



HAL
open science

Forecasting of flash-flood human impacts integrating the social vulnerability dynamics

Galateia Terti

► **To cite this version:**

Galateia Terti. Forecasting of flash-flood human impacts integrating the social vulnerability dynamics. Hydrology. Université Grenoble Alpes, 2017. English. NNT : 2017GREAU004 . tel-01639161

HAL Id: tel-01639161

<https://theses.hal.science/tel-01639161>

Submitted on 20 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Géographie**

Arrêté ministériel : 25 Mai 2016

Présentée par

Galatea TERTI

Thèse dirigée par **Sandrine ANQUETIN**

et codirigée par **Isabelle RUIN** et **Jonathan GOURLEY**

préparée au sein de l' **Institut des Géosciences de l'Environnement (IGE, UMR5001-UR252, UGA - CNRS - IRD - Grenoble-INP)**
dans l'Ecole Doctorale **Terre Univers Environnement**

Forecasting of flash-flood human impacts integrating the social vulnerability dynamics

Thèse soutenue publiquement le **27 Mars 2017**,
devant le jury composé de :

Mme, Julie, DUGDALE

Maître de conférences, UGA, LIG, Grenoble, Présidente

Mme, Léna, SANDERS

Directrice de Recherche CNRS, Géographie-Cités, Paris, Rapporteuse

M, Murugesu, SIVAPALAN

Professor of Geography and GIScience, University of Illinois at Urbana-Champaign, USA, Rapporteur

M, Eric, GAUME

Directeur Département Géotechnique, Environnement, Risques naturels et Sciences de la Terre, IFSTTAR, Nantes, Examineur

M, Richard, LAGANIER

Professeur Université Paris Diderot, PRODIG, Paris, Examineur

Mme, Sandrine, ANQUETIN

Directrice de Recherche CNRS, IGE, Grenoble, Directrice de thèse

Mme, Isabelle, RUIN

Chargée de Recherche CNRS, IGE, Grenoble, Co-Directrice de thèse

M, Jonathan, GOURLEY

Research Hydrometeorologist, NOAA/National Severe Storms Laboratory, Norman, Oklahoma, USA, Co-Directeur de thèse



Acknowledgements

I would like to offer a sincere Thank You to my advisors Dr. S. Anquetin, Dr. I. Ruin and Dr. J. Gourley, for believing in me and for giving to me the opportunity to work with them in such a multifaceted and multidisciplinary topic. Among other attributes, Sandrine's organizational skills, Isabelle's generousness, and JJ's leading profile compose a great supervising scheme. Each of you - in your own special way - provided to me important advice and help on both science and life. I would like to acknowledge all of the scientific discussions but also the family-warmth moments we shared. Thanks to you I was very much welcomed in both France and the United States, and I gained a great experience in collaborating in an international environment. Thank you for all the opportunities for research advancement, international interactions and traveling. I appreciate your patience and listening to my concerns and perfectionism. I will be always grateful for the support (and funding) as well as the respect and trust you showed to me during the last three years.

I would like to express my appreciation to my committee (Dr. L. Sanders, Dr. S. Murugesu, Dr. J. Dugdale, Dr. E. Gaume and Dr. R. Laganier) for the insightful comments and the valuable discussion during my thesis defense. It was a big pleasure to exchange opinions and research experiences with such high-quality scientists. I hope we will meet scientifically again in the future.

I would like to acknowledge also the following people who provided expertise and information that were crucial for the completion of this dissertation: Ami Arthur (NOAA/National Severe Storms Laboratory, and Cooperative Institute for Mesoscale Meteorological Studies, U.S.), who was always willing to help me in the gathering and processing of GIS data; Pierre Kirstetter (NOAA/National Severe Storms Laboratory, and Atmospheric Radar Research Center, U.S.), who was very open to discuss my topic and share his experience in predictive modeling; Zac Flamig (NOAA/National Severe Storms Laboratory, and Atmospheric Radar Research Center, U.S.), who was an ideal person for brainstorming and productive criticism; and Juliette Blanchet (Univ. Grenoble Alpes, CNRS, IGE, France), who replied to my statistical concerns and struggles in a direct and effective way. Ami,

Pierre, Zac and Juliette, you are very precious to me because you were there to reinforce me in times of doubt, enhancing my confidence in my research.

At this point, I would like to express my gratitude to the people back in Greece who believed in me and supported my decision to study abroad. Mostly, I thank Dr. Galiatsatou, researcher in the Aristotle University of Thessaloniki (Greece), who motivated me to pursue an international research program and expected nothing less than excellence from me. Panagiota Galiatsatou is an excellent example for other researchers to follow.

Since the Ph.D. degree is not only an academic exercise, but also a personal journey in life, my biggest Thank You goes to my family (“prior” and “new”) and my best friends back in Greece. Many thanks to my father for supporting me to follow my dreams and find my own way; my brothers for being great supporters from the first day of my life to the present, sharing motivating calls with me wherever I am; and especially, my mother for being next to me day and night with love, patience and encouragement. Mum, without you I would never be the person I am and I would never know the value of setting high goals and the beauty of trying to meet them.

Special thanks to my husband for following my “craziness” and deciding to start a new life with me far from our home, relatives and friends. My dear, your support was my driving force to leave from Greece and do something that I love with all my heart. I am very fortunate to have such a generous person in my life. Every day you remind me that happiness is in small things. You are there to bring rationalism back when I am frustrated and celebrate with me every little thing that I achieve. I am truly thankful that you are always by my side with such a positive and optimistic attitude to life, and I promise to do the same for you.

My best friend Argiro, I recognize your support from the first day in the polytechnic school to the present. I am grateful for your friendship that is beyond distance. You taught me that true friends are never apart despite the miles that may lie between them.

To continue, I would like to say a big Thank You to the people who came to my life during the last three/four years and will be forever in my heart. These include people from the LTHE and 3SR laboratories, the Greek community in Grenoble, the U.S. NWC and the FLASH project and all the others who shared with me moments of real friendship.

Many thanks to Eleni Kolokitha who welcomed me in Grenoble and was my only family during my first six months in France. My sweet Elenitsa, thank you for sharing with me your company and your prior experience with the HYDROHASARDS Master. I appreciate so much your advice and support from the first to the last day of this journey. Thanks to my colleagues and friends Saif, Melody, Damien, Aurelien, Audre and Jeremy for integrating me in the French team in IGE (previously LTHE), speaking with me in English and translating for me conversations, jokes and gossips. Saif, Damien and Aurelien it was a great pleasure to be with you on the “big days” of our defense!

I am also thankful for the colleagues and friends that I met in the U.S. Thanks to Elizabeth, Jess, Zac and Race, who welcomed me in my arrival in Oklahoma and helped me to adapt in the U.S. Jess and Zac, thank you for sharing with me all those Oklahoman experiences such as theatre, Christmas lights seeing, hiking, storm chasing, shooting. I am so thankful for the beautiful days we have spent together at work or on our travels. I do appreciate very much all the people who drove me around

in Norman, helped me with the various in-processing steps and supported me during my stay in Oklahoma like Elizabeth, Jess, Pierre and his wife Gyslaine, Ami, Keli, Manab, Humberto, Brandon, Eric and of course my advisor JJ and his wife Stephanie. I feel deep appreciation for each of you who opened your home to me and included me in family gatherings and special moments such as Thanksgiving, Christmas and New Years. I owe my warmest thanks to my first teacher in English and also family friend Koula Stefou who welcomed me in her home in New York and offered to me a mother's hug during my stay in the U.S.

Last but not least I would like to give my big thanks to the girls who lived with me in the Traditions East Square Apartment in Norman and became my family during my stay in the U.S. On the top of all, thank you Pauline Zaunet (from France), Marine Giroud (form France), and Paulina Cwik (from Poland) for being my European company in Oklahoma, supporting me with love, laughs, and motivating hugs. Pauline, thanks for filling me with joy with your delicious French burgers; Marine, thank you for making me feel like home with your amazing desserts; Paulina, thanks for reminding to me the shine of Sunday with your yummy pancakes and of course, your Polish vodka! It was always a relief to know that I will cook, eat or talk with you when coming back from work. My girls, I appreciate every single moment of our "GalateiOUS" "TertiOUS" philosophical discussions and I will never forget our motto: LIVE, LOVE, LAUGH.

**“Quand tu prendras le chemin d’Ithaque,
souhaite que la route soit longue,
pleine d’aventures, pleine d’enseignements.**

*Les Lestrygons et les Cyclopes,
ne les crains pas, ni la colère de Poséidon,
jamais tu ne trouveras rien de tel sur ton chemin,
si ta pensée reste élevée, si une émotion rare
étreint ton esprit et ton corps.*

*Les Lestrygons et les Cyclopes,
tu ne les rencontreras pas, ni l’irascible Poséidon,
si tu ne les transportes pas dans ton âme,
si ton âme ne les fait surgir devant toi.*

Souhaite que la route soit longue.
*Que nombreux soient les matins d’été
où - avec quel plaisir et quelle joie,
tu découvriras des ports que tu n’as jamais vus;
arrête-toi dans les comptoirs phéniciens
pour te procurer de précieuses marchandises,
ambre, corail, ébène, nacre,
et capiteux parfums de toutes sortes,
le plus que tu pourras de capiteux parfums;
visite aussi beaucoup de villes égyptiennes,
et n’aie de cesse de t’instruire auprès
de ceux qui savent.*

Garde toujours Ithaque présente à ton esprit.

*Y parvenir est ta destination finale.
Mais ne te hâte surtout pas dans ton voyage.
Mieux vaut le prolonger pendant des années;
et n’aborder dans l’île que dans ta vieillesse,
riche de ce que tu auras gagné en chemin,
sans attendre d’Ithaque aucun autre bienfait.*

Ithaque t’a offert ce beau voyage.

Sans elle, tu n’aurais pas pris la route.

Elle n’a rien de plus à t’apporter.

Et même si elle est pauvre, Ithaque ne t’a pas trompé.

**Sage comme tu l’es, avec une expérience pareille,
Tu as sûrement déjà compris ce que les Ithagues signifient”.**

C.P. Cavafy, 1911

(Traduction de Dominique Grandmont, En attendant les barbares et autres poèmes, Gallimard.)

***“As you set out for Ithaka
hope the voyage is a long one,
full of adventure, full of discovery.***

*Laistrygonians and Cyclops,
angry Poseidon; don't be afraid of them:
you'll never find things like that on your way
as long as you keep your thoughts raised high,
as long as a rare excitement
stirs your spirit and your body.*

*Laistrygonians and Cyclops,
wild Poseidon; you won't encounter them
unless you bring them along inside your soul,
unless your soul sets them up in front of you.*

Hope the voyage is a long one.

*May there be many a summer morning when,
with what pleasure, what joy,
you come into harbors seen for the first time;*

*may you stop at Phoenician trading stations
to buy fine things,
mother of pearl and coral, amber and ebony,
sensual perfume of every kind,
as many sensual perfumes as you can;
and may you visit many Egyptian cities
to gather stores of knowledge from their scholars.*

Keep Ithaka always in your mind.

Arriving there is what you are destined for.

But do not hurry the journey at all.

*Better if it lasts for years,
so you are old by the time you reach the island,
wealthy with all you have gained on the way,
not expecting Ithaka to make you rich.*

Ithaka gave you the marvelous journey.

Without her you would not have set out.

She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.

***Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean”.***

C.P. Cavafy, Collected Poems, 1911

(Translated by Edmund Keeley and Philip Sherrard. Edited by George Savidis. Revised Edition. Princeton University Press, 1992)

Preface

Every PhD student is dreaming of a robust model or theory, or maybe a well-defined experiment to work with, and so, ending up with nice conclusions and good results for a successful defense. It took three years for me to realize that fighting with unknown theories or concepts that may not apply to the point of your research, and even not arriving to fancy models or applications, is what brings you from a student to a researcher. And in my opinion, this is the most fascinating part of the process. Because this is what leads you to personalize the project that, initially, was your advisor(s)' idea. Nobody says it is an easy process; it can be a long or even a tough road that nobody else can travel for you. But as Cavafy explains in his poem "Ithaka", there are many treasures to meet through the way, and at the end there is always a new part of yourself waiting for you to discover it. Though an academic exercise, being a PhD student is definitely a way of life. Looking back to whom I was when I started this PhD program and who I became three years later makes me realize how many treasures, but also hindrances, I have met the last years. It was back in 2012 when I firstly left Greece to follow the Greek-French Master program "HYDROHASARS". At that moment, I had no idea how much my life would change in the following years; how many new "homes" were waiting for me out there. So, the "target", named as "Ithaka" by Cavafy, is what brings us into an amazing journey that has a lot to teach to us. Ithaka reflects the goals that we set, and as so, it is the forcing power that keeps us motivated and focused but also open to anything new. In the following lines, let me do a brief flashback and make a stop with you to some of the most challenging but also beautiful places of the three-year travel to **my personal Ithaka: My PhD.**

“Multi-PhD”: Multi-cultural, Multi-disciplinary, Multi-supervised

“Multi-cultural”

From “*Bonjour ça va*” to “*Hi, how’s it going*”, and feeling that as “*καλημερα, τι κανεις*”, this project has been a life-changing experience for me. Traveling and working for the half of my PhD in France and the other half in the U.S. taught me that people across different countries and continents may be very similar in their needs and attitudes. These may be then expressed through different traditions and habits depending on geographic and cultural specificities. The most usual question I get about this experience is what I did find different between Greece, France and the U.S.. Well, the first thing that comes to my mind is all the fun interactions you get with the supervisors abroad. Restricted to formal relationships, in Greece there is a significant distance between advisors and students. In Greece, formal addresses like “Mr X” are supposed to present particular respect to someone. In this context, students are expected to call the professors with their last names and talk in plural. This probably sounds to you as a small detail arising mainly from differences in languages. However, it is based on a deep mindset about academic, and by inference, social hierarchy supported also by linguistic and behavioral rules. Thus, I was positively surprised to be invited to the place of my advisors from the beginning of my interaction with them. Sharing family and friendly moments with advisors not only does not reduce the respect you have for them but instead, it increases the appreciation you get for the complete personality you get to know: a real person, that can combine research, family, career, professional and personal dreams in one life. Like the moment that you meet your advisor in his/her pajamas; your Greek conservatism is crushed leaving place for true and honest communications.

At the beginning of my arrival in France, I caught myself looking for small or big differences in every aspect: food, lunch time, sleeping hours, leisure, mentality. Generally, people have the tendency to focus on differences when they enter a new environment. That is because we understand the world through comparisons. As my advisors say, “It all depends on the scale you are looking at.” You may think there is a big difference between the Greek and French cuisine when, for example, you compare an “overcooked” Greek steak seasoned and sprinkled with salt and pepper with a plain, almost “raw” French steak or fillet. However, French and Greek food fall in the same category called as “European” from people outside Europe, when comparing it with the American diet. Although clichés are here to defeat me, I strongly believe that we cannot or at least we should not generalize across countries. Did you know that French people do not necessarily need olive oil or generally grease to “fry” their egg? I have to admit that as far as the frying of an egg, Greece may be much closer to China than France, and this fact is confirmed by my Chinese and French flatmates, respectively. Very soon I realized that even if there is always probability to find differences in habits from county to country, it is also highly probable to find out important similarities. It is that moment when you join people in a family gathering and you become a witness of all this laughter and love coming out from similar jokes, stories and quips among relatives and friends around the table. During these three years I had the luck to experience with others important days like Christmas, Thanksgiving, 4th of July. Those moments were touching my heart since on the face of people that I was meeting for the first time,

and I had no common biomes or even common language, I could see my family; the smile of my husband, the jokes of my father or brothers, the warmth of my mum. . . .

Traveling around, meeting or living with new people and getting new experiences in work, and life in general, is a wonderful process. It teaches you not only to accept but also to appreciate the difference! Sure, the “difference”, in the sense of something new, may give you a hard time until you get familiar with that. One could say that the most challenging part is when you arrive at a new place and you are called to build a new life. Far from what you were used to having in your daily life - your friends or family - you have to create a new routine. This routine has to be adapted in the culture of the new place, but still reflect your habits and personality. When this new place is not related to leisure travels but instead it is linked with a new working environment, you may need to speed the process of adaptation in order to come back to a regular working rhythm as soon as possible. What a conflict! Though, once you get into the new reality, a whole world full of new friends and colleagues is opening to you; absolutely a new family. Getting to know new people and that each of them has his own values, beliefs, and traditions to share may fascinate you! Every single person that I met either in France or the U.S. during my “multi-cultural” PhD gave me a new lesson for my life. The friendships acquired made me realize every day something that people often say but they rarely feel: that “life is beautiful”. Thus, if you ask me now, I strongly believe that adapting to new things is not the biggest difficulty you will have to deal with if you move in different places during your PhD. Leaving a place can be even harder than arriving to a new one. As Miriam Adeney wrote: *“You will never be completely at home again, because part of your heart always will be elsewhere. That is the price you pay for the richness of loving and knowing people in more than one place”*.

“Multi-disciplinary”

Studying risk to natural hazards has been always interesting to me. It is that power of nature that impresses me in two forms: Firstly, as physical phenomena, complex in their nature and inherently unpredictable and secondly, as “enemies” of humans and their environment, challenging to face. As civil engineers in the polytechnic school, they made us to believe that natural hazards can be mitigated through structural measures. Then, the subject to investigate is how to optimize the resistance of infrastructure to such severe or extreme natural occurrences. During my Master studies in France I met, for the first time, the idea that the effects of some kinds of hazards, like flash flooding, cannot be reduced by solely undertaking technical measures. Such phenomena may be so sudden and severe, and their dynamics are in interplay with people. In fact this interplay is what creates the disaster when human life and livelihoods are threatened. True story: That idea about individuals and their socio-economic constraints being a key factor for a hazard to become a disaster, which was delivered in the 1970s from political ecology, came to my ears, for the first time, in 2012. Since then, every day I am penetrating deeper and deeper in the “vulnerability and risk to natural hazards” world, fascinated by the multifaceted and multidisciplinary nature of the problem.

Within multidisciplinary (also called interdisciplinary) topics lurk some difficulties. In my opinion, the biggest troubles arise when a word used to explain a process or phenomenon in one discipline may mean something very different in another. Even worse, a term may be unknown for people with a certain specialty, and as so, it tends to be ignored when studying a complex multidisciplinary

issue. For example, have you ever thought that “vulnerability” as a word may have not so much to tell to meteorologists? Instead the term “impact” makes much more sense for them when referring to potentiality for losses from weather phenomena such as storms. It didn’t take a long time to realize such difficulties in communication. The University of Grenoble in France organizes annually a series of presentations for the PhD students belonging to a specific laboratory. Each laboratory conducts its own “journée de thèse” where its students present and discuss their subject with the ultimate goal to increase interactions and connectivity between researchers with similar interests. My participation in the first “journée de thèse” was just after the first three months of my PhD ,and similarly to the rest of the first-year PhD students of my lab I was called to present my topic in five minutes. I remember feeling so proud to present my plans to work on a multidisciplinary issue. Excited to answer possible questions after my presentation, I beckoned positively to that professor raising his hand. “Will you use numbers in your study?” he asked. Obviously, for a laboratory comprised mainly from hydrologists and climate scientists my work sounded too much “social”. And to make the translation, “social” for hydrologist engineers means “theoretical”. So, I guess the term “human vulnerability factors” could not find a place in their mind.

Of course, at that time I couldn’t do this kind of understanding. At a first glance, I was thinking that probably I didn’t make a good presentation, and I tried to figure out how I could improve that in the future. It was about one year later when I attended a multidisciplinary workshop for extreme weather challenges, and I got the opportunity to participate in working groups made up of social and physical scientists such as psychologists, sociologists and meteorologists as well as stakeholders. That experience helped me to regain my confidence by confirming my thinking related to social and physical aspects of my research. That was my first “official” touch with the “babylonia” that exists in such multidisciplinary collaborations, and after that I experienced similar cases in other conferences or seminars. It was apparent that for social scientists I was just an engineer with an interest on social aspects of physical phenomena whereas hydrometeorologists and practitioners characterized my work as social study. So, which is exactly my specialty? Based on my small experience gained the last three years, it comes out that some scientists need to specifically work on transferring knowledge from one discipline to the other, while translating research to a well-defined terminology commonly understood by the involved actors and researchers. Hopefully geography accommodates such multidisciplinary works, though probably a new autonomous “multi-discipline” is to be built. I am definitely interested to work and become an expert on this potential practice but I always struggle to explain that to others with one representative word. These years I tried to find a special name to characterize the “multi-discipline” on which I am working, but without great success. Still, I am not giving up.

Among other reasons, interdisciplinary topics are becoming attractive to researchers and practitioners because they are usually society-centered, meaning that they aim to solve a real-world problem crucial for the well-being of modern societies. The “marriage” of different disciplines is necessary to address the complexity of the real-world systems and their direct or indirect interactions. So, interdisciplinary projects are often linked with operational efforts. The involvement of stakeholders in research leads to a transdiscipline teamwork challenged in many aspects. In my opinion, operationalization may have either positive or negative effects on a research project. It is an incredible feeling to know that the product of your research intends to a big scope like protecting the environment or saving human lives. The potentiality for application of your research findings in practices vital for the

prosperity of society is a driving force increasing your motivation for further advancements. Though, when operational work becomes a goal in itself, research opportunities may be ignored. In this case, research quality may be degraded in favor of time- and/or cost-effective solutions that are of interest for a certain audience in the public or private sector.

My first interaction with the operational world and especially operational weather forecasting was at the second half of my first PhD year when I started my visit in the U.S. National Weather Center (NWC). I still remember the first research group meeting that I joined; scientists involved in the FLASH project were discussing their accomplishments and weaknesses in generating high-resolution hydrologic simulations contributing to the realm of hydrologic forecasting. Many ideas of how to improve and/or facilitate the work of human forecasters were falling one after the other on the table. PhD students were arguing for the ability of hydrographs, radar-based estimations and other hydro-meteorological stuff that sounded like a big noise in my ears at that moment. I felt uncomfortable. I was thinking, “What I am doing here? I am just a PhD student in a university.” When I heard my name from my advisor, I wanted to shout: “Oh, wait, I am not such an expert! I don’t know how to make it work for real?!” As you can imagine, the call was just to introduce myself and my research goals to the rest of the group. Though, I could say that I felt kind of stressed for the next three or four following meetings.

Eventually, I found the whole process of thinking how to produce research that could be potentially incorporated in operational forecasting very interesting. Until then, I had the impression that the users of a research product (e.g., in this case operational forecasters) would be the ones who would set specific demands from researchers. In my mind, meeting the forecasters would be the first step in a work-flow towards a predictive model. However, as I was observing researchers working towards operational tools, I realized that scientifically sound prototypes have to be studied and delivered to forecasters before getting any feedback from them. So, a big pre-operational effort is required, including valuable basic research. We could view the final output as a compromise between scientific and empirical knowledge. The most important then is to play this game fairly. I mean that obviously operationalization creates possible constraints (e.g., type of data to be used, time of model run, format of the output). Although these limitations should be seriously considered during the model designing, they should not constrain or reduce the research substance, but instead they should leave space for scientifically innovative efforts. But why I am telling all that? Well, it sounds probably obvious but dealing with services and their requirements is not trivial. I feel very lucky to get to know this world in such a smooth way. In my case, I have been somehow “protected” by the fact that I was conducting basic, not yet operational, research work during my thesis. This fact gave me more freedom than others to spend substantial time on exploring concepts, data or methodologies that have been interesting to me. Still, there were moments that I intensely felt the conflict between my research curiosity and the need to deliver outcomes relevant to meet certain requirements and interests related to operational forecasting. Who knows how much pressure a PhD student, funded to conduct operational research, may have! Trust me, communicating with decision-makers, understanding their needs and translating them into well-formed research and user-friendly outputs may be a very challenging task even for highly experienced researchers.

“Multi-supervised”

For me, advisors and students in a PhD project are like parents and children in a family. I keep telling to my friends and/or other PhD students that you can always observe similarities in the personality of a PhD student and his/her advisor(s) as you can identify common characteristics between parents and their children. In my opinion, it is this “chemistry” or “similarity” (i.e., how closely attitudes, values, interests and personality match between people) - as it is called in social psychology - that leads the advisor to choose the student and vice versa. In my case, my “PhD family” consists of three parents: Sandrine, Isabelle and JJ, and one child: me. My three parents are all different between themselves but they aim on the same thing: to help me to grow up and mature scientifically. I do not think this text is the most relevant place to give concrete examples, but I can tell that as a child, I have seen myself behaving similarly to some of my parents depending on the challenge I had to face in my project. Certainly, following the advice of scientists with different backgrounds and academic experiences is not trivial. Sometimes it gets tricky to converge the opinions and find the “best” (or at least a good) research direction for a multidisciplinary issue. To make it more challenging, imagine four researchers (well, I mean three experienced and a young one) coming from different schools, and generally involved in differently-oriented projects collaborating on a multidisciplinary project over long distance. Long skype meetings causing strong headaches were a reality for my advisors and me these three years.

As in every family, there can be always some disagreements or even fights between the members. But each member supports the other on every goal. And as everyone else, my strongest motivation these three years has been to make my family proud of me. Every time I was getting tired of studying, I was regaining my motivation and power by reminding to myself how much Sandrine appreciated it when I knew deeply the details of my work; how happy Isabelle was as I was finding a good way to execute interesting results and disseminate my research; and how trustful JJ became as I was becoming capable to present a complete story from my analysis. Good parents want for their children to become independent and confident members in the society. This is why good PhD advisors encourage their students to disseminate their work and be exposed to and criticized from the scientific community through conferences and publications. And of course, as it happens in the society, it is not the “parents” nor the “friends”, but the rest of the scientific community that will evaluate how much maturity I gained for my next research steps, and how ready I became for the rest of my professional life through this PhD. Though, for me it will be always the process to the fulfillment that matters. Tony Fahkry wrote: *“The journey to achieve a goal is governed by: the person you become along the way, the skills acquired, the connections made and the inner growth which takes place”*.

Résumé

Au XXI^e siècle, la prévision de l'aléa hydrométéorologique et des impacts associés aux crues rapides demeurent un défi pour les prévisionnistes et les services de secours. Les mesures structurelles et / ou les avancées des systèmes de prévision hydrologique ne garantissent pas, à elles seules, la réduction des décès lors de ces phénomènes d'inondation rapide. La littérature souligne la nécessité d'intégrer d'autres facteurs, liés aux processus de vulnérabilité sociaux et comportementaux, afin de mieux prendre en compte les risques encourus par les populations lors de ces épisodes extrêmes.

Cette dissertation conduit une analyse théorique couplés à ceux de une analyse des accidents historiques mortels afin d'expliquer les interactions qui existent entre les processus hydrométéorologiques et sociaux responsables de l'apparition de vulnérabilités humaines lors de crues rapides aux États-Unis. Des données d'enquêtes liées aux crues rapides sont examinées afin d'élaborer un système de classification des circonstances du décès (en voiture, à l'extérieur, à proximité d'un cours d'eau, dans un camping, dans un bâtiment ou en mobile-home). L'objectif est d'établir un lien entre la conception des vulnérabilités et l'estimation des pertes humaines liées à ces catastrophes naturelles. "Random forest" est utilisé et est basé sur un arbre de décision, qui permet d'évaluer la probabilité d'occurrence de décès pour une circonstance donnée en fonction d'indicateurs spatio-temporels. Un système de prévision des décès liés à l'usage de la voiture lors des crues rapides, circonstance la plus répandue, est donc proposé en s'appuyant sur les indicateurs initialement identifiés lors de l'étude théorique.

Les résultats confirment que la vulnérabilité humaine et le risque associé varient de façon dynamique et infra journalière, et en fonction de la résonance spatio-temporelle entre la dynamique sociale et la dynamique d'exposition aux dangers. Par exemple, on constate que les jeunes et les personnes d'âge moyen sont plus susceptibles de se retrouver pris au piège des crues rapides particulièrement soudaines (par exemple, une durée de près de 5 heures) pendant les horaires de travail ou de loisirs en extérieur. Les personnes âgées sont quant à elles plus susceptibles de périr à l'intérieur des bâtiments, lors d'inondations plus longues, et surtout pendant la nuit lorsque les opérations de sauvetage et /

ou d'évacuation sont rendues difficiles. Ces résultats mettent en évidence l'importance d'examiner la situation d'exposition aux risques en tenant compte de la vulnérabilité dynamique, plutôt que de se concentrer sur les conceptualisations génériques et statiques. Ce concept de vulnérabilité dynamique est l'objectif de modélisation développée dans cette thèse pour des vulnérabilités liés aux véhicules. À partir de l'étude de cas sur les crues rapides survenues en mai 2015, et en analysant principalement les états du Texas et de l'Oklahoma, principaux états infectés par ces évènements, le modèle montre des résultats prometteurs en termes d'identification spatio-temporelle des circonstances dangereuses. Cependant, des seuils critiques pour la prédiction des incidents liés aux véhicules doivent être étudiés plus en profondeur en intégrant des sensibilités locales non encore résolues par le modèle.

Le modèle établi peut être appliqué, à une résolution journalière ou horaire, pour chaque comté du continent américain. Nous envisageons cette approche comme une première étape afin de fournir un système de prévision des crues rapides et des risques associés sur le continent américain. Il est important que la communauté scientifique spécialisée dans l'étude des crues éclairées récolte des données à plus haute résolution lorsque ces épisodes entraînent des risques mortels, et ce afin d'appuyer la modélisation des complexités temporelles et spatiales associées aux pertes humaines causées par les futures inondations soudaines.

Mots-clés: Crue rapide, impacts humains, facteurs de vulnérabilité, prédiction par apprentissage virtuel, cartographie dynamique des risques

Abstract

IN the 21st century the prediction of and subsequent response to impacts due to sudden onset and localized flash flooding events remain a challenge for forecasters and emergency managers. Structural measures and/or advances in hydrological forecasting systems alone do not guarantee reduction of fatalities during short-fuse flood events. The literature highlights the need for the integration of additional factors related to social and behavioral vulnerability processes to better capture risk of people during flash floods.

This dissertation conducts a theoretical analysis as well as an analysis of flash flood-specific historic fatalities to explain complex and dynamic interactions between hydrometeorological, spatial and social processes responsible for the occurrence of human life-threatening situations during the “event” phase of flash floods in the United States (U.S.). Individual-by-individual fatality records are examined in order to develop a classification system of circumstances (i.e., vehicle-related, outside/close to streams, campsite, permanent buildings, and mobile homes). The ultimate goal is to link human vulnerability conceptualizations with realistic forecasts of prominent human losses from flash flood hazards. Random forest, a well-known decision-tree based ensemble machine learning algorithm for classification is adopted to assess the likelihood of fatality occurrence for a given circumstance as a function of representative indicators at the county-level and daily or hourly time steps. Starting from the most prevalent circumstance of fatalities raised from both the literature review and the impact-based analysis, flash flood events with lethal vehicle-related accidents are the subject to predict.

The findings confirm that human vulnerability and the subsequent risk to flash flooding, vary dynamically depending on the space-time resonance between that social and hazard dynamics. For example, it is found that younger and middle-aged people are more probable to get trapped from very fast flash floods (e.g., duration close to 5 hours) while participating in daytime outdoor activities (e.g., vehicle-related, recreational). In contrary, older people are more likely to perish from longer flooding inside buildings, and especially in twilight and darkness hours when rescue and/or

evacuation operations are hindered. This reasoning places the importance of situational examination of dynamic vulnerability over generic and static conceptualizations, and guides the development of flash flood-specific modeling of vehicle-related human risk in this thesis. Based on the case study of May 2015 flash floods with a focus in Texas and Oklahoma, the model shows promising results in terms of identifying dangerous circumstances in space and time. Though, critical thresholds for the prediction of vehicle-related incidents need to be further investigated integrating local sensitivities, not yet captured by the model.

The developed model can be applied on a daily or hourly basis for every U.S. county. We vision this approach as a first effort to provide a prediction system to support emergency preparedness and response to flash flood disasters over the conterminous U.S. It is recommended that the flash flood disaster science community and practitioners conduct data collection with more details for the life-threatening scene, and at finer resolutions to support modeling of local temporal and spatial complexities associated with human losses from flash flooding in the future.

Keywords: Flash flood, human impacts, vulnerability factors, machine-learning predictions, dynamic risk mapping

Contents

ACKNOWLEDGEMENTS	III
PREFACE	IX
RÉSUMÉ	XVII
ABSTRACT	XXI
ABBREVIATIONS	XXXI
LIST OF FIGURES	XXXV
LIST OF TABLES	XLIII
I. GENERAL INTRODUCTION	1
1. Flash flood forecasting and warning in the United States	3
1.1. The FLASH project	4
1.2. Scale of the study	5
2. Vulnerability and Risk as means to understand and predict human losses from natural hazards	7
2.1. How is the hazard depicted?	8
2.2. What does vulnerability mean?	8
2.3. How to operationalize vulnerability?	10
2.4. Which is the relevant conceptual model to link vulnerability and risk?	12
2.5. How to pass from vulnerability and risk assessment to impact prediction?	15
3. Research Hypothesis and Objectives	17
4. Document structure	18
CONCEPT AND DATA FOR FLASH FLOOD AND IMPACT-SPECIFIC VULNERABILITY DEFINITION	20
Résumé	22
Abstract	23

II.	DYNAMIC VULNERABILITY FACTORS FOR IMPACT-BASED FLASH FLOOD PREDICTION	25
1.	Introduction	25
2.	Flash flood spatial and temporal context	27
2.1.	Space-time scale of flash flood	27
2.2.	Is the flash flood spatio-temporal scale one of the vulnerability driving forces?	28
3.	Understanding social vulnerability to flash floods	29
3.1.	Considering direct human impacts as symptoms of social vulnerability	29
3.2.	Social processes and the embedded vulnerability factors	30
4.	Conceptual model for the assessment of vulnerability to flash flood	33
4.1.	Towards a definition of dynamic vulnerability to flash flooding	33
4.2.	Model concepts and the embedded functions	37
5.	Summary and Conclusions	39
III.	DATA COLLECTION AND PROCESSING	41
1.	Introduction	41
2.	Impact data	43
2.1.	The <i>Storm Data</i> for flash flood casualties in the United States	45
2.2.	Individual fatalities reclassification	48
2.3.	Flash flood event database for binary classification	51
3.	Supplementing flash flood reports with extra datasets for flash flood human losses prediction	53
3.1.	Hazard-related data	55
a.	Magnitude and time of the flash flood event	55
b.	Duration of the flash flood event	56
c.	Magnitude and duration of the rainfall event	56
d.	Flash flood severity	57
e.	Flood hazard areas	58
3.2.	Territorial data	59
a.	Local emergency services	59
b.	Distribution of human population.	60
c.	Road and river network	61
d.	Recreational areas	62
e.	Special needs buildings	62
3.3.	Social data	62
a.	Age	63
b.	Gender	64
c.	Family status.	65
d.	Poverty status	66
e.	Education	67
f.	Ethnicity/citizenship	67
g.	Language	67
h.	Disability	67
i.	Special needs population	68
j.	Home ownership	68
k.	Length of residence	68
l.	Mobile housing structure.	69
m.	Year structure built	69
n.	Phone availability	69
o.	Vehicle availability	69

p. Vehicles	70
q. Travel time to work	71
r. Time arriving at work	71
4. Summary and Conclusions	71
TOWARDS SITUATIONAL EXAMINATION OF HUMAN VULNERABILITY TO FLASH FLOOD _____	72
Résumé	74
Abstract	75
IV. A SITUATION-BASED ANALYSIS OF FLASH FLOOD FATALITIES IN THE UNITED STATES _____	77
1. Introduction	77
2. Statistical analysis of flash flood fatalities	79
2.1. Circumstances of the fatalities	79
2.2. Temporal characteristics of the fatalities	81
a. Annual and monthly distribution	81
b. Distribution based on the timing and the duration of the causative event	84
2.3. Socio-demographic profile of the victims	88
3. Spatial distribution of flash flood fatalities	91
4. Summary and Conclusions	93
TOWARDS PROBABILISTIC PREDICTION OF FLASH FLOOD HUMAN IMPACTS _____	94
Résumé	96
Abstract	97
V. MACHINE LEARNING PREDICTIONS OF FLASH FOOD HUMAN RISK RELATED TO VEHICLES _____	99
1. Introduction	99
2. Methodology	101
2.1. Target variable for vehicle-related fatal flash flood events	101
2.2. Candidate predictors of flash-flood specific human risk in the <i>vehicle</i> circumstance	103
2.3. Classification method	105
3. Random forest model for vehicle-related casualties	107
3.1. Selecting candidate predictors	107
3.2. Model training	110
3.3. Model performance	111
3.4. Variables importance	112
4. Building a prototype: May 2015 flash floods	115
4.1. Deadly flash floods in Texas and Oklahoma in May 2015	117
4.2. Mapping dynamic human risk related to vehicles	118
a. Daily estimations of vehicle-related human risk	118
b. Hourly estimations of vehicle-related human risk	121
5. Summary and Conclusions	124
VI. GENERAL CONCLUSION _____	127
1. General review	127
2. Summary of research findings	128
2.1. Research question #1: Which social processes interact with the flash flood hazard defining the dynamic human vulnerability?	128
2.2. Research question #2: Who is the most vulnerable in terms of loss of life to flash flooding?	129

2.3. Research question #3: How to quantify the dynamic relationship between the flash flood magnitude and human vulnerability? 130

3. With an eye to the future 131

3.1. Prospective work 131

3.2. Next research directions 133

References 137

Appendix 153

A. FREQUENCY OF FLASH FLOOD VICTIMS AND FATAL FLASH FLOOD EVENTS BY STATE_____ 155

B. TYPE AND SOURCES OF COLLECTED DATA, AND THEIR ROLE IN THE ASSESSMENT OF HUMAN VULNERABILITY AND RISK TO FLASH FLOOD_____ 159

C. AVAILABLE U.S. CENSUS SURVEYS AND THEIR GEOGRAPHICAL REPRESENTATION_____ 165

D. INDICATORS AND PROXY VARIABLES FOR VEHICLE-RELATED HUMAN RISK TO FLASH FLOOD_____ 169

Abbreviations

Name	Description
ABS	Agent Based Simulations
ACS	American Community Survey of the U.S. Census Bureau
AI	Artificial Intelligence
All	All of the five circumstances of flash flood fatality
AUC	Area Under the Receiver Operating Characteristic curve
BBC	Vulnerability framework of the authors Bogardi, Birkmann and Cardona
CA	Camping/Recreational areas-related
CACI	Consolidated Analysis Centers, Inc.
CC	Coping Capacity to flash flood
CIMMS	Cooperative Institute for Mesoscale Meteorological Studies
CONUS	Conterminous United States
CRED	Center for Research on the Epidemiology of Disasters
CREST	Coupled Routing and Excess Storage
CST	Central Standard Time
DDI	Disaster Deficit Index
E	Exposure to flash flood
EDAC	Earth Data Analysis Center
EM-DAT	International Emergency Events Database
EOCs	Emergency Operation Centers
EPA	Environmental Protection Agency

ESRI	Environmental Systems Research Institute
FEMA	Federal Emergency Management Agency
FF	Flash Flood
FIPs	Federal Information Processing Standards
FIRMs	Flood Insurance Rate Maps
FLASH	Flooded Locations and Simulated Hydrographs
GEOID	Geographic Identifier for each U.S. county created by the U.S. Census Bureau
GIS	Geographic Information System
HSIP	Homeland Security Infrastructure Program
IPCC	Intergovernmental Panel on Climate Change
KS test	Kolmogorov-Smirnov statistic test
LSRs	Local Storm Reports
MAF	Master Address File
MAGMA	Socially Relevant Multi-Agent systems
MH	Mobile Home-related
MRMS	Multi-Radar Multi-Sensor
NCDC	National Climatic Data Center
NCEI	National Centers for Environmental Information
NCEP	National Centers for Environmental Prediction
NFHL	National Flood Hazard Layer
NFIP	National Flood Insurance Program
NHD	National Hydrography Dataset
NOAA	National Oceanic and Atmospheric Administration
NSSL	National Severe Storm Laboratory
NWPS	National Wilderness Preservation System
NWS	National Weather Service
OFDA	Office for Foreign Disaster Assistance
OMB	Office of Management and Budget
OOB	Out-Of-Bag
ORNL	Oak Ridge National Laboratory
OT	Other/ Unknown
OU	Outside/Open or close to streams areas-related
PAR	Pressure And Release
PB	Permanent Building-related

PhD	Doctor of Philosophy
PUMS	Public Use Microdata Sample
PVI	Prevalent Vulnerability Index
QPE	Quantitative Precipitation Estimates
QPF	Quantitative Precipitation Forecasts
RF	Random Forest technique
RFCs	River Forecast Centers
RMI	Risk Management Index
ROC	Receiver Operating Characteristic
S	Sensitivity to flash flood
SF	Summary File
SHAVE	Severe Hazards Analysis and Verification Experiment
SoVI	Social Vulnerability Index
TAR	Third Assessment Report
TGS	TechniGraphics, Inc. multi-national company
TIGER	Topologically Integrated Geographic Encoding and Reference database
TNR	True Negative Rate
TPR	True Positive Rate
U.S.	United States
UN	United Nations
UNU-EHS	United Nations University - Institute for Environment and Human Security
USGS	U.S. Geological Survey
VE	Vehicle-related
VIF	Variance Inflation Factor
VPR	Vertical Profiles of Reflectivity
WCDR	World Conference on Disaster Reduction
WFOs	Weather Forecast Offices
WMO	World Meteorological Organization
WPC	Weather Prediction Center

List of Figures

I.1	Fatalities due to weather hazards for 2015 with 10- and 30-year averages. Source: U.S. Natural Hazard Statistics compiled by the Office of Services and the National Climatic Data Center from information contained in <i>Storm Data</i> . The <i>Storm Data</i> were collected from the NWS forecast offices in the 50 states, Puerto Rico, Guam and the Virgin Islands. Available at http://www.nws.noaa.gov/om/hazstats.shtml . Retrieved on 31 October, 2016.	2
I.2	The mean annual maximum unit discharge simulated by the Coupled Routing and Excess Storage (CREST) distributed hydrologic model of the FLASH system from 2002-2011. Source: Flamig [2016].	5
I.3	Daytime versus nighttime population distribution information for lower Manhattan in New York City, disaggregated to a 90-m grid (LandScan USA dataset). Source: Aubrecht et al. [2013].	6
I.4	General concept of the interdisciplinary approach toward flash flood human impacts prediction adopted in this dissertation.	7
I.5	Vulnerability framework developed by Turner et al. [2003]. Source: Turner et al. [2003] in Birkmann [2006a].	13
I.6	BBC conceptual framework developed by Bogardi and Birkmann [2004]. Source: Fekete [2010] after Birkmann [2006b].	15
II.1	Temporal and spatial scales of various hydrological processes. Figure adapted from Blöschl and Sivapalan, 1995.	28
II.2	Conceptual model of dynamic social vulnerability to Flash Flood events	36
II.3	Human behavior as a function of cognitive processes and contingent circumstances.	38

III.1	General flowchart of data processing. Parallelograms highlighted with bold black borders represent new datasets created after processing the raw data. Data sources are noted with acronyms: HSIP (Homeland Security Infrastructure Program), NHD (National Hydrography Dataset), EDAC (Earth Data Analysis Center), NWPS (National Wilderness Preservation System), Census (U.S. Census Bureau), TIGER (Topologically Integrated Geographic Encoding and Reference), FLASH (Flooded Locations and Simulated Hydrographs).	44
III.2	State-by-state percentage of flash flood victims. Percentages of victims are estimated to the total fatalities (n=1,075) reported in the <i>Storm Data</i> from 1996 to 2014 (see Annex A for the states abbreviations and estimates details). Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	46
III.3	State-by-state ratio of flash flood victims per fatal flash flood event. Ratios are estimated by dividing the number of fatalities with the number of events that led to fatalities as they are reported in the <i>Storm Data</i> from 1996 to 2014 (see Annex A for the states abbreviations and estimates details). Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	47
III.4	A) Percent of flash flood victims by circumstance. Percentages are estimated to the total 1,075 fatalities from 1996 to 2014 in whole U.S. and Puerto Rico (dark grey) and the 551 fatalities from 2001 to 2011 in the conterminous U.S. (light grey), respectively. The values on the top of the bars indicate the raw number of fatalities in each circumstance. B) Percent of fatal flash flood events by circumstance. The values on the top of the bars indicate the raw number of fatal flash flooding events in each circumstance. Some of the total 705 (1996-2014) and 385 (2001-2011) unique fatal events led to fatalities in more than one circumstances. Estimates are based on the 1996-2014 reclassified <i>Storm Data</i> fatality files	52
III.5	County-by-county number of flash flood events reported in the <i>Storm Data</i> from 2001 to 2011. Counties with fatal flash flood events that according to the <i>Storm Data</i> led to one or more fatalities are highlighted with red line. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	53
III.6	County-by-county distribution of the simulated daily unit peak discharge ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) for May 26 2015 in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	56
III.7	County-by-county distribution of the calculated mean flashiness in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	58
III.8	County-by-county distribution of the calculated high flood hazard risk area (km^2) in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	59
III.9	County-by-county distribution of the number of emergency operation centers in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	60

III.10	County-by-county distribution of the calculated number of road-river intersections in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	61
III.11	County-by-county distribution of median age of commuters traveling to work by vehicle. Data are missing for 21 counties over the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	64
III.12	County-by-county distribution of the number of male people in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	65
III.13	County-by-county distribution of the household size in the conterminous U.S.. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	66
III.14	County-by-county distribution of the available vehicles in the conterminous U.S.. The aggregate number of vehicles refers to the total number of vehicles available in the total households of the county. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	70
IV.1	A) Percent of flash flood victims by circumstance. Percentages are estimated to the total 1,036 fatalities with defined circumstance. The values on the top of the bar indicate the row number of fatalities in each circumstance. B) Ratio of flash flood victims per event with fatalities in each circumstance. The values on the top of the bar indicate the raw number of fatal flash flooding events in each circumstance. Some of the total 679 unique fatal events led to fatalities in more than one circumstances.	80
IV.2	Annual frequency of flash flood victims (n=1,075) and fatal events (n=705) from 1996 to 2014. The dashed red horizontal line represents the annual fatality median and the solid grey line represents the annual fatal event median. The values on the top of the fatality bars indicate the ratio of flash flood victims per fatal event for each year.	81
IV.3	Percentage of flash flood victims by circumstance of the death for each year. The proportions of each circumstance (if exists) in a certain year are estimated by dividing the victims in that year and circumstance to the total victims in the specific year.	82
IV.4	Monthly frequency of flash flood victims (n=1,075) and fatal events (n=705) from 1996 to 2014. The dashed red horizontal line represents the monthly fatality median and the solid grey line represents the monthly fatal event median. The values on the top of the fatality bars indicate the ratio of flash flood victims per fatal event for each month.	83
IV.5	Percentage of flash flood victims by circumstance of the death for each month. The proportions of each circumstance (if exists) in a certain month are estimated by dividing the victims in that month and circumstance to the total victims in the specific month.	84
IV.6	Cumulative percentage of flash flood victims with the duration of the causative fatal flash flood event from 1996 to 2014. The duration is estimated as the difference between begin and end time of the flash flood event and rounded in the closest value. The percentages (red dots) are calculated for each hour to the total fatalities for which the duration of the causative event can be estimated (n=954).	85

IV.7 Distribution of the flash flood duration discretized by the circumstance of the related deaths. Boxplots are constructed for each circumstance separately for the fatalities for which the duration of the causative event can be estimated and it is less or equal to 30 hours (n=921). 86

IV.8 Percentage of flash flood victims by timing of flash flood occurrence. The flash flood-occurrence time is represented by hourly intervals (e.g., flash flood occurring at 24:01 is presented with a bar between 24:00 and 01:00). Percentages of flash flood victims (gradient colored bars) in each hourly interval are estimated to the total fatalities from 1996 to 2014 (n=1,075). The color in the bars depicts the number of victims associated with events occurring in that hourly interval. 87

IV.9 Percentage of flash flood victims by the local timing of flash flood occurrence discretized by the circumstance of the related deaths. The flash flood occurrence time is represented by hourly intervals (e.g., a flash flood event occurring at 24:01 is assigned to the interval 24-01). Percentages of flash flooding victims (gradient colored bars) in each hourly interval are estimated to the total fatalities in each circumstance. The color in the bars depicts the number of victims associated with events occurring in that hourly interval and that circumstance. 88

IV.10 Percentage of flash flood victims and population by gender. A) Percentages of flash flood victims (dark grey bars) in each age group are estimated to the total fatalities with known gender from 1996 to 2014 (n=1,051). The percentage of population (light grey bars) is based on the U.S. Decennial Census: 2010. B) Percentages of flash flood victims in each gender are estimated to the total fatalities in each circumstance separately. The values on the top of the bars indicate the raw number of victims of that gender and in that circumstance. The sum of the values for female and male bars gives the total victims with known gender in each circumstance. The points represent the percentage of male and female population based on the U.S. Decennial Census: 2010. 89

IV.11 Distribution of age by the gender of flash flood victims discretized by circumstance of the death. Boxplots are constructed for each circumstance separately for the victims with known both age and gender from 1996 to 2014 (n=1,003). 90

IV.12 State-level flash flood fatalities for each circumstance (form A to F) represented as difference between the percent of fatalities in each state and the percent of the fatalities in all U.S. for the fatalities related to a specific circumstance. The number in each state indicates the exact value of the percent of difference for the states which had fatalities occurring in the depicted circumstance and refer to states with more than 5 fatalities in total from 1996 to 2014. 92

V.1 Schematic illustration of the process followed to formulate the target variable for the vehicle-related (VE) machine learning classification. 102

V.2	County-level distribution of flash flood events represented as variation from the mean (i.e., 13 events per county) for the period of 2001-2011. Counties with fatal flash flood events that according to the <i>Storm Data</i> led to one or more vehicle-related fatalities are highlighted with red solid line. Map colors based on http://www.ColorBrewer.org , by Cynthia A. Brewer, Penn State.	102
V.3	Schematic illustration of the random forest algorithm using as example binary classification for the occurrence or non-occurrence of vehicle-related (VE) fatality in flash flood events.	107
V.4	Spearman's rank correlation for the 12 variables extracted after VIF analysis (see Table III for the meaning of the variables).	109
V.5	Modeling steps (from A to F) for training (orange boxes), testing (green boxes), and applying (purple box) the random forest classifier for vehicle-related fatalities prediction.	110
V.6	Receiver Operating Characteristic (ROC) curve for the final random forest model. Sensitivity in y-axis corresponds to true positive rate (TPR) known also as hit rate. X-axis demonstrates specificity known also as True Negative Rate (TNR). False alarm cases in forecasting are estimated as 1-Specificity. The area under the curve (AUC) is shaded in light grey. The 95% Confidence Intervals for the estimated AUC are computed with 2000 stratified bootstrap replicates as defined by [DeLong et al., 1988]. The cutoff probabilities of 0.5 and 0.4, and the corresponding specificity and sensitivity values, are illustrated in black and red font, respectively. The best cutoff probability (closest top left point) for this model is estimated as 0.494 (blue point on the curve).	112
V.7	Rank of features by variable importance based on A) the out-of-bag decrease of accuracy and B) the out-of-bag decrease of Gini index.	113
V.8	Monthly observed precipitation (mm) for May 2015 in the conterminous U.S estimated by NOAA. Source: Advanced Hydrologic Prediction Service, NWS. Available at http://water.weather.gov/precip/	115
V.9	Maximum county-level daily forecast probability of vehicle-related casualty due to flash flooding, estimated by the random forest model for each day between the 16th and the 31st of May 2015.	116
V.10	Percent of flash flood victims by day of reported fatality. Percentages are estimated to the total 34 fatalities reported in the <i>Storm Data</i> for Texas and Oklahoma in May 2015. The values on the top of the bars indicate the raw number of fatalities in each day.	117
V.11	Percent of flash flood victims by circumstance. Percentages are estimated to the total 30 fatalities reported for Texas and Oklahoma, from May 16 to May 31, in the <i>Storm Data</i> . Especially, vehicle-related fatalities were reported on May 18, 20, 23, 24, 25, 26, and 29. The values on the right of the bars indicate the raw number of fatalities in each circumstance. The estimations are based on reclassification of the fatalities circumstances reported in the <i>Storm Data</i> fatality file, available at https://www.ncdc.noaa.gov/stormevents/ftp.jsp . The fatality circumstances are classified as proposed in Chapter III.	118

V.12 County-level daily forecast of vehicle-related human risk for each day (from A to D) from May 23, 2015 to May 26, 2015 for Texas and Oklahoma States: the likelihood of flash flood vehicle-related casualty for individuals is predicted by the random forest model for each county and day (i.e., 1%-100%), and is assigned to four categories from low to very high. Counties with daily unit peak discharge < 2 ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) are also assigned to the low likelihood category. 120

V.13 County-level hourly forecast of vehicle-related human risk for May 26, 2015 for Texas and Oklahoma States: the likelihood of flash flood vehicle-related casualty for individuals is predicted by the random forest model for each county and hour (i.e., 1%-100%), and is assigned to four categories from low to very high. Counties with hourly unit peak discharge < 2 ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) are also assigned to the low likelihood category. The illustration includes twelve hourly maps from 00:00 to 23:00 using a 2-hours step (from left to the right). 122

V.14 Estimated hourly predictors, and vehicle-related risk predictions by local time (Central Standard Time), for flash flooding in Harris County (Texas) on May 26, 2015. The left Y-axis refers to the simulated unit peak discharge ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) and the commuters (*people* $\times 10,000$). The predicted likelihood of flash flood casualty is presented in percentages on the right Y-axis rendered over the corresponding categories of risk. We recall that predictions which correspond to unit peak discharge < 2 ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) are masked in the low likelihood category. 123

V.15 Up: Random forest predictions for vehicle-related human risk from flash flooding at 4 a.m. on the 26th of May (2015) in Texas and Oklahoma. The illustration focuses in Harris County (Texas), and is rendered over a reference map from Google Maps to provide details of the existing road network in the area. Down: Examples of observed vehicle-related incidents from flash flooding on the highways and downtown road network in Houston area, as shared through social media. 124

VI.1 Agents and their interactions in the future agent-based model for flash flood crisis. 134

C.1 Hierarchy of Select Geographic Entities in the American Community Survey. Geographic areas with star (*) refer to areas for which data are only available in the 5-year estimates. Double star (**) notifies areas for which 5-year estimates are only available, and the first release was in 2012 (for the 2007-2011 5-year estimates). Source: Geographic Hierarchy as organized by the U.S. Census Bureau, available at <http://www.census.gov/programs-surveys/acs/geography-acs/concepts-definitions.html>. Retrieved on December 30, 2016. 167

List of Tables

I.1	Difference between a <i>Flood Watch</i> and a <i>Flood Warning</i> issued by the National Weather Service. Source: Definitions available at http://www.floodsafety.noaa.gov/ . Retrieved on October 31, 2016.	4
I.2	Research questions and the corresponding objectives that guide this thesis.	18
II.1	Social vulnerability factors.	31
II.2	Definitions of the three dynamic vulnerability components.	34
III.1	Categories of flash flooding fatalities' circumstances before and after reclassification.	49
III.2	Examples of reclassified cases in the <i>Storm Data</i> (1996-2014).	50
III.3	Number of reclassified cases of flash flooding fatalities' circumstance and percentages to the total 1,075 reported fatalities.	51
V.1	Risk indicators and the related proxy variables to serve as candidate predictors for flash flood events with vehicle-related incidents. Details on the proxies sources and processing are additionally provided in Table D.1 of Annex D.	103
V.2	Variance Inflation Factor (VIF) for the proxy variables with $VIF < \text{threshold}=2$. The variables are sorted from the ones with the least to the ones with the most variance explained by the other predictor variables in the regression.	108
V.3	Description of the nine candidate predictors.	109
V.4	Predictive performance of alternative models on the test dataset. Full model is the selected optimal model including all of the nine predictors (see Table V.3). Additional models are built by removing one by one the least important predictors. The reduced model includes six predictors (i.e., EOCs, household size, and median age predictors are excluded).	114

A.1 Frequency and percentages of flash flood fatalities and fatal flash flood events by state for the 50 states, the District of Columbia, and the territory of Puerto Rico (including their ranks in fatalities) for the period 1996-2014. Ratios are estimated by dividing the number of fatalities with the number of fatal flash flood events in each state. 155

B.1 Summary of collected data types and sources, and their role in the assessment of human vulnerability and risk to flash flood. Indicators are proposed as relevant to one or more circumstances (ALL: all circumstances, VE: Vehicle-related, OU: Outside/Open or close to stream area, CA: Camping/Recreational area, PB: Permanent Building, MH: Mobile Home). 160

C.1 U.S. Census Bureau’s surveys explored for data availability and suitability in this dissertation. 166

D.1 Summary of processing and interpretation of proxy variables to serve as candidate predictors for flash flood events with vehicle-related human losses. 169

General introduction

Flash floods are high-impact, occasionally catastrophic events that result from the intersection of hydrometeorological extremes and society at small space-time scales, generally on the order of hours [AMS, 2000; Ruin et al., 2009]. The impacts of these events include both damage to property and threat to life [French et al., 1983; Staes et al., 1994; Grunfest and Handmer, 2001; Petersen, 2001; Jonkman, 2005; Vinet, 2008; Gaume et al., 2009; Llasat et al., 2010]. Flash flood severity is shaped by a range of factors including topography, land cover, seasonality, the distribution of flood structures (e.g., dams, bridges, culverts) and critical infrastructure (e.g., schools, hospitals, electricity industry), and human population. Because flash flood events are generally very localized in space and time, they are difficult to forecast with precision and can subsequently leave people uninformed and subject to surprise in the midst of their daily activities.

In the United States (U.S.) flood is the second most devastating hazard after heat in terms of number of fatalities, with flash floods to account for the majority of those [Ashley and Ashley, 2008]. Recent examples of fatal flash flood events in the U.S. include the Albert Pike campground flood in Arkansas that killed 20 campers on June 11, 2010 [Holmes Jr and Wagner, 2011], and the Oklahoma City flash flood on May 31, 2013 that killed 13 people [Yussouf et al., 2016]. Another case of major flash flooding in the U.S. occurred in May 2015, with Oklahoma and Texas to be among the most impacted states. On May 14, 2015, prior to extensive flooding beginning around May 24, flash flood warnings were issued for counties in southeast Texas. At least 34 people lost their lives in flash floods from May 6 to 29, including 30 victims in Texas and 4 in Oklahoma. On September 14, 2015 in Hildale, Utah, 19 people were killed in a flash flood event characterized as the most deadly weather disaster in Utah history [Avila, 2016]. In 2015, the National Weather Service (NWS) reported 176 flood-related fatalities, noticeably overriding the 10- and 30-year average fatalities per year (Figure I.1). Sixty-seven percent of those 176 flood fatalities were attributed to flash flooding¹. Only heat wave has a higher 30-year average number of fatalities than flooding.

1. Data available online at <https://www.ncdc.noaa.gov/stormevents/ftp.jsp>.

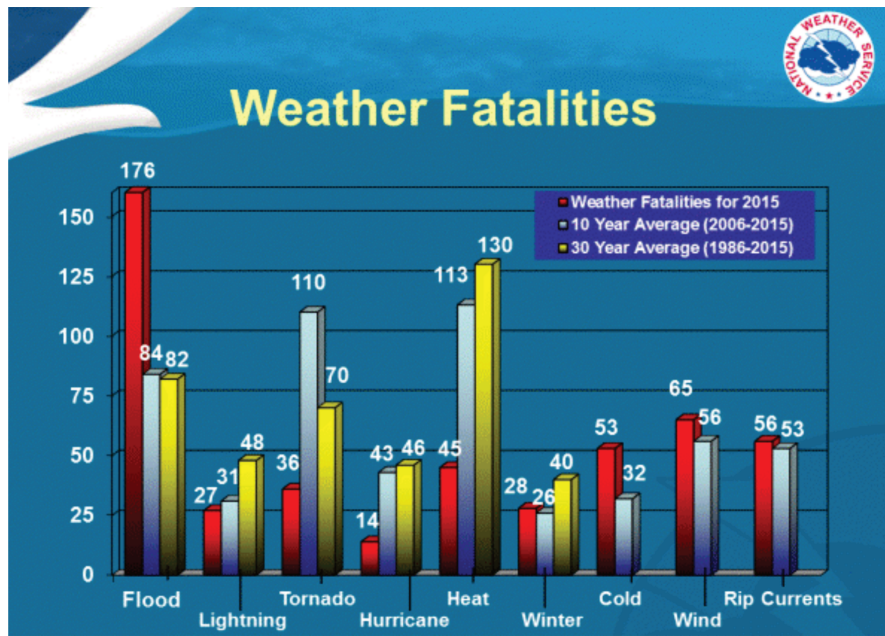


Figure I.1 – Fatalities due to weather hazards for 2015 with 10- and 30-year averages. Source: U.S. Natural Hazard Statistics compiled by the Office of Services and the National Climatic Data Center from information contained in *Storm Data*. The *Storm Data* were collected from the NWS forecast offices in the 50 states, Puerto Rico, Guam and the Virgin Islands. Available at <http://www.nws.noaa.gov/om/hazstats.shtml>. Retrieved on 31 October, 2016.

Worldwide, a survey on disaster risk reduction conducted by the World Meteorological Organization (WMO) found that 105 out of 139 responding countries identified flash floods as the second most common hazard after strong winds. Interestingly, 91 of the total 99 countries that responded as issuing flash flood warnings, highlighted the demand for improvements in the warnings [WMO, 2008]. Given the sudden onset nature of flash flood phenomenon, the integration of flash flood risk in prevention measures like the ones used in regular flooding (land-use planning, flood zoning) is questionable. Thus, the development of advanced forecasting and warning tools is of higher consideration to support emergency response for public protection. In the next section, the existing flash flood warning system in the U.S. is briefly discussed to present the engagement of the current research in human impacts forecasting and monitoring of human risk from flash flooding. Section 2 presents how vulnerability and risk research can complement natural hazards science towards a better realization of the circumstances leading to flash-flood specific human impacts. Especially, in the introductory part of this thesis critical points regarding the state-of-the-art in vulnerability to natural hazards are discussed to frame our research intentions.

This dissertation is focused on the primary consequence of flash flood events: human losses. The area of study refers to the U.S. which commonly experiences dangerous flash floods leading to an important amount of fatalities each year. The focal point of the research is to understand the main circumstances in which the casualties occur and to go toward the prediction of their occurrence.

1 Flash flood forecasting and warning in the United States

Advances in flash flood forecasts and warnings are very important to increase preparedness and response capacity of the local authorities and population to flash flood crisis. Hydrologic and natural hazard sciences show a great interest in the improvement of tools for monitoring and forecasting flash floods recognizing flash flooding as an extremely sudden release of water producing high risk for human life [Carpenter et al., 1999; Borga, 2009; Marchi et al., 2010; Borga et al., 2011; Hapuarachchi et al., 2011]. The U.S. NWS Glossary (2009) defines a flash flood as: “*a rapid and extreme flow of high water into a normally dry area, or a rapid water level rise in a stream or creek above a predetermined flood level, beginning within six hours of the causative event (e.g., intense rainfall, dam failure, ice jam)*”. This definition is not universally accepted in the scientific literature [Gaume et al., 2009]. Though, it is in accordance with definitions given from the WMO which marks the time scale between four to six hours, and it serves as the starting point for the organization of operational flash flood forecasting and monitoring in the U.S. In the U.S. NWS, the timescale of six hours is used to divide operational responsibility between local weather forecast offices that issue flash flood warnings and regional river forecast centers that issue river flood warnings.

In the U.S., hydrometeorological hazards are communicated to the public by the NWS federal governmental agency. In terms of spatial and temporal scales of the forecasts, the forecasting and alerting responsibilities are distributed according to time scales: the longer scale forecasts and the short-term predictions. The NWS Weather Prediction Center (WPC), part of the National Centers for Environmental Prediction (NCEP) is responsible for long-range interpretation of automated weather forecast guidance, providing 0-72 hours (0-3 days) forecasts of heavy or excessive rainfall (i.e., quantitative precipitation estimate exceeding flash flood guidance), and 0-168 hours (0-7 days) quantitative precipitation forecasts (QPF). Thirteen regional River Forecast Centers (RFCs), mostly employing hydrologists, cover all 50 states and the U.S. territories focusing primarily on riverine flooding. At a local level, the flash flood alerting enterprise is administered by 122 NWS Weather Forecast Offices (WFOs). Meteorologists in the WFOs issue point-based forecasts for their local areas of responsibility. These point forecasts include probabilities of precipitation and QPFs. WFOs issue *flash flood watches* when there is a fifty to eighty percent chance of flooding conditions in the next 48 hours¹. Warnings are issued when flash flooding is considered as “*imminent or likely*” over a period generally less than six hours (or up to twelve hours in specific cases) (Table I.1).

If impacts arise from the interplay of multiple, natural and social factors, then in principle at least, an early warning system should address all of the factors relevant to the particular risk. Solely based on hydrometeorological products, the aforementioned warning messages are restricted to generalized advice to the public. Obviously, despite technological advances in forecasting have largely improved watch-warning systems during the last decades, the prediction of prominent impacts of this phenomenon remains a big challenge. Forecasters of the NWS do their best to collect as many information as possible (e.g., hydrometeorological products, social media), to make sense of the situation before and after a flash flood occurrence. They are also responsible for collecting ground truth reports to validate their warnings. Still, a more concrete tool that could inform forecasters of the

1. For documentation on the WFO hydrologic products specification visit <http://www.nws.noaa.gov/directives/sym/pd01009022curr.pdf>.

Table I.1 – Difference between a *Flood Watch* and a *Flood Warning* issued by the National Weather Service. Source: Definitions available at <http://www.floodsafety.noaa.gov/>. Retrieved on October 31, 2016.

Flood product	Interpretation
<i>Flash Flood Warning</i>	Take Action! <i>A Flash Flood Warning is issued when a flash flood is imminent or occurring. If you are in a flood prone area move immediately to high ground. A flash flood is a sudden violent flood that can take from minutes to hours to develop. It is even possible to experience a flash flood in areas not immediately receiving rain.</i>
<i>Flood Warning</i>	Take Action! <i>A Flood Warning is issued when the hazardous weather event is imminent or already happening. A Flood Warning is issued when flooding is imminent or occurring.</i>
<i>Flood Watch</i>	Be Prepared: <i>A Flood Watch is issued when conditions are favorable for a specific hazardous weather event to occur. A Flood Watch is issued when conditions are favorable for flooding. It does not mean flooding will occur, but it is possible.</i>
<i>Flood Advisory</i>	Be Aware: <i>An Flood Advisory is issued when a specific weather event that is forecast to occur may become a nuisance. A Flood Advisory is issued when flooding is not expected to be bad enough to issue a warning. However, it may cause significant inconvenience, and if caution is not exercised, it could lead to situations that may threaten life and/or property.</i>

potential risk to people is required to go towards more targeted warnings and protection actions in specific areas.

Initiatives towards improving tools for flash flood modeling and prediction across the U.S. include the Flooded Locations and Simulated Hydrographs (FLASH) project that was launched in 2012 in the University of Oklahoma [Gourley et al., 2017]. The FLASH project presented in the next section promotes the development of probabilistic impact focused outputs to advance the state-of-science in operational flash flood predictions by integrating human impacts aspects with FLASH hydrometeorological products.

This thesis was partly conducted at the University of Oklahoma, and was partly funded by the FLASH project, with the aim to conduct basis research for a nationwide prediction effort for forecasters and emergency managers to target their warnings on anticipated human impacts during flash flood events in the U.S..

1.1 The FLASH project

The FLASH team¹, led by Dr. Gourley, is comprised of researchers and students who use an interdisciplinary and collaborative approach to advance the accuracy, timing, and specificity of flash flood warnings in the U.S., with the ultimate goal to save lives and protect infrastructure from flash flood hazards. The project is funded by the Disaster Relief Appropriations Act of 2013 (P.L. 113-2), providing support to the Cooperative Institute for Mesoscale Meteorological Studies (CIMMS) at the University of Oklahoma. The FLASH system generates hydrometeorological products at flash flood scale in real-time across the conterminous U.S. including rainfall average recurrence intervals, rainfall-to-flash flood guidance ratios, and distributed hydrologic model-based discharge forecasts

1. For information on the FLASH system and the team members visit <https://blog.nssl.noaa.gov/flash/>.

[Gourley et al., 2017; Flamig, 2016]. Especially, it uses the National Severe Storm Laboratory’s (NSSL’s) Multi-Radar Multi-Sensor (MRMS) forcing and produces flash flood forecasts at 1-km/5-min resolution through direct, forward simulation (Figure I.2). The developed FLASH products are displayed online in a real-time demonstration environment¹.

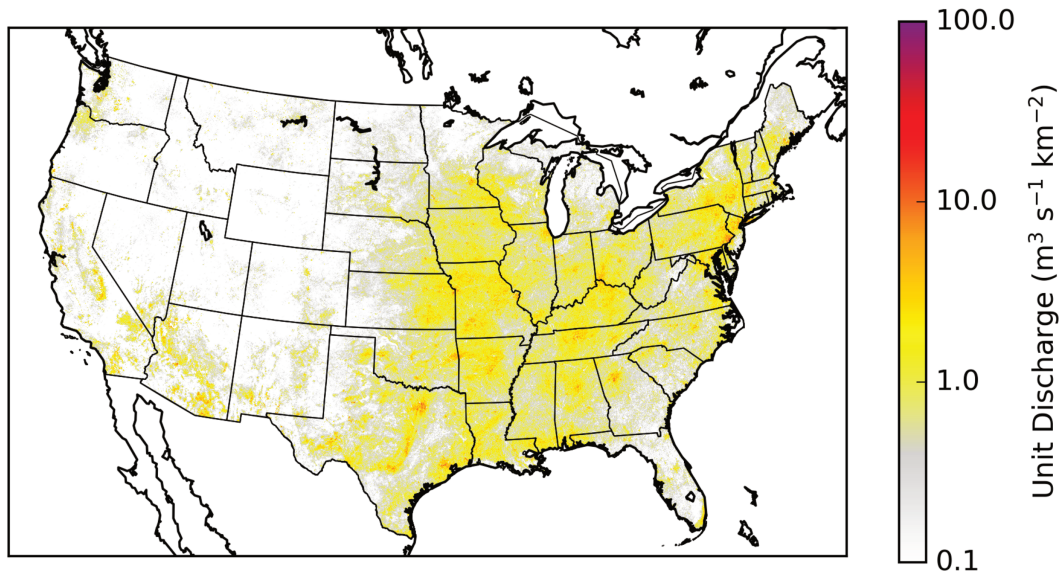


Figure I.2 – The mean annual maximum unit discharge simulated by the Coupled Routing and Excess Storage (CREST) distributed hydrologic model of the FLASH system from 2002-2011. Source: Flamig [2016].

To evaluate the tools developed within the FLASH hydrologic modeling framework, the MRMS Hydro Experiment is organized annually [Martinaitis et al., 2017]. During the experiment, NWS forecasters work with research scientists to assess emerging hydrometeorological concepts and products to enhance flash flood watches and warnings. The experience of FLASH scientists gained by working for years in close collaboration with operational forecasters in the National Weather Center (NWC), provide valuable insights related to operational needs and constrains guiding our research. Such needs refer primarily to the scale of analysis and applicability discussed in the next section.

The conceptual and methodological developments of this thesis were briefly discussed in a free-conversation style with the NWS forecasters participated in the “Hydro Experiment 2015” held at the NWC from 6 July to 24 July, 2015². The forecasters recognize the need for a tool to synthesize information about the exposed population and the surrounding environment, and the importance of quantifying their relationship with the hydrologic forecasts, in order to have on hands an auxiliary metric (probabilistic or not) to use in a warning decision support role.

2. Details on the *Hydro 2015 Operations Plan* are available at http://blog.nssl.noaa.gov/flash/wp-content/uploads/sites/7/2014/06/HMT-hydro_exp_plan_2015.pdf.

1.2 Scale of the study

Within the concept of FLASH, this dissertation aims to predict the dynamics of human impacts with respect to hydrological forecast outputs across the conterminous U.S.. Dynamic human vulner-

1. Product description and real-time forecasts are available at <http://flash.ou.edu>.

ability, and eventually risk, rely on the hypothesis that since people and their activities are rarely static but move across space at different times (Figure I.3), the dominant vulnerability factors also change correspondingly. Human impacts depend on the intersection of the timing and the location of the flash flood event with the social profile and activity of the exposed people. For example, the lives of active people (e.g., workers, students) are expected to be threatened when flash flooding interrupts daytime commutes, whereas residential flooding is more likely to increase the exposure of people during nighttime rest hours, and especially vulnerable population such as the elderly, when incapable to evacuate.

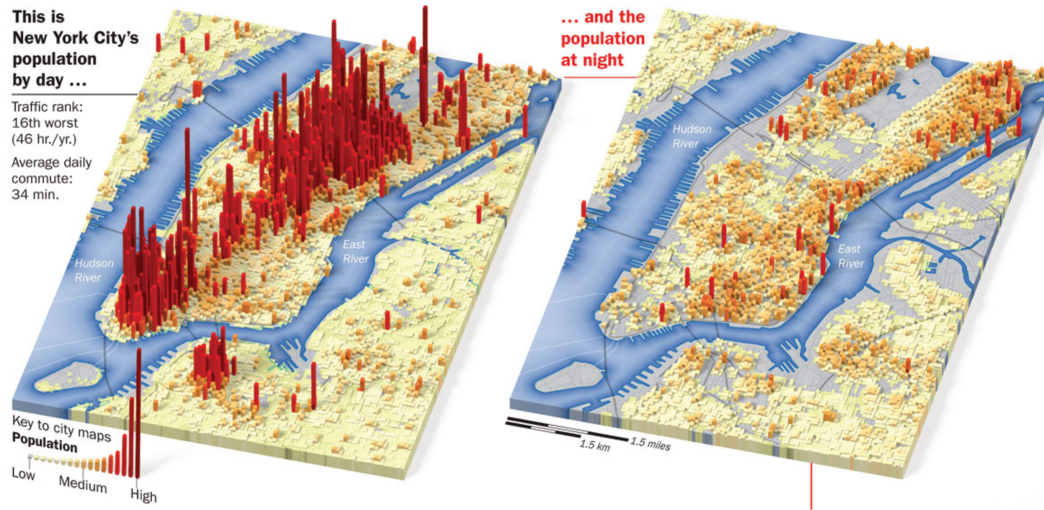


Figure I.3 – Daytime versus nighttime population distribution information for lower Manhattan in New York City, disaggregated to a 90-m grid (LandScan USA dataset). Source: Aubrecht et al. [2013].

Explaining the social-environmental interactions that drive the distribution of flash flood human impacts across space and time requires an interdisciplinary approach to be adopted in this dissertation. Especially, the study seeks to combine knowledge and data from hydrology and human geography to integrate the physical and social dynamics leading to forecasts and understanding of circumstance-specific human losses from flash flooding (Figure I.4). The variety of the origin of the data (e.g., hydrological, meteorological, spatial, demographic) to be included in the analysis suggests the divergence of spatial and temporal resolutions. For example, although discharge forecasts from FLASH are available in the order of kilometer and minutes, information for the human population in areas exposed to flash flooding may refer to much coarser administrative geographic units over U.S., usually valid for a number of years.

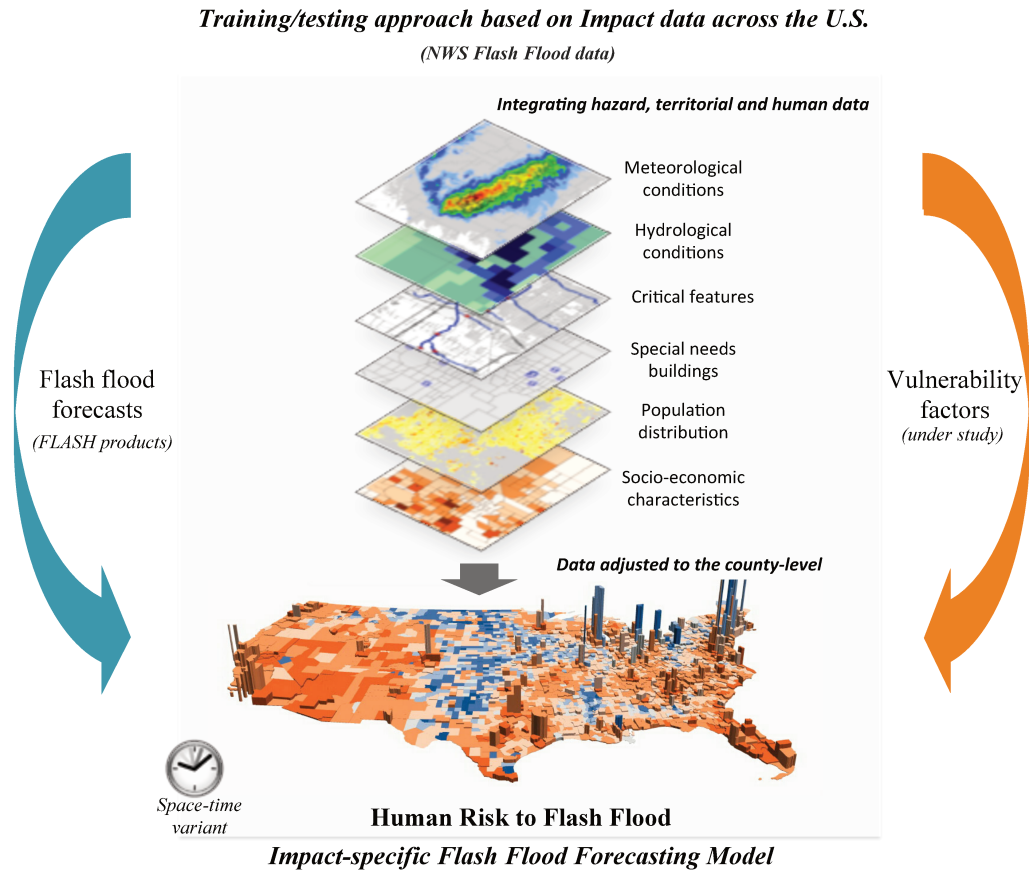


Figure I.4 – General concept of the interdisciplinary approach toward flash flood human impacts prediction adopted in this dissertation.

In this dissertation, the need for a nationwide prototype research implies the identification of a spatial and temporal resolution relevant to adjust the inhomogeneous spatial and temporal reference units of the input data, and to support the application of the methodological developments over the U.S.. The main challenge that arises in this research is to compromise between capturing localized dynamic vulnerability and hazard processes, still distributed over the U.S., and the use of nationwide readily available data (usually defined in coarser resolution) for a reproducible research.

2 Vulnerability and Risk as means to understand and predict human losses from natural hazards

Risk concept introduces the occurrence of outcomes resulting from the interaction of a physical phenomenon (e.g., flash flood) and social vulnerability in certain locations and times. The Internationally agreed glossary of basic terms related to disaster management defines risk as “*expected losses (of lives, persons injured, property damaged, and economic activity disrupted) due to a particular hazard for a given area and reference period*” [DHA, 1992]. Therefore, the terms human risk or human impacts may be used interchangeably in the present study, with a focus on the likelihood of loss of life during flash flooding.

In the last decades the science of vulnerability has been gaining space in Disaster and risk research bringing together experts from different social and geophysical disciplines. The former secretary general of United Nations (UN), Kofi Annan, highlights the difference between hazard and disaster, with

the latter to exist only when people's lives and livelihoods are swept away by a hazard [Annan, 2003]. Since the 70s O'Keefe et al. [1976] and the tenants of the political ecology approach have stated that disasters do not come out from the natural hazard itself. Instead they argue that social, economic, and political constrains of individuals are the main drivers of the increase of human vulnerability and the related hazards impacts. Disaster or disaster risk when expressed in human terms (loss of lives, people affected) is therefore the outcome of a hazard, and depends on the physical, social, economic properties of the system that is exposed to and interacts with the hazard. Examples may refer to physical disabilities of people, risk awareness and perception, lack of financial resources or social capital for emergency response, poor constructions.

Researchers working on multidisciplinary studies related to natural hazards and climate variability or change, often have different interpretations of vulnerability, but also of terms that are supposed to be well-defined like hazard and risk. Birkmann [2006b] provides a list of essential terminology in the assessment of vulnerability and the disaster risk reduction domain in general (Birkmann, 2006b, p. 453). With such an existing profusion of terms, in the following pages we propose to clarify the meaning of the ones that are selected as working definition for our study.

2.1 How is the hazard depicted?

As a determinant of risk, the notion of hazard is defined as *“a dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage”* [UN/ISDR, 2009]. Especially for hydrometeorological hazards (e.g., thunderstorm, tornado, drought, coastal flood, flash flood), the dangerous phenomenon corresponds to processes of atmospheric, hydrological and/or oceanographic nature. The natural hazard under consideration is commonly defined in probabilistic terms as *“the probability of occurrence associated with an extreme event that can cause a failure”* [UNDRO, 1991; Plate, 2002] or *“the probability of occurrence, within a specific period of time in a given area, of a potentially damaging natural phenomenon”* [Cardona et al., 2003].

In this thesis flash flood hazard is considered as a natural process that exceeds specific thresholds and become a threat for human losses and damages in the human-environmental system. As an alternative to the probabilistic representation, we may use the magnitude (e.g., discharge, accumulated rainfall) or intensity (e.g., rainfall rate, time to peak discharge) of the hazard event, may be combined with factors such as speed of onset, duration and spatial extent, as proxies of the occurrence of a hydrometeorological extreme at specified locations and times. The hazard can be combined then with vulnerability proxies (e.g., exposed people, sensitive characteristics of the environment) leading to an integrated dynamic estimation of human risk.

2.2 What does vulnerability mean?

In the context of risk to natural hazards, vulnerability describes both the social processes driving the potential for harm and/or characteristics of individuals or groups of people that make them susceptible to be harmed physically and/or psychologically; a concept that evolved out of the social sciences in the 1970s as an alternative to the hazard or techno-centered paradigme developed in

the 1940s based on which disaster risk is mostly due to a natural cause and a lack of perception or adjustment to it [White and Haas, 1975; Schneiderbauer and Ehrlich, 2004]. From that time vulnerability has taken various definitions depending on the research objective and the author's background with the ultimate goal to analyze the human-environmental conditions and interactions within socio-ecological systems threatened and impacted by a stressor [Adger, 2006]. The majority of definitions in the literature tend to view vulnerability either i) as a pre-existing state of the social system defined independently of the hazard occurrence, or ii) as a potential for losses caused by the system's exposure to a particular hazard and its sensitivity to specific impacts. The former idea is supported mainly by purely social-oriented scientists representing vulnerability through a set of selected socio-economic characteristics that reveal the inherent fragility of the system [Allen, 2003; Sarewitz et al., 2003], whereas the latter is promoted by climate change community which links vulnerability with the likelihood of impact occurrence considering the specificity of the hazard under study. The Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report (TAR) describes vulnerability with the following statement:

“The degree to which a system is susceptible to, or unable to cope with, adverse effects of climate change, including climate variability and extremes. Vulnerability is a function of the character, magnitude, and rate of climate variation to which a system is exposed, its sensitivity, and its adaptive capacity.” (IPCC, 2001, p. 995)

In the climate change perspective, McCarthy [2001] defines sensitivity as *“the degree to which a system is affected, either adversely or beneficially, by climate-related stimuli. The effect may be direct (e.g., a change in crop yield in response to a change in the mean, range, or variability of temperature) or indirect (e.g., damages caused by an increase in the frequency of coastal flooding due to sea level rise)”*. Thus, sensitivity is a term that explains the degree to which a system is modified or affected by hazards. Adaptive capacity is *“the ability of a system to evolve in order to accommodate environmental hazards or policy change and to expand the range of variability with which it can cope”* [Adger, 2006].

Conceptualizations that consider social vulnerability as an inherent property of the society may adequately explain social groups that are the most fragile from an economic or physical point of view (or as a result of other types of marginalization processes). However, they focus on the negative side of the vulnerability concept, ignoring positive aspects related to people's strength to deal with the hazard or their capacities for self-protection [Wisner, 2003; Blaikie et al., 2014]. Bohle [2001] highlights that vulnerability can not be described without considering the capacity to anticipate, cope with, resist and recover from the impact of a hazard, defined as the internal side of vulnerability. The term “coping” is used in his conceptual framework to represent coping and response capacities as they emerge from the Crisis and Conflict Theory (e.g., control of assets and resources, capacity to manage crisis situations), Action Theory Approaches (e.g., how people act freely as result of socio-economic or governmental constrains) and Model of Access to Assets (e.g., access to resources/assets) [Bohle, 2001]. According to Bohle, exposure to specific risks and shocks is one of the vulnerability components described as “external” side of vulnerability, which together with the internal vulnerability side compose the “double structure of vulnerability”. This external aspect encompasses spatial exposure but also social inequities, population dynamics, and incapacity of people to obtain assets that can alter the exposure of individuals and households to risk.

In this study, social vulnerability is considered as a whole, integrating the pre-existing social conditions that make people susceptible to the prominent hazard (inherent social vulnerability), and the hazard-specific sensitivities emerging from the interaction of the exposed people and the hazard in specified circumstances across space and time. The term social is used to bound vulnerability factors to human beings [Blaikie et al., 2014]. Building on the IPCC definition, vulnerability is viewed as a process defined by a set of social sub-processes related to human exposure, sensitivity and coping capacity that evolve in time and thus, can not be assessed in advance or mapped as a static picture independently of the hazard (see Chapter II). The term “coping” instead of “adaptive” capacity is used to describe the short-term ability of people to deal with the strength of the hazard (here flash flood).

2.3 How to operationalize vulnerability?

To support decision makers and mitigation planning for disaster risk reduction, the conceptualizations of vulnerability need to be translated in specific metrics or measurement tools to be used for risk assessment in practice. Currently, social vulnerability assessment research is driven by the selection of relevant vulnerability indicators and criteria. In the 2005 World Conference on Disaster Reduction (WCDR), the international community underlined that it is important to *“develop systems of indicators of disaster risk and vulnerability at national and sub-national scales that will enable decision-makers to assess the impact of disasters on social, economic, and environmental conditions and disseminate the results to decision-makers, the public and population at risk”* [ISDR, 2005]. Following explanations given by Gallopin [1997], Birkmann [2006b] defines an indicator of vulnerability to natural hazards as:

“a variable which is an operational representation of a characteristic or quality of a system able to provide information regarding the susceptibility, coping capacity and resilience of a system to an impact of an albeit ill-defined event linked with a hazard of natural origin.”

Birkmann [2006b] highlights the fact that indicators may be defined differently by the various authors in the literature, though they always intend to have a special meaning for the estimation of a certain quality or characteristic of a system based on interpretations about the relationship between the indicator and the phenomenon of interest. That means that every variable that is selected to be included in the vulnerability assessment should indicate a specific aspect or process of the underlying vulnerability to the hazard(s) under study [Gallopin, 1997]. Certainly, there are limitations on the selection and use of indicators. According to Cutter et al. [2009] the main problems discussed in the literature can be summarized as following:

- Research requirements to assign complex concepts and interactions into a set of variables (compromise between good knowledge of the system and its specifics, and the need for simplicity in applications).
- Difficulty to quantify some vulnerability concepts and interactions between the social vulnerability processes (e.g., social networking, cognition, trust in the government and warnings).

- Data availability and resources constrain the selection of input variables relevant to explain the natural and social processes on a scale that is suitable for the purposes of the analysis (geographic unit, timeframe).
- Lack of compatibility between the methods of indicators aggregation, and reproducibility of the indexes. Various indexes are constructed following different approaches driven from the available variables, geography, and certain study-specific interpretations.

In the frame of the socially oriented studies where vulnerability is considered as an internal property of the society, Cutter et al. [2009] highlights some broad indicators that appear frequently in the literature using different proxies: the socioeconomic status (e.g., wealth or poverty); the age; the special needs populations (e.g., people in hospitals); the gender; and the race and/or the ethnicity are some of the most commonly used characteristics [Tierney et al., 2001; Center, 2002]. These indicators summarize social dependencies and economic disadvantages of the population through indexes assigned to geographic units varying from block groups to states [Cutter et al., 2000; Cutter et al., 2003]. Especially, the eleven social vulnerability indicators proposed by Cutter et al. [2003] have largely been used in various studies presented the literature [Rygel et al., 2006; Azar and Rain, 2007].

Given the multifaceted nature of vulnerability to different hazards, adopting indicators that may be available through the literature is not the best practice. In the words of Brooks et al. [2005], *“vulnerability depends critically on context, and the factors that make a system vulnerable to a hazard will depend on the nature of the system and the type of hazard in question”*. For example, flood insurance may indicate the existence of preparedness measures specifically reducing vulnerability to flood but not to other hazards (e.g., wind). Flood insurance can not directly reduce vulnerability during flooding but may facilitate the recovery process after a flood disaster [Tunstall, 2009; Zhong et al., 2013]. In addition to that, such flood prevention measures may be relevant for some countries and type of economies but not for others (e.g., economic differences between developed and developing countries, differences in insurance policy between Europe and U.S.). Rufat et al. [2015] review 67 flood-hazard case studies (1997-2013) to present the main factors considered when assessing social vulnerability to floods. Their results show that the demographic and socio-economic characteristics, and health and coping capacity issues are the most frequently used ones in the quantification of social vulnerability. Though, the frequency varies depending on the flood type (e.g., riverine or flash flood), disaster phase (e.g., response or recovery) and place of application (e.g., developed or developing country) [Rufat et al., 2015]. In this perspective, studying social vulnerability to a specific temporal and spatial context of the flood hazard is a key step to identifying relevant and measurable indicators [Fekete, 2010]. It also helps to explain the causative processes avoiding generalizations and simplifications in vulnerability assessment and mapping.

Downing [2004] argues that a relevant conceptual model has to be developed to be used as the basis for indicator development and selection. The conceptual model should be guided from the objective of the analysis, and be in respect to the spatial and temporal bounds of the phenomenon (e.g., scales of flood phenomenon) and application (e.g., vulnerability of certain administrative units) of interest. Thus, defining the objective of the analysis implies the identification of the area or territory and hazard of interest, and the scope of the study (e.g., preventing human impacts or economic losses).

In this thesis, a dynamic vulnerability conceptual model is developed to explain the main factors and their interactions, related to the vulnerability of individuals during the flash flood crisis. Based on the identified flash flood-specific vulnerability factors, nationwide available data in the U.S. will be gathered and explored for their relevance to serve as indicators of human risk to flash flood hazard at the county level^a, and daily or sub-daily time steps.

^a. In the U.S., a county is a political and geographic subdivision of a state and is used for the level of local government. At the 2000 U.S. Census, the median land area of U.S. counties is 1,610 km². Documentation of the 2000 U.S. Census Bureau geographic entities is available online at <http://www.census.gov/prod/2001pubs/mso-01icdp.pdf>.

2.4 Which is the relevant conceptual model to link vulnerability and risk?

Several conceptual frameworks or models have been developed in the literature to set vulnerability in the context of risk analysis, and help researchers to systematize the measurement of vulnerability aspects relevant to the hazard and the social system of interest. Birkmann [2006b] and Cutter et al. [2009] both provide a review and/or criticism of well-known conceptual models in vulnerability research. Here, we discuss the main attributes of core previous conceptual models that lead the way for framing vulnerability in our study: i) the components included in the vulnerability term (e.g., exposure, sensitivity, capacity), ii) the characterization of vulnerability (e.g., process or characteristic), and iii) the dynamics considered (e.g., temporal or spatial interactions between vulnerability components and the hazard).

One of the most well-known conceptual frameworks in the literature that views risk as the intersection of processes associated with the natural hazard event and vulnerability is the pressure and release model (PAR model) presented in the *At Risk* volume [Blaikie et al., 1994; Wisner et al., 2004]. In this framework, vulnerability is considered as a progressive process from root causes (e.g., limited access to resources), to dynamic pressures (e.g., lack of local institutions or training, socio-demographic changes such as rapid urbanization) to unsafe conditions (e.g., unprotected infrastructure, special groups at risk, low income levels, lack of disaster preparedness). Though, the model is criticized for not explicitly addressing the interactions between social and natural systems [Cutter et al., 2009].

Integrated multidisciplinary approaches that combine vulnerability of a system with exposure to particular hazards are applied in vulnerability (or risk) mapping for identifying particularly vulnerable (or critical) regions [o'Brien et al., 2004; Metzger et al., 2005]. In this direction, Cutter [1996] presents the hazard-of-place approach to integrate the place-based interaction between potential exposure and societal vulnerability with a specific focus on particular places or regions [Cutter et al., 2000; Cutter et al., 2006]. In this framework exposure is differentiated from social vulnerability, and it is defined as biophysical vulnerability influenced by the geographic context (i.e., site and situation of the place, proximity to hazard). Thus, social vulnerability is focused on the community ability to cope with, respond to or recover from one or more hazards, and it is mainly described by the economic, demographic, and housing characteristics of the place. This perspective of social vulnerability agrees more with the definition of vulnerability as an intrinsic characteristic of the place or the system of interest (independent of the hazard type). Unfortunately, it does not address the dynamics of vulnerability emerging from the embedded socio-environmental interactions, and the intersection of human activities with the hazard dynamics.

Turner et al. [2003] presents an interesting dynamic approach considering interactions between human and environmental processes operating in local, regional or global spatial levels. Yet, temporal dynamics are not explicitly apparent in the proposed feedback-loop system. The framework provides a scheme to link vulnerability with the broader human-environmental conditions within the global environmental change context. Opposed to Cutter’s approach, vulnerability in Turner’s framework encompasses i) exposure (i.e., threatened individuals, households, ecosystems) to a threat of a specific magnitude, duration and frequency, ii) sensitivity (i.e., social and natural capital and endowments), and iii) coping capacities, with the last to be extended to the system’s resistance and ability for adaptation, named resilience (Figure I.5). Given the complexities associated with the interactions and inter-dependencies illustrated in the model, the applicability of Turner’s conceptual model in research applications is questionable [Cutter et al., 2009].

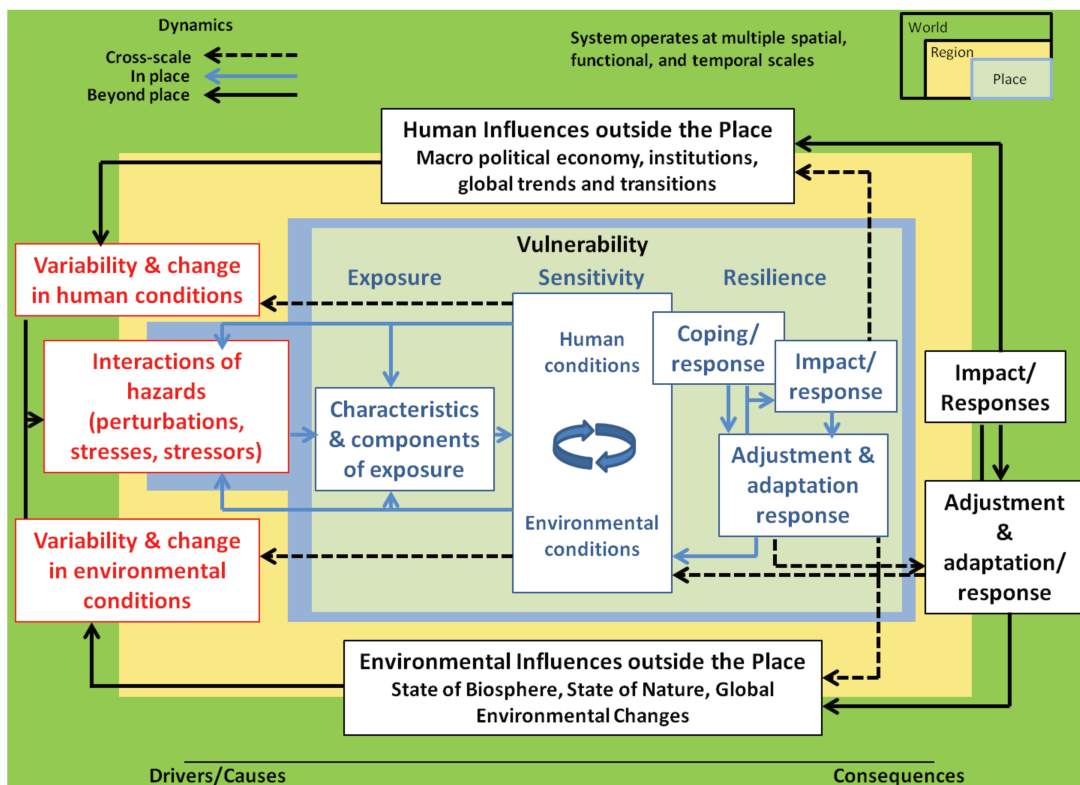


Figure I.5 – Vulnerability framework developed by Turner et al. [2003]. Source: Turner et al. [2003] in Birkmann [2006a].

In the domain of disaster risk reduction Bollin et al. [2003] adopted the conceptual framework initially developed by Davidson and Shah [1997] for the assessment of a city’s disaster risk in case of earthquake, to present vulnerability as one of the determinants of disaster risk at the level of commune in Latin America. In this framework, vulnerability describes the physical, social, economic and environmental factors that affect the potentiality for impacts on the infrastructure (e.g., damages), the individuals (e.g., injury, death), the economy (e.g., disruption of economic flows) or the socio-political system (e.g., deficiencies of individual institutions and organizations) within a community. The main characteristic of this framework is that vulnerability is distinguished from the exposure and capacity/measures components under the disaster risk umbrella. Thus, risk is viewed as the sum of hazard (natural occurrence and/or intensity of a hazardous phenomenon; in terms of probability and severity), exposure (i.e., infrastructure, population or systems that are subject to the physical

presence of the hazard), and vulnerability (i.e., susceptibility of the exposed elements to damage or loss), minus the capacity (e.g., urban planning, emergency response, communications) of the exposed system to deal with the hazard.

Within the context of sustainability, the BBC framework developed by Bogardi and Birkmann [2004] at the Institute of Environment and Human Security of the United Nations University (UNU-EHS) integrates environmental, social and economic aspects of human safety with the occurrence of a natural phenomenon related to a specific hazard. In this framework, the economic, social or environmental risk arises from the combination of the hazard with the corresponding vulnerability sphere; both viewed as sequential components in the risk assessment procedure. In contrast with the disaster risk community, the BBC framework (Bogardi and Birkmann Conceptual framework) defines exposure, susceptibility and coping capacities explicitly as elements of vulnerability (Figure I.6). This framework implies for consideration of the specifics of the hazard and distinguishes two temporal phases for the overall reduction of risk: just before the hazard strikes (preparedness), and after the hazard effects start to be apparent (emergency response). Fekete [2010] applies the BBC model to assess social vulnerability of German population to river flooding. He focuses especially on the social sphere and based on the BBC framework he selected river flood-relevant indicators that explained social susceptibility at the county-level; thus providing comparisons of social vulnerability estimations among the counties of Germany. An interesting contribution of this framework is that vulnerability is viewed as a process rather than a simple characteristic, encompassing capacities that may reduce vulnerability. In this perspective, exposure, susceptibility and capacities are interconnected vulnerability processes that mediate the anticipated risk. The common attribute of the BBC framework and Bollin's model is that they consider social vulnerability and hazard as separate components in the realm of risk. Although the dependencies and interactions with the hazard are broadly discussed, they often presume a linear relationship between those two contributors and risk in practice. However, we believe that social vulnerability processes intersect with the hazard occurrence in a possibly complex and non linear manner revealing a dangerous scene for the exposed human or environmental system.

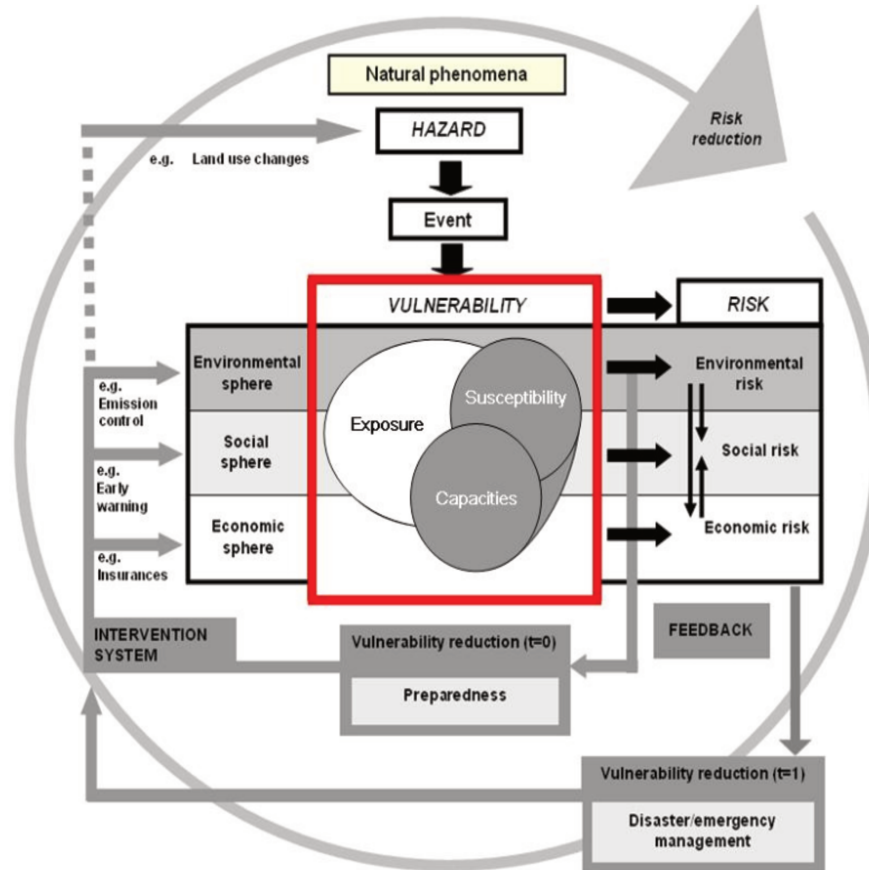


Figure I.6 – BBC conceptual framework developed by Bogardi and Birkmann [2004]. Source: Fekete [2010] after Birkmann [2006b].

The conceptual frameworks developed in the literature are the entry point for qualitative description and understanding of the concept of vulnerability and its relation to risk. The existing models may provide useful insights for the construction of a conceptual model dedicated to the hazard and social system of interest with explicit description of the involved processes.

This thesis considers a process-oriented flash flood and impact-specific conceptual framework to link social vulnerability with human impacts occurring during flash flood events. Building upon the dynamics of the coupled human-environmental system introduced by Turner et al. [2003], we argue that social vulnerability, and subsequently human risk to flash flood, is not a constant estimation but an evolving outcome that depends on the interaction of physical and social processes at specified spatial and temporal units. Dynamic vulnerability encompasses exposure, sensitivity and coping capacity of individuals that all vary over time and space, making people and property susceptible to harm when they are exposed to a specific flash flood occurrence.

2.5 How to pass from vulnerability and risk assessment to impact prediction?

Risk assessment represents the methodological framework for determining the level of risk *by analyzing potential hazards and evaluating existing conditions of vulnerability that together could potentially harm exposed people* [UN/ISDR, 2009]. Generally, there is no uniquely defined methodology

to combine indicators of risk in the literature. In recent years, most of the studies aggregate vulnerability indicators to compose indexes at the national level for international or global-oriented projects [UNDP, 2004; Cardona, 2005]. Examples of indexes that aim to assess vulnerability and risk towards natural hazards include the *Disaster Deficit Index* (DDI) for the expected financial loss and capacity [Cardona, 2005], the *Prevalent Vulnerability Index* (PVI) for the socio-economic fragility and infrastructural capacity to recover from natural hazards [Cardona, 2005], and the *Risk Management Index* (RMI) for the capacity related to risk identification and reduction, disaster management and financial protection [Carreño et al., 2007]. At the national level, indexes use only one numeric value to describe an entire country ignoring the possible spatial variability of vulnerability within that country and/or its temporal evolution. Cutter et al. [2003] developed the *Social Vulnerability Index* (SoVI) based on the social dimensions of the PAR model to quantify the relative socio-economic and demographic quality of a place as a means of understanding vulnerability at the county level.

In majority, the aforementioned approaches view vulnerability as a static metric of social aspects that is studied independently of the hazard, as they aim to be generic and multi-hazard. However, in fast-evolving events such as flash floods, human impacts depend not only on variables such as the magnitude of the natural hazard and the vulnerability of those affected, but also on how these factors evolve and intersect in space and time. These contextual factors can alter the scale distribution, and magnitude of impacts on people. Thus, indexes may be descriptive of the existing and foreseen conditions of the coupled human-environmental system, but they can not serve as predictive tools for specific impacts [Cutter et al., 2009].

To predict combinations of physical and social characteristics and processes favorable for the outbreak of impacts (e.g., fatalities, injuries, damages) within a flash flood or other hazard event, several variables have to be related to observed impact data. Therefore, we need an integrated approach allowing to test interrelationships between social and physical indicators with respect to their ability to explain past human impact occurrences in the geographic unit of interest. Methodologically, this approach can be carried out through “*supervised*” machine learning techniques, where the dependent attribute or label (e.g., occurrence of human losses), is defined as the variable to be predicted, and is part of the dataset inserted in the machine learning algorithm [Kohavi and Provost, 1998]. The goal of the algorithm is then to learn general rules that map the inputs to desired outputs. This is not the case in “*unsupervised*” learning where the label is not specified as part of the dataset, and the machine learning algorithm is allowed to cluster cases drawn from the dataset into classes naturally driven from the data [Kohavi and Provost, 1998].

Machine learning is a technique belonging to the broader field of artificial intelligence (AI), that “*gives computers the ability to learn without being explicitly programmed*” [Simon, 2013]. Although there is a disagreement on how exactly to define “learning”, a concrete way of doing so is to consider learning as a process of acquiring knowledge that the learner can use to develop a set of rules [Quinlan, 1986]. Rapid rise of data availability during the last century discourage manual interpretations by the human brain. On the other hand, computers and subsequently machine learning, are designed to perform repetitive assignments such as developing sets of rules based on the analysis of big datasets (learning). According to Quinlan [1986], “*domain specialists*” and “*knowledge engineers*” need to collaborate to create explicit rules outlining and defining the knowledge available about the operation of a particular “*expert system*” in the modern world, but this approach may lead to “*a few rules per*

man day”, opposed to the computers and by inference, the machine learning techniques, that can rapidly process thousands of rules.

Forecasting of flash flood impacts relies on the already complicated *expert system* of weather forecasting. Machine learning may therefore serve as a promising tool to complement hydrologic forecasts with vulnerability-related variables used as predictors, in order to capture the complex and dynamic rules related to life-threatening situations during flash flooding. That means that risk indicators should first be chosen based on theoretical knowledge before being sorted out by machine learning algorithms to help identifying complex patterns and relationships that would not be detectable through vulnerability hypothesis and two-dimensional statistics. Machine learning algorithms include support vector machines [Vapnik and Cortes, 1995], artificial neural networks [Rojas, 2013], and regression or classification trees [Breiman et al., 1984; Quinlan, 1986], to name but a few. Such techniques have been applied in hydrological and meteorological studies including extreme rainfall [e.g., Nayak and Ghosh, 2013], and tornado development from mesocyclones [e.g., Trafalis et al., 2014]. Clark [2016] used machine learning models to forecast the probability of flash flooding given a set of atmospheric and hydrologic conditions in the contiguous U.S., and explore their applicability in operational forecasting. Recently, data-driven models have been further applied in assessing flood damage based on multiple variables describing the flooding hydrology and warnings, building characteristics and precaution measures, and the socio-economic status of private households [Merz et al., 2013].

In this dissertation, a machine-learning technique is applied to a compiled database with indicators about the hydrometeorology of the flash flood event, and the infrastructure and socio-demographics of the exposed county to produce, for the first time, automatic probabilistic forecasts of flash flood fatalities for a given life-threatening circumstance.

3 Research Hypothesis and Objectives

The main goal of this PhD research is to propose a conceptual and methodological framework to link social vulnerability conceptualizations with realistic forecasts of prominent impacts from flash flood hazards. Especially, we see this study as entry point for a forecasting system to anticipate potential human losses, with a focus on the most prevalent circumstance of fatalities: vehicle-related incidents.

In this direction, we hypothesize that:

- i. Integrating flash flood impact reports with extra datasets describing critical elements of the hazard, the population and the environment in the counties exposed to those flash floods, would provide us a base for building a machine-learning approach for flash flood human risk prediction.
- ii. Machine-learning approaches are relevant to link flash flood occurrence with the potentiality of circumstance-specific casualties by integrating exposure, sensitivity and coping capacity characteristics defining human vulnerability to flash flooding.

To develop our approach and test this hypothesis core research questions need to be answered. Table I.2 summarizes the three main questions that pose three challenging objectives that lead this thesis.

Table I.2 – Research questions and the corresponding objectives that guide this thesis.

Keywords	Research question	Objective
Social vulnerability	Which social processes interact with the flash flood hazard and define the dynamic variability of human vulnerability across space and time?	To build a conceptual vulnerability model that captures the dynamic interplay between relevant vulnerability factors and the spatiotemporal scales of flash flood events.
Human impacts	Who is the most vulnerable to flash flooding in terms of loss of life?	To analyze the situations of historic fatality events as determined by the victims' profile and activity and the spatio-temporal context of the causative flash flooding.
Forecasting of prominent human losses	How to quantify the relationship between the magnitude of the flash flood and proxies revealing the exposure and susceptibility of people at certain circumstances at the time of the event?	To model and map dynamically circumstance-specific human risk predictions revealing the time-variant exposure to a given flash flood forecast.

4 Document structure

The first and the last chapters serve as the general introduction and general conclusion of this thesis, respectively. The rest of the dissertation is organized in four chapters which document the integrated hazard-vulnerability approach toward flash flood human losses prediction.

After framing the research intentions in the present chapter, the second and third chapters describe the main concepts and data adopted in this research, respectively. **Chapter II** presents findings from a literature review on past flood impact and vulnerability studies and lists the driving factors that control social vulnerability to flash floods. Human vulnerability is defined and contextualized to short-fuse floods to consider the spatial and temporal specificity of the flash flood hazard. Then, a conceptual model for the assessment of social vulnerability and by inference, of human risk to flash flood events is developed and presented accordingly.

Chapter III presents the collection of nationwide data from various sources in the U.S., and their pre-processing to serve as indicators of the main vulnerability processes identified in the developed conceptual model at the scale of the analysis (county-level). Especially, historic flash flood impacts records are supplemented with a set of variables that may explain the hazard occurrence but also the

sensitive characteristics of the exposed population and built environment corresponding to the set of flash flood events under study.

Chapter IV explores the situations of the fatality events as determined by the victims' profile and activity and the spatio-temporal context of the flash flooding in the U.S.. Impact data presented in Chapter III are used to carry out a statistical and geospatial analysis of historic human losses in the U.S.. Especially, 1075 flash flood-specific human losses from 1996-2014 on the scale of the U.S. are classified to six main categories/circumstances and are investigated correspondingly. This analysis provides interesting insights for the understanding of human vulnerability in past flash flood events, and the prediction of human losses in future events; investigated in Chapter V of the thesis.

Chapter V discusses methodological developments allowing the integration of physical and social dynamics leading to model forecasts of circumstance-specific human losses during a flash flood. A machine learning predictive approach is developed based on the flash flood database prepared for binary classification in Chapter III. Especially, a Random Forest classifier is applied to assess the likelihood of fatality occurrence for a given circumstance (i.e., vehicle related incidents) as a function of representative indicators. Details on the classification method and the modeling steps are presented in this chapter. Chapter V provides also descriptions of the final model performance and the contribution of each indicator/predictor in the prediction of vehicle-related fatalities at the county-level across U.S. In the continue, the catastrophic flash floods of May 2015 in the states of Texas and Oklahoma are used as a case study to map the dynamics of the estimated probabilistic risk on a daily and hourly scale, by applying the developed model. Achievements and limitations of the developed modeling tool are finally discussed to evaluate the ability of the proposed approach to realistically capture severe cases of flash flooding in terms of vehicle-related incidents.

Each of the Chapters II, IV, and V focuses on one of the three research objectives presented in Table I.2, and correspond to the three main publications produced during the dissertation. The general conclusions of this thesis in **Chapter VI** are presented in terms of general review and summary of findings in each of the three research questions (Table I.2). The summary and conclusions are followed by a discussion on the perspectives of the current research, as well as potential research directions for determining casualties from flash flood hazards in the future.

CONCEPT AND DATA FOR FLASH FLOOD AND IMPACT-SPECIFIC VULNERABILITY DEFINITION

Chapter II: Dynamic vulnerability factors for impact-based flash flood prediction

Chapter II is based on a paper published in *Natural Hazards, Journal of the International Society for the Prevention and Mitigation of Natural Hazards*: Terti, G., Ruin, I., Anquetin, S. and Gourley, J.J., 2015. Dynamic vulnerability factors for impact-based flash flood prediction. *Natural Hazards*, 79(3), pp.1481-1497.

Chapter III: Data collection and processing

Résumé

Les circonstances sociologiques et humaines qui transforment un événement naturel en catastrophe mortelle peuvent être expliquées par la vulnérabilité sociale. Mais quelles sont les caractéristiques spatio-temporelle de la vulnérabilité (i.e., la vulnérabilité dynamique) qui influencent la façon dont les individus sont touchés par un aléa naturel particulier? Le Chapitre II présente une analyse critique des études antérieures réalisées sur les impacts humains et les vulnérabilités associées, liés aux crues rapides. L'étude est motivée par l'hypothèse que la résonance des échelles spatio-temporelles associées à l'aléa avec celles liées à la localisation des personnes, en connaissant leurs caractéristiques socio-économiques, révèle différents types de vulnérabilités et définit la zone la plus exposée en termes de mortalité. Sur la base de cette idée, un modèle conceptuel pour évaluer la vulnérabilité face à ces aléas naturels est développé et présenté ici. La principale avancée, par rapport aux études précédentes, est l'introduction du concept de la variabilité spatiale et temporelle de la vulnérabilité et du risque associé. Cela signifie que le modèle proposé ne considère pas la vulnérabilité sociale comme un synopsis statique, décrit par une seule carte, mais comme un processus en constante évolution dérivée de l'interaction des dynamiques sociales et physiques. Ce concept de vulnérabilité dynamique est essentiel pour identifier les variables pertinentes à utiliser pour évaluer les risques associés aux crues éclair, la cartographie dynamique et la prévision. La collecte de telles variables à partir de données nationales aux États-Unis est présentée dans le Chapitre III. Le modèle développé sert de base à l'élaboration d'une approche méthodologique pour quantifier les risques humains, en tenant compte des circonstances dans lesquelles les victimes ont perdu la vie lors de l'épisode de crue. De manière plus précise, les données d'impact de ces aléas sont complétées par des données qui décrivent les caractéristiques de l'aléa, de la population exposée et de l'environnement, dans les comtés ayant déjà subi des crues rapides. C'est sur cette base de données ainsi compilée que la conceptualisation du modèle de prévision de l'occurrence des crues rapides, et des éventuelles victimes associées, sera construite en intégrant l'ensemble des données définissant l'aléa et la vulnérabilité humaine face aux crues éclair (i.e. exposition, capacité d'adaptation, ...).

Abstract

Social vulnerability explains the sociological and human-dependent circumstances that translate a natural event into a deadly disaster. But, what are the space-time characteristics of vulnerability (i.e., dynamic vulnerability) that influence how people are impacted by a specific natural hazard? Chapter II presents a critical analysis of previous flood-related human impact and vulnerability studies to better understand and summarize the human-related factors that determine the impacts from flash flood events. The study is motivated by the hypothesis that the intersection of the spatiotemporal context of the flash flood hazard with the distribution of people and their characteristics across space and time reveals different paths of vulnerability and defines the most probable space of an exposed area in terms of deadly impacts. Based on this idea, a conceptual model for assessing vulnerability to flash flooding is developed and presented herein. The most important advance of the current research in comparison with previous efforts in vulnerability assessment is the introduction of the concept of the spatial and temporal variability of vulnerability and the subsequent risk. This means that the proposed conceptual model does not consider social vulnerability as a static synopsis that can be described by a single map, but as an ever-evolving process derived from the interaction of social and physical dynamics. The dynamic perspective of vulnerability is key for the identification of pertinent variables to be used for flash flood risk assessment and dynamic mapping, and prediction. The gathering of such variables from nationwide datasets in the U.S. is presented in Chapter III. The developed conceptual model is used as the basis to build a methodological approach towards quantification of human risk considering the circumstances in which people lost their life in past flash flood events. Especially, flash flood impact data are supplemented with extra data that may describe the characteristics of the hazard and the exposed population and built environment at counties exposed to historic flash flood occurrences. It is hypothesized that the compiled database will support the incorporation of our conceptualizations in a statistical approach to link flash flood occurrence with the potentiality of casualties by integrating exposure, sensitivity and coping capacity characteristics defining human vulnerability to flash flooding.

Dynamic vulnerability factors for impact-based flash flood prediction

1 Introduction

Flash floods are the most dangerous floods since they can occur with little or no warning, restricting the anticipation time of effective response [Creutin et al., 2013]. For different types of flood (e.g., river flood, flash flood), there may be differences in the drivers of human vulnerability depending on the way that the hazard intervenes social processes, and whether it creates opportunities for anticipation or not. When scrutinizing the socio-demographic variables in flash flood-related human losses, for example, it becomes obvious that they reflect the space-time distribution of everyday life activities (e.g., commuting to work in a vehicle), revealing both the dynamics of exposure and the difficulty to adapt patterned movements to fast-changing and potentially dangerous conditions. In fact, in such fast-evolving events, impacts depend not only on variables such as the magnitude of the natural hazard and the vulnerability of those affected, but also on how these factors evolve and intersect in space and time. In the case of flooding fatalities, for instance, the elderly are often thought to be the most vulnerable. But when fatalities are mapped against basin scale and response time, it has been shown that in fact it is young motorists who are most likely to be killed in flash flooding in small catchments, whereas the elderly most frequently perish in their homes from large-scale fluvial flooding [Ruin et al., 2008]. As a consequence, generic vulnerability factors addressing the overall fragility of populations with poor biophysical, social, and/or financial capital fail to capture the variability of the situations in which people become vulnerable and perish in flash flooding conditions.

To date, there has been very little work on the identification of vulnerability factors that are specific to short-fuse weather events and even fewer studies on the intersection between human behavior and flash floods [Ruin et al., 2008; Creutin et al., 2009]. The objective of this study is to introduce a new conceptual framework for the analysis of social vulnerability to short-fuse weather events

(here flash floods) taking into account the spatial and temporal characteristics of the natural hazard and its interaction with the social dynamics at the daily and subdaily timescales. This conceptual framework is a necessary first step toward the development of a dynamic vulnerability model to be coupled with flash flood forecasting tools for the development of impact-specific flash flood forecasting products that will be implemented at first in the U.S.. Such flash flood forecasting products include impact-explicit maps that would inform forecasters and emergency managers for the likelihood of human impacts to occur at an area exposed to flash flooding at a certain time step. To capture the spatiotemporal variability of the human impacts, this probabilistic outcome should be the product of the magnitude of the flash flood forecast with the vulnerability of the exposed infrastructure and people at that time.

To address the objective, we begin this thesis with a literature review of impact assessments with a primary focus of flash flood events in Europe, North America, and Australia. In addition to data availability, common features in terms of human development indices and living conditions permit making parallels between Europe and North America and contribute to the applicability of this paper's findings to these regions. For example, we assume that similar economic conditions (i.e., high-income countries) have similar capabilities to assign financial resources to flood risk forecast, mitigation, protection, and recovery. The ability of high-income countries for more advanced flood risk prevention and management could explain the fact that although they are affected more frequently by natural catastrophes they have relatively fewer fatalities [Jonkman, 2005]. If the conceptual model proposed in this study is generalizable to the types of countries described above, the development of a specific flash flood impact-forecasting tool implies careful consideration of the cultural specificities for the country of reference. In other words, certain classification and thresholds of the variables to be used to explain the vulnerability processes of the conceptual framework need to be adapted depending on both the hydrometeorological thresholds and the habits of population in the area. Data availability and resolution pose additional constraints to the transferability of the concepts and methods in this research.

This chapter addresses the following central questions that enforce an innovative perspective of vulnerability assessment:

- i. How do the space and time scales of flash flood events interact with vulnerability and influence the magnitude and type of human impacts?
- ii. What are the human-dependent processes (i.e., vulnerability factors) that are related to flash flood human risk?
- iii. What are the interactions between the flash flood and the social vulnerability processes that determine the dynamic variability of vulnerability across space and time?

The primary outcome of this part of the study is an integrated conceptual vulnerability model that seeks to capture the dynamic interplay between the identified space-time vulnerability factors and the spatiotemporal scales of flash flood events. The chapter is structured in the following manner. First, the spatial and temporal aspects of flash flood events are discussed in the context of how they pertain to social vulnerability. The next section presents findings from a literature review on past flood impact and vulnerability studies and lists the driving factors that control social vulnerability to flash floods. Then, we provide insights into the definition of dynamic vulnerability and present a conceptual model for the assessment of social vulnerability to flash flood events. The final section provides concluding

remarks and discusses methodological challenges in dynamic vulnerability assessment introducing the approach adopted in this dissertation.

2 Flash flood spatial and temporal context

2.1 Space-time scale of flash flood

Blöschl and Sivapalan [1995] state that “hydrological processes occur at a wide range of scales, from unsaturated flow in a 1 m soil profile to floods in river systems of a million square kilometres; from flash floods of several minutes duration to flow in aquifers over hundreds of years”. The hydrological processes in Figure II.1 are classified by Blöschl and Sivapalan [1995] after Anderson, Burt, et al. [1990] based on the spatial and temporal ranges at which they actually operate. This type of scale is referred to as “intrinsic” in the literature and is differentiated from the “observational” scale that processes are measured or sampled, although the two types of scales sometimes converge [Anderson, Burt, et al., 1990; Wu and Li, 2006]. The spatial scale on the graph refers to the length that each process extends to whereas the temporal range represents its characteristic response time to the triggering hydrometeorological process of a specific duration and extent.

Flash flood events result from several of the processes across space-time scales presented in Figure II.1. For example, the runoff from infiltration excess acts almost instantaneously at a very small length scale whereas saturation excess runoff operates at a certain catchment area characterized by a longer response time. According to the “space-time correspondence principle”, mesoscale or synoptic scale meteorological events are associated with slower streamflow responses whereas localized thunderstorms yield fast responses in smaller, headwater catchments [Wu and Li, 2006]. The faster the response, the higher the demand for details to detect the underlying short-term vulnerability patterns and processes. Fast responses are also associated with higher complexity in the human environmental system under study [Kienberger, 2007].

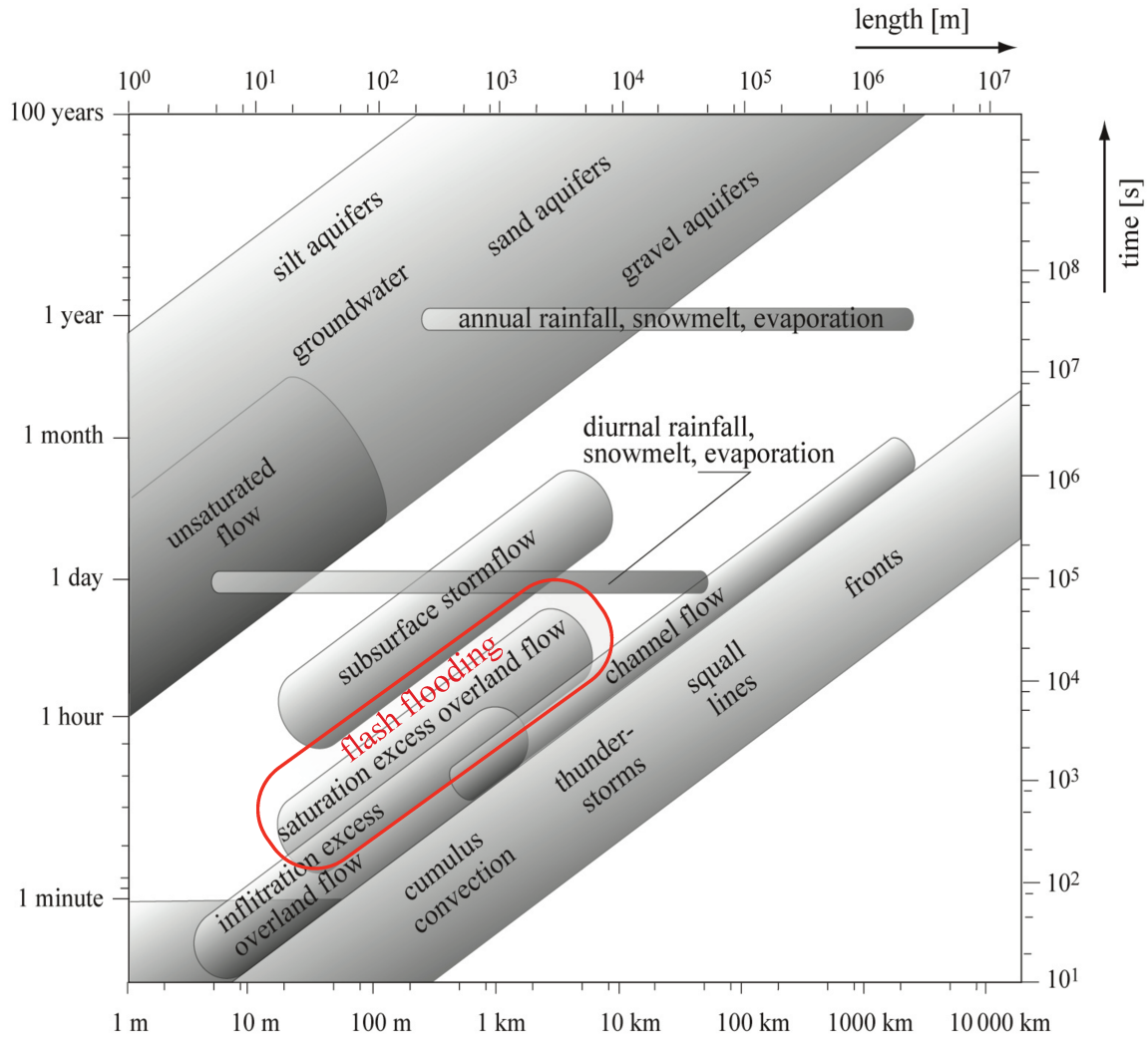


Figure II.1 – Temporal and spatial scales of various hydrological processes. Figure adapted from Blöschl and Sivapalan, 1995.

Specific meteorological factors (i.e., rainfall rates, storm motion), hydrologic parameters (e.g., antecedent soil moisture, surface imperviousness), and geomorphologic characteristics (e.g., relief, soil depth and type) with flooding events are also tied to specific space-time scales, but their quantification goes beyond the scope of the present study. Here, we focus on the scaling of the resulting floods rather than the meteorological and hydrologic factors that drive them. Flash flooding happens quickly on a time scale of minutes to a few hours usually over small drainage areas on the order of a few hundred square kilometers. These are the most dangerous floods because they can occur with little or no warning and are often associated with fast-moving waters that can sweep humans and their cars off their intended path [Montz and Grunfest, 2002]. Hereafter, we constrain our space-time window to a small one over which social vulnerability has to be analyzed in order to observe the coupled physical-social patterns that lead to high human impacts.

2.2 Is the flash flood spatio-temporal scale one of the vulnerability driving forces?

The type of flood and especially its space-time dynamics has an effect on the type of impacts and level of harm suffered [Jonkman, 2003; Ryan and Hanes, 2009; Vinet et al., 2012]. As reported in

Jonkman [2005], the severity of impact varies not only with the place where the event happens but also with the type of hazard as shown by the difference in mortality rate between flood and flash flood. The same study reveals that the difference between the two flood types not only contribute to the type and magnitude of losses such distinct phenomena trigger but they also play a role in the emergence of specific forms of vulnerability that are not relevant in the case of general flooding. The effect of the flash flood spatio-temporal specificity on vulnerability is summarized as follows:

- In contrast with river flooding where the proximity to streams and rivers indicate a potential risk level, the spatial distribution of small drainage areas prone to flash flooding limits the efficacy of flood zoning measures for flash flood prevention. Therefore, the development of advanced warning systems is preferred.
- The small spatial and temporal scales associated with flash flooding hinder the forecasting ability to predict their precise locations with sufficient warning lead-time. Unlike river floods, producing and disseminating accurate and timely forecasts that meet the human security needs remain a challenge [Montz and Grunfest, 2002].
- Flash flood dynamics such as the hydrologic response time to rainfall or concentration time and the peak magnitude of streamflow vary with catchment size and basin physiographic and geomorphological characteristics. In general, small catchments (few km²) are characterized by short response times limiting the anticipation time for effective response [Creutin et al., 2009]. Flash flood events have a greater chance to trap people in their vehicles or during activities outside, especially during times of the day when commuters are on the road, i.e., rush hour [Ruin et al., 2008]. Outdoor locations such as the road networks are where most of the fatalities occur with flash flood events [Sharif et al., 2012; Diakakis and Deligiannakis, 2013]. On the other hand, river floods occurring at large hydrological scales (hundreds to thousand km²) are more typically responsible for building damages, evacuations, and inside drowning of physically vulnerable populations [Ruin et al., 2008].

3 Understanding social vulnerability to flash floods

3.1 Considering direct human impacts as symptoms of social vulnerability

Social studies of vulnerability have traditionally focused on the loss of life or physical and mental health problems [Enarson, 2007]. Data are available concerning lethal consequences of flash flood events (e.g., *Storm Data* reports from U.S. NWS's National Climatic Data Center (NCDC)), but these measurable phenomena constitute a small subset of social impacts, many of which are difficult to classify, operationalize and measure. Usually, demographic characteristics (e.g., age and gender) of the flash flood victims are analyzed to explore the possible factors that made them susceptible to flooding [French et al., 1983; Coates, 1999; Ashley and Ashley, 2008; Cutter et al., 2009; Sharif et al., 2012; Doocy et al., 2013]. Flood casualties vary by region and with the flood intensity. However, the following findings about vulnerability have emerged from the analysis of flood consequences on lives and property during a flood event.

- Lethal consequences are mostly related to flash floods than to river floods [Jonkman, 2005; Ashley and Ashley, 2008]. Jonkman [2005] analyses 632 flood events reported in the OFDA-CRED International Emergency Events Database (EM-DAT) maintained by the Center for Research on the Epidemiology of Disasters in Brussels (CRED) in cooperation with the U.S. Office for Foreign Disaster Assistance (OFDA). For the period 1975-2001, Jonkman [2005] shows that unlike river floods, flash flood events are characterized by a high mortality rate per event (5.6% versus 0.47% for river flooding), meaning that even if more people are affected by river flooding, much less perish in those circumstances.
- Most of the flood fatalities in the U.S. [Staes et al., 1994; Ashley and Ashley, 2008; Maples and Tiefenbacher, 2009; Sharif et al., 2012], Australia [Coates, 1999; FitzGerald et al., 2010] and Europe [Jonkman and Kelman, 2005; Diakakis and Deligiannakis, 2013] are vehicle-related with drowning being the main cause of death [French et al., 1983; Ryan and Hanes, 2009]. Since the velocity of the water is usually very high, flash flooding is much more dangerous for motorists [Ruin et al., 2008]. According to FEMA [2015] only 0.15 m of swift-moving water is needed to move a car and 0.61 m of moving water can carry larger vehicles.
- Most of the flash flood impacts take place during the “event” phase and most fatalities happen together with the peak of the hydrological event and sometimes even before the official warnings are issued [Duclos et al., 1991; Staes et al., 1994; Jonkman and Kelman, 2005; Ruin et al., 2008; Sharif et al., 2012]. [Špitalar et al., 2014] found that there was a peak in flash flood fatalities shortly after dusk that was attributed to the motorists’ lack of visibility of the floodwaters.
- As revealed in analyzing flood victims’ profiles with respect to catchment size and response time [Ruin et al., 2008], the scales of flash flood impacts interfere with the daily routines of people, potentially transforming daily activities and individual space-time paths into dangerous circumstances. For instance, fast-reacting catchments have shown to preferably affect middle-aged male drivers who are generally not considered as part of the vulnerable population. Understanding whose routines are the most at risk and why is key in forecasting human impacts from flash flood events.

3.2 Social processes and the embedded vulnerability factors

Based on prior studies of flood fatalities and social vulnerability to flooding and natural hazards in general, this section provides an outline of the main factors that possibly make a person (or household) vulnerable to flash flood events (Table II.1). The term “factor” is used here to qualitatively describe the underlying reasons leading to losses during a flash flood event. Some of the cited studies do not apply specifically to flash flood events. The primary factors identified through a literature review are separated in four major categories based on the nature of the social process to which they are related: 1) Land Use, 2) Risk Governance, 3) Individuals’ Status, and 4) Cognition processes.

Here, “Land Use” processes refer to the management of the natural environment to become built or semi-natural habitats where specific human activities may be located. The characteristics of this man-made environment may also affect the safety of people and their emergency response in flash flood circumstances. “Risk Governance” is related to institutional policies for flood risk preparedness

and response. The third and fourth categories refer to social and mental processes at the individual or household level. “Individuals’ Status” describes their position in relation to others in regard to social or professional standing. The socio-demographic profile of people (e.g., age, gender, profession, housing ownership, family ties, health) is a primary factor that shapes the everyday life priorities and constrains and defines individuals’ predisposition to crisis situations. Lastly, “Cognition”, as a mental process of understanding through experience, thoughts and sensation, forms the conscious and unconscious mechanisms of individuals.

Table II.1 – Social vulnerability factors.

Social sub-process	Vulnerability factor
Land use	
<i>Urban development</i>	The <i>densely built environment</i> (e.g., dense road network) that impedes evacuation and rescue processes within a flood event for example due to traffic jams [Tapsell et al., 2005; Calianno et al., 2013].
	The <i>quality of the structures</i> (e.g., poor road or building constructions) that is related to hazardous rescues or loss of life [Tapsell et al., 2005; Jonkman, 2003].
	The <i>buildings’ design</i> (e.g., number of floors or existence of roof openings) that is related to the people’s ability to escape from floodwaters [Priest et al., 2008].
<i>Building usage</i>	The <i>use of buildings</i> that determines the evacuation feasibility. For example, nursing homes, schools or hospitals constitute “special needs” places from where population evacuation or removal is problematic [Cutter et al., 2000; Vinet et al., 2012].
Risk governance	
<i>Flood risk prevention</i>	The <i>existence of official flood prevention plans and measures</i> (e.g., flood zoning) that affects risk awareness and preparedness for evacuation [Duclos et al., 1991].
<i>Official emergency management</i>	The <i>efficiency of the official emergency response and support</i> that defines the timely evacuation and rescue that are the dominant response actions during a flood event [Lindell and Perry, 1991].
<i>FF forecasting and warning</i>	The <i>existence of official and timely warnings</i> that determines the population’s ability to undertake proper protection actions [Staes et al., 1994].
	The <i>dissemination capability</i> that plays a significant role on informing people and also making them aware of the danger [Sharif et al., 2012].

Table II.1 Continued

Social sub-process	Vulnerability factor
Individual's status	The <i>quality of the warning</i> that affects people's trust of the warnings and subsequent protection actions during the flash flood [Vinet et al., 2012].
<i>Socioeconomical dependencies</i>	The <i>dependency on others</i> to perform self-protective actions such as evacuation [Clark et al., 1998; Cutter et al., 2003; Chakraborty et al., 2005; Azar and Rain, 2007].
	The <i>need for care-giving</i> (i.e., supply of assistance to others) that reduces the ability of timely self-protection and/or evacuation [Fothergill, 1998; Cutter et al., 2000; Cutter et al., 2003; Wisner et al., 2004].
	The <i>linguistic skills</i> that affect the ability to receive and understand warnings and emergency advice [Fekete, 2010; Wilhelmi and Morss, 2013].
	The <i>financial ability</i> to have access to resources permitting preparedness measures and evacuation or rescue means [Enarson, 2007; Fekete, 2010; Wilhelmi and Morss, 2013].
<i>Daily routine</i>	The <i>daily mobility</i> related to professional activity that creates differences in population density across space and time (i.e., different distribution of exposure due to movements from residential to commercial, industrial or service-related areas and vice versa) [Belmonte et al., 2011].
	The <i>flexibility to reschedule the daily life work activities</i> [Golden, 2001] that affects the decision of people to postpone their self-protection and/or drive under bad weather conditions [Ruin, 2010].
<i>Social capital</i>	The <i>social ties</i> that increase the chance of receiving warnings, safety advice and external help from relatives, friends or acquaintances during the emergency phase of the flood event [Duclos et al., 1991; Wilhelmi and Morss, 2013].
<i>Physical conditions</i>	The <i>physical strength and health</i> that define the ability to stand or move through flood waters in order to escape from the flooded area and prevent injuries or drowning [Blaikie et al., 1994; Tobin, 1997; Clark et al., 1998; King, MacGregor, et al., 2000; Jonkman et al., 2002; Tapsell et al., 2002; Wu et al., 2002; Haki, 2003; Chakraborty et al., 2005; Azar and Rain, 2007; Cutter et al., 2003; McGuire et al., 2007; Müller et al., 2011; Vinet et al., 2012].

Table II.1 Continued

Social sub-process	Vulnerability factor
Cognition	
<i>Risk perception</i>	The <i>risk awareness</i> built over the long term from education, communication and experience influences people's decision-making related to flood risk (from preparedness to post-event response) [Burton et al., 1978; Fischhoff et al., 1978; Blaikie et al., 1994; Slovic, 2000; Ruin, 2007; Ashley and Ashley, 2008; Carroll et al., 2009; Fekete, 2010].
<i>Cognitive mapping</i>	The <i>familiarity with the environment</i> related to risky behavior such as driving into flash flood waters that exacerbates drowning potentiality, especially after dusk [Montz and Gruntfest, 2002; Ruin and Lutoff, 2004; Ruin et al., 2007; Ashley and Ashley, 2008; Maples and Tiefenbacher, 2009; Ryan and Hanes, 2009; Vinet et al., 2012; Špitalar et al., 2014].
	The <i>emotional attachment</i> to locale and belongings such as a house or personal items that affect the willingness to evacuate [Peacock et al., 1997a; King, MacGregor, et al., 2000; Jonkman and Kelman, 2005; Carroll et al., 2009].

4 Conceptual model for the assessment of vulnerability to flash flood

4.1 Towards a definition of dynamic vulnerability to flash flooding

Despite the variety of definitions depending on the specific scientific field [Birkmann, 2006b], vulnerability generally encompasses the nature and level of exposure of a system to an undesirable and/or unexpected change (i.e., a natural hazard). Sensitivity defines the level of effects on the exposed system, and resilience explains the capacity of the system to adapt and/or cope with the changes [Turner et al., 2003; Adger, 2006]. By adding a dynamic dimension to vulnerability, we intend to account for the evolution and interactions in space and time between the natural hazard and social characteristics and processes presented in section 2.1 and section 2.2, respectively. Dynamic vulnerability encompasses the complex links between social and natural processes as well as their interactions that make people and property susceptible to harm when they are exposed to a specific flash flood event. Three components of dynamic vulnerability are proposed in Table II.2.

Although discussed in the literature [Adger, 2006; Turner et al., 2003], the dynamic character of vulnerability is still not represented in applications to flooding [Wu et al., 2002; Fekete, 2009; Müller et al., 2011; Wilhelmi and Morss, 2013]. This means that vulnerability is considered as a

Table II.2 – Definitions of the three dynamic vulnerability components.

Vulnerability component	Definition for dynamic vulnerability
Exposure (E)	The space and time intersection between a socioecological system (e.g., people and their natural or built environment) and the occurrence of a threat of a specific nature and magnitude (e.g., flood characteristics). Unlike previous studies that view exposure as an external feature that does not constitute a component of vulnerability [Davidson and Shah, 1997], we consider exposure to be part of the vulnerability definition.
Sensitivity (S)	The preexisting and ever-evolving conditions of the exposed elements (e.g., people, buildings, roads) that influence the degree to which they might be impacted. It is defined similarly to the concept of susceptibility as an intrinsic part of vulnerability [UN/ISDR, 2009]. “Condition” refers to structural attributes of the built environment (e.g., building construction material) as well as to human characteristics (e.g., disabilities due to old age or poor health) that explain the quality and reliability of the exposed system.
Coping capacity (CC)	The short-term ability to deal with the strength of the perturbation [Smit and Wandel, 2006]. This differs from adaptive capacity that represents the longer-term ability of a system to respond to and recover from an event [Adger, 2006]. It is mostly used to characterize the capabilities of individuals and societies to deal with adverse conditions to avoid or lessen loss. Mitigation measures conducted at all levels (from individuals to institutions) are thus crucial parameters that drive the CC of population. Usually, CC refers to material resources and social capital [Hanifan, 1916] that enable people to avoid being harmed. In our study, we introduce the mental and cognitive processes of individuals and their interaction with social capital. For example, the individual’s perception of risk is a mental process that could change the decision to evacuate independently of the availability of a car.

key but static element of the system where exposure, sensitivity, and coping capacity are neither interacting with each other, nor evolving during an event or even from one specific event to the next one. However, the three components of vulnerability vary throughout the day, from one day to the other and according to the space and location under concern. The variability of exposure depends on the different occupancy of the same space as a function of time of the day (“quantity” of elements at risk). For example, more people are at work during the working hours, on the road during the rush hours, at home during the rest hours, and often at recreational places during the holidays and weekends [Belmonte et al., 2011]. The variability of sensitivity and coping capacity depends on the different contributions of the individual’s characteristics, short- to long-term priorities (e.g., individual socio-economic status and/or daily constraints), and the way they deal with the natural and social crisis circumstances. For example, there are some classes of workers who are employed by

time-sensitive businesses and may not be freely capable to adapt their scheduled activities in case of bad weather conditions. As they may be less adaptable to change their normal work-related journey in fear of losing their job, they could be considered a sensitive population when flooding conditions happen around commuting hours [Ruin, 2010].

The interaction between the flash flooding circumstances and the variable direct or indirect contributions of the embedded social vulnerability factors (see section 3.2) is the core of the developed conceptual vulnerability model presented in this study (see Figure II.2). This model integrates the primary factors that need to be considered in the vulnerability analysis to predict the level of human impacts for a specific flash flood event.

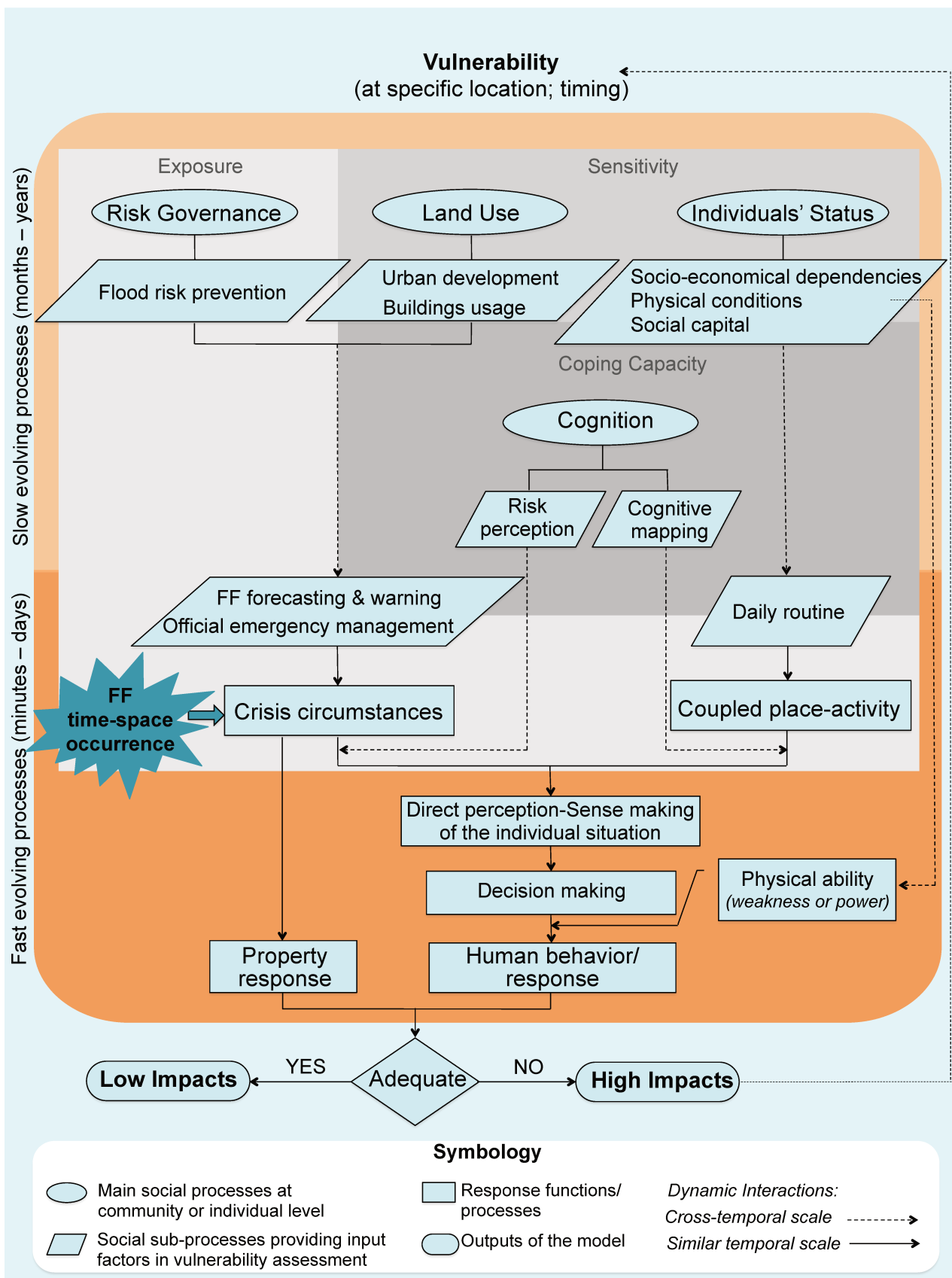


Figure II.2 – Conceptual model of dynamic social vulnerability to Flash Flood events

4.2 Model concepts and the embedded functions

What is really happening in the short duration of flash flooding? To understand the potential impacts of flash flood, we focus on the coupled physical and social circumstances that interplay during the event (i.e., fast evolving processes) and their relationship with the slow evolving processes. “Slow” processes are characterized by rates of change ranging from months to years. They influence the pre-existing conditions of the flooding event (i.e., governmental, socio-economic, physical and cognitive processes, in section 3.2). “Fast evolving processes” take place on the order of minutes to days and therefore can interact with the dynamics of the flash flood. Compared to previous vulnerability models [Turner et al., 2003], the main outcome of this new model is to take into account the individual behavioral processes in the vulnerability assessment model. The behavior of people during the flash flooding is determined by the coincidence of the flash flood event (i.e., “Crisis circumstances” in Figure II.2) with the daily schedule of individuals (i.e., “Coupled place-activity” in Figure II.2). “Crisis circumstances” refer to the flood (e.g., water depth, spatial extent of inundation, speed of the flow) and the timing of the flood onset and also to the official warning and emergency system’s response. “Crisis Circumstances” also define the property response that poses an extra threat to people’s safety (e.g., the collapse of an old building when the fast moving water impinges on it).

The “Coupled place-activity” concept refers to processes representing the daily mobility and sequence of activities of people including where they are (e.g., inside a building, driving on the road) and what they are doing (e.g., working, resting) at the different times of the day and across the days of the week. This concept evolves out of time geography that describes the sequential path (also called life path) of personal human events (with time and place as dimensions) that marks the history of a person [Gamow, 1970] within a situational context [Hägerstrand, 1970]. Hägerstrand [1970] stated that “life paths become captured within a net of constraints, some of which are imposed by physiological and physical necessities and some imposed by private and common decisions”.

In the framework of natural hazards, perception of environmental cues and warning messages strongly depend on contingent conditions (e.g., rush hours when there are errands to run and children to pick up and lots of other cars on the road, or working hours when people feel they must be at work regardless of the conditions) [Ruin, 2007; Ruin, 2010]. Likewise, the nature and dynamics of the individuals’ reactions will differ according to the location and activity they were performing when they felt the need for action, and their capability to connect with their relatives or to have social interactions allowing a group response [Gruntfest and Drainage, 1977; Mileti, 1995; Drobek, 2000; Lindell and Perry, 2003; Ruin et al., 2014]. Those contextual factors result from long-term Land use factors (i.e., the decisions taken at the national or community level concerning land use planning and management), “Risk Governance” factors (i.e., the risk prevention policy), and “Individuals’ Status” characteristics related to the individual’s position in life, attitudes, values and worldviews (Table II.1). Over the long-term, societies shape their surroundings by making strategic choices that drive individual’s decisions in terms of residential mobility related to the job market, for instance. Those slow-evolving processes condition the daily routine of individuals and the way they deal with the range of perturbations that may affect their daily project or tasks when they are faced with unusual hydro-meteorological circumstances.

At this point it has to be noticed that the human behavior when faced with a life-threatening event is not a simply defined process related directly to the crisis and the coupled place-activity processes. Intermediate cognitive processes play a fundamental role in the choice of behavior (Figure II.3).

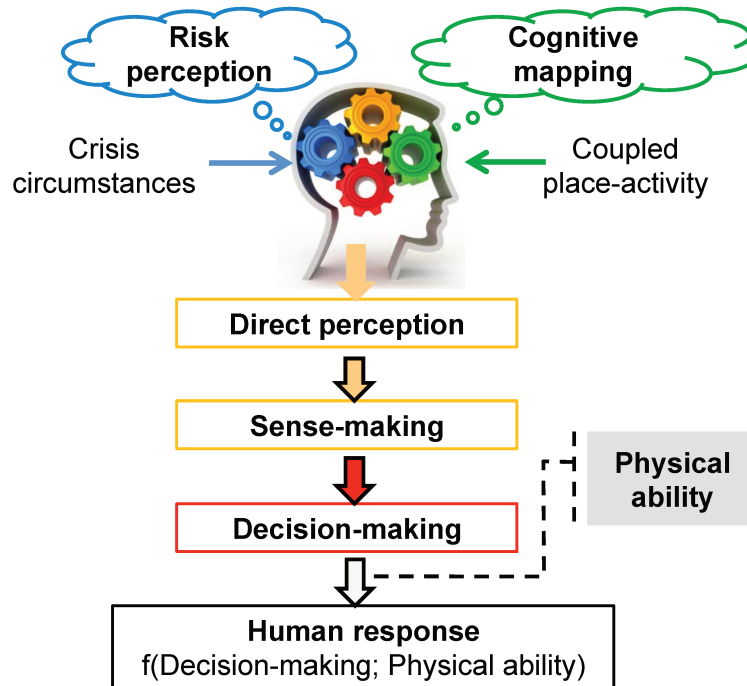


Figure II.3 – Human behavior as a function of cognitive processes and contingent circumstances.

The cognitive processes described in section 3.2 interact with the “Crisis Circumstances” and the “Coupled place-activity” processes and shape the individuals’ instantaneous perception. Risk perception impacts each individual’s subjective judgment about the severity of the flood risk. For example, a person who has previous flooding experience might perceive the danger differently than someone who has never faced such an event. Thus, the experienced people have a better chance to notice, understand and interpret the actual circumstances according to the mental model (cognitive map) that organizes their stored knowledge and beliefs.

A direct perception (through the 5 senses) can then be translated into the final sense of a situation to give (or not) an alert signal referring to higher or lower sense of danger. In fact, cognitive mapping (i.e., mental maps or schemes that organize the information on different subjects in everyone’s brain) is a factor that not only contributes to the understanding of the situation but also to the behavioral response. Among the stored information, spatial attributes or relationships between places are important elements that will partly drive people’s spatial behavior and the way they will react to perturbations. For instance, when a perturbation is faced on daily routes, the familiarity with the place may help to find an alternative and safer route. Nevertheless, when on the way home other emotional concerns like attachment to the place may intervene by lowering the perceived distance and/or difficulty to use this usual route, even when submerged, to get home. At the coarser scale (e.g., county, census tract, block group scale), variables extracted from the census data can be related to such processes. For instance, length of residence in the same area could be used as a proxy for familiarity with a place or with the type of extreme weather events that the area is prone to. Similarly, home or business ownership and family status (need for care-giving) can also inform us on

the personal concerns that may hinder the perception of danger for oneself when all the attention is dedicated to family safety or securing their belongings.

These interactions are very important because they define the decisions and actual actions taken by the people during the flood event. Decision-making is the transition from the sense that people make of their situation to a course of action among several alternative possibilities. However, the final reaction (i.e., response) of people during a hypothetical flash flood event is not only determined by the decision that they have made according to all the pre-mentioned functions but also by the physical ability that they have to implement their decisions (e.g., the physical ability to move upstairs while being in a multi-storey building).

5 Summary and Conclusions

This chapter argues that human impacts related to flash flood events present high variability and diversity from place to place not only due to differences in hydro-meteorological circumstances but also due to the space-time variability of people's exposure and capacity to react. Rapidness of flash flooding (short time between the rainfall and surface hydrological response) lessens the available time for effective warning and protective actions revealing special forms of a population's vulnerability. The occurrence of this type of flood in small catchments (a few km²) with short duration (minutes to hours) interacts with the spatial distribution and temporality of social vulnerability patterns.

In this study, prior flood-related human impact and vulnerability studies were reviewed to explore the factors that influence how individuals or households experience flash flooding. It is highlighted that since people and their activities are rarely static but move across space at different time intervals, the dominant vulnerability factors also change correspondingly. Human impacts depend on the intersection of the timing and the location of the flash flood event with the social profile and activity of the exposed people. For instance, flash flood occurrence during darkness inhibits rescue operations and safe driving due to limited visibility. Commonly, nighttime hours are associated with rest hours when people are at home, but in winter when days are short, dusk comes early and may interact with rush hours to exacerbate the surprising character of flash flood events. This, in turn, lessens the capacity of people to make sense of the situation and respond effectively. A conceptual vulnerability model is developed in this thesis to capture such complex interactions of the contextual vulnerability factors and promote the dynamic mapping of human vulnerability to flash flood.

We recognize that space and time-varying vulnerability computational modeling is a challenge. The data availability and the scale of application pose a critical dilemma on the methods to be chosen. Hereafter, we adopt a multi-variate modeling approach that enables the quantification of interactions and effects of multiple vulnerability variables based on a statistical analysis of impact observations [e.g., Merz et al., 2013]. Based on statistical analysis and the U.S *Storm Data* datasets, we will test the influence of the event timing, the severity of the observed/forecasted rainfall-runoff and of selected exposure, sensitivity and coping capacity variables at the county scale and daily and sub-daily time-steps. Vulnerability variables will be selected based on their representation of the factors listed in Table II.1, on the number of fatalities classified by accidents' circumstances, and victims' profile. This method, based on publicly available national datasets, can support a nation-

wide operational prediction tool for forecasters and emergency managers to target their warnings on anticipated impacts using the model combined with the forecasted timing of magnitude of the natural hazard (flash flood in this case).

CHAPTER *III*

Data collection and processing

1 Introduction

Based on the conceptual framework discussed in Chapter II, two main types of data are identified as crucial for the understanding and assessment of human risk to flash flood. The first type refers to human impacts from past flash flood events. Historic human losses can help understanding the circumstances in which people became vulnerable to flash flood hazards. Impact observations can be also the base for the development of data mining approaches for assessing human losses due to flash flooding and predict their occurrence in future events. In the U.S., observed impacts from flash flooding are officially recorded by the National Oceanic and Atmospheric Administration's (NOAA's) National Centers for Environmental Information (NCEI) *Storm Events Database* known as *Storm Data*. Although not faultless [Gall et al., 2009], *Storm Data* is the most extensive nationwide database in the U.S., recording four types of impacts (i.e., fatalities, injuries, and property and crop damages) for forty eight weather related events (e.g., Hail, Heat, Hurricane, Flood, Flash Flood, Tornado, Tsunami, Wildfire etc.)¹. From 2008 to 2010 summer months, the Severe Hazards Analysis and Verification Experiment (SHAVE) conducted at the NSSL collected flash flooding reports using public survey responses to a questionnaire in the entire U.S. [Gourley et al., 2010]. The collected information is point based referring to the residential address of the responders to the telephone surveys. Although the severity of flash flooding is described by a detailed impact classification proposed by Caliano et al. [2013], the data only focus on material damages and economic losses. Given the short time of data collection it was not possible to use it for the current study.

The second data type answers to the need for a set of variables that depict the critical hazard characteristics and the sensitive characteristics of the exposed population and built environment. Such information is not included in the existing nationwide impact datasets in the U.S. yet, and

1. Documentation of the *Storm Data* is available online at <http://www.ncdc.noaa.gov/stormevents/details.jsp?type=eventtype>.

need to be supplemented from other sources. Today there is no comprehensive catalog of proxy data derived from the quantitative analysis of human impact observations that can be used to understand and predict the vulnerability of people when facing flash flood events. The majority of flash flood applications adopts generic vulnerability indicators that do not adequately describe the vulnerability of people during the crisis phase; instead they describe the social groups that are the most fragile from an economic point of view (or as a result of other types of marginalization processes) [Wilhelmi and Morss, 2013; Karagiorgos et al., 2016]. These indicators are not sufficiently specific to deal with social and physical dynamics that interact/emerge during short-fuse and localized events like flash floods. To fill this gap, we investigate nationwide available datasets in the U.S. to quantify the main vulnerability factors related to the individuals' status, land use, risk governance and cognition processes influencing the exposure, sensitivity and coping capacity of people during flash floods, as presented in Chapter II. The indicators quantifying vulnerability and the prominent human risk related to loss of life from flash flooding are considered according to the following criteria:

- The **temporal phase of the event**: Some indicators can be indicative of vulnerability in the preparation or the recovery but not in the emergency phase of the hazard [Kuhlicke et al., 2011; Rufat et al., 2015]. As an example, gender is used as a proxy with different meanings depending on the stage of a disaster. Being female is often considered as a factor of vulnerability because they generally have lower incomes which may involve more difficulties in the recovery phase [Morrow, 1999; Cutter et al., 2003]. But during the “event” phase of flash flooding, men have been observed to adopt riskier behaviors than women by entering floodwaters, which make them more vulnerable during that phase [Ashley and Ashley, 2008; Becker et al., 2015; Diakakis and Deligiannakis, 2013; Jonkman and Kelman, 2005; Sharif et al., 2012].(MISSING: Ryan and Hanes 1995)
- The **circumstance of the life-threatening incident**: There are proxies that are specific to loss of life circumstances. For example, characteristics of buildings such as their integrity and distance to a nearby stream relates to the indoor loss of life circumstances. Other attributes such as the road network density or travel time to work are associated with daily mobility and environmental familiarity factors contributing to vehicle-related incidents [Ruin et al., 2007].
- The **interaction between the social and flood dynamics**: The rapidity and intensity of the runoff plays an important role in shaping specific life-threatening circumstances. The responses of small and flashy catchments (few square kilometers) have sufficient power to trigger loss of life among people who are not protected by permanent structures. These include mobile people (e.g., drivers, pedestrians, recreationists), campers, and residents in mobile homes. It seems logical that data depicting the flow of commuters at the time of peak runoff would be very indicative of the potentiality of vehicle-related accidents. Conversely, the location of the nighttime population is more relevant to evaluate vulnerability in cases of drowning from extended flooding in houses, when residents may get surprised in their sleep.

In the following sections, we present the available data sources and the process that has been followed to create a complete database i) to analyze and understand vulnerability situations associated with historic flash flood fatality events, and ii) to support a statistical approach for predicting human impacts emerging from the interaction of social and physical dynamics. These two goals correspond to the second and third research objectives, respectively, as presented in Table I.2.

2 Impact data

Deadly and non-deadly flash flood reports in the U.S. are obtained from the NOAA/NCEI *Storm Data* online¹. The *Storm Data* publish two complementary files each year:

- i. An *event details* file with information about the weather event and the respective event narratives.
- ii. A *fatality* file with details about each death resulting from the events.

Storm Data events are recorded based on a specified point, however for many flash floods early in the study period the point was missing from the data set and the flash flood event was considered to be “countywide”. Although after October 2006 an effort has been made to report the locations of impacted regions using bounding polygons independent of the county polygons, the accuracy of the storm based polygons is unknown. Especially, at the time of the analysis, the NWS issued some warnings about the validity of the longitude and latitude coordinates stored in the database. Therefore, to avoid spatial vagueness and inconsistencies between the *Storm Data* files, and maximize the amount of available records, the county reference is the finest spatial resolution used in this study.

The following subsections present the preparation of the *Storm Data* files for flash-flood specific analysis in the continue to this thesis. Further examination of the situations where people lost their lives, is proposed to better understand the link between certain social, geographic and hydro-meteorological parameters at stake in the occurrence of flash flood human losses in the U.S. (see Chapter IV). Moreover, the impact data are processed to be used as input to integrate extra datasets and build an impact-based modeling approach in Chapter V. Figure III.1 illustrates the general processing of the gathered data files, and the links between the impact data and the various extra datasets.

1. Digital data available at <http://www.ncdc.noaa.gov/stormevents/ftp.jsp>.

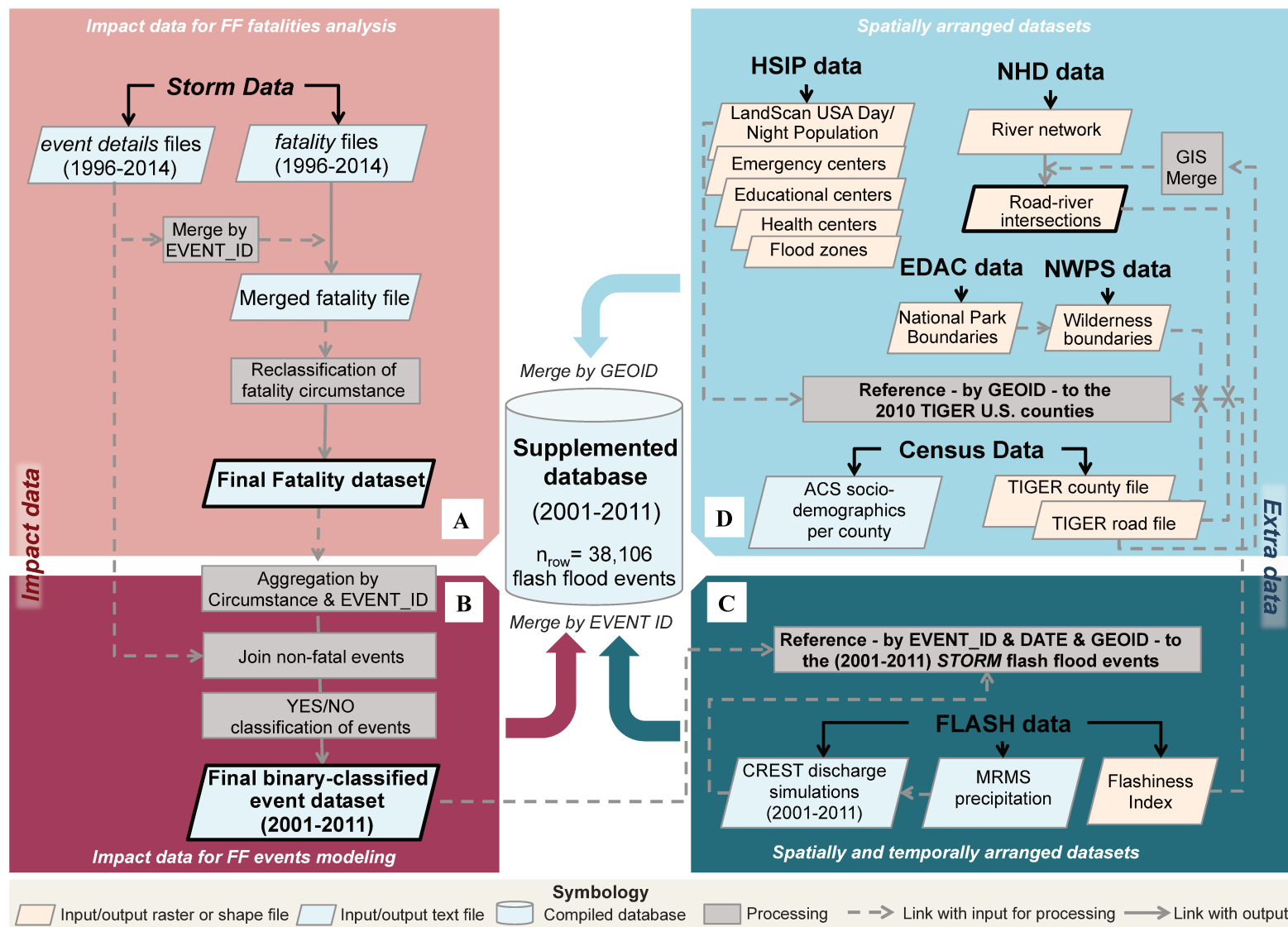


Figure III.1 – General flowchart of data processing. Parallelograms highlighted with bold black borders represent new datasets created after processing the raw data. Data sources are noted with acronyms: HSIP (Homeland Security Infrastructure Program), NHD (National Hydrography Dataset), EDAC (Earth Data Analysis Center), NWPS (National Wilderness Preservation System), Census (U.S. Census Bureau), TIGER (Topologically Integrated Geographic Encoding and Reference), FLASH (Flooded Locations and Simulated Hydrographs).

2.1 The *Storm Data* for flash flood casualties in the United States

Information on flash flooding fatalities and the related events are extracted from the 1996-2014 *Storm Data*¹. In total 1,048,683 event reports for all the weather hazard types are listed from 1996 to 2014. About 6% of those refer to flash flooding in the 19-year period. The 19-year data are selected from the 1950-2014 *Storm Data* due to accessibility in digital format provided only after 1996.

According to the NWS [2016] a recorded flash flood must have posed a potential threat to life or property and had a report of moving water with a depth greater than 0.15 m or more than 0.91 m of standing water. From 1996 to 2014, 63,176 flash flood events have been reported across the entire U.S including the non-contiguous states of Alaska, Hawaii and the territory of Puerto Rico. Forecasters at local National Weather Service (NWS) offices collect flash flood reports primarily from official sources including law enforcement and first-response authorities (e.g., Fire Department, Dept. of Highways, County official, Park/Forest Service, 911 Call Center), emergency managers, and trained spotters. Additionally, flash flood reports are gathered through other sources outside the NWS such as mass media (e.g., newspaper, broadcast or social media, amateur radio), insurance companies and members of the general public. The NWS makes an effort to use the best information available. Though, underreporting can occur especially for low-impact events (i.e., with small spatial extension or very few losses) usually not well documented by the media or public [Curran et al., 2000]. This source of inaccuracy in the *Storm Data* is discussed in previous studies [Ashley and Ashley, 2008], and assumed to be the main uncertainty source taking into account that almost 97% of the flash flood events between 1996 and 2014 are events in which less than five people died. From the 63,176 reported flash flooding events, 1.6% includes at least one human impact (i.e., direct or indirect injury or fatality). The database includes 705 flash flooding events with fatalities and 417 with injuries, yielding a total of 1,075 fatalities and 6,028 injuries.

1. While the database continues to grow, this study uses the content of March 2015, at which time the *Storm Data* contained flash flood reports from 1996 to 2014.

The 19-year dataset covers the 50 states, the District of Columbia and the territory of Puerto Rico. Only the District of Columbia (DC), and the states of Massachusetts (MA) and Rhode Islands (RI), did not experience fatal flash flood events from 1996 to 2014 (Figure III.2). In total 49 states or equivalents, and 482 out of the 3,221 counties¹ experienced fatal flash flood events. Ranking the states by frequency of fatalities shows that Texas (TX) is the first state with 19.3% of the total fatalities for the 19-yr period dataset followed by Arizona (AZ) (6.4%), Missouri (MO) (5.8%), Arkansas (AR) (5.2%), and Pennsylvania (PA) (4.6%) (see Annex A). In Texas, fatalities resulted from 136 fatal events between 1996 and 2014. The most deadly events in the study period occurred in fall and summer of 1998 in south Texas. On October of 1998, flash floods triggered by heavy rainfall killed eleven people in Bexar County. The event lasted 28 hours and the widespread flash flood became a major river flood affecting seven river basins². Almost half of the fatalities in Texas occurred during 2007, 1998, and 1997 with the most devastating years being 2007 and 1998 that account for the 20.8% and the 19.3% of the 207 fatalities in Texas, respectively.

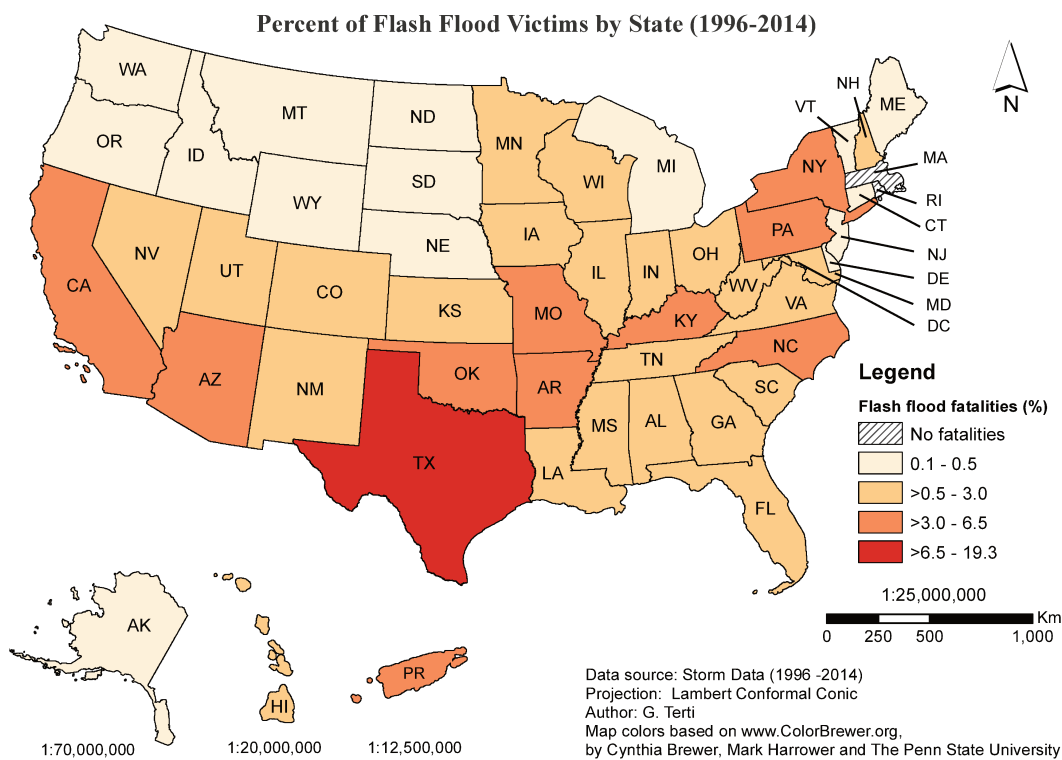


Figure III.2 – State-by-state percentage of flash flood victims. Percentages of victims are estimated to the total fatalities (n=1,075) reported in the *Storm Data* from 1996 to 2014 (see Annex A for the states abbreviations and estimates details). Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

When considering the ratio of fatalities per fatal event, other states draw the attention like Wyoming (WY), Arkansas (AR), North Carolina (NC) which may suffered less flash flood events but some with higher mortality (Figure III.3).

1. Enumeration based on the Census Bureau 2010.
 2. A full post event description is provided by the NWS Assessment available at www.nws.gov/assessments/pdfs/txflood.pdf. Retrieved on March 20, 2015.

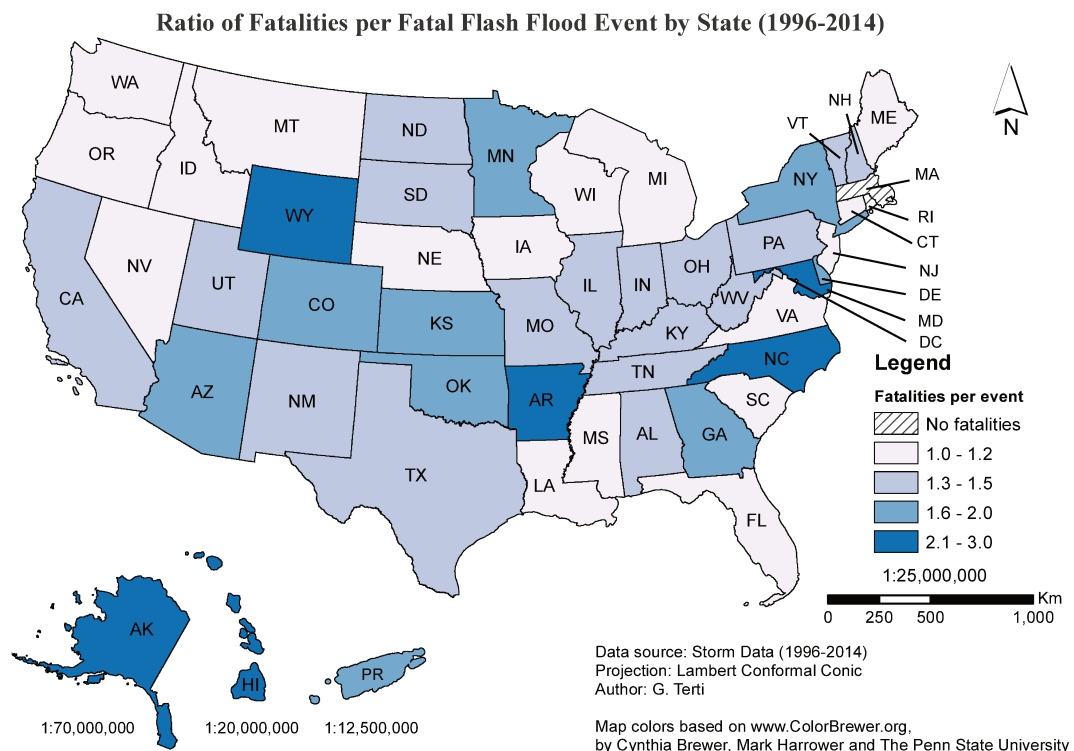


Figure III.3 – State-by-state ratio of flash flood victims per fatal flash flood event. Ratios are estimated by dividing the number of fatalities with the number of events that led to fatalities as they are reported in the *Storm Data* from 1996 to 2014 (see Annex A for the states abbreviations and estimates details). Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

The human impact database used for further analysis in this thesis is compiled based on annual *fatality* files from 1996 to 2014 with extra records and details added when available for specific flash flooding events¹. Because the date and time of the fatal incidents are not consistently determined in the original *fatality* database, these fields were supplemented based on the *event details* files to cross analyze later the fatalities with the temporal context of the flash flooding event that led to them. The event narratives provided in these files range from very coarse description of the flooding to very detailed ones with information about the victims and the specific conditions that led to casualties. For instance, a narrative may be very descriptive: “*Heavy rainfall in the Omaha metropolitan area caused areas of flash flooding, especially near the Saddle Creek and Center Streets area, near the intersections of 96th and Q Streets, 17th and Ames Avenue and 108th and Q streets. It was near that last location where a 29-year-old male apparently drowned after his car became stalled in flood waters and he was swept away in a nearby drainage ditch as he was walking for help. His body was found more than 20 miles downstream in the Missouri River after it was surmised that his body traveled from the ditch into the Big Papillion Creek and then eventually into the Missouri River. . .*”. Sometimes there were no information about the victim(s) in the available narratives of the causative flash flood event, and the specific circumstance of the fatality could not be categorized. As an example of such a narrative in the *Storm Data*, “*. . . Approximately 90 percent of the roads in the county were damaged with 65 percent of the roads had major damage. There were 25 low water crossings that were totally washed out and numerous more needed repairs. There were up to 100 homes and businesses that*

1. Event narratives for past flash flood reports available at <http://www.ncdc.noaa.gov/stormevents/eventdetails>. (box A in Figure III.1). Retrieved on April 15, 2015.

were inundated by flood waters. Numerous cars were flooded or washed away. Most of the homes flooded were along Mitchell Creek and Rubidoux River near downtown Waynesville. Over 100 people were rescued from swift and high water. There were two flash flood fatalities which occurred near downtown Waynesville”.

The focus of this thesis is constrained to fatalities due to the availability of details concerning victims (e.g., age, gender, . . .), and especially the circumstance of the fatalities. Although sometimes included as comments in the *event* narratives of *Storm Data*, details about other nonfatal impacts from flash floods, such as injuries or rescues, are not provided in a coherent database on the U.S. scale, yet. The final fatality dataset consists of 1,075 individual fatalities with the following attributes:

- i. The circumstance that the fatality occurred (reclassified as presented in the next subsection) that explains where the victim was (e.g., inside a building, driving on the road) and what the victim was doing (e.g., working, trying to reach home) at the time of the fatal incident
- ii. The age and gender of the victim (if provided)
- iii. The year and month of the fatality
- iv. The state and county that the fatality occurred within
- v. The local beginning and end time of the flash flooding event responsible for the fatality that provides the onset of the flash flooding occurrence and the duration of the event

2.2 Individual fatalities reclassification

To prepare the fatality data for further analysis, we examine the individual-by-individual fatality records for both direct (98%) and indirect (2%) losses from 1996 to 2014, and based on additional details noted in the corresponding flash flooding event narrative (when available in the *event details* files), we generalize the 13 categories of the location/activity of the perished people into six circumstances that adequately explain the framework of the majority of deaths (Table III.1).

The “in water” category was mainly distinguished from the “outside” category in the *Storm Data* publication depending on whether the victim had purposely entered flash flood waters or had fallen or swept into them accidentally [Ashley and Ashley, 2008]. In terms of vulnerability, however, these two categories are identical since they are both dominated by situations in which people underestimated the dangerousness of the flash floods in areas close to streams or rivers and walked through the floodwaters to reach some destination like home. In the “in water” category there were also many cases of children or teenagers who walked or played in the floodwaters close to streams. All these cases were therefore assigned to the “outside/open or close to streams areas” class (Table III.2).

When people entered the flood to escape, for example, from a trapped vehicle or a flooded home the deaths were reclassified as “vehicle-related” and “permanent building-related”, respectively, to better explain the original, causative circumstances. Vehicle-related circumstances in *Storm Data* represent weather-induced fatal incidents rather than traffic accidents. The “permanent home”, “permanent structure” and “business” categories were merged into the “permanent building-related” circumstance. Finally, cases for which there was no clear information for the location or the context of the fatality (although other important details of the victim’s profile was available) were registered as

Table III.1 – Categories of flash flooding fatalities’ circumstances before and after reclassification.

Category	Code
<i>Location defined in the Storm Data before reclassification</i>	
Vehicle/Towed Trailer	VE
In water	IW
Outside/open areas	OU
Permanent home	PH
Mobile/Trailer home	MH
Camping	CA
Boating	BO
Permanent structure	PS
Business	BU
Ball field	BF
Under tree	UT
Other	OT
Unknown	NA
<i>Circumstance defined in the compiled database after reclassification</i>	
Vehicle-related	VE
Outside/Open or close to streams areas-related	OU
Camping/Recreational areas-related	CA
Permanent Building-related	PB
Mobile Home-related	MH
Other/Unknown	OT

“other/unknown” to be further considered in the analysis of the available fatality and event-related variables.

The number of classified fatalities in each circumstance is presented in Table III.3. The compiled database with six circumstances of flash flood fatalities from 1996 to 2014 is now part of the unified flash flood database described in Gourley et al. [2013] and is publicly available through FLASH¹. The reclassified fatality dataset will be used for a comprehensive statistical and spatial description of the circumstances leading to flash flood fatalities presented in Chapter IV.

1. The fatality data are available online at <http://blog.nssl.noaa.gov/flash/database/>.

Table III.2 – Examples of reclassified cases in the *Storm Data* (1996-2014).

Previous location	Circumstance	Narrative
In water (IW)	Vehicle-related (VE)	<i>“A woman drowned after attempting to cross a flooded low-water crossing in her truck. The truck was found on January 15, but the woman’s body was not located until the 18th. She was found off CR 147 near the Marak community. Water was also reported over roads near Davilla”.</i>
In water (IW)	Outside/close to streams areas (OU)	<i>“One to three inches of rain fell over Shannon County. All low areas that typically flood during periods of excessive rainfall were flooded. A 14 year old male attempted to cross a flooded creek in extreme north-east Shannon County near the community of Bunker. He lost grip of a cable he was using to keep stable in the swift flowing water. He was swept downstream where he drowned”.</i>
In water (IW)	Permanent building (PB)	<i>“... Around 400 people were evacuated in Logan County, about 100 in Boone County, and about 40 residents in Wayne County. A 57 year old woman from Whitman of Logan County died when she tried to evacuate her home around 0015 EST on the 16th. She tried to wade through the flood waters from Whitman Creek, but drowned”.</i>
Unknown (NA)	Vehicle-related (VE)	<i>“A 36 year old mother and her 16 year old daughter drowned in an SUV at the underpass of Interstate 45 and Tellepsen Road. The victims were discovered several hours later as flood waters began to slowly recede when law enforcement noticed the top of their vehicle. The SUV was pulled out of the water and the two victims were discovered in the back of the vehicle”.</i>
Unknown (NA)	Outside/close to streams areas (OU)	<i>“... Around 5:45 p.m., a young girl fell into a drainage ditch and was swept into a culvert at a botanical garden in Mansfield. The mother of the young girl jumped into the culvert to rescue her and was also swept away. The young girl was swept through the culvert and was rescued. Unfortunately, the mother did not survive...”.</i>

Table III.3 – Number of reclassified cases of flash flooding fatalities' circumstance and percentages to the total 1,075 reported fatalities.

	Circumstance after reclassification:						TOTAL	TOTAL (%)
	VE	OU	CA	PB	MH	OT		
Previous location:								
VE	496	3	0	0	3	0	502	46.7
IW	99	116	26	14	1	22	278	25.9
OU	12	66	11	2	2	0	93	8.7
PH	0	2	0	37	0	0	39	3.6
MH	0	0	0	0	21	2	23	2.1
CA	0	0	30	0	0	0	30	2.8
BO	0	4	4	0	0	0	8	0.7
PS	0	0	0	5	0	0	5	0.5
BU	0	0	0	1	0	0	1	0.1
BF	1	0	0	0	0	0	1	0.1
UT	1	0	0	0	0	0	1	0.1
OT	3	19	1	0	0	5	28	2.6
NA	42	10	0	2	2	10	66	6.1
TOTAL	654	220	72	61	29	39	1,075	100
TOTAL (%)	60.8	20.5	6.7	5.7	2.7	3.6	100	–

2.3 Flash flood event database for binary classification

Part of the reclassified fatality dataset presented in section 2.2 is used to create an event-based dataset with flash flood reports with or without human losses from 2001 to 2011 (box B in Figure III.1). A total of 551 fatalities resulting from 385 flash flood events from 2001 to 2011 in the conterminous U.S. are discriminated by circumstance and aggregated by causative flash flood event to create a statistical sample for each of the circumstances. Although almost half in size from the whole dataset, the new sample presents a similar distribution of the fatality circumstances dominated by vehicle-related incidents (Figure III.4).

In a second step, the dataset including 385 fatal flash flood events from 2001 to 2011 is supplemented with non fatal flash flood events reported in the *Storm Data* during those ten years. In total 38,106 flash flood reports with or without human losses are listed in this dataset. This allows us to classify each flash flood as event with fatality or as event without fatality when examining the number of fatalities in each flash flood event for each circumstance separately. As shown in Figure III.4B, the sample size for fatal flash flood events in each circumstance ranges from 259 events that included vehicle-related victims to 12 events associated with deaths in mobile homes (see values presented on the top of the light grey bars in Figure III.4B).

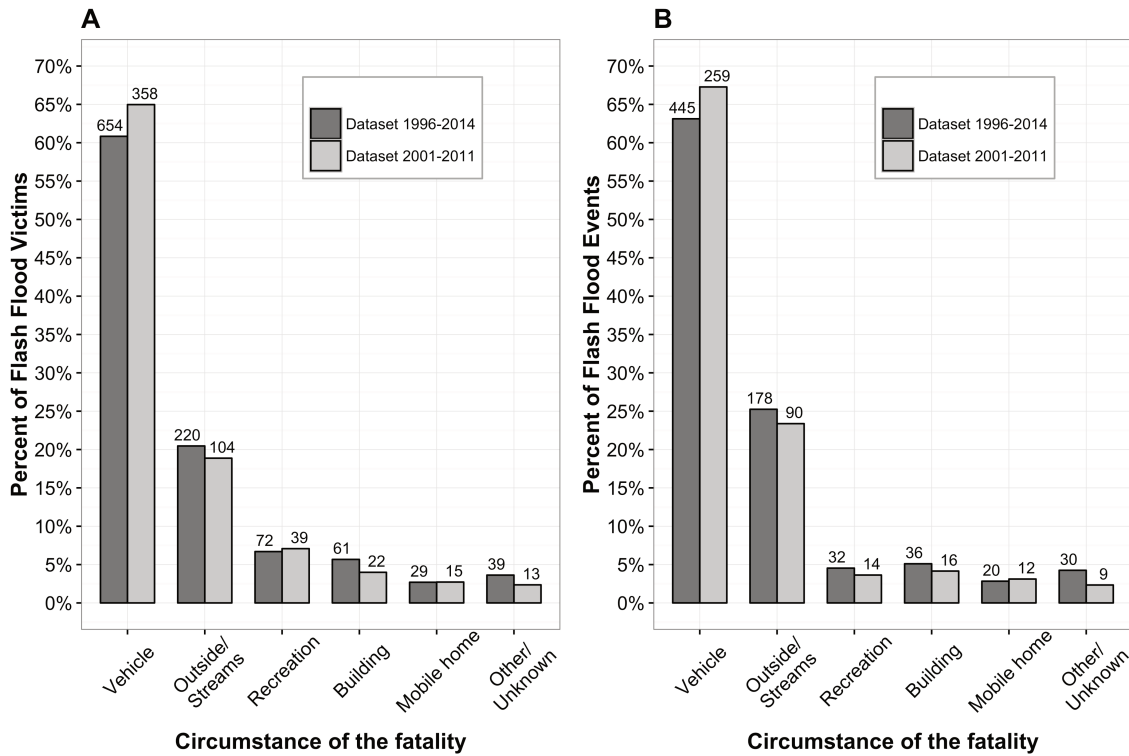


Figure III.4 – A) Percent of flash flood victims by circumstance. Percentages are estimated to the total 1,075 fatalities from 1996 to 2014 in whole U.S. and Puerto Rico (dark grey) and the 551 fatalities from 2001 to 2011 in the conterminous U.S. (light grey), respectively. The values on the top of the bars indicate the raw number of fatalities in each circumstance. B) Percent of fatal flash flood events by circumstance. The values on the top of the bars indicate the raw number of fatal flash flooding events in each circumstance. Some of the total 705 (1996-2014) and 385 (2001-2011) unique fatal events led to fatalities in more than one circumstances. Estimates are based on the 1996-2014 reclassified *Storm Data* fatality files

The final binary-classified event dataset will be the base for supplementing extra datasets as described in the next section, with the ultimate goal to be used for modeling circumstance-specific human losses in Chapter V. The 10 years are selected from the 19-year period (1996-2014) due to the restricted availability of radar-based precipitation reanalysis and hydrological outputs, used to describe the hazard dynamics of past events in this study (see section 3.1). For the same reason, the spatial reference of the database is constrained to the counties of the conterminous U.S.¹ (Figure III.5).

1. Hawaii, Alaska and Puerto Rico are excluded.

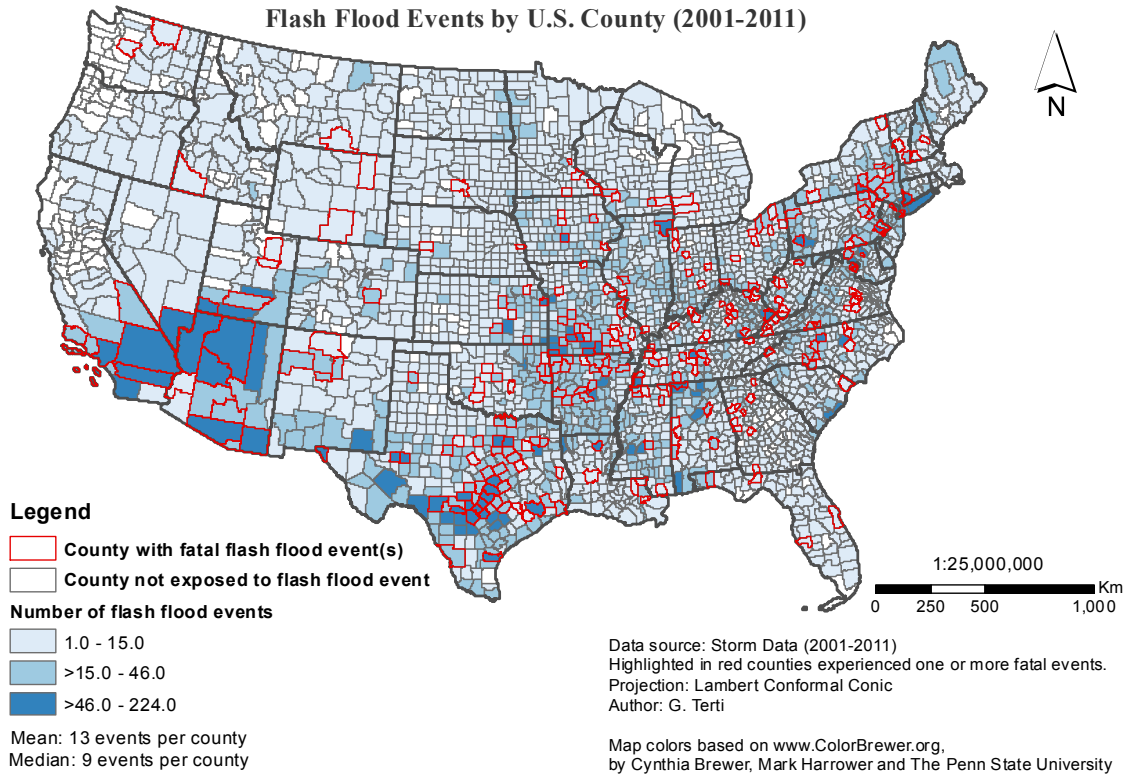


Figure III.5 – County-by-county number of flash flood events reported in the *Storm Data* from 2001 to 2011. Counties with fatal flash flood events that according to the *Storm Data* led to one or more fatalities are highlighted with red line. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

3 Supplementing flash flood reports with extra datasets for flash flood human losses prediction

The extra datasets to be supplemented to the final binary-classified event dataset described in section 2.3 can be split in two categories based on the characteristic(s) used as reference to bring all the data in common spatial and temporal (when applicable) resolution: i) the temporally and spatially arranged data that are redefined based on both the date and the county reported for each flash flood event (2001-2011) (box C in Figure III.1), and ii) the spatially arranged data that are adjusted to the county affected by each flash flood event (2001-2011) (box D in Figure III.1). The first category includes hydrometeorological variables (e.g., unit peak discharge, accumulated rainfall) from FLASH used to describe the hydrological and meteorological conditions associated with a certain flash flood occurrence in the study period. In this category, the new data are paired with the flash flood events reported from 2001 to 2011 using the corresponding event identifier (“Event ID”) as defined in the *Storm Data* (Figure III.1). The FLASH data include also estimations of the flashiness index, an indicator of flash flood severity, that is provided at every grid point over the conterminous U.S. with a spatial resolution of 1 km. This variable falls in the second category: the spatially arranged data.

In the second category, the new variables are processed and referred to the county level to be supplemented with the compiled database based on the unique geographic identifier of the reported county in each flash flood record (“GEOID”). The GEOID of each county in the flash flood event

dataset is produced by combining the state and county Federal Information Processing Standards (FIPS) codes¹ provided in the *Storm Data*. The information were screened for possible inconsistencies and the county identifiers were manually corrected when needed to ensure that the codes correspond to the same county across the files. Especially, the counties from the 2010 Topologically Integrated Geographic Encoding and Reference (TIGER) shapefile² are used as auxiliary data to spatially join all the data at the county-level³.

A bunch of spatial data are accessible through the Homeland Security Infrastructure Program (HSIP) of the U.S. Department of Homeland Security⁴ with NOAA's federal access. The 2013 HSIP Freedom Database compiles geospatially enabled data from various sources to support planning, situational awareness, threat and impact analysis, modeling emergencies, and decision making during response and recovery operations in the U.S.. Data include information about the territory such as grid-based estimations of people in day and night time, and especially the built environment (e.g., location of schools, hospitals). Also, some of the available data may refer to factors related to the geomorphology of the territory (e.g., flood hazard zones), and the emergency service (i.e., location of official emergency centers). Additional datasets used in this study are the national parks and recreational areas provided by the Earth Data Analysis Center (EDAC)⁵, and by the National Wilderness Preservation System (NWPS)⁶. Features of the natural environment such as the river network are obtained from the National Hydrography Dataset (NHD) which is available online at the NHDPlus application (Version 2.1) built by the U.S. Environmental Protection Agency (EPA) assisted by the U.S. Geological Survey, and Horizon Systems Corporation⁷. These data are combined with the 2010 TIGER road data to extract intersections related to low water crossings and bridges across the U.S. Finally, socio-economic and demographic characteristics of people and households are extracted for each county from Census surveys/programs estimates aggregated at the county-level by the U.S. Census Bureau.

The kind and sources of the data gathered to provide inputs for analysis in this thesis, as well as the risk indicators that they represent, are summarized in Table B.1 (see Annex B). The final supplemented database composes a set of 38,106 flash flood events (unique Event IDs) accompanied with more than 400 proxy variables the majority of which are obtained from the added datasets, and some of them are kept from the *Storm Data*. There are fifty-eight missing values in the dataset mainly because of lack of commuting information for the people of some counties in the Census data. All of them correspond to nonlethal flash flood events. In the following subsections the variables

1. Federal Information Processing Standards (FIPS) are publicly announced standards developed by the United States federal government for use in computer systems by non-military government agencies and government contractors (<https://www.nist.gov/itl/popular-links/federal-information-processing-standards-fips>).

2. Geospatial data maintained by the Census Bureau's Master Address File/Topologically Integrated Geographic Encoding and Reference (MAF/TIGER) database publicly available at <https://www.census.gov/geo/mapsdata/data/tiger.html>.

3. The unique geographic identifiers (GEOIDs) created by the U.S. Census Bureau for 2010 are available at <https://www.census.gov/geo/reference/codes/cou.html>.

4. Documentation of the HSIP products is available at <https://gii.dhs.gov/HIFLD/public/HSIP-Gold-Freedom-One-Page-2015.pdf>

5. Data developed at the University of New Mexico and are available at <https://catalog.data.gov/dataset/national-park-boundariesf0a4c>.

6. Data managed by the Bureau of Land Management, Fish and Wildlife Service. Available at <http://www.wilderness.net/NWPS/geography>.

7. Data available for download at http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php. For more documentation on the NHDPlusV2 see ftp://ftp.horizon-systems.com/NHDplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.

obtained from the various sources are grouped by the nature of indicators that they bring into the analysis (i.e., hazard-related, territorial, social), and presented accordingly.

3.1 Hazard-related data

Although the NCEI *Storm Data* provides a monthly publication with details on storm occurrences that were reported to the NWS, it does not accommodate a consistent description of the hydrological response from heavy rainfall. The reported storm events are listed in chronological order by state with information about the local time of occurrence, the estimated casualties and damages, and the character of the Storm (e.g., coastal flood, flash flood). The description of the events is supported by photographs, illustrations and narratives. However, the narratives for the meteorological event and the related flash flood occurrence do not always include information on the magnitude of the hydrological hazard. Thus, information that could be of interest when examining vulnerability of people to flash flood waters such as the flood water depth and velocity, is not easy to retrieve. To overcome the lack of hazard information in the flash flood reports dataset, hazard-related variables are obtained by data generated by the FLASH system [Gourley et al., 2017].

a) Magnitude and time of the flash flood event

The magnitude of each reported flash flood occurrence is described by distributed hydrological model-based discharge forecasts. Especially, the unit peak discharge (i.e., discharge normalized by the cell's upstream drainage area in $\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) was computed by running the Coupled Routing and Excess Storage (CREST) distributed hydrological model [Wang et al., 2011] with kinematic wave routing (at 0.01×0.01 degree resolution over the conterminous U.S.). The hydrological model is forced with the NSSL's MRMS 5-minute precipitation rates and provides unit peak discharge simulations on a daily scale from 2001 to 2011 (for more information on the hydrological modeling and its principles see Flamig [2016]). The maximum (as well as the mean and median) unit discharge in the county where the event occurred was extracted for each flash flood event reported in the *Storm Data* from 2001 to 2011. The timing (in local hour) of the simulated daily unit peak discharge assigned to each reported flash flood occurrence was also registered in the database. CREST simulations are also used to provide hydrological data for a series of flash flood events in May 2015 which are outside the time window of the reported flash floods used for model building, and they can therefore provide for a case study to visualize and evaluate the model predictions (Chapter V). Figure III.6 shows an example of simulated daily unit discharge assigned to each county for May 26 2015. Higher values are concentrated in southeastern Texas which experienced major flash flooding on that day.

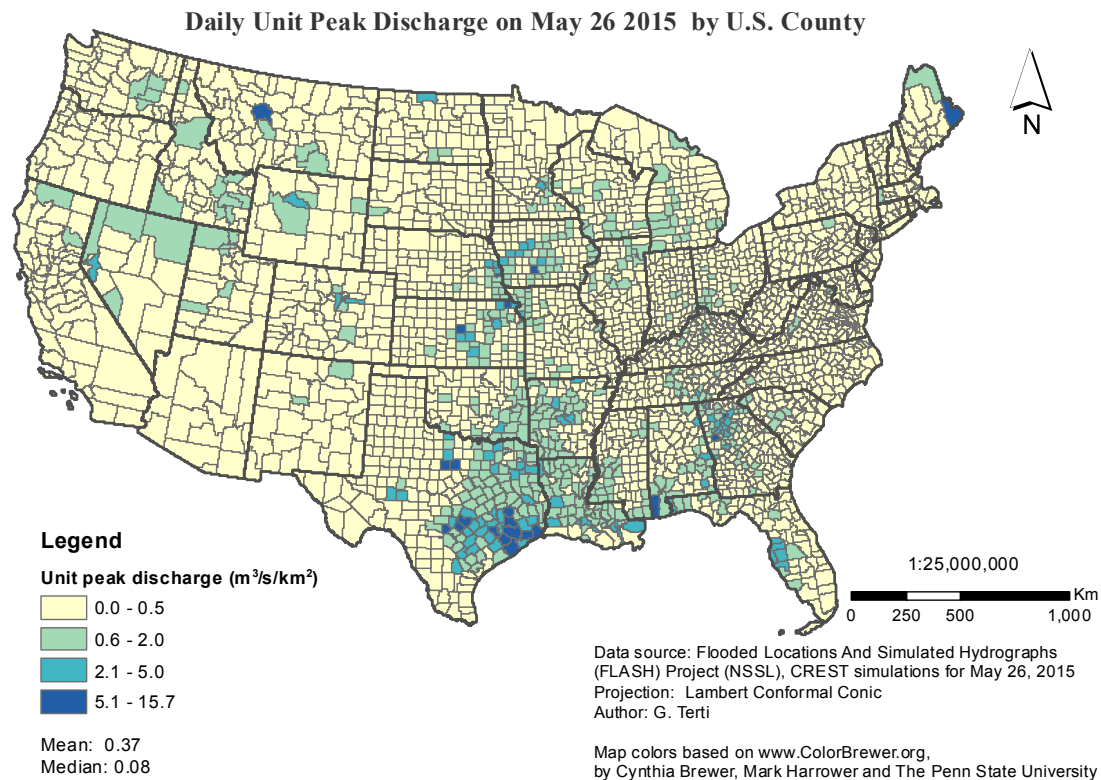


Figure III.6 – County-by-county distribution of the simulated daily unit peak discharge ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) for May 26 2015 in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

b) Duration of the flash flood event

The only hazard information that we are able to obtain directly from the *Storm Data* is the duration of the flash flood event, estimated as the difference between the reported local beginning and end time of the event. Although flash flood duration can be revealing for the dynamic of flash flooding as it relates to social impacts, this variable is not applicable in a forecasting mode since it can not be assessed prior to a flash flood occurrence.

c) Magnitude and duration of the rainfall event

The MRMS project started by the National Oceanic and Atmospheric Administration's (NOAA's) NSSL improving largely radar based quantitative precipitation estimates by generating frequent Quantitative Precipitation Estimation (QPE) updates without human intervention available every five minutes [Zhang et al., 2016]. Compared to other rain rate algorithms, the main advantage of MRMS is the ability to determine on a grid cell basis (0.01 degree) which rainfall rate - reflectivity (i.e., R-Z) relationship is the most applicable. Zhang et al. [2011] describes the decision tree classification that the system operates to choose among snow, hail, warm rain, convective rain, or stratiform rain based on 3D reflectivity data, surface temperature, surface wet bulb temperature, the column vertically integrated liquid density, and if a warm rain process was identified in the vertical profiles of reflectivity (VPR). To enhance the hazard description in our dataset, the accumulated precipitation (mm) and the duration of precipitation (hours) are aggregated from the high resolution

MRMS precipitation estimates (1-km every 5 minutes) at the county level based on the county and the date in which a specific flash flood observation was reported.

d) Flash flood severity

Except for the hydrometeorological facts, the severity of the hydrological response depends also on basin characteristics driven by topological and geomorphological parameters. Flash flood-prone locations in the U.S. have been recently studied within the FLASH system to support regional and community planning and mitigation. In this objective, a new variable called “flashiness” is defined by Saharia et al. [2016] as the difference between the peak discharge and action stage discharge¹ divided by flood rise time and basin area. Thus, high flashiness depicts basins that have high conditional probability of having a large-magnitude discharge in a short period of time. Flashiness is estimated based on streamflow observations for 70,596 flooding events provided by the U.S. Geological Survey (USGS) for 1,649 gauged locations (basins), and is used after to identify severe flash flooding in ungauged basins based on spatially distributed variables describing basin topography, hydroclimatology, geology, and geomorphology (see Table 1 in Saharia et al. [2016]). The flashiness variable is scaled between 0 and 1 for a given basin (estimated as the median of all the event-level flashiness values for a basin) and presented as point data over the conterminous U.S. Because our spatial resolution is the county, the original flashiness point data are converted to a 1-km raster². The mean in each U.S. county is then calculated by applying zonal statistics to the 1-km float flashiness raster based on the county administrative boundaries. By doing so, we assume that the mean flashiness in the county is an indication of fast rise times depicting counties where the anticipation time for protective actions may be more limited than others in a future flash flood event. According to Figure III.7, the estimations highlight flash flood hotspots such as the: i) West coast, ii) southeast Arizona, iii) Front range along New Mexico and Colorado, iv) Flash Flood Alley in Texas, v) Missouri Valley region, and vi) Appalachians extending into the more populated regions of the Northeast [Saharia et al., 2016].

1. Action stage is defined as the stage at which NWS forecasters take mitigation action for possible significant hydrological activity, and it is usually associated with bankful conditions.

2. The ArcGIS geographic information system (GIS) is used for this analysis.

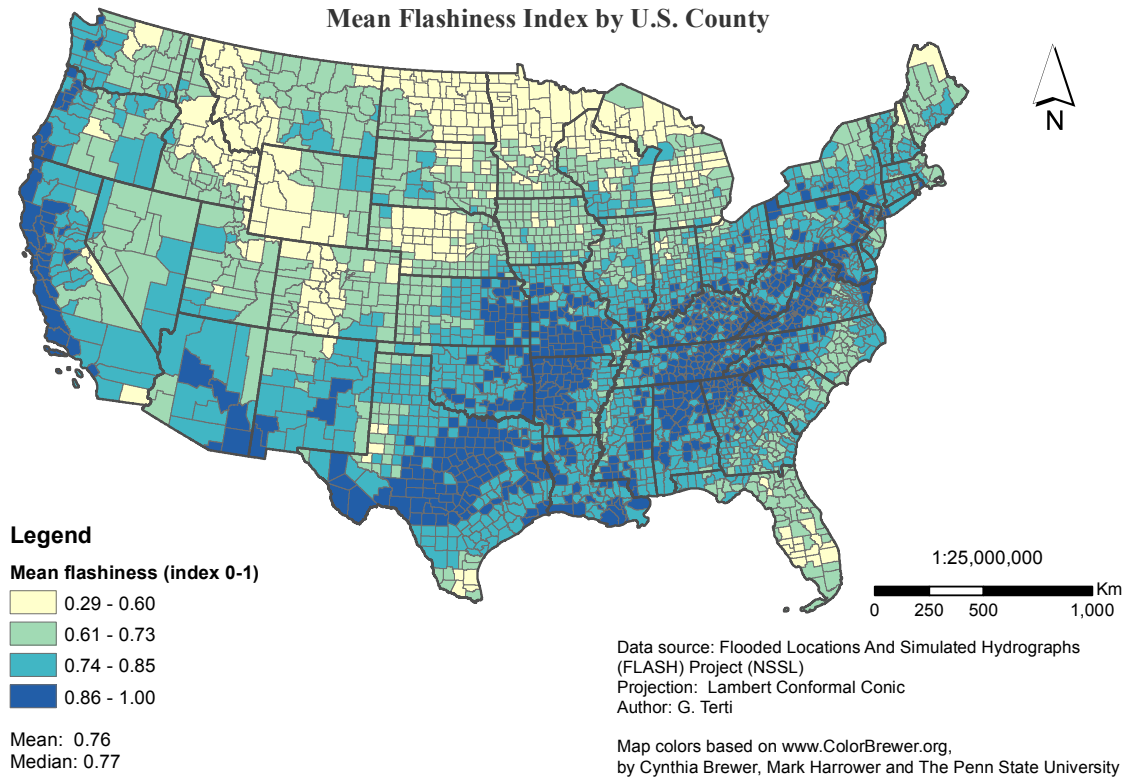


Figure III.7 – County-by-county distribution of the calculated mean flashiness in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

e) Flood hazard areas

Flood zones in the U.S. are also identified by the Federal Emergency Management Agency’s (FEMA’s) National Flood Insurance Program (NFIP) characterizing land areas in terms of their flood hazard risk. Flood maps, known officially as Flood Insurance Rate Maps (FIRMs), show areas of “high” and “moderate-to-low” flood risk with the aim to reduce loss of life and damages due to flooding. The National Flood Hazard Layer (NFHL) is the digital database that contains the flood hazard mapping data. The region of interest is extracted from the geodatabase and is converted in polygon shapefiles. Both the high and moderate-to-low flood risk areas (km²) and their percentage to the total county area are calculated for each U.S. county after dissolving the flood hazard areas geodatabase based on the county boundaries¹. In general, bigger highly flood-prone areas are concentrated in the counties of south U.S., and especially in the West coast, Arizona, and southeastern Texas (Figure III.8). Compared to the flashiness index, major flood-prone areas are also estimated in Florida and eastern coastal states where North Atlantic tropical cyclones are responsible for large flooding [Villarini et al., 2014]. This difference indicates the role of orography in the generation of large unit discharges favorable for flash flooding, when combined with highly moist air [Konrad, 2001].

1. The 2010 TIGER counties were firstly clipped with the boundaries of the conterminous U.S. based on the National Atlas states shapefile, and then the land areas provided in the TIGER shapefiles and the flood hazard areas provided in the NFHL were recalculated.

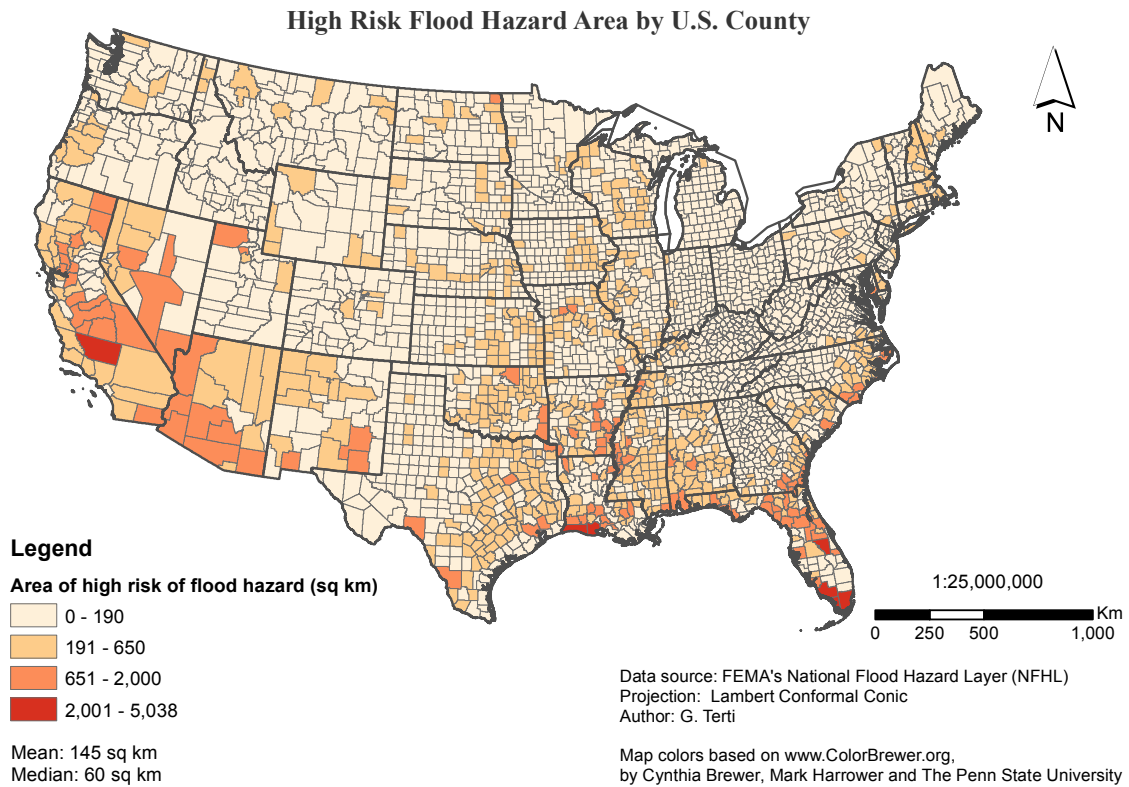


Figure III.8 – County-by-county distribution of the calculated high flood hazard risk area (km²) in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

3.2 Territorial data

Critical features that may be related to the exposure of people, and their capacity to respond to flash flood occurrences in certain circumstances in the exposed territory (e.g., road flash flooding, buildings flooding), include roads, recreational areas, and critical buildings such as schools and hospitals among others. In this section available geospatial data are explored to quantify the land use and risk governance vulnerability processes described in Chapter II (Table II.1).

a) Local emergency services

We hypothesize that the existence of local emergency services contribute to more timely and efficient response leading to successful evacuation and rescues from flash flooding. Point data illustrating the Local Emergency Operation Centers (EOCs) in the U.S. are produced by the TechniGraphics, Inc. multi-national company (TGS) (currently named as Consolidated Analysis Centers, Inc. (CACI))¹. The original point shapefile is spatially joined with the county polygons, and the number of EOCs is counted for every county². At the county-level the number of EOCs present low variability across the conterminous U.S. with a mean of 2 centers per county (Figure III.9). The number of EOCs increases for counties in the South California and Central Oklahoma. A cluster of EOCs well-above the mean

1. Information about the services of CACI are available at <http://www.caci.com/fcc/Geospatial/capabilities.shtml>.

2. Analysis based on the Summary Statistics tool of ArcGIS software.

is concentrated in north-west corner and especially, in Vermont, New Hampshire, Massachusetts, and Connecticut.

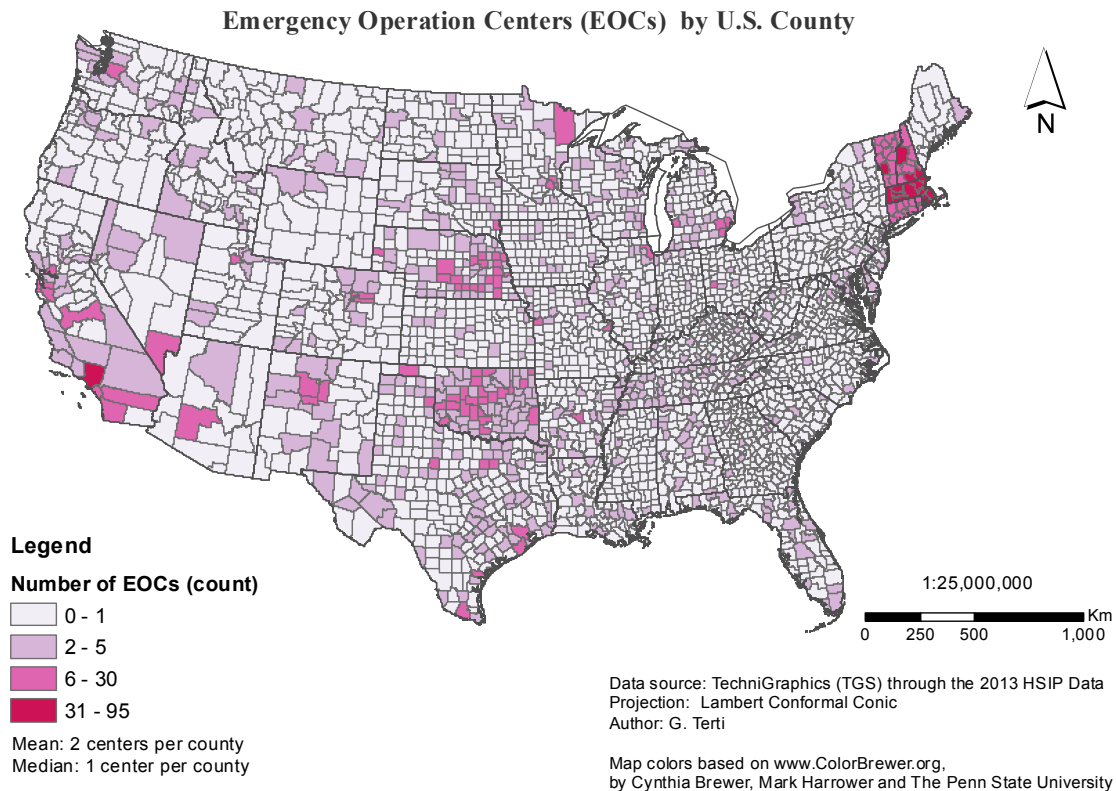


Figure III.9 – County-by-county distribution of the number of emergency operation centers in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

b) Distribution of human population

The dynamic exposure of people in space and time is a key factor for the evolution of human risk to flash flooding. The LandScan USATM Raster Datasets developed by the Oak Ridge National Laboratory (ORNL) provide Day/Night Population estimates¹. LandScan USA is a multi-dimensional dasymetric modeling approach, which allowed the creation of a very high-resolution population distribution data both over space and time. At a spatial resolution of 3 arc seconds (90 m), the database contains both nighttime residential as well as baseline daytime population distribution that incorporates movement of workers and students [Bhaduri et al., 2007]. The county-level daytime and nighttime population is estimated based on the county polygons. The daytime and nighttime population density (people/km²) is then calculated for each U.S. county by dividing the daytime and nighttime population by the estimated county land area, respectively, to be related with the exposure of people at those periods.

1. Documentation of the LandScan data is available at <https://www.ornl.gov/ornl/careerssci/landscan/>.

c) Road and river network

Searching for features related to road flooding circumstances, the TIGER/Line shapefiles contain linear features such as roads, railroads, and hydrological network¹. The All Roads Shapefiles downloaded for the U.S. counties in this study include all the primary, secondary, local neighborhood, and rural roads, and also smaller streets such as city streets, vehicular trails (4wd), ramps, service drives, alleys, parking lot roads, private roads for service vehicles². These data were used to calculate the total length of road network (km) in every county. The county-level road density (km/km²) is also calculated by dividing the total road length in the county (km) with the estimated land area of the county (km²) to quantify the “densely built environment” factor in circumstances related to road flooding. In addition to that, we attempt to identify other critical features for flash flooding such as low-water crossings and bridges that are associated with the majority of vehicle-related fatalities from floods in the U.S. [Kellar and Schmidlin, 2012]. To do so, we extract the linear river network provided by the NHD. Then, at the U.S. scale, the road shapefile is merged with the NHD hydrographic network shapefile, and the road-river intersection points are identified. The number of the road-river intersections are finally calculated for each county. A swarm of crossings are concentrated in the West, but many other counties across the coterminous U.S. are estimated as well-above the county average (Figure III.10).

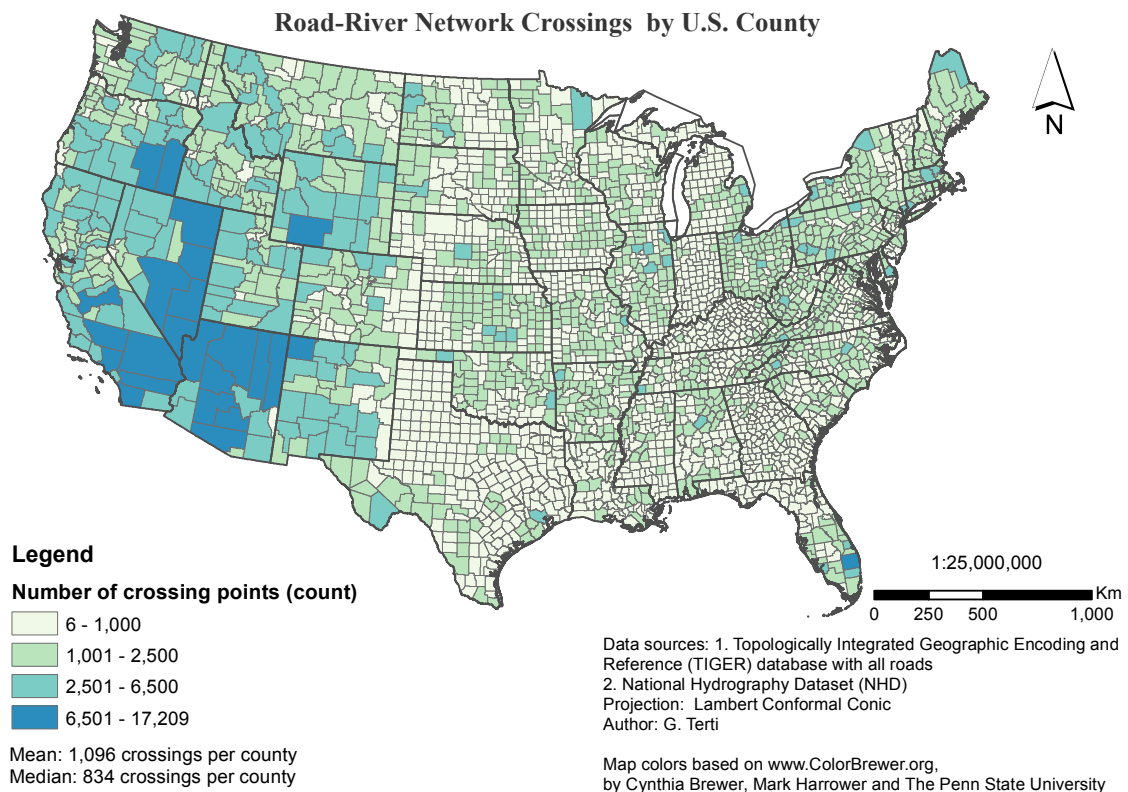


Figure III.10 – County-by-county distribution of the calculated number of road-river intersections in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

1. For technical documentation of the 2010 TIGER/Line Shapefiles see <http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2010/TGRSHP10SF1.pdf>.

2. Geospatial data available at <https://www.census.gov/geo/maps-data/data/tiger-line.html>.

d) Recreational areas

As we have discussed earlier the type of place matters in the way that people are constrained in their response to the flooding situation. In recreational areas, and especially when performing outdoor activities in the wild, people may have less access to information; a fact that might hinder self-protection actions. To capture both national and state-defined recreational areas in the U.S., the National Park Boundaries shapefiles developed by the EDAC are merged with the Wilderness Boundary Data from the NWPS. Additional data are provided by the Environmental Systems Research Institute's (ESRI's) Parks¹. The merged park shapefiles are intersected with the counties shapefile, and are then dissolved by county. The total recreational area (km²) per county is calculated, and then divided with the total estimated county land area to produce percentages of recreational area to the total area in the county for all the counties in the conterminous U.S.

e) Special needs buildings

Special needs buildings that may require special attention for massive evacuation during a flash flood event include education and health centers among others. The 2013 HSIP Freedom Data provide geodatabases with the geolocation of education and health centers in the U.S., originally developed by the ORNL. Educational features include daycare centers, private and public schools, colleges and universities, and multiple campus college institutions. The data point files are spatially joined with the counties and the number of centers are counted for every county². Similar process has been followed to calculate the number of public health centers (hospitals) in every county.

3.3 Social data

The spatial scale of the analysis and application poses constraints on the representation of some of the social vulnerability processes (especially the cognitive ones) presented in Chapter II. Especially, no large-scale survey or dataset is available to directly provide up-to-date information on the level of flash flood risk awareness or the capability of response from the exposed population at the U.S. scale. Instead, the literature based questionnaire surveys on flood risk knowledge, perception and behaviors, establish the links to socio-demographic characteristics such as age and gender [Drobot et al., 2007; Knocke and Kolivras, 2007; Franklin et al., 2014; Morss et al., 2015; Gissing et al., 2016; Lazrus et al., 2016]. Therefore, we propose to explore the suitability of publicly available census data to be considered as proxies for behavioral response in flash flood circumstances. In general, findings from flash flood or flood human-impact studies [Jonkman and Kelman, 2005; Ashley and Ashley, 2008; Jonkman et al., 2009; Maples and Tiefenbacher, 2009; Diakakis and Deligiannakis, 2013; Doocy et al., 2013; Becker et al., 2015; Sharif et al., 2014] are crosschecked with arguments from the literature on social vulnerability to flooding and natural hazards in general [Adger, 2006; Rygel et al., 2006; Cutter et al., 2003; Cutter et al., 2009; Fekete, 2010; Kuhlicke et al., 2011; Zhong et al., 2013].

1. State and local parks from the ArcGIS Maps and Data DVD available through the FLASH system.
2. Analysis based on the Summary Statistics tool of ArcGIS software.

The selection of relevant vulnerability indicators expressed through specific proxies provided by census data is a trade-off between the social vulnerability processes to be represented and the availability of data at the required spatio-temporal scale and resolution. In particular, we searched among nationwide population and household surveys conducted by the U.S. Census Bureau that provide population data at the county level (see Annex C). The criteria to choose between the available U.S. census data were related to the following requirements:

- **Spatial adequacy.** Census data should refer to geographic units such as counties to be consistent with the county-level compiled flash flood reports dataset described in section 2.3. Though, the consistency of the data at the national scale (i.e., same proxies available for all the geographic units in the U.S.) is important to allow the transferability of the current approach in future advances. For example, selected proxies could be elaborated at smaller geographic units such as census tracts and block groups¹ to better capture the local variability of vulnerability at the neighborhood level (if relevant) in future work.
- **Temporal adequacy.** Census data should provide up to date information relevant for the years of the analysis. Often data releases (e.g., every three or five years) ensure data updating in future applications.
- **Content adequacy.** Census data is desired to cover the socio-economic aspects related to the flash flood-specific vulnerability definition. For example, commuting characteristics for the population of specific geographic units is essential information for understanding and quantifying the interruption of people's intended paths due to flash flooding while they perform everyday life activities such as travelling from and to work.

Based on the pre-mentioned criteria the American Community Survey (ACS) 5-years pre-tabulated sample estimations for the period 2006-2010 was selected as the most relevant dataset for this study. The ACS pre-tabulated estimates are freely accessible for download through the American FactFinder application². For ease, in the remaining sections of this thesis, the terms ACS 5-years pre-tabulated 2006-2010 estimations, ACS 5-year estimates or simply ACS data, are used interchangeably. Several proxy variables are extracted from the ACS estimates. Below, we introduce the main attributes of social vulnerability assumptions selected to be explained by the obtained data.

a) Age

Both young and elderly may have limited physical strength to withstand floodwaters in flash flood circumstances (e.g., limited stability in moving water) and depend on others for protective actions when inside a building (e.g., evacuation) or outside (e.g., rescue from street floodwaters) [Clark et al., 1998; Coates, 1999; Morrow, 1999; Tapsell et al., 2002; Cutter et al., 2003; Ashley and Ashley, 2008; Vinet et al., 2012]. Old people may also be less informed when being alone at home or can be less willing to abandon their personal possessions [Coninx and Bachus, 2009] and/or their home [Peacock et al., 1997a] for evacuation purposes. Especially, for circumstances related to flooding along road networks, young and middle-aged active population can be considered the most vulnerable since they

1. Census Tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people. Block groups are statistical divisions of census tracts, are generally defined to contain between 600 and 3,000 people.

2. Data access at <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>.

are overrepresented in vehicle-related fatalities from past flood events [Jonkman and Kelman, 2005; Diakakis and Deligiannakis, 2013]. In fact, active population in urban areas is more likely to rely on everyday routines such as travelling to work or school by car and thus, their potentiality to meet flood disruptions is increased [Ruin et al., 2007; Shabou, 2016]. Young drivers may be also less aware of flash flood risk and more prone to risky behaviors [Drobot et al., 2007; Knocke and Kolivras, 2007; Ruin et al., 2007]. The ACS 5-year (2006-2010) estimates provide both counts of people and percents of people to the total population of the county for thirteen age classes with a 5-year step (e.g., under 5, from 5 to 9 , . . . , 85 years and over) that can be summarized in fewer classes for the purposes of the analysis. In addition to that, the median age of the residents and workers are provided for both the residential and workplace¹ county, with details on the median age of workers by means of transportation to work (e.g., car, public transportation, walked). Figure III.11 presents the distribution of median age of people using a vehicle (drove alone or carpooled) in their daily commutes to work by the county of workplace.

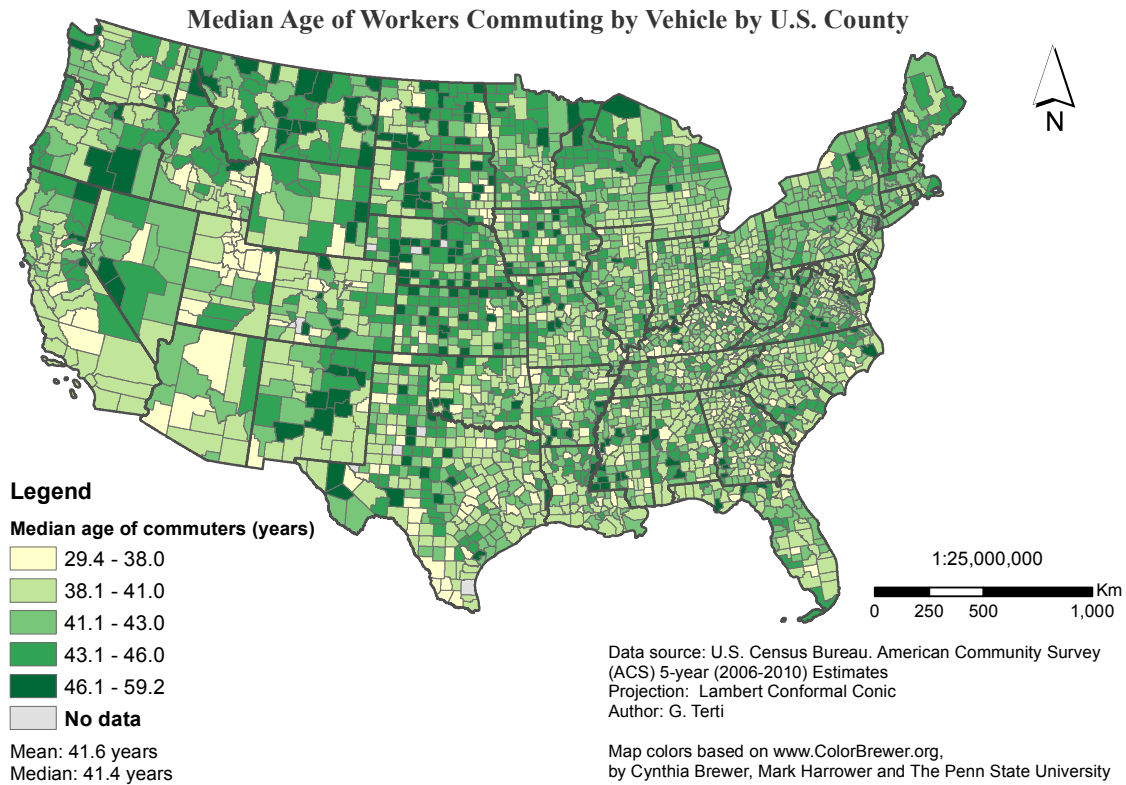


Figure III.11 – County-by-county distribution of median age of commuters traveling to work by vehicle. Data are missing for 21 counties over the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

b) Gender

Females are supposed to be more likely to hold low-status jobs and have less access to resources for evacuation due to economical constraints. Considering the “need for care-giving” vulnerability factor presented in Table II.1, women tend to ignore their safety due to their responsibility for children and elderly people in the household [Fothergill, 1998; Blaikie et al., 2014]. Thus, they may

1. The usual geographic location at which workers carried out their occupational activities during the employment status reference week. The terms worksite, workplace, and place of work are interchangeable in the ACS data.

constitute vulnerable group of people for certain circumstances such as flooding in mobile homes mainly occupied by housewives or low-income females. However, women are hypothesized to have higher risk perception and preparedness for action [Martens and Ramm, 2007]. Especially, they are considered as more likely to believe warnings and be more willing to respond [Phillips and Morrow, 2007]. In contrast, males are supposed to be more prone to risk-taking behavior than females, and especially entering floodwaters on foot or in vehicle [French et al., 1983; Coates, 1999; Ashley and Ashley, 2008]. Thus, they may be the most vulnerable when referring to flash flooding on the road network. The literature shows that men are overrepresented in vehicle-related drowning worldwide [Diakakis and Deligiannakis, 2013; Doocy et al., 2013; Jonkman and Kelman, 2005; Sharif et al., 2012]. The number of people of a specific gender, as well as the percent of males and females to the total population in the county, respectively, are included in the ACS data. The mean and median percent of males per U.S. county is 49.9% and 49.5%, respectively. Figure III.12 shows that independently of the county size, the number of males is big in highly urbanized areas as for example, in the Capital area and Houston-Galveston area in Texas. Other combined information such as sex of workers and aggregate travel time to work of workers by sex, are available in ACS detailed tables. Such information are explored for possible use in certain circumstances of flash flooding related to the daily mobility vulnerability factor (Table II.1).

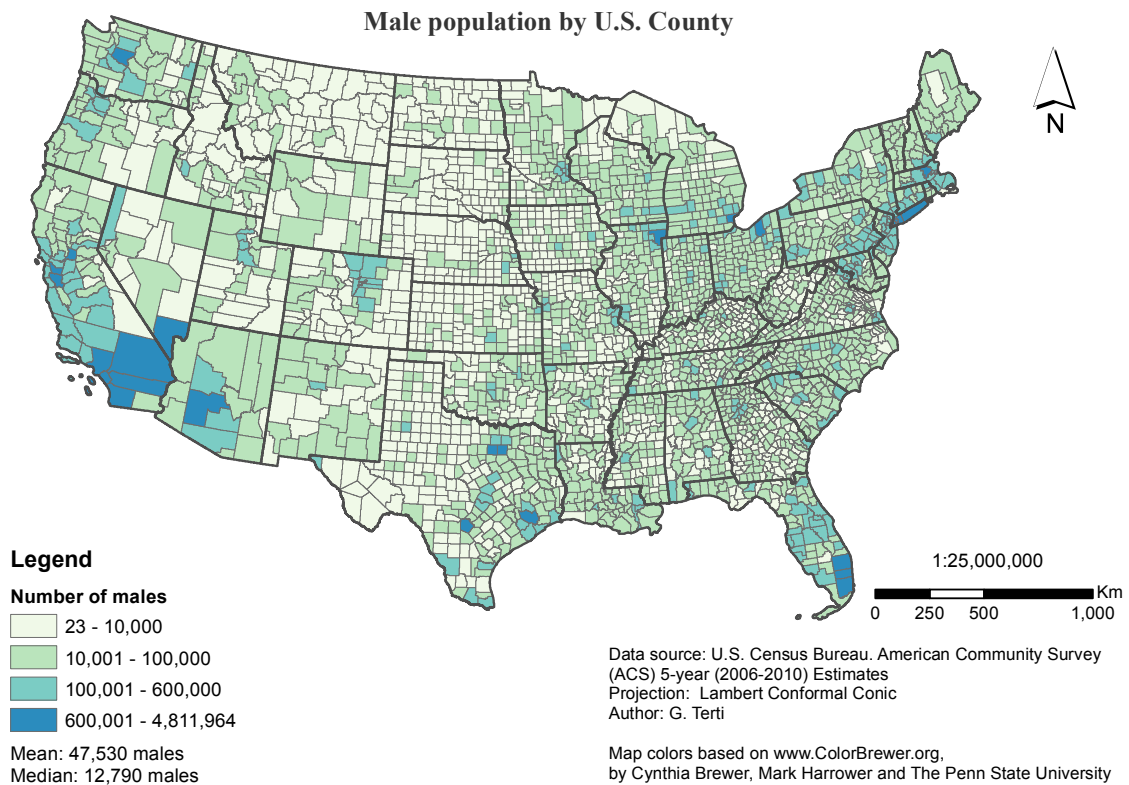


Figure III.12 – County-by-county distribution of the number of male people in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

c) Family status

People living alone are usually considered as dependent on others to evacuate if being isolated at home during the flood occurrence [Fekete, 2010]. In family households, parents are assumed as

very likely to put themselves into dangerous situations during flooding, given that they are generally expected to ignore their self-protection to protect their children [Tapsell et al., 2002]. The literature review presented by Cutter et al. [2003] suggests that especially single-parent households and large families “often have limited finances to outsource care for dependents, and thus must juggle work responsibilities and care for family members”. Thus, single parents or members of big households may have less flexibility to reschedule work-related activities and thus, face flash flood risk while travelling to work under adverse weather. Also, household members may decide to enter floodwaters to gather with and/or help the rest of the family. The ACS data profile for demographic characteristics and housing units includes information about the household type such as the number (and also percentage) of family and single-parent households, as well as the average household size in the county (Figure III.13).

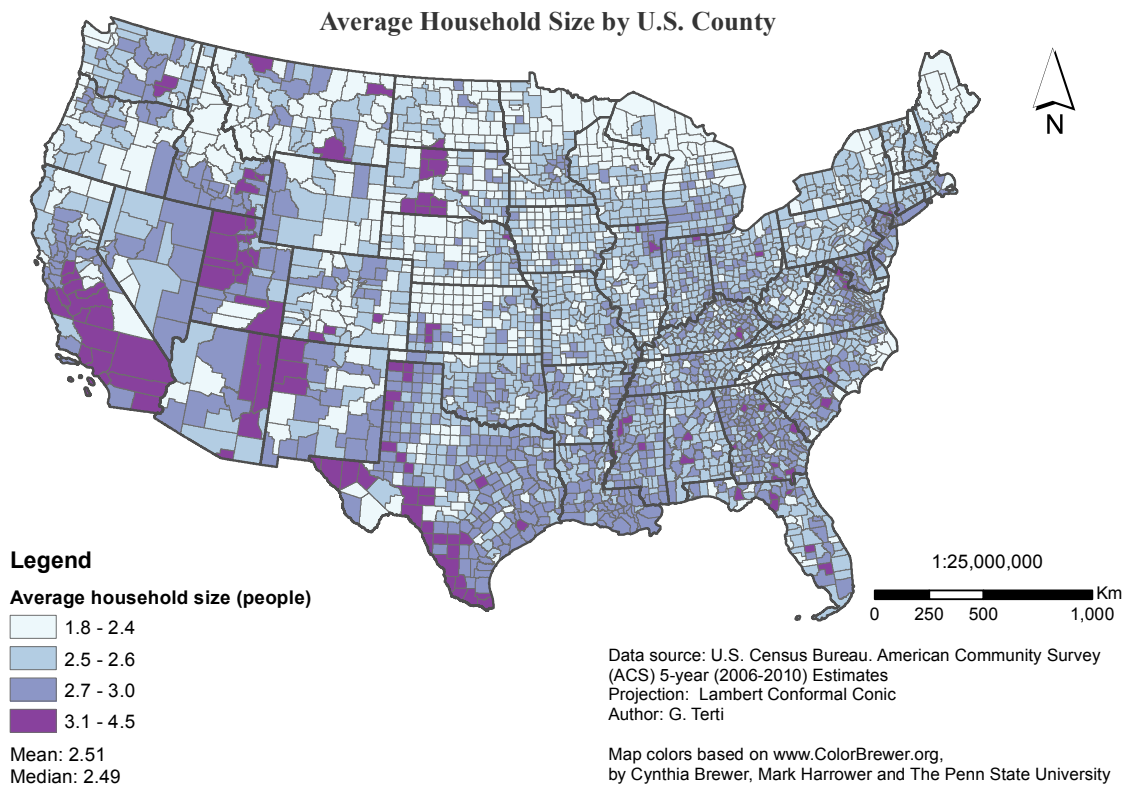


Figure III.13 – County-by-county distribution of the household size in the conterminous U.S.. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

d) Poverty status

People living in poverty conditions are considered highly vulnerable in terms of reduced economical ability to access preparedness measures and resources for evacuation in case of building flooding [Few, 2003]. The 2010 ACS 5-year estimates provide the number of people with income below the poverty level for population for whom poverty status is determined¹. The information refers to the poverty status of the last twelve months.

1. Following the Office of Management and Budget’s (OMB’s) Directive 14, the Census Bureau uses a set of money income thresholds that vary by family size and composition to determine who is in poverty. If the total income for a family or unrelated individual falls below the relevant poverty threshold, then the family (and every individual in it) or unrelated individual is considered in poverty.

e) Education

Lower education (e.g., without high school diploma) may constrain the ability to understand warnings [Cutter et al., 2003; Fekete, 2010]. In addition to that, people with less than a high school diploma are the least likely (about 17.5% in 2004) to work in occupations in which they are flexible to vary their work schedules [McMenamin, 2007]. Thus, they may decide to drive through flooded ways in their effort to deal with work-related constraints. The educational attainment for population 25 years and over in the county is included in the selected social characteristics table of the ACS data. Especially, the educational level is provided in seven classes (e.g., less than 9th grade, 9th to 12th grade without diploma, high school graduate, . . . , graduate of professional degree). The information is presented as count of people or percent of people in a certain education class to the total population 25 years and over in the county.

f) Ethnicity/citizenship

Potential cultural constraints of foreign population may hinder situational awareness related to the forthcoming weather, especially for foreigners who have low skills in the language of the warnings [Trujillo-Pagán, 2007]. This can be determinant for decisions taken by those people during flash flooding such as driving through flooded or barricaded roads. The ACS data include counts of native population, foreign born population, foreign born naturalized U.S. citizens and foreign born not U.S. citizens in the county. It is interesting that this information is discriminated by means of transportation to work (e.g., car/truck/van drove alone, car/truck/van carpooled, walked, worked at home) and thus, foreigner commuters can be tested for their relevance in certain circumstances of flash flood exposure (e.g., vehicle-related, outside).

g) Language

Households that speak English may have more chance to receive and understand national or local broadcastings and watches and thus, be efficiently warned. On the other hand, language difficulties may lead to limited or no reception of warnings and emergency advice [Fekete, 2009; Wilhelmi and Morss, 2013]. The number of people living in households that speak English only or households in which the main language is other than English, and also speak English less than “very well”, is available in the ACS data. The information refers to population five years and over in the county, and it is also provided as percentage to the total population with five or more years. Similarly to above, details about the main language spoken in the household are also available by means of transportation to work and especially, for workers in the workplace county.

h) Disability

In crisis, disabled people need additional assistance to engage in protective actions [Rygel et al., 2006]. This factor can be exacerbated when disabled people are isolated at home during flash

flooding. Information about disability of the total civilian non institutionalized population¹ are generally included in the ACS data but were not available in the 2010 5-year estimates (2006-2010). Thus, the number of disabled people, and the number of people 65 years and over with disability as well as the corresponding percentages to the total non institutionalized population in the county were extracted from the 2012 ACS 5-year estimates (2008-2012).

i) Special needs population

Similarly to the disabled people, population living in group quarters² may have high dependency on others to respond adequately to flash flood events. Especially, since this group living arrangement is owned or managed by an entity or organization, people living there are dependent on higher-level decisions and thus, are considered to have needs for special treatment during the flash flood crisis. Emergency managers need to focus on such high population concentration areas for early evacuation [Morrow, 1999; Rygel et al., 2006]. The ACS data provides the number of people living in group quarters including all people not living in housing units and thus, may be used as indicator of special needs population.

j) Home ownership

Home ownership is usually considered as an indication of more financial resources and awareness towards preparedness measures for flood hazard [Cutter et al., 2003; Fekete, 2010]. In certain cases however, such as a necessary evacuation from a flooded house, homeowners may feel more attached to their place, and so they probably appear reluctant to leaving their property. The number of people in owner and renter occupied housing units, respectively, is available in the ACS estimates.

k) Length of residence

New residents might have less experience with local floods and probably limited social capital in terms of neighborhood networks to receive information and/or external help for evacuation or rescue [Fekete, 2010]. They may have also limited knowledge of the local area and the national, regional or local warning and emergency system. However, in certain circumstances such as road flooding, longer residents probably feel familiarity with the local road network, and underestimate the flood risk while driving [Maples and Tiefenbacher, 2009; Diakakis and Deligiannakis, 2013]. The 2010 ACS estimates provide the number of owner and renter households by the year that the householder moved into the housing unit, an information that could probably explain how long the residents are in the same housing unit. This information is provided in six intervals of years (e.g., householder moved in the house in 2005 or later, in 2000 to 2004, in 1990 to 1999, . . . , in 1970 to 1979, in 1969 or earlier). The data are provided for householders that were living in another house in the same county or in

1. All U.S. civilians not residing in institutional group quarters facilities such as correctional institutions, juvenile facilities, skilled nursing facilities, and other long-term care living arrangements.

2. The Census Bureau classifies all people not living in housing units as living in group quarters. There are two types of group quarters: institutional group quarters (e.g., correctional facilities for adults, nursing homes, and hospice facilities) and noninstitutional group quarters (e.g., college/university student housing, military quarters).

a house at another county, respectively, providing some more details for the assumptions related to the familiarity with the area of residence.

l) Mobile housing structure

Mobile structures are less resistant than permanent buildings to flash flood waters increasing vulnerability for those inside. Mobile homes may reveal also economic constraints for their residents to access resources for flood preparedness and evacuation [Cutter et al., 2003; Chakraborty et al., 2005]. The number of movable housing units such as mobile homes and vans in the county, are included in the selected housing characteristics of ACS estimates.

m) Year structure built

New structures are expected to be more likely to withstand floodwaters. In respect to the “quality of the structures” vulnerability factor presented in Table II.1, old homes might be in bad condition creating hazardous situations for their residents during flooding, and especially when flash flood occurrence surprises residents during nighttime rest hours. The age of housing units can be indirectly extracted by the “year structure built” variable of the ACS data. The ACS 5-year 2006-2010 estimates include nine classes of construction years for housing units (e.g., built in 2005 or later, built from 2000 to 2004, . . . , built from 1940 to 1949, built in 1939 or earlier). The county-level count and also the percent of housing units in each class can be summarized in fewer classes relevant for further analysis (e.g., built before 1980, built after 2000).

n) Phone availability

The existence of telephone service in the household may indicate the ability to access information and/or call for help and evacuation when being blocked from floodwaters at home [Chakraborty et al., 2005]. Though, this assumption is questionable because warning messages can now be delivered directly to individuals through cell phones and personal data assistants [Phillips and Morrow, 2007]. Nowadays, many people rely on their cell phones and they do not maintain a land line. Data for individuals’ phone availability is difficult to find. The number of households without phone available in the county is the only related information at the county-level in the ACS data. The information is also given as percentage to the total households in the county.

o) Vehicle availability

The existence of vehicle(s) may facilitate evacuation and create better opportunities for rescue activities in flash flood circumstances related to buildings [Chakraborty et al., 2005]. The ACS 5-year estimates provide county-level counts and percents of households for four classes starting from households without vehicle available to households to 1, 2, or 3 and more vehicles available. The availability of vehicles in the household increases also the chance for driving in flooded roads. Other variables that might be more representative of the number of vehicles used in daily travels are explored below to be used especially for road flooding circumstances.

p) Vehicles

Personal vehicles are very much involved in flood vehicle-related deaths possibly due to drivers' confidence in the safety of their automobile or their driving capabilities [Maples and Tiefenbacher, 2009; Diakakis and Deligiannakis, 2013]. Especially, the amount of vehicles used in daily commuting, or being available to be used to reach a destination or retrieve family members (and/or property) during flash flooding can be related to the likelihood of people to get trapped in a vehicle-related incident. Information about the aggregated number of vehicles available in the total households in the county are available in the ACS data. In addition to that, different tables of the ACS data provide the aggregate number of vehicles (e.g., car, truck or van) used in commuting by workers 16 years and over, giving an indication of the probable daily exposure for workers whose mean of transportation is vehicle. Figure III.14 indicates large number of vehicles in various urban counties spread in the conterminous U.S.. A characteristic example is the urban corridor that extends from Dallas to San Antonio, and the Houston area in south-central and southeastern Texas, respectively. A visual comparison with the map in Figure III.12, the distribution of vehicles presents similarities with the distribution of men over the U.S. counties. Further analysis reveals that the total number of vehicles, and the number of vehicles used in commuting in the county, are both highly correlated with the number of males in the county (Pearson's correlation > 0.9).

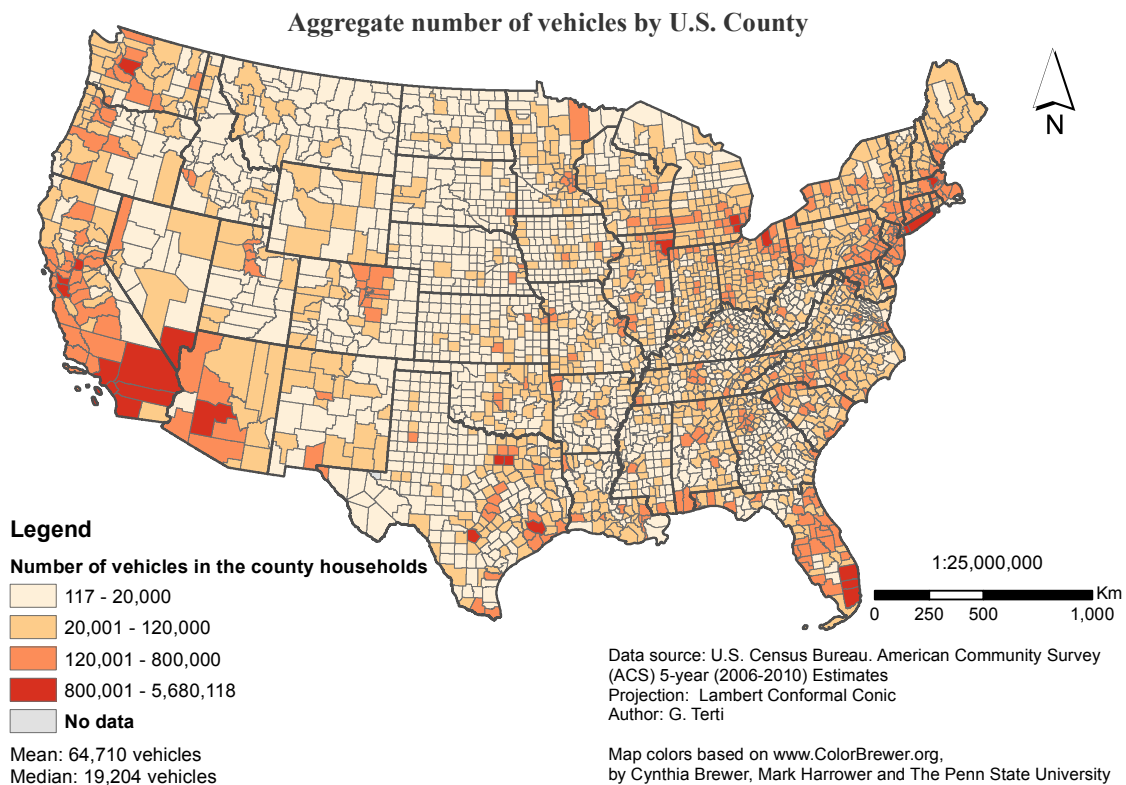


Figure III.14 – County-by-county distribution of the available vehicles in the conterminous U.S.. The aggregate number of vehicles refers to the total number of vehicles available in the total households of the county. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

q) Travel time to work

Longer journeys suggest higher likelihood of exposure to flooded roads [Shabou, 2016]. Ruin et al. [2007] state that flash flood risk perception of motorists correlates to their everyday experience of the road networks. Especially, they found that commuters risk perception is increased for short daily travels. Information about the length of commuting for the county population are provided by the ACS data either as aggregated travel time to work (in minutes) or as the number of commuters in twelve time intervals of 5 or 10 minutes of traveling (e.g., workers traveling less than 5 minutes, 5 to 9 minutes, . . . , more than 90 minutes).

r) Time arriving at work

When compared with the time of the flash flood occurrence, the commuters that arrive at work at that time are indicative of the exposure of work-related mobile population [Ruin, 2010]. On the other hand, the time that the majority of population arrives at work in the county may highlight critical hours (e.g., rush hours) in a specific county to be considered when a flash flood event occurs. The ACS 5-year estimates include commuting details such as the number of commuters that arrive at work at specific time intervals. In fact, fourteen time intervals are provided with a time step of thirty minutes (e.g., 5:00 a.m. to 5:29 a.m., 5:30 a.m. to 5:59 a.m.) except for the two longer intervals concerning early morning and late evening hours (i.e., 12:00 a.m. to 4:59 a.m., and 4:00 p.m. to 11:59 p.m) when the temporal resolution is reduced due to confidentiality reasons. The information is also discretized by means of transportation to work for the county of the workplace providing interesting inputs when studying vehicle-related human impacts from flash flood occurrences.

4 Summary and Conclusions

Quantifying human vulnerability in terms of loss of life risk during short-fuse dynamic flash flooding requires an integrated interdisciplinary approach where many factors may interplay and coincide. The lack of comprehensive datasets to describe the hydrometeorological response and the characteristics of areas and people exposed to those responses associated with certain impacts at specific times and locations, adds a big challenge towards realistic assessments. In this chapter, we described our effort to collect nationwide data from various sources in the U.S. and merge them into one common database using as reference the countywide occurrence of historic flash flood events.

Firstly, we compiled a 19-year dataset with 1,075 individual flash flood-specific fatalities (1996-2014) from annual files available in the *Storm Data* at the time of the analysis. The *fatality* files were merged with the *event* files, based on the ID of each flash flood event reported from 1996 to 2014 in the *Storm Data*, to get complementary information about the timing, the duration and the narratives related to the causative flash flood events. The circumstances of the elaborated fatality records were inspected through the corresponding event narratives, and a classification scheme was set to further analyze vulnerability situations associated with historic losses from flash floods in the remainder of this thesis (Chapter IV). In a second step, the spatial and temporal coverage of the dataset was restricted to ten years (2001-2011) in the conterminous U.S. to match with the spatio-

temporal availability of additional data and especially, the hydrological description of the reported flash flood events through simulated discharges. The individual fatalities in a given circumstance were aggregated for each of the 385 fatal flash flood event reported from 2001 to 2011. Then, the non-fatal flash flood events reported for the same time period in the *Storm Data* were added to the list of fatal flash flood reports composing a dataset of 38,106 flash flood records. In that way, each flash flood record could be classified as “yes” or “no” depending if the specific event led to one or more fatalities in a given circumstance or did not include reported fatalities at all, respectively. This provides for a binary classified dataset of flash flood events that can be used to formulate the target variable to understand or predict when assessing the occurrence of human losses from flash flooding in a certain circumstance.

A total of 38,106 flash flood events composed the final event-based dataset to be supplemented with extra information. The selection of information relevant to serve as candidate indicators for the assessment of human losses to flash flood hazard (Chapter V), was supported by a review of literature on vulnerability and human impact studies, and interpretations of the author based on the flash flood spatial and temporal specificity described in Chapter II. About thirteen different databases were downloaded and edited to provide indicators related to hydrological and meteorological responses, geomorphological characteristics, information for the available official emergency response, and other spatial and socio-demographic attributes. The biggest challenge in this process was to integrate data produced at different scales and/or resolution, expressed in different units and in terms relevant to a specific discipline. Data processing was mainly supported by the R project for statistical computing and ArcGIS GIS platform from Esri to refer the gathered data to the county administrative unit, and join the variables to develop a consistent database. Especially, the event identifier, the county in which a specific flash flood event was reported in the data, and the date of the flash flood occurrence (when applicable), were the reference information for the final merging of all the datasets. Different origin data are assumed to contribute to different aspects of the assessment of human vulnerability depicting the overall view of human risk during the “event” phase of flash floods.

TOWARDS SITUATIONAL EXAMINATION OF HUMAN VULNERABILITY TO FLASH FLOOD

Chapter IV: A situation-based analysis of flash flood fatalities in the United States

Chapter IV is based on a paper published in *Bulletin of American Meteorological Society (BAMS)*, *Journal of the American Meteorology Society*: Terti, G., Ruin, I., Anquetin, S. and Gourley, J.J., 2016. A Situation-based Analysis of Flash Flood Fatalities in the United States. *Bulletin of the American Meteorological Society*, doi: 10.1175/BAMS-D-15-00276.1, in press.

Résumé

LE Chapitre IV propose une analyse des circonstances de 1,075 décès dus aux crues rapides enregistrées entre 1996 et 2014 aux États-Unis. Cette étude décrit les circonstances des décès à partir du profil et de l'activité des victimes ainsi que du contexte spatio-temporel de l'aléa. Sur la base du reclassement des circonstances des décès (lieu / activité) effectué dans le Chapitre III, nous étudions statistiquement l'heure de début de l'épisode, la durée et l'emplacement ainsi que l'âge et le sexe des victimes. En accord avec d'autres études, plus de 60% des décès rapportés sont liés à des véhicules impliquant principalement des hommes. Une analyse géospatiale indique qu'ils sont les usagers les plus fréquents dans les états du sud. De plus, 21% des décès surviennent à l'extérieur, généralement dans les quartiers situés à proximité immédiate des cours d'eau, où les victimes présentent un comportement à risque élevé, comme, par exemple, le nettoyage des canalisations et même des activités "aquatiques" dans les eaux en crue. La vulnérabilité humaine varie de façon dynamique sur une base infra-journalière et dépend des caractéristiques de la crue. Par exemple, la plupart des décès liés au camping sont associés à des événements rapides (moins de 5 heures de durée), se produisent plus fréquemment après minuit. Ils ont un impact sur les jeunes femmes et les jeunes hommes. En revanche, les décès liés à l'inondation de bâtiments sont le plus souvent associés à des événements de plus longue durée et touchent les personnes âgées. Un examen situationnel plutôt que générique de la vulnérabilité est nécessaire pour saisir de façon réaliste les cas à risques pour les individus lors des crues de courte durée.

Abstract

Chapter IV investigates the circumstances of 1,075 fatalities from flash flooding recorded from 1996 to 2014 across the United States. This study provides insights on the situation of the fatality events as determined by the victims' profile and activity and the spatio-temporal context of the flooding. Based on the re-classification of the individual fatality circumstance (i.e., location/activity) performed in Chapter III, we explore statistically the timing, the duration and location of the flash flood event, and the age and gender of the victims. In agreement with other studies, more than 60% of the reported fatalities are related to vehicles involving mainly males. A geospatial analysis indicates these are most common in southern states. Further, 21% of fatalities occur outdoors, typically in neighborhoods near streams, where the victims were exhibiting high risk-taking behaviour such as cleaning out drains, and even playing in the floodwaters. Human vulnerability varies dynamically on a sub-daily basis and depends on social and natural factors of the flash flood. For example, most campsite-related fatalities are associated with very fast responding flash flood events (less than 5 hours duration), occur more commonly after midnight, and impact younger females and males alike. On the other hand, fatalities related to inundation of permanent buildings are most commonly associated with longer duration events and impacted the elderly. Situational rather than generic examination of vulnerability is required to realistically capture risky cases for individuals during short-fuse flood events.

A situation-based analysis of flash flood fatalities in the United States

1 Introduction

Flash floods are often associated with rapid rises in water levels and fast-moving waters that can sweep humans and their cars off their intended path [Montz and Grunfest, 2002; Jonkman and Vrijling, 2008; Ruin et al., 2009; FEMA, 2015]. Human impact studies are sometimes hazard-specific but only a few focus on flash floods [Mooney, 1983; French et al., 1983; Staes et al., 1994]. A review of the literature shows that:

- (i) in most of the natural hazard mortality studies worldwide, flash flooding information is merged with other types of floods for analysis [Coates, 1999; Ahern et al., 2005; Borden and Cutter, 2008; FitzGerald et al., 2010; Kellar and Schmidlin, 2012],
- (ii) many studies are case-specific or are restricted to the analysis of fatality data obtained from a limited number of flood events in specific regions [Staes et al., 1994; Jonkman and Kelman, 2005; Jonkman et al., 2009; Maples and Tiefenbacher, 2009; Sharif et al., 2014], and
- (iii) when fatal accident circumstances are investigated, studies either focus on one specific type of circumstance (often the vehicle-related one), or spatial and temporal patterns specific to the various circumstances are rarely addressed [Coates, 1999; Ashley and Ashley, 2008; Maples and Tiefenbacher, 2009; Sharif et al., 2012; Diakakis and Deligiannakis, 2013].

Because flash flooding events can be distinguished from riverine floods by their fast response to rainfall and resulting impacts signature [Jonkman and Kelman, 2005; Gourley et al., 2013], this dissertation proposes analyzing flash flood-specific impact datasets to identify the conjunction of social and physical circumstances leading to those impacts. As described in Chapter III, data on

flash flood-related casualties from 1996-2014 are derived from the NCEI *Storm Data*. Although not unbiased, *Storm Data* is the most comprehensive nationwide database for flash flooding events and the resulting impacts (i.e. fatalities, injuries and damages) [Gall et al., 2009]. Currently, our study is restricted to the analysis of fatalities due to the availability of details concerning victims (e.g., age, gender, location). Although sometimes included as comments in the event narratives of *Storm Data*, details about other nonfatal impacts from flash floods, such as injuries or rescues, are not provided in a coherent database on the U.S. scale, yet.

Rather than using the claim of the “deadliest flood type” to study flash floods separately, we address specific aspects of vulnerability that are not relevant in the case of general flooding [Terti et al., 2015]. Chapter II stated that the intersection of the spatio-temporal context of the flash flooding phenomena with the distribution of people and their socio-demographic characteristics reveals various paths of vulnerability through the expression of different accidents’ circumstances (i.e., vehicle-related, inside buildings, open-air). In the proposed conceptual vulnerability model, we use the term “coupled place-activity” to point out that the nature and dynamics of the individuals’ reactions will differ according to the location and activity they were performing when they felt the need for action, and their capability to connect with their relatives or to have social interactions allowing a group response [Ruin et al., 2014].

Previous analyses highlighted the importance of the location and activity of the exposed individuals during a flash flooding event on the distribution of impacts [Ashley and Ashley, 2008; Ruin et al., 2009]. Ashley and Ashley [2008] analyzed 4586 flood fatalities included in *Storm Data* for the period 1959-2005 to provide conclusions on the vulnerable states and populations in the contiguous U.S. Examining the frequency of all flood-related fatalities by location revealed that 63% were associated with vehicles whereas a number of deaths happened “in water” (9%) in cases where the victims intentionally entered the flood waters. Špitalar et al. [2014] used a unified flash flood observational database compiled at the National Severe Storms Laboratory (NSSL) [Gourley et al., 2013] to analyze spatial, temporal and hydrological parameters with human impacts. In their study, physical attributes related to 21,549 events in U.S. (2006-2012) were cross-analyzed with the aggregated number of fatal events weighted with the fatalities. Their findings propose late evening flash flooding occurrences as the most devastating in terms of injuries and fatalities. Further investigation of the vehicle-related casualties showed that visibility but also rush hour habits contribute to more impactful flash floods. The aforementioned studies do not analyze the profile of victims in certain circumstances. However, they reveal that certain behaviors and attitudes are embedded in the fatal scene, inviting future research on the socio-spatio-temporal characteristics of the circumstances and identification of the vulnerability factors.

Jonkman and Kelman [2005] proposed a categorization of the causes and circumstances for 247 deaths caused by 13 small-scale flood events in Europe and the U.S. Their classification is a valuable contribution towards a more consistent comparison between different fatal flood events. The reclassification of the fatalities circumstances in the present study does not intend to present statistics on the exact reason or location of the losses. We rather attempt to contextualize prominent responses and behaviors of the victims using smaller number of classes that will facilitate more targeted warning and prediction approaches in the future. The purpose is to identify the circumstances that can be described by certain physical attributes of the exposed environment (e.g., road network,

campsites, mobile homes) and/or socio-demographic characteristics of the exposed population (e.g., family status, work travels) to serve as vulnerability predictors associated with human risk during flash flood crisis. Today, very little is known about the distribution of flash flood-specific human losses under certain circumstances and/or on a sub-daily basis in the U.S. Unlike previous work in mortality data analysis, information about the victims and the spatio-temporal context of the fatal flash flooding events are disaggregated for each of the circumstances. The analysis addresses the following questions:

- i. What are the predominant circumstances associated with the occurrence of fatalities during flash flooding events?
- ii. What is the temporal distribution of flash flooding fatalities for the different circumstances?
- iii. Who is the most vulnerable to flash flooding in terms of loss of life; are the same patterns revealed when discretizing by circumstance and/or time of the day?
- iv. What is the sub-state and sub-county distribution of circumstance-specific fatalities across the entire U.S.?

We believe that having the circumstances as the center point of the analysis is fundamental to superimpose situational against generic vulnerability assessment. The spatial analysis improves the picture of the geographic distribution of flash flooding fatalities in the U.S. The results of this research can contribute to the development of more targeted warning and prediction approaches to prevent human losses during flash floods, as introduced in Chapter V.

2 Statistical analysis of flash flood fatalities

2.1 Circumstances of the fatalities

This study is based on information about fatalities caused by flash flood events reported in nineteen annual *fatality* and *event* files of the *Storm Data* (for data description see section 2.1 of Chapter III). Examining the event narratives associated with 1,075 individual deaths reported from 1996 to 2014 in *Storm Data*, allowed us to assign each fatality to one of the six prevalent circumstances: (i) vehicles; ii) permanent buildings like homes or businesses; iii) mobile homes; iv) campsite or recreational areas; v) outside/open air and close to streams/rivers areas; and vi) other/unknown. The number of classified fatalities in each circumstance is presented in Table III.3.

After re-classification, 61% of the total 1,075 fatalities occurred in circumstances related to vehicles. If we remove the 39 fatalities for which the location or activity of the victim could not be defined (Table III.3), the vehicle-related circumstances account for 63% of the fatalities with known circumstance (Figure IV.1A). Despite differences in the exact percentages due to the data temporal and spatial coverage, these findings agree with previous studies stating that most of the flood fatalities in the U.S. [Mooney, 1983; Staes et al., 1994; Ashley and Ashley, 2008; Maples and Tiefenbacher, 2009; Sharif et al., 2014; Špitalar et al., 2014], Australia [Coates, 1999; FitzGerald et al., 2010], and Europe [Jonkman and Kelman, 2005; Diakakis and Deligiannakis, 2013] are vehicle-related with drowning being the main cause of death [French et al., 1983; Ryan and Hanes, 2009].

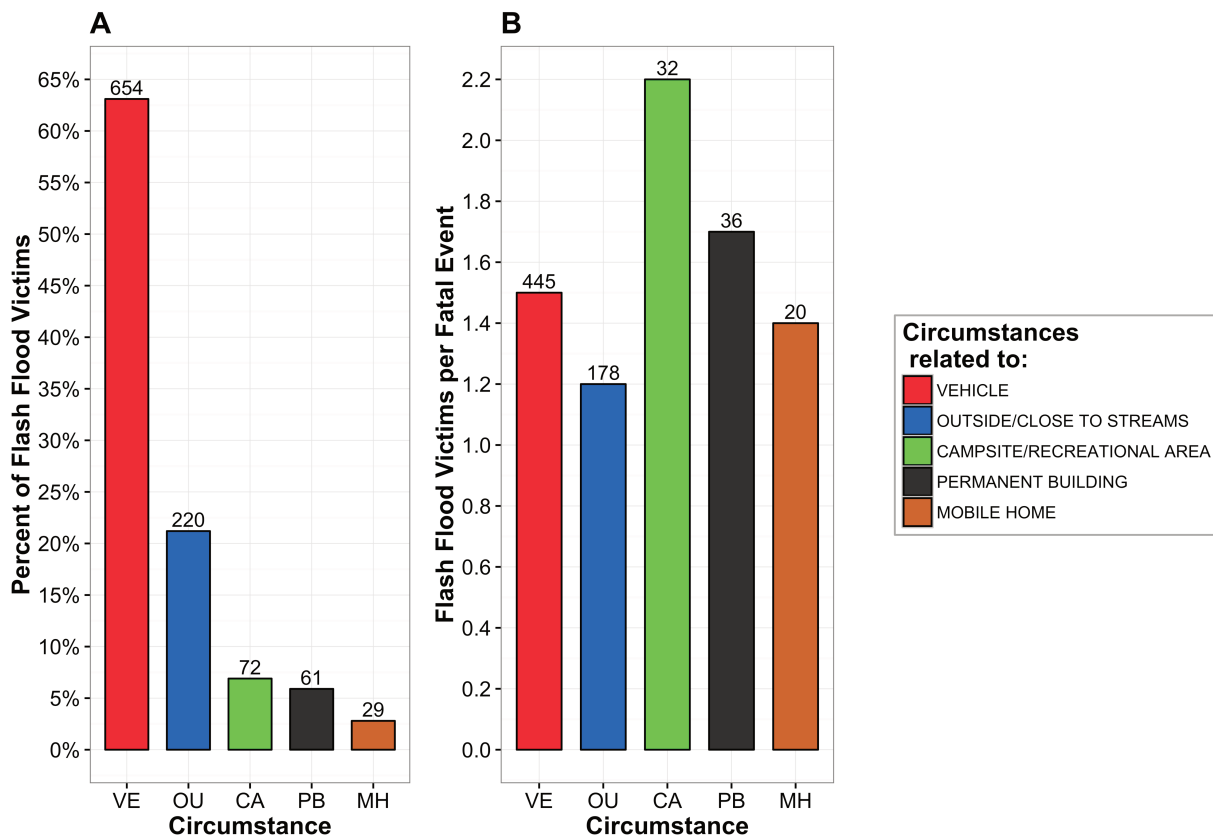


Figure IV.1 – A) Percent of flash flood victims by circumstance. Percentages are estimated to the total 1,036 fatalities with defined circumstance. The values on the top of the bar indicate the row number of fatalities in each circumstance. B) Ratio of flash flood victims per event with fatalities in each circumstance. The values on the top of the bar indicate the raw number of fatal flash flooding events in each circumstance. Some of the total 679 unique fatal events led to fatalities in more than one circumstances.

The second dominating circumstance is the outside/open air or close to streams areas. This category is mainly composed of cases in which people were swept into creeks while trying to walk through swollen portions of a creek or already flooded intersections to reach their relatives and/or belongings (e.g., home). Out of the 84 (from the total 220) deaths for which additional information could be extracted for the motivation of the activity, only 13% of them were related to rescue or evacuation processes. Qualitative analysis of actual behavioral response to flood have shown that such conflicting priorities (e.g., staying safe versus saving your belongings or livelihood) are common and often require the intervention of a third party who has no emotional nor financial involvement in the situation for people to be restored to their common sense [Jonkman and Kelman, 2005; Ashley and Ashley, 2008; Ruin et al., 2014].

Although less events are associated with campsite/recreational area-related circumstances, when they happen they lead to more losses per event compared to the other circumstances (Figure IV.1B). Given the fast onset of flash flooding events, people in recreational areas may be more subject to surprise as their remote locations lower their chances to be weather aware and warned, limiting the speed of their response. Moreover, people are more likely to be grouped together in campsites rather than traveling alone. The outside/open air or close to streams events, on the other hand, are more typical to impact individuals rather than groups of people.

2.2 Temporal characteristics of the fatalities

a) Annual and monthly distribution

The annual number of flash flooding fatalities varies significantly from a minimum of 19 in 2012 to a maximum of 118 in 1998 with an average of 57 fatalities per year from 1996-2014. In particular, the median of reported victims is 59 victims per year, and the median of the fatal events is 38 events per year (Figure IV.2). The most deadly events in 1998 were related to the “Great October Flood” that drowned at least 25 people in multiple counties in South Central Texas [NOAA, 1999]. According to the *Storm Data* episode narratives, huge livestock losses on the order of tens of thousands, and almost 3000 destroyed and 8,000 damaged homes characterized that event. Nearly 1,000 mobile homes were destroyed and another 3,000 were damaged.

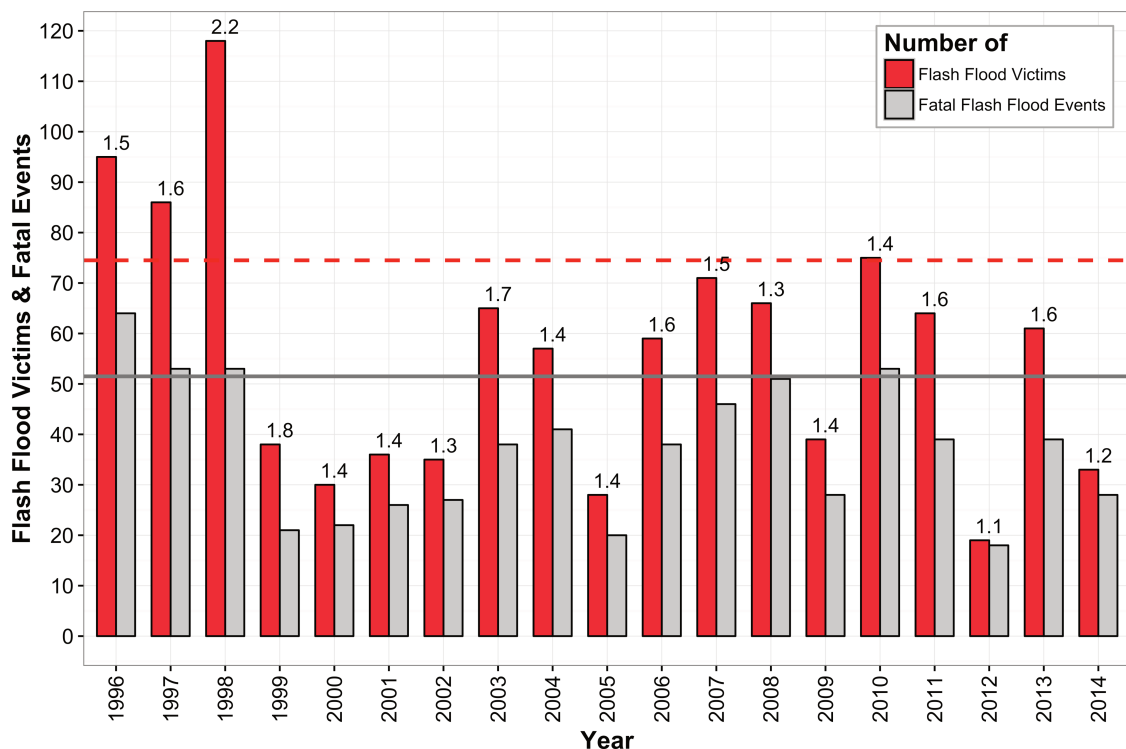


Figure IV.2 – Annual frequency of flash flood victims (n=1,075) and fatal events (n=705) from 1996 to 2014. The dashed red horizontal line represents the annual fatality median and the solid grey line represents the annual fatal event median. The values on the top of the fatality bars indicate the ratio of flash flood victims per fatal event for each year.

Flash flooding events and associated fatalities show no clear trend from 1996 to 2014. A similar observation was noted for the flood fatalities from 1959 to 2005 and for vehicle-related flood fatalities from 1995 to 2005 across the U.S., presented by Ashley and Ashley [2008] and Kellar and Schmidlin [2012]. As the *Storm Data* continues to grow after the execution of the current analysis, 118 more human losses related to 70 reported flash flood events are recorded for the study area in 2015¹. The lack of decreasing fatalities can be partly explained by increases in exposure due to population growth especially in the South and West during the study period [U.S. Census Bureau, 2016]. However, it also indicates that despite recent technological improvements in the forecasting and watch-warning

1. Data available online at <https://www.ncdc.noaa.gov/stormevents/ftp.jsp>.

system, they did not necessarily lead to prevention of losses from flash floods. This flash flood-specific finding is in contrast to other weather-related hazards losses such as tornadoes [Brooks and Doswell III, 2002] and lightning [Ashley and Gilson, 2009] in the U.S, which show a stabilized or even decreasing trend with time, respectively [Ashley, 2007; Ashley and Strader, 2016].

Flash flood impacts are strongly dependent on the social response driven from both the communication of risk and the individual decision-making during the event. Disaggregation of the annual number of fatalities into specific circumstances reveals that almost every year a majority of flash flood victims died after deciding to use their vehicle under extreme weather conditions (Figure IV.3). An exception to this finding is in 2012 and 2013 when the OU fatalities equaled or exceeded the VE ones. The CA events reached a peak in 2010 that was a result of a deadly flash flood that occurred in the Albert Pike Campground on the 11th of June. Twenty campers sleeping in tents lost their lives as a river rose rapidly in response to more than 175 mm of rain that fell during the night.

Flash flood impacts are strongly dependent on the social response driven from both the communication of risk and the individual decision-making during the event. Disaggregation of the annual number of fatalities into specific circumstances reveals that almost every year a majority of flash flood victims died after deciding to use their vehicle under extreme weather conditions (Figure IV.3). An exception to this finding is in 2012 and 2013 when the OU fatalities equaled or exceeded the VE ones. The CA events reached a peak in 2010 that was a result of a deadly flash flood that occurred in the Albert Pike Campground on the 11th of June. Twenty campers sleeping in tents lost their lives as a river rose rapidly in response to more than 175 mm of rain that fell during the night.

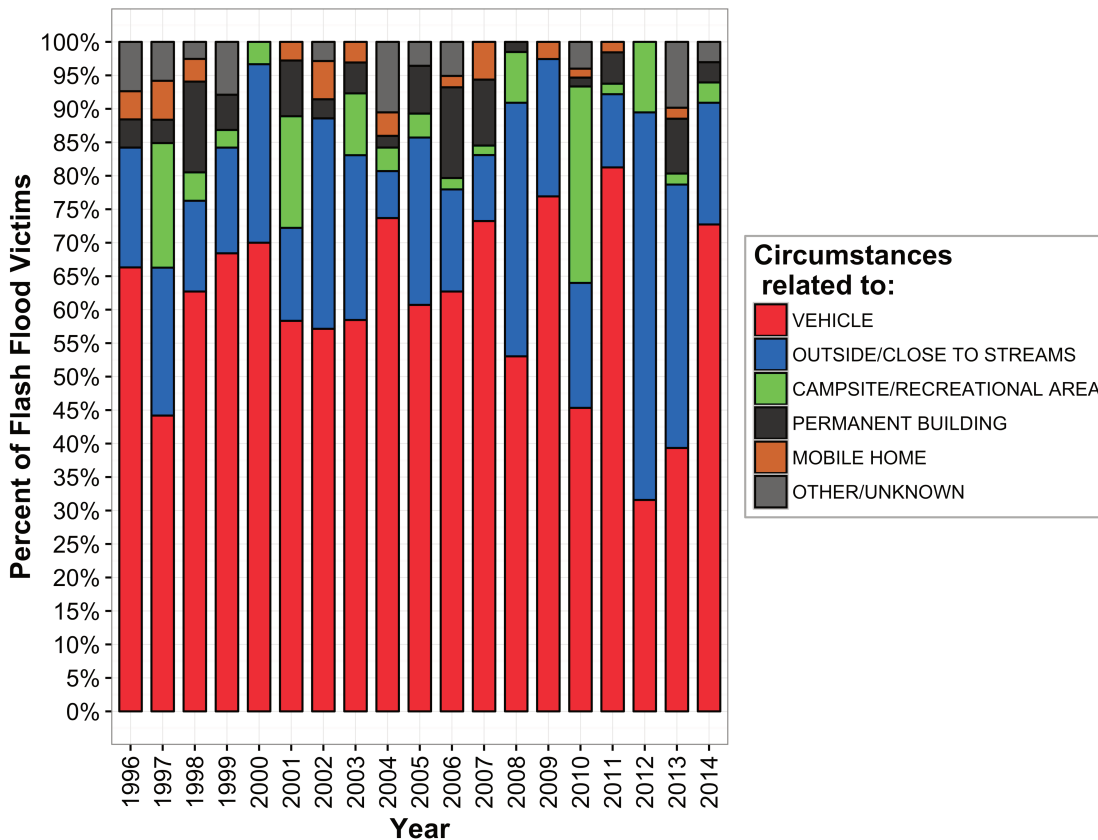


Figure IV.3 – Percentage of flash flood victims by circumstance of the death for each year. The proportions of each circumstance (if exists) in a certain year are estimated by dividing the victims in that year and circumstance to the total victims in the specific year.

The monthly analysis indicates that the peak in fatalities and fatal events occurs from May through September with a peak in August (Figure IV.4). Sixty-six percent of the annual fatalities and 65% of the flash flooding fatal events occurred during this warm season period. The warm season preference of both flash flooding fatalities and fatal events has been attributed to the spatio-temporal distribution of heavy precipitation in previous analyses [Ashley and Ashley, 2008; Sharif et al., 2014; Špitalar et al., 2014]. Figure IV.4 also shows that although less in number, fall events during October were particularly deadly with almost two fatalities/event. The total 77 fatalities were recorded in 40 flash flooding events that happened in the states of Texas, Missouri, and Puerto Rico throughout the 19-year period. Almost half of these fatalities resulted from just three meteorological episodes, all of which occurred in 1998: the “Great October Flood” in Texas, the October 4th event in Missouri and the October 22nd tropical-storm related flash flood in Puerto Rico [NOAA, 1998].

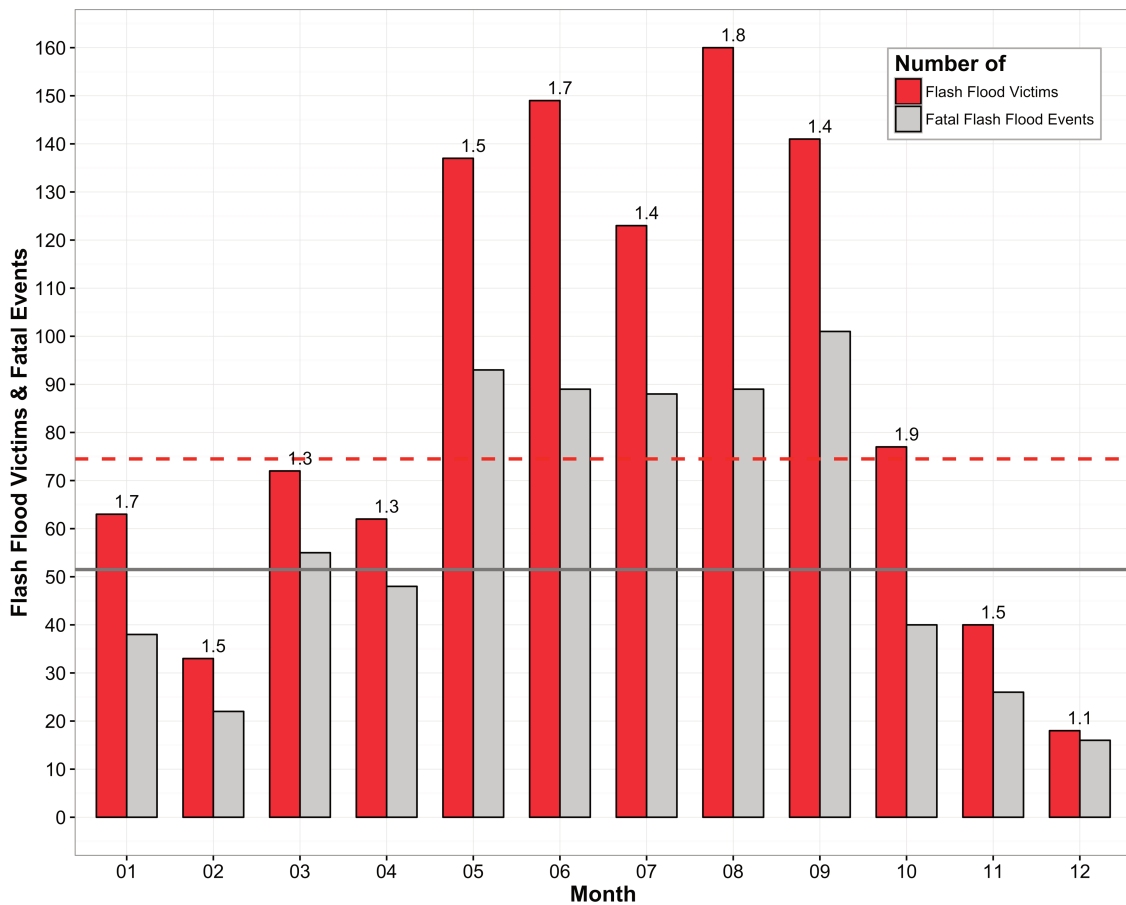


Figure IV.4 – Monthly frequency of flash flood victims (n=1,075) and fatal events (n=705) from 1996 to 2014. The dashed red horizontal line represents the monthly fatality median and the solid grey line represents the monthly fatal event median. The values on the top of the fatality bars indicate the ratio of flash flood victims per fatal event for each month.

When monthly fatalities are examined by circumstance it reveals that in all months except for July more than the half of the monthly fatalities are related to vehicles (Figure IV.5). The domination of vehicle-related incidents through the years and months confirms that the association of the majority of fatalities in the dataset with vehicle-related circumstances (Figure IV.1A) is not specific to rare, individual events, and warrants additional investigation into social spatio-temporal behaviors.

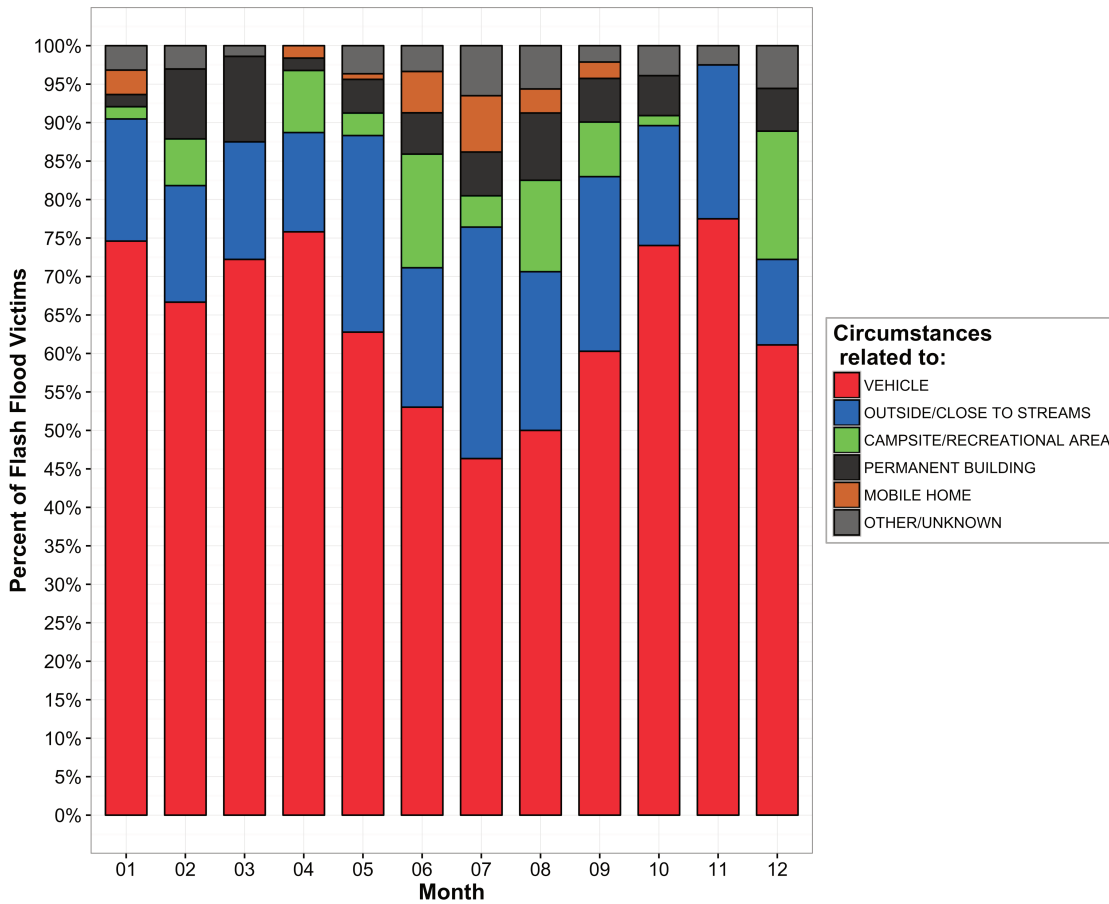


Figure IV.5 – Percentage of flash flood victims by circumstance of the death for each month. The proportions of each circumstance (if exists) in a certain month are estimated by dividing the victims in that month and circumstance to the total victims in the specific month.

b) Distribution based on the timing and the duration of the causative event

Flash flood duration is estimated as the difference between the beginning and the end time of the causative event available for 954 out of the total 1,075 fatalities in our dataset. The calculated duration is rounded to closest hour. The cumulative percentages of the flash flooding fatalities per hour illustrates that 54% of the fatalities are associated with very short events (five hours or less) whereas 77% happened in events that lasted ten hours or less (Figure IV.6). This result agrees with the analysis of Špitalar et al. [2014] who show that most of the flash flood events with casualties in U.S. occur within five hours. It should be noted that the event durations are often reported for an entire episode rather than the specific, causative flash flooding event. For instance, a land-falling tropical storm can take days to traverse a region and yield multiple impacts and fatalities. This type of flash flooding event will be handled with a single report in *Storm Data* that will inevitably cover the event’s duration from landfall until the time of its final impact. As such, caution is warranted on interpreting results for the long-duration reports in contrast to the events duration that occur within approximately ten hours.

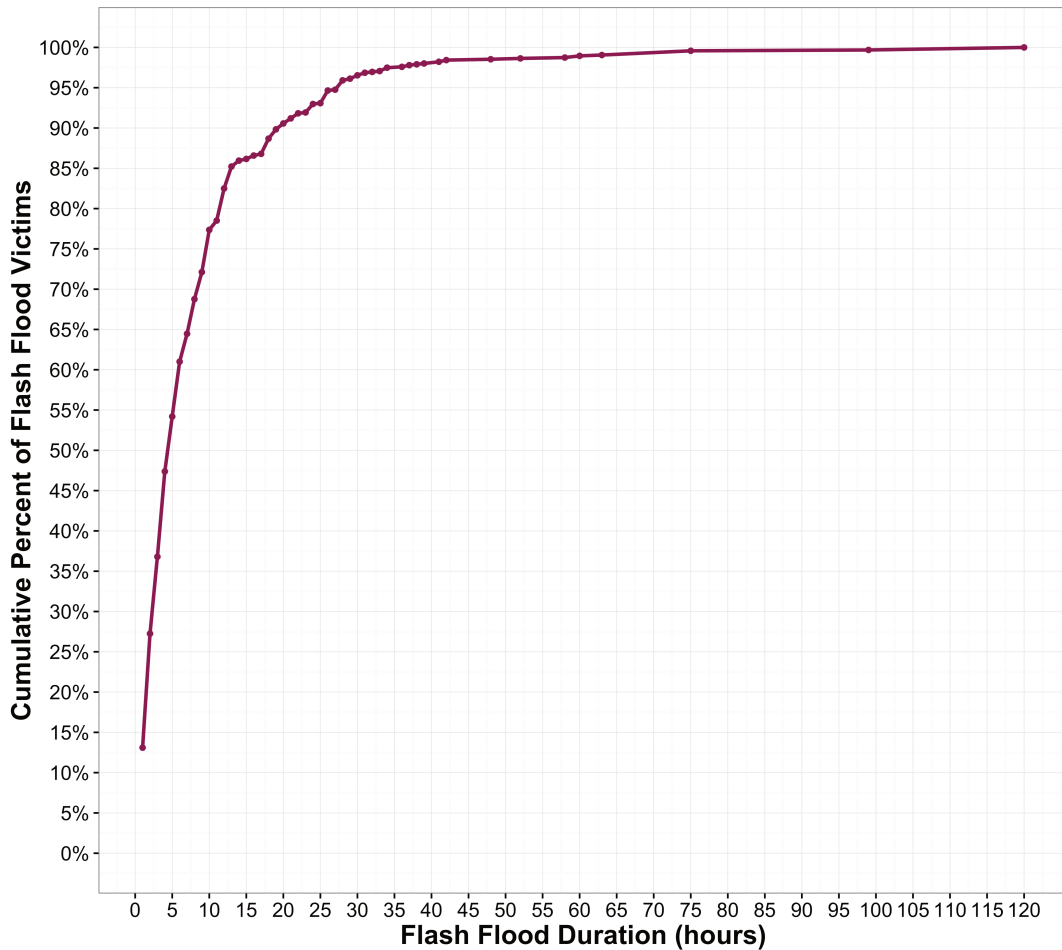


Figure IV.6 – Cumulative percentage of flash flood victims with the duration of the causative fatal flash flood event from 1996 to 2014. The duration is estimated as the difference between begin and end time of the flash flood event and rounded in the closest value. The percentages (red dots) are calculated for each hour to the total fatalities for which the duration of the causative event can be estimated (n=954).

The distributions of the flash flood fatalities for the specific circumstances show the fatalities decrease sharply as the event duration increases. There are differences between fatalities that are related to outside and inside activities (Figure IV.7). Specifically, on average, people in outdoor circumstances such as the vehicle-related, open air/close to streams and camping/recreational categories drowned in very fast and dynamic flash flooding events with durations close to five hours. The distribution of the flash flooding duration in the camping circumstance presents much less variation indicating that it is the very fast-reacting events that surprise and trap people in recreational areas. Rapid flash floods imply limited time for warning, anticipation, and reaction [Creutin et al., 2009; Špitalar et al., 2014]. On the other hand, fatalities that are related to permanent buildings or mobile homes have durations with central tendencies more than 7.5 hours. The median of the distribution of victims being in a permanent structure is about eight hours, meaning that longer duration flash floods are the threatening ones for people inside.

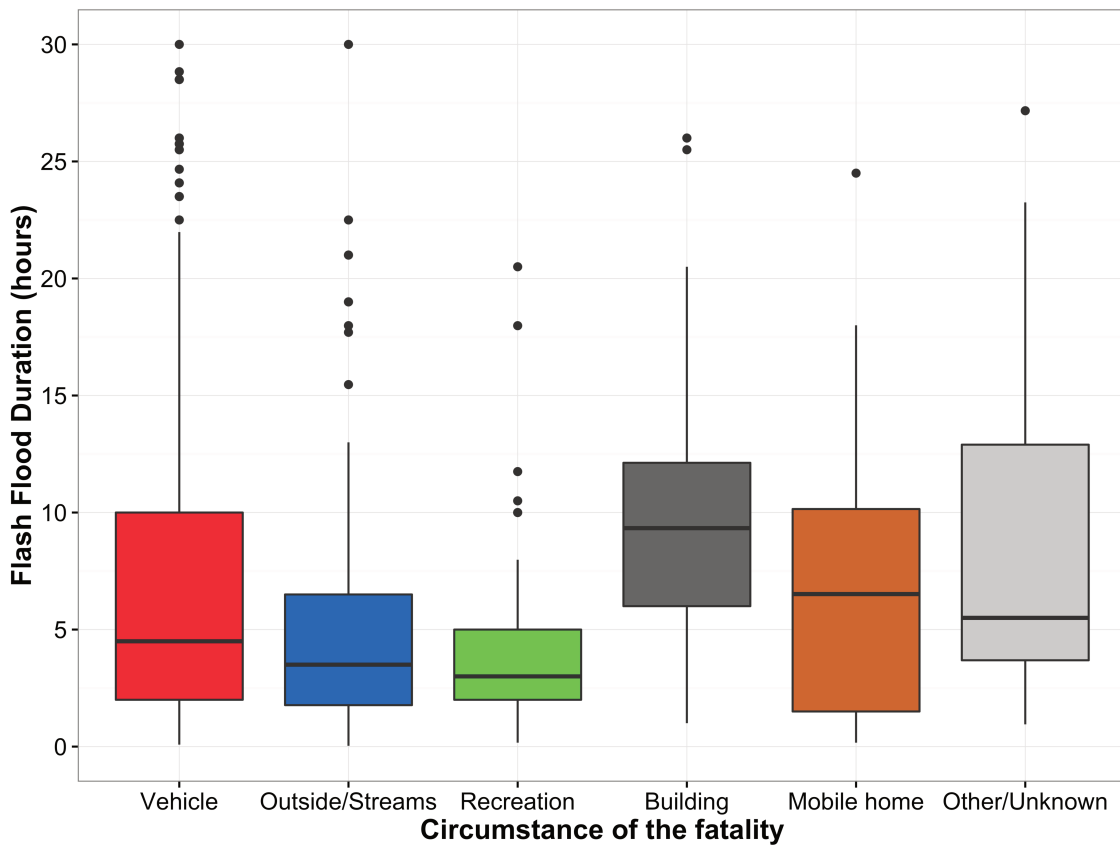


Figure IV.7 – Distribution of the flash flood duration discretized by the circumstance of the related deaths. Boxplots are constructed for each circumstance separately for the fatalities for which the duration of the causative event can be estimated and it is less or equal to 30 hours (n=921).

Flash floods occur at any time of day, but when associated with the hours of twilight and darkness they become more severe in terms of lethal impacts [Špítalar et al., 2014]. In our study, 38% of the 1,075 fatalities occurred during nighttime hours (after 2100 until 0600 local time) (Figure IV.8). To better understand the social risk factors, the distribution of fatalities by the local hour of day of the flash flooding occurrence is presented by circumstance in Figure IV.9. When viewing flash flooding fatalities by circumstance, differences appear in the timing of the vehicle-related and outside/open air events. In this study, 41% of the vehicle-related fatalities are associated with nighttime events whereas only 24% of the outside/open air fatalities occur in darkness. Being unable to view the details of the floodwaters such as its mere presence, depth, or movement yields greater vehicle-related flash flooding fatalities during darkness. The secondary peaks of vehicle-related fatalities in events that occurred early in the morning (between 0500 and 0600) and late evening (between 1900 and 2100) indicate that people often refuse to change their daily mobility schedule (i.e., commuting to/from work) even with the occurrence of floodwaters on the roadways. In many cases, victims did not enter the flooded roads accidentally but they intended to drive through low-water crossings or bridges to reach their home probably motivated by their confidence in their vehicle and driving capabilities in a familiar area [Drobot et al., 2007; Maples and Tiefenbacher, 2009; Ashley and Ashley, 2008].

On the other hand, 76% of the outside/open air events occur during daylight hours. The *Storm Data* narratives reveal that in many outside/open air circumstances adults were swept away while performing some cleanup operation to protect their property. Also children and teenagers were swept into creeks while playing near high waters. The daylight conditions probably enabled the victims to

participate in such outdoor activities. Risky behaviors can occur when people feel familiar with their environment or prioritize their property savings [Ruin et al., 2014]. Concerning camping/recreational area fatalities, there are two maxima in the timing of flash flooding events. The secondary maximum occurs during daylight hours closer to the climatological peak of heavy rainfall and flash flooding. These are events that trap people in recreational areas such as in canyons in the desert Southwest. The primary nocturnal peak occurs between 0100 and 0200, during the evening. These are nighttime rest hours where the surprising nature of the flash flood is exacerbated by low visibility and rescue operations are further hindered.

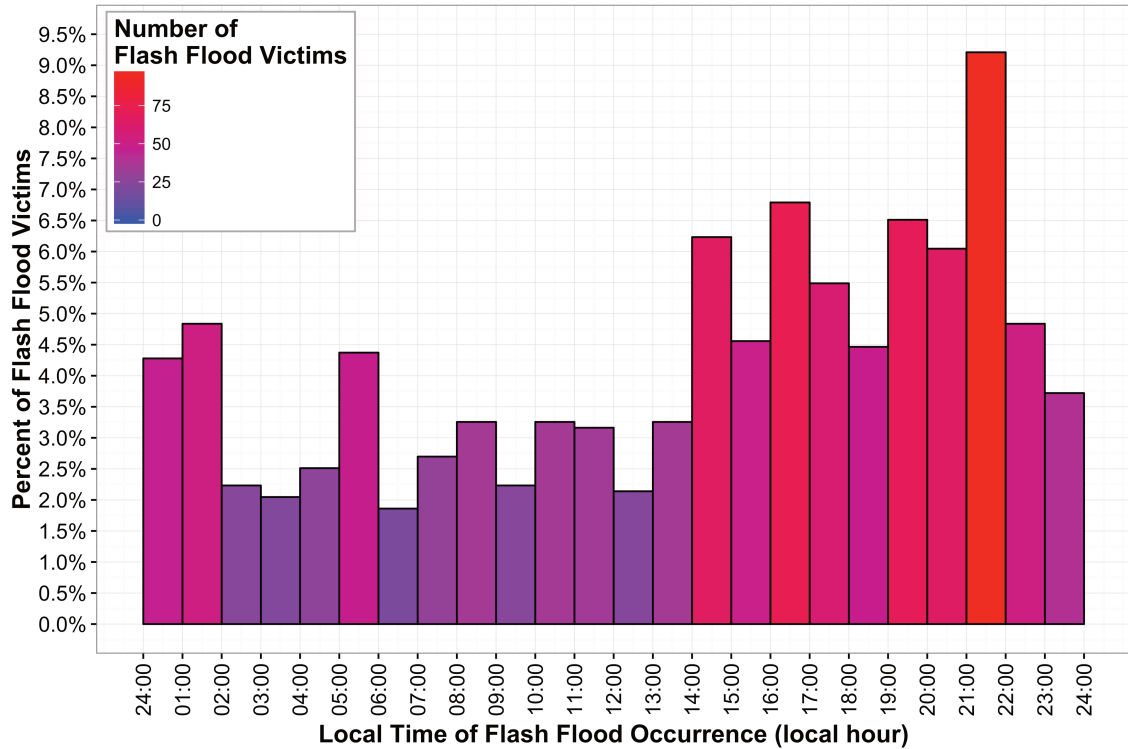


Figure IV.8 – Percentage of flash flood victims by timing of flash flood occurrence. The flash flood-occurrence time is represented by hourly intervals (e.g., flash flood occurring at 24:01 is presented with a bar between 24:00 and 01:00). Percentages of flash flood victims (gradient colored bars) in each hourly interval are estimated to the total fatalities from 1996 to 2014 (n=1,075). The color in the bars depicts the number of victims associated with events occurring in that hourly interval.

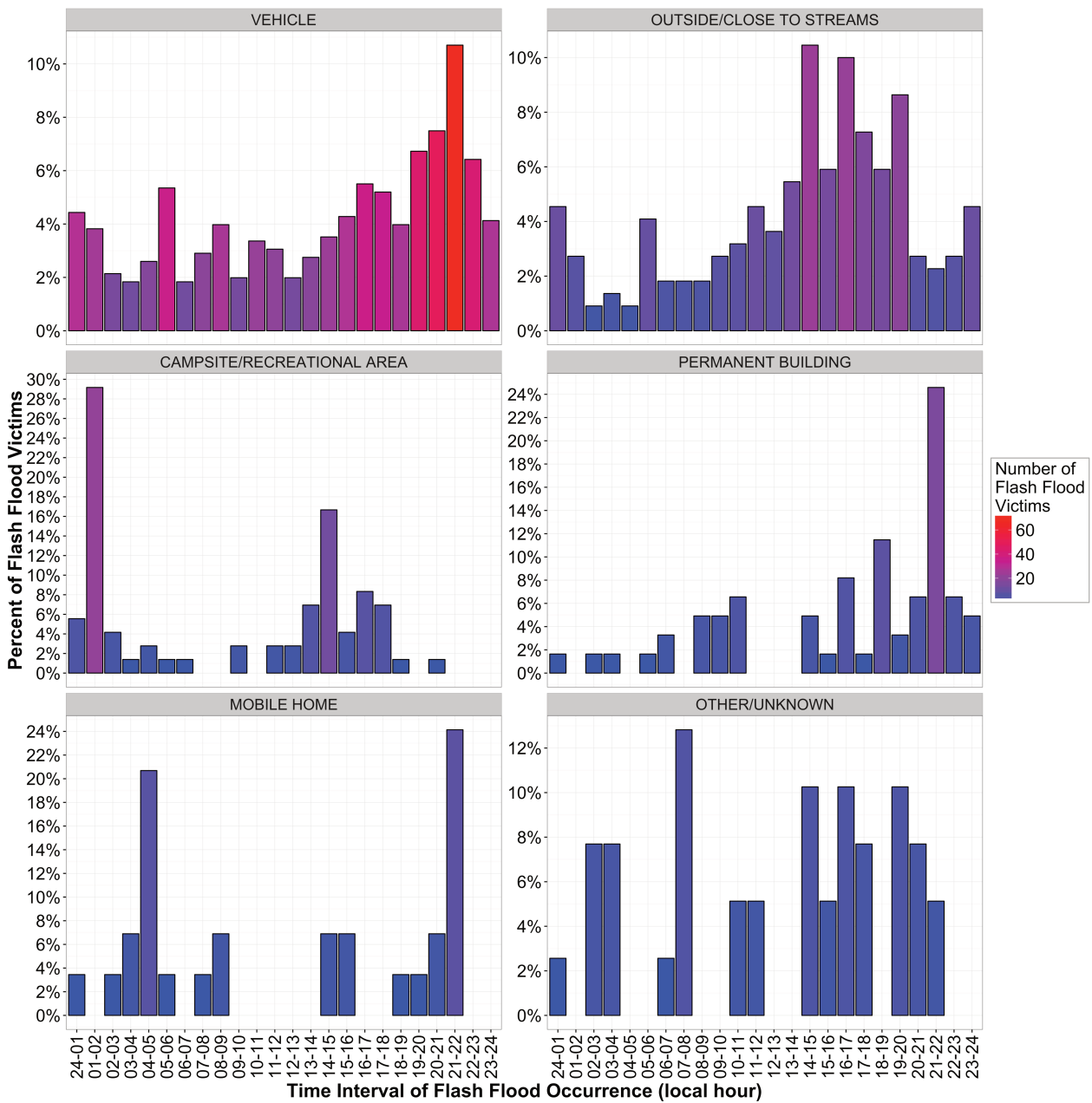


Figure IV.9 – Percentage of flash flood victims by the local timing of flash flood occurrence discretized by the circumstance of the related deaths. The flash flood occurrence time is represented by hourly intervals (e.g., a flash flood event occurring at 24:01 is assigned to the interval 24-01). Percentages of flash flooding victims (gradient colored bars) in each hourly interval are estimated to the total fatalities in each circumstance. The color in the bars depicts the number of victims associated with events occurring in that hourly interval and that circumstance.

2.3 Socio-demographic profile of the victims

Age and gender of the victims are known for 94% and 98% of the flash flooding fatalities, respectively. Sixty-one percent of the victims with defined gender were males (Figure IV.10A). A one-sample chi-square test with proportions pre-defined according to the distribution of males and females in the U.S. population [U.S. Census Bureau, 2010] reveals that vulnerability of males is statistically significant at alpha 0.05 ($p < 0.05$). Further analysis has shown that males exceed the number of females in all the fatality circumstances except for a small difference in mobile home cases (Figure IV.10B). In the literature, overrepresentation of males is usually assigned to risky behavior

or the fact that they are more exposed as frequently employed in emergency services [Coates, 1999; Jonkman and Vrijling, 2008; Doocy et al., 2013].

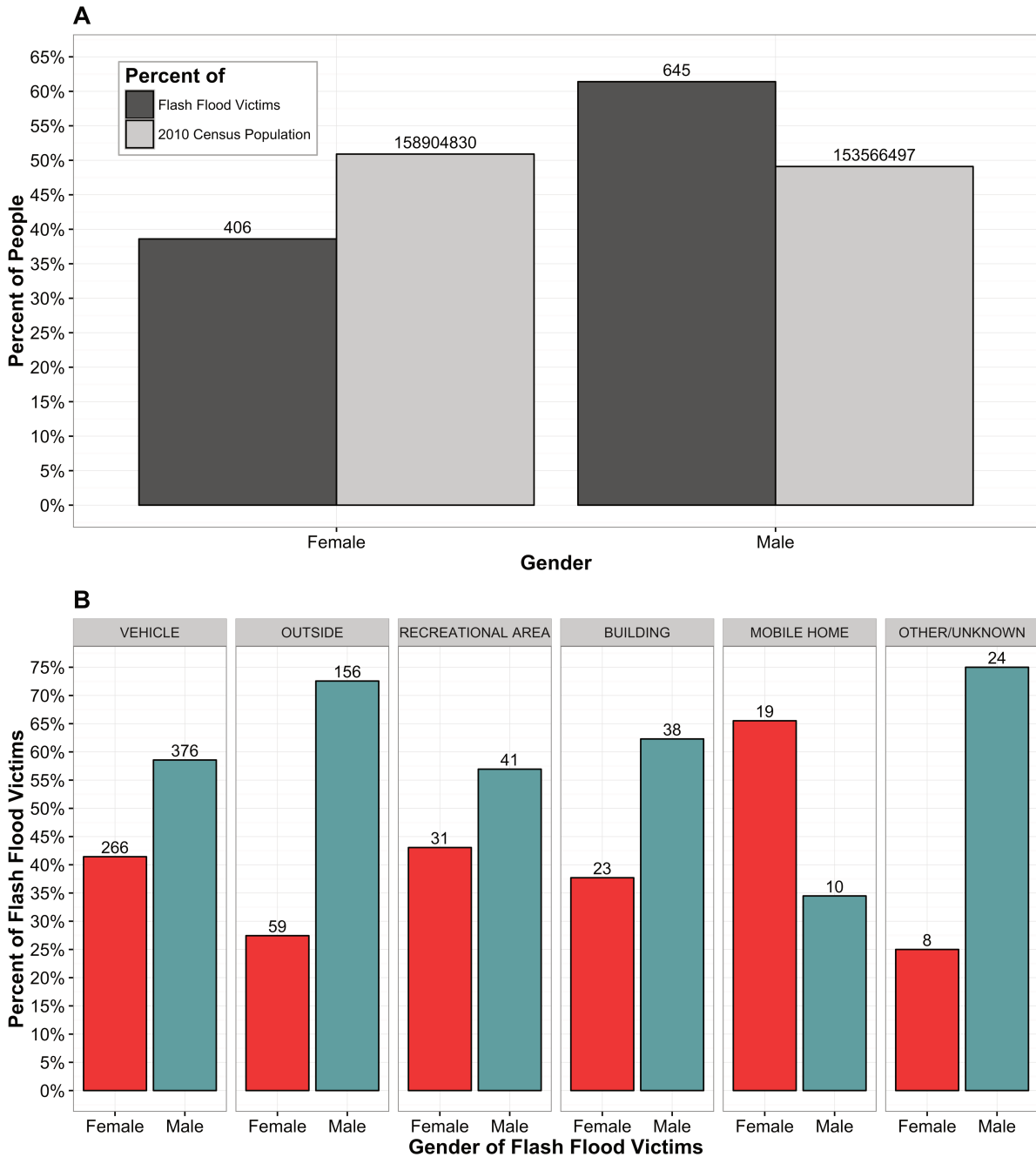


Figure IV.10 – Percentage of flash flood victims and population by gender. A) Percentages of flash flood victims (dark grey bars) in each age group are estimated to the total fatalities with known gender from 1996 to 2014 (n=1,051). The percentage of population (light grey bars) is based on the U.S. Decennial Census: 2010. B) Percentages of flash flood victims in each gender are estimated to the total fatalities in each circumstance separately. The values on the top of the bars indicate the raw number of victims of that gender and in that circumstance. The sum of the values for female and male bars gives the total victims with known gender in each circumstance. The points represent the percentage of male and female population based on the U.S. Decennial Census: 2010.

The victims’ age ranges from babies (noted as 1 yr old in the dataset) to elderly (i.e., 93 yrs old). Ashley and Ashley [2008] found enhanced vulnerability for those in the 10-29 yr, and >60 yr categories when compared with the U.S. Census data in 2000. Boxplots of age distribution for

males and females in each fatality circumstance show that almost 50% of the males in vehicle-related fatalities were between 24 and 62 years old (Figure IV.11). This finding can be partly explained by the fact that 80% of licensed drivers are aged between 20 and 64 according to the Federal Highway Administration [FHWA, 2009]. The outside/open air circumstances are more likely to take the lives of younger people with a median age for females of 29 and just 22 for males. Many of the outside/open air events are a result of children who either investigate or even play in dangerous floodwaters; these fatalities could be largely preventable through educational campaigns [Ronan and Johnston, 2001]. Vehicle-related events tend to strike the middle-aged; outside/open air events are connected to the young; the permanent building events tend to impact the elderly. The median ages for female and male victims in this category are 65 and 53, respectively. Elderly people inside buildings may have limited mobility, be emotionally attached to their belongings, or be unaware of the dangerous situation, which makes them vulnerable to flash flooding events even in their own homes [Peacock et al., 1997b].

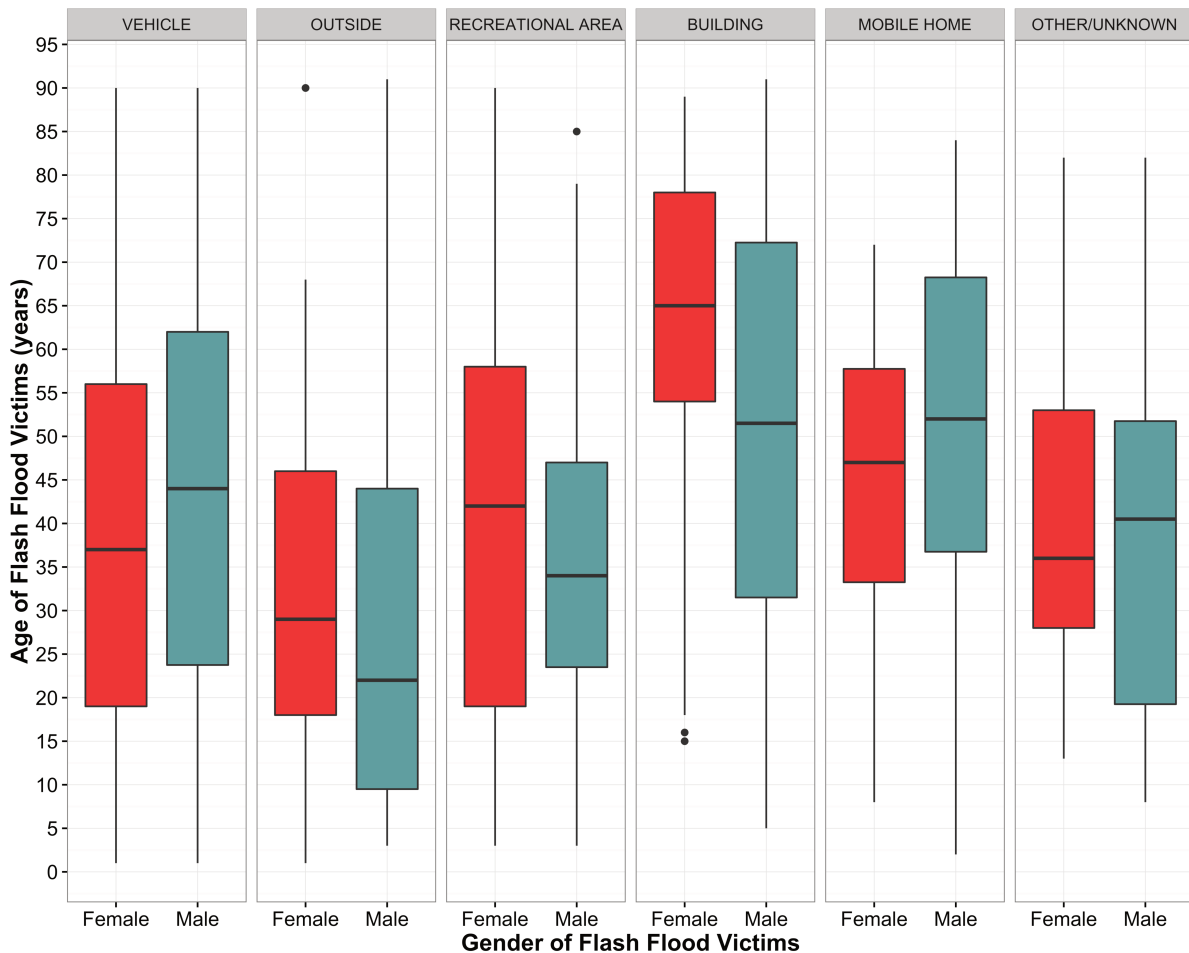
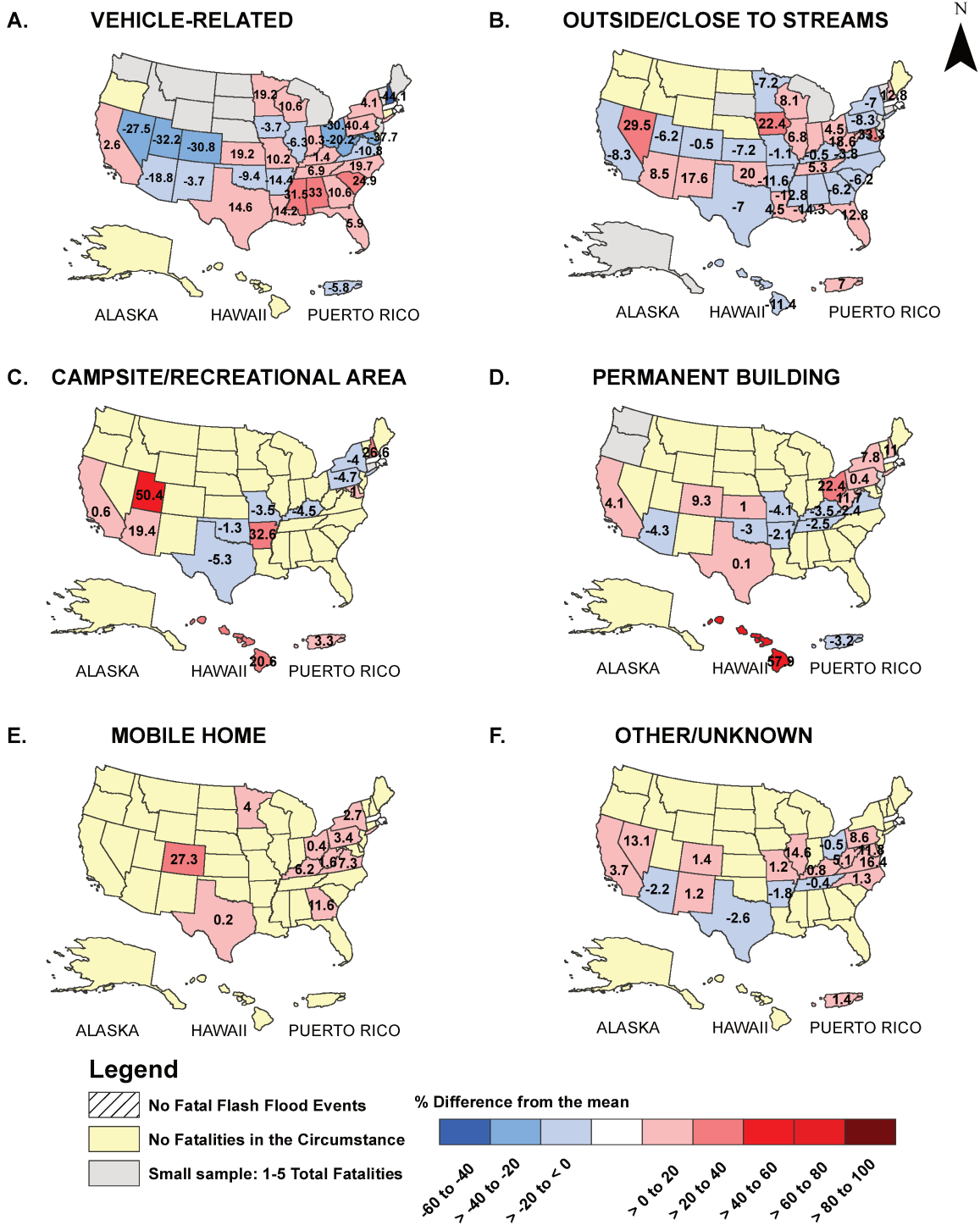


Figure IV.11 – Distribution of age by the gender of flash flood victims discretized by circumstance of the death. Boxplots are constructed for each circumstance separately for the victims with known both age and gender from 1996 to 2014 (n=1,003).

3 Spatial distribution of flash flood fatalities

From 1996-2014, there were fatalities reported in 49 U.S. states and territories with the exception of Rhode Island, Massachusetts, and the District of Columbia. The state-based analysis of vehicle-related fatalities reveals some patterns (Figure IV.12). First, there is a dearth of reports in a large swath of the intermountain West. The hotspot for vehicle-related fatalities extends from Texas eastward into the South reaching maximum positive anomalies in Alabama (33%) and Mississippi (32%). Central and South-central Texas holds the most extreme rainfalls (i.e., rates less than 48h) that led to some of the greatest flood peaks nationwide [O'Connor and Costa, 2004]. In the literature, the domination of vehicle-related flash flood fatalities at those areas have been mainly related to increases in exposure associated with rising population densities in urban areas with physiography susceptible to flash flooding (e.g., Flash Flood Alley) [Sharif et al., 2012]. While additional research is warranted on this topic, it is likely that this increased exposure combined with intense rainfall rates and the prevalence of low-water crossings extends vehicle-related fatality occurrences eastward across the South. As a general attitude of motorists, they may show incorrect confidence on driving through usually dry low-water crossings, and may underestimate the risk when being experienced in crossing successfully flooded intersections. The outside/open air circumstance in Figure IV.12B reveals no significant regional preferences, indicating risky behaviors such as playing in floodwaters, taking photographs, or cleaning out a drain are problematic on a national basis.

The sample sizes with the camping/recreational area events in Figure IV.12C are smaller, but there are very clear regions that are particularly vulnerable to flash flooding fatality events in recreational settings. Canyon hiking and camping in the states of Utah and Arizona claim the lives of many during the warm season. Most of these victims, several of whom are foreign, are not familiar with their environment and do not readily recognize a hazardous situation. This problem is exacerbated by the nature of the flash flooding events that can cause damage and impacts to areas well downstream from the causative rainfall. Many canyons have sheer, steep walls, making a quick escape very difficult. In 1997, ten hikers perished in Antelope Canyon in northern Arizona from a flash flood that reached a stage of 3.4 m. The sun was still out when they entered the canyon, but heavy rainfall was occurring in the headwaters of the basin. As the water rose, the curvy limestone walls that have been shaped by water over time were practically impossible to climb without a rope, trapping the hikers. The states of Arkansas and Hawaii also appear as being vulnerable to camping/recreational area flash flooding events. Arkansas stands out primarily due to the Albert Pike campground flood that killed 20 people on June 10, 2010 [Holmes Jr and Wagner, 2011]. These camping-related fatalities could be mitigated through more active local awareness activities, alerting systems, and escape routes (e.g., permanently placed ropes). There are no strong regional signals with permanent building or mobile home fatalities. However, the states of Hawaii and Ohio have large positive anomalies for permanent building fatalities, and Colorado stands out in mobile home fatality circumstances.



Difference between the percent of fatalities in each state and the percent of the fatalities in all U.S. for the fatalities related to a specific circumstance from 1996 to 2014.

Data source: Storm Data (1996-2014)
 Projection: Lambert Conformal Conic
 Author: G.Terti

Figure IV.12 – State-level flash flood fatalities for each circumstance (form A to F) represented as difference between the percent of fatalities in each state and the percent of the fatalities in all U.S. for the fatalities related to a specific circumstance. The number in each state indicates the exact value of the percent of difference for the states which had fatalities occurring in the depicted circumstance and refer to states with more than 5 fatalities in total from 1996 to 2014.

4 Summary and Conclusions

In this chapter, we investigated the circumstance of 1,075 flash flood-specific human losses from 1996-2014 on the scale of the U.S.. The compiled fatality database used for this analysis is now part of the unified flash flood database described in Gourley et al. [2013] and is publicly available at <http://blog.nssl.noaa.gov/flash/database/>. One recommendation coming out of this study is for the National Weather Service to consider classifying each flash flood fatality into the categories presented herein. The purpose of our analysis was to explore if different vulnerability paths occur depending on the situation, as determined by the victims' profile and activity, and the spatio-temporal context of the flash flooding. Indeed, we found that the circumstances associated with flash flooding fatalities have certain characteristics related to season, time of the day, duration of the flood, location, and tends to be associated with specific age and gender groups. Especially, the results suggest the following patterns:

- Most of flash flood victims were involved in vehicle-related circumstances followed by the “outside” incidents (almost 61% and 21% of the total 1,075 fatalities, respectively). In many cases, both circumstances were characterized by unnecessary risk-taking behavior on behalf of the victims. It occurs that an important amount of losses could be prevented if people had a different approach and decided to stay inside instead of being active and mobile during the flash flood event.
- In agreement with the spatio-temporal distribution of heavy precipitation, the majority of fatal flash flood events (65% of the 705 total fatal flash floods) and the related fatalities (65% of the total 1,075 fatalities) occurred in the warm season from May through September. From a social point of view, it is interesting to notice that outdoor circumstances such as OU and CA strongly contribute to the total monthly fatalities during this summer period, compared to the rest of the year where VE victims dominate with noticeably higher percentages.
- People in outdoor circumstances such as VE, OU and CA were more likely to drown in very fast and dynamic FF events with duration close to 5 hours. On the contrary, people in buildings and mobile homes were more likely to get threatened by longer flash floods.
- Outside/close to streams incidents were more likely during daylight hours associated mainly with young males (e.g., children and teenagers that were swept into creeks while playing near high waters). In general, younger and middle-aged people were more probable to get trapped while participating in daytime outdoor activities (e.g., VE, OU, CA), whereas older people were more likely to perish inside buildings, and especially in twilight and darkness hours when rescue and/or evacuation operations are hindered.
- In agreement with previous analysis of flood fatalities in U.S. [French et al., 1983; Ashley and Ashley, 2008; Kellar and Schmidlin, 2012; Sharif et al., 2014], the majority of flash flood victims from 1996 to 2014 were males (61% of the total 1,051 victims with known gender). Doocy et al. [2013] presented a historical review of flood fatalities (1080-2009) worldwide showing that in contrary to the developing countries, greater proportions of males compared to female fatalities are observed in most of the developed countries. In our analysis, it occurs that men were 2.6 and 1.4 times more likely than women to be involved in OU and VE fatal incidents, respectively.

We have conducted this analysis in preparation for more sophisticated and targeted alerting systems that will incorporate these socio-demographic characteristics. Future targeted alerts can be communicated when we can collocate the location of risky incidents in space (e.g., roads, campsites, mobile homes) with specific vulnerable groups (e.g., certain age groups, gender). The findings highlight the importance of situation-specific assessment of flash flooding fatalities to guide the development of flash flood-specific human risk modeling. In this direction, recording as much details as possible for the life-threatening scene in *Storm Data*, and especially placing emphasis on the profile and intentions of people involved, is of high importance for future methodological developments.

The next chapter will focus on a statistical classification model can be applied to obtain trends and patterns in the probability of a fatality to occur in certain circumstances. Based on the reclassified dataset supplemented with other variables describing the storm event, the spatial distribution and socio-demographics of the exposed population, and the exposed built environment (Chapter III, section 3), such a probabilistic approach serves as a promising method to quantify the time and space-dependent human risk factors using representative indicators. We expect that this human impact-based predictive approach will contribute to renewing alerting systems, making them more specific and effective in triggering timely preventive actions by the public, finally leading to a decrease in the trend of fatalities caused by flash flooding.

TOWARDS PROBABILISTIC PREDICTION OF FLASH FLOOD HUMAN IMPACTS

Chapter V: Machine learning predictions of flash flood human risk related to vehicles

Chapter V is based on a paper under review in *Risk Analysis*, Journal of Society for Risk Analysis: Terti, G., Ruin, I., Gourley, J.J., Kirstetter, P.-E., Flamig, Z., Blanchet, J., Arthur, A., and Anquetin, S., 2017. Towards Probabilistic Prediction of Flash Flood Human Impacts. *Risk Analysis*, in review.

Résumé

Les crues rapides frappent les populations de manière rapide et violente et révèlent plusieurs situations de vulnérabilité liées à des circonstances d'accidents différentes (e.g. en véhicules, à l'intérieur des bâtiments). La prévision des impacts humains lors des ces phénomènes extrêmes reste très difficile pour les prévisionnistes et les services de secours car les outils actuels ne tiennent pas compte des facteurs de vulnérabilité sociale. D'autre part, l'évaluation traditionnelle de la vulnérabilité considère rarement la dynamique de l'aléa et ses interactions avec la société en situation de crise. Le Chapitre V se concentre sur les développements conceptuels et méthodologiques permettant l'intégration des dynamiques physiques et sociales, conduisant à des modèles de prévision des impacts humains face aux crues éclair. Pour atteindre cet objectif, un classificateur de Random forest est appliqué pour évaluer la probabilité d'occurrence de décès, pour une circonstance donnée, en fonction d'indicateurs représentatifs. Dans un premier temps, on choisit la vulnérabilité liée aux véhicules, en effet l'analyse menée précédemment et la littérature indiquent que la plupart des décès en cas de crues rapides sont issus de cette catégorie (Chapitre IV). La méthodologie est développée à partir d'une base de données sur les crues rapides, avec ou sans perte humaine entre 2001 et 2011 aux États-Unis, complétée par d'autres variables décrivant, à l'échelle du comté, l'aléa (Chapitre III). Les inondations de mai 2015 dans les états du Texas et de l'Oklahoma sont utilisées comme étude de cas pour cartographier la dynamique du risque associé aux véhicules, à une résolution quotidienne et horaire pour chaque comté dans la zone d'étude. Les résultats indiquent l'importance de l'évaluation du risque humain, en fonction du temps et de l'espace, pour ces phénomènes. La nécessité d'une collecte plus systématique des impacts humains est également mise en évidence afin de faire progresser les modèles prédictifs basés sur les victimes des crues éclair, dans le but d'utiliser des approches d'apprentissage automatiques à l'avenir.

Abstract

Flash floods disrupt people's intended paths in a rapid and violent manner, revealing various vulnerability situations through the expression of different accidents' circumstances (e.g., vehicle-related, inside buildings). Forecasting human impacts from flash flooding remains very challenging for forecasters and emergency managers because today's forecasting tools do not consider social vulnerability factors. On the other hand, traditional vulnerability assessment rarely considers the dynamics of the hazard and their interactions with society in crisis situations. Chapter V focuses on conceptual and methodological developments allowing the integration of physical and social dynamics leading to model forecasts of circumstance-specific human losses during a flash flood. To reach this objective, a Random Forest classifier is applied to assess the likelihood of fatality occurrence for a given circumstance as a function of representative indicators. To begin with, vehicle-related circumstance is chosen as previous analysis and literature both indicate that most fatalities from flash flooding fall in this category (Chapter IV). The methodology is developed using a database of flash flood events, with and without human losses from 2001 to 2011 in the United States which has been supplemented with other variables describing the storm event, the spatial distribution of the sensitive characteristics of the exposed population and built environment at the county level (Chapter III, section 3). The catastrophic flash floods of May 2015 in the states of Texas and Oklahoma are used as a case study to map the dynamics of the estimated probabilistic vehicle-related human risk on a daily and hourly level for each county in the study area. The results indicate the importance of time and space-dependent human risk assessment for short-fuse flood events. The need for more systematic human impact data collection is also highlighted to advance impact-based predictive models for flash flood casualties using machine-learning approaches in the future.

Machine learning predictions of flash food human risk related to vehicles

1 Introduction

Technological advances in forecasting the potential for flash flooding have largely improved watch-warning systems during recent decades. Hydrometeorologists work on the challenging issue of modeling physical processes associated with the occurrence and magnitude of flash floods. A suite of hydrometeorological products operating at high spatiotemporal resolutions has been developed to support operational forecasters when issuing flash flood warnings in the U.S. [Gourley et al., 2017]. However, such advancements cannot yet address the occurrence of life-threatening situations emerging from the conjunction of the hazard, still difficult to predict, and social vulnerabilities that evolve in space and time.

Currently, social vulnerability modeling research is dominated by the construction of indexes summarizing social dependencies and economic disadvantages of the population in geographic units varying from block groups to states [Clark et al., 1998; Cutter et al., 2000; Cutter, 2003; Tapsell et al., 2002]. While there is a lot of research on analyzing flood impacts and understanding the underlying causes of social vulnerability to flood hazards [Jonkman and Kelman, 2005; Ashley and Ashley, 2008; Sharif et al., 2012; Doocy et al., 2013; Sharif et al., 2014], establishing specific vulnerability metrics remains rare. Being strongly influenced by pioneering studies [Clark et al., 1998; Cutter et al., 2000; Tapsell et al., 2002; Cutter, 2003], social vulnerability quantification in cases of flooding relies on either data-reduction techniques such as factor analysis [Rygel et al., 2006] or arithmetic methods such as standardization scores [Wu et al., 2002; Chakraborty et al., 2005; Wilhelmi and Morss, 2013] to compose indicator-based aggregated social vulnerability measures and maps. With these approaches, social vulnerability is treated separately and is then merged with the hazard information (provided

through flood hazard maps or scores) only as a final step to provide a static map of integrated socio-economical risk [e.g., [Wilhelmi and Morss, 2013](#); [Koks et al., 2015](#)].

Indicators are chosen based on theoretical knowledge (deductive approach) or data-driven analysis (inductive approach) whereas links with impact-related observations are rarely considered [[Zahran et al., 2008](#); [Fekete, 2009](#)]. [Zahran et al. \[2008\]](#) analyzed 832 flood events in Texas from 1997-2001 to explore the intersection of population vulnerability characteristics and aggregated flood casualties at the county level. Adopting a multiple regression analysis their study reveals that flood casualties are dependent on certain social vulnerability patterns. It was found that flood deaths and injuries in Texas are positively correlated with socially vulnerable populations, whereas they are reduced with the increase of structural and non-structural flood mitigation strategies in the exposed communities. Still, social vulnerability in that analysis is described in a static way in terms of racial minorities and economic status, inviting further research on the integration of more hazard and circumstance-specific vulnerability predictors.

Following the analysis presented in Chapter IV, the primary focus hereafter is the vehicle-related circumstance where the majority of people perish while inside their vehicle or are attempting to escape from a vehicle being swept away in flash flood waters [[Kellar and Schmidlin, 2012](#); [Sharif et al., 2012](#); [Diakakis and Deligiannakis, 2013](#); [Terti et al., 2016](#)]. An empirically-guided, predictive approach is adopted to estimate the likelihood of one or more vehicle-related fatality incidents to occur in a specific flash flood event given the conjunction of supplemented characteristics about the hydrometeorology of the event and the infrastructure and demography of the exposed county. Random forest (RF) [[Breiman, 2001](#)], a well-known decision-tree based ensemble machine-learning algorithm for classification and regression is adopted for this analysis. Tree-based models recursively split the data space into sub-spaces according to the behavior of a target variable. The succession of binary splits leads to a set of tree branches subdividing the data space into disjoint partitions of the target variable. The splits are selected to maximize the homogeneity or purity of the target variable in the leaves.

Such modeling is a powerful tool with recent, increasing use in hydrological and meteorological research. Classification tree analysis has been used in hydrograph analysis to identify the effect of various hydro-meteorological variables and certain thresholds on the type of catchment response [[Ali et al., 2010](#)], as well as in seasonal streamflow forecasting considering large-scale climatic predictors and their nonlinear interactions [[Wei and Watkins, 2011](#)]. [Clark \[2016\]](#) used machine learning models to forecast the probability of flash flooding given a set of atmospheric and hydrologic conditions in the conterminous U.S., and explored their applicability in operational forecasting. Recently, regression tree models have been further applied in assessing flood damage based on multiple variables describing the flooding hydrology and warnings, building characteristics and precaution measures, and the socio-economic status of private households [[Merz et al., 2013](#)]. Compared to other advanced statistical approaches such as logistic regression, random forest algorithm does not rely on any linear or other relationship between the input predictor variables and the target variable, and it is not sensitive to outliers, being able to handle nonlinear and complex high-order interactions [[Breiman, 2001](#); [Merz et al., 2013](#)].

Building upon prior theoretical and empirical knowledge this chapter addresses the following questions:

- i. How can social and physical proxy variables at the county level inform a circumstance-specific vulnerability metric available at temporal and spatial scales relevant to flash flood emergency response?
- ii. How to use historic fatal and non-fatal flash flood reports as the basis to quantify the relationship between the magnitude of the flash flood and proxies revealing the vehicle-related vulnerability of people at the time of the event?
- iii. How can human risk predictions be estimated and mapped dynamically to reveal the time-variant exposure to a given flash flood forecast?

The chapter consists of three main parts. First, we present the flash flood human impact data used to create the target variable in the analysis and the supplemented extra variables treated as vehicle-related risk predictors. In that part, our conceptual and methodological approach for flash flood specific human vulnerability is refined to reflect the occurrence of human losses in the vehicle-related circumstance. Then, section 3 describes the process to select certain independent predictor variables to insert in the random forest algorithm, and the performance of the final classifier on predicting flash flood events with vehicle-related human losses is assessed. Section 4 applies the built classifier for a series of flash flood events that occurred in Texas and Oklahoma during May 2015. This section presents a prototype towards vehicle-related risk prediction by providing dynamic maps at the county level. The final section discusses the achievements and limitations of the current work and proposes key future steps for the improvement of machine-learning based prediction of human risk to flash flood threat.

2 Methodology

2.1 Target variable for vehicle-related fatal flash flood events

In this study, the challenge is to differentiate flash flood events based on the occurrence or non-occurrence of vehicle-related fatalities. The data were prepared for binary classification as described in section 2.3 of Chapter III. The process for formulating the target variable in the vehicle-related circumstance is recalled in Figure V.1¹. The 10-year dataset includes 38,106 unique rows (i.e., flash flood events) the majority of which (99%) is labeled as “NO_EVENT” event. Each of the 38,106 flash flood events is attributed to a specific county in the conterminous U.S. Hereafter we called “exposed county”, counties where at least one flash flood event has been reported between 2001 and 2011. Over this period, 2,899 of the total 3,109 counties in the conterminous U.S. are concerned, with a mean of about 13 events per county and up to 224 events for the most exposed. The latter are

1. Pictures used in this illustration are available at online media. The picture on the top (“EVENT”) refers to a vehicle-related rescue conducted during the flash flood event of May 25, 2013 in Austin (Texas), and is available online at <http://www.desdemonadespair.net/2013/05/record-rainfall-causes-severe-flooding.html>. The picture on the bottom (“NO_EVENT”) illustrates flood waters in a mobile home park in Pelham, Ala (Alabama) on April 6, 2014 (photo gallery available at <http://illuminatimindcontrol.com/extreme-weather-severe-thunderstorms-tornado-and-flash-flooding-swamps-u-s-southeast-many-homes-damaged-child-found-dead-in-mississippi-1-dead-in-georgia-dozens-injured-more-stormy-weather-fo/>).

located in Arizona, southern California, southeastern and central Texas, and southwestern Missouri (Figure V.2). The 259 flash flood events that led to one or more vehicle-related fatalities from 2001 to 2011 are distributed in 204 counties mostly located in the southern U.S. (red hatch in Figure V.2).

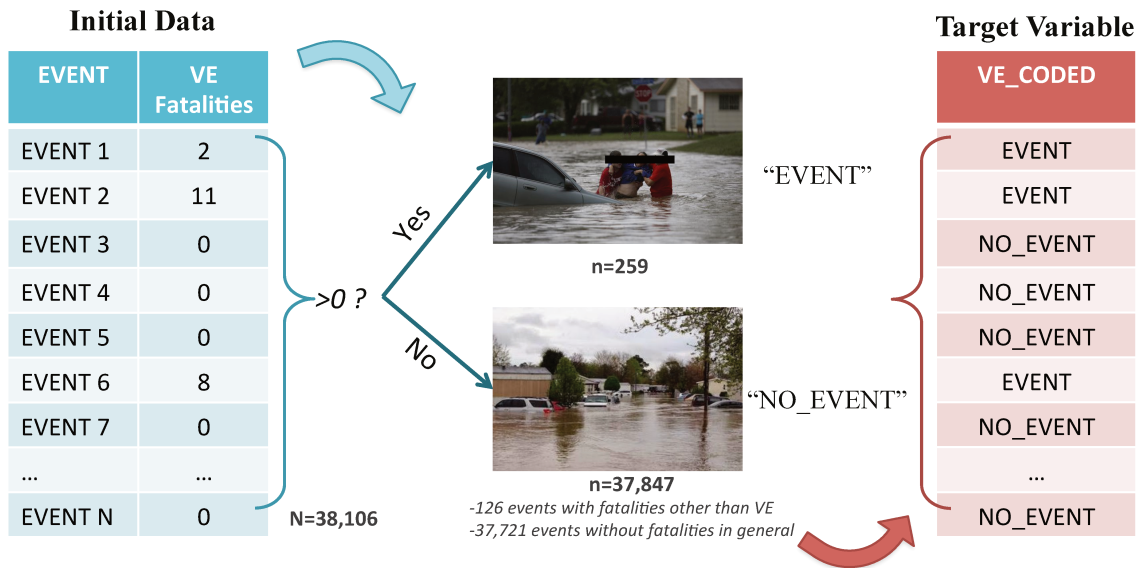


Figure V.1 – Schematic illustration of the process followed to formulate the target variable for the vehicle-related (VE) machine learning classification.

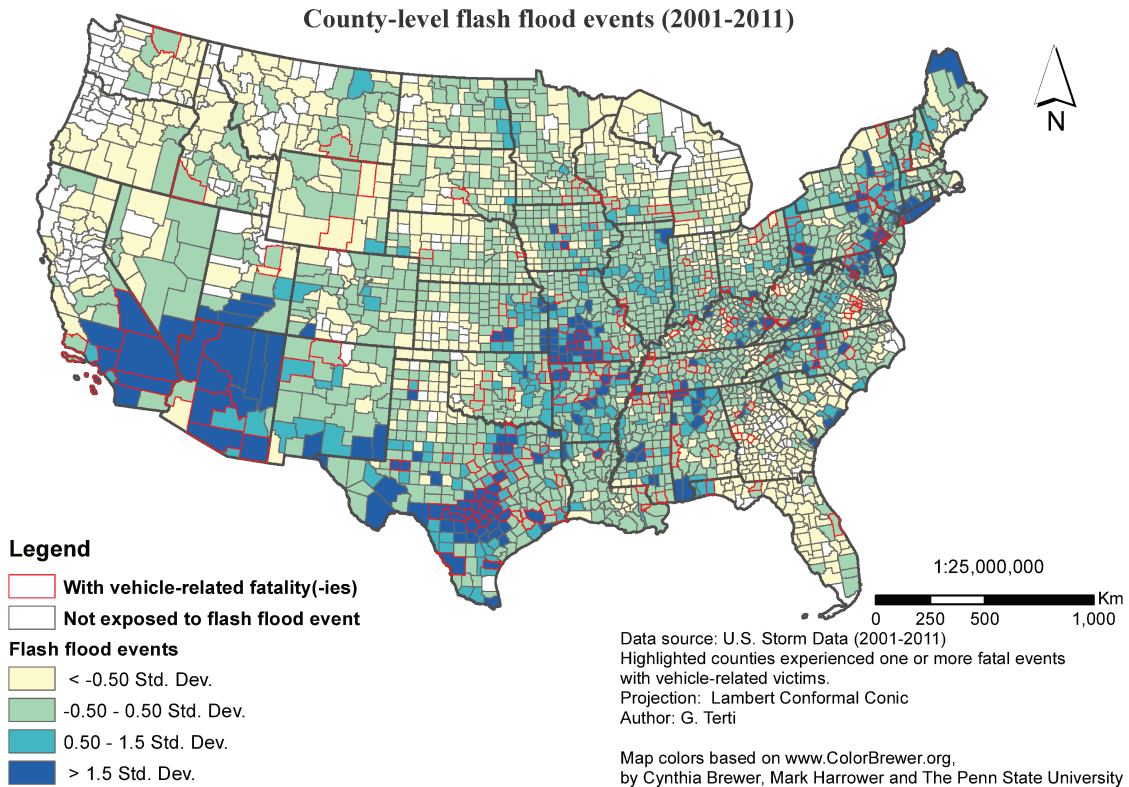


Figure V.2 – County-level distribution of flash flood events represented as variation from the mean (i.e., 13 events per county) for the period of 2001-2011. Counties with fatal flash flood events that according to the *Storm Data* led to one or more vehicle-related fatalities are highlighted with red solid line. Map colors based on <http://www.ColorBrewer.org>, by Cynthia A. Brewer, Penn State.

2.2 Candidate predictors of flash-flood specific human risk in the *vehicle* circumstance

Indicators relevant for the vehicle-related circumstance are chosen based on the interpretation of the data discussed in section 3 of Chapter III. Table V.1 is an excerpt of the total list of the gathered proxy variables included in the supplemented database, proposed for modeling the vehicle-related human losses at the county level. Details on the pre-processing and the contribution of each variable in depicting the vehicle-related risk situation in each exposed county are summarized in Table D.1 (Annex D).

In the U.S., private vehicle is the predominant transportation mode for work-related and other travels [NHTS, 2009]. Integrating a proxy representing the flow of commuters at the time of the event is crucial for this circumstance. Therefore, we combined indicators concerning the “time of flash flood occurrence” and the “time arriving at work” from Table B.1 (Annex B), to create a new indicator referred to as “commuters”. Based on the time of the simulated unit peak discharge for a given flash flood event, each reported flash flood was assigned to a 30-min time step interval. Each flash flood event was then supplemented with the number of workers that arrived at work by vehicle in the exposed county during the certain time interval that includes the occurrence of the flood peak. For evening and nighttime hours, the temporal resolution in the census data at the county level is reduced due to confidentiality reasons, thus leading to increasing bias during this period. To avoid further subjectivity in the analysis, the number of commuters assigned to the evening-night time events is kept constant, assuming that this would be the highest possible exposure of commuters for that event. Despite this limitation, this new variable enables a more realistic representation of the people exposed in vehicles during a flash flood event in the specific county. Commuting plays an important role in the overall vulnerability since work-related travels during a normal daily routine are more likely to be continued under adverse weather conditions in contrast to leisure trips that can be more easily rescheduled [Kilpeläinen and Summala, 2007; Cools and Creemers, 2013].

Table V.1 – Risk indicators and the related proxy variables to serve as candidate predictors for flash flood events with vehicle-related incidents. Details on the proxies sources and processing are additionally provided in Table D.1 of Annex D.

Variable	Units & Reference
<i>Indicator 1: Magnitude of the flash flood event</i>	
Daily unit peak discharge	$m^3 s^{-1} km^{-2}$ in the event and the exposed county
<i>Indicator 2: Duration of the flash flood event</i>	
Duration of the flash flood event	hours
<i>Indicator 3: Magnitude of the rainfall event</i>	
Maximum accumulated precipitation	mm in the exposed county at the reported day(s)
<i>Indicator 4: Duration of the rainfall event</i>	
Maximum duration of precipitation	hours (for MRMS precipitation >1.0 mm in the day of the reported flash flood event)

Table II.1 Continued

Variable	Units & Reference
<i>Indicator 5: Flood hazard areas</i>	
Area of high risk of flood hazard	km^2 and % to the total area of the exposed county
Area of moderate-to-low risk of flood hazard	county
<i>Indicator 6: Flood severity</i>	
Mean flashiness	<i>index</i> (value between 0-1) in the exposed county
<i>Indicator 7: Official emergency service</i>	
Number of Local Emergency Operation Centers (EOCs)	<i>count</i> in the exposed county
<i>Indicator 8: Distribution of population</i>	
Daytime population density	$people/km^2$ in the exposed county
<i>Indicator 9: Road network</i>	
Road length	km in the exposed county
Road density	km/km^2 of the exposed county
<i>Indicator 10: River-road network intersections</i>	
Number of river-road crossings	<i>count</i> in the exposed county
<i>Indicator 11: Age</i>	
People:	
-14 years or under (youth)	
-15 to 34 years (new drivers and young adults)	<i>count and %</i> to the total residential population of the exposed county
-35 to 59 years (middle-aged active adults)	
-60 years or over (retired and elderly)	
Median age of residents	<i>years</i> in the exposed county
Median age of workers	<i>years</i> in the workplace exposed county
Median age of workers commuting by vehicle	<i>years</i> in the workplace exposed county
<i>Indicator 12: Gender</i>	
Males	<i>count and %</i> to the total population of the exposed county
<i>Indicator 13: Household family status</i>	
Average household size	<i>persons per household</i> in the exposed county
Number of family households (i.e., families)	<i>count and %</i> to the total number of households in the exposed county
Number of single-parent families (i.e., with either male or female householder)	

Table II.1 Continued

Variable	Units & Reference
<i>Indicator 14: Educational attainment</i>	
People educated with less than 9th grade	<i>count and %</i> to the total population 25 years and over in the exposed county
People graduated from high school or equivalent	
<i>Indicator 15: Ethnicity/citizenship</i>	
Number of foreign, not U.S. citizen commuters by private vehicle (drove alone or carpooled)	<i>count</i> in the workplace exposed county
<i>Indicator 16: Language</i>	
Number of people who speak other than English languages at home, and speak English less than “very well”	<i>count</i> for population over 5 years in the exposed county
Number of commuters by private vehicle who speak other than English languages at home, and speak English less than “very well”	<i>count</i> for workers 16 years and over in the workplace exposed county
<i>Indicator 17: Vehicles</i>	
Aggregate number of vehicles	<i>count</i> for the total households in the exposed county
Aggregate number of vehicles used in commuting	
<i>Indicator 18: Travel time to work</i>	
Number of commuters who are traveling to go to work from 5 to 90 or more minutes estimated in 11 classes (e.g., 5 to 9 min, 10 to 14 min, . . . , 60 to 89 min, more than 90 min)	<i>count</i> in the exposed county
<i>Indicator 19: Commuters</i>	
Number of commuters that arrive at work by vehicle at the time of the peak discharge	<i>count</i> for workers 16 years and over who do not work at home in the workplace exposed county

2.3 Classification method

Random forests grow many binary classification trees that may be weak classifiers by themselves. These are combined with the ultimate goal of obtaining a learner with higher accuracy [Dietterich, 2000]. Data consists of a given training set $(X, Y) = \{(X_1, Y_1), \dots, (X_N, Y_N)\}$ with N independent observations (e.g., flash flood events). The vector X_j is composed of p input predictors $(X_j^1, X_j^2, \dots, X_j^p)$ where $X_j \in \mathbb{R}^p$, and Y_j is the target variable that we are trying to classify or understand (i.e., “EVENT” or “NO_EVENT”). Breiman [2001] defines random forests as:

a classifier that consists of a collection of tree-structured classifiers $\{h(X, \Theta_k), k = 1, \dots\}$ where Θ_k is a random vector generated for the k_{th} tree independent from the past random vectors $\Theta_1, \dots, \Theta_{k-1}$ but with the same distribution, and each tree gives a unit vote for the most popular class at input X .

Each tree in the forest is grown with additional splitter variables until all terminal nodes of the tree (also called leaves) are purely one class or the other (Figure V.3). When the data reach an internal node, they follow one of the two branches if the value X^p of the predictor used at that node is greater than some threshold or the other branch if X^p is less than or equal to the same threshold. The main principle of random forests is randomization that is applied in two levels: i) each tree in the ensemble forest is built from a new training sample drawn randomly with replacement (i.e., a bootstrap sample) from the N cases in the original training set (X, Y) , and ii) the split in each node during the construction of the tree is the best split of a random subset m_{try} of all variables ($m_{try} < p$) [Liaw and Wiener, 2002].

As a result of the inherent randomness, the individual trees are almost independent. Bootstrapping makes the ensemble less immune to changes in data and avoids overfitting [Touw et al., 2012]. It also allows for an internal validation during the model training. As the forest is built on training data, each tree is tested on the samples not used in building that tree. Similar to a validation set, the predictions on the data points not included in the bootstrap sample (called “out-of-bag” or OOB sample) are aggregated and the error rate is thus estimated (OOB error) [Hastie et al., 2009]. The predictions of the trees in the final forest are aggregated using a dataset that is independent from the training sample. Each tree provides, for instance, a classification for each new flash flood event depending on where it lands in the tree. At the end, the random forest algorithm retains the classification having the most votes (over all the trees in the forest). Probabilities of a vehicle-related fatality are computed through the total number of votes. In our case, a probability threshold of 0.5 is used as a dichotomous event vs. non-event definition.

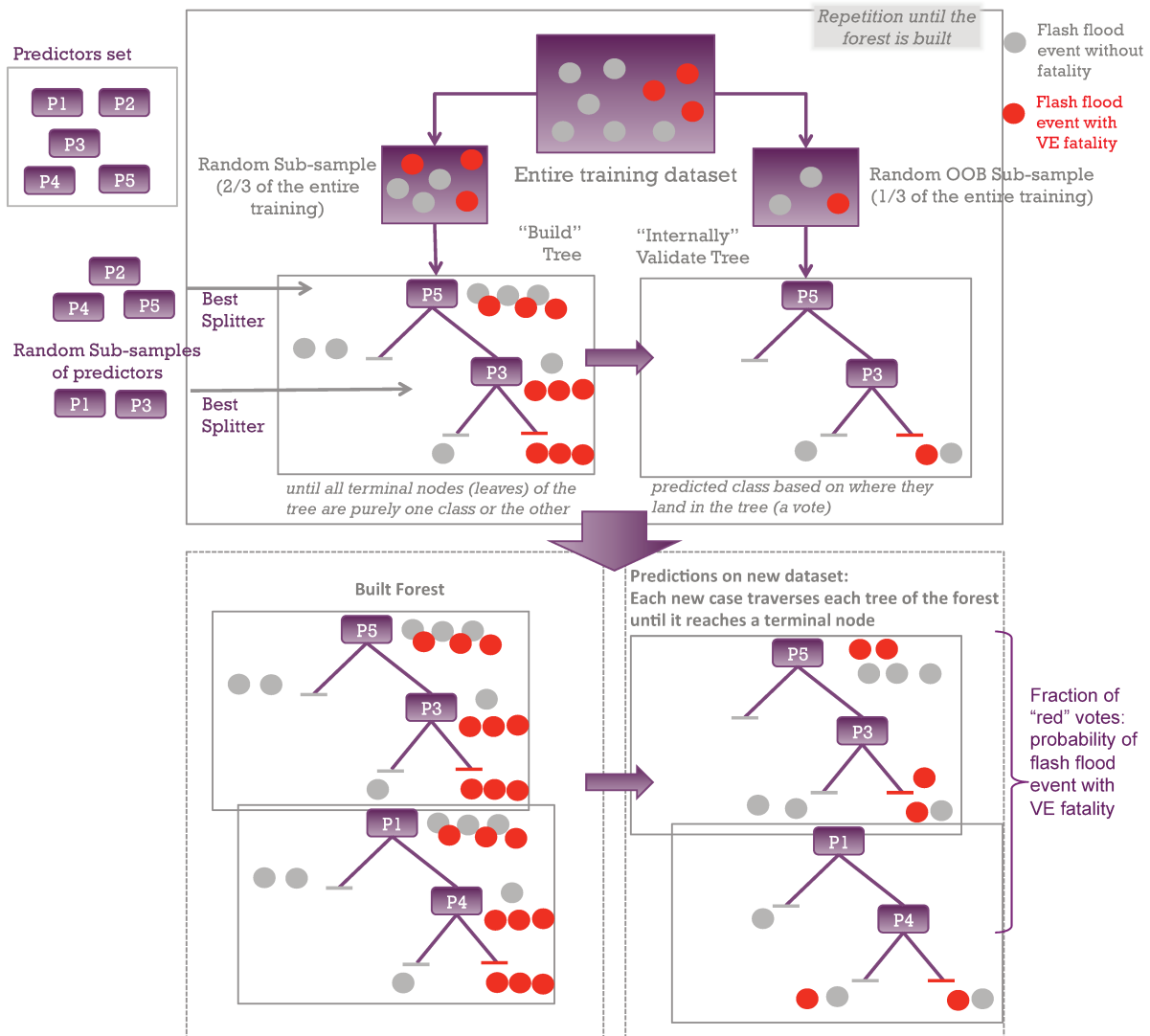


Figure V.3 – Schematic illustration of the random forest algorithm using as example binary classification for the occurrence or non-occurrence of vehicle-related (VE) fatality in flash flood events.

3 Random forest model for vehicle-related casualties

3.1 Selecting candidate predictors

Both conceptual and data-driven criteria need to be considered when selecting proxy variables as relevant predictors in modeling the occurrence of vehicle-related flash flood fatalities. First, the proxies proposed in Table D.1 were screened for collinearity. Correlation between some of the proxies is expected either because the variables are compositional or because they are different expressions of the same underlying processes, though hidden dependencies may also exist. For the variables for which percentages could be estimated (Table D.1), a two-sample Kolmogorov-Smirnov (KS) statistic test was performed [Conover, 1971] to compare the cumulative distributions of the proxies in the two datasets (i.e., fatal flash floods with vehicle-related fatalities/non fatal flash floods). It was found that percentages did not present good behavior in terms of separated distributions between the “EVENT” and “NO_EVENT” classes, so they were excluded from the analysis.

Although the random forest algorithm does not suffer from multi-collinearity issues, redundant variables complicate the evaluation of the effect of each variable to the target variable [Breiman, 2001; Dormann et al., 2013]. To detect and remove dependent variables, the variance inflation factor (VIF) is computed as: $VIF = 1/(1 - R^2)$ where R is the multiple correlation coefficient resulting from regressing linearly a predictor variable against all other predictor variables¹ [Dormann et al., 2013; Naimi et al., 2014]. VIF equal to 1 indicates no collinearity, whereas increasing values (> 1) entail increasing correlation between the variables. The procedure is described as follows: i) compute VIF for all the 41 variables from Table D.1, excluding the one with the highest VIF ; ii) repeat the stepwise procedure until no variables with VIF greater than 2 remain [Zuur et al., 2010]. At the end, 12 variables are found and kept for further analysis (Table V.2).

Table V.2 – Variance Inflation Factor (VIF) for the proxy variables with $VIF < \text{threshold}=2$. The variables are sorted from the ones with the least to the ones with the most variance explained by the other predictor variables in the regression.

Proxy variable	VIF
1. Mean flashiness	1.06
2. Area of moderate-to-low risk of flood hazard	1.10
3. Median age of workers commuting by vehicle	1.15
4. Maximum duration of precipitation	1.15
5. Daily unit peak discharge	1.18
6. Average household size	1.18
7. Area of high risk of flood hazard	1.31
8. Number of local emergency operation centers	1.33
9. Daytime population density	1.41
10. Number of river-road crossings	1.57
11. Road density	1.82
12. Number of commuters that arrive at work by vehicle at the time of the peak discharge	1.88

The Spearman rank correlation coefficient (r_s) illustrated in Figure V.4 accounts for monotonic (possibly non-linear) relationships between the variables. High Spearman correlations suggest either strong non-linearity between the variables or strong outliers in the data. Pairwise Spearman correlations indicated that daytime population density was highly correlated with the commuters ($r_s = 0.85$) that arrived at work close to the peak discharge time and the road density ($r_s = 0.81$), where the latter two were also correlated to each other ($r_s > 0.6$). These three variables represent similar exposure aspects in the vehicle-related vulnerability assessment. Therefore, we decided to keep only the number of commuters estimated at the hydrologic peak time as input for the random forest model. This variable is the most dynamic and flash flood-specific one compared to the other two correlated variables, and their distributions in the “EVENT” and “NO_EVENT” classes present an adequate distinction (p-value <0.05 in the KS test). Similarly, the area of moderate/low risk of the flood hazard variable is excluded as being highly correlated with the high-risk area variable. The final set of candidate predictors consists of nine uncorrelated variables that are standardized as $(\text{number} - \text{mean})/\text{standard deviation}$ to avoid scale effects (Table V.3).

1. The functions “vifcor” and “vifstep” from “Uncertainty Analysis for Species Distribution Models” package (usdm-package) built by Babak Naimi for R interface were used for this analysis. Documentation of the usdm-package is available at <https://cran.r-project.org/web/packages/usdm/usdm.pdf>.

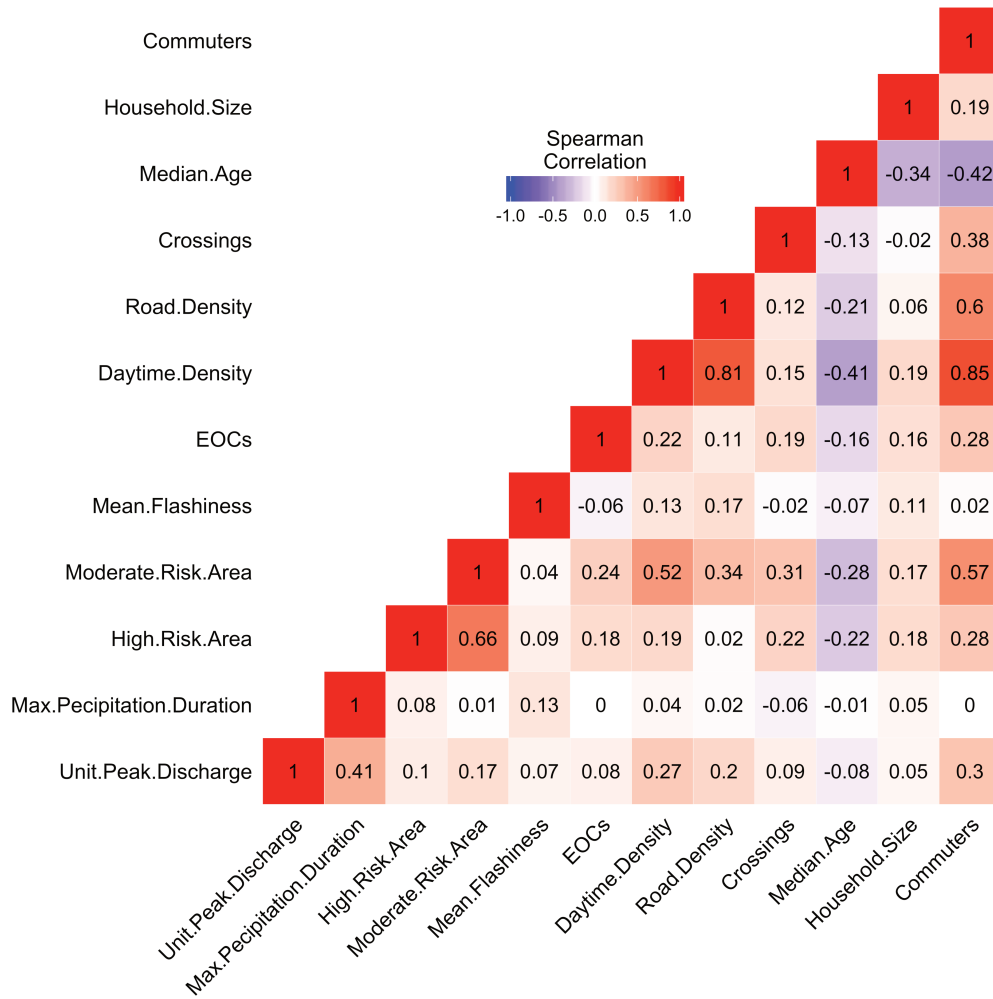


Figure V.4 – Spearman’s rank correlation for the 12 variables extracted after VIF analysis (see Table III for the meaning of the variables).

Table V.3 – Description of the nine candidate predictors.

Proxy variable	Range	After standard-ization
1. Daily unit peak discharge	$2.2 \cdot 10^{-4}$ -99.3 ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$)	-0.7-13.8
2. Maximum duration of precipitation	0.0-23.9 (hours)	-1.6-3.9
3. Area of high risk of flood hazard	0.0-5,037.9 (km^2)	-0.7-18.3
4. Mean flashiness	0.3-1.0 (index)	-5.0-2.0
5. Number of local emergency operation centers	0.0-95.0 (count)	-0.5-19.0
6. Number of river-road crossings	6.0-17,209.0 (count)	-0.7-7.9
7. Median age of workers commuting by vehicle	29.4-59.2 (years)	-4.6-7.2
8. Average household size	1.9-4.0 (persons per household)	-2.7-6.0
9. Number of commuters that arrive at work by vehicle at the time of the peak discharge	0.0-471,245.0 (count)	-0.3-27.6

3.2 Model training

The set of 38,048 flash flood events was sampled randomly to provide a training set of 28,536 observations. The remaining 25% of the total events comprises a separate test set, only used to assess the prediction performance of the final model. The main statistical problem is the extreme imbalance between the frequency of flash flood events with vehicle-related fatalities and non-fatal flash flood events (i.e., 1% of class “EVENT” and 99% of class “NO_EVENT”) from 2001 to 2011. A common approach to deal with severe class imbalance in which the main interest is to forecast the rare class is undersampling [Yap et al., 2014]. In this study, the two classes in the training set are randomly subset so that their class frequencies match the least prevalent class (i.e., 199 cases of “EVENT” in the training set) while maintaining the true ratio between classes in the test set. To increase variability of the subsamples, undersampling was thus repeated for 20 bootstrap resamples (steps A to C in Figure V.5). One-thousand trees were built for each random forest in each bootstrap iteration according to the process described in section 2.3 [Liaw and Wiener, 2002]. The algorithm compares five variables, selected randomly from the total number of nine predictor variables, to identify the best splitter in each node. The optimal model is selected to maximize the Area Under the Receiver Operating Characteristic (ROC) curve (AUC) across the resamples¹ [Mason and Graham, 2002] (see processes in step C in Figure V.5).

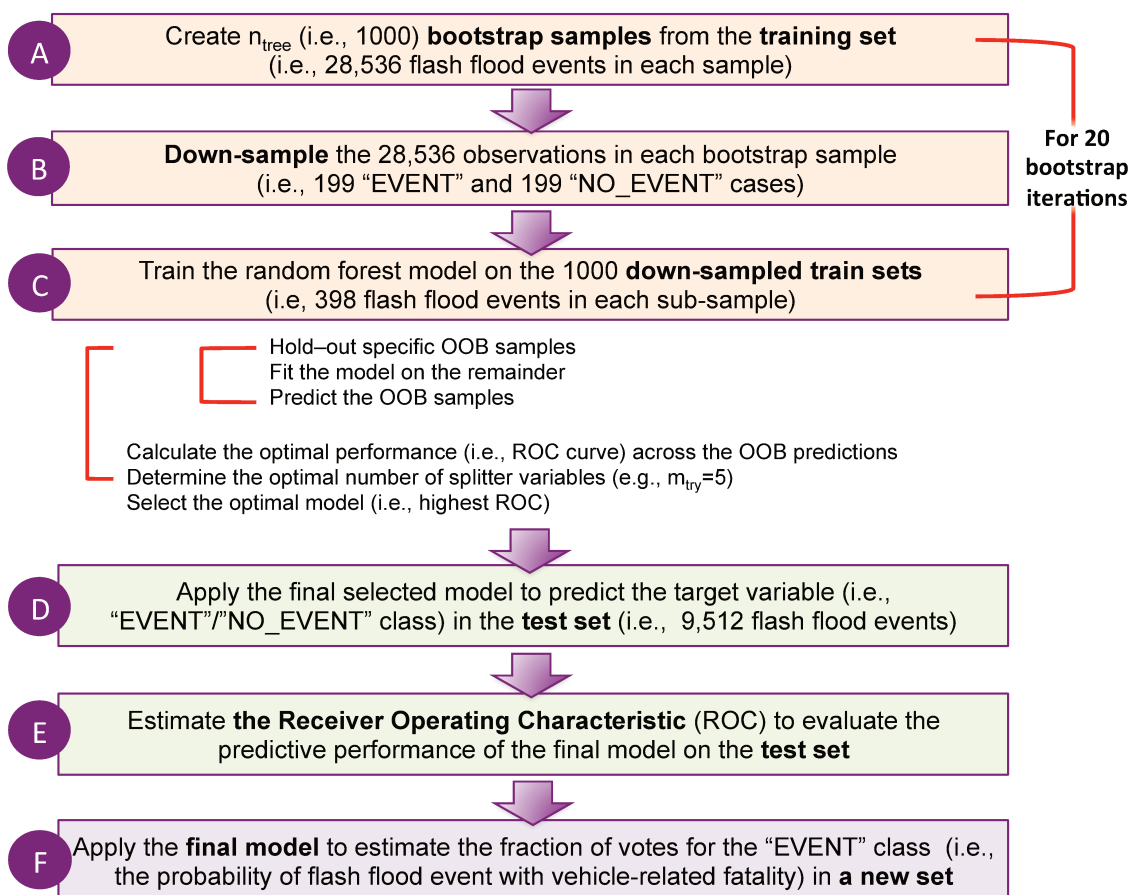


Figure V.5 – Modeling steps (from A to F) for training (orange boxes), testing (green boxes), and applying (purple box) the random forest classifier for vehicle-related fatalities prediction.

1. The R package about random forests used in this study is described in Liaw and Wiener [2002]. The randomForest package provides an R interface for the Fortran programs developed by Breiman and Cutler, available at http://www.stat.berkeley.edu/~breiman/RandomForests/cc_papers.htm.

3.3 Model performance

The internal evaluation of the final random forests model shows that the OOB error is about 39%. There is no typical value to evaluate the OOB error rate since it totally depends on the training data and the model. Class probabilities are estimated for the independent test dataset comprised of 60 flash flood events in “EVENT” class and 9,452 “NO_EVENT” cases (step D in Figure V.5). The model performance is quantified based on the AUC, estimated as equal to 0.7 for this classifier [Robin et al., 2011] (i.e., step E in Figure V.5). An AUC value of 0.5 corresponds to random guessing (i.e., the diagonal line on the ROC curve) and a value of less than 0.5 indicates discrimination worse than random chance (Figure V.6).

While the predicted probability is a continuous value between 0 and 1, it is often desirable to provide a binary prediction of whether the event will or will not occur to better understand the performance of the binary classifier. The perfect model would be pointed in the left upper corner of the ROC area where both the sensitivity $P(\hat{Y} = EVENT|Y = EVENT)$ ¹ and the specificity $P(\hat{Y} = NO_EVENT|Y = NO_EVENT)$ are equal to 1. ROC curve illustrates the performance of the classifier system as its discrimination threshold is varied. The end-users can then decide what is the best trade off between the hit rate and false alarms. Figure V.6 shows that for a 0.5 probability cutoff the model classifies correctly the 73% “EVENT” and the 62% “NO_EVENT” cases of the test dataset². If hit rate (i.e., sensitivity) and false alarm (i.e, 1-specificity) have the same importance, for example, then the best cutoff probability minimizes the Euclidean distance between the ROC curve and the upper left corner of the graph which in our case is close to the 50% probability threshold (blue point in Figure V.6) [Robin et al., 2011]. Forecasters and decision makers can further decide if they prefer to maximize the hit rate at the cost of increasing false alarms when issuing warnings for flash flood risk related to vehicles. In other words, they may select to warn and respond to vehicle-related threats when the modeled probability of vehicle fatality exceeds 40%. According to Figure V.6, for this threshold the random forest classifier assigns class “EVENT” when the predicted probability is > 0.4 and by doing so, it classifies correctly the 87% of “EVENT” of the test dataset (red point in Figure V.6). However, the probability of no impact events to be classified as events with vehicle fatality in the test dataset is also increasing to 0.57.

1. Sensitivity or probability of detection is the conditional probability of the predicted target variable (\hat{Y}) to be “EVENT” given that the observed class was “EVENT”.

2. The functions “roc”, “plot”, and “coords” from “Display and Analyze ROC Curves” package (pROC-package) built by Xavier Robin and co-authors for R interface were used for this analysis. Documentation of the pROC-package is available at <https://cran.r-project.org/web/packages/pROC/pROC.pdf>.

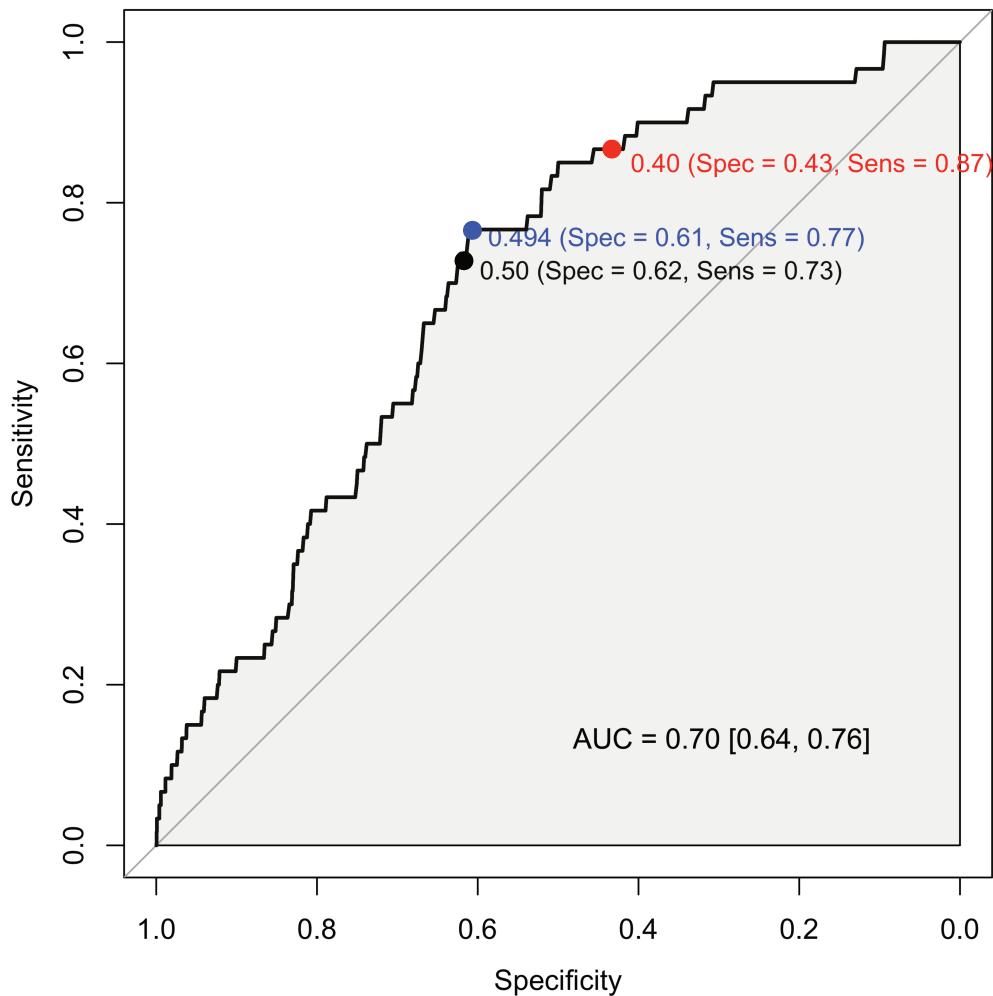


Figure V.6 – Receiver Operating Characteristic (ROC) curve for the final random forest model. Sensitivity in y-axis corresponds to true positive rate (TPR) known also as hit rate. X-axis demonstrates specificity known also as True Negative Rate (TNR). False alarm cases in forecasting are estimated as 1-Specificity. The area under the curve (AUC) is shaded in light grey. The 95% Confidence Intervals for the estimated AUC are computed with 2000 stratified bootstrap replicates as defined by [DeLong et al., 1988]. The cutoff probabilities of 0.5 and 0.4, and the corresponding specificity and sensitivity values, are illustrated in black and red font, respectively. The best cutoff probability (closest top left point) for this model is estimated as 0.494 (blue point on the curve).

3.4 Variables importance

The Mean Accuracy Decrease and Mean Decrease Gini through the training process is used to get information about the contribution of each candidate predictor to the model [Genuer et al., 2010]. In every tree grown in the forest, OOB samples are used to measure prediction accuracy. After the values of each variable are permuted randomly, the OOB accuracy is re-estimated and subtracted from the OOB accuracy based on the corresponding original variable. The differences for each variable are averaged over all trees in the forest and normalized by the standard error to provide permutation importance or “Mean Decrease Accuracy” of the variable. In Figure V.7A, household size, emergency operation centers and median age appear to be the least important for the estimated model. The

more the accuracy of the random forest decreases due to the exclusion (or permutation) of a single variable, the more important that variable is ranked.

Every time a predictor variable is used to split a node, the Gini index for the two descendent nodes are calculated and subtracted from that of the original node. The Gini decreases are summed and normalized at the end of the calculation to provide a measure of node impurity for each variable. Figure V.7B shows that peak unit discharge plays the most important role in partitioning the flash flood events in events with and without fatalities. This variable and the maximum precipitation duration are both dynamic and describe the magnitude of the natural hazard. The mean basin flashiness, high risk area, and crossings variables describe the hydrological response of the land surface to heavy rainfall and flooding as it relates to social impacts. The flashiness variable is statistically modeled across the U.S. and is based on observed streamflows and geomorphological variables [Saharia et al., 2016]. The high-risk area is a variable that represents flood prone areas based on FEMA’s flood zones. The crossings variable represents potentially dangerous low-water crossings. The commuters, median age and household size describe the social vulnerability variables of the model.

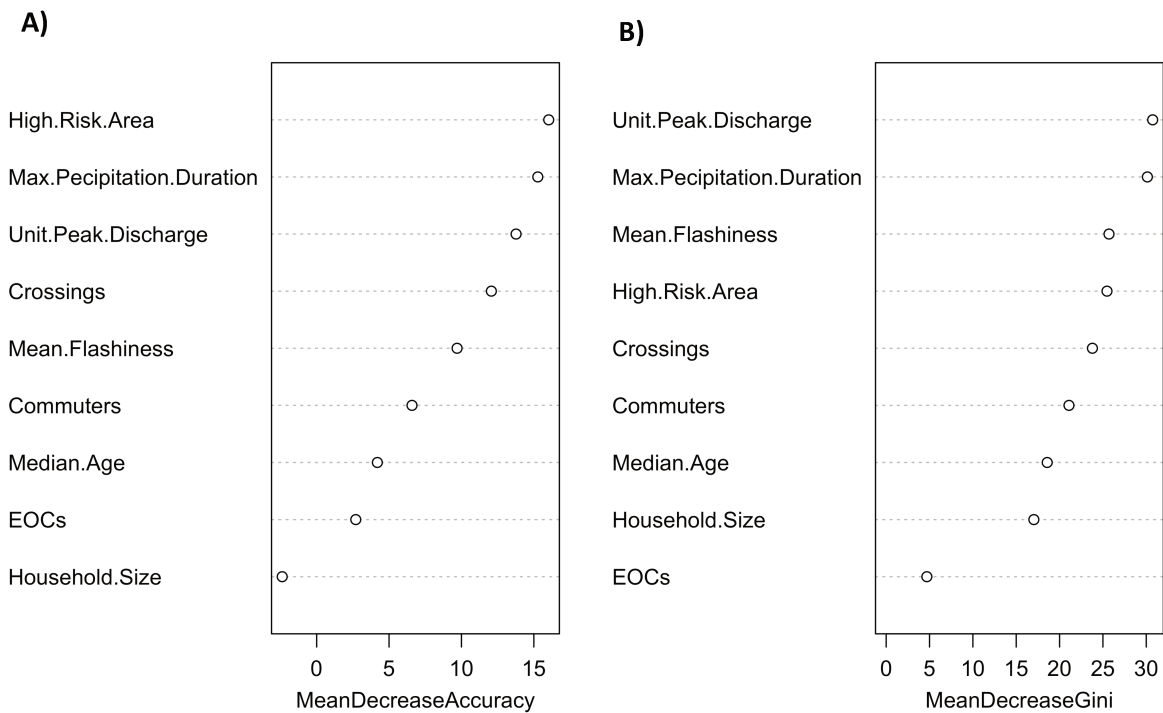


Figure V.7 – Rank of features by variable importance based on A) the out-of-bag decrease of accuracy and B) the out-of-bag decrease of Gini index.

The number of emergency operation centers (EOCs in Figure V.7B) is the weakest variable in terms of node purity and thus, it could be removed from the model. To evaluate the effect of the least important variables on the predictions, we estimated new models by removing, one-by-one, all of the last three variables of Figure V.7. The model without the EOCs variable is almost indicative of the full model in terms of AUC and sensitivity estimated based on the test dataset (Table V.4). When the “household size” and “median age” variables are additionally excluded, the ability of the

model to predict the probability of flash flood events with vehicle-related fatalities in the validation set is slightly reduced.

Even if some variables do not constitute very strong predictors, it appears that considering all possible interactions between them may lead to a better model for vehicle incidents in flash flood events. Since the number of variables is not large enough to cause increase in the experimental run time, the 9 predictors are all kept in the final model. In the following section, the final model will be applied to a new set of flash flood events that occurred in May 2015 in the conterminous U.S., and therefore they are independent of the training and testing datasets used during the model building procedure (step F in Figure V.5). This study serves as a first step towards visualization and evaluation of the outputs produced by the model, exploring its strengths and weaknesses in predicting the probability of vehicle-related casualties in counties exposed to flash flooding.

Table V.4 – Predictive performance of alternative models on the test dataset. Full model is the selected optimal model including all of the nine predictors (see Table V.3). Additional models are built by removing one by one the least important predictors. The reduced model includes six predictors (i.e., EOCs, household size, and median age predictors are excluded).

Random Forest model	AUC	Specificity	Sensitivity
Full model (9 predictors)	0.703	0.620	0.733
Model without <i>EOCs</i> (8 predictors)	0.700	0.620	0.733
Model without <i>Average household size</i> (8 predictors)	0.691	0.614	0.717
Model without <i>Median age</i> (8 predictors)	0.703	0.620	0.700
Reduced model (6 predictors)	0.694	0.621	0.633

4 Building a prototype: May 2015 flash floods

May 2015 was the wettest May on record and the all-time wettest month in 121-years of record in the conterminous U.S.¹. From the West to the East, about 37 states including California, Nevada, Arizona, Colorado, Kansas, Oklahoma, and Texas experienced flash flooding, to name but a few. Fortunately, despite the severe flood damages across the flooded states, fatalities were limited to three of them: Louisiana, Oklahoma, and Texas². Oklahoma and Texas each had their wettest month of any month on record with precipitation totals more than twice the long-term average. The Oklahoma Mesonet measured 367 mm of May total rainfall averaged over the state compared to the prior monthly statewide record of 273 mm observed in October 1941 (Figure V.8). In Texas the average accumulated rainfall in May 2015 (230 mm) also exceeded the previous record wet month of June 2004 (170 mm) [Nielsen-Gammon, 2015]. Those states were hit by major flash flooding, leading to massive human losses, especially at the second half on the month.

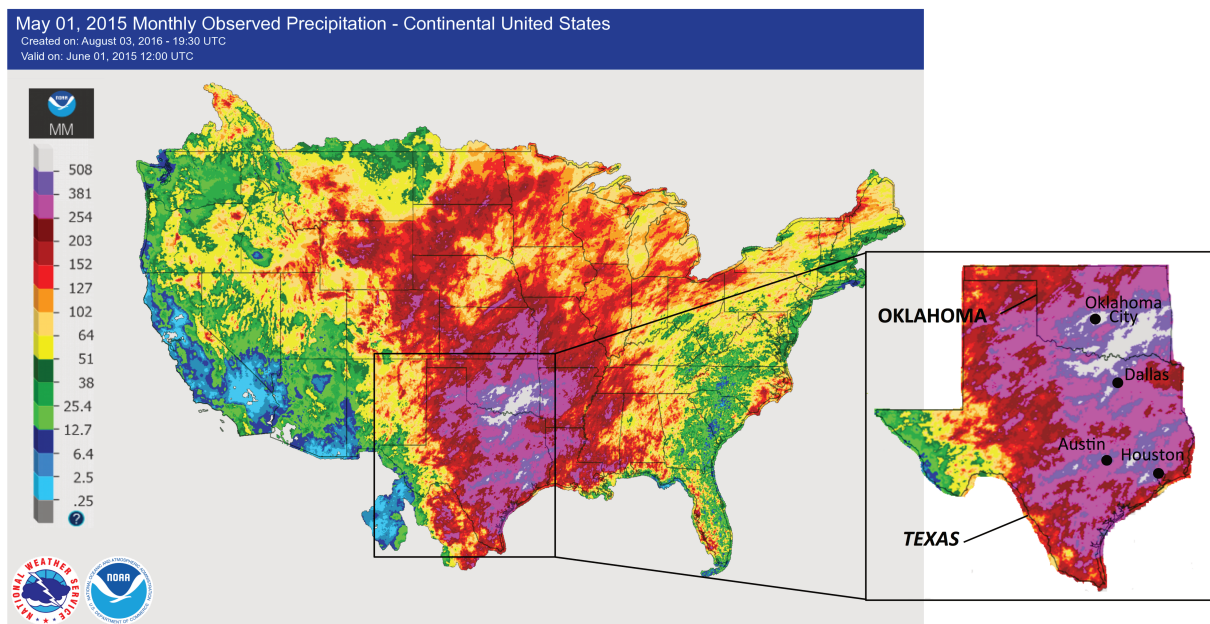


Figure V.8 – Monthly observed precipitation (mm) for May 2015 in the conterminous U.S. estimated by NOAA. Source: Advanced Hydrologic Prediction Service, NWS. Available at <http://water.weather.gov/precip/>.

The model built in section 3.2 is used to estimate the probability of vehicle-related fatality occurrence during the flash flood events of May 2015 in the conterminous U.S. The hydrologic conditions were described on a daily basis by extracting simulated unit peak discharge in each county, accessible from May 16 and after. Values for all predictors listed in Table V.3 were extracted for each of the 3,109 counties in the conterminous U.S.³. The model was then applied for each day from May 16 and after to compute daily the “EVENT” class probabilities providing a probabilistic assessment of vehicle-related human risk in each U.S. county (step F in Figure V.5). Figure V.9 presents the maximum value of the daily probabilities produced by the random forest classifier in each county in

1. Climatological rankings according to the NOAA, available at <https://www.ncdc.noaa.gov/temp-and-precip/climatological-rankings/>.

2. Considering the conterminous U.S., the *Storm Data* for May 2015 report 1 fatality in Louisiana, 4 fatalities in Oklahoma, and 30 fatalities in Texas.

3. Enumeration based on the Census Bureau 2010.

the 16 days of simulations. We should recall that the random forest classifier was trained on reported flash flood events¹. As a result, when new cases, probably with mild hydrological responses, traverse each tree of the forest, they may reach a terminal node characterized as “EVENT” because of the dominance of other static predictors (e.g., river-water crossings) in the county. Thus, a threshold of daily unit peak discharge will be set to deal with high predicted probabilities in days with low daily unit peak discharge in the continue of the study (see subsection 4.2). In the next sections we focus on Oklahoma and Texas which encountered the majority of human impacts and especially, vehicle-related fatalities, to discuss the model outputs and explore the dynamics of vehicle-related human risk from day-to-day. Then, the case of flash flooding on May 26, 2015 will be used to study hourly predictions of casualties related to vehicles.

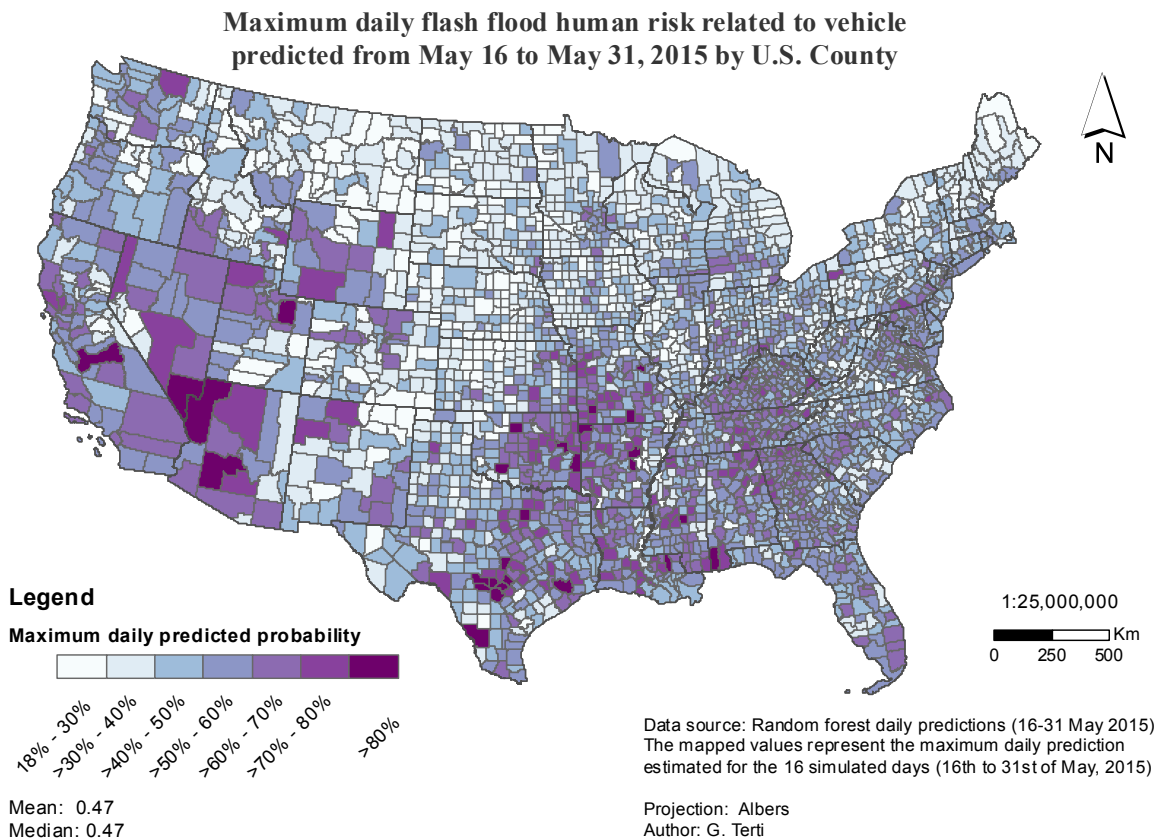


Figure V.9 – Maximum county-level daily forecast probability of vehicle-related casualty due to flash flooding, estimated by the random forest model for each day between the 16th and the 31st of May 2015.

1. A flash flood event is reported by the NWS when it has posed a potential threat to life or property, and had a report of moving water with a depth greater than 0.15 m or more than 0.91 m of standing water. Documentation available at <http://www.nws.noaa.gov/directives/>.

4.1 Deadly flash floods in Texas and Oklahoma in May 2015

On May 14, 2015, prior to extensive flooding beginning around May 24, flash flood warnings were issued for counties in southeast Texas. At least 34 people lost their lives in flash floods from May 6 to 29, including 30 victims in Texas and 4 in Oklahoma (Figure V.10). Eighty-eight percent of these fatalities occurred from May 18 to 29, and more than fifty percent of those occurred in vehicle-related circumstances. First-response authorities carried out hundreds of water rescues involving mainly stranded motorists who attempted to drive through high water. Submerged streets and flooded or destroyed homes drew a multi-day hazardous scene with thousands of stranded, occasionally abandoned vehicles, and homeless people along the Blanco river as well as the urban centers of Dallas and Houston in Texas.

Especially after May 16 that refers to the case study presented in the next subsection, the *Storm Data* reported 27 victims in Texas and 3 in Oklahoma. Fifty-seven percent of those deaths were related to vehicles (Figure V.11). Individual fatalities for those events were reclassified as proposed in subsection 2.2 of Chapter III.

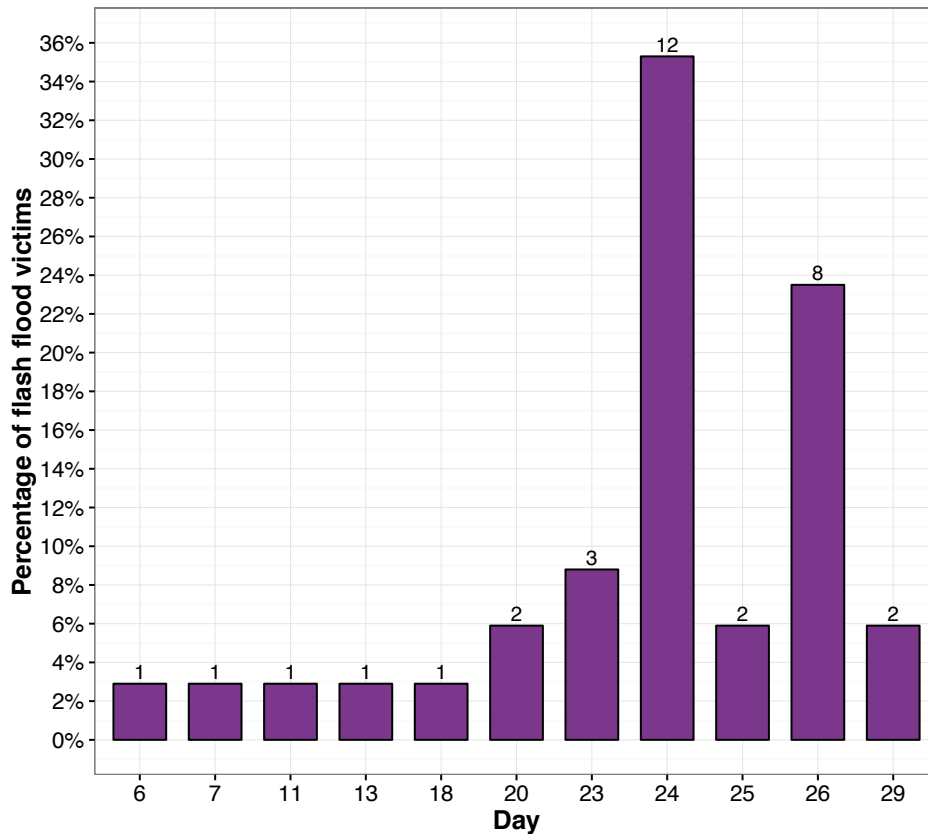


Figure V.10 – Percent of flash flood victims by day of reported fatality. Percentages are estimated to the total 34 fatalities reported in the *Storm Data* for Texas and Oklahoma in May 2015. The values on the top of the bars indicate the raw number of fatalities in each day.

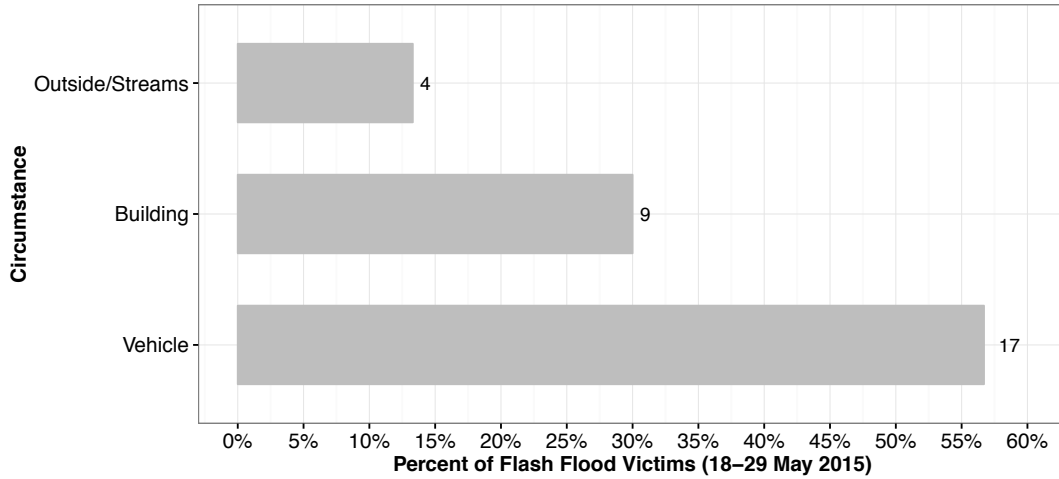


Figure V.11 – Percent of flash flood victims by circumstance. Percentages are estimated to the total 30 fatalities reported for Texas and Oklahoma, from May 16 to May 31, in the *Storm Data*. Especially, vehicle-related fatalities were reported on May 18, 20, 23, 24, 25, 26, and 29. The values on the right of the bars indicate the raw number of fatalities in each circumstance. The estimations are based on reclassification of the fatalities circumstances reported in the *Storm Data* fatality file, available at <https://www.ncdc.noaa.gov/stormevents/ftp.jsp>. The fatality circumstances are classified as proposed in Chapter III.

4.2 Mapping dynamic human risk related to vehicles

a) Daily estimations of vehicle-related human risk

In this case study, daily risk maps are constructed with a focus on the 254 counties in Texas and 77 counties in Oklahoma. Figure V.12 presents daily maps from May 23 to 26 when the majority of vehicle-related fatalities occurred (12 fatalities). The estimated probabilities are equally distributed in four categories: i) low likelihood: ≤ 0.25 , ii) moderate likelihood: $> 0.25 - \leq 0.50$, iii) high likelihood: $> 0.50 - \leq 0.75$, and iv) very high likelihood: > 0.75 . To prevent overrepresentation of dynamic risk in counties with possibly high values of static (e.g., flashiness) or semi-static predictors (e.g., commuters) but no actual flash flooding, the probabilities in counties with low daily unit peak discharge ($< 2 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$) are mapped in the low likelihood category. The counties with vehicle-related victims are extracted depending on the fatality day reported in the *Storm Data* fatality file and highlighted with red boundaries on the produced daily risk maps. Local storm reports (LSRs) are also mapped with red dots to illustrate flash flood emergency issues such as road flooding, closures, and rescues. LSRs are preliminary reports issued in near real time by local NWS forecast offices and serve as the initial source for reports in *Storm Data*¹.

On May 23, the model predicts higher probabilities for vehicle-related incidents in two main areas along the western Oklahoma-Texas boundary and central Texas. In the first area, the eastbound and westbound lanes of Interstate 40 were closed because of flooding in counties predicted as high-moderate likelihood (Figure V.12A). According to the media, nearly every low-lying bridge in Elk City was flooded². Also, Oklahoma City and surrounding cities' fire departments responded to more

1. LSR are available online at <https://mesonet.agron.iastate.edu/lsr/>.

2. Weather news on <https://weather.com/storms/severe/news/southern-plains-flooding-texas-arkansas-oklahoma>. Retrieved on August 10, 2016.

than 100 vehicles stuck in high water in the evening¹. In Texas, a 42-year-old man died in his vehicle along the Blanco River near downtown Blanco². In the same county, another male victim (81-year-old) was swept away in floodwaters while trying to escape his car. The next day on the 24th, high likelihood of vehicle-related accidents are predicted from the Eastern border of Oklahoma to the Central-Southern counties of Texas. In fact, two vehicle-related fatalities occurred in the highlighted counties (Hays and Medina) in Texas (Figure V.12B). A 29-year-old man was washed away with his vehicle and an 18-year-old girl was swept away while driving back home. In Oklahoma, no fatalities were reported but several local storm reports indicate numerous roads flooding and submerged cars.

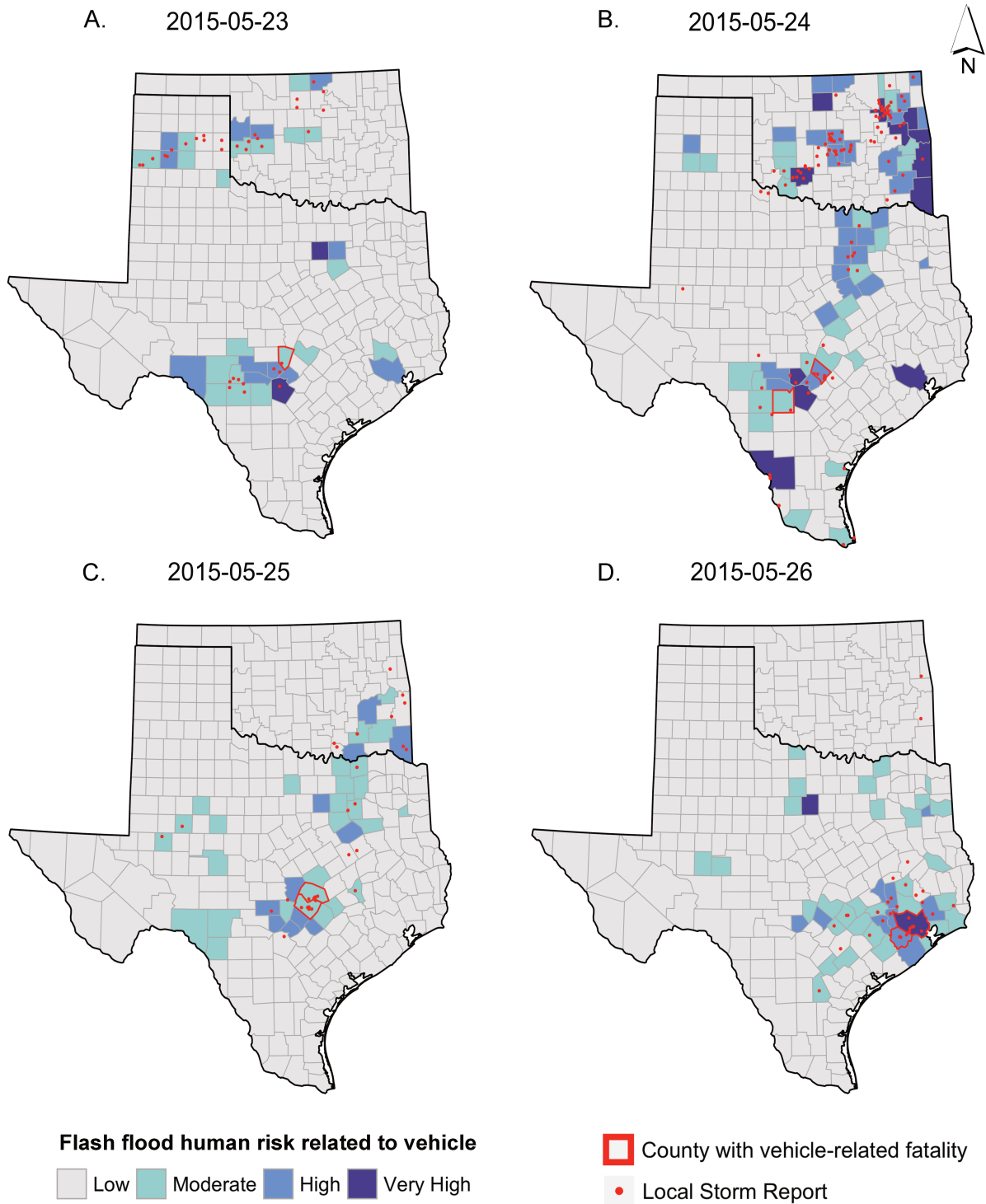
On May 25, the spatial pattern of the predicted vulnerability remains similar and notable but with lower values. A cluster of higher probabilities occurred in central Texas where multiple water rescues were reported (Figure V.12C). Two males (23 and 55-year-old) died when their vehicles were swept away in Travis and Williamson Counties, respectively. On May 26, risk for motorists according to the developed model is concentrated in southeastern Texas with highest probabilities estimated around Harris County in Houston, Texas (Figure V.12D). Actually, hundreds of vehicles were stranded in floodwaters after daylight in the Houston area. Four fatalities occurred that were directly related to vehicles in Harris County. Three more fatalities resulted from the capsizing of a Houston Fire Department rescue boat while rescuing stranded motorists³. In Fort Bend County, a 73-year-old woman lost her life while driving to work and was found dead about 50 m from her submerged car⁴. It appears that the model performs better for widespread precipitation and flash flooding than for more localized events.

1. Online news at <http://newsok.com/article/5422080>. Retrieved on August 10, 2016.

2. According to *Storm Data* event narratives available at <https://www.ncdc.noaa.gov/stormevents/eventdetails.jsp?id=581620>.

3. The three fatalities are classified as vehicle-related based on the reclassification approach presented in Chapter III.

4. Details from online media available at <http://edition.cnn.com/2015/05/27/us/severe-weather/>.



Data source: Random forest daily predictions (23-26 May 2015)
 Projection: Albers
 Author: G. Terti

Figure V.12 – County-level daily forecast of vehicle-related human risk for each day (from A to D) from May 23, 2015 to May 26, 2015 for Texas and Oklahoma States: the likelihood of flash flood vehicle-related casualty for individuals is predicted by the random forest model for each county and day (i.e., 1%-100%), and is assigned to four categories from low to very high. Counties with daily unit peak discharge $< 2 \text{ (m}^3 \text{ s}^{-1} \text{ km}^{-2}\text{)}$ are also assigned to the low likelihood category.

b) Hourly estimations of vehicle-related human risk

Forcing the random forest classifier with hourly unit peak discharge simulations produces hourly predictions for the catastrophic day of May 26. Other time-variant predictors are also adjusted at the hourly scale to be inserted as inputs in the model. The number of commuters is related to the corresponding hour associated with a certain unit peak discharge in each county; assuming that this discharge is representative of a potential flash flood event at that specific time. The duration of precipitation is considered as increasing with an hourly step, starting one hour before the first simulated hourly unit peak discharge of the day which begin at midnight. The hourly prediction maps highlight southeastern Texas counties with the highest probabilities to occur in the morning hours (Figure V.13). Although the exact time of the fatalities occurred on May 26 is not easy to be determined through the available data or media, the *Storm Data* report the causative events within the first hours of the day.

According to the daily map for May 26, vehicle-related human risk is high and very high for Fort Bend and Harris Counties, respectively, where vehicle-related fatalities were actually reported. Interestingly, the hourly maps show some variability for this area with the probabilities to fall on the high, moderate, and low categories at specific hours during the day. The highest probabilities in Harris Counties are estimated mainly from 03:00 a.m. to 09:00 a.m., and reduces throughout the day, revealing the conjunction of commuters and flood dynamics during morning commuting hours (Figure V.14). In fact, on the morning of May 26, the NWS in Houston/Galveston issued a flash flood emergency for southwest Harris County (which includes the city of Houston) and northeast Fort Bend County. A flash flood emergency is only issued in the most life-threatening rainfall situations. Especially, that was the first flash flood emergency for Harris County history, and it was announced at 10:52 p.m. on the 25th [Talbot, 2015]. During the morning hours on the 26th, people shared hundreds of pictures and comments through social media such as twitter, instagram and facebook, indicating road flooding as the dominating life-threatening circumstance (Figure V.15). A flood warning has been issued for the entire Houston metro area until 2:45 p.m of the 26th. The Houston NWS announced strong messages to prevent drivers from trying to cross flooded roads, getting emergency responders and themselves into high risk:

“MOST FLOOD DEATHS OCCUR IN AUTOMOBILES. NEVER DRIVE YOUR VEHICLE INTO AREAS WHERE THE WATER COVERS THE ROADWAY. FLOOD WATERS ARE USUALLY DEEPER THAN THEY APPEAR. JUST ONE FOOT OF FLOWING WATER IS POWERFUL ENOUGH TO SWEEP VEHICLES OFF THE ROAD. WHEN ENCOUNTERING FLOODED ROADS MAKE THE SMART CHOICE...TURN AROUND...DON'T DROWN”

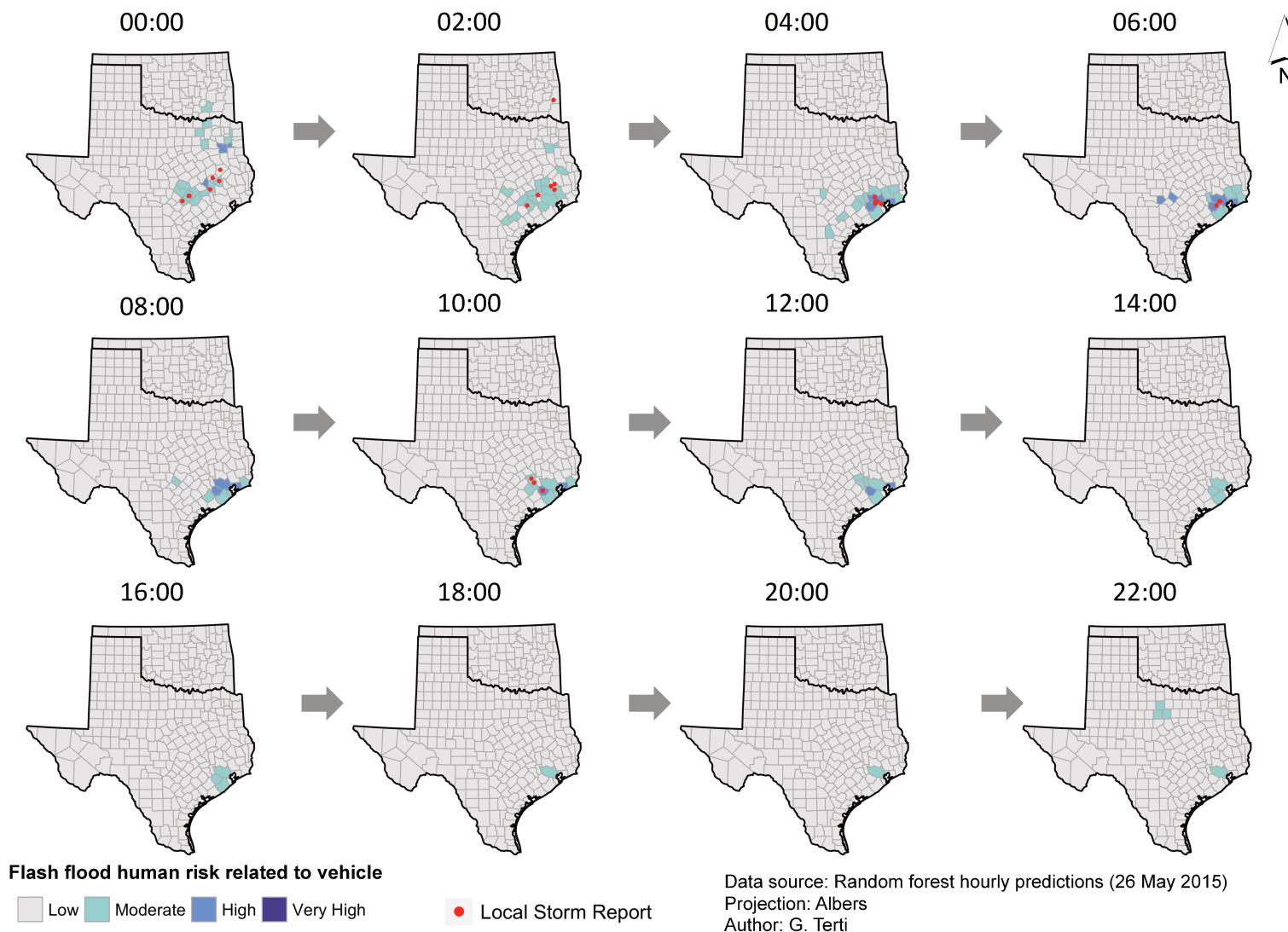


Figure V.13 – County-level hourly forecast of vehicle-related human risk for May 26, 2015 for Texas and Oklahoma States: the likelihood of flash flood vehicle-related casualty for individuals is predicted by the random forest model for each county and hour (i.e., 1%-100%), and is assigned to four categories from low to very high. Counties with hourly unit peak discharge $< 2 \text{ (m}^3 \text{ s}^{-1} \text{ km}^{-2}\text{)}$ are also assigned to the low likelihood category. The illustration includes twelve hourly maps from 00:00 to 23:00 using a 2-hours step (from left to the right).

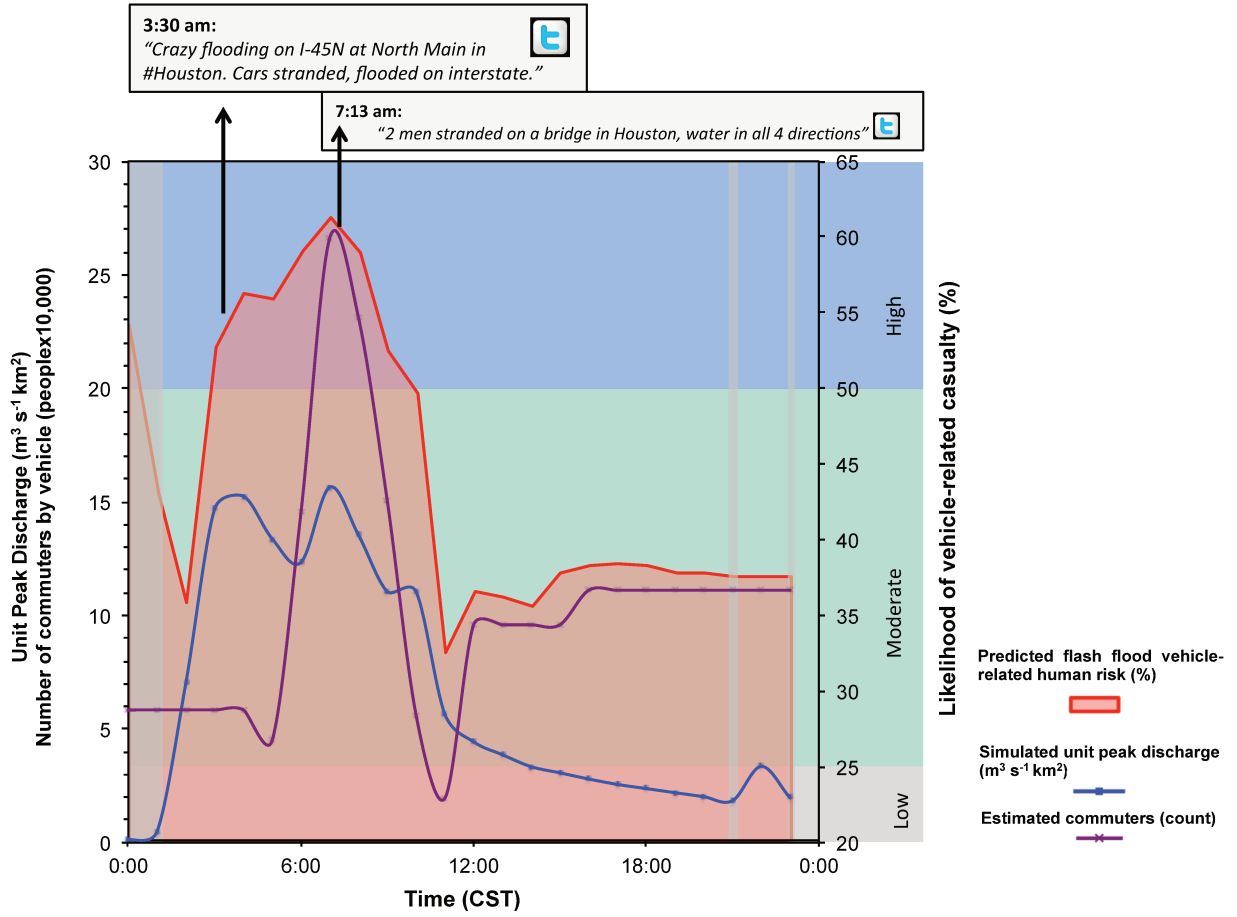


Figure V.14 – Estimated hourly predictors, and vehicle-related risk predictions by local time (Central Standard Time), for flash flooding in Harris County (Texas) on May 26, 2015. The left Y-axis refers to the simulated unit peak discharge ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) and the commuters (*people* $\times 10,000$). The predicted likelihood of flash flood casualty is presented in percentages on the right Y-axis rendered over the corresponding categories of risk. We recall that predictions which correspond to unit peak discharge < 2 ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) are masked in the low likelihood category.

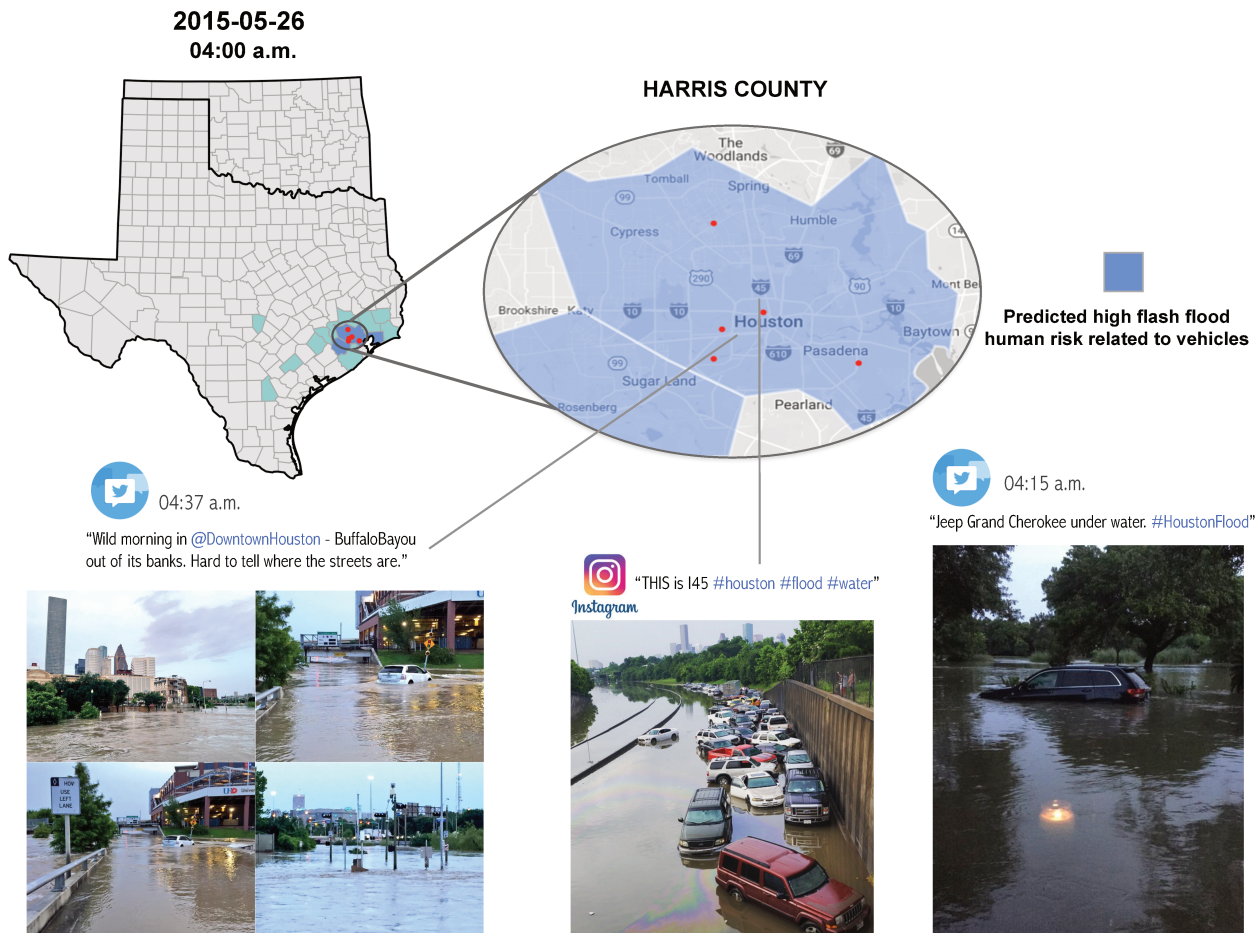


Figure V.15 – Up: Random forest predictions for vehicle-related human risk from flash flooding at 4 a.m. on the 26th of May (2015) in Texas and Oklahoma. The illustration focuses in Harris County (Texas), and is rendered over a reference map from Google Maps to provide details of the existing road network in the area. Down: Examples of observed vehicle-related incidents from flash flooding on the highways and downtown road network in Houston area, as shared through social media. Posts are available online at <https://www.washingtonpost.com/news/capital-weather-gang/wp/2015/05/26/much-of-houston-under-water-after-11-inches-of-rain-devastating-flood/> (retrieved on November 27, 2016).

5 Summary and Conclusions

This chapter proposes a machine-learning modeling approach to produce probabilistic forecasts of human losses from flash flooding at the county-level, and daily or hourly steps. We vision this approach as a forecasting system to anticipate potential human losses with a focus on the most prevalent circumstance of fatalities: vehicle-related incidents. The study builds on the social vulnerability research with two main contributions: i) human vulnerability aspects are integrated, for the first time, with hydrological forecasts to account for the evolution of human risk to flash flood hazard in time and space, and ii) historic losses are involved in the modeling procedure to link vulnerability conceptualizations with human impact observations. Random forests classifiers deal with nonlinear, complex interactions between the predictors and were selected for this study. This method provides for the identification of the variables that best represent the interplay between the natural hazard and human vulnerability processes for commuters during flash flooding events.

In this study, uncertainty in quantification of human risk related to vehicles is accounted for by treating the occurrence of flash flood fatalities in a probabilistic way. Compared to previous studies, human risk is illustrated as an evolving likelihood of vehicle-related incidents overcoming the one-sided static generalization of social vulnerability from county to county. Validation of the developed model is not a straightforward exercise. Rare events such as flash floods with casualties are difficult to predict. The same conjunction of socio-hydrological conditions identified as lethal in past flash flood events may not result in fatalities during a future event due to differing circumstances at a very local level. More precise impact data are needed to calibrate and/or verify the model outputs. Integrating social media and crowdsourcing datasets in the modeling process could provide a valuable contribution to the model performance. Based on the case study presented in this chapter, the model shows promising results in terms of locating dangerous circumstances in space and time. Higher probabilities are adequately predicted for extended county-level flash flooding while the model seems to overestimate vehicle-related risk during very localized events. Critical thresholds for the prediction of vehicle-related incidents need to be further investigated integrating local sensitivities.

When analyzing the variable importance in the model, it is shown that peak unit discharge plays the most important role in partitioning the flash flood events in events with and without vehicle-related fatalities. Being dynamic, this variable and the maximum precipitation both describe the magnitude of the natural hazard. Especially, it is because these dynamic variables have been determined in much higher spatial and temporal resolutions than the county-level demographics, that they probably can inherently capture some local conditions crucial for the occurrence of life-threatening scenes. Still, other hydrological variables (e.g., flood water velocity), not available in this study, might be stronger predictors for vehicle-related incidents. The area under the ROC curve (AUC) used to evaluate the final classifier is estimated as 0.7 indicating a moderate predictive performance; inviting further improvements with the predictors. Particularly, the results presented here are subject to the following limitations and inherent uncertainties:

- *Data uncertainties:* Flash flood events and the recorded human losses in *Storm Data* are subject to undercounting [Curran et al., 2000; Ashley and Ashley, 2008; Sharif et al., 2012]. Because the model uses a binary classification, under-reporting is generally not a problem. Socio-demographic information from the U.S. Census Bureau and other data sources used as inputs in this study may add further inaccuracies. Sometimes data provided by the American Community Survey are characterized from large margins of error adding further concerns about the quality and precision of model inputs [Spielman et al., 2014]. Lastly, the hydrologic model simulations are subject to uncertainties due to inadequate model physics representations and forcing from weather radar-based rainfall estimates [Flamig, 2016].
- *Scale limitations:* The need for large number of observations to construct an adequate statistical sample for the machine-learning algorithm necessitates the consideration of many years of flash flood event observations within a large geographic area (i.e., whole U.S.). That means that regional differences and local specificity that may convert an initially moderate risk flash flood event to a catastrophic event are not considered in the current analysis.
- *Resolution constraints:* The fact that reports on flash flood fatalities are not spatially explicit complicates the supplementation with other extra datasets available at higher resolution than the county. Local and sometimes dynamic information defined on the order of a few kilometers

and/or with high temporal resolution (e.g., population density, unit discharge) are aggregated, losing details that may contribute to the occurrence of a lethal scene.

Rather than presenting this model as an established relationship between the selected predictors, we envision an adaptive approach that evolves with data updates and improves with experience. Opportunities for future work are discussed in Chapter VI. For example, it would be interesting to further refine the exposed areas within the counties based on the extent of the hydrological forecast. Collection of more spatially precise candidate predictors can then be forced according to the extent of the impacted area. Casualties depend on many parameters such as personal strengths and last-minute decisions. Discrimination between lethal and non-lethal events is very difficult especially for flash flood events with less than 5 fatalities. Exploring other classification criteria of the target variable might enable a more refined clustering of the severity of flash flooding in terms of impacts. We recommend that the flash flood disaster science community and practitioners conduct data collection with more details and at finer resolutions to better capture local temporal and spatial complexities associated with human losses from flash flooding.

General conclusion

1 General review

Within the context of risk management, vulnerability assessment complements hazard assessment to inform actions related to prediction and early warning and ultimately, risk reduction. For forecasters and emergency managers the prediction and warning of human impacts due to such a sudden onset and localized event like flash flooding is a big challenge. Despite unavoidable biases and scale issues, this PhD work represents a first attempt to provide a prediction system to support emergency preparedness and response to flash flood disasters. The research is motivated by the hypothesis that the intersection of the spatiotemporal context of the hazard with the distribution of people and their characteristics across space and time reveals different paths of vulnerability and defines the occurrence of human losses. But how far can vulnerability and risk research go towards prediction of anticipated human impacts? Are the available data adequate to capture complexities associated with both physical and social processes that not only overlay, but also interact in space and time? In this study, we explored readily available datasets across the U.S., and we adopted a modeling approach to support a nationwide prediction effort for NWS forecasters and emergency managers to target their warnings on anticipated human impacts, forcing the model with hydrologic forecasts.

Specifically, throughout this PhD project, we have addressed the following objectives, presented in the General Introduction:

- Constructing a conceptual vulnerability model that captures the dynamic interplay between the flash flood hazards and social vulnerability factors at relevant spatio-temporal scales [Terti et al., 2015].
- Analyzing the situations of historic fatality events as determined by the victim's profile and fatality circumstance, and the spatio-temporal context of the causative flash flood event reported over the U.S. [Terti et al., 2016].

- Modeling and dynamic mapping of human risk predictions, based on the circumstance of the anticipated incidents (here vehicle-related), and the time-variant exposure to a given flash flood forecast [Terti et al., in review].

2 Summary of research findings

2.1 Research question #1: Which social processes interact with the flash flood hazard defining the dynamic human vulnerability?

To decide which social processes are related with people's vulnerability to short-fuse weather events (and especially, flash floods), impact assessments of flood and flash flood events in Europe, North America, and Australia were primarily reviewed. Prior vulnerability studies have been also critically reviewed to conceptualize vulnerability to a flash-flood and impact specific context with a focus on the emergency phase of the event. Thus, findings from this theoretical analysis implies for applicability in developed regions such as Europe and North America, where common features in terms of human development indexes and living conditions call for similar priorities in the distribution of assets at the individual or institutional level. The primary vulnerability factors identified through the literature review and conceptual analysis are separated in four major categories based on the nature of the social process to which they are related:

- Land use: The management of the natural environment to become built where specific human activities may be located. The characteristics of this man-made environment (e.g., dense road network, "special needs" buildings) may affect the safety of people and their emergency response in flash flood circumstances.
- Risk governance: Institutional policies for flood risk preparedness and response (e.g., official emergency services, dissemination of timely warnings).
- Individuals' status: People's position in relation to others in regard to social or professional standing. The socio-demographic profile of people (e.g., age, gender, profession, housing ownership, family ties, and health) is a primary factor that shapes the everyday life priorities and constrains and defines individuals' predisposition to crisis situations.
- Cognition: Mental process of understanding through experience, thoughts, and sensation, forms the conscious and unconscious mechanisms of individuals (e.g., risk awareness, familiarity with roads, emotional attachment to property).

This study argues that human impacts related to flash flood events present high variability and diversity from place to place not only due to differences in hydro-meteorological circumstances but also due to the space-time variability of people's exposure and capacity to react. Based on this idea, a conceptual model for assessing impact-focused vulnerability to flash flooding was developed and presented in Chapter II. The conceptual model developed in this dissertation, was used as the base to identify exposure, sensitivity and coping capacity variables to serve as vulnerability indicators for the assessment of human risk at the county level, and daily or sub-daily time steps at the scale of U.S. The selection and collection of indicators based on nationwide datasets, as presented in Chapter III, intended to quantify the vulnerability factors summarized in the presented conceptual framework.

Further analysis of human losses associated specifically with flash flood occurrences in the U.S, and classified by accidents' circumstances and victim's profile, was required for more targeted selection of relevant proxies to describe each indicator depending on the risk situation.

2.2 Research question #2: Who is the most vulnerable in terms of loss of life to flash flooding?

Identifying the social, spatial, and temporal framework of the historic human losses from flash floods is key to gaining a deeper understanding of the contextual risk factors, and thus to advance vulnerability assessment and future prevention policies. In this dissertation, 1075 flash flood-specific human losses reported from 1996 to 2014 on the scale of the U.S. were classified in six main categories/circumstances that explain the majority of the deaths, and were investigated correspondingly. The purpose was to explore if different vulnerability paths occur depending on the situation as determined by the victims' profile and activity and the spatio-temporal context of the flooding. The statistical analysis suggests the following patterns that seek to answer the questions set in the introduction of Chapter IV:

- The majority of flash flood victims are involved in vehicle-related circumstances followed by the "outside" incidents (almost 61% and 21% of the total 1,075 fatalities, respectively). In many cases, both circumstances are characterized by unnecessary risk-taking behavior on behalf of the victims. In our analysis, men were 2.6 and 1.4 times more likely than women to be involved in "vehicle" and "outside" fatal incidents, respectively. Given that the median age of males is 22 years for "outside/close to streams" circumstances and 44 years for vehicle-related circumstances, it appears that the generic concept of elderly people being the most vulnerable to natural hazards is not categorically supported in the flash flood context. Older people are more likely to perish inside buildings, and especially in twilight and darkness hours when rescue and/or evacuation operations are hindered, whereas younger and middle-aged people are more likely to get trapped while participating in daytime outdoor activities. In particular, "outside" incidents are more likely during daylight hours associated mainly with young males (e.g., children and teenagers that were swept into creeks while playing near high waters). On the other hand, low visibility of dusk conditions is an exacerbating factor for middle-aged active population to be involved in vehicle-related accidents.
- Concerning the flood dynamics, people in outdoor circumstances such as vehicle-related, outside in neighborhoods and recreational areas (e.g., campsites, canyons) are more likely to drown in fast and dynamic flash flood events with duration close to 5 hours. On the contrary, people in buildings and mobile homes are more likely to get threatened by longer flash floods.
- There is no clear trend in the annual fatalities or circumstances during the study period. In agreement with the spatio-temporal distribution of heavy precipitation, the majority of fatal flash flood events (65% of the 705 total fatal flash floods) and the related fatalities (65% of the total 1075 fatalities) occurred in the warm season from May through September. From a social point of view, it is interesting to notice that outdoor circumstances such as "outside" and "recreational", strongly contribute to the total monthly fatalities during this summer period

compared to the rest of the year where "vehicle-related" victims dominate with noticeably higher percentages.

- Geographically, the state-level analysis of the fatalities circumstances reveals the higher susceptibility of people living in southern states to perish in vehicle-related circumstances. Further investigations are needed to determine if this is more due to cultural factors such as risk-taking behavior and confidence in automobiles or other reasons more related to higher urbanization and higher exposure of the road networks. Certain physical attributes of the exposed environment (e.g., canyons, campsites, proximity to rivers) can be strong vulnerability predictors for certain circumstances such as "recreational" and "outside".

The above findings confirmed the hypothesis raised from the theoretical analysis presented in Chapter II, that situational rather than generic examination of vulnerability is required to realistically capture risky cases during short fuse flood events. The patterns identified in the *Storm Data* generated the idea to use such impacts observations for modeling situational human risk towards prediction of circumstance-specific human losses in future events. The results from the statistical analysis further supported the construction of vulnerability assumptions when selecting proxy variables to represent circumstance-specific risk indicators in our methodological framework. The compiled database used for this analysis is now part of the unified flash flood database, publicly available at <http://blog.nssl.noaa.gov/flash/database/> for future research studies.

2.3 Research question #3: How to quantify the dynamic relationship between the flash flood magnitude and human vulnerability?

Modeling human risk to flash flood events, is a challenging task that requires i) a comprehensive dataset with variables explaining the hydrometeorological response, and the characteristics of the exposed areas and people associated with certain disaster occurrences, and ii) machine learning methods capable to deal with nonlinear and complex interactions between the variables. To support such modeling approach a list of *Storm Data* flash flood events, with and without human losses from 2001 to 2011 in the contiguous U.S. was supplemented with other variables describing the storm event, the spatial distribution of the sensitive characteristics of the exposed population and built environment at the county level (Chapter III). As presented in Chapter V, the flash flood event database was prepared for binary classification separating events with one or more vehicle-related fatalities and events without reported fatality. Then, random forest, a decision-tree based ensemble machine learning algorithm for classification was adopted to assess the likelihood of fatality occurrence for a given circumstance as a function of representative indicators. Starting from the most prevalent circumstance of fatalities raised from the analysis presented in Chapter IV, flash flood events with lethal vehicle-related accidents were the target to predict. The classification method applied in Chapter V reveals the following:

- The internal evaluation that was conducted through the model building shows that the classifier does not present very high predictive performance (AUC=0.7). Though, this first result towards predicting vehicle-related losses in a set of unseen flash flood events is encouraging. Given that the coarse-resolution of the predictors (e.g., county level) may be unable to explain salient local natural and social processes, and that fatal flash floods are extremely rare

events to predict, the random forest method could not provide predictive skills comparable to other more explicitly-defined applications (e.g., prediction of an e-mail as spam or not based on characteristic keywords-predictors). Considering the uncertainties introduced by the data, but also the modeling method itself, we elaborate a probabilistic configuration of human risk related to vehicles.

- Concerning the importance of variables, it appears that the more dynamic the variable, the more determinant it becomes in the classification of fatal and non-fatal events. Indeed, variables that describe the social aspects (e.g., median age, household size) are considered as weaker predictors in the model. However, when these variable are excluded from the predictors set, the ability of the model to predict the probability of flash flood events with vehicle-related fatalities in the validation set is reduced. There is a plausible signal that considering all possible interactions between the probably weak predictors may lead to a better model for predicting vehicle incidents in flash flood events. Still, more work is required to advance the representation on human vulnerability in the predictive methodology.

The advantage of the methodology proposed in this dissertation is that, in contrary to previous studies, human vulnerability, and the subsequent risk, is illustrated as a time-variant likelihood of vehicle-related incidents advancing static generalizations of human vulnerability. The developed model can be applied on a daily or sub-daily basis for every county in the conterminous U.S. by forcing the random forest classifier presented in Chapter V with daily or hourly hydrological forecasts, respectively. When examining the May 2015 flash floods in Texas and Oklahoma, the model results are encouraging in identifying vehicle-related human risk in space and time (Chapter V). Especially, in cases of extended county-level flash flooding, the model sufficiently predicts high probabilities. Though, the vehicle-related risk tends to be overestimated during very localized events that may largely depend on local sensitivities not yet captured by the model.

3 With an eye to the future

3.1 Prospective work

The use of observed human losses and other reported impacts, as guidance to build and adjust machine learning models is a promising approach to better link social vulnerability conceptualizations with realistic forecasts of prominent impacts from flash flood hazards. However, there is a lot of work to be done to establish a model that has high predictive power for the occurrence of casualties in future flash flood events across the U.S.

A logical following this work is to include other types of human impacts (e.g., injuries, rescues) in the analysis. This would enhance the statistical analysis of vulnerability situations explored in Chapter IV. In addition to that, considering other human impacts would enlarge the size of the statistical sample for events with human impacts in the machine-learning based risk analysis presented in Chapter V. Unfortunately, systematic classification of impacts other than fatalities in not a straightforward exercise in the *Storm Data*. Although, initially, we explored the potentiality to classify the circumstance of nonfatal impacts when details were included in the event narratives of *Storm Data*, it appeared that the majority of the events lacked information about the location of the

accident, and especially, they did not detail on the profile of the people involved. One recommendation arising from that limitation is for the NWS to promote a more human-centered reporting approach for weather disasters in the U.S.

Another advancement of the current methodology would be to improve the spatial resolution of the analysis. Given the locality and complexity of the flash flood hazard, the practicality of the county-level modeling is questionable. As described in Chapter III, at the time of the analysis the most reliable spatial reference for the reported impacts in the *Storm Data* was the county. To avoid spatial vagueness and inconsistencies between the *Storm Data* files, and to maximize the amount of available records, the county reference was used in the predictive modeling. Though, when the accuracy of the bounding polygons currently adopted by the NWS to report impacted areas in flash flood events, allow for it, it would be interesting to bring all the data in finer resolution. Furthermore, the reported bounding polygons could be crosschecked with the extent of the hydrologic forecast to delineate even more specific exposed areas. This would provide for the collection of more spatially precise predictors to be used as input in the machine learning model training. For instance, data that were already available on the order of a few kilometers (e.g., population density, unit discharge) could be then more valuable for describing the exposure related to a certain flash flood event. Census data from the ACS could be then extracted and aggregated from smaller geographic units such as block groups to allow for a better representation of the socio-economic and demographic variability of the exposed people. Additionally, it would be very interesting to proceed to an assessment of uncertainties associated with the input data (when feasible). The ACS data for example, provide the margin of error for the socio-demographic estimates allowing for a brief description of the measurement inaccuracy and uncertainty.

Future work should further expand the methodological developments presented in this dissertation in the modeling of other life-threatening circumstances (e.g., flooding inside buildings, in mobile homes, at recreation areas) identified as challenging for the response of individuals to flash flooding. Though, it was realized that the biggest caveat of the adopted machine learning approach was the need for large number of observations to train a robust model. Thus, the rarity of flash flood events with fatalities in circumstances other than vehicle-related make the prediction of incidents in those circumstances a real statistical challenge. Instead of using a binary classification of events (e.g., with and without fatalities) in a given circumstance, other unsupervised machine learning methods such as clustering algorithms [Huang, 1998; Kim et al., 2004; Plant and Böhm, 2011] should be probably explored. Based on additional information that could be potentially available in the data (e.g., considering both the circumstance of the fatal accident and the profile of victims, duration of the causative event etc.) such clustering methods may could classify the historic flash flood events of the *Storm Data* in different categories. If that would work, then, instead of modeling every circumstance separately, we could build a multiclass classification approach for predicting different clusters of flash flood events.

Given the strong dependency of machine learning methods on the data inserted to them, it is apparent that they are not a panacea in the realization of previously-unseen conditions. Experts engagement is a necessity to compensate the scarcity of large and suitable data at the scale of the flash flood disasters. It is suggested, therefore, that the model developed in this study may be presented as a prototype for forecasters of the NWS to visualize the capabilities of the included

information and outputs on human risk forecasting. A participatory approach, involving forecasters and emergency managers, is a strong recommendation not only to fit the model objectives and outputs to their needs, but also to get feedback on potential adjustments and improvements of the modeling itself based on experts' knowledge and experience in the area of their responsibility.

3.2 Next research directions

One question raised from the current study is how well social vulnerability related to short-fuse weather hazards can be represented through aggregated socio-demographic characteristics available in the census data. Although, traditionally vulnerability assessments related to natural hazards propose the statistical analysis of socio-demography revealing economic and physical fragility in certain geographic units (e.g., block groups, census tracts, counties) [Cutter et al., 2003; Wilhelmi and Morss, 2013], it is an obvious omission that risk perception and other cognitive factors related to decision making in crisis situation are not directly represented through such vulnerability indicators. In fact, the machine learning model trained on the *Storm Data* flash flood reports indicates social variables at the county level as weak predictors of the vehicle-related fatalities. Since information on flood risk awareness and perception is only available through questionnaire-based local studies [Drobot et al., 2007; Morss et al., 2015], their effect on the prominent impacts during flash flooding should be explored at the local level.

Other approaches with different data requirements and spatial applicability could be elaborated in future research. The proposed approach could be more experimental with the development and implementation of an agent-based model to test our vulnerability hypotheses at smaller scales: over a small city or a neighborhood with resolutions of the order of meters and minutes [Chen et al., 2006; Taillandier et al., 2010; Beck et al., 2014]. Agent-based simulations (ABS) “*focus on the concept of the emergence of a system. A group of agents are defined, which follow a set of rules and, in their interaction, whilst following these rules, the behavior of the system emerges*” [Nagarajan et al., 2012]. Agents are autonomous, goal-oriented, software components that perform a task and interact with other agents and environment [Albino et al., 2007]. Recent studies showed interesting results with agent-based modeling to describe the interactions between people and flood hazard, and to assess vulnerability of individuals during flooding [Dawson et al., 2011; Lumbroso and Tagg, 2011]. This type of modeling is particularly appropriate to test the hypotheses listed in Table II.1 of this thesis, about the influence of cognitive processes on crisis behaviors among other factors. Nevertheless this approach can also integrate outcomes from the statistical analysis conducted on coarse resolution datasets (e.g., analysis of *Storm Data* presented in Chapter IV) to see how these affect the distribution of human impacts at the small scale.

Based on the present research findings, the agent-based model may include six main agents: Flash flood, road network, buildings, mobile home parks, recreation areas, and people. Some of the agents may interact and define the state of each other in each time step (Figure VI.1). The flash flood agent represents the magnitude and the timing of the flash flood event and directly affects the state of the infrastructure elements in the environment (e.g., roads become flooded or totally impassible/closed). People (i.e., human agents) can change their behavior after perceiving the state of the environmental agents. For example, they may decide to change direction in order to reach

their target (e.g., residential or workplace building) when the road network that they use is closed or they may change totally their destination (e.g., go to a shelter instead of the residential building that could be the primary target). The response of the people in the model can also change with the reception of a warning when they decide to follow the received message. The final objective of this model would be to reproduce daily life mobility choices made on a) a usual weekday and a usual weekend and on b) a weekday and a weekend at which a flash flood occurs, and to test how the number of people involved in certain accidental circumstances vary through the day (i.e., diurnal weekday differences) or through the week (i.e., differences between weekday and weekend).

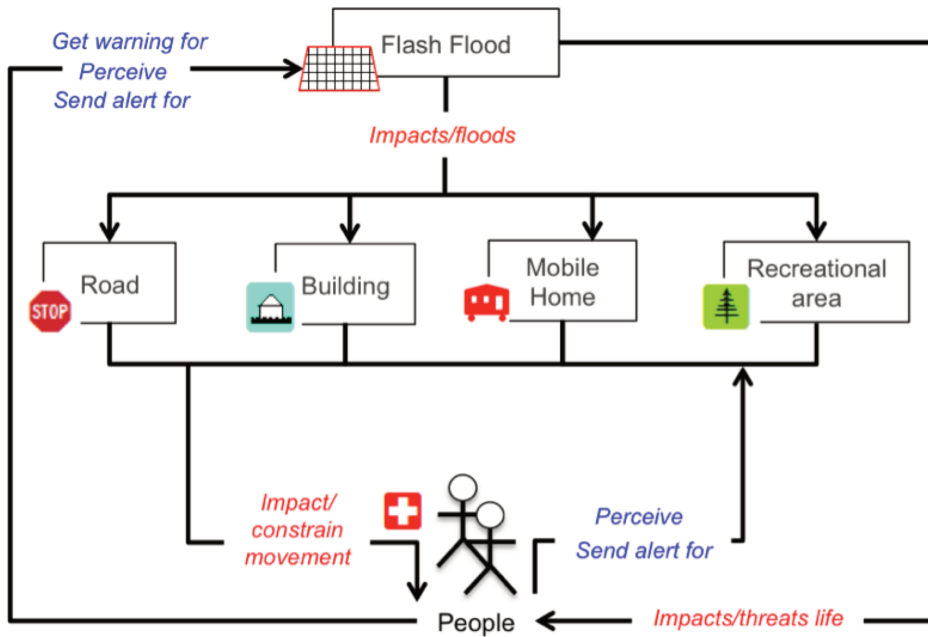


Figure VI.1 – Agents and their interactions in the future agent-based model for flash flood crisis.

Focusing on small scale and high-resolution case studies, this complementary method is expected to better capture the timing and spatial distribution of the overall human impacts. This modeling approach was initiated through this dissertation, and is under development in collaboration with Julie Dugdale and Carole Adam, both multi-agent systems experts from MAGMA (Socially Relevant Multi-Agent systems) team, and Elise Beck, researcher from Pacte-Social Science laboratory in Grenoble, France.

By conducting these two, integrating approaches (i.e., statistical modeling and agent-based modeling), feedback from each method can help to refine the final impact-based forecasting product. The agent-based model would be primarily designed for scientific and not operational purposes, but it could be a useful tool for stakeholders as well as scientists to explore and visualize the performance of the flash flood warnings and their contribution to the reduction of losses in flash flood incidents. Eventually, simulations using both approaches can be performed on selected past flash flood cases in the U.S. in order to compare the outputs of both approaches with detailed local impact datasets available for specific flash flood events. In the future, European flash flood cases could be also examined to investigate if there are differences in the occurrence of casualties, emerging from potential differences in the dynamics of flooding in European and U.S. territories, as well as differences in anticipation decisions of individuals from different cultural contexts. It is hypothesized that such comparisons

of case studies would allow for an holistic understanding of the contextual factors embedded in the small-scale physical and human processes driving the distribution of flash flood human impacts.

REFERENCES

- Adger, W. N. [2006]. "Vulnerability". *Global environmental change*. Vol. 16. no. 3, pp. 268–281.
- Ahern, M., Kovats, R. S., Wilkinson, P., Few, R., and Matthies, F. [2005]. "Global health impacts of floods: epidemiologic evidence". *Epidemiologic reviews*. Vol. 27. no. 1, pp. 36–46.
- Albino, V., Carbonara, N., and Giannoccaro, I. [2007]. "Supply chain cooperation in industrial districts: A simulation analysis". *European Journal of Operational Research*. Vol. 177. no. 1, pp. 261–280.
- Ali, G. A., Roy, A. G., Turmel, M.-C., and Courchesne, F. [2010]. "Multivariate analysis as a tool to infer hydrologic response types and controlling variables in a humid temperate catchment". *Hydrological processes*. Vol. 24. no. 20, pp. 2912–2923.
- Allen, K. [2003]. "Vulnerability reduction and the community-based approach". *Natural disasters and development in a globalizing world*. P. 170.
- AMS, A. M. S. [2000]. "Policy statement: Prediction and mitigation of flash floods." *Bull. Amer. Meteor. Soc.*. Vol. 81, pp. 1338–1340.
- Anderson, M., Burt, T., et al. [1990]. "Subsurface runoff." *Process studies in hillslope hydrology*. Pp. 365–400.
- Annan, K. [2003]. "Message for the international day for disaster reduction 8 October 2003". *Source: http://www.unisdr.org/eng/public_aware/world_camp/2003/pa-camp03-sg-eng.htm*. Last assessed. Vol. 15. no. 11, p. 2005.
- Ashley, S. T. and Ashley, W. S. [2008]. "Flood fatalities in the United States". *Journal of Applied Meteorology and Climatology*. Vol. 47. no. 3, pp. 805–818.
- Ashley, W. S. [2007]. "Spatial and temporal analysis of tornado fatalities in the United States: 1880-2005". *Weather and Forecasting*. Vol. 22. no. 6, pp. 1214–1228.
- Ashley, W. S. and Gilson, C. W. [2009]. "A reassessment of US lightning mortality". *Bulletin of the American Meteorological Society*. Vol. 90. no. 10, pp. 1501–1518.
- Ashley, W. S. and Strader, S. M. [2016]. "Recipe for disaster: how the dynamic ingredients of risk and exposure are changing the tornado disaster landscape". *Bulletin of the American Meteorological Society*. Vol. 97. no. 5, pp. 767–786.
- Aubrecht, C., Özceylan, D., Steinnocher, K., and Freire, S. [2013]. "Multi-level geospatial modeling of human exposure patterns and vulnerability indicators". *Natural hazards*. Vol. 68. no. 1, pp. 147–163.
- Avila, L. [2016]. "The 2015 Eastern North Pacific Hurricane Season: A Very Active Year". *Weatherwise*. Vol. 69. no. 3, pp. 36–42.
- Azar, D. and Rain, D. [2007]. "Identifying population vulnerable to hydrological hazards in San Juan, Puerto Rico". *GeoJournal*. Vol. 69. no. 1-2, pp. 23–43.
- Beck, E., Dugdale, J., Van Truong, H., Adam, C., and Colbeau-Justin, L. [2014]. "Crisis mobility of pedestrians: from survey to modelling, lessons from Lebanon and Argentina". *International Conference on Information Systems for Crisis Response and Management in Mediterranean Countries*. Springer, pp. 57–70.
- Becker, J., Taylor, H., Doody, B., Wright, K., Grunfest, E., and Webber, D [2015]. "A Review of People's Behavior in and around Floodwater". *Weather, Climate, and Society*. Vol. 7. no. 4, pp. 321–332.
- Belmonte, A. M. C., López-García, M. J., and Soriano-García, J. [2011]. "Mapping temporally-variable exposure to flooding in small Mediterranean basins using land-use indicators". *Applied Geography*. Vol. 31. no. 1, pp. 136–145.

- Bhaduri, B., Bright, E., Coleman, P., and Urban, M. L. [2007]. "LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics". *GeoJournal*. Vol. 69. no. 1-2, pp. 103–117.
- Birkmann, J. [2006a]. "Indicators and criteria for measuring vulnerability: Theoretical bases and requirements". *Measuring vulnerability to natural hazards: Towards disaster resilient societies*. Pp. 55–77.
- Birkmann, J. [2006b]. "Measuring vulnerability to natural hazards: towards disaster resilient societies". no. ISBN-10 92-808-1135-5. United Nations University Press.
- Blaikie, P., Cannon, T., Davis, I., and Wisner, B. [1994]. "At risk: natural hazards, people's vulnerability and disasters". London: Routledge.
- Blaikie, P., Cannon, T., Davis, I., and Wisner, B. [2014]. "At risk: natural hazards, people's vulnerability and disasters". Routledge.
- Blöschl, G. and Sivapalan, M [1995]. "Scale issues in hydrological modelling: a review". *Hydrological processes*. Vol. 9. no. 3-4, pp. 251–290.
- Bogardi, J. and Birkmann, J. [2004]. "Vulnerability assessment: the first step towards sustainable risk reduction". *Disaster and Society—From Hazard Assessment to Risk Reduction*. Logos Verlag Berlin, Berlin. Pp. 75–82.
- Bohle, H.-G. [2001]. "Vulnerability and criticality: perspectives from social geography". *IHDP update*. Vol. 2. no. 01, pp. 3–5.
- Bollin, C., Cárdenas, C., Hahn, H., and Vatsa, K. S. [2003]. "Disaster risk management by communities and local governments". Tech. rep. Inter-American Development Bank.
- Borden, K. A. and Cutter, S. L. [2008]. "Spatial patterns of natural hazards mortality in the United States". *International Journal of Health Geographics*. Vol. 7. no. 1, p. 1.
- Borga, M., Anagnostou, E., Blöschl, G, and Creutin, J.-D. [2011]. "Flash flood forecasting, warning and risk management: the HYDRATE project". *Environmental Science & Policy*. Vol. 14. no. 7, pp. 834–844.
- Borga, M. [2009]. "Realtime guidance for flash flood risk management". Tech. rep. University of Padua.
- Breiman, L. [2001]. "Random forests". *Machine learning*. Vol. 45. no. 1, pp. 5–32.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. [1984]. "Classification and regression trees". CRC press.
- Brooks, H. E. and Doswell III, C. A. [2002]. "Deaths in the 3 May 1999 Oklahoma City tornado from a historical perspective". *Weather and Forecasting*. Vol. 17. no. 3, pp. 354–361.
- Brooks, N., Adger, W. N., and Kelly, P. M. [2005]. "The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation". *Global environmental change*. Vol. 15. no. 2, pp. 151–163.
- Burton, I., Kates, R. W., and White, G. F. [1978]. "The Environment as Hazard".
- Calianno, M., Ruin, I., and Gourley, J. J. [2013]. "Supplementing flash flood reports with impact classifications". *Journal of Hydrology*. Vol. 477, pp. 1–16.
- Cardona, O., Hurtado, J., Duque, G, Moreno, A, Chardon, A., Velásquez, L., and Prieto, S. [2003]. "The notion of disaster risk: conceptual framework for integrated management". *IADB/IDEA Program on Indicators for Disaster Risk Management*.

- Cardona, O. D. [2005]. “Indicators of Disaster Risk and Risk Management: Program for Latin America and the Caribbean: Summary Report”. Tech. rep. Inter-American Development Bank.
- Carpenter, T., Sperflage, J., Georgakakos, K., Sweeney, T, and Fread, D. [1999]. “National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems”. *Journal of Hydrology*. Vol. 224. no. 1, pp. 21–44.
- Carreño, M. L., Cardona, O. D., and Barbat, A. H. [2007]. “A disaster risk management performance index”. *Natural Hazards*. Vol. 41. no. 1, pp. 1–20.
- Carroll, B., Morbey, H., Balogh, R., and Araoz, G. [2009]. “Flooded homes, broken bonds, the meaning of home, psychological processes and their impact on psychological health in a disaster”. *Health & Place*. Vol. 15. no. 2, pp. 540–547.
- Center, H. [2002]. “Human links to coastal disasters”. *The H. John Heinz III Center for*.
- Chakraborty, J., Tobin, G. A., and Montz, B. E. [2005]. “Population evacuation: assessing spatial variability in geophysical risk and social vulnerability to natural hazards”. *Natural Hazards Review*. Vol. 6. no. 1, pp. 23–33.
- Chen, X., Meaker, J. W., and Zhan, F. B. [2006]. “Agent-based modeling and analysis of hurricane evacuation procedures for the Florida Keys”. *Natural Hazards*. Vol. 38. no. 3, pp. 321–338.
- Clark, G. E., Moser, S. C., Ratick, S. J., Dow, K., Meyer, W. B., Emani, S., Jin, W., Kasperson, J. X., Kasperson, R. E., and Schwarz, H. E. [1998]. “Assessing the vulnerability of coastal communities to extreme storms: the case of Revere, MA., USA”. *Mitigation and adaptation strategies for global change*. Vol. 3. no. 1, pp. 59–82.
- Clark, R. [2016]. “Machine learning predictions of flash floods”. PhD thesis. University of Oklahoma, School of Meteorology.
- Coates, L. [1999]. “Flood fatalities in Australia, 1788-1996”. *Australian Geographer*. Vol. 30. no. 3, pp. 391–408.
- Coninx, I. and Bachus, K. [2009]. “Exploring social flood impacts. Delphi study results”.
- Conover, W. [1971]. “Practical Nonparametric Statistics. New York”.
- Cools, M. and Creemers, L. [2013]. “The dual role of weather forecasts on changes in activity-travel behavior”. *Journal of Transport Geography*. Vol. 28, pp. 167–175.
- Creutin, J. D., Borga, M., Lutoff, C., Scolobig, A., Ruin, I., and Créton-Cazanave, L. [2009]. “Catchment dynamics and social response during flash floods: the potential of radar rainfall monitoring for warning procedures”. *Meteorological Applications*. Vol. 16. no. 1, pp. 115–125.
- Creutin, J. D., Borga, M., Grunfest, E., Lutoff, C., Zocatelli, D., and Ruin, I. [2013]. “A space and time framework for analyzing human anticipation of flash floods”. *Journal of Hydrology*. Vol. 482, pp. 14–24.
- Curran, E. B., Holle, R. L., and López, R. E. [2000]. “Lightning casualties and damages in the United States from 1959 to 1994”. *Journal of Climate*. Vol. 13. no. 19, pp. 3448–3464.
- Cutter, S., Boruff, B. J., and Shirley, W. L. [2006]. “Social vulnerability to environmental hazards”. *Hazards, Vulnerability, and Environmental Justice*. Pp. 115–132.
- Cutter, S. L. [1996]. “Vulnerability to environmental hazards”. *Progress in Human Geography*. Vol. 20. no. 4, pp. 529–539.

- Cutter, S. L. [2003]. “The vulnerability of science and the science of vulnerability”. *Annals of the Association of American Geographers*. Vol. 93. no. 1, pp. 1–12.
- Cutter, S. L., Mitchell, J. T., and Scott, M. S. [2000]. “Revealing the vulnerability of people and places: a case study of Georgetown County, South Carolina”. *Annals of the association of American Geographers*. Vol. 90. no. 4, pp. 713–737.
- Cutter, S. L., Boruff, B. J., and Shirley, W. L. [2003]. “Social vulnerability to environmental hazards”. *Social science quarterly*. Vol. 84. no. 2, pp. 242–261.
- Cutter, S. L., Emrich, C. T., Webb, J. J., and Morath, D. [2009]. “Social vulnerability to climate variability hazards: a review of the literature”. *Final Report to Oxfam America*. Vol. 5.
- Davidson, R. A. and Shah, H. C. [1997]. “An urban earthquake disaster risk index”. John A. Blume Earthquake Engineering Center Stanford University.
- Dawson, R. J., Peppe, R., and Wang, M. [2011]. “An agent-based model for risk-based flood incident management”. *Natural Hazards*. Vol. 59. no. 1, pp. 167–189.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. [1988]. “Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach”. *Biometrics*. Pp. 837–845.
- DHA, U. [1992]. “Internationally agreed glossary of basic terms related to disaster management”. *UN DHA (United Nations Department of Humanitarian Affairs), Geneva*.
- Diakakis, M. and Deligiannakis, G. [2013]. “Vehicle-related flood fatalities in Greece”. *Environmental Hazards*. Vol. 12. no. 3-4, pp. 278–290.
- Dietterich, T. G. [2000]. “Ensemble methods in machine learning”. International workshop on multiple classifier systems. Springer, pp. 1–15.
- Doocy, S, Daniels, A, Murray, S, and Kirsch, T. [2013]. “The human impact of floods: a historical review of events 1980–2009 and systematic literature review PLOS Curr”. *Disasters*. Vol. 5.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., et al. [2013]. “Collinearity: a review of methods to deal with it and a simulation study evaluating their performance”. *Ecography*. Vol. 36. no. 1, pp. 27–46.
- Downing, T [2004]. “What have we learned regarding a vulnerability science”. Science in Support of Adaptation to Climate Change. Recommendations for an Adaptation Science Agenda and a Collection of Papers Presented at a Side Event of the 10th Session of the Conference of the Parties to the United Nations Framework Convention on Climate Change, Buenos Aires. Pp. 18–21.
- Drobek, I. [2000]. “RESPONSES TO FLOOD WARNINGS”. *Floods*. Vol. 1, p. 361.
- Drobot, S. D., Benight, C, and Grunfest, E. [2007]. “Risk factors for driving into flooded roads”. *Environmental Hazards*. Vol. 7. no. 3, pp. 227–234.
- Duclos, P, Vidonne, O, Beuf, P, Perray, P, and Stoebner, A [1991]. “Flash flood disaster-nîmes, France, 1988”. *European journal of epidemiology*. Vol. 7. no. 4, pp. 365–371.
- Enarson, E. [2007]. “Identifying and addressing social vulnerabilities”. *Emergency Management: principles and practice for local government*,. Pp. 257–278.
- Fekete, A. [2009]. “Validation of a social vulnerability index in context to river-floods in Germany”. *Natural Hazards and Earth System Sciences*. Vol. 9. no. 2, pp. 393–403.

- Fekete, A. [2010]. "Assessment of Social Vulnerability for River Floods in Germany". United Nations University, Institute for Environment and Human Security (UNU-EHS).
- FEMA [2015]. "Flood: Basic Safety Tips". Available online at <https://www.ready.gov/floods>. Accessed 1 July 2016.
- Few, R. [2003]. "Flooding, vulnerability and coping strategies: local responses to a global threat". *Progress in Development Studies*. Vol. 3. no. 1, pp. 43–58.
- FHWA, F. H. A. [2009]. "Highway Statistics 2009". Policy and Governmental Affairs, Highway Policy Information, Office of Highway Policy Information. Available online at <http://www.ready.gov/floods>. Accessed 1 July 2016.
- Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., and Combs, B. [1978]. "How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits". *Policy sciences*. Vol. 9. no. 2, pp. 127–152.
- FitzGerald, G., Du, W., Jamal, A., Clark, M., and Hou, X.-y. [2010]. "Flood fatalities in contemporary Australia (1997–2008)". *Emergency Medicine Australasia*. Vol. 22. no. 2, pp. 180–186.
- Flamig, L. Z. [2016]. "A high resolution distributed hydrologic model climatology over the conterminous United States focused on flash flooding". PhD thesis. University of Oklahoma, School of Meteorology.
- Fothergill, A. [1998]. "The neglect of gender in disaster work: an overview of the literature". *The Gendered Terrain of Disaster: Through Women's Eyes*. Westport, CT, Praeger Publishers. pp11-25.
- Franklin, R. C., King, J. C., Aitken, P. J., and Leggat, P. A. [2014]. "Washed away: Assessing community perceptions of flooding and prevention strategies: a North Queensland example". *Natural Hazards*. Vol. 73. no. 3, pp. 1977–1998.
- French, J., Ing, R., Von Allmen, S., and Wood, R. [1983]. "Mortality from flash floods: a review of National Weather Service reports, 1969-81." *Public Health Reports*. Vol. 98. no. 6, p. 584.
- Gall, M., Borden, K., and Cutter, S. [2009]. "When do losses count? Six fallacies of loss data from natural hazards". *Bull. Am. Meteorol. Soc.* Vol. 90, pp. 1–11.
- Gallopin, G. C. [1997]. "Indicators and their use: information for decision-making". *SCOPE-SCIENTIFIC COMMITTEE ON PROBLEMS OF THE ENVIRONMENT INTERNATIONAL COUNCIL OF SCIENTIFIC UNIONS*. Vol. 58, pp. 13–27.
- Gamow, G. [1970]. "My world line: An informal autobiography." *My world line: An informal autobiography., by Gamow, G.. New York, NY (USA): Viking Press, 178 p..* Vol. 1.
- Gardner, E., Kimpel, T., and Zhao, Y. [2010]. "AMERICAN COMMUNITY SURVEY User Guide ACS Publication No."
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., Blaškovičová, L., Blöschl, G., Borga, M., Dumitrescu, A., et al. [2009]. "A compilation of data on European flash floods". *Journal of Hydrology*. Vol. 367. no. 1, pp. 70–78.
- Genuer, R., Poggi, J.-M., and Tuleau-Malot, C. [2010]. "Variable selection using random forests". *Pattern Recognition Letters*. Vol. 31. no. 14, pp. 2225–2236.
- Gissing, A., Haynes, K., Coates, L., and Keys, C. [2016]. "Motorist behaviour during the 2015 Shoalhaven floods". *Australian Journal of Emergency Management*. Vol. 31, pp. 23–27.
- Golden, L. [2001]. "Flexible work schedules which workers get them?": *American Behavioral Scientist*. Vol. 44. no. 7, pp. 1157–1178.

- Gourley, J., Erlingis, J., Smith, T., Ortega, K., and Hong, Y. [2010]. “Remote collection and analysis of witness reports on flash floods”. *Journal of hydrology*. Vol. 394. no. 1, pp. 53–62.
- Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P.-E., et al. [2013]. “A unified flash flood database across the United States”. *Bulletin of the American Meteorological Society*. Vol. 94. no. 6, pp. 799–805.
- Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P.-E., Clark III, R. A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., et al. [2017]. “The Flooded Locations And Simulated Hydrographs (FLASH) project: improving the tools for flash flood monitoring and prediction across the United States”. *Bulletin of the American Meteorological Society*. in press.
- Gruntfest, E. and Handmer, J. [2001]. “Coping with flash floods”. Vol. 77. Springer Science & Business Media.
- Gruntfest, E. C. and Drainage, U. [1977]. “WHAT PEOPLE DID DURING THÉ BIG-THOMPSON FLOOD”. Tech. rep. Working paper.
- Hägerstrand, T. [1970]. “Wath about people in regional science?”. *Regional Science Association*. Vol. 24.
- Haki, Z. G. [2003]. “Assessment of social vulnerability using geographic information systems: Pendik, Istanbul case study”. PhD thesis. MIDDLE EAST TECHNICAL UNIVERSITY.
- Hanifan, L. J. [1916]. “The rural school community center”. *The Annals of the American Academy of Political and Social Science*. Vol. 67, pp. 130–138.
- Hapuarachchi, H., Wang, Q., and Pagano, T. [2011]. “A review of advances in flash flood forecasting”. *Hydrological Processes*. Vol. 25. no. 18, pp. 2771–2784.
- Hastie, T., Tibshirani, R., and Friedman, J. [2009]. “Unsupervised learning”. *The elements of statistical learning*. Springer, pp. 485–585.
- Holmes Jr, R. R. and Wagner, D. M. [2011]. “Flood of June 11, 2010, in the Upper Little Missouri River watershed, Arkansas”. *US Geological Survey Scientific Investigations Report*. Vol. 5194, p. 31.
- Huang, Z. [1998]. “Extensions to the k-means algorithm for clustering large data sets with categorical values”. *Data mining and knowledge discovery*. Vol. 2. no. 3, pp. 283–304.
- IPCC, I. P. O. C. C. [2001]. “Climate change 2007: impacts, adaptation and vulnerability”. *Genebra, Suíça*.
- ISDR, U. [2005]. “Hyogo framework for action 2005-2015: building the resilience of nations and communities to disasters”. Extract from the final report of the World Conference on Disaster Reduction (A/CONF. 206/6). Vol. 380.
- Jonkman, S. N. [2005]. “Global perspectives on loss of human life caused by floods”. *Natural hazards*. Vol. 34. no. 2, pp. 151–175.
- Jonkman, S. N. and Kelman, I. [2005]. “An analysis of the causes and circumstances of flood disaster deaths”. *Disasters*. Vol. 29. no. 1, pp. 75–97.
- Jonkman, S. N., Maaskant, B., Boyd, E., and Levitan, M. L. [2009]. “Loss of life caused by the flooding of New Orleans after Hurricane Katrina: analysis of the relationship between flood characteristics and mortality”. *Risk Analysis*. Vol. 29. no. 5, pp. 676–698.
- Jonkman, S. [2003]. “Loss of life caused by floods: an overview of mortality statistics for worldwide floods”. Tech. rep. Delft Cluster.
- Jonkman, S. and Vrijling, J. [2008]. “Loss of life due to floods”. *Journal of flood risk management*. Vol. 1. no. 1, pp. 43–56.

- Jonkman, S., Van Gelder, P., and Vrijling, J. [2002]. “Loss of life models for sea and river floods”. *Flood defence*. Vol. 1, pp. 196–206.
- Karagiorgos, K., Thaler, T., Hübl, J., Maris, F., and Fuchs, S. [2016]. “Multi-vulnerability analysis for flash flood risk management”. *Natural Hazards*. Vol. 82. no. 1, pp. 63–87.
- Kellar, D. and Schmidlin, T. [2012]. “Vehicle-related flood deaths in the United States, 1995–2005”. *Journal of Flood Risk Management*. Vol. 5. no. 2, pp. 153–163.
- Kienberger, S. [2007]. “Assessing the vulnerability to natural hazards on the provincial/community level in Mozambique: the contribution of GIScience and remote sensing”. Joint CIG/ISPRS conference on geomatics for disaster and risk management, Toronto, Canada. Citeseer.
- Kilpeläinen, M. and Summala, H. [2007]. “Effects of weather and weather forecasts on driver behaviour”. *Transportation research part F: traffic psychology and behaviour*. Vol. 10. no. 4, pp. 288–299.
- Kim, D.-W., Lee, K. H., and Lee, D. [2004]. “Fuzzy clustering of categorical data using fuzzy centroids”. *Pattern Recognition Letters*. Vol. 25. no. 11, pp. 1263–1271.
- King, D., MacGregor, C., et al. [2000]. “Using social indicators to measure community vulnerability to natural hazards”. *Australian Journal of Emergency Management, The*. Vol. 15. no. 3, p. 52.
- Knocke, E. T. and Kolivras, K. N. [2007]. “Flash flood awareness in southwest Virginia”. *Risk analysis*. Vol. 27. no. 1, pp. 155–169.
- Kohavi, R. and Provost, F. [1998]. “Glossary of terms”. *Machine Learning*. Vol. 30. no. 2-3, pp. 271–274.
- Koks, E. E., Jongman, B., Husby, T. G., and Botzen, W. J. [2015]. “Combining hazard, exposure and social vulnerability to provide lessons for flood risk management”. *Environmental Science & Policy*. Vol. 47, pp. 42–52.
- Konrad, C. E. [2001]. “The most extreme precipitation events over the eastern United States from 1950 to 1996: considerations of scale”. *Journal of Hydrometeorology*. Vol. 2. no. 3, pp. 309–325.
- Kuhlicke, C., Scolobig, A., Tapsell, S., Steinführer, A., and De Marchi, B. [2011]. “Contextualizing social vulnerability: findings from case studies across Europe”. *Natural Hazards*. Vol. 58. no. 2, pp. 789–810.
- Lazrus, H., Morss, R. E., Demuth, J. L., Lazo, J. K., and Bostrom, A. [2016]. ““Know What to Do If You Encounter a Flash Flood”: Mental Models Analysis for Improving Flash Flood Risk Communication and Public Decision Making”. *Risk analysis*. Vol. 36. no. 2, pp. 411–427.
- Liaw, A. and Wiener, M. [2002]. “Classification and regression by randomForest”. *R news*. Vol. 2. no. 3, pp. 18–22.
- Lindell, M. K. and Perry, R. W. [1991]. “Understanding evacuation behavior: an editorial introduction”. *International journal of mass emergencies and disasters*. Vol. 9. no. 2, pp. 133–136.
- Lindell, M. K. and Perry, R. W. [2003]. “Communicating environmental risk in multiethnic communities”. Vol. 7. Sage Publications.
- Llasat, M., Llasat-Botija, M., Prat, M., Porcú, F., Price, C, Mugnai, A, Lagouvardos, K, Kotroni, V, Katsanos, D, Michaelides, S, et al. [2010]. “High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database”. *Advances in Geosciences*. Vol. 23, pp. 47–55.
- Lumbroso, D. and Tagg, A [2011]. “Evacuation and loss of life modelling to enhance emergency response”.

- Maples, L. Z. and Tiefenbacher, J. P. [2009]. “Landscape, development, technology and drivers: The geography of drownings associated with automobiles in Texas floods, 1950–2004”. *Applied geography*. Vol. 29. no. 2, pp. 224–234.
- Marchi, L., Borga, M., Preciso, E., and Gaume, E. [2010]. “Characterisation of selected extreme flash floods in Europe and implications for flood risk management”. *Journal of Hydrology*. Vol. 394. no. 1, pp. 118–133.
- Martens, T and Ramm, K [2007]. “Newsletter 2: Weitere Ergebnisse der Telefonbefragung”.
- Martinaitis, S. M., Gourley, J. J., Flamig, Z. L., Argyle, E. M., Clark III, R. A., Arthur, A., Smith, B. R., Erlingis, J. M., Perfater, S., and Albright, B. [2017]. “The HMT Multi-Radar Multi-Sensor Hydro Experiment”. *Bulletin of the American Meteorological Society*.
- Mason, S. J. and Graham, N. E. [2002]. “Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation”. *Quarterly Journal of the Royal Meteorological Society*. Vol. 128. no. 584, pp. 2145–2166.
- McCarthy, J. J. [2001]. “Climate change 2001: impacts, adaptation, and vulnerability: contribution of Working Group II to the third assessment report of the Intergovernmental Panel on Climate Change”. Cambridge University Press.
- McGuire, L. C., Ford, E. S., and Okoro, C. A. [2007]. “Natural disasters and older US adults with disabilities: implications for evacuation”. *Disasters*. Vol. 31. no. 1, pp. 49–56.
- McMenamin, T. M. [2007]. “Time to work: recent trends in shift work and flexible schedules, A”. *Monthly Lab. Rev.*. Vol. 130, p. 3.
- Merz, B, Kreibich, H, and Lall, U [2013]. “Multi-variate flood damage assessment: a tree-based data-mining approach”. *Natural Hazards and Earth System Sciences*. Vol. 13. no. 1, pp. 53–64.
- Metzger, M. J., Leemans, R., and Schröter, D. [2005]. “A multidisciplinary multi-scale framework for assessing vulnerabilities to global change”. *International Journal of Applied Earth Observation and Geoinformation*. Vol. 7. no. 4, pp. 253–267.
- Mileti, D. S. [1995]. “Factors related to flood warning response”. US-Italy Research Workshop on the Hydrometeorology, Impacts, and Management of Extreme Floods. Citeseer, pp. 1–17.
- Montz, B. E. and Gruntfest, E. [2002]. “Flash flood mitigation: recommendations for research and applications”. *Global Environmental Change Part B: Environmental Hazards*. Vol. 4. no. 1, pp. 15–22.
- Mooney, L. [1983]. “Applications and implications of fatality statistics to the flash flood problems”. Proceedings of the 5th Conference on Hydrometeorology. Pp. 127–129.
- Morrow, B. H. [1999]. “Identifying and mapping community vulnerability”. *Disasters*. Vol. 23. no. 1, pp. 1–18.
- Morss, R. E., Mulder, K. J., Lazo, J. K., and Demuth, J. L. [2015]. “How do people perceive, understand, and anticipate responding to flash flood risks and warnings? Results from a public survey in Boulder, Colorado, USA”. *Journal of Hydrology*.
- Müller, A, Reiter, J, and Weiland, U [2011]. “Assessment of urban vulnerability towards floods using an indicator-based approach—a case study for Santiago de Chile.” *Natural Hazards & Earth System Sciences*. Vol. 11. no. 8.
- Nagarajan, M., Shaw, D., and Albores, P. [2012]. “Disseminating a warning message to evacuate: A simulation study of the behaviour of neighbours”. *European journal of operational research*. Vol. 220. no. 3, pp. 810–819.

- Naimi, B., Hamm, N. A., Groen, T. A., Skidmore, A. K., and Toxopeus, A. G. [2014]. “Where is positional uncertainty a problem for species distribution modelling?”: *Ecography*. Vol. 37. no. 2, pp. 191–203.
- Nayak, M. A. and Ghosh, S. [2013]. “Prediction of extreme rainfall event using weather pattern recognition and support vector machine classifier”. *Theoretical and applied climatology*. Vol. 114. no. 3-4, pp. 583–603.
- NHTS, N. H. T. S. [2009]. “Summary of Travel Trends: 2009 National Household Travel Survey”. U.S. Department of Transportation. Federal Highway Administration. Technical Report No. FHWA- PL-11-022, 2011. Available online at <http://nhts.ornl.gov/2009/pub/stt.pdf>. Accessed 1 July 2016.
- Nielsen-Gammon, J. W. [2015]. “The Faucet: Informal Attribution of the May 2015 Record-Setting Texas Rains”. 40th NOAA Annual Climate Diagnostics and Prediction Workshop. Science and Technology Infusion Climate Bulletin, pp. 8–10.
- NOAA [1998]. “Storm Data and Unusual Weather Phenomena with Late Reports and Corrections, October 1998”. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Climatic Data Center Rep. Vol. 40, No. 10. 146 pp. Available online at <https://www1.ncdc.noaa.gov/pub/orders/IPS/IPS-229ECA4D-4FE0-4343-B3D9-2A2C80583A13.pdf>. Accessed 1 July 2016.
- NOAA [1999]. “South Texas Floods October 17-22, 1998”. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service Assessment, 34 pp. Available online at <http://www.nws.noaa.gov/om/assessments/pdfs/txflood.pdf>. Accessed 1 July 2016.
- NWS [2016]. “Storm data preparation. National Weather Service Instruction 10-1605”. NOAA/NWS. Available online at www.nws.noaa.gov/directives/sym/pd01016005curr.pdf. Accessed 30 December 2016.
- o’Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., Javed, A., Bhadwal, S., Barg, S., Nygaard, L., et al. [2004]. “Mapping vulnerability to multiple stressors: climate change and globalization in India”. *Global environmental change*. Vol. 14. no. 4, pp. 303–313.
- O’Connor, J. E. and Costa, J. E. [2004]. “The world’s largest floods, past and present: their causes and magnitudes”. Vol. 1254. US Geological Survey.
- O’Keefe, P., Westgate, K., and Wisner, B. [1976]. “Taking the naturalness out of natural disasters”. *Nature*. Vol. 260, pp. 566–567.
- Peacock, W. G., Morrow, B. H., and Gladwin, H. [1997a]. “Hurricane Andrew: Ethnicity, gender, and the sociology of disasters”. Psychology Press.
- Peacock, W. G., Morrow, B. H., and Gladwin, H. [1997b]. “Warning and evacuation: A night for hard houses”. *Hurricane Andrew: ethnicity, gender, and the sociology of disasters, New York: Routledge*. Pp. 52–74.
- Petersen, M. S. [2001]. “Impacts of flash floods”. *Coping with flash floods*. Springer, pp. 11–13.
- Phillips, B. D. and Morrow, B. H. [2007]. “Social science research needs: Focus on vulnerable populations, forecasting, and warnings”. *Natural Hazards Review*. Vol. 8. no. 3, pp. 61–68.
- Plant, C. and Böhm, C. [2011]. “Inconco: interpretable clustering of numerical and categorical objects”. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 1127–1135.
- Plate, E. J. [2002]. “Risk management for hydraulic systems under hydrological loads”. *Risk, reliability, uncertainty, and robustness of water resource systems*. Pp. 209–220.
- Priest, S. J., Tapsell, S. M., Penning-Rowsell, E. C., Viavattene, C., and Wilson, T. [2008]. “Building models to estimate loss of life for flood events: executive summary”.
- Quinlan, J. R. [1986]. “Induction of decision trees”. *Machine learning*. Vol. 1. no. 1, pp. 81–106.

- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. [2011]. “pROC: an open-source package for R and S+ to analyze and compare ROC curves”. *BMC bioinformatics*. Vol. 12. no. 1, p. 1.
- Rojas, R. [2013]. “Neural networks: a systematic introduction”. Springer Science & Business Media.
- Ronan, K. R. and Johnston, D. M. [2001]. “Correlates of hazard education programs for youth”. *Risk Analysis*. Vol. 21. no. 6, pp. 1055–1064.
- Rufat, S., Tate, E., Burton, C. G., and Maroof, A. S. [2015]. “Social vulnerability to floods: Review of case studies and implications for measurement”. *International Journal of Disaster Risk Reduction*. Vol. 14, pp. 470–486.
- Ruin, I. [2007]. “Conduite à contre-courant. Les pratiques de mobilité dans le Gard: facteur de vulnérabilité aux crues rapides”. PhD thesis. Université Joseph-Fourier-Grenoble I.
- Ruin, I. [2010]. “Conduite à contre-courant et crues rapides, le conflit du quotidien et de l’exceptionnel”. *Annales de géographie*. no. 4. Armand Colin, pp. 419–432.
- Ruin, I. and Lutoff, C. [2004]. “Vulnérabilité face aux crues rapides et mobilités des populations en temps de crise”. *La Houille Blanche*. Vol. 6, pp. 114–119.
- Ruin, I., Gaillard, J.-C., and Lutoff, C. [2007]. “How to get there? Assessing motorists’s flash flood risk perception on daily itineraries”. *Environmental hazards*. Vol. 7. no. 3, pp. 235–244.
- Ruin, I., Creutin, J.-D., Anquetin, S., and Lutoff, C. [2008]. “Human exposure to flash floods—Relation between flood parameters and human vulnerability during a storm of September 2002 in Southern France”. *Journal of Hydrology*. Vol. 361. no. 1, pp. 199–213.
- Ruin, I., Creutin, J.-D., Anquetin, S., Grunfest, E., and Lutoff, C. [2009]. “Human vulnerability to flash floods: Addressing physical exposure and behavioural questions”. Flood risk management: research and practice proceedings of the European Conference on Flood Risk Management Research into Practice (FLOODrisk 2008), Oxford, UK, 30 September–2 October 2008. Pp. 1005–1012.
- Ruin, I., Lutoff, C., Boudevillain, B., Creutin, J.-D., Anquetin, S., Rojo, M. B., Boissier, L., Bonnifait, L., Borga, M., Colbeau-Justin, L, et al. [2014]. “Social and hydrological responses to extreme precipitations: an interdisciplinary strategy for postflood investigation”. *Weather, climate, and society*. Vol. 6. no. 1, pp. 135–153.
- Ryan, T and Hanes, S [2009]. “North Texas flash flood characteristics”. National Weather Service, Fort Worth Forecast Office, 1–14. Available online at <http://www.srh.noaa.gov/images/fwd/pdf/ffpaper.pdf>. Accessed 1 July 2016.
- Rygel, L., O’Sullivan, D., and Yarnal, B. [2006]. “A method for constructing a social vulnerability index: an application to hurricane storm surges in a developed country”. *Mitigation and adaptation strategies for global change*. Vol. 11. no. 3, pp. 741–764.
- Saharia, M, Kirstetter, P., Vergara, H, Gourley, J., Hong, Y, and Giroud, M [2016]. “Mapping Flash Flood Severity in the United States”. *Journal of hydrometeorology*. doi: <http://dx.doi.org/10.1175/JHM-D-16-0082.1>.
- Sarewitz, D., Pielke, R., and Keykhah, M. [2003]. “Vulnerability and risk: some thoughts from a political and policy perspective”. *Risk analysis*. Vol. 23. no. 4, pp. 805–810.
- Schneiderbauer, S. and Ehrlich, D. [2004]. “Risk, hazard and people’s vulnerability to natural hazards”. *A Review of Definitions, Concepts and Data. European Commission Joint Research Centre. EUR*. Vol. 21410, p. 40.

- Shabou, S. [2016]. “Extrêmes hydro-météorologiques and Exposition sur les routes. Contribution à MobRISK: Modèle de simulation de l’exposition des mobilités quotidiennes aux crues rapides”. PhD thesis. Université Grenoble Alpes, Géographie.
- Shankar, V., Mannering, F., and Barfield, W. [1995]. “Effect of roadway geometrics and environmental factors on rural freeway accident frequencies”. *Accident Analysis & Prevention*. Vol. 27. no. 3, pp. 371–389.
- Sharif, H. O., Hossain, M. M., Jackson, T., and Bin-Shafique, S. [2012]. “Person-place-time analysis of vehicle fatalities caused by flash floods in Texas”. *Geomatics, Natural Hazards and Risk*. Vol. 3. no. 4, pp. 311–323.
- Sharif, H. O., Jackson, T. L., Hossain, M. M., and Zane, D. [2014]. “Analysis of flood fatalities in Texas”. *Natural Hazards Review*. Vol. 16. no. 1, p. 04014016.
- Simon, P. [2013]. “Too Big to Ignore: The Business Case for Big Data”. Vol. 72. John Wiley & Sons.
- Slovic, P. E. [2000]. “The perception of risk.” Earthscan publications.
- Smit, B. and Wandel, J. [2006]. “Adaptation, adaptive capacity and vulnerability”. *Global environmental change*. Vol. 16. no. 3, pp. 282–292.
- Spielman, S. E., Folch, D., and Nagle, N. [2014]. “Patterns and causes of uncertainty in the American Community Survey”. *Applied Geography*. Vol. 46, pp. 147–157.
- Špitalar, M., Gourley, J. J., Lutoff, C., Kirstetter, P.-E., Brilly, M., and Carr, N. [2014]. “Analysis of flash flood parameters and human impacts in the US from 2006 to 2012”. *Journal of Hydrology*. Vol. 519, pp. 863–870.
- Staes, C., Orengo, J. C., Malilay, J., Rullan, J., and Noji, E. [1994]. “Deaths due to flash floods in Puerto Rico, January 1992: implications for prevention”. *International journal of epidemiology*. Vol. 23. no. 5, pp. 968–975.
- Taillandier, P., Vo, D.-A., Amouroux, E., and Drogoul, A. [2010]. “GAMA: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control”. International Conference on Principles and Practice of Multi-Agent Systems. Springer, pp. 242–258.
- Talbott, M. D. [2015]. “2015 Memorial day flood in Harris County, Texas”. THC-IT-2015 Conference and Exhibition. Texas Hurricane Center for Innovative Technology, pp. 1–12.
- Tapsell, S., Tunstall, S., Green, C., and Fernández-Bilbao, A. [2005]. “Task 11: Social Indicator Set”. Tech. rep. FLOODsite, Integrated Flood Risk Analysis and Management Methodologies.
- Tapsell, S. M., Penning-Rowsell, E. C., Tunstall, S. M., and Wilson, T. [2002]. “Vulnerability to flooding: health and social dimensions”. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 360. no. 1796, pp. 1511–1525.
- Terti, G., Ruin, I., Anquetin, S., and Gourley, J. J. [2015]. “Dynamic vulnerability factors for impact-based flash flood prediction”. *Natural Hazards*. Vol. 79. no. 3, pp. 1481–1497.
- Terti, G., Ruin, I., Anquetin, S., and Gourley, J. J. [2016]. “A Situation-based Analysis of Flash Flood Fatalities in the United States”. *Bulletin of the American Meteorological Society*. doi:10.1175/BAMS-D-15-00276.1.
- Terti, G., Ruin, I., Gourley, J. J., Kirstetter, P.-E., Flamig, Z. L., Blanchet, J., Arthur, A., and Anquetin, S. [in review]. “Towards Probabilistic Prediction of Flash Flood Human Impacts”. *Risk Analysis*.
- Tierney, K. J., Lindell, M. K., and Perry, R. W. [2001]. “Facing the unexpected: Disaster preparedness and response in the United States”. Joseph Henry Press.

- Tobin, G. A. [1997]. "Natural hazards: explanation and integration". Guilford Press.
- Touw, W. G., Bayjanov, J. R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., and Hijum, S. A. van [2012]. "Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle?": *Briefings in bioinformatics*. bbs034.
- Trafalis, T. B., Adrianto, I., Richman, M. B., and Lakshmivarahan, S [2014]. "Machine-learning classifiers for imbalanced tornado data". *Computational Management Science*. Vol. 11. no. 4, pp. 403–418.
- Trujillo-Pagán, N. [2007]. "Katrina—¿Latinos: Vulnerability and disasters in relief and recovery". *Through the eye of Katrina: Social justice in the United States*. Pp. 147–168.
- Tunstall, S [2009]. "Vulnerability and flooding: a re-analysis of FHRC data. Country Report for England and Wales". Tech. rep. Middlesex University.
- Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., et al. [2003]. "A framework for vulnerability analysis in sustainability science". *Proceedings of the national academy of sciences*. Vol. 100. no. 14, pp. 8074–8079.
- UNDP, U. N. D. P. [2004]. "Reducing Disaster Risk: A Challenge for Development—a Global Report". United Nations. Bureau for Crisis Prevention and Recovery.
- UNDRO, O. o. t. U. N. D. R. C.-o. [1991]. "Mitigating Natural Disasters: Phenomena, Effects and Options". *A Manual for Policy Makers and Planners*. Pp. 3–28.
- UN/ISDR, I. S. f. D. R. [2009]. "Terminology on Disaster Risk Redution". United Nations International Strategy for Disaster Reduction (UNISDR), Geneva, Switzerland.
- U.S. Census Bureau [2010]. "Profile of General Population and Housing Characteristics". United States Census Bureau FactFinder. Available online at <https://www.fhwa.dot.gov/policyinformation/pubs/hf/p111028/chapter4.cfm>. Accessed 23 June 2015.
- U.S. Census Bureau [2016]. "U.S. and world population clock". Available online at www.census.gov/popclock/?intcmp=home_pop. Accessed 30 December 2016.
- Vapnik, V. and Cortes, C. [1995]. "Support Vector Networks, machine learning 20, 273-297".
- Versini, P.-A., Gaume, E., and Andrieu, H. [2010]. "Assessment of the susceptibility of roads to flooding based on geographical information—test in a flash flood prone area (the Gard region, France)". *Natural Hazards and Earth System Sciences*. Vol. 10. no. 4, pp. 793–803.
- Villarini, G., Goska, R., Smith, J. A., and Vecchi, G. A. [2014]. "North Atlantic tropical cyclones and US flooding". *Bulletin of the American Meteorological Society*. Vol. 95. no. 9, pp. 1381–1388.
- Vinet, F, Lumbroso, D, Defossez, S, and Boissier, L [2012]. "A comparative analysis of the loss of life during two recent floods in France: the sea surge caused by the storm Xynthia and the flash flood in Var". *Natural hazards*. Vol. 61. no. 3, pp. 1179–1201.
- Vinet, F. [2008]. "Geographical analysis of damage due to flash floods in southern France: The cases of 12–13 November 1999 and 8–9 September 2002". *Applied Geography*. Vol. 28. no. 4, pp. 323–336.
- Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., et al. [2011]. "The coupled routing and excess storage (CREST) distributed hydrological model". *Hydrological sciences journal*. Vol. 56. no. 1, pp. 84–98.
- Wei, W. and Watkins, D. W. [2011]. "Data mining methods for hydroclimatic forecasting". *Advances in Water Resources*. Vol. 34. no. 11, pp. 1390–1400.

- White, G. F. and Haas, J. E. [1975]. “Assessment of Research on Natural Hazards”.
- Wilhelmi, O. V. and Morss, R. E. [2013]. “Integrated analysis of societal vulnerability in an extreme precipitation event: A Fort Collins case study”. *Environmental science & policy*. Vol. 26, pp. 49–62.
- Wisner, B. [2003]. “Disaster risk reduction in megacities: making the most of human and social capital”. *Building safer cities: The future of disaster risk*. Pp. 181–196.
- Wisner, B., Blaikie, P., Cannon, T., and Davis, I. [2004]. “At risk: natural hazards, people’s vulnerability and disasters”. 2nd ed. London-New York: Routledge.
- WMO [2008]. “Capacity Assessment of National Meteorological and Hydrological Services in Support of Disaster Risk Reduction”. Tech. rep. World Meteorological Organization.
- Wu, J. and Li, H. [2006]. “Concepts of scale and scaling”. *Scaling and uncertainty analysis in ecology*. Springer, pp. 3–15.
- Wu, S.-Y., Yarnal, B., and Fisher, A. [2002]. “Vulnerability of coastal communities to sea-level rise: a case study of Cape May County, New Jersey, USA”. *Climate Research*. Vol. 22. no. 3, pp. 255–270.
- Yap, B. W., Rani, K. A., Rahman, H. A. A., Fong, S., Khairudin, Z., and Abdullah, N. N. [2014]. “An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets”. *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. Springer, pp. 13–22.
- Yussouf, N., Kain, J. S., and Clark, A. J. [2016]. “Short-term Probabilistic Forecasts of the 31 May 2013 Oklahoma Tornado and Flash Flood Event Using a Continuous-Update-Cycle Storm-scale Ensemble System”. *Weather and Forecasting*. no. 2016.
- Zahran, S., Brody, S. D., Peacock, W. G., Vedlitz, A., and Grover, H. [2008]. “Social vulnerability and the natural and built environment: a model of flood casualties in Texas”. *Disasters*. Vol. 32. no. 4, pp. 537–560.
- Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., Van Cooten, S., Kelleher, K., Kitzmiller, D., Ding, F., et al. [2011]. “National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans”. *Bulletin of the American Meteorological Society*. Vol. 92. no. 10, p. 1321.
- Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., et al. [2016]. “Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities”. *Bulletin of the American Meteorological Society*. Vol. 97. no. 4, pp. 621–638.
- Zhong, S., Clark, M., Hou, X.-Y., Zang, Y.-L., and FitzGerald, G. [2013]. “2010–2011 Queensland floods: using Haddon’s Matrix to define and categorise public safety strategies”. *Emergency Medicine Australasia*. Vol. 25. no. 4, pp. 345–352.
- Zuur, A. F., Ieno, E. N., and Elphick, C. S. [2010]. “A protocol for data exploration to avoid common statistical problems”. *Methods in Ecology and Evolution*. Vol. 1. no. 1, pp. 3–14.

APPENDIX

Frequency of flash flood victims and fatal flash flood events by state

From 1996 to 2014, the *Storm Data* reported 705 flash flood events that led to one or more fatalities. The frequency of fatal flash flood events and the caused fatalities is summarized by state and sorted by the number of fatalities in decreasing order (Table A.1). The analysis includes the 48 conterminous states, as well as the states of Alaska and Hawaii, the District of Columbia, and the territory of Puerto Rico.

Table A.1 – Frequency and percentages of flash flood fatalities and fatal flash flood events by state for the 50 states, the District of Columbia, and the territory of Puerto Rico (including their ranks in fatalities) for the period 1996-2014. Ratios are estimated by dividing the number of fatalities with the number of fatal flash flood events in each state.

State	Abbrev	Fatalities	Fatalities (%)	Events	Events (%)	Ratio
TEXAS	TX	207	19.3	136	19.3	1.5
ARIZONA	AZ	69	6.4	40	5.7	1.7
MISSOURI	MO	62	5.8	44	6.2	1.4
ARKANSAS	AR	56	5.2	26	3.7	2.2
PENNSYLVANIA	PA	49	4.6	32	4.5	1.5
KENTUCKY	KY	45	4.2	32	4.5	1.4
NORTH CAROLINA	NC	41	3.8	18	2.6	2.3
CALIFORNIA	CA	41	3.8	29	4.1	1.4

Table A.1 Continued

State	Abbrev	Fatalities	Fatalities (%)	Events	Events (%)	Ratio
PUERTO RICO	PR	40	3.7	25	3.5	1.6
NEW YORK	NY	37	3.4	21	3	1.8
OKLAHOMA	OK	37	3.4	21	3	1.8
OHIO	OH	32	3	22	3.1	1.5
TENNESSEE	TN	31	2.9	22	3.1	1.4
VIRGINIA	VA	30	2.8	24	3.4	1.2
WEST VIRGINIA	WV	23	2.1	17	2.4	1.4
NEW MEXICO	NM	21	2	15	2.1	1.4
COLORADO	CO	20	1.9	11	1.6	1.8
INDIANA	IN	18	1.7	14	2	1.3
ALABAMA	AL	16	1.5	12	1.7	1.3
MINNESOTA	MN	15	1.4	8	1.1	1.9
KANSAS	KS	15	1.4	9	1.3	1.7
GEORGIA	GA	14	1.3	9	1.3	1.6
UTAH	UT	14	1.3	11	1.6	1.3
MARYLAND	MD	13	1.2	5	0.7	2.6
MISSISSIPPI	MS	13	1.2	12	1.7	1.1
HAWAII	HI	11	1	4	0.6	2.8
ILLINOIS	IL	11	1	8	1.1	1.4
FLORIDA	FL	9	0.8	8	1.1	1.1
LOUISIANA	LA	8	0.7	8	1.1	1
SOUTH CAROLINA	SC	7	0.7	6	0.9	1.2
WISCONSIN	WI	7	0.7	6	0.9	1.2
IOWA	IA	7	0.7	7	1	1
NEW HAMPSHIRE	NH	6	0.6	4	0.6	1.5
NEVADA	NV	6	0.6	6	0.9	1
WYOMING	WY	5	0.5	2	0.3	2.5
MICHIGAN	MI	5	0.5	4	0.6	1.2
NEW JERSEY	NJ	5	0.5	4	0.6	1.2
SOUTH DAKOTA	SD	4	0.4	3	0.4	1.3
ALASKA	AK	3	0.3	1	0.1	3

Table A.1 Continued

State	Abbrev	Fatalities	Fatalities (%)	Events	Events (%)	Ratio
NORTH DAKOTA	ND	3	0.3	2	0.3	1.5
VERMONT	VT	3	0.3	2	0.3	1.5
MAINE	ME	3	0.3	3	0.4	1
NEBRASKA	NE	3	0.3	3	0.4	1
DELAWARE	DE	2	0.2	1	0.1	2
CONNECTICUT	CT	2	0.2	2	0.3	1
IDAHO	ID	2	0.2	2	0.3	1
WASHINGTON	WA	2	0.2	2	0.3	1
MONTANA	MT	1	0.1	1	0.1	1
OREGON	OR	1	0.1	1	0.1	1
DIST OF COLUMBIA	DC	0		0		
MASSACHUSSETS	MA	0		0		
RHODE ISLAND	RI	0		0		

Type and sources of collected data, and their role in the assessment of human vulnerability and risk to flash flood

In this study, about thirteen different databases are gathered and processed to provide indicators related to hydrological and meteorological responses, geomorphological characteristics, information for the available official emergency response, and other spatial and socio-demographic attributes. According to Table B.1, different origin data are assumed to contribute to different aspects of the assessment of human vulnerability, depicting the overall view of human risk during the “event” phase of flash floods. The data are originally provided in various spatial resolutions (e.g., 1-km grid cell, point geodata, county administrative units), and are all adjusted to the county-level to proceed for the analysis.

Table B.1 – Summary of collected data types and sources, and their role in the assessment of human vulnerability and risk to flash flood. Indicators are proposed as relevant to one or more circumstances (ALL: all circumstances, VE: Vehicle-related, OU: Outside/Open or close to stream area, CA: Camping/Recreational area, PB: Permanent Building, MH: Mobile Home).

Category of variables	Source/Origin	Indicators to be represented	Contribution
Hydrological conditions	CREST model simulations developed by the OU and the NSSL (NOAA), FLASH (https://blog.nssl.noaa.gov/flash/)	1. Magnitude of the flash flood event (ALL) 2. Time of flash flood occurrence (ALL)	Incorporation of the hazard spatial and temporal dynamics that alter human exposure to flash flooding
	Storm Data maintained by NOAA's NWS (https://www.ncdc.noaa.gov/stormevents/ftp.jsp)	3. Duration of flash flood event (ALL)	
Meteorological conditions	MRMS precipitation estimates developed by the CIMMS, and OU (http://mrms.ou.edu), FLASH	4. Magnitude of the rainfall event (ALL) 5. Duration of the rainfall event (ALL with focus on outside: VE, OU and CA)	
Geomorphological characteristics	National Flood Hazard Layer (NFHL) from FEMA's NFIP (https://www.fema.gov/national-flood-hazard-layer-nfhl), HSIP	6. Flood hazard areas (ALL)	Incorporation of flood predicament of the exposed area that can pre-define the potentiality for human exposure and coping capacities
	Flashiness Index developed in the NSSL (NOAA), FLASH	7. Flood severity (ALL)	

Table B.1 Continued

Category of variables	Source/Origin	Indicators to be represented	Contribution
Emergency re- sponse	Geospatial data produced by the TGS (currently CACI International Inc.) (http://www.caci.com/fcc/Geospatial/capabilities.shtml), HSIP	8. Official emergency services (ALL)	Incorporation of official emergency resources that support population's capacity to cope with flash flood crisis
Spatial information for the natural and built environment	LandScan USA Day/Night Population developed by the ORNL (http://web.ornl.gov/sci/landscan/), HSIP	9. Distribution of population (Daytime: ALL with focus on VE, OU; Nighttime: PB, MH)	Incorporation of densities of people and physical features that determine the spatial distribution of human exposure to flash flooding
	MAF/TIGER database (https://www.census.gov/geo/maps-data/data/tiger.html), Census Bureau	10. Road network (VE)	
	National Hydrography Dataset, NHDPlus (http://www.horizon-systems.com/nhdplus/NHDplusV2_data.php)/2010 TIGER Roads, Census Bureau	11. River-road network intersections (VE)	
	Wilderness Boundary Data, NWPS (http://www.wilderness.net/NWPS/geography)	12. Recreational areas (CA)	
	National Park Boundaries, EDAC (https://catalog.data.gov/dataset/national-park-boundariesf0a4c)		

Table B.1 Continued

Category of variables	Source/Origin	Indicators to be represented	Contribution
	Education and Public Health Centers from the ORNL, HSIP (https://gii.dhs.gov/HIFLD/public/HSIP-Gold-Freedom-One-Page-2015.pdf)	13. Special needs buildings (PB)	Incorporation of critical usage information for buildings that affect population's sensitivity and coping capacity during flash floods
Socio-economic and demographic characteristics for the population and households	ACS 5-year (2006-2010) Estimates provided at the American FactFinder (http://factfinder.census.gov/), Census Bureau	14. Age (ALL; for commuting workers: VE, OU; for workers working at home: PB) 15. Gender (Males:VE, OU; Females:MH) 16. Household family status (VE, PB) 17. Poverty Status (PB, MH) 18. Educational attainment (ALL) 19. Ethnicity/citizenship (VE, PB, MH) 20. Language (VE, OU, PB, MH) 21. Disability (PB, MH) 22. Special Needs population (PB) 23. Home ownership (PB, MH) 24. Length of residence (PB) 25. Mobile housing structure (MH)	Incorporation of the population and household sensitive characteristics that affect people's ability to react adequately, and the potential for better coping responses when exposed to flash flooding

Table B.1 Continued

Category of variables	Source/Origin	Indicators to be represented	Contribution
		26. Year housing structure built (PB)	
		27. Household phone availability (PB, MH)	
		28. Household vehicle availability (PB, MH)	
		29. Vehicles (VE)	
		30. Travel time to work (VE, OU)	
		31. Time arriving at work (VE, OU)	

Available U.S. Census Surveys and their geographical representation

The American Community Survey (ACS) is a part of the U.S. Census Bureau's Decennial Census Program and is designed to provide more current demographic, social, economic, and housing estimates throughout the decade (Table C.1). ACS estimates are periodic estimates that describe the average characteristics of population and housing over a period of data collection. For the 2010 5-year data products used in this study, interviews from January 1, 2006 through December 31, 2010 were applied. ACS questionnaires are mailed out each month to a sample of residences. The monthly survey data are then combined to represent the characteristics of a population over periods of 1, 3, or 5 years. While all areas are continuously sampled each year, ACS data are not available for all areas in the 1-year and 3- year datasets due to confidentiality reasons [Gardner et al., 2010]. All areas are represented in the 5-year data products starting in 2010 (i.e., the 2005-2009 ACS release). Especially, the 5-year estimates are available for many distinct geographies including the nation, all 50 states, District of Columbia, and Puerto Rico, counties, places, census tracts (Figure C.1). In this study, human vulnerability to flash flood is assessed for the county-level population. As shown in Figure C.1 counties are connected to states because a state is comprised of many counties, and a county can never cross a state boundary.

Table C.1 – U.S. Census Bureau’s surveys explored for data availability and suitability in this dissertation.

Format	Dataset	Temporal resolution	Spatial resolution
<i>American Community Survey (annual survey)</i>			
Untabulated Sample individual-based data (sample of about 1% of the U.S. population)	Public Use Microdata Sample (PUMS)	1-year ACS PUMS	Region, Division, State, and Public Use Microdata Areas (PUMAs) with population >100,000
		3-years ACS PUMS	
		5-years ACS PUMS	
Pre-tabulated Sample data summarized by geographical unit (sample of about 2% of the U.S. population)	ACS Files	ACS 1-year Estimates	Areas with population >65,000
		ACS 3-years Estimates	Areas with population >20,000
		ACS 5-years Estimates	States, Counties, . . . , up to Census Tracts and Block Groups
<i>Decennial Census (Survey conducted every 10 years)</i>			
Pre-tabulated 100% survey data summarized by geographical unit	Decennial census SF1	Every 10 years period (e.g. 2000; 2010)	States, Counties, For most subjects, statistics for Census Block Groups and Blocks are also shown
	Decennial census SF2		Similar to SF1. To preserve confidentiality, only geographic entities with a population of at least 100 for the specified group are available in the summary file
Pre-tabulated sample survey data summarized by geographical unit	Decennial census SF3		Up to Census Tracts and Block Groups (but for selected subjects) until 2000 /replaced by ACS survey after 2005

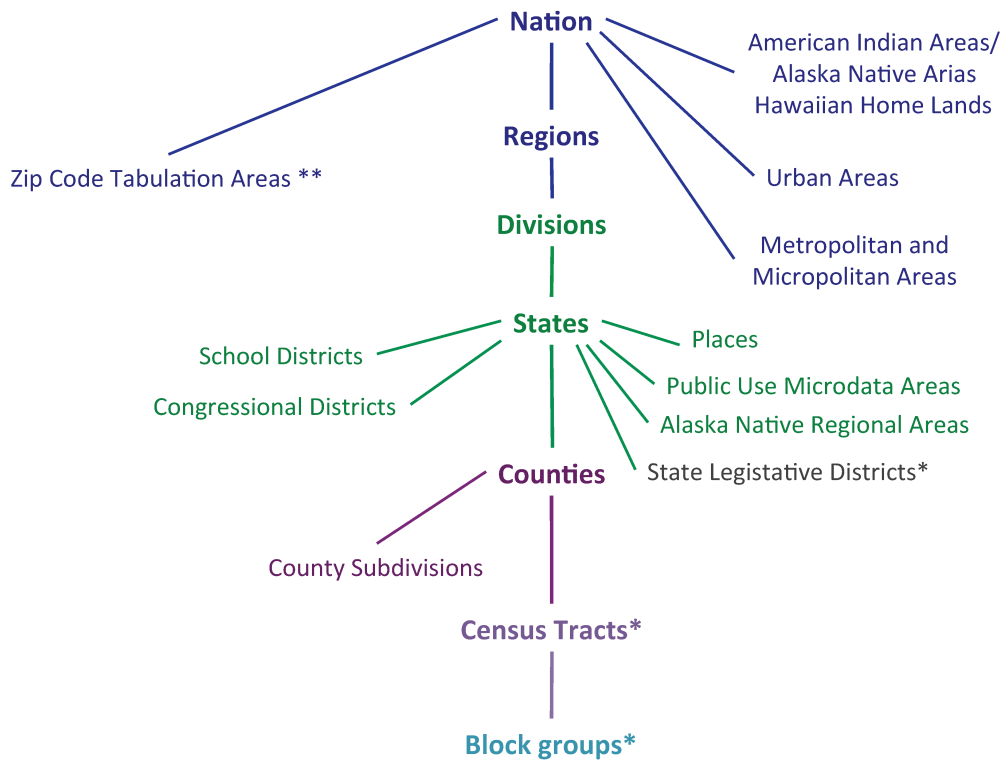


Figure C.1 – Hierarchy of Select Geographic Entities in the American Community Survey. Geographic areas with star (*) refer to areas for which data are only available in the 5-year estimates. Double star (**) notifies areas for which 5-year estimates are only available, and the first release was in 2012 (for the 2007-2011 5-year estimates). Source: Geographic Hierarchy as organized by the U.S. Census Bureau, available at <http://www.census.gov/programs-surveys/acs/geography-acs/concepts-definitions.html>. Retrieved on December 30, 2016.

Indicators and proxy variables for vehicle-related human risk to flash flood

To model the occurrence of human losses from flash flooding in vehicles we choose proxy variables relevant to such circumstances. Table D.1 summarizes the 41 variables selected from the supplemented database presented in Chapter III as the most representative ones of vehicle-related human risk to flash floods.

Table D.1 – Summary of processing and interpretation of proxy variables to serve as candidate predictors for flash flood events with vehicle-related human losses.

Variables	Processing	Risk hypothesis
<i>Indicator 1: Magnitude of the flash flood event</i>		
Daily Unit Peak Discharge ($\text{m}^3 \text{ s}^{-1} \text{ km}^{-2}$)	Computed by running the CREST distributed hydrologic model. The unit peak discharge for each day was stored, and has been aggregated for the county that the event occurred at the reported day(s).	Higher magnitudes are associated with higher water levels that inundate and expose a larger area to flooding.
<i>Indicator 2: Duration of the flash flood event</i>		
Duration of the flash flood event (in hours)	Estimated as the difference between the beginning and end local time of the flash flood event (e.g., 4, 1.17 hours) when provided in the <i>Storm Data</i> .	Higher magnitudes are associated with higher water levels that inundate and expose a larger area to flooding.

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 3: Magnitude of the rainfall event</i>		
Maximum accumulated precipitation (mm)	Extracted from MRMS system providing precipitation rate estimates across the CONUS at 1-km resolution with updates every five min. Aggregated for the county where the event occurred at the reported day(s).	Maximum rainfall is associated with adverse weather and road conditions exacerbating traffic accidents and vehicle-related risk [Shankar et al., 1995].
<i>Indicator 4: Duration of the rainfall event</i>		
Maximum duration of precipitation (in hours)	Estimated number of hours of the MRMS Precipitation >1.0mm in the day of the reported flash flood event.	The longer the precipitation lasts the more likely is the occurrence of floodwaters on impervious surfaces such as roads creating dangerous conditions for motorists.
<i>Indicator 5: Flood hazard areas</i>		
Area of high risk of flood hazard (in km ² and % to the total county area)	Calculated for each U.S. county after dissolving the flood hazard areas geodatabase based on the 2010 counties.	The existence of areas sensitive to flood risk indicates higher likelihood of severe flash flooding and impacts on the road network and its users.
Area of moderate to low risk of flood hazard (in km ² and % to the total county area)		
<i>Indicator 6: Flood severity</i>		
Mean flashiness (index)	Calculated mean flashiness index (i.e., values between 0-1) for each county. The original flashiness point data were converted to a 1-km float raster and after to a 1-km integer raster to calculate the mean in each U.S. county.	High flashiness index reveals the potentiality of high-magnitude discharge in a short period of time associated with severe flooding and limited anticipation time for people [Saharia et al., 2016].
<i>Indicator 7: Official emergency service</i>		
Number of Local Emergency Operation Centers (EOCs) (count)	Counted in each U.S. county using summary statistics on the spatially joined 2010 TIGER counties and EOC shapefiles from the HSIP.	The existence of local emergency services can likely contribute to more timely and efficient response leading to successful rescues from vehicles [Sharif et al., 2014].

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 8: Distribution of population</i>		
Daytime Population Density (people/km ²)	Calculated for each U.S. county by dividing the daytime population (provided by the ORNL's 90m-cell daytime population data) by the county land area from the 2010 TIGER counties shapefile.	The daily mobility and routine that creates differences in population density across space during the day defines the distribution of exposure [Belmonte et al., 2011].
<i>Indicator 9: Road network</i>		
Road Length (km)	Calculated for each road feature from the 2010 TIGER/Line road shapefile and aggregated by 2010 U.S. county.	The exposure of roads is inseparable linked with exposure of vehicle users. Road network sensitivity to inundation impedes rescue operations and limits the response capacity of drivers and passengers during flooding [Versini et al., 2010].
Road Density (km/km ²)	Calculated for each U.S. county by dividing the estimated road length (km) by the calculated county land area (km ²).	
<i>Indicator 10: River-road network intersections</i>		
Number of River-Road intersections (count)	Calculated as the intersection points of the merged 2010 TIGER road and NHD river/stream network shapefiles and aggregated for each U.S. county.	Crossings like bridges and low-water crossings are features sensitive to flash flooding largely linked to vehicle-related deaths in the U.S. [Kellar and Schmidlin, 2012]. Especially, when associated with low-visibility hours, drivers' ability to evaluate the conditions on high-risk locations of the road network is subsequently reduced [Jonkman and Kelman, 2005; Diakakis and Deligiannakis, 2013].

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 11: Age</i>		
People (count and % to the total county residential population):	Estimated for the each U.S. county by grouping age sub-groups provided in table DP05 of the county-level 2010 ACS 5-year estimates.	Very young and old population is always susceptible due to their physical constraints, and their dependency on others to deal with or escape from floodwaters [Coates, 1999; Morrow, 1999; Ashley and Ashley, 2008; Fekete, 2010]. But in majority young and middle-aged active population is more likely to be involved in vehicle-related incidents [Ruin et al., 2007; Kellar and Schmidlin, 2012; Diakakis and Deligiannakis, 2013; Terti et al., 2016]. Young drivers may be also less aware of flash flood risk [Knocke and Kolivras, 2007] and more confident to undertake risky behaviors towards crossing flooded roadways [Drobot et al., 2007].
-14 years or under (youth)		
-15 to 34 years (new drivers and young adults)		
-35 to 59 years (middle-aged active adults)		
-60 years or over (retired and elderly)		
Median age of residents (years)	Extracted from table DP05 of the county-level 2010 ACS 5-year estimates.	
Median age of workers (in the workplace county) (years)	Extracted from table B08503 of the county-level 2010 ACS 5-year estimates.	
Median age of workers commuting to work by vehicle in the workplace county (years)	Estimated by grouping the carpooled and drove alone classes from table B08503 of the county-level 2010 ACS 5-year estimates for workplace geography.	
<i>Indicator 12: Gender</i>		
Males (count and % to the total county population)	Extracted from table DP05 of the county-level 2010 ACS 5-year estimates.	Males are supposed to be more likely to be involved in emergency activities or to undertake risky behavior associated with entering floodwaters in vehicle [Coates, 1999; Jonkman and Kellman, 2005; Ashley and Ashley, 2008; FitzGerald et al., 2010; Kellar and Schmidlin, 2012; Doocy et al., 2013; Terti et al., 2016], and especially, driving through already barricaded roads [Diakakis and Deligiannakis, 2013; Gissing et al., 2016].

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 13: Household family status</i>		
Average household size	Extracted from table DP02 of the county-level 2010 ACS 5-year estimates.	Family responsibilities and dependencies can lead to unexpected mobility under extreme weather conditions. Someone may try to cross flooded locations in the effort to reach and help the rest of the household members during flash floods [Fekete, 2010; Ruin et al., 2014]. Single parents may have more pressure for care giving that along with parents' tendency to ignore their self-protection to protect their children can lead them to enter flashy waters [Tapsell et al., 2002].
Number of family households (i.e., families) (count and % to the total number of households)		
Number of single-parent families (i.e., with either male or female householder) (count and % to the total number of households)		
<i>Indicator 14: Educational attainment</i>		
People educated with less than 9th grade (count and % to the total population 25 years and over)	Extracted from table DP02 of the county-level 2010 ACS 5-year estimates.	Lower education may reduce the ability to understand warnings [Cutter et al., 2003; Fekete, 2010]. People with less than a high school diploma are the least likely (about 17.5% in 2004) to work in occupations in which they are flexible to vary their work schedules [McMenamin, 2007] and thus, may feel the need to drive through potentially flooded ways.
People graduated from high school or equivalent (count and % to the total population 25 years and over)		
<i>Indicator 15: Ethnicity/Citizenship</i>		
Number of foreign born, not U.S. citizen commuters by private vehicle (drove alone or carpooled; in the workplace county)	Estimated by grouping the drove alone and carpooled classes from table B08511 of the county-level 2010 ACS 5-year estimates for workplace geography.	Probable cultural or language constraints of foreign commuters may hinder situational awareness related to the forthcoming weather and driving conditions [Ruin et al., 2007; Maples and Tiefenbacher, 2009].

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 16: Language</i>		
Number of people who speak other than English languages at home, and speak English less than "very well" (for population over 5 years) (count)	Extracted from table DP02 of the county-level 2010 ACS 5-year estimates.	Language difficulties can lead to limited or no reception of warnings and emergency advice [Fekete, 2010; Wilhelmi and Morss, 2013].
Number of commuters by private vehicle who speak other than English languages at home, and speak English less than "very well" (count) (for workers 16 years and over in the workplace county)	Estimated by grouping drove alone and carpooled classes for workers who speak other than English languages at home and speaking English less than "very well" from table B08513 of the county-level 2010 ACS 5-year estimates.	
<i>Indicator 17: Vehicles</i>		
Aggregate number of vehicles available in the total households (count)	Extracted from table B25046 of the county-level 2010 ACS 5-year estimates.	The amount of vehicles used in daily commuting or being available to be used to reach a destination, or retrieve family members (and/or property) during flooding can be related to the likelihood of people to get trapped in a car-related incident. The use of private four-wheel vehicles in driving through flooded ways is mainly attributed to the drivers' confidence in automobile safety or personal driving capabilities, and underestimation of risk [Diakakis and Deligiannakis, 2013; Franklin et al., 2014; Gissing et al., 2016].
Aggregate number of vehicles used in commuting by workers (count)	Extracted from table B08015 of the county-level 2010 ACS 5-year estimates.	

Table II.1 Continued

Variables	Processing	Risk hypothesis
<i>Indicator 18: Travel time to work</i>		
<p>Number of commuters who are travelling to go to work from 5 to 90 or more minutes estimated in 11 classes (e.g., 5 to 9 min, 10 to 14 min, . . . , 60 to 89 min, more than 90 min) (count)</p>	<p>Extracted from table B08303 of the county-level 2010 ACS 5-year estimates.</p>	<p>Longer journeys suggest higher likelihood of exposure to flooded roads. Also, commuters who are familiar with long everyday travels on certain roads may be more likely to underestimate the level of risk associated with voluntary entering floodwater [Ruin et al., 2007; Maples and Tiefenbacher, 2009].</p>
<i>Indicator 19: Commuters</i>		
<p>Number of commuters that arrive at work by vehicle in a time interval that covers the time of the unit peak discharge associated with a certain flash flood event in the exposed county (for workers 16 years and over who do not work at home in the workplace county) (count)</p>	<p>Estimated by assigning the number of workers arriving during a given time interval at work to each flash flood event for which the CREST simulated unit peak discharge has been recorded in the same time interval. The drove alone and carpooled classes of commuters were grouped for each given time interval from table B08532 of the 2010 ACS 5-year estimates</p>	<p>The conjunction of daily mobility related to professional activity of people with the occurrence of unusual hydrometeorological circumstances increases the vehicle-related spatio-temporal exposure [Ruin, 2007].</p>

Prévision des impacts humains conséquences des crues rapides intégrant le concept de vulnérabilité sociale dynamique.

À u XXI^e siècle, la prévision de l'aléa hydrométéorologique et des impacts associés aux crues rapides demeurent un défi pour les prévisionnistes et les services de secours. Les mesures structurelles et / ou les avancées des systèmes de prévision hydrologique ne garantissent pas, à elles seules, la réduction des décès lors de ces phénomènes d'inondation rapide. La littérature souligne la nécessité d'intégrer d'autres facteurs, liés aux processus de vulnérabilité sociaux et comportementaux, afin de mieux prendre en compte les risques encourus par les populations lors de ces épisodes extrêmes. Cette dissertation conduit une analyse théorique couplée à ceux de une analyse des accidents historiques mortels afin d'expliquer les interactions qui existent entre les processus hydrométéorologiques et sociaux responsables de l'apparition de vulnérabilités humaines lors de crues rapides aux États-Unis. Des données d'enquêtes liées aux crues rapides sont examinées afin d'élaborer un système de classification des circonstances du décès (en voiture, à l'extérieur, à proximité d'un cours d'eau, dans un camping, dans un bâtiment ou en mobile-home). L'objectif est d'établir un lien entre la conception des vulnérabilités et l'estimation des pertes humaines liées à ces catastrophes naturelles. "Random forest" est utilisé et est basé sur un arbre de décision, qui permet d'évaluer la probabilité d'occurrence de décès pour une circonstance donnée en fonction d'indicateurs spatio-temporels. Un système de prévision des décès liés à l'usage de la voiture lors des crues rapides, circonstance la plus répandue, est donc proposé en s'appuyant sur les indicateurs initialement identifiés lors de l'étude théorique. Les résultats confirment que la vulnérabilité humaine et le risque associé varient de façon dynamique et infra journalière, et en fonction de la résonance spatio-temporelle entre la dynamique sociale et la dynamique d'exposition aux dangers. Par exemple, on constate que les jeunes et les personnes d'âge moyen sont plus susceptibles de se retrouver pris au piège des crues rapides particulièrement soudaines (par exemple, une durée de près de 5 heures) pendant les horaires de travail ou de loisirs en extérieur. Les personnes âgées sont quant à elles plus susceptibles de périr à l'intérieur des bâtiments, lors d'inondations plus longues, et surtout pendant la nuit lorsque les opérations de sauvetage et / ou d'évacuation sont rendues difficiles. Ces résultats mettent en évidence l'importance d'examiner la situation d'exposition aux risques en tenant compte de la vulnérabilité dynamique, plutôt que de se concentrer sur les conceptualisations génériques et statiques. Ce concept de vulnérabilité dynamique est l'objectif de modélisation développée dans cette thèse pour des vulnérabilités liés aux véhicules. À partir de l'étude de cas sur les crues rapides survenues en mai 2015, et en analysant principalement les états du Texas et de l'Oklahoma, principaux états infectés par ces événements, le modèle montre des résultats prometteurs en termes d'identification spatio-temporelle des circonstances dangereuses. Cependant, des seuils critiques pour la prédiction des incidents liés aux véhicules doivent être étudiés plus en profondeur en intégrant des sensibilités locales non encore résolues par le modèle. Le modèle établi peut être appliqué, à une résolution journalière ou horaire, pour chaque comté du continent américain. Nous envisageons cette approche comme une première étape afin de fournir un système de prévision des crues rapides et des risques associés sur le continent américain. Il est important que la communauté scientifique spécialisée dans l'étude des crues éclairs récoltent des données à plus haute résolution lorsque ces épisodes entraînent des risques mortels, et ce afin d'appuyer la modélisation des complexités temporelles et spatiales associées aux pertes humaines causées par les futures inondations soudaines.

Mots clés : Crue rapide, impacts humains, facteurs de vulnérabilité, prédiction par apprentissage virtuel, cartographie dynamique des risques.

Forecasting of flash-flood human impacts integrating the social vulnerability dynamics.

I n the 21st century the prediction of and subsequent response to impacts due to sudden onset and localized flash flooding events remain a challenge for forecasters and emergency managers. Structural measures and/or advances in hydrological forecasting systems alone do not guarantee reduction of fatalities during short-fuse flood events. The literature highlights the need for the integration of additional factors related to social and behavioral vulnerability processes to better capture risk of people during flash floods. This dissertation conducts a theoretical analysis as well as an analysis of flash flood-specific historic fatalities to explain complex and dynamic interactions between hydrometeorological, spatial and social processes responsible for the occurrence of human life-threatening situations during the "event" phase of flash floods in the United States (U.S.). Individual-by-individual fatality records are examined in order to develop a classification system of circumstances (i.e., vehicle-related, outside/close to streams, campsite, permanent buildings, and mobile homes). The ultimate goal is to link human vulnerability conceptualizations with realistic forecasts of prominent human losses from flash flood hazards. Random forest, a well-known decision-tree based ensemble machine learning algorithm for classification is adopted to assess the likelihood of fatality occurrence for a given circumstance as a function of representative indicators at the county-level and daily or hourly time steps. Starting from the most prevalent circumstance of fatalities raised from both the literature review and the impact-based analysis, flash flood events with lethal vehicle-related accidents are the subject to predict. The findings confirm that human vulnerability and the subsequent risk to flash flooding, vary dynamically depending on the space-time resonance between that social and hazard dynamics. For example, it is found that younger and middle-aged people are more probable to get trapped from very fast flash floods (e.g., duration close to 5 hours) while participating in daytime outdoor activities (e.g., vehicle-related, recreational). In contrary, older people are more likely to perish from longer flooding inside buildings, and especially in twilight and darkness hours when rescue and/or evacuation operations are hindered. This reasoning places the importance of situational examination of dynamic vulnerability over generic and static conceptualizations, and guides the development of flash flood-specific modeling of vehicle-related human risk in this thesis. Based on the case study of May 2015 flash floods with a focus in Texas and Oklahoma, the model shows promising results in terms of identifying dangerous circumstances in space and time. Though, critical thresholds for the prediction of vehicle-related incidents need to be further investigated integrating local sensitivities, not yet captured by the model. The developed model can be applied on a daily or hourly basis for every U.S. county. We vision this approach as a first effort to provide a prediction system to support emergency preparedness and response to flash flood disasters over the conterminous U.S. It is recommended that the flash flood disaster science community and practitioners conduct data collection with more details for the life-threatening scene, and at finer resolutions to support modeling of local temporal and spatial complexities associated with human losses from flash flooding in the future.

Keywords : Flash flood, human impacts, vulnerability factors, machine-learning predictions, dynamic risk mapping.