduces a minor degree of subjectivity of a peer in drawing the distinctions. Further, as the whole spectrum of aspects from sources to consequences is covered the grading of a document then also depends on the peer's scientific background and experience (which will likely not be at the same technical level for all aspects). The same accounts for the concordance of the peers in the intrinsic quality dimensions 'Accuracy' and 'Objectivity'.

Inasmuch as there is perfect agreement in the assessments of accuracy, the author's and the peer's experience was that a common occurrence is the need for indirect inference of the data and methods used within technical reports from governmental authorities who draw on their own data and standard methods (as given by official guidelines, e.g. for the assessment of return periods) and presuppose clearness about this. Generally, assessing the accuracy of event documentations that aim at compiling information rather than presenting scientific results (data-method-result - conclusion type of analysis) poses some challenges to a peer. It is then up to the peer to decide whether the in-text (and sometimes in-figure) references allow for a sound judgement of the quality of data and methods used. Similarly, assessing the objectivity of the conclusions reached is a largely investigative process that refers rather to circumstantial than hard evidences. It is then particularly a matter of the technical background and expertise of the peer that allows identifying flaws or misjudgements. For both 'Accuracy' and 'Objectivity' the assessment of the quality is based on a mean perception over the entire document and therefore differences are introduced by the need to weigh all aspects presented. Considering the above mentioned degrees of freedom in the interpretation of the score definitions the agreement reached amongst the peers can be considered as substantial.

The least agreement is given for the representational quality dimensions 'Ease of understanding' and 'Concise representation'. For 'Concise representation' the special case occurs that more scores are deviating rather than being fully concordant leading to a low kappa. However, the observed differences in the quality scores assigned by the peers at no instance exceed more than one unit and they exhibit no bias towards lower or higher scores. The result indicates that the interpretation of the definitions in both dimensions is influenced more by the peer's mental framework and expectations towards form and content of a document, i.e. by the application of standards for scientific publications. The discussion showed that most of the uncertainty in the scoring occurs when the peer's initial

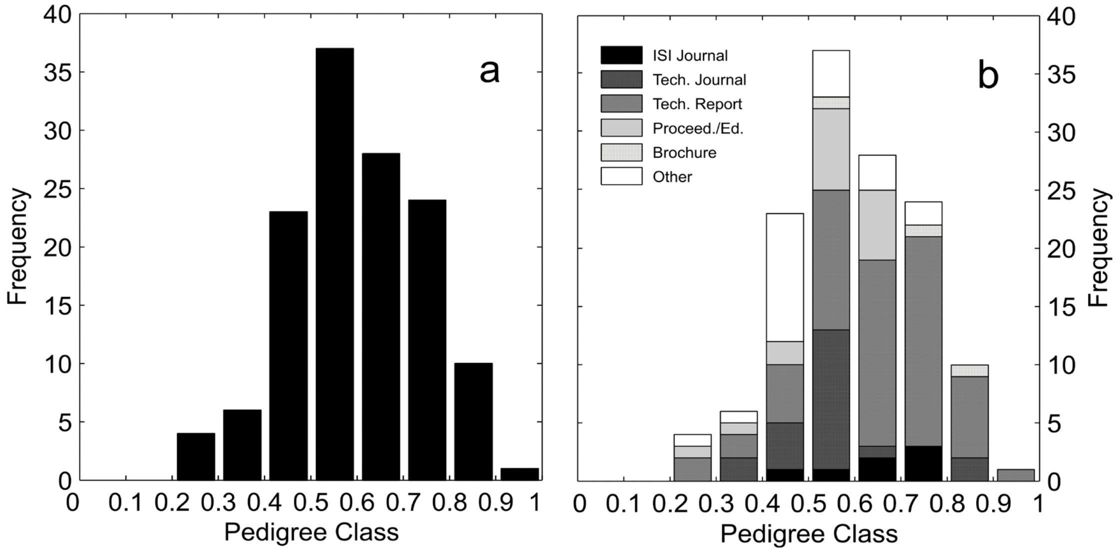

Figure 4.3: Histogram of the pedigree results for all evaluated flood reports (a), and further stratified by reference type (b).

expectation differs from the actual definitions of the scores in the quality dimensions. The representational quality criteria are defined to assess the overall quality of the document prescribing a central role towards the objectives of a report. Deferring the objectives is in turn an investigative process as they are frequently not stated and have to be inferred by the reader mostly from the title of the document. Once inferred it is then still a matter of the peer's expectation towards the comprehensiveness and form at which content is presented. This opens a wide interpretational space for the peers and helps explaining the discordances in the scoring. As the QAF is designed to allow assessing the quality of a heterogeneous group of material (from long technical reports over short ministerial statements to full journal articles) the definition of an adequate size of a document in the conciseness dimension cannot be accompanied by a clear page limit. Therefore the discordances in this QD have to be accepted as a source of uncertainty.

In order to assess the effect of peer disagreement on the overall pedigree we compare the resulting P values (Fig. 4.2). The maximum difference encountered is +0.13 equalling a score difference of four (a difference of one score leads to an alteration of P by 0.03 units). A slight bias towards lower pedigrees can be observed, however, taken the small sample size no statistically significant inference can be drawn. Taking the absolute value of all differences the mean difference between the P is 0.06 and therefore less than 10% of the P. Considering the above spread of scores that is introduced by a random error (due to peers subjectivity without bias) this is a very reasonable result.

4.3.2 Quality distribution and document characteristics

For the 133 assessments undertaken in the course of this study the pedigree ranges between a minimum of $P = 0.23$ and maximum of $P = 0.97$, with a median of $P = 0.57$. Figure 4.3a provides an overview of the spread of the pedigree showing a distribution slightly skewed towards higher quality. The bulk of the documents (67%) reach medium pedigrees in the range of $0.5 \leq P < 0.8$ highlighting that the overall quality of the documents is generally good. One quarter of all documents reach less than 50% of the maximum quality (low quality) with only four reports being evaluated in the range of $0.2 \leq P < 0.3$. As a result of the stringent inclusion criteria applied during the systematic search (Uhlemann et al., 2013a), reports that do not or hardly meet any of the quality criteria are not present in the evaluated material as a report that scores a pedigree below 0.2 (reaching a maximum of 6 out of 30 points) would likely not be accessible, be of low intrinsic quality and generally fail to provide information on the SPRC of the event and therefore not fulfil the inclusion criteria that relate to the task at hand. Eleven reports were evaluated to be of high quality ($0.8 \leq P < 1.0$) and one report can be considered as qualitatively near perfect (Gräfe, 2004).

In Fig. 4.3b we show the distribution of the pedigree stratified by the type of the document. Technical reports form the largest group of flood event documentations. Generally, technical reports (most of which are produced by governmental agencies) span the whole spectrum of quality; however, they exhibit a tendency towards higher quality, i.e. they form the largest part of documents that are assessed of being of medium to high quality. They provide a format that is obviously more suited for the documentation of flood events. In contrast, contributions in national technical journals that are also used as medium for publishing by governmental agencies tend towards lower qualities. Reports produced in the academic/scientific environment that are published in ISI-listed journals are evaluated as medium quality with no deviations to poor or high overall quality. That means they fulfil standards in presentation and the intrinsic quality is high, however often they are constrained by the contextual scope, i.e. in most cases they focus on only one particular aspect like the hydrometeorological conditions of the flood.

4.3.3 Components of quality

The overall quality of each report as expressed by the pedigree is reflective of the scores reached in the single quality dimensions. In order to add meaning to the summary measure P and to gain more understanding of the features that determine report quality, we analyse the quality scores reached in the single quality dimensions and their correlations.

A first survey of the composition of the pedigree by its ten dimensions reveals that an overall low pedigree also means that the scoring in any of the single quality dimensions is likewise low, i.e. a report that is of low overall quality is also comprised of evenly low scores in the single quality dimensions. Effects of one quality dimension excelling the others by far are not encountered as they are not completely independent. This holds except for the contextual criteria, where it is likely that only one or two dimensions have been within the scope of the report. A report that is of low representational quality will likely also be of lower contextual and intrinsic quality. Similarly, as the overall pedigree increases also the scores in the quality dimensions largely increase. Only in the range of 0.4 to 0.5 the first quality dimensions reach the maximum score of 3, often this is either in the intrinsic or representational criteria. For reports of $P > 0.6$ additionally the contextual quality dimensions reach high scores.

In the following we will analyse the four quality criteria and their dimensions in more detail. Table 4.3 presents an overview of the performance of all documents in the single quality dimensions and highlights the frequencies at which certain quality levels are reached. Additionally, Fig. 4.4 illustrates the scores that were reached by the documents per QD in relation to the overall document quality.

Accessibility

The majority of flood event specific reports (62.4%) can be found and retrieved using standard search terms on standard search engines within the national framework provided. However, only 17.3% of all documents are both searchable and open accessible in their full text. 14% of the material can only be found using non-standard searches, i.e. through searching the online portals of flood relevant institutions (where in fact the document may be available as free download) (score 1). 6% of the material used for this study is actually inaccessible and became available either by chance (donations etc.) or was found in reference lists. In Uhlemann et al. (2013a) we already provided a detailed analysis on the accessibility of literature relevant for the task at hand and the evaluation of the subset of material used within this study reflects the results presented for the entire set of literature.

Comparing the quality level reached in this dimension with the scores reached in the contextual dimensions highlights, that the degree to which the document is easily accessible has no implication for the contextual depth of the document. I.e., open access documents (one open access journal, two proceedings paper, three international river commission reports, four governmental press releases, the others: states or federal government reports by authorities with technical expertise; all published past 1996) likewise contain few to many aspects relevant for understanding a flood event and the overall pedigree exhibits some spread (see box-plot for score class 3 of accessibility in Fig. 4.2). However, both the representational as well as the intrinsic quality dimensions are graded as good to complete (scores 2-3) for open access documents indicating that the degree of quality control increases (irrespective of peer review). This may be attributed to the author's or authority's additional effort towards meeting quality standards upon the document's enhanced public exposure.

Table 4.3: Distribution of the scores per quality dimension, given in percentage of documents per score class. Bold numbers indicate the scores reached by the majority of documents.

Quality Criteria	Access. Qual.	Representational Quality			Contextual Quality			Intrinsic Quality		
Quality Dim.	Access-ibility	Inter-pret.	Ease of Un-derst.	Concise Repres.	Sources	Path-ways	Recept./ Cons.	Accu-racy	Objec-tivity	Repu-tation
Score	Absolute frequency									
0	8	0	0	3	21	22	92	4	1	2
1	19	115	19	35	37	61	24	24	7	10
2	83	5	72	53	44	45	8	62	36	89
3	23	13	42	42	31	5	9	43	89	32
Score	Relative frequency (%)									
0	6.0	0	0	2.3	15.8	16.5	69.2	3.0	0.8	1.5
1	14.3	86.5	14.3	26.3	27.8	45.9	18.0	18.0	5,3	7.0
2	62.4	3.8	54.1	39.8	33.1	33.8	6.0	46.6	27.1	66.9
3	17.3	9.8	31.6	31.6	23.3	3.8	6.8	32.3	66.9	24.1

Representational quality

Uhlemann et al. (2013a) already highlighted that the majority of documents compiled in the systematic search is drafted in the language of the producing body, in this case in German (90.3% for the subset used), a main characteristic of grey flood reporting. Less than 10% are completely published in English and within the German documents 4% are accompanied by English titles, captions and/or summaries. The language component however does not influence the overall quality of the documents. The spread of pedigrees for the German documents reflects the pedigree distribution with a median of $P = 0.57$. The five documents that are accompanied by English annotations are of a slightly better quality (one governmental technical report, four contributions to national technical journals with respective publishing requirements). The 13 documents that are published fully in English (eleven articles in ISI Journals or proceedings, two European Commission (EC) contributions) are in turn characterised by lower scores in the contextual quality dimensions and none exceeds a P of 0.8. Two main reasons can be identified: the journal articles mostly present a detailed analysis of the meteorological aspects of the floods however they do not treat aspects of the catchment conditions, pathways and/or consequences. The proceedings papers in turn treat all aspects but, due to the condensed format, in a rather brief manner. The EC contributions concern official statements of the European Union towards political actions in the aftermath of the August flood in 2002. Here, the special effect occurs that the content provided scores as insignificant in all contextual quality dimensions.

The results obtained within the other two quality dimensions of the representational quality criteria (‘Ease of understanding’ and ‘Concise Representation’) indicate that the flood specific reports available for this paper are at the majority qualitatively good in their representational characteristics. This means that the quality of the use of language and terminology as well as the structure of the documents is generally good and also that the content is largely presented making adequate use of figures, tables etc. and providing a balanced report in terms of length of the document with respect to its objectives. However, it has to be noted that quite a number of reports also suffer from a too condensed presentation (26.3%), e.g. the presented content is not representative in its volume or presentation to fulfil the aims of the report. The most frequent occurrence in this context is the overall brevity of the report (less than one to a few pages of text). Some reports present results largely by figures, tables or photos that are insufficiently accompanied by explanatory text. Clearly, if a report lacks in its conciseness and structure the fitness for use is likewise limited. Figure 4.4 shows a clear correlation of both dimen-

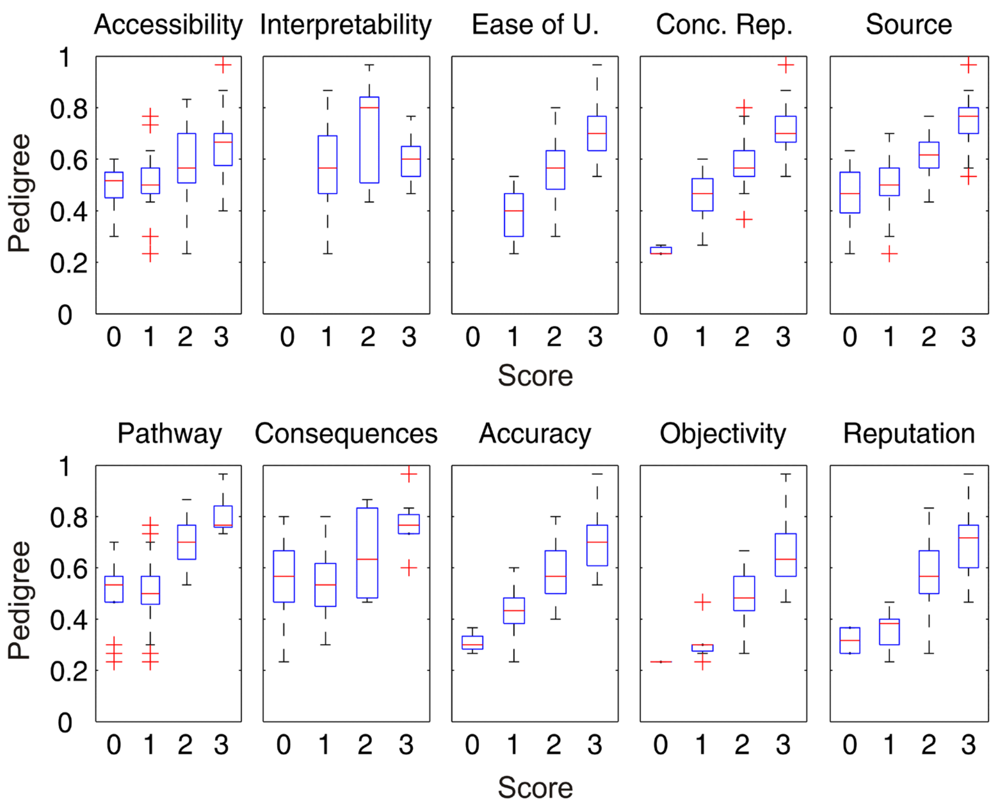

Figure 4.4: Boxplots showing the distribution of the overall pedigrees reached by the documents in a particular score class of a quality dimension. The sample size per box can be inferred from Table 4.3.

sions with the overall quality of the documents and a pair wise correlation with the score class 3 of the contextual dimensions and accuracy. Those reports that are of an overall good quality are exclusively well written and well structured.

Contextual quality

The summary statistics in Table 4.3 show that the contextual dimensions are assessed heterogeneously. 83.5% of all reports include an analysis of the hydrometeorological causes of the flood, whereby one third contains detailed (score 2) and slightly less than one third either complete (score 3) to limited information (score 1). Documents that capture the hydrometeorological origin of a flood rather coarsely are of a generally broad and mostly condensed nature (conciseness scores 1 or 2). Most reports in this score category present a summary of the hydrometeorology based on secondary sources and provide a largely qualitative and descriptive characterisation of the flood initiating conditions (7 out of 37 present own data along the analysis). Reports that were assigned a quality score of 2 can be categorized into two types: first, reports with a focus on the atmospheric aspects providing detailed and mostly also quantitative analysis (however omitting the other aspects of flood initiating conditions) and, second, reports that provide a more detailed description of the entire spectrum of the flood generating processes, however, not fully quantitative.

The pathways of the flood event are mostly treated partially and hardly any document fulfils the quality expectations towards a comprehensive documentation of the flood routing and particularly of the inundation processes. Most documents fail to provide a systematic documentation of the entire flood affected areas and the mechanisms that led to inundation (which is mandatory for quality score 3). Only 5 reports attempt to document the flood extent fully, however mostly in a descriptive manner. 2 reports on the summer flood in 1954 actually provide a mapped overview of the flood affected area. Frequently routing charts are presented together with analyses of the wave propagation in the river. Also breach occurrences or volume and effects of retention on the flood crest are documented providing a considerable added value as this information cannot be obtained in a data based analysis that relies on series of discharges or water levels. The scores assessed for sources and pathways show a slight positive correlation. Most reports present a combined analysis of the hydrometeorology and the flood wave propagation and often at comparable depth.

Almost 70% of all documents do not include a documentation or analysis of the consequences of the flood event. If considered, the information is mostly treated as an add-on in the report (18%, score 1) presenting only the official overall damage (or an estimate of it) and/or giving examples of damages at (arbitrary) locations not allowing for an overview of the amount and types of damages (or their spatial pattern). However, a number of documents, even though not delivering any descriptive or quantitative analysis of the damages (and therefore score 0), include photo documentations. In case of verifications of particular damage occurrences they may still provide useful information. The nine reports that contain detailed information on the consequences of the flood are all of a high quality. Except for one report on the August flood in 2002 that exclusively treats the damages of the event (therefore reaching only $P = 0.6$) in all other cases the sources and pathways are also treated in detail and $P \geq 0.73$.

Intrinsic quality

The intrinsic quality of a document represents the trustworthiness of the content presented. The accuracy of the data and methods used varies amongst the material; however, the majority of all documents (79%) are of medium to good accuracy. As indicated in Section 4.3.2, a number of low to medium scores can be attributed to the fact that the document under validation failed to highlight the data or information sources that were used and/or did not document which methods were used to obtain the results and conclusions presented. Most documents present their results in an objective manner and only in a few cases direct evidence on falsifications or unjustified interpretations could be found. For three documents clear bias in the interpretation of the flood event could be found. These are 1) politically motivated, as for two reports from the former GDR where the actual inundation effects and damages of the described flood events were concealed to emphasize the successful management of the flood, and 2) they are the result of biased positions of the author particularly in the discussion of flood engineering vs. restoration of flood plains (1980s) and since the 1990s also in the discussion of the effect of climate change on the frequency and magnitude of river floods. In both cases the conclusions reached are not supported by the presented data analysis. The other five reports of score class 1 generally failed to meet their objectives. Generally, reports of low objectivity are also of an overall low quality. Once the quality in the other quality dimensions is given the reports are mostly presented in an unbiased, objective manner. Also, if a report is accurate (score 3) it is always fully objective (score 3), however, not vice versa.

The reputation of a document is judged based on the authorship as well as on the level of independent reviewing that the document has been subjected to. As highlighted earlier, most reports were not published in a peer-reviewed journal. Therefore the majority of documents only receive the second highest score (67%). For peer reviewed material (which includes a limited number of technical national journals too) and technical reports published by any international river commission or by a consortium of different agencies/stakeholders a thorough quality control was assumed. The box-plot for score class 3 in Fig. 4.4 highlights that the documents of high reputation are generally of higher quality. Assessing the type of documents underneath this distribution reveals that these higher quality documents of $P \geq 0.7$ are nearly exclusively non-ISI publications highlighting that the format of technical reports better suits the needs for event documentation.

4.3.4 Application: Event analysis example

In the following we will demonstrate the applicability of the quality assessment in the context of understanding trans-basin flood events in Germany. In order to improve the readability of the text we will use the record number of a report instead of the full citation to a document. The record number is a unique identifier which allows the identification of the document via the endnote database and evaluation tables provided in the data supplement. A summary for other flood events can be quickly obtained by querying the data supplement for the rank of the trans-basin floods (Uhlemann et al., 2010a) which serves as a unique identifier, too. In the case of a document containing analyses of more than one flood it is listed several times; each time annotated with the respective trans-basin flood rank for which it has been evaluated using the QAF.

A full assessment of all flood events is beyond the scope of this paper and we use the example of the flood event in October-November 1998. According to the analysis of Uhlemann et al. (2010a), the flood event of autumn 1998 ranks as the 5th strongest trans-basin flood in Germany in the period 1952-2002. It affected all river basins in Germany except for the

Table 4.4: Overview of the flood event reports available for the flood in Oct.-Nov. 1998 and their quality assessment. For each document the respective record number, the issuing body, the Pedigree and the vector of the scores assigned in each quality dimension (in the order given by the QAF, aligned to spatial scope of the report) are given. The contextual dimensions are highlighted in bold.

			QD-Vector aligned to spatial scope of report			
Record No.	Issuing Body*	P	EU	National	Regional	Local
# 482	DWD	0.57	[3|1-2-1|**2-0-0**|3-3-2]			
# 159	BfG	0.67		[2|1-3-2|**2-2-0**|3-3-2]		
# 94	BfG	0.47		[2|1-2-2|**0-1-0**|2-2-2]		
# 148	LA BW	0.80		[2 | 1 – 2 – 2 | **3 – 3 – 3** | 3 – 3 – 2]		
# 28	LA RP	0.67		[3|1-3-2|**2-2-0**|2-3-2]		
# 486	LA BY	0.50			[1|1-2-1|**2-1-0**|2-3-2]	
# 37	LA NI	0.23			[1|1-1-0|**1-1-0**|1-0-1]	
# 30	Union	0.63				[1|1-3-3|**2-2-0**|2-3-2]

* Abbreviations: DWD – German Meteorological Service, BfG – Federal Institute for Hydrology, LA – State authorities of Baden-Wurttemberg (BW), Rhineland-Palatinate (RP) and Lower Saxony (NI), Union – Water management company for the Ruhr catchment.

Odra, lasted for 14 days from October 29th to November 11th 1998 and is the only autumn flood in the record. Compared to other flood events in the 90s (Rhine floods in 1993 and 1995, Odra flood in 1997, Danube flood in 1999) the event has received limited public and scientific attention. In total ten documents were identified in Uhlemann et al. (2012, 2013a), eight of which belong to the group of special reports. These reports have been evaluated in the course of this study using the quality assessment framework. Table 4.4 provides an overview of all studies and the results of the quality assessment. We illustrate the spatial scope of the reports by aligning the vector of scores assigned in each quality dimension to four classes of spatial extent (see chapter 4.3.2). Three of the reports cover the flood on a national scale with the report of the German Meteorological Service (#482) making also reference to other extreme weather and flood occurrences on EU scale. Four reports have a largely regional focus with (# 148 extending in the analysis both to national and regional aspects) and one report documents the flood in the smaller catchment of the river Ruhr (#30). The latter is the only report issued by a local water management company and all other reports have been issued by governmental authorities on federal or state level. The reports are exclusively drafted in German. The regional to supra-regional reports can be easily accessed as they are indexed in public access catalogues and two of them are even fully open access. In turn, the three reports with a regional to local scope are exclusively available as downloads (#30) or web-text (#37 & #486) at the respective sites of the providers but have not been indexed.

The result of the quality assessment for the documents of the 1998 event reveals a wide spectrum of document quality that resembles the distribution of P in the parent population, with $P_{min} = 0.23$ and $P_{max} = 0.8$ and a median of $P \sim 0.6$. Except for #37 all reports present their analysis in a largely accurate and objective manner, however, most exhibit minor to substantial shortcomings in the way they are presented (in most cases due to a very condensed presentation and in case of #148 due to deficiencies in the structure of the otherwise very detailed and useful content). Considering the low quality of #37 we exclude the study from further analysis. The content presented in the reports largely focuses on the analysis of the sources and pathways of the flood event. Only #148 includes direct notions on the damages encountered thereby constituting one of the most complete documentations of damages that can be found in the entire set of documents analysed in the course of this paper.

In the following we will present a synthesis of the information contained in the documents on the SPRC of the 1998 flood event, highlighting where information is complementary, (non-) confirmative or missing. As indicated, the information contained in the seven reports we use for the synthesis can be considered as largely trustworthy (on average) with differences in *P* resulting from various levels of depth in the contextual dimensions. In order to keep the scope of this paper the synthesis is necessarily kept short.

Sources: In the course of a persistent predominantly westerly circulation from mid to end October, connected with storm activities, a series of frontal systems transported moist Atlantic air masses over the central European continent (#486, #28, #148). #482 highlight that the scale of the frontal system extended on a W-E extension from the British isles to Germany causing also flooding e.g. in England/Wales. The flood in Germany was caused by a series of widespread extreme heavy rainfall events (locally very heterogeneous) (all reports) that met widespread saturated soil conditions due to an anomalous sum of precipitation of the month preceding the flood (all reports). The heavy rainfall of the flood was influenced by orographic enhancement and foehn (increasing the temperature and pressure gradient) (#486). Both the sums of precipitation in the month of October as well as the event precipitation sums (sub-hourly, hourly, daily) were unprecedented in several places (#148, #30) and were assessed with very high return periods or %-anomalies for many regions (all reports).

Pathways: The runoff from the flood initiating rainfall events met already high flow conditions in smaller catchments of central north Germany and also Bavaria (#30, #486) and low to mean flow conditions in the main rivers – Rhine, Danube, Elbe, Weser, Ems (#159). Only #159 provides some few notations on Elbe, Danube, Ems and Weser (main rivers). The flooding exhibited some flash flood characteristics in smaller catchments of Baden-Wurttemberg (#159, #148). Two main flood waves were registered, with the first larger in western and central regions (#148, #28), i.e. mid-altitude mountains, and the second larger in alpine (#159) and south-eastern catchments (#486). The main rivers were affected at a gradient south-north, with the upper and middle Rhine experiencing Q(T<5a) (#148, #28) and higher high flows with increasing contributions from tributaries Neckar, Main, Moselle. Generally all reports confirm that the most severely flood affected regions were predominantly small to medium sized (mountainous) catchments. Technical retention options were not operated in river Rhine (#28). However, a substantial amount of flood retention basins and dams were in flood operation in mountainous catchments (#148, #30). #28 confirms that the upper Rhine was not and the middle Rhine only lightly affected by inundations the latter resulting in no consequences for adjacent communities. A detailed description (no map) of inundation mechanisms and areas is only provided in #148 for Baden-Wurttemberg.

Receptors and consequences: An estimate of the overall national damage is given in #148 with 250Mio DM. Both #148 and #159 are concordant in their overview of the most affected geographical regions on the national scale. The overview of affected regions (#94, #148, #159) indicates that substantially more areas were affected by flood damages than are included in the reporting. Two centres are identified: 1) the north German basins of Weser and Ems and 2) the northern part of the state of Baden-Wurttemberg and adjacent regions. For the latter, #148 provides a very detailed documentation of the types of damages and sectors affected with special sections on particularly affected local regions (city reports). However, for the state of Bavaria #486 provides no damage analyses. Similarly, no document provides a damage description for the basins of Weser and Ems.

4.4 Summary and conclusions

Written sources that aim at documenting and analysing a particular flood event in the past in Germany are published at vast majority as grey literature (e.g., as technical reports) and therefore outside of the scholarly publication routes. In consequence, the application of event specific documentations in flood research has been constrained by barriers in accessibility and concerns of credibility towards these sources and by limited awareness of their content and its applicability. The problem of accessibility has been discussed in detail in Uhlemann et al., (2013a). By presenting an open access database on the bibliographic metadata of flood relevant sources they have significantly reduced the barriers in accessibility.

In this study we address the concerns of credibility for the first time and present a quality assessment framework for written sources from a user's perspective, i.e. we assess the documents' fitness for use to enhance the understanding of trans-basin floods in Germany in the period 1952-2002. By this approach we have filled the methodological gap that existed for quality assessment of grey literature. Our results highlight that, in order to obtain a complete understanding of the content presented in a document, its trustworthiness and applicability for a particular research question, more than the intrinsic quality measures common to approaches of critical research assessment need to be applied, i.e., the quality of a document is determined by accessibility, representational, contextual and intrinsic characteristics. The user's perspective is an essential strength of the framework as the quality tag attached to a document is reflective of the actual task at hand. Therefore it does not label a document as good or bad per se rather than its fitness for a particular use.

Our validation of the framework by two independent peers highlights the objectivity and transparency of the approach. We have shown that the overall agreement reached between the peers is substantial, that disagreement is in the range of deviances that can be expected from the degrees of freedom in the interpretation of the score definitions and from the influences of the peers own mental framework and perceptions, and that the overall effect on the quality measure *P* is low.

For the first time, we provide a picture of the overall quality of the body of event specific reports for the considered task at hand. We show that the majority of flood event specific reports are of good quality, i.e. they are well enough drafted as well as largely accurate and objective which allows applying their content for the interpretation and understanding of large-scale flood events and processes in Germany. Our assessment highlights that the majority of reports present a substantial amount of information on the flood initiating processes and the processes of wave propagation and inundation. The complete display of the inundation extent is addressed rarely and out of 105 reports only two reports provide a mapped overview of the inundation area. The effects of the flood on the exposed receptors and the degrees and types of damages that occurred in the course of the event are less frequently and often not completely covered. The application of the approach of this study to the flood in October/November 1998 highlights the opportunity of deploying complementary, confirmative or non-confirmative information available in a range of documents. Using the pedigree quality measure conjointly with the quality dimensions allows gaining a quick overview of the contextual aspects covered by a report for any particular trans-basin flood event.

A natural extension of the example application presented is the combination of data-based and model-based results with the quality-labelled information of the reports resulting essentially in an uncertainty assessment of the available knowledge. We want to stress that the scheme provides a guideline for the overall quality to be expected and the depth of information presented by a report. It does not free any user of critically checking the accuracy of single information entities before using them in scholarly argument. Evidence-based or related methods are a natural successor of the results of this study that can assist in combining quantitative and qualitative measures of uncertainty (e.g., van der Sluijs et al. (2005)).

The analyses and conclusions presented in this study are valid for flood reporting in Germany. The framework of quality assessment can certainly be adapted to other countries as well as to other natural hazards. In that respect it will be particularly interesting to apply the approach also in other European countries in order to gain an understanding and overview of the flood relevant material residing off scientific shelves and, if applicable, to compile a comprehensive inventory of information (that may include data) and knowledge available for floods in central Europe.

We see the main benefit of our study in its contribution towards reducing the barriers for using information of flood event reports in the context of flood research. Inasmuch as the literature presented and evaluated relates to the particular aspect of trans-basin floods in Germany, the task at hand was phrased widely enough so that the contextual aspects can be interrogated also for related research questions. We provide the results of this study as an open access data supplement and therefore any user can quickly reference into and, by means of the quality dimensions and adjunct quality measure, access the literature in a more directed way.

Acknowledgements. We thank the GFZ German Research Centre for Geosciences and the University of Potsdam for their financial support. The main author is grateful for partial funding through the excellence in teaching programme (Junior Teaching Professionals) of the Potsdam Graduate School and a scholarship by the equal opportunities office, both University of Potsdam. We want to thank the authorities and respective individuals who provided information and access to relevant publications for this study, assisted in their retrieval and for donations. We dedicate our special thanks to Dr. Florian Elmer and Dr. Philip Bubeck for taking over the peer roles in the kappa test and for their valuable comments. The work of the students Florian Betz and Sylvia Wesser who assisted in the application of an early version of the quality assessment are gratefully acknowledged.

Data description. The data used for this publication is freely available as data supplement under the creative commons license and can be permanently addressed following the doi given in Uhlemann (2013c).

Appendix

Table A1: Definitions of the quality dimensions and their respective scores for the quality categories 'Accessibility' and 'Representational Quality'.

Category	Accessibility Quality	Representational Quality		
Dimension	Accessibility	Interpretability	Ease of understanding	Concise Representation
Definition Score	The extent to which the report is available, and easily and quickly retrievable.	The extent to which report is in appropriate language(s) and can be comprehended.	The extent to which report is presented in an intelligible and clear manner without ambiguity and can be easily comprehended. Relating to formal aspects of style, structure, writing.	The extent to which the report is compactly presented without being overwhelming or too coarse considering its own objectives. Balanced use of fig./tab./photos.
0	Citation not indexed in any OPAC/ SCI. Document not retrievable online or via IL. Donation, archive material.	Language other than English and that of the region described.	Poorly presented, serious flaws in usage of language and terminology, almost not comprehensible, extremely poor appearance of text, figures.	Few lines or paragraphs only, not representative for objectives, or content cannot be related to objectives.
1	Citation not directly indexed in OPAC/SCI (articles in non-indexed journals) or with terms other than SS or indexed in non-public catalogues. However retrievable through IL.	Native language of the region of interest + English title (optional)	Largely comprehensible in use of language, few ambiguities. However, not well structured document (e.g. title not meaningful, no headings, inconsistent order of topics, poor appearance/ readability of parts of document, e.g., poor figure quality).	Strongly condensed content for objectives given. Or too much unnecessary detail (inadequate figure-text-ratio: Report consists mostly of photo documentation or figures in relation to explanatory text).
2	Citation indexed in OPAC/SCI, can be found through SS. IL, no OA. OR: Document not indexed but retrievable as OA from institutional homepage.	Native language of the region of interest + English title, abstract/ summary, table/figure captions	Succinct piece of writing with a fairly clear structure. Some shortcomings in appearance, terminology, and structure.	Largely balanced report (with respect to detail and figure/ table use), yet inconsistent in parts (variance between the parts).
3	Citation indexed in OPAC/SCI and open access to full text. Found through SS.	All English	Succinct piece of writing with a clear terminology, appearance and structure (meaningful title, summary, clearly stated objectives and a logical order).	Well balanced reporting in all parts (appropriate figure/text ratio). Appropriate size of document with respect to objectives.

Abbreviations: OPAC - Online public access catalogue, SCI – Science Citation Index, OA – Open access, IL – Interlibrary loan, SS-Systematic search

Table A2: Definitions of the quality dimensions and their respective scores for the quality category 'Contextual Quality'.

Category	Contextual Quality		
Dimension	Source	Pathway	Receptors/Consequence
Definition Score	To which degree are the atmospheric processes, catchment state and runoff processes described that have lead to the flood?	To which degree are the processes of flood propagation in the river and the inundation effects described?	To which degree is the occurrence or absences of damages described (separated by affected sectors (private households, business) and damage/loss types (direct, indirect; fatalities)?
0	Not/insignificant part of the report	Not/ insignificant part of the report	Not/ insignificant part of the report
1	Rough typecasting of the hydrometeorological causes. The spatial and temporal dependencies/developments of the flood formation processes are only coarsely sketched. Largely condensed and descriptive with little quantitative evidence.	Rough typecasting of flood wave development. Little coherence in the description of the flood propagation and inundation (or no notion on inundation effects, i.e. only routing charts and/or return periods given). Reference to single aspects, like one dyke breach.	Damages/no damages only described for arbitrary locations, often qualitative only, not distinguishing types of damages and affected sectors. AND/OR: Overall damage estimate is given without further differentiation.
2	Either quantitative, detailed analysis of one aspect of the flood formation OR mostly qualitative analysis of all relevant processes (atmospheric, catchment state and runoff) that have lead to the flood. Evidence on spatial and temporal courses is given.	The flood wave propagation is described and a more or less coherent picture on the course of the flood over time and space is drawn. Notions on defence failures or ad-hoc measures, mostly qualitative. Some indications on flood extent or inundation mechanisms.	Estimation on the overall (expected) damage is given. Some damage numbers are provided distinguishing the affected sectors and types of damages allowing for limited conclusions on the flood impact in the area. OR: spatial reference to unaffected regions
3	The proportions of and the spatial and temporal correlations amongst the flood generating processes are given (quantitative analysis). The dominating CP(s) before and during the event are described. At best indications of advective, convective components, orographic enhancement.	The flood propagation, the influence of dyke breaches, peak amplifications, retention etc. are clearly documented in their causes and in their effect on the flood crest. The course of the flood wave can be clearly understood in its spatial and temporal development. Flood inundation extent is fully documented.	The affected sectors and types of damages are documented in a spatially explicit manner and allowing identifying hotspots of damage and unaffected regions.

Abbreviations: CP – Circulation pattern

Table A3: Definitions of the quality dimensions and their respective scores for the quality category 'Intrinsic Quality'.

Category	Intrinsic Quality		
Dimension	Accuracy	Objectivity	Reputation
Definition Score	To which degree are the analyses and results reliable (i.e. quality and amount of data and/or sources used, application of appropriate methods)?	To which degree is the report unbiased (unprejudiced) and impartial (independent of political, religious, personal, institutional, business interests) and therefore objective?	Is the author/institution reputable, i.e. does it hold scientific or professional experience in the field? To which degree has the document undergone quality control (reviewing)?
0	No notation on the origin of data/sources and methods on which the results are based	Strong bias, the report clearly falsifies results (faked statistics; strong exaggerations or understatements)	No verifiable source or author. Source or author cannot be identified or affiliated. Quality control of the document cannot be assumed.
1	Quality and amount of data and/or information sources unclear or mainly raw data (uncertified/ unofficial) used. Limited documentation of methods. Results reproducible to a limited degree.	Some aspects are objective, some not; biases can be either inferred directly or assumed from the choice of results that is presented, i.e. results/interpretation not fully supported by evidence.	Reports from institutions or authors with technical reputation but mostly not directly flood related, i.e. core experience in related disciplines. Degree of quality control is either unclear or is assumed to be low.
2	Largely replicable analysis, minor inconsistencies (i.e. documentation of methods); most data used is official/certified, largely sufficient amount of data or information from secondary (reputable) sources.	Most aspects are objective, no apparent falsifications, exaggerations etc. However influence of interests on (the choice of) results cannot be fully excluded. Results/ conclusions largely supported by evidence.	Reports from institutions/authors with technical reputation in hydrology. Some quality control of the document can be assumed (internal quality check, however no peer reviewing).
3	Official/certified data/sources of adequate amount. Standard and documented methods applied; discussion on accuracy of data and/or uncertainties provided.	Impartial, objective work. Results presented supported by evidence. No evidence of political, religious, personal, institutional, business interests on results of the study.	Report from authors or institutions with clear flood/hydrologic expertise; (Peer) reviewed article or consortium work that has been cross-checked by several resorts

Chapter 5:

Discussion, conclusions and outlook

5.1 Summary of achievements

The objective of this study was to investigate flood events that at the same time affect more than one river basin and to provide the data and information basis on which knowledge about this type of floods can be generated in order to improve any future flood risk assessment. In this study a sequential approach has been followed starting with the identification of trans-basin floods and a consecutive identification and quality labelling of relevant documentations for the events. In the following, the main findings of this work are summarized according to the research questions proposed:

1. **How can trans-basin floods be identified in Germany?**

In order to systematically capture the simultaneous occurrence of flooding in Germany we developed a method based on time series of mean daily discharge at a maximum number of stations of even time series length (1952-2002). The method identifies the spatial and temporal dependencies amongst peak discharges based on physical reasoning. That is, we consider both the option that in the course of a synoptical weather pattern the onset of a flood can be at geographically distant locations not belonging to the same river basin and that is time-lagged due to the drift velocity of the weather system, and we consider the expected times of concentration in the catchments and travel times in the channel network that lead to time lagged recordings of peak discharges as the flood wave progresses downstream. We can show that these signatures are preserved in the time series of river discharge which allows defining coherence strictly in terms of the timing of the peak discharges. The presented adaptive algorithm for event identification is based on an initial exceedance of a threshold of a 10 year flood at one location and it consecutively pools all related peak recordings first within a fixed time window around that peak (both forward and backward, accounting for dependent peaks both in nested and unnested catchments) followed by a moving time window that accounts for the evolution of the flood in the various river networks. In result, this study, for the first time, presents a complete and consistent set of trans-basin floods for Germany in the period between 1952 and 2002. Each flood is characterised by a specific value for the timing, the location and the magnitude of discharges within the entire river network.

2. **How can we define a measure of severity for trans-basin floods that can be used for a frequency analysis and what are the frequency characteristics of the observed sample of trans-basin floods?**

Estimating the probability of a flood scenario is a basic requirement for any flood risk assessment. Assigning this estimate to a spatial pattern of flooding necessitates the definition of an adequate quantitative measure that represents both the intensity and extent of the scenarios. In this study we propose a weighted cumulative index that accounts for the spatial extent of the flood and the individual flood magnitudes reached and therefore allows quantifying the overall event severity and ranking of the events. By regionalising the flood peaks to the river network we remove the geographical bias caused by the location of the gauges. The index cumulates only those regions that are affected by inundations (i.e. those exceeding the bankfull flow under the assumption of no flood defences). The measure therefore provides a first best estimate that can be used for the estimation of event probability.

Our results indicate that in Germany trans-basin floods are a frequent phenomenon, with 80 events detected in the entire 51-year period. We can show that seasonality and consequently different process regimes have a marked effect on the flood intensity, location and frequency of the events. The set is dominated by events that were recorded in the hydrological winter and the most severe floods according to our severity index are also winter floods. By analysing the occurrence of events over time we find that the events tend to cluster in time, with periods of frequent, often even multiple floods per year, and periods with few occurrences. We find an increase in the percentage of winter flood occurrences in the second half of the observation period which coincides with an increase in frequency of extreme trans-basin floods. It will be important to consider these empirical findings in the methodological development of a full spatial flood risk assessment.

3. **Can and should flood event documentations be used as another source of data for understanding trans-basin flood events in their sources, pathways and consequences on particular receptors?**

For Germany, a large body of literature has been produced that contains relevant information on any of the aspects sources, pathways, receptors and consequences of trans-basin floods. The systematic search approach developed in our study shows that this body can be efficiently searched for with tools widely available to the scientific community. For the majority of the flood events we can obtain event documentations and especially the most severe floods have received extensive coverage. The largest part of the material can be considered grey literature, i.e., literature not controlled by commercial publishers.

We address the concerns of credibility towards flood event specific reports and propose a novel framework for the quality assessment of written sources. The framework is designed to be generally applicable for any natural hazard event documentation and, from a user's perspective, assesses the quality of a document addressing accessibility as well as representational, contextual, and intrinsic dimensions of quality. We can show that the majority of flood event specific reports are of a good quality, i.e. they are well enough drafted, largely accurate and objective, and contain a substantial amount of information on the sources, pathways and receptors/consequences of the floods. We therefore recommend including this information in the process of knowledge building for flood risk assessment. To facilitate this process we have published the data and results from this study in an open access format which allows using our results as a knowledge base for trans-basin flood events in Germany.

5.2 Discussion, synthesis and directions for further research

This work addresses fundamental questions on the nature and occurrence of trans-basin floods in Germany and with the development of a consistent and complete set of trans-basin floods and the quality labelled collection of documentations on these events a data and information base has been obtained that contributes to an improved understanding of trans-basin floods and their role in flood risk assessment. From the tasks addressed in this work and the synthesis of the results a number of aspects can be identified that need further discussion and that point to directions for further research. Where appropriate a discussion of the results with respect to recent research findings that have emerged in the course of this work is included.

5.2.1 Modelling spatial dependence

For a spatial risk assessment that is directly based on river flow data it is necessary to identify a statistical model that describes the joint distribution of flood events in multiple locations allowing for extrapolation to more extreme levels than observed. This also pertains to trans-basin floods in Germany and it is necessary to discuss how the empirical findings of our study can be usefully extended into a full spatial flood risk assessment.

Recently, two approaches have been proposed. The first defines the statistical model of dependence on a predefined set of events (Ghizzoni et al., 2012). We call this an informed or event based approach. The second is an uninformed approach which can be described as a multivariate peaks over threshold approach. It derives the spatial dependence structure using the entire series of daily flow data under the condition that one site exceeds a (high) threshold (Heffernan and Tawn, 2004; Keef et al., 2009a; b). The multivariate dependence structure is modelled using a copula function in both approaches and Ghizzoni et al. (2012) additionally employ the t-skew distribution. The main difference is that the parameters which relate to the hydrologic processes like the temporal lags between peak observations are applied in the step of event selection in the first approach and the multivariate model is fitted a-posteriori to the vector of peaks for the selected events. In the second approach the temporal lags are included directly in the statistical model. For both approaches the advantage over methods that assume independence between locations has been clearly shown highlighting the substantial overestimation of risk in the independent case.

The question arises: Which approach is best suited for a spatial flood risk assessment in Germany given the findings of the nature of flood occurrence in our study? In as much as it seems obvious that with the event set already created the informed approach should be a natural successor of our study, it is useful to evaluate the uninformed approach as well.

The latter has found wider application (Keef et al., 2009a, b; Mendes and Pericchi, 2009; Lamb et al., 2010; Faulkner et al., 2012; Neal et al., 2012; Keef et al., 2013) and has been extended in multiple ways. Besides including the temporal lag in the dependence model, Keef et al. (2009b) also account for missing data in the procedure. Recently, the model has been extended to account for the typically observed clustering of events in time which leads to an invalidation of the assumption of a Poisson distribution (Keef et al., 2013). The latter is particularly interesting considering the strong clustering observed in trans-basin flood events in Germany. A general strength of the model is the inclusion of the entire time series in the dependence model. For one this increases the number of data points that can be used to define dependence and further it allows modelling also those events that are spatially not necessarily large. This is charming when the risk assessment is to include also the probability of accumulated damage due to a number of small events within a fixed period of time (1 year).

The informed approach carries its particular advantages in the opportunity to use a maximum of process knowledge in the identification of the event set which is input into the statistical dependence model. Further, the dependence model is considerably less complicated in its setup, i.e. only the copula or t-skew function has to be established without any additional parameter estimation. However, some difficulties arise as the method tends to underestimate the frequency of rather localized events (Ghizzoni et al., 2012) and overestimates that of very large events which might be due to the limited number of observations (comparable to our case).

Both approaches have not been used on the same set of data so far, which limits the options of comparison of the model results. For a start we would recommend to commence with the more straightforward informed approach. However care has to be taken on the effect of the different process regimes, the year to year dependence and the observed increase in flood frequency and magnitude on the dependence structures. For the process types splitting the sample might provide a solution.

The method of Keef et al. (2009a, b) has recently been applied in a probabilistic framework mapping flood risk using a hydraulic model in order to derive probabilities for patterns of inundation (Neal et al., 2012) (for a small scale application though) and in a full risk assessment using a simple model chain from inundation to damage modelling for a regional application in NE England (Lamb et al., 2010; Environment Agency, 2011). Both studies highlight that the bulk of uncertainties in the risk assessment is associated with the flood impact model chain and the necessary simplification of processes needed in order to allow calculating inundations and/or damage for a very large number of events. In the light of the substantial improvements still needed to reduce the uncertainties in the modelling of flood impact (on the large scale), the question for the choice of the dependence model seems more like a technical side note.

5.2.2 Flood magnitude

One critical aspect in the assessment of spatial flood risk is the assignment of a probability estimate to a spatial pattern of flooding. This necessitates the definition of an adequate quantitative measure that represents both the intensity and extent of the scenarios. In this study we proposed a weighted index that accounts for the spatial extent of the flood and the individual flood magnitudes reached. We regionalise the flood peaks to the river network adjusting the station network to affected river length. This allows us to remove any geographical bias that may result from interpreting the severity of an event based on the number of stations exceeding a particular flood threshold. It further allows mapping the flood event completely, highlighting the regions affected by inundations (i.e. those exceeding the bankfull flow under the assumption of no flood defences), regions with a flood response but no inundation and regions that did not show any reaction during the event. The measure therefore provides a first best estimate for the assessment of the overall event probability. However, by using the median annual flood as threshold for assuming bankfull flow, we impose the strong assumption of near natural river morphology which does not hold for most of the major rivers considered in this study.

This highlights an important fact: As the damage caused by a flood depends on the vulnerability of the inundated region, the probability of the hazard scenario may severely differ from that of the damage scenario. Therefore, the only regionally unbiased measure suitable to attribute probabilities to the flood scenarios might actually be damage, given e.g. as total economic damage during one event. This necessitates extending the spatial dependence model by transforming the flow pattern into a pattern of inundation using a hydraulic model and in consequence the translation into damage estimates. For a national approach this still poses severe challenges on data availability to set up the hydraulic model (parameters for channel morphology, floodplain topography including knowledge about defence structures) and for the estimation of the damage for a large number of events to be modelled. Therefore this approach comes at the cost of introducing numerous additional sources of uncertainty in the estimate. To our knowledge, no comparison of the effect of using different measures to derive the probability estimate for large scale flood scenarios has been presented so far.

5.2.3 Understanding trans-basin floods

Floods are the result of highly complex and inextricably linked processes involving atmospheric conditions, runoff generation and concentration in the catchment, in-channel wave propagation and vulnerability in the flood plain. Understanding these processes and interactions is at the heart of flood risk management where the question needs to be addressed which measures are best suited to reduce flood risk. This becomes even more important when the behaviour of the system is dynamic and when changes in the flood risk are observed or expected.

In this study assess whether documentations and reports on flood events can provide a valuable data source for understanding flood events. We conduct a systematic search for relevant material and find that the majority of events has been reported on and that especially the most severe floods in the record are very well covered by reports. The majority of reports present a substantial amount of information on the flood initiating processes and the processes of wave propagation. The complete display of the inundation extent is addressed rarely, however, information on breach occurrences or volume and effects of retention on the flood peak are documented frequently. The degrees and types of damages that occurred in the course of the event are less frequently covered. Our results highlight on one hand that documentations provide a considerable amount of flood event information, however, that the resulting information is inevitably incomplete and inhomogeneous with respect to the depth of information and spatial coverage of an event by documentation.

Considering, that the more common way of approaching the issue of flood process understanding is a (observational) data or model based analysis the question may arise: What is the advantage of collecting already pre-processed information like that in event documentation over a data based analysis and is, what at a first glance may appear as puzzle solving, tedious work of collecting documentary evidence from a fragmented information landscape, worth its salt?

We will address this question first from the point of understanding the sources of floods. Linking floods to the physical processes is a common way of inferring the dominant flood producing processes. Using either a data based statistical or a simulation based approach, the aim is to provide an (automated) classification of the underlying hydro-meteorological processes and their space-time characteristics. For the European regional scale a number of studies have been presented analysing single process domains like circulation patterns and precipitation (Bárdossy and Filiz, 2005; Jacobeit et al., 2006; Petrow et al., 2007; Müller et al., 2009; Prudhomme and Genevier, 2010) and a few analysing combinations of processes (Merz and Blöschl, 2003; Viglione et al., 2010b). Recently, Nied et al. (2013) present an automatic classification method for large scale soil moisture patterns for the Elbe river basin that are linked to flood events in the region. A complete quantitative approach integrating all processes has not been presented for the national scale. In a parallel study to the work presented here we have also undertaken an analysis of the flood governing processes in two ways, 1) using a qualitative approach deriving a process typology based on hydrometeorologic indicators from time series of rainfall, discharge and Reanalysis data (Uhlemann et al., 2009), and 2) using a conceptual semi-distributed hydrologic model aiming at a quantification of the space-time dynamics of the processes (Uhlemann et al., 2010b). The conducted work and presented studies highlight some limitations of the approaches that aim at large scale flood understanding: Concerning the role of the catchment storages for flood initiation, their dynamics and spatial patterns, some critical parameters are hard to address as observations are either not available, scarce, or of limited temporal or spatial resolution (hence requiring model assumptions). These concern soil frost, snow accumulation and melt, and snow water equivalent. For rainfall formation the relative contributions of advective processes, convection and orographic enhancement have also not been investigated in detail and require strong disciplinary skills in their identification. On a general note, for an integrated hydro-meteorologic analysis the effort in data acquisition, spatial statistical analysis and/or model set up, calibration, and validation is still challenging on a large scale.

The event of October/November 1998 which has been used as a case study for assembling the information contained in flood event reports (Chapter 4) highlights that the role of a number of the outlined critical parameters as highlighted above, i.e. the influence of orographic enhancement and foehn on the precipitation field, are treated in the event documentation at hand. For another example, the flood event of March 1988, we can infer from the event documentation that the flood peaks have been largely produced from the water stored in the snow cover, i.e. they were the result of an extremely high snow water equivalent and a considerably light rainfall at the onset of the event.

For a complete risk assessment understanding the sources of flood events provides answers to only one third of the processes that lead to a certain flood impact. As we have outlined earlier, the understanding and modelling of the pathways (i.e., modelling the inundation in the floodplain and responsible mechanisms like dike breaches or overtopping) and the resulting damages (e.g., number of affected buildings, economic damage) is confronted with stronger limitations in both the availability of data and in the available large scale modelling approaches. For a more detailed evaluation of hydraulic models for large scale flood risk assessments refer to Falter et al. (2013) and for a review on data availability and quality as well as the state of models for damage estimation refer to Merz et al. (2010).

Returning to the example event of Oct./Nov. 1998, the reports in the database, for example, contribute an overview of the flood affected regions and an estimate of the overall damage. For the flood of 1988 the operation of flood retention polders in the upper Rhine has been documented, its effect on the flood wave as well as the installation of mobile flood protection walls in the city of Cologne. Both ad-hoc defence measures resulted in the protection of the city of Cologne that would have otherwise been flooded, increasing the total damage of the event.

Returning to the question proposed: Based on the review of the current state of understanding of trans-basin floods (in Germany), both from the perspective of data-based or model-based approaches as well as from the perspective of event documentations, we can conclude that the knowledge on the general processes as well as on the single events is still incomplete. Both approaches are worth their effort and we see the highest potential for improvement in a combination of both, which should be used complementary and possibly in an integrated fashion. The systematic collection of event documentation for a large number of events opens the opportunity to include information from documentation in the process of model development or validation. The often detailed information contained in flood event reports can assist in accounting for those particularities in flood generation and impact that cannot be obtained from data or modelling alone. In a recent study, Van Dyck and Willems (2013) highlight the great potential of including information from a systematic collection of documentary evidence on a set of flood events in a probabilistic modelling framework for assessing the risk over large geographical regions. They searching for information both from public media (newspapers, broadcasting …) and flood reports from public authorities, scientific reports etc. The systematic approach allows them to account for the inevitable incompleteness in the record of documentations. From the analysis of the entries in the database they obtain the spatial extent of the events in terms of areas affected by inundation. This allows them calibrating the otherwise conceptually simple probabilistic model for large scale flooding and allows for a fast flood risk assessment (case study Belgium).

Vice versa, process knowledge and in particular classifications of dominant processes based on statistical analysis or simulation can provide systematic information that can be used to complement the knowledge obtained from flood reports and should be usefully combined in an event knowledge base.

5.2.4 A knowledge base on trans-basin floods

This study has assembled a wide array of different data sets that provide the basis for a deeper understanding of widespread flooding in Germany and can assist in a national flood risk assessment. We present the results of this study in the form of open access publications which includes the publication of the data produced in chapters 3 and 4 ensuring long term availability and accessibility of the data and the adjunct documentation of the methodology used to create the data. Both the data publications as well as the research publications are cross referenced. Each of the data sets produced can be related using the single flood events as a common identifier (through the event rank or date). All sets together provide a knowledge base on trans-basin flood events in Germany that includes notations on the main characteristics of each event, i.e., date and duration, severity, the geographical space affected, as well as the amount, quality, depth (given by the scores in the contextual quality dimensions of reports) and location (spatial scope of a report) of information available on the sources, pathways, receptors and consequences. Therefore the results of this study are the first step into a structured deposition of content.

In order to tap the knowledge base to its fullest potential the next necessary step is the development of a conceptual and technical framework that allows for knowledge integration. This framework needs to be flexible enough to allow for best use both with respect to extending and updating the knowledge base as well as for querying the content for other questions than the ones followed in this study. This framework could be very simple, using common identifiers to link separate data sets. This is the current technical state. However, at its best it provides the relational capacities to deal with a multitude of different sources (data and text) allowing also to search for single information entities within these sources.

For the latter a number of obstacles have to be overcome and will probably not be achieved easily. At the moment, the documents that have been obtained in the course of this study cannot be accessed directly and they are included in their meta-data only. So far any user of the data base who is interested in a particular entity of information, i.e. what was the influence of snowmelt in the flood genesis of event x, can only defer the approximate depth of information in event documentation from the score given in the quality assessment in the contextual category. This allows addressing the documents of relevance in a directed way; however, it does not grant immediate full text access necessitating the acquisition of the document by the user through any information provider.

Digital full text documents are a precondition to apply advanced text-processing tools (text mining, semantic search tools) that facilitate the efficiency of any search query and therefore improve largely the speed of information reception including that of machine information processing. For now, any information within the documents can be processed only intellectually. Another problem is that of language which imposes a strong barrier on information access in an international research context as 85% of the material included in the dataset is written in German.

The current barriers in language and accessibility of the collected documentations may necessitate the step of an a-priori synthesis of flood event information. This synthesis will be particularly useful on the level of the single flood events; however it will require the definition of standard parameters and a method for combining evidence from multiple sources (including the geographic referencing). This can facilitate the use of the knowledge base on two aspects. The standardised form allows for a very fast longitudinal overview on the availability of information about a flood event and it allows for comparing events and possibly classification along processes or single factors. This will largely facilitate the use of the knowledge base. The initial cost (work load) for setting up the synthesis and further in maintaining the knowledge base (including new data…) is however relatively high. However, if designed well the opportunity to deploy the standardised attributes in tasks such as the rapid assessment of a new event are quite promising.

5.2.5 Transferability to other regions and periods

This study presents the results of a national assessment of trans-basin floods in Germany and for the period 1952-2002. As floods are not confined to administrative borders and as they occur throughout time it is interesting to discuss the options of extending the analyses of this study both to other regions and periods.

In principle, the transferability of the method for event detection to smaller spatial (basin) scale is easily done and has been shown in a recent study for the Elbe basin (Nied et al., 2012). It requires a check on the length of the time window and independence criteria chosen, i.e., Nied et al. (2013) use the same time window as in our study and a 4 day independence criterion. The sensitivity analysis of our method shows that the number of stations can be reduced by as much as 50 stations without losing much of the information content. For the national context this allows extending the period of analysis for ten years back in time or even further (1920s), however with some caution due to regional biases in the station network for early years. As the station network for the chosen major rivers (catchment size >500km²) is being continuously maintained the option of extending the analysis for the period past 2002 is easily possible leaving the effort to data acquisition. The recommendation of extending the event set in time is self evident as this increases the statistical sample for any frequency assessment. More importantly, it also allows verifying and possibly better explaining the detected temporal dependence of events with periods of high and low flood activity and the shift towards more frequent winter floods (and therefore towards more frequent severe trans-basin events) in the current set of observations.

Extending the analysis to a central European context, i.e. considering the entire river basins of Rhine, Elbe, Danube and Odra poses some data problem as in particular river flow data for the eastern parts of the Odra and Elbe are more difficult to access. This highlights the importance for the hydrologic community (both public and administrative) to continue the efforts in reducing barriers in hydrometeorological data exchange and the importance of river flow archives like the Global Runoff Data Centre (GRDC) or the European Water Archive (EWA) for research (see Hannah et al. (2011) for a review). Given that a sufficient amount of time series can be obtained, the method of event detection would need to be adopted. As the length of the river network increases, cases may occur where the run out of a flood wave of a previous event is overlapped by the onset of a new event. This can be solved by assigning a river specific hierarchical station identifier to each peak and explicitly accounting for the flood wave propagation in the pooling process of peaks and in the definition of the independence criterion.

Extending the search for documentation to other European countries requires a systematic approach as presented in chapter 3. In particular it requires an initial scan of the institutional and information landscape in order to identify the potential sources for obtaining the literature. We commend a careful mapping of these landscapes (both of the administrative landscape as well of the information landscape in form of libraries and their cataloguing systems) over time in order to understand any bias that may arise from the search and in order to ensure completeness. The effort may vary from country to country and in particular for the eastern countries we expect similar breaks in the number of available studies as we have found them for East Germany due to the political changes in 1990.

In order to ensure long-term applicability, extending and updating the knowledge base both for past flood events (closing the gap to historic hydrology) as well as for future events will be of high importance. Based on our results we expect that adding material to the database on events after 2002 or any future event will likely become easier and that the amount of material will be more abundant, including coverage also of low-impact floods. This will likewise account for other European countries.

The framework of quality assessment has been designed to be generic and can certainly be adapted for use in other countries as well as to other natural hazards. For flood risk it will be particularly interesting to apply the approach also in other European countries in order to gain an understanding and overview of the quality and depth of flood relevant material residing off scientific shelves and, if applicable, to compile a comprehensive inventory of information (that may include data) and knowledge available for floods in Central Europe.

5.3 Concluding outlook

National flood risk models will soon be available in greater numbers and for a diversity of countries. They will provide a range of possible ways, from rainfall-runoff model chains over spatial dependence models, of how to approach the question of the risk of widespread flooding. Yet, despite the technical and methodological improvements that are on the way, any risk assessment will require a profound knowledge of the underlying causes of risk. This knowledge is inevitably grounded on the experience and data gained on past flood events. New events will provide more and possibly better data; however, the extant data of the past cannot be increased in its volume today.

I think that a future risk assessment requires a more thorough and unprejudiced consideration of all sources of data and experience available on past flood events. Flood knowledge produced on non-academic desks is a valuable and abundant source too frequently ignored in scholarly argument. The use of effective search strategies and the careful assessment of the quality of data and information are cornerstones of daily scientific work. It should be easy to extend these competencies to sources off scientific sight; and in presenting the results of this work I add proof to the thought. It is true that flood event documentations alone will not provide answers to many questions in flood risk assessment, but in combination with observations and simulations they may provide the added value needed for improving flood risk models.

References

Auger, C. P.: Information sources in grey literature, Bowker Saur, London, 1998.

Bailin, A. and Grafstein, A.: The critical assessment of research: traditional and new methods of evaluation, Chandos, Oxford, 121 pp., 2010.

Bárdossy, A. and Filiz, F.: Identification of flood producing atmospheric circulation patterns, J. Hydrol., 313, 48–57, 2005.

Barnolas, M. and Llasat, M. C.: A flood geodatabase and its climatological applications: the case of Catalonia for the last century, Nat. Hazards Earth Syst. Sci., 7, 271–281, doi:10.5194/nhess-7-271-2007, 2007.

Barredo, J. I.: Major flood disasters in Europe: 1950–2005, Nat. Hazards, 42, 125–148, 2007.

Barredo, J. I.: Normalised flood losses in Europe: 1970–2006, Nat. Hazards Earth Syst. Sci., 9, 97–104, doi:10.5194/nhess-9-97-2009, 2009.

Belz, J. U., Braehmer, G., Buitefeld, H., Engel, H., Grabher, R., Hodel, H., Krahe, P., Lammersen, R., Larina, M., Mendel, H.-G., Meuser, A., Müller, G., Plonka, B., Pfister, L., and Van Vuuren, W.: Das Abflussregime des Rheins und seiner Nebenflüsse im 20. Jahrhundert – Analyse, Veränderungen, Trends, Schriftenreihe der KHR, Bd. 1–22, Koblenz and Lelystad, 2007.

Beurton, S. and Thieken, A. H.: Seasonality of floods in Germany, Hydrol. Sci. J., 54, 62–76, 2009.

BfG (Bundesanstalt für Gewässerkunde): Publikationen, Koblenz, available at: http://www.bafg.de (last access: 17 August 2012), Wissenstransfer/Publikationen, 2012.

Blong, R.: A Review of Damage Intensity Scales, Nat. Hazards, 29, 57–76, 2003.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R.: Introduction to Meta-Analysis, John Wiley & Sons, Ltd., Chichester, UK, 421 pp., 2009.

Borgman, C.: The Conundrum of Sharing Research Data, J. Am. Soc. Inf. Sci. Tec., 63, 1059–1078, doi:10.1002/asi.22634, 2011.

Brázdil, R., Kundzewicz, Z. W., and Benito, G.: Historical hydrology for studying flood risk in Europe / L'hydrologie historique pour une meilleure connaissance du risque inondation en Europe, Hydrol. Sci. J., 51, 739–764, 2006.

Bragues, G.: Wiki-philosophizing in a marketplace of ideas: Evaluating Wikipedia's entries on seven great minds, MediaTropes eJournal, 2, 117–158, 2009.

Burton, I.: Forensic Disaster Investigations in Depth: A New Case Study Model, Environment Magazine, 52, 36–41, 2010.

Caspary, H. J.: Recent winter floods in Germany caused by changes in the atmospheric circulation across Europe, Phys. Chem. Earth, 20, 459–462, 1995.

Caspary, H. J.: Increased risk of river flooding in southwest Germany caused by changes of the atmospheric circulation across Europe, in: PIK Report, European Conference on Advances in Flood Research, Potsdam, 288–299, 2000.

Centre for Evidence-Based Conservation: European Guide to good Practice in Knowledge Management - Part 1: Knowledge Management Framework, European Committee for Standardization, CWA 14924-1, Brussels, 2004.

Centre for Evidence-Based Conservation: Guidelines for Systematic Review in Environmental Management, Environmental Evidence, available at: www.environmentalevidence.org (last access: 27 September 2012), 2010.

Christensen, H. J., and Christensen, O. B.: Severe Summertime Flooding in Europe, Nature, 421, 805–806, 2002.

Cohen, J.: A coefficient of agreement for nominal scales, Educat. Psychol. Measure., 20, 37–46, doi:10.1177/001316446002000104, 1960.

Cooper, H. M., Hedges, L. V., and Valentine, J. C. (Eds.): The handbook of research synthesis and meta-analysis, 2nd Edn., Russell Sage Foundation, New York, 615 pp., 2009.

Cunderlik, J. M. and Burn, D. H.: The use of flood regime information in regional flood frequency analysis, Hydrol. Sci. J., 47, 77–92, 2001.

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J. L., and Blöschl, G.: Socio-hydrology: conceptualising human-flood interactions, Hydrol. Earth

Syst. Sci. Discuss., 10, 4515-4536, doi:10.5194/hessd-10-4515-2013, 2013a.

Di Baldassarre, G., Kooy, M., Kemerink, J. S., and Brandimarte, L.: Towards understanding the dynamic behaviour of floodplains as human-water systems, Hydrol. Earth Syst. Sci. Discuss., 10, 3869-3895, doi:10.5194/hessd-10-3869-2013, 2013b

Disse, M. and Engel, H.: Flood Events in the Rhine Basin: Genesis, Influences and Mitigation, Nat. Hazards, 23, 271–290, 2001.

Duckstein, L., Bárdossy, A., and Bogárdi, I.: Linkage between the occurrence of daily atmospheric circulation patterns and floods: an Arizona case study, J. Hydrol., 143, 413–428, 1993.

Environment Agency: The risk of widespread flooding - Capturing spatial patterns in flood risk from rivers and coasts: Methodology report, Environment Agency, Bristol, UK, 2011.

European Commission: Towards better access to scientific information: Boosting the benefits of public investments in research, European Commission, Brussels, available at: http://ec.europa.eu/research/science-society/documentlibrary/pdf06/era-communication-towards-better-access-to-scientific-informationen.pdf, 2012.

European Union: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks, Official Journal of the European Union, L 288/27, 2007.

Evans, J. and Reimer, J.: Open Access and Global Participation in Science, Science, 323, 1025–1025, doi:10.1126/science.1154562, 2009.

Falter, D., Vorogushyn, S., Lhomme, J., Apel, H., Gouldby, B., and Merz, B.: Hydraulic model evaluation for large-scale flood risk assessments, Hydrol. Process., 27, 1331-1340, 10.1002/hyp.9553, 2013.

Farace, D. and Schöpfel, J.: Introduction Grey Literature, in: Grey Literature in Library and Information Studies, edited by: Farace, D. and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 1–7, 2010.

Faulkner, D., Keef, C., and Martin, J.: Setting design inflows to hydrodynamic flood models using a dependence model, Hydrol. Res., 43, 663-674, 2012.

Fleiss, J. L. and Cohen, J.: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educat. Psychol. Measure., 33, 613–619, doi:10.1177/001316447303300309, 1973.

FLORIS: Flood risks and safety in the Netherlands (Floris). Floris study - Full report, Ministerie van Verkeer en Waterstaat, DWW-2006-014, 2005.

Frei, C., Davies, H. C., Gurtz, J., and Schär, C.: Climate dynamics and extreme precipitation and flood events in Central Europe, Integr. Assess., 1, 281–299, 2000.

Funtowicz, S. O. and Ravetz, J. R.: Uncertainty and Quality in Science for Policy, Kluwer, Dordrecht, 1990.

Genova, F.: The many faces of SIMBAD, in: 16th Annual Conference on Astronomical Data Analysis Software and Systems, 15–18 October 2006, edited by: Shaw, R. H. and Bell, F., DJ, Astronomical Data Analysis Software and Systems XVI, Astronomical Society of the Pacific Conference Series, Tucson, AZ, 145–152, 2007.

Ghizzoni, T., Roth, G., and Rudari, R.: Multisite flooding hazard assessment in the Upper Mississippi River, J. Hydrol., 412-413, 101-113, doi:10.1016/j.jhydrol.2011.06.004, 2012.

Glaser, R.: Klimageschichte Mitteleuropas, Primus Verlag, Darmstadt, 2001.

Glaser, R. and Stangl, H.: Historical floods in the Dutch Rhine Delta, Nat. Hazards Earth Syst. Sci., 3, 605–613, doi:10.5194/nhess-3-605-2003, 2003.

GO Consortium: The Geneontology, available at: http://www.geneontology.org (last access: 31 August 2012).

Gottschalk, L., Krasovskaia, I., Leblois, E., and Sauquet, E.: Mapping mean and variance of runoff in a river basin, Hydrol. Earth Syst. Sci., 10, 469–484, doi:10.5194/hess-10-469-2006, 2006.

Gräfe, H.: Ereignisanalyse – Hochwasser August 2002 in den Osterzgebirgsflüssen, Sächsisches Landesamt für Umwelt und Geologie, Dresden, 2004.

Grünewald, U.: Extreme hydro(meteoro-)logical events in the Elbe basin, Österreichische Wasser- und Abfallwirtschaft, 58, 27–34, 2006 (in German).

Haendel, M. A., Vasilevsky, N. A., and Wirz, J. A.: Dealing with Data: A Case Study on Information and Data Management Literacy, PLoS Biol., 10, e1001339, doi:10.1371/journal.pbio.1001339, 2012.

Haeussler, M., Gerner, M., and Bergman, C. M.: Annotating genes and genomes with DNA sequences extracted from biomedical articles, Bioinformatics, 27, 980–986, doi:10.1093/bioinformatics/btr043, 2011.

Hall, J. W., Sayers, P. B., and Dawson, R. J.: National-scale assessment of current and future flood risk in England and Wales, Nat. Hazards, 36, 147-164, 2005.

Hannah, D., Demuth, S., Looser, U., and Prudhomme, C.: Largescale river flow archives: importance, current status and future needs, Hydrol. Process., 25, 1191–1200, doi:10.1002/hyp.7794, 2011.

Heffernan, J. E. and Tawn, J. A.: A conditional approach for multivariate extreme values, J. Roy. Stat. Soc. B, 66, 497-546, 2004.

Higgins, J. P. T. and Green, S.: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (updated March 2011), The Cochrane Collaboration, available at: http://www.cochrane-handbook.org/ (last access: 21 June 2012), 2011.

Hirschboeck, K. K.: Flood hydroclimatology, in: Flood geomorphology, edited by: Baker, V. R., Kochel, R. C., and Patton, P. C., Wiley, New York, 27-49, 1988.

Hjørland, B.: Evaluation of an information source illustrated by a case study: Effect of screening for breast cancer, J. Am. Soc. Inf. Sci. Technol., 62, 1892–1898, doi:10.1002/asi.21606, 2011.

Hjørland, B.: Methods for evaluating information sources: An annotated catalogue, J. Inform. Sci., 38, 258–268, doi:10.1177/0165551512439178, 2012.

Houghton, J.: Costs and Benefits of Data Provision. Report to the Australian National Data Service, Centre for Strategic Economic Studies Victoria University, Melbourne, Australia, available at: http://ands.org.au/resource/houghton-cost-benefit-study.pdf (last access: 11 September 2012), 2011.

Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D. P., Kania, R., Schaeffer, M., St Pierre, S., Twigger, S., White, O., and Yon Rhee, S.: Big data: The future of biocuration, Nature, 455, 47–50, doi:10.1038/455047a, 2008.

Hübl, J., Kienholz, H., and Loipersberger, A. (Eds.): DOMODIS – Documentation of mountain disasters, State of discussion in the european mountain areas, Interpraevent, Schriftenreihe 1, Handbuch 1, Klagenfurt, 40 pp., 2002.

IKSR: Rheinatlas, Internationale Kommision zum Schutz des Rheins, Koblenz, 2001.

IRDR: Forensic Investigations of Disasters. The FORIN Project, Integrated Research on Disaster Risk (IRDR), Beijing, available at: http://www.irdrinternational.org/wp-content/uploads/2012/2006/FORIN-REPORT web.pdf (last access: 11 October 2012), 2011.

Jacobeit, J., Glaser, R., Luterbacher, J., and Wanner, H.: Links between flood events in central Europe since AD 1500 and largescale atmospheric circulation modes, Geophys. Res. Lett., 30, 1172, doi:10.1029/2002GL016433, 2003.

Jacobeit, J., Philipp, A., and Nonnenmacher, M.: Atmospheric circulation dynamics linked with prominent discharge events in Central Europe, Hydrol. Sc. J., 51, 946–965, 2006.

Keef, C., Svensson, C., and Tawn, J. A.: Spatial dependence in extreme river flows and precipitation for Great Britain, J. Hydrol., 378, 240–252, 2009a.

Keef, C., Tawn, J. A., and Svensson, C.: Spatial risk assessment for extreme river flows, J. Roy. Stat. Soc. C-App., 58, 601-618, 2009b.

Keef, C., Tawn, J. A., and Lamb, R.: Estimating the probability of widespread flood events, Environmetrics, 24, 13-21, doi:10.1002/env.2190, 2013.

Kraemer, H. C., Periyakoil, V. S., and Noda, A.: Kappa coefficients in medical research, Stat. Medicine, 21, 2109–2129, doi:10.1002/sim.1180, 2002.

Kron, W.: Flood risk = Hazard * values * vulnerability, Water Int., 30, 58–68, 2005.

Lamb, R., Keef, C., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C.: A new method to assess the risk of local and widespread flooding on rivers and coasts, J. Flood Risk Manage., 3, 323-336, doi:10.1111/j.1753-318X.2010.01081.x, 2010.

Llasat, M.-C., Barriendos, M., Barrera, A., and Rigo, T.: Floods in Catalonia (NE Spain) since the 14th century. Climatological and meteorological aspects from historical documentary sources and old instrumental records, J. Hydrol., 313, 32–47, 2005.

Lowry, R.: Kappa as a measure of concordance in categorical sorting, available at: http://www.vassarstats.net/kappa.html (last access: 5 November 2012), 2012.

Luzi, D.: Grey documents in open archives, in: Grey Literature in Library and Information Studies edited by: Farace, D. and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 2010.

MacDonald, B. H., Wells, P. G., Cordes, R., Hutton, G. R. G., Cossarini, D. M., and Soomai, S.: The use and influence of information produced as Grey Literature by international, intergovernmental marine organizations: Overview and current research, in: Grey Literature in Library and Information Studies, edited by: Farace, D. and Schöpfel, J., De Gruyter/Saur, Berlin/New York, 167–180, 2010.

Mackenzie Owen, J. S.: The expanding horizon of Grey Literature, 3rd International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information, 13–14 November 1997, Luxemburg, 1997.

Madnick, S. E., Wang, R. Y., Lee, Y. W., and Zhu, H.: Overview and framework for data and information quality research, ACM J. Data Inf. Quality, 1, 1–22, doi:10.1145/1515693.1516680, 2009.

Marx, V.: My data are your data, Nat. Biotechnol., 30, 509–511, doi:10.1038/nbt.2243, 2012.

Mendes, B. V. D. and Pericchi, L. R.: Assessing conditional extremal risk of flooding in Puerto Rico, Stoch. Env. Res. Risk A., 23, 399-410, doi:310.1007/s00477-00008-00220-z, 2009.

Merz, B. and Thieken, A. H.: Flood risk analysis: Concepts and challenges, Österreichische Wasser- und Abwasserwirtschaft, 56, 27–34, 2004.

Merz, B.: Hochwasserrisiken - Grenzen und Möglichkeiten der Risikoabschätzung, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 2006.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. H.: Review article 'Assessment of economic flood damage', Nat. Hazards Earth Syst. Sci., 10, 1697-1724, 2010.

Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Hundecha, Y.: HESS Opinions “More efforts and scientific rigour are needed to attribute trends in flood time series”, Hydrol. Earth Syst. Sci., 16, 1379–1387, doi:10.5194/hess-16-1379-2012, 2012.

Merz, R. and Blöschl, G.: A process typology of regional floods, Water Resour. Res., 39, 1340, doi:10.1029/2002WR001952, 2003.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, Water Resour. Res., 44, W08432, doi:10.1029/2007WR006744, 2008a.

Merz, R. and Blöschl, G.: Flood frequency hydrology: 2. Combining data evidence, Water Resour. Res., 44, W08433, doi:10.1029/2007WR006745, 2008b.

Milly, P., Wetherald, R., Dunne, K., and Delworth, T.: Increasing risk of great floods in a changing climate, Nature, 415, 514–517, 2002.

Mitchell, J. K.: European River Floods in a Changing World, Risk Anal., 23, 567–574, 2003.

Mudelsee, M. M. B., Tetzlaff, G., and Grünewald, U.: Extreme floods in central Europe over the past 500 years: Role of cy clone pathway “Zugstrasse Vb”, J. Geophys. Res., 109, D23101, doi:10.1029/2004JD005034, 2004.

Mudelsee, M., Deutsch, M., Borngen, M., and Tetzlaff, G.: Trends in flood risk of the River Werra (Germany) over the past 500 years, Hydrolog. Sci. J., 51, 818–833, 2006.

Müller, M., Kašpar, M., and Matschullat, J.: Heavy rains and extreme rainfall-runoff events in Central Europe from 1951 to 2002, Nat. Hazards Earth Syst. Sci., 9, 441–450, doi:10.5194/nhess-9-441-2009, 2009.

Navratil, O., Albert, M.-B., Hérouin, E., and Gresillon, J.-M.: Determination of bankfull discharge magnitude and frequency: comparison of methods on 16 gravel-bed river reaches, Earth Surf. Proc. Land., 31, 1345–1363, 2006.

Neal, J., Keef, C., Bates, P., Beven, K., and Leedal, D.: Probabilistic flood risk mapping including spatial dependence, Hydrol. Process., published online, doi:10.1002/hyp.9572, 2012.

Nied, M., Hundecha, Y., and Merz, B.: Flood-initiating catchment conditions: a spatio-temporal analysis of large-scale soil moisture patterns in the Elbe river basin, Hydrol. Earth Syst. Sci., 17, 1401-1414, doi:10.5194/hess-17-1401-2013, 2013.

Norris, R., Nichols, S. J., Ransom, G., Webb, A., Stewardson, M., Liston, P., and Mugodo, J.: Causal criteria analysis. Methods manual: a systematic approach to evaluate causality in environmental science, eWater Cooperative Research Centre, Canberra, 2008.

OBO Foundry: The Open Biological and Biomedical Ontologies, available at: http://www.obofoundry.org/ (last access: 31 August 2012).

Osenberg, C. W., Sarnelle, O., and Goldberg, D. E.: Meta-analysis in ecology: Concepts, statistics, and applications, Ecology, 80, 1103–1104, 1999.

Pauling, A. and Paeth, H.: On the variability of return periods of European winter precipitation extremes over the last three centuries, Clim. Past, 3, 65–76, doi:10.5194/cp-3-65-2007, 2007.

Petit, F. and Pauquet, A.: Bankfull discharge recurrence interval in gravel-bed rivers, Earth Surf. Proc. Land., 22, 685–693, 1997.

Petrow, T., Thieken, A. H., Kreibich, H., Merz, B., and Bahlburg, C. H.: Improvements on flood alleviation in Germany: Lessons learned from the Elbe flood in August 2002, Environ. Manage., 38, 717–732, 2006.

Petrow, T., Merz, B., Lindenschmidt, K.-E., and Thieken, A. H.: Aspects of seasonality and flood generating circulation patterns in a mountainous catchment in south-eastern Germany, Hydrol. Earth Syst. Sci., 11, 1455–1468, doi:10.5194/hess-11-1455-2007, 2007.

Petrow, T. and Merz, B.: Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, J. Hydrol., 371, 129–141, 2009.

Petrow, T., Zimmer, J., and Merz, B.: Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns, Nat. Hazards Earth Syst. Sci., 9, 1409–1423, doi:10.5194/nhess-9-1409-2009, 2009.

Philipp, A. and Jacobeit, J.: Das Hochwasserereignis in Mitteleuropa im August 2002 aus klimatologischer Perspektive, Petermann. Geogr. Mitt., 147, 50–53, 2003.

Pohl, R.: Historische Hochwasser aus dem Erzgebirge: von der Gottleuba bis zur Mulde, Dresdner Wasserbauliche Mitteilungen, Heft 28, Technische Universität Dresden, Inst. für Wasserbau und Techn. Hydromechanik, Dresden, 2004.

Prudhomme, C. and Genevier, M.: Can atmospheric circulation be linked to flooding in Europe?, Hydrol. Process., in press, doi:10.1002/hyp.7879, 2010.

Ranger, S. L.: Grey Literature in Special Libraries: Access and Use, in: GL-Conference series, Sixth International Conference on Grey Literature: Work on Grey in Progress, New York Academy of Medicine (USA), 182 pp., 2004.

Reiche A.: Setting the new standard in modelling flood in Germany and across Europe. AIRCurrents, Feb. 2012, available at: www.airworldwide.com, last accessed: 22.4.2013, 2012.

Renear, A. H. and Palmer, C. L.: Strategic Reading, Ontologies, and the Future of Scientific Publishing, Science, 325, 828–832, doi:10.1126/science.1157784, 2009.

RMS (Risk Management Solutions): Germany river flood. https://support.rms.com/Publications/Germany_Flood.pdf, last accessed: 22.4.2013, 2006.

Rodda, H. J.: The Development and Application of a Flood Risk Model for the Czech Republic, Nat. Hazards, 36, 207–220, 2005.

Rothstein, H. R. and Hopewell, S.: Grey Literature, in: The handbook of research synthesis and meta-analysis, edited by: Cooper, H. M., Hedges, L. V., and Valentine, J. C., Russell Sage Foundation, New York, 103–125, 2009.

Samuels, P. and Gouldby, B.: Language of Risk – Project Definitions (Second Edition), FLOODsite Project, Document Number: T32-04-01, 2009.

Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., de Daruvar, A., de Matos, P., Dix, I., Edmunds, S., Evelo, C. T., Forster, M. J., Gaudet, P., Gilbert, J., Goble, C., Griffin, J. L., Jacob, D., Kleinjans, J., Harland, L., Haug, K., Hermjakob, H., Sui, S. J. H., Laederach, A., Liang, S., Marshall, S., McGrath, A., Merrill, E., Reilly, D., Roux, M., Shamu, C. E., Shang, C. A., Steinbeck, C., Trefethen, A., Williams-Jones, B., Wolstencroft, K., Xenarios, I., and Hide, W.: Toward interoperable bioscience data, Nat. Genet., 44, 121–126, 2012.

Seglen, P. O.: Why the impact factor of journals should not be used for evaluating research, British Medical J., 314, 498–502, 1997.

Shorthouse, C. A. and Arnell, N. W.: Spatial and temporal variability in European river flows and the North Atlantic oscillation in: IAHS Publ., FRIEND'97 – Regional hydrology: Concepts and models for sustainable water resource management, Postojna, Slovenia, 1997.

Simons, K.: The Misused Impact Factor, Science, 322, 165, doi:10.1126/science.1165316, 2008.

Smith, M.: Communicating with data: new roles for scientists, publishers and librarians, Learn. Publ., 24, 203–205, doi:10.1087/20110308, 2011.

Stanescu, V. A.: Outstanding floods in Europe. A regionalisation and comparison, International Conference on Flood Estimation, 6–8 March 2002, Berne, Switzerland, 697–706, 2002.

Strahler, A. N.: Quantitative geomorphology of drainage basins and channel networks, in: Handbook of Applied Hydrology, edited by: Chow, V. T., MacGraw Hill, New York, 39–76, 1964.

Strong, D. M., Lee, Y. W., and Wang, R. Y.: Data quality in context, Commun. ACM, 40, 103–110, doi:10.1145/253769.253804, 1997.

Sturm, K., Glaser, R., Jacobeit, J., Deutsch, M., Brázdil, R., Pfister, C., Luterbacher, J., and Wanner, H.: Hochwasser in Mitteleuropa seit 1500 und ihre Beziehungen zur atmosphärischen Zirkulation, Petermann. Geogr. Mitt., 145, 14–23, 2001.

Svensson, C., Kundzewicz, Z., and Maurer, T.: Trend detection in river flow series: 2. Flood and low-flow index series, Hydrolog. Sci. J., 50, 811–824, 2005.

The Scientist: Whither Science Publishing?, The Scientist, 1 August 2012, available at: http://the-scientist.com/2012/2008/2001/whither-science-publishing/ (last access: 10 September 2012), 2012.

Thieken, A. H., Kreibich, H., Müller, M., and Merz, B.: Coping with floods: preparedness, response and recovery of flood-affected residents in Germany in 2002, Hydrolog. Sci. J., 52, 1016–1037, 2007.

Thieken, A. H., Seifert, I., Elmer, F., Maiwald, H., Haubrock, S., Schwarz, J., Müller, M., and Seifert, J.: Standardisierte Erfassung und Bewertung von Hochwasserschäden, Hydrol. Wasserbewirts., 53, 198–207, 2009.

Uhlemann, S., Zimmer, J., Thieken, A. H., and Merz, B.: Understanding trans-basin floods, Geophys. Res. Abstr., Vol. 11, EGU2009-8906, 2009.

Uhlemann, S., Thieken, A. H., and Merz, B.: A consistent set of trans-basin floods in Germany between 1952–2002, Hydrol. Earth Syst. Sci., 14, 1277–1295, doi:10.5194/hess-14-1277-2010, 2010a.

Uhlemann, S., Zimmer, J., Hundecha, Y., and Merz, B.: Quantifying the hydrometeorological origins and space-time characteristics of trans-basin floods, Geophys. Res. Abstr., Vol. 12, EGU2010-13351, 2010b.

Uhlemann, S.: Supplement to: Data Expansion: the potential of grey literature for understanding floods. Deutsches GeoForschungsZentrum GFZ, doi:10.5880/GFZ.5.4.2012.001, 2012.

Uhlemann, S., Bertelmann, R., and Merz, B.: Data expansion: the potential of grey literature for understanding floods, Hydrol. Earth Syst. Sci., 17, 895-911, doi:10.5194/hess-17-895-2013, 2013a.

Uhlemann, S., Thieken, A. H., and Merz, B.: A quality assessment framework for natural hazard event documentations: application to trans-basin flood reports in Germany, Nat. Hazards Earth Syst. Sci. Discuss., 1, 143–192, doi:10.5194/nhessd-1-143-2013, 2013b.

Uhlemann, S.: Supplement to: A quality assessment framework for natural hazard event documentations: Application to trans-basin flood reports in Germany, Deutsches GeoForschungsZentrum GFZ, doi:10.5880/GFZ.5.4.2013.001, 2013c.

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 1 – Rainfall periods and flood development, Weather, 58, 371–377, 2003a.

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., and Pinto, J. G.: The central European floods of August 2002: Part 2 – Synoptic causes and considerations with respect to climate change, Weather, 58, 434–442, 2003b.

Van der Sluijs, J. P., Risbey, J. S., and Ravetz, J.: Uncertainty Assessment of Voc Emissions from Paint in the Netherlands Using the Nusap System, Environ. Monitor. Assess., 105, 229–259, 2001.

Van der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., and Ravetz, J.: Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: The NUSAP system, Risk Anal., 25, 481–492, 2005.

Van Dyck, J. and Willems, P.: Probabilistic flood risk assessment over large geographical regions, Water Resour. Res., accepted article, doi:10.1002/wrcr.20149, 2013.

Van Noorden, R.: Trouble at the text mine, Nature, 483, 134–135, doi:10.1038/483134a, 2012.

Viglione, A., Borga, M., Balabanis, P., and Blöschl, G.: Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy, J. Hydrol., 394, 63–77, 2010a.

Viglione, A., Chirico, G. B., Woods, R., and Blöschl, G.: Generalised synthesis of space-time variability in flood response: An analytical framework, J. Hydrol., 394, 198-212, doi:10.1016/j.jhydrol.2010.05.047, 2010b.

Vogt, J., Soille, P., de Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M., Haarstrup, P., and Bamps, C.: A pan-European river and catchment database, European Commission, EUR 22920 EN – Joint Research Centre, 2007.

Vorogushyn, S., Merz, B., Lindenschmidt, K.-E., and Apel, H.: A new methodology for flood hazard assessment considering dike breaches, Water Resour. Res., 46, W08541, doi:10.1029/2009WR008475, 2010.

Wang, R. Y. and Strong, D. M.: Beyond accuracy: what data quality means to data consumers, J. Manage. Inf. Syst., 12, 5–33, 1996.

Warnick, W.: Tailoring access to the source: preprints, grey literature and journal articles, Nature Webdebates, available at: http://www.nature.com/nature/debates/e-access/Articles/warnick.html (last access: 12 July 2012), 2001.

Weed, D. L.: Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related, Int. J. Epidemiol., 29, 387–390, doi:10.1093/ije/29.3.387, 2000.

Weintraub, I.: The role of grey literature in the sciences, available at: http://library.brooklyn.cuny.edu/access/greyliter.htm (last access: 9 November 2010), 2000.

White, G. F., Kates, R. W., and Burton, I.: Knowing better and losing even more: the use of knowledge in hazard management, Environmental Hazards, 3, 81–92, 2001.

Whitlock, M., McPeek, M., Rausher, M., Rieseberg, L., and Moore, A.: Data Archiving, Am. Nat., 175, 145–146, doi:10.1086/650340, 2010.

Wilkerson, G. V.: Improved bankfull discharge prediction using 2-year recurrence period discharge, J. Am. Water Resour. As., 44, 243–258, 2008.

Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, available at: http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf, Berlin, 90 pp., 2012.

Wood, J., Andersson, T., Bachem, A., Best, C., Genova, F., Lopez, D. R., Los, W., Marinucci, M., Romary, L., Van de Sompel, H., Vigen, J., and Wittenburg, P.: Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data – A submission to the European Commission, European Union, available at: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf (last access: 11 September 2012), 2010.

Author’s declaration

I prepared this dissertation without illegal assistance. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. This dissertation has not been presented to any other University for examination, neither in Germany nor in another country.

Steffi Uhlemann

Potsdam, April 2013