Response reviewer 1

Authors: We thank the reviewer for her/his time to read the paper and provide comments. We found the comments and suggestions helpful, and we have revised the paper accordingly. We think that the revised paper is a significant improved.

Reviewer 1: Terminology: as general assumption of our community (e. g. Merz et al., 2010, de Moel et al., 200), and following what authors write at P2 L43-46, flood impact is one of the three components of flood risk. Therefore, I would avoid to use the statement "flood risk and impact assessment", because flood impact is somehow included in flood risk. Please, go through the manuscript (include title) and correct these cases.

Authors: Here we intended to make the distinction between flood risk (predictive and probabilistic analysis conducted before an event) from descriptive flood impact assessment (observational and deterministic assessment conducted after an event). We have therefore added clarifying text in the introduction as follows: "We make the distinction between flood risk, as the probabilistic analysis of the *potential* (predictive) impacts of floods and flood impact assessment, as the post-event assessment of (descriptive) impact from an *actual* flood event."

Reviewer 1: Introduction: I would add some more detailed explanations of the different machine
learning methods. Just some sentences, but it can help in order to have clear in mind, in the rest of the manuscript, what decision trees, neural network, etc. are. In addition, as general comments, I would add some sentences (and references) which state that machine learning methods can really improve estimations, in case of large datasets available: up to now, this concept is taken for granted, but citing some studies that demonstrate it could improve the manuscript, in my opinion.

Authors: Thank you for this comment. We intended to highlight applicability and relevance of ML rather than details of specific methods. However, we have added some additional description of the methods and references, to give the reader a bit more context to the ML methods mentioned, and reference to learn more.

Reviewer 1: Structure: in order to be consistent with the definition of flood risk at P2 L43-49, I would suggest to analyse hazard (current Ch. 3) before exposure (current Ch. 2).

Authors: We agree with the reviewer that the order of the chapters should correspond to the order in which the different components are introduced. We therefore changed the order in which we introduce the different components in the introduction.

Reviewer 1: Ch. 5: I would add, as an issue, the difficulty to use machine learning methods: they require a quite high degree of knowledges in order to really appreciate improvements in flood risk estimation and to avoid errors, that can easily be done by not-experts.

Authors: We would like to thank the reviewer for this suggestion. We think the hype around machine learning can cause misuse and unwarranted trust by users in the models. We added an entire new paragraph to address this issue in chapter 5.

Reviewer 1: P3 L89-90: Why traditional process models were not displaced by machine learning methods? Please explain, or refer to specific following Sections.

Authors: That's an interesting question. This is implicitly already answered throughout the paper when we discuss future changes and obstacles for these changes to occur. However, we also added a sentence about this directly on page 3 so the reader doesn't have to wait for an answer.

Reviewer 1: P11 L352-353: I would suggest to remove the sentence.

Authors: We considered removing this sentence but since this is the only appropriate place to reference to table 2 (see comment below), we decided to keep the sentence.

Reviewer 1: P12 L383-384: please add reference for this statement

Authors: This sentence is a prediction/perspective that is then motivated in the rest of the paragraph. Within this motivation we added a few additional references now that would make this statement more convincing.

Reviewer 1: Table 2 is never cited in the text. Please correct

Authors: We added a reference to this table

Reviewer 1: As suggestion, I list some recent papers I found, which use machine learning methods in the flood risk assessment, and can be useful to cite in order to strengthen some concepts:

Authors: Thank you for these suggestions. These papers can help as references throughout the paper and have been included as additional motivation for existing text.

Reviewer 2: Lack of adequate referencing

Key references of this work are the following: a. Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2 b. The two works: Solomatine, D.P., Ostfield, A. Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics, 10 (1), 3-22; Dibike, Y.B., Solomatine, D.P., 2001. River flow forecasting using artificial neural networks. Physics and Chemistry of the Earth Part B: Hydrology, Oceans and Atmosphere. Volume 26, Issue 1, Pages 1-7. c. The ebook "GFDRR. 2018. Machine Learning for Disaster Risk Management. Washington, DC: GFDRR.

May be others are also included, but I do believe a much wider bibliography is needed to support this work. In fact, the first two are surely crucial supporting references, but are not adequately representing and supporting the submitted research/perspective, considering that after 2001 and 2006 tremendous advancements and scientific production was developed not only for ML in general, but within the topic of ML for floods specifically (see quick search on SCOPUS below in Figure 1 and Figure 2).

Authors: Thank you for this feedback. We have added 21 additional references to further support the paper. As such and while still not exhaustive, the list of 96 references included in our manuscript cover key contributions relevant to understand current state of the art in applications of machine learning methods for flood risk and impact analysis.

Reviewer 2: Summary and review of state of the art ML

I'd suggest authors to better introduce ad categorize major concepts, procedures and tools of ML. In the introduction the reference to Bishop's book is then followed by few specific examples (see also specific comments). I'd see here a flow chart or summary table to improve the manuscript while addressing this general remark. I'm sure authors can benefit and extrapolate the work already done and cited within the GFDRR book on the topic. Additional scientific referencing can further strengthen this important part of the manuscript where the reader will be guided with key concise definitions and adequate referencing on state of the art ML for earth/geo/water science and flood risk management in particular. Please see this general comment as a further extension of general comment n.1, a surely expected contribution by authors to better support and introduce the final findings/perspectives of this work.

Authors: We described the state of the art of machine learning for flood risk assessments with numerous references within specific sections on exposure, hazard and impact. We choose this structure so we can, for each topic, transition easily from what is already being done to what could be done in the future, and directly discuss the difference between science and application. This gap between science and application causes a lot of grey area in the state-of-the-art description and makes it therefore necessary to do this in the main body of the paper. In a traditional research paper this information would indeed all be expected in the introduction, but this paper is a perspective on the latest scientific advances and how they can be extrapolated and be applied. The introduction therefore remains on a somewhat higher level of abstraction. We have further added a description of the paper structure up front in the in the introduction.

As for the state of the art on machine learning techniques themselves, we deliberately focused on applications of these methods (i.e. what can you do with it) rather than theory (i.e. how do they work). For each method presented, we included references to papers and books that describe the specifics of the algorithms and underlying theory. Our target audience are flood risk management experts interested to understand the potential applications of these methods. We believe that covering theory of the ML methods would be beyond the scope of the paper, and detract readers from the main messages communicated. We did expand the description of the advances in machine learning to better highlight what is possible with the latest innovations that wasn't possible before (see also response to reviewer 1).

Reviewer 2: Structure of the manuscript

Sections 2, 3 and 4 should be merged into a Section 2 with subsections. I see the three components "exposure, hazard and impact" as a unique core section with subsections related to descriptive versus predictive assessment models and related comments.

Authors: We merged sections 2, 3 and 4 into one section.

Reviewer 2: Scientific soundness of the "Perspectives" section

Section 5 seems to be a bit general. As requested to Invited perspectives my opinion is that authors miss to explicitly include in the paper sound arguments, facts, published research studies to support the conclusive remarks. Those remarks remain, in fact, general and simplicistic relying on few selected, yet relevant, references (mostly the GFDRR that is not even a research work).

Authors: Section 5 only contains general perspectives that apply to multiple topics and are seen throughout the paper or summaries of key points. Specific perspectives are covered in the exposure, hazard and impact sections and sometimes repeated here and classified by future likelihood. There is argumentation for the future likelihood. We have now added references to these earlier chapters in the perspectives chapter. In this way we now emphasize where the general comments are coming from and that they often arise from the specific chapters. We also changed the name of the chapter from "perspectives" to "general perspectives" so that we take away the expectation that only this chapter has perspectives.

Reviewer 2: I understand this paper, as written in the Acknowledgement section, is the result of a "2 week long intensive collaboration during the Understanding Risk Field lab on urban flooding in Chiang Mai, but "Out of the context" and the intent is to share these ourcomes with the scientific audience, but the submitted manuscript seems not to capture the surely significant value of authors' knowledge and experience as well as the value of the Understanding Risk Field lab
workshop discussion and reasoning.

Authors: We have re-written the acknowledgement section to better reflect the role of the event as a convening and facilitating platforms for multi-disciplinary experts, from which emerged much of this manuscript.

Reviewer 2: Line 16-20: I think the second sentence is redundant, merge into one sentence.

Authors: We have modified this sentence to clarify that one refers to the past and the other to the future.

Reviewer 2: Line 27: "avoid repetitions just here rephrase" and line 29: "same comment as before"

Authors: We removed some of the repetition.

Reviewer 2: Line 33: I'd expect a clear definition of machine learning as well a solid reference to this statement

Authors: We repeated the reference from line 35 here now because that also covers this statement. A more solid definition of machine learning is provided in line 69-70 .

Reviewer 2: Line 39: Again, as remarked before, I see redundancies. You are repeating that advantages vs limitations shall be considered. No references added here also makes the introuction weak. See general comment n.1

Authors: The second sentence was removed and the first has been integrated with the one prior.

Reviewer 2: Line 42: hazard, exposure and impact are key concepts of flood risk. Why mentioning here floods and society. And, again, why no references supporting key statements of the introduction?

Authors: Exposure and impact are only relevant for floods when the relationship between floods and society is studied. The "interaction between floods and society" is a less technical term then just "flood risk". We added a reference to this key concept now.

Reviewer 2: Line 51: I don't find a close link between this introductory sentence of ...application of floof risk and impact assessment, and the disaster managementy cycle

Authors: The disaster management cycle is a paradigm that includes everything that is done to fight the negative consequences of floods. Every flood risk and impact assessment should therefore have a benefit somewhere in this paradigm. The second line also refers directly to the first sentence already making the link clear, we strengthened that reference even more by repeating the word "different" before applications.

Reviewer 2: Line 57-58: you are here referring to recostruction of real events. I'd make this clear and cite relevant papers

Authors: We are not referring to reconstructing hazard events. For a reconstruction you need to know what happened. We are talking about having a descriptive hazard (e.g. remote sensing based observation) and then developing a model to predict the impact. We rewrote this sentence to clarify it.

Reviewer 2: Line 63: there are two references by Wagenaar et al 2019 (make sure the manuscript not accepted is removed from the list)

Authors: The first is already published. The second is under final review and will be published shortly. Since it is very relevant to the accompanying statement, we feel that it should be included as a reference, and should be made available to readers as relevant literature. Our understanding is that we should reference relevant literature, even if "unpublished" or "grey" literature.

Reviewer 2: Line 69: improve the introduction to machine learning, from definition to key concepts, procedures, tools etc

Authors: The paragraph starts with a definition, then clarifies this definition with an example. Procedures and tools go a bit unnecessarily deep for this paper see the response to the general comment "Summary and review of state-of-the-art ML".

Reviewer 2: Line 73 a reference is needed here, otherwise it becomes quite unclear the reading

Authors: We have added a reference.

Reviewer 2: Line 80 -85: I think this should be absolutely expanded, supported by more references. I advise the use of a summary table or flow chart. This paper can's comply with the title without a deep description of state of the art ML. See General comment n.2

Authors: See the response to general comment 2. The paragraph is expanded to make a clarify what the new methods can do compared to the old methods. However, we want to stick to the choice to not get into the topic of how the techniques work as it isn't necessary to answer the question in the title.

Reviewer 2: Line 129: reference?

Authors: The first part of the sentence is a conclusion of our literature research that is shown just before this sentence where we found no more advanced applications. The second part of the sentence is a perspective. This perspective is now supported with some additional reasoning.

Reviewer 2: Line 136: references?

Authors: This is about a categorization we defined for this paper. See introduction so it won't be possible/necessary to reference to that.

Reviewer 2: Line 168: there are tons of papers on the use of crowdsouced data for flood hazard modelling and mapping.

Authors: We referenced here to 4 recent papers about this and added a sentence about something that isn't done yet but could potentially be done in the future.

Reviewer 2: Line 207: this is not included in the reference list. The reference should be this https://arxiv.org/abs/1901.09583 but I don't find it proper for supporting this statement

Authors: That is indeed the correct reference, and thank you for pointing out that it is missing from the reference list. In the referenced paper, the authors suggest that hybrid ML-physical models will contribute to flood forecasting. Specifically, ML models 'responsible for calibration, error corrections, and perhaps additional processes that were not well modelled'. In this respect, we feel that using the paper to support our own argument (see below) is justified. "Hybrid component models assign machine learning a specific task in the modelling process that is either highly complex or not well understood. Examples of this include using machine learning for error correctors (see, for example, studies by Abrahart et al. 2007 and Google Research - Nevo et al. 2019)"

Reviewer 2: Line 208-215: See General comments n.1 and n.2 I suggest inserting more references and also more clear descriptions of the ML modelling approaches.

Authors: We already reference to 5 papers in this paragraph and all key ideas are already supported by at least 1 reference. We added a newer reference on the real-time control of flood defences and systems. We don't intend to get into more detail about it, this is just one of many applications in this paper and this level of detail is appropriate for a high level understanding of the possibilities.

Reviewer 2: Line 220: all capital letters to define acronym

Authors: We changed this

Reviewer 2: Line 244: This may be generally true, but social networks are providing highe amount of geotagged images and videos. This is in contrast to what authors stated in the hazard assessment when citing crowdsource data.

Authors: This sentenced refers to labelled data that could be applied for training descriptive impact models (not for hazard assessment). Data from social networks is I expect only very rarely labelled with an object damage label that can be applied for training impact models. For descriptive hazard modelling the potential of social media data is already mentioned.

Reviewer 2: Line 248-253: This is quite superficial and not fully developed. What about drones, webcams and other sensors? Just to cite an example to trigger the discussion

Authors: We added a sentence that change detection might be applied to estimate damage from other angles then top view.

Reviewer 2: Line 281: check English

Authors: We will rewrite this sentence

Reviewer 2: Line 304: This entire section doesn not provide insightful comments and remarks on what's coming next. It is not, thus, respecting what was expected from the paper goal and title.

The lack of proper introduction, definition and review of ML methods is probably affecting the robusteness of this section that doesn't go deeper in the discussion of perspectives of ML for flood risk management. See General comment n.3

Authors: See the response to general comment 3. In short the perspectives are included all over the paper and this section only summarizes general trends and classifies the likelihood of specific predictions earlier in the paper using argumentation. We added references to earlier section to better to make this link between this chapter and the rest of the paper clearer.

Reviewer 2: Line 323: These 4 references are all missing from the reference list

Authors: Thanks for pointing this out. We added them now.

Reviewer 2: Line 359: too general. Socia media will help what? exposure, hazard, impact or what?? You can't here just mention flood risk management

Authors: We agree this is not well formulated. We rephrased this and referred to earlier chapters.

Invited perspectives:

How machine learning will change flood risk and impact assessment

Dennis Wagenaar[1,2], Alex Curran[1,3], Mariano Balbi[4], Alok Bhardwaj[5], Robert Soden[6,7,8] ,Emir Hartato[9], Gizem Mestav Sarica[10], Laddaporn Ruangpan[11], Giuseppe Molinario[7], David Lallemant[5,8]

1 Deltares, Delft, the Netherlands
2 VU University, Amsterdam, the Netherlands
3 Delft University of Technology, Delft, the Netherlands
4 University of Buenos Aires, Buenos Aires, Argentina
5 Earth Observatory of Singapore, Nanyang Technological University, Singapore
6 Columbia University, New York City, New York, United States of America
7 GFDRR, World Bank Group, Washington D.C., United States of America
8 Co-Risk Labs, Oakland, California, United States of America
9 Planet, San Francisco, United States of America
10 Institute of Catastrophe Risk Management, Nanyang Technological University, Singapore
11 IHE Delft Institute for water education, Delft, the Netherlands

Correspondence to: Dennis Wagenaar (dennis.wagenaar@deltares.nl)

Abstract. Increasing amounts of data, together with more computing power and better machine learning algorithms to analyse the data are causing changes in almost every aspect of our lives. This trend is expected to continue as more data keeps becomes becoming available, computing power keeps increases improving and machine learning algorithms keep improveimproving as well. Flood risk and impact assessments are also being influenced by this trend, particularly in areas such as the development of mitigation measures, emergency response preparation, and flood recovery planning. Machine learning methods have the potential to improve accuracy as well as reduce calculating time and model development cost. It is expected that in the future more applications become feasible and many process models and traditional observation methods will be replaced by machine learning. Examples of this include the use of machine learning on remote sensing data to estimate exposure or on social media data to improve flood response. Some improvements may require new data collection efforts, such as for the modelling of flood damages or defence failures. In other fieldscomponents, machine learning may not always be suitable or should be applied complementary to process models, for example in hydrodynamic applications. Overall, machine learning is likely to drastically improve future flood risk and impact assessments, but issues such as applicability, bias and ethics must be considered carefully to avoid misuse. This paper presents some of the current developments on the application of machine learning for flood risk and impact assessment,in this field and highlights some key needs and challenges.

1.Introduction

Exponentially increasing computing power and data, as well as rapidly improving machine learning algorithms to analyse this data have been changing many aspects of our lives (Manyika et al., 2011). . These trends are expected to continue and will undoubtedly keep affecting many scientific, commercial and social sectors (Manyika et al., 2011). Flood risk and impact assessments are no exception to this trend. Flooding yearly affects more people than any other natural hazard types (Jonkman, 2005) and the impact and frequency of flooding events is expected to increase in the future due to urban

1

development and climate change (Kundzewicz et al., 2014). It is therefore an opportunity for researchers and flood managers

40   to tap into the potential of machine learning, taking advantage of their strengths while being cognisant~~-~~ of their limitations. It is also important to anticipate improvements in the capabilities of machine learning methods, so as to plan for forthcoming changes in flood modelling. ~~At the same time, being aware and transparent about the limitations of machine learning methods in the field is of paramount importance.~~

45   When assessing the interaction between floods and society, three different components can be recognized: exposure, hazard, ~~exposure~~ and impact (Kron, 2002). Exposure refers to the characteristics of the people and assets that can be affected by flooding. Hazards are the physical characteristics of a flood such as the extent, water depth, duration and flow velocity. ~~Exposure refers to the characteristics of the people and assets that can be affected by flooding, and~~ ~~i~~Impacts are the effects the hazard has on the exposure. To assess these three components, we make the distinction between flood risk, as the

50   probabilistic analysis of the *potential (predictive) impacts* of floods and flood impact assessment, as the post-event assessment of (descriptive) impact from an *actual* flood event." ~~All three components can be assessed both in descriptive and predictive manners. A predictive assessment is a long term or short term forecast, while a descriptive assessment is an analysis of the current or past state. Therefore, flood risk assessment can be considered a particular type of predictive impact assessment in which the probability of an event is also accounted for.~~ Table 1 provides examples of predictive and

55   descriptive assessments in relation to the hazard, exposure and impact components.

Flood risk and impact assessments have many different applications. A useful paradigm to look at these different applications is the 'disaster management cycle' (Khan et al., 2008; National Research Council, 2006) (Fig 1). This cycle delineates the phases between events, i.e. the immediate response to an event, the long-term recovery, the mitigation to

60   prevent future events and the preparation prior to a new forecasted event.

In the response phase, the focus is typically on descriptive hazard, exposure and impact assessments (e.g. Klemas, 2015), sometimes complemented with predictive models if the event descriptive information isn't available yet (e.g. a predictive model estimating the number of people affected can be ~~based~~ fed by~~on~~ a descriptive hazard model of the flood extent). The

65   challenge in this phase is mostly data reliability.   In the recovery phase, descriptive assessments are often used for payouts (e.g. indemnity insurance), and one of the main challenges is ensuring these payouts are timely and reliable. In the mitigation phase, probabilistic predictive models are used (e.g. Wagenaar et al., 2019), typically for the design of risk-reduction interventions ranging from protective infrastructure to insurance products. The challenge in this phase is model reliability and uncertainties about future developments (e.g. uncertainty in future exposure). In the preparation phase, predictive models

70   are used for emergency planning (e.g. Coughlan de Perez et al., 2016), where the challenge is the reliability, availability and communication of data. Machine learning is capable of generating more reliable and faster models that can help solve some of the current challenges in the disaster management cycle but could also provide new opportunities.

Machine learning algorithms can find patterns in data and use these patterns to make predictions about new data (Bishop, 2006). For example, when providing a machine learning algorithm with aerial images of either urban or rural areas and corresponding labels (urban or rural) it can build the capacity to classify new unlabelled aerial images as either urban or rural. Features in the above example would be different components of the aerial images (i.e. pixel tone and locations) and the target variable would be the label (i.e. urban or rural). When a precise value is required as opposed to a label, it is called a 'regression task'(e.g. Bishop, 2006). An example of this is in flood damage modelling, where features such as water depth, flow velocity and building materials can be used to predict a target variable such as monetary economic damage based on historical records (e.g. Merz et al., 2013; Wagenaar et al., 2017). Due to the use of labelled training data (e.g. classified images or historic damage examples), regression and classification are called supervised learning tasks. Machine learning method categories also include unsupervised learning and reinforcement learning, (see GFDRR, 2008). However, such methods are not discussed in this paper because they are expected to have a smaller short-term impact on the field of flood risk and impact- assessments.

The simplest machine learning algorithms have been used for a long time and are often known as basic statistical techniques (e.g. linear regression: Legendre, 1805; Gauss, 1809). More sophisticated machine learning techniques that emerged in the 1980s and 1990s (e.g. Decision Trees and Neural Networks) can find more complex non-linear patterns (Breimann et al., 1984; Rumelhart et al., 1986). Recent advances in machine learning (e.g. convolutional neural networks) make computer vision and other advanced applications possible (Krizhevsky et al., 2012). The more advanced techniques such as decision trees, neural networks and especially convolutional neural networks can find more complex patterns. This is because they allow for more complex non-linear functions to be fitted to the data. Such complex functions require a large number of model coefficients to be set during the training of the model. To set all these coefficients a lot of training examples are required. In some cases the number of training examples can be reduced with transfer learning techniques (Olivas et al., 2010). These techniques make it possible to re-use knowledge gained from other problems to train a model on a smaller training data set.

From the beginning, machine learning has been used in predictive flood hazard modelling (Solomatine & Ostfield, 2008) mostly as a faster and simpler alternative to process models. A simple example of this is the prediction of river discharge based on upstream rainfall data (e.g. Dibike & Solomatine, 2001). This type of modelling has been practiced for a long time but hasn't displaced the traditional process models. This is probably because the methods aren't sufficiently better than traditional methods to offset some disadvantages as discussed in the predictive hazard section. In recent years, more data is becoming available through remote sensing and other sources. This impulse of new data combined with machine algorithms

could lead to changes in flood risk and impact -assessment. Some of these changes have already been highlighted by major international organizations such as the World Bank and others (GFDRR, 2018).

This invited perspective paper ~~provides~~ starts with a perspective per risk assessment component as defined in table 1. These specific perspectives start with a description of the traditional approach for the assessments. Followed by a ~~review~~literature review on how machine learning techniques are currently being developed to improve the traditional approach and then proceed to speculate on potential future improvements. This is followed by a general perspectives chapter in which general trends that come back in the different components are identified and discussed. This includes common challenges (i.e. data limitations, transferability, ethics and bias) and ends with some speculation about the likelihood of future developments. ~~and analysis of the implications that machine learning methods are having, and will have in the future, for flood risk and impact assessments. We will go through each of the components of flood risk and impact assessment as outlined in table 1 and provide an analysis of how current assessments are done without machine learning, current applications of machine learning and potential future improvements. Then we will discuss ethics and bias questions related to these developments and draw some general conclusions.~~

2. Perspective per component

2.1 Exposure assessment

2.1.1 Descriptive exposure assessments

Descriptive exposure assessments consist of detecting and characterizing spatial features such as current buildings and infrastructure, agriculture fields, roads and other infrastructure. Traditionally this has been done by population censuses, building counts and conventional mapping techniques that require ground surveys. Remote sensing is currently changing this. The manual digitization of building footprints using aerial or satellite images has become increasingly common. This digitization of building footprints, but also notably of other features such as roads, has been done manually by "crowds" of mappers in "mapathons", for example using the OpenStreetMap platform. Machine learning is very likely going to drastically change this. Research to automatically labelling remote sensing data has already been going on for some (e.g. Heermann & Khazenie, 1992; Giacinto & Roli, 2001). It is currently already being used to label build-up areas based on nighttime lights (Goldblatt et al., 2018) or satellite images (Goldblatt et al., 2016). Furthermore, algorithms are already being used to automatically label buildings (Sermanet et al., 2013; Alshehhi et al., 2017; GFDRR, 2018) and map roads (Gao et al., 2019) using aerial / satellite imagery. This will reduce the need for manual detection and will probably provide global availability of such building footprints and road information in the near future.

4

Part of an exposure assessment is the observation of asset features relevant for risk analysis. For example, building materials, building occupancy (e.g. residential or industrial), building height, ground floor elevation, poverty rates in the population, etc. This information is typically not available, but could be very valuable as input for impact models (e.g. Merz et al., 2014; Wagenaar et al., 2017; Schröter et al., 2014) or, for example, to account for poverty in cost-benefit analyses (e.g. Kind et al., 2016). Similarly, ground floor elevation information could radically improve urban pluvial flood damage modelling as damage from small-scale floods is very sensitive to such variables.

Some work has already been carried out on detecting poverty (Watmough et al., 2019) and building heights (Saadi & Bensaibi, 2014) by satellite imagery. Another source of this building feature information could be 360-degree street view images combined with computer vision techniques. Such images are available in, for example, the open source streetview data platform Mapillary (Neuhold et al., 2017). Such techniques are already starting to impact earthquake risk assessments, such as in Guatemala, where 360 degree imagery was fed into Mapillary algorithms in order to automatically detect "soft story" buildings; those most likely to collapse in an earthquake. This was done by having the machine learning algorithm detect features that were indicators of large openings on the ground floor of buildings (large doors, garage doors, shop windows, etc.) (GFDRR, 2018). Computer vision techniques from street level imagery are currently limited to detecting such relatively simple features. However, but based on recent advances seen in other computer vision applications (e.g. facial recognition), it may is likely that in the future it will be possible be able to detect more complex building features alsoin the future. For computer vision models to detect complex information like ground floor elevation or building materials, it would be necessary to provide labelled examples to the algorithms. Such labelled examples are sometimes already available for some areas, e.g. ground floor elevation (Bouwer et al., 2017) or building materials (Schröter et al., 2018).

2.2 1.2 Predictive exposure assessments

Predictive exposure mapping consists of estimates of future exposure. This mostly includes modelling to predict urban growth and other changes in land-use. It is required for evaluating flood mitigation measures (e.g. Wagenaar et al., 2019) because such measures typically need to function for a long time and should therefore still perform as required after predicted land-use changes. Land-use changes affect the impact of a flood because more damage may occur for the same flood hazard and the flood hazard may become greater because of changes in impervious area and therefore rainfall-runoff (Triantakonstantis and Mountrakis, 2013; Mestav Sarica et al., 2019). Predictive exposure assessments for flood risk and impact assessments are currently often not carried out spatially, but rather GDP growth projections are applied to estimate future total exposure values (e.g. van der Most et al., 2014; Wagenaar et al., 2019). This is enough for some studies but if large land-use changes are expected a land-use change or urban growth model is required.

170 Urban growth has been modelled with simple machine learning models in the past (e.g. logistic regression) (Samardzic-Petrovic, 2017). The use of Cellular Automata (CA) models has become more common recently (Naghibi et al., 2016). These models assign cells as either urban or non-urban based on specific transition rules. Determining the optimum transition rules is a critical issue for CA modelling (Aarthi and Gnanappazham, 2019). This is sometimes difficult because of human bias, heterogeneity and nonlinear relations between driving factors and urban expansion (Naghibi et al., 2016; Xu et

175 al., 2019). To overcome these limitations, machine learning algorithms such as artificial neural networks have been integrated with traditional CA to model urban growth (Aarthi and Gnanappazham, 2019; Naghibi et al., 2016). They then use historical land-use changes (e.g. Song et al., 2015) to learn the transition rules. Complex machine learning models have also been directly applied to urban growth modelling without the CA model structure (Pal and Ghosh, 2017). These improvements, together with more data about past land-use changes and additional computation power, are expected to

180 provide better future land-use maps and make high-resolution future land-use maps globally available.

~~3~~2.2.~~.~~ Hazards assessment

~~3~~2.~~1~~2.1 Descriptive hazard assessments

Descriptive flood hazard assessment focuses primarily on the response phase, i.e. in estimating current inundation extents

185 and depths to assist both emergency responders and those affected directly. This is traditionally achieved using optical remote sensing data, local sensors or manually collected data from observers on the ground. However, the rise of two major data sources, Synthetic Aperture Radar (SAR) and social media, provides a number of opportunities for machine learning to improve upon current flood detection methods.

190 During a flood event, affected populations frequently produce 'user-generated content' or 'crowd-sourced' data from social media posts or apps where citizens can report floods (UrbanRiskLab, 2019). This is especially the case in urban areas where internet and social media penetration are higher compared to rural areas. This data is often 'tagged' temporally and spatially and can be used by machine learning algorithms for applications such as nowcasting by searching for certain keywords like "flood" (e.g. see Tkachenko et al. 2017, Bischke et al. 2017, Lopez-Fuentes et al. 2017). The method is currently used to

195 map real-time flood extents in several countries (Eilander et al., 2016). Potential future machine learning and computer vision techniques could be extended to estimate water depths and other flood characteristics from posted photos.

Remotely-sensed optical data is often used to identify the extents of flooding, but optical sensors are not functional during periods of cloud-cover or at night. Furthermore, the temporal resolution often prevents the observation of flash floods. SAR

200 data using the microwave wavelengths of the electro-magnetic spectrum can help overcome these problems by providing additional imagery during the night or during cloud cover. Adding night-time and cloud-cover images will provide a higher

total temporal resolution. Flood extents are currently determined with statistical methods using thresholds to subsequently identify flood extents e.g. by using Bayesian method on SAR amplitude time-series data (Lin et al. 2019). Advanced machine learning classification methods are being developed to improve this process, but in order to train them it is necessary to have manually labelled images as training data. Collection of this labelled flood extent information is the main challenge for automatic detection moving forward. Manual methods could harness the power of the crowd, as people are connected through the internet or with mapathons. These approaches could have game-changing implications for the training of machine learning algorithms. Already mapathons are often 'trainathons', where mappers are not only manual digitizers, but also labellers and trainers for automated machine learning methods for the future.

3.2~~2.2.2~~ Predictive hazards assessments

Predictive flood hazard assessments consist of predicting future floods and their characteristics such as extents, inundation depths, durations, flow velocities, waves and water levels in rivers or seas. These assessments are applied for short-term forecasting in the preparation phase (preparing for imminent events) and long-term risk analyses for use in flood risk management (mitigation phase).

In flood-forecasting, traditional methods of predicting hazard variables can involve a chain of hydrologic and hydraulic models that describe the physical processes. Although such models provide system understanding, they often have high computational and data requirements. Therefore, the use of process models may not always be feasible or necessary in the preparation stage of a disaster. At that moment, accurate and timely outputs become more important than system understanding, and the use of 'black-box' machine learning models (e.g. Campolo et al., 2009) is becoming more widespread (Mosavi et al., 2018). The increased speed can create a trade-off with the robustness of forecast models, as changes to the hydraulic system (such as a new structure that could be easily implemented into a hydraulic model) cannot be directly introduced into a trained machine learning model. In addition, machine learning models might not perform well in predicting extremes far outside past observations, since it has not been trained against such extremes.

A review of flood forecasting methods using machine learning by Mosavi et al. (2018) highlights trends such as component and ensemble models (collectively termed 'hybrid models', Corzo & Solomatine, 2014). Hybrid component models assign machine learning a specific task in the modelling process that is either highly complex or not well understood. Examples of this include using machine learning for error correctors (see, for example, studies by Abrahart et al. 2007 and Google Research - Nevo et al. 2019) or flows subject to human influence (Yaseen et al., 2019). Hybrid ensemble methods often use machine learning models to supplement process models, providing robust predictions and uncertainty ranges (Solomatine & Ostfeld, 2007). Such methods benefit from the speed and ability to deal with non-linear multi-variable problems of machine learning modelling and the process understanding available in conventional modelling. The review by Mosavi et al. (2018)

does not consider gridded / spatial forecasting techniques, but advanced machine learning techniques are starting to be developed for precipitation pattern nowcasting (Shi et al. 2015) and flood extents prediction (Chang et al. 2018). Another application of machine learning in the preparation phase is in the real-time control of flood defences and systems (e.g. Lobbrecht & Solomatine, 2002; Castelletti et al., 2010). For example, Lobbrecht & Solomatine (2002) used machine learning methods to optimise control decisions in the event of communication network breakdowns during extreme storm events.

Another major application for machine learning in long-term risk analysis is 'surrogate' modelling (Ong et al. 2008), in which the outputs from process models are used to train computationally less-intensive machine learning models. This can be applied to speed up different types of process models applied in predictive hazard modelling. For example, in flood defence analysis and design, classical reliability techniques such as First order Order Rreliability Mmethods (FORM) and Monte-Carlo simulations (Steenbergen et al. 2004), or large-scale risk analyses that utilise them (Curran et al. 2019), can be replicated using a relatively small amount of evaluations as samples (Chojaczyk et al. 2015, Kingston et al. 2011). However, surrogate models may be particularly susceptible to extrapolation problems, where input data outside the range of the training data is introduced (Ghalkhani et al., 2013).

In the mitigation phase a chain of hydrologic and hydraulic models that describe the physical processes is typically applied (e.g. Wagenaar et al., 2019). In general, system understanding is required to assess proposed or potential future changes. In such cases, data-driven approaches are typically not applicable as there is no data about how the system behaves after the changes occur and hence simulation models are required that describe the physical system.

42.3. Flood impact assessment

4.12.3.1 Descriptive impact assessments

Descriptive impact assessments consist of making estimates of the flood impact after or during an event. This is traditionally done with manually collected data from observers on the ground. However, such manual ground inspections are slow and require people to enter the disaster area. Remote sensing can be used to get a very quick first impression of the damage to help with disaster response. Such techniques have already been applied, for earthquake and wind damage (e.g. Menderes et al., 2016). For flooding, this is often more difficult because damage inside buildings is difficult to obtain either from aerial-based or space-based sensors. Only when buildings completely collapse or are removed by strong flows does remote sensing become feasible. This is, for example, the case with flash floods, tsunamis or some storm surges. 360-degree streetview images collected after a flood could potentially be used for damage assessment. Machine learning techniques could then eventually be used to give a quick first estimate of the damage.

The use of machine learning techniques for automatic detection of damages from remote sensing information (aerial or streetview) requires labelled training data from manually collected data from observers on the ground. This data is currently rare. An approach could be to start using remote sensing data to manually label the impact. A way to get around this limitation is to detect changes in pre- and post-flood using high-resolution satellite images for urban areas where many buildings are damaged. Pixels with changed information will denote the damage that happened due to the floods. Eventually this data can then be used as training data for cases where only the post-flood images are available within a short time interval after the flood event. This method would however only be relevant for catastrophic floods because it doesn't address the fact that most damage remains not observable from top-view. On top of that this approach introduces significant new error: (1) error in the change detection signal, (2) error in relating the change to damage, (3) error in training a new model based on those damage labels. Imagery from different angles (e.g. from streetview or drones) might be more useful for change detection, however this data would also be more difficult to acquire.

4.22.3.2 Predictive impact assessments

Predictive flood impact assessments include models that translate hazard and exposure information into socio-economic impacts of the flood. This can include information such as   monetary flood damage, casualties, buildings damaged, crop damage, disease outbreak, building materials needed, recovery time, health monitoring of key structures or indirect damage (damages that occur in a different spatial and/or temporal setting than the originating event).

Most predictive flood damage modelling relies on depth-damage functions that describe a relationship between the water depth and monetary flood damage (Merz et al., 2010). They are either based on historical flood damage records (e.g. Thieken et al., 2008; Kreibich et al., 2010) or on expert estimates (e.g. Penning-Rowsell et al., 2005). In practice, many more variables than water depth have an influence on the flood damage (Cammerer et al., 2013, Wagenaar et al., 2016). Therefore, in the scientific literature there has been a move towards multi-variable flood damage models that use many variables (e.g. flood duration, velocity, building materials, socio-economic status of inhabitants etc.) instead of just water depth (e.g. Merz et al., 2013; Spekkers et al., 2014 Chinh, 2015; Kreibich et al., 2017; Wagenaar et al., 2017; Carisi et al., 2018; Amadio et al., 2019). These models are based on data and machine learning. The problem lies with insufficient data availability to train machine learning models and that using the models requires a lot of feature data about flood and building characteristics plus socio-economic data about inhabitants (Wagenaar et al., 2017). In the future we expect more data about features to become available from computer vision applied to street view, satellite or drone images (see descriptive exposure section). This would improve the quality of such models, could make it easier to apply them and make the development possible for more areas.

Machine learning could also be applied to predict disease outbreak after floods by combining remote sensing, meteorological, and socio-economic data (e.g. Mayfield et al., 2018; Carvajal et al., 2018; Modu et al., 2017; Yomwan et al., 2015; Tiwari et al., 2013; Shively et al., 2015). In a flood event, there is an increased risk of infectious diseases among survivors and displaced persons such as measles, diarrhea, acute respiratory infections and malaria can be responsible for many deaths (Lignon, 2006). Predictive modelling of such diseases is rarely carried out, and current approaches mostly focus on simple regression models or process models that simulate the spread of pollutants in the water. One major challenge is that the degree to which such epidemics occur, is associated with the regional endemicity of specific diseases, the nature and scope of the disaster, the level of public health infrastructure in place both before and after the event, and the level and efficacy of disaster response (Ivers & Ryan, 2006). Machine learning models could take such complex processes better into account.

Machine learning can be used for structural health monitoring, this has applications in the preparation phase (Pyayt et al. 2014, Jonkman et al. 2018) and in the long-term reliability analysis required in the mitigation phase (Prendergast et al. 2018, Klerk et al. 2019). Structural health monitoring is in the preparation phase is often done by manual inspections of the infrastructure on the ground, for example in the Netherlands there is a large network of volunteers that can be activated in case of high river levels to inspect the dikes. In the mitigation phase this is done by geotechnical process models fed by observations from the ground (e.g. De Waal, 2016), this is for example applied to decide on dike strengthening. Machine learning algorithms can help detect damage patterns from sensor data and are currently being used for the monitoring of flood defence structures such as dikes (Pyayt et al. 2011). Similar methods have also been applied to bridges (Neves et al. 2017). The use of both machine learning algorithms and traditional techniques for damage detection during floods is still very scarce (Prendergast et al., 2018, Pyayt et al., 2011); however, integration of structural health monitoring with flood early warning systems is a very promising field of development for machine learning techniques but would also requiring training data.

Indirect damages and business interruption are often taken into account simply through a scaling factor of the direct damage (e.g. Wagenaar et al., 2019). More complex models to quantify such damages include input-output models and general equilibrium models (e.g. Koks et al., 2016). To quantify indirect damages, such as business interruption losses, estimating the time it will take for different assets to be back in full or partial functionality is required. These post-disaster restoration models have started to be formalized in the last few years, primarily focused on earthquake disasters (Kang et al., 2018; Burton et al., 2018). Due to a lack of gathered empirical data on post-disaster recovery, the use of data-intensive machine learning techniques has not yet made an impact on this discipline. However, the need of probabilistically quantifying recovery will require the use of statistical models for calibration or assessments of recovery times, and that might be possible in the near future with the use of new remote sensing and crowd-sourcing technologies to obtain the empirical feature data needed.

53. General Perspectives

335 ### 35.1 Common challengesData limitations

Many machine learning applications in flood risk and impact modelling appear to be limited by a lack of data, especially training data needed to build effective machine learning models. This is especially true since the field of flood risk analysis is concerned primarily with extreme events, which are rare, and data-collection during such events is often logistically difficult. The increase in the amount of data around the world does not necessarily imply that this problem will be resolved

340 in the future. Some data is simply not collected or there are measurement definition or quality issues. To fulfil the potential of machine learning, new data collection efforts will be required, along with data standardization protocols. This will take collaboration between different organisations and stakeholders, setting of data standards and a willingness to share. This problem is common to ~~damage~~ impact data collected ~~for natural hazards~~(see 2.3.1 and 2.3.2), labelled flood extent data (see 2.2.1), social media hazard data (see 2.2.1) and first floor elevation data (see 2.1.1).

345 ### 35.1.2 Transferability of data

A critical assumption behind machine learning techniques is that the data being used to train a model is representative of the situation the model needs to be applied in. For example, a dataset on damage to concrete buildings is not fully applicable to modelling the damage to thatched huts. It is therefore important to collect heterogenous datasets that cover a large spectrum of potential situations (Wagenaar et al., 2018). Data that isn't fully applicable can still have some value, for example through

350 domain adaptation or transfer learning (GFDRR, 2018) but applicable data is always required as well. Wagenaar et al. (~~2019~~2020) showed that sample selection bias correction, a form of domain adaptation, helps to improve machine learning impact models in a transfer setting. Furthermore, it is important to work on efficient ways to communicate the applicability of data-driven models.

35.1.3 Ethics and Bias

355 Significant attention is currently being given to questions of the ethics and bias of machine-learning systems across a variety of domains, including facial recognition (Keyes, 2018), automated weaponry (Suchman et al., 2016), criminal justice (Eubanks, 2018) and search engines (Noble 2018). A number of technology companies and research institutions have developed guidelines for evaluating machine-learning systems, but this work is still evolving. Despite similar potential for negative impacts of these tools in flood risk management (Soden et al., 2019), the community has not given these issues as

360 much attention. Such concerns include the potential for reinforcing existing social inequalities and the reduced role of human judgement in modelling processes. These are risks that need to be weighed seriously against the potential benefits of machine learning and explored in greater detail.

Bias~~es in machine learning~~ ~~problems~~ can occur because of datasets that~~, for a number of reasons,~~ ~~are~~ do not fully
~~representative~~ represent the phenomena which they are meant to describe (e.g. people are accidentally excluded). ~~In addition~~For example, we ~~tend to~~often measure what we have data for, rather than measure what matters most, or ~~potentially use~~ training datasets that reinforce past problems. For example, if certain settlements aren't detected in exposure maps~~,~~ because they use different construction practices than the settlements used in training datasets, they may not receive emergency aid in the event of a flood. These problems can be mitigated by ensuring ~~should be addressed by ensuring~~ modelers understand the context of what they are attempting to model. Other ~~e~~Ethical issues ~~for~~ raised by machine learning in the flood management context ~~could~~ include data ownership, transparency, consent~~,~~ and privacy. For example, some people may object to having their home labelled "vulnerable" on a vulnerability map used by first responders. Privacy concerns may be aggravated by machine learning and other big data techniques. Ethics problems should be addressed by carefully weighing the benefits of collecting certain data against the related privacy costs, in collaboration with people who may be affected by the outcomes of decisions based on machine-learning tools.~~.~~ ~~The people that are being modelled should be included in this decision making.~~

An additional ethical concern regarding machine learning in flood risk assessment is misuse of models. In some sectors great advances have been made with machine learning (e.g. facial recognition, self-driving cars). This success for some tasks, can lead to an awe-inspiring general attitude towards the techniques (Ames, 2018; Narayanan, 2019). This hype sometimes leads to unwarranted trust in the techniques for tasks machine learning is not (yet) suitable for (Narayanan, 2019). For example, many companies are currently using machine learning for hiring decisions despite well-documented failings of these tools. (Narayanan, 2019; Raghavan et al., 2019). In order to avoid such misuse in flood risk assessment, it is important that machine learning implementations are transparent and supervised by independent machine learning and flood risk assessment experts.

Importantly, ~~f~~Flood risk assessments are highly data reliant, and the increased attention to questions of ethics and bias in machine learning systems might serve as an opportunity to drive conversations in our field about the limits of disaster data more broadly. Many of the sources of bias or ethical concerns in machine-learning systems originate in, or share common roots with, other kinds of data used to understand disaster risks ~~and impacts~~. This includes issues such as 1) property values driving ~~where~~ what areas gets protected~~ areas~~, 2) privacy concerns (which may be aggravated by ML and other big data techniques), 3) how the lack of gender/age/ethnicity disaggregated data on disaster risk masks differential vulnerabilities, and 4) the importance of public participation and the voice of residents of areas portrayed by models as "at risk". Detailed analyses of specific cases (e.g. Soden & Kauffman, 2019) are urgently needed to make further progress in understanding the consequences of the assessment methods we use to understand disasters.
~~An additional ethical concern regarding machine learning in flood risk assessment is misuse of models. In some sectors great advances have been made with machine learning (e.g. facial recognition, self-driving cars). This success for some tasks, can~~

400

405

~~3~~5.2 ~~4~~ Future predictions

In the following section we draw some general conclusions about how machine learning will change flood risk and impact assessments. Table 2 provides an overview of these predictions.

~~5~~3.2~~4~~.1 Very likely changes

410 A few of the trends seem inevitable, primarily in cases where recent technological advances or data that recently became available make next steps obvious. A good example of this is the automatic detection of building footprints and roads from high resolution remote sensing imagery (see 2.1.1). This is already possible and will, especially in data-poor areas, drastically improve the quality of the first response and risk calculation. Further advances in the use of machine learning in ~~for flood risk management~~ descriptive hazard assessment through social media are also inevitable (see 2.2.1), given the

415 amount of data available to social media companies.

~~5~~3.2~~4~~.2 Likely and potential changes

This is the category that can be shaped the most by individual innovators and the majority of the advances discussed in this paper fall under this category. In this case, the innovation still experiences some kind of obstacle that prevents widespread

420 application. It is typically difficult to predict whether such obstacles can be truly removed in the future and how long that will take. Because flood risk and impact assessments are a relatively small field, the obstacles are often economic feasibility that is difficult to assess combined with conservative users. An example of this is the large-scale collection of impact data which is required for both descriptive and predictive impact modelling (see 2.3.1 and 2.3.3) or the training data required for descriptive hazard assessments (see 2.2.1). Sometimes the obstacle is also technical feasibility, for example whether it will

425 really be possible to extract first floor elevation levels from streetview (see 2.1.2). Innovations are also interdependent, for example, when building feature information can be automatically extracted from streetview, impact models will become easier to train and easier to run and it will make more sense to start collecting the required impact data.

13

53.24.3 Unlikely changes

For some processes, machine learning may not be the best solution from a theoretical perspective. For example, the
430 processes of how water flows are very well known and can be well approximated with existing equations. It, therefore, does not always make sense to pick a machine learning approach. Another situation when machine learning is not applicable is when a system is being modelled on which predictions need to be made that cannot have been seen in the data or when we know from an exploratory data analysis that we have no data for it (GFDRR, 2018). For example, how a system may behave under never seen discharges or after new infrastructure has been built (e.g. new dam in the river). In these cases, machine
435 learning may play a role in some components of the model, but process models will very likely remain crucial in simulating the never before seen conditions. Especially for predictive hazard models (see 2.2.2), there are many elements that are unlikely to change with the advance of machine learning.

53.24.4 New practices in flood risk and impact assessments

Most change to flood risk and impact assessments discussed in this manuscript relate to better models. Such cheaper, faster
440 and more accurate models could possibly yield new practices in flood risk and impact assessments. Cheaper models would make flood risk and impact assessments feasible to carry out for a larger group of users and are therefore likely to make emergency aid and investments in mitigation measures more efficient. Faster methods may speed up emergency response and recovery, especially when manually collected data from observers on the ground are replaced by remote earth observation. More accurate models may lead to more early actions being feasible (Coughlan de Perez et al. 2014) and hence
445 early actions can be carried out that couldn't be carried out before to new model applications that are difficult with currently available models. For example, more targeted measures during the preparation and response phase of a flood. Such new measures include providing emergency payouts even before the event to the most vulnerable people (e.g. Reuters, 2019), prioritization of emergency measures in buildings, targeted disease outbreak prevention (Coughlan de Perez et al. 2014), early shipping of the right emergency goods (Coughlan de Perez et al. 2014) and prioritization of early harvesting of crops.

450

References

Abrahart, R. J. and See, L. M.: Neural network modelling of non-linear hydrological relationships, Hydrol. Earth Syst. Sci., 11, 1563–1579, https://doi.org/10.5194/hess-11-1563-2007, 2007.

Aarthi, A. D. and Gnanappazham, L., "Comparison of Urban Growth Modeling Using Deep Belief and Neural Network Based Cellular Automata Model - A Case Study of Chennai Metropolitan Area, Tamil Nadu, India," Journal of Geographic Information System, 11, 1-16, 2019.

Alshehhi, R., Marpu, P.R., Woon, W., Dalla Maru, M., 2017. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. ISPRS Journal of Photogrammetry and Remote Sensing. Volume 130, August 2017, Pages 139-149

Amadio, Mattia; Scorzini, Anna Rita; Carisi, Francesca; Essenfelder, Arthur H.; Domeneghetti, Alessio; Mysiak, Jaroslav; Castellarin, Attilio (2019): Testing empirical and synthetic flood damage models: the case of Italy. In Nat. Hazards Earth Syst. Sci. 19 (3), pp. 661–678. DOI: 10.5194/nhess-19-661-2019

Ames, M. G. (2018). Deconstructing the algorithmic sublime. Big Data & Society. https://doi.org/10.1177/2053951718779194

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

Bouwer, L.M., Haasnoot, M., Wagenaar, D. and Roscoe, K., (2017). Assessment of alternative flood mitigation strategies for the C-7 Basin in Miami, Florida. Deltares. 1230718.

15

490 Breiman, L., Friedman, J. H.; Olshen, R. A.; Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. ISBN 978-0-412-04841-8.

Cammerer, H., A.H. Thieken, J. Lammel (2013): Adaptability and transferability of flood loss functions in residential areas. - Nat. Hazards Earth Syst. Sci. 13: 3063-3081.

495

Campolo, M., Soldati, A., and Andreussi, P.: Artificial neural network approach to flood forecasting in the River Arno, Hydrolog. Sci. J., 48, 381–398, https://doi.org/10.1623/hysj.48.3.381.45286, 2003

500 Carisi, F., Schröter, K., Domeneghetti, A., Kreibich, H., and Castellarin, A.: Development and assessment of uni- and multivariable flood loss models for Emilia-Romagna (Italy), Nat. Hazards Earth Syst. Sci., 18, 2057–2079, https://doi.org/10.5194/nhess-18-2057-2018, 2018

Castelletti, A., S. Galelli, M. Restelli, and R. Soncini-Sessa (2010), Tree-based reinforcement learning for optimal water 505 reservoir operation, Water Resour. Res., 46, W09507, doi:10.1029/2009WR008898

Carvajal, T.M., Viacrusis, K.M., Hernandez, L.F.T., Ho, H.T., Amalin, D.M. and Watanabe, K., 2018. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. BMC infectious diseases, 18(1), p.183.

510

Chinh, D., Gain, A., Dung, N., Haase, D., and Kreibich, H.: Multi-Variate Analyses of Flood Loss in Can Tho City, Mekong Delta, Water, 8, 6, https://doi.org/10.3390/w8010006, 2015

Chojaczyk, A., Teixeira, A. P., Neves, L. C., Cardoso, J. B., and Guedes Soares C., 2015. "Review and application of 515 Artificial Neural Networks models in reliability analysis of steel structures," Struct. Saf., vol. 52, pp. 78–89, Jan. 2015.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., and Zsoter, E.: Action-based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549-3560, 520 https://doi.org/10.5194/hess-20-3549-2016, 2016.

16

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., and Suarez, P.: Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts, Nat. Hazards Earth Syst. Sci., 15, 895–904, https://doi.org/10.5194/nhess-15-895-2015, 2015.

525

Curran, A.; de Bruijn, K.M.; Klerk, W.J.; Kok, M. Large Scale Flood Hazard Analysis by Including Defence Failures on the Dutch River System. Water 2019, 11, 1732.

De Waal, J.P., 2016. Basisrapport WBI 2017. Deltares 1230086-002

530

Dibike, Y.B., Solomatine, D.P., 2001. River flow forecasting using artificial neural networks. Physics and Chemistry of the Earth Part B: Hydrology, Oceans and Atmosphere. Volume 26, Issue 1, Pages 1-7.

Eilander, D., Trambauer, P., Wagemaker, J., Van Loenen, A., Harvesting social media for generation of near real-time flood
535   maps. 12th International Conference on Hydroinformatics, HIC 2016

Eubanks, V., 2018. Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

Gao, X., Klaiber, C., Patel, D., Underwood, J., 2019. AI is supercharging the creation of maps around the world.
540   Tech@Facebook. https://tech.fb.com/ai-is-supercharging-the-creation-of-maps-around-the-world/. Accessed: August 21, 2019

Gauss, C.F.,1809.. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientum
GFDRR. 2018. Machine Learning for Disaster Risk Management. Washington, DC: GFDRR. License: Creative Commons
545   Attribution CC BY 3.0

Giacinto, G.and Roli, F.: Design of effective neural network ensembles for image classification purposes, Image Vis. Comput., 19, 699–707, https://doi.org/10.1016/S0262-8856(01)00045-2, 2001

550   Goldblatt, R., Stuhlmacher, M.F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A.K., Cheng, W., Balling, R., 2018. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. Remote sensing of environment 205(2018) 253-275.

Goldblatt, R., You, W., Hanson, G., Khandelwal, A., 2016. Detecting the boundaries of urban areas in india: A dataset for
555   pixel-based image classification in google earth engine. Remote Sensing, Volume 8, Issue 8

17

Heermann, P. D. and Khazenie, N.: Classification of multispectral remote sensing data using a back-propagation neural network, IEEE T. Geosci. Remote., 30, 81–88, 1992.

560 Ivers, L.C. and Ryan, E.T., 2006. Infectious diseases of severe weather-related and flood-related natural disasters. Current opinion in infectious diseases, 19(5), pp.408-414.

Jonkman, 2005. Global Perspectives on Loss of Human Life Caused by Floods. Natural Hazards, Volume 34, Issue 2, pp 151-175.

565

Jonkman, S. N., Voortman, H. G. , Klerk, W. J., van Vuren, S., "Developments in the management of flood defences and hydraulic infrastructure in the Netherlands," Struct. Infrastruct. Eng., vol. 14, no. 7, pp. 895–910, Jul. 2018.

Keyes, O., 2018. The misgendering machines: Trans/HCI implications of automatic gender recognition. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), p.88
570

Khan, A., Khan, H., & Vasilescu, L. (2008). Disaster Management CYCLE – a theoretical approach, Management and Marketing Journal, 6, issue 1, p. 43-50, https://EconPapers.repec.org/RePEc:aio:manmar:v:6:y:2008:i:1:p:43-50.

575 Kind, J., Botzen, W.J., Aerts, C.J.H., (2016) Accounting for risk aversion, income distribution and social welfare in cost-benefit analysis for flood risk management. WIREs Clim Change2016. doi: 10.1002/wcc.446

Kingston, G. B., Rajabalinejad, M. Gouldby, B. P., and. Van Gelder, P.H.A.J.M, 2011. "Computational intelligence methods for the efficient reliability analysis of complex flood defence structures," Struct. Saf., vol. 33, no. 1, pp. 64–73, Jan. 580 2011.

Klemas, V., (2015) Remote Sensing of Floods and Flood-Prone Areas: An Overview. Journal of Coastal Research: Volume 31, Issue 4: pp. 1005 – 1013.

585 Klerk, W., Schweckendiek, T., Den Heijer, F., Kok, M., 2019. Value of information of Structural Health Monitoring in Asset Management of Flood Defences. Infrastructures 2019, 4(3), 56; https://doi.org/10.3390/infrastructures4030056

Koks E.E., Carrera, L., Jonkeren, O., Aerts, J.C.J.H., Husby, T.G., Thissen, M., Standardi, G., Mysiak, J., 2016. Regional disaster impact analysis: comparing input-output and computable general equilibrium models. Nat. Hazards Earth Syst. Sci., 16, 1911–1924, 2016. doi:10.5194/nhess-16-1911-2016

Kreibich, H., Botto, A., Merz, B., and Schröter, K.: Probabilistic, Multivariable Flood Loss Modeling on the Mesoscale with BT-FLEMO, Risk Anal., 37, 774–787, https://doi.org/10.1111/risa.12650, 2017

Kreibich, H., Seifert, I., Merz, B., and Thieken, A.: Development of FLEMOcs – a new model for the estimation of flood losses in the commercial sector, Hydrolog. Sci. J., 55, 1302–1313, 2010.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012)

Kron, W. (2002) Flood Risk = Hazard × Exposure × Vulnerability. In: Wu, M., et al., Eds., Flood Defence, Science Press, New York, 82-97.

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.L., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, R., Kron, W., Benito, G., Honda, Y., Takahashi, K., Sherstyukov, B., 2014. Flood risk and climate change: global and regional perspectives, Hydrological Sciences Journal, 59:1, 1-28, DOI: 10.1080/02626667.2013.857411

Legendre, A.M.,1805. Nouvelles méthodes pour la détermination des orbites des comètes, Firmin Didot, Paris, 1805. "Sur la Méthode des moindres quarrés"

Lin, Y.N., Yun, S., Bhardwaj, A., Hill, E.M, 2019. Urban Flood Detection with Sentinel-1 Multi-Temporal Synthetic Aperture Radar (SAR) Observations in a Bayesian Framework: A Case Study for Hurricane Matthew. Remote Sens. 2019, 11, 1778; doi:10.3390/rs11151778

Lobbrecht A., Solomatine, D., 2002. Machine Learning in Real-Time Control of Water Systems. Urban Water 4, 2002, 283-289

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., 2011. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. May 2011.

Mayfield, H.J., Smith, C.S., Lowry, J.H., Watson, C.H., Baker, M.G., Kama, M., Nilles, E.J. and Lau, C.L., 2018. Predictive risk mapping of an environmentally-driven infectious disease using spatial Bayesian networks: A case study of leptospirosis in Fiji. PLoS neglected tropical diseases, 12(10), p.e0006857.

625

Menderes, A., Erener, A., Sarp, G., 2015. Automatic Detection of Damaged Buildings after Earthquake Hazard by Using Remote Sensing and Information Technologies. Procedia Earth and Planetary Science. Volume 15, 2015, Pages 257-262. https://doi.org/10.1016/j.proeps.2015.08.063

630   Merz, B., Kreibich, H., and Lall, U., (2013), Multi-variate flood damage assessment: a tree-based data-mining approach, Nat. Hazards Earth Syst. Sci., 13, 53-64, 2013. doi:10.5194/nhess-13-53-2013.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A., (2010), Review article "Assessment of economic flood damage", Nat. Hazards Earth Syst. Sci., 10, 1697-1724, doi:10.5194/nhess-10-1697-2010, 2010.

635

Mestav Sarica, G., Zhu, T. and T.- Pan, C., "Flood Exposure of Shenzhen from Past to Future: A Spatio-Temporal Approach using Urban Growth Modeling," Proceedings of the 7th Annual International Conference on Architecture and Civil Engineering, 400–405, 2019.

640   Modu, B., Polovina, N., Lan, Y., Konur, S., Asyhari, A. and Peng, Y., 2017. Towards a Predictive Analytics-Based Intelligent Malaria Outbreak Warning System. Applied Sciences, 7(8), p.836.

Mosavi, A., Ozturk, P., Kwok-wing Chau, 2018. Review: Flood Prediction Using Machine Learning Models: Literature Review. Water 2018, 10, 1536; doi:10.3390/w10111536

645

Naghibi, F., Delavar, M. R., and Pijanowski, B. "Urban Growth Modeling Using Cellular Automata with Multi-Temporal Remote Sensing Images Calibrated by the Artificial Bee Colony Optimization Algorithm," Sensors, 16, 2122, 2016.

Narayanan, A., 2019. How to recognize AI snake oil. https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-
650   snakeoil.pdf (accessed January 27, 2020)

National Research Council. (2006). Introduction Facing hazards and disasters: Understanding human dimensions. Washington, DC: The National Academies Press. https://doi.org/10.17226/11671.

655   Neuhold, G., Ollmann, T., Rota, Bulo, S, Kontschieder, P., 2017. The Mapillary Vistas for Semantic Understanding of Street Scenes. International Conf. on Computer Vision (ICCV) 2017 / July, 2017

Neves, C., González, I., Leander, J., Karoumi, R., 2017 "Structural health monitoring of bridges: a model-free ANN-based approach to damage detection," J. Civ. Struct. Heal. Monit., vol. 7, no. 5, pp. 689–702, Nov. 2017.

660

Nevo, S., Anisimov, V., El-Yaniv, R., Giencke, P., Gigi, Y., Hassidim, A., Mushe, Z., Schlesinger, M., Shalev, G., Tirumali, A., Wiesel, A., Zlydenko, O., Matias, Y., 2019. Machine Learning for Flood Forecasting at Scale. arXiv:1901.09583

Noble, S.U., 2018. Algorithms of oppression: How search engines reinforce racism. nyu Press.

665

Olivas, E. S., Guerrero, J. D., Martinez-Sober, M., Magdalena-Benedito, J. R., & Serrano López, A. J. (2010). Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (2 Volumes) (pp. 1-852). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-766-9

670   Pal, S. and Ghosh, S. K., "Rule based End-to-End Learning Framework for Urban Growth Prediction," CoRR, abs/1711.10801, 2017.

Penning-Rowsell, E. C., Johnson, C., and Tunstall, S.: The benefits of Flood and Coastal Risk Management: A Manual of Assessment Techniques, Middlesex University Press, London, UK, 2005

675

Prendergast, L. J., Limongelli, M. P., Ademovic, N., Anžlin, A., Gavin, K., Zanini, M., 2018. "Structural Health Monitoring for Performance Assessment of Bridges under Flooding and Seismic Actions," Struct. Eng. Int., vol. 28, no. 3, pp. 296–307, Jul. 2018.

680   Pyayt, A., Mokhov, I., Lang, B., Krzhizhanovskaya, V., Meijer, R., 2011. "Machine learning methods for environmental monitoring and flood Protection," World Acad. Sci. Eng. Technol., vol. 78, 2011.

Pyayt, A.L., Kozionov, A.P., Mokhov, I.I., Lang, B., Meijer, R.J., Krzhizhanovskaya, V.V., Sloot, P.M.A., 2014. "Time-frequency methods for structural health monitoring.," Sensors (Basel)., vol. 14, no. 3, pp. 5147–73, Mar. 2014.

685

Raghavan, M., Barocas, S., Kleinberg, J., Levy, K., 2019. Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices. arXiv:1906.09208. DOI: 10.1145/3351095.3372828

21

Reuters, 2019. Bangladesh tries new way to aid flood-hit families: cash up front.
690 https://www.preventionweb.net/news/view/66899

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature volume 323, pages 533–536 (1986)

695 Saadi S., Bensaibi, M., 2014. Detection of Buildings height using satellite monoscopic image. 2nd EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING. DOI:10.13140/2.1.4985.6005

Samardzic-Petrovic, M., Kovacevic, M. , Bajat, B. and Dragicevic, S., 2017. "Machine Learning Techniques for Modelling Short Term Land-Use Change," ISPRS International Journal of Geo-Information, 6, 387, 2017.
700

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F. and Merz, B.(2014), How useful are complex flood damage models? Water Resour. Res. 50, 3378–3395. doi:10.1002/2013WR014396, 2014.

Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., Nagel, C., and Kreibich, H.: Flood loss estimation 705 using 3D city models and remote sensing data, Environ. Modell. Softw., 105, 118–131, https://doi.org/10.1016/j.envsoft.2018.03.032, 2018.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. "Overfeat: Integrated recognition, localization and detection using convolutional networks," CoRR, vol. abs/1312.6229, 2013.
710

Shively, G., Sununtnasuk, C., Brown, M., 2015. Environmental variability and child growth in Nepal. Health & Place. Volume 35, September 2015, Pages 37-51

Soden, R., Kauffman, N. (2019). Infrastructuring the Imaginary: How Sea-level Rise Comes to Matter in The San Francisco 715 Bay Area. In Proceedings of the *2019 CHI Conference on Human Factors in Computing Systems.*

Soden, R., Wagenaar, D. Luo, D., Tijssen, A. 2019. Taking Ethics, Fairness, and Bias Seriously in Machine Learning for Disaster Risk Management. Workshop Paper. Presented at NeurIPS 2019 Workshop on Machine Learning for the Developing World. Vancouver, CA.
720

Solomatine, D.P., Ostfield, A. Data-driven modelling: some past experiences and new approaches. Journal of hydroinformatics 10 (1), 3-22

Song, X., Sexton, J.O., Huang, C., Channan, S., Townshend, J.R., 2015. Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover. Remote Sensing of Environment

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: Decision-tree analysis of factors influencing rainfall related building structure and content damage, Nat. Hazards Earth Syst. Sci., 14, 2531–2547, https://doi.org/10.5194/nhess-14-2531-2014, 2014.

Steenbergen, H.M.G.M., Lassing, B.L., Vrouwenvelder, A.C.W.M., Waarts, P. H., 2004: "Reliability analysis of flood defence systems,"

Suchman, L.A. and Weber, J., 2016. Human-machine autonomies. Autonomous Weapons Systems (pp. 75-102). Cambridge, UK: Cambridge University Press.

Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., and Merz, B.: Development and evaluation of FLEMOps – A new flood loss esimation model for the private sector, WIT Trans. Ecol. Envir., 118, 315–324, 2008.

Tiwari, S., Jacoby, H., Skoufias, E., 2013. Monsoon babies: rainfall shocks and child nutrition in Nepal. Policy Reasearch Working Paper. The World Bank.

Triantakonstantis, D., and Mountrakis, G., "Urban Growth Prediction: A Review of Computational Models and Human Perceptions," Journal of Geographic Information System, 04(06), pp. 555–587, 2013.

UrbanRiskLab, 2019. https://urbanrisklab.org/work#/riskmap/. Accessed: 12-08-2019

Van der Most, H., Tanczos, I., De Bruijn, K.M. and Wagenaar, D.J.,(2014), New, Risk-Based standards for flood protection in the Netherlands. 6th International Conference on Flood Management (ICFM6), September 2014, Sao Paulo, Brazil. 2014.

Wagenaar, D., de Jong, J., and Bouwer, L. M.(2017), Multi-variable flood damage modelling with limited data using supervised learning approaches, Nat. Hazards Earth Syst. Sci., 17, 1683-1696, https://doi.org/10.5194/nhess-17-1683-2017, 2017.

755 Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L.M., Kreibich, H., 2018. Regional and Temporal Transferability of Multivariable Flood Damage Models. Water Resources Research. Volume 54, Issue 5, Pages 3688-3703. https://doi.org/10.1029/2017WR022233

Wagenaar, D.J., Dahm, R.J., Diermanse, F.L.M., Dias, W.P.S., Dissanayake, D.M.S.S., Vajja, 760 H.P, Gehrels, J.C., Bouwer, L.M., 2019. Evaluating adaptation measures for reducing flood risk: A case study in the city of Colombo, Sri Lanka. International Journal of Disaster Risk Reduction. Volume 37, July 2019, 101162.

Wagenaar, D.J., De Bruijn, K.M., Bouwer, L.M. and De Moel, H.,(2016), Uncertainty in flood damage estimates and its 765 potential effect on investment decisions. Natural Hazards and Earth System Sciences, 16(1), 1-14, 2016.

Wagenaar, D.J., Hermawan, T., Van den Homberg, M., Aerts, J.C.J.H., Kreibich, H., De Moel, H., Bouwer, L.M., 2020. Improved transferability of multi-variable damage models through sample selection bias correction. Manuscript submitted for publication.
770

Watmough, G., R., Marcinko, C.L.J., Sullivan, C., Tschirhart, K., Mutuo, P.K., Palm, C.A., Svenning, J., 2019. Socioecologically informed use of remote sensing data to predict rural household poverty. PNAS January 22, 2019 116 (4) 1213-1218; first published January 7, 2019 https://doi.org/10.1073/pnas.1812969116

775 Xu, T., Gao, J. and Coco, G., "Simulation of urban expansion via integrating artificial neural network with Markov chain – cellular automata," International Journal of Geographical. Information Science, 33:10, pp. 1960-1983, 2019. (DOI: 10.1080/13658816.2019.1600701)

Yomwan, P., Cao, C., Rakwatin, P., Suphamitmongkol, W., Tian, R. and Saokarn, A., 2015. A study of waterborne diseases 780 during flooding using Radarsat-2 imagery and a back propagation neural network algorithm. Geomatics, Natural Hazards and Risk, 6(4), pp.289-307.
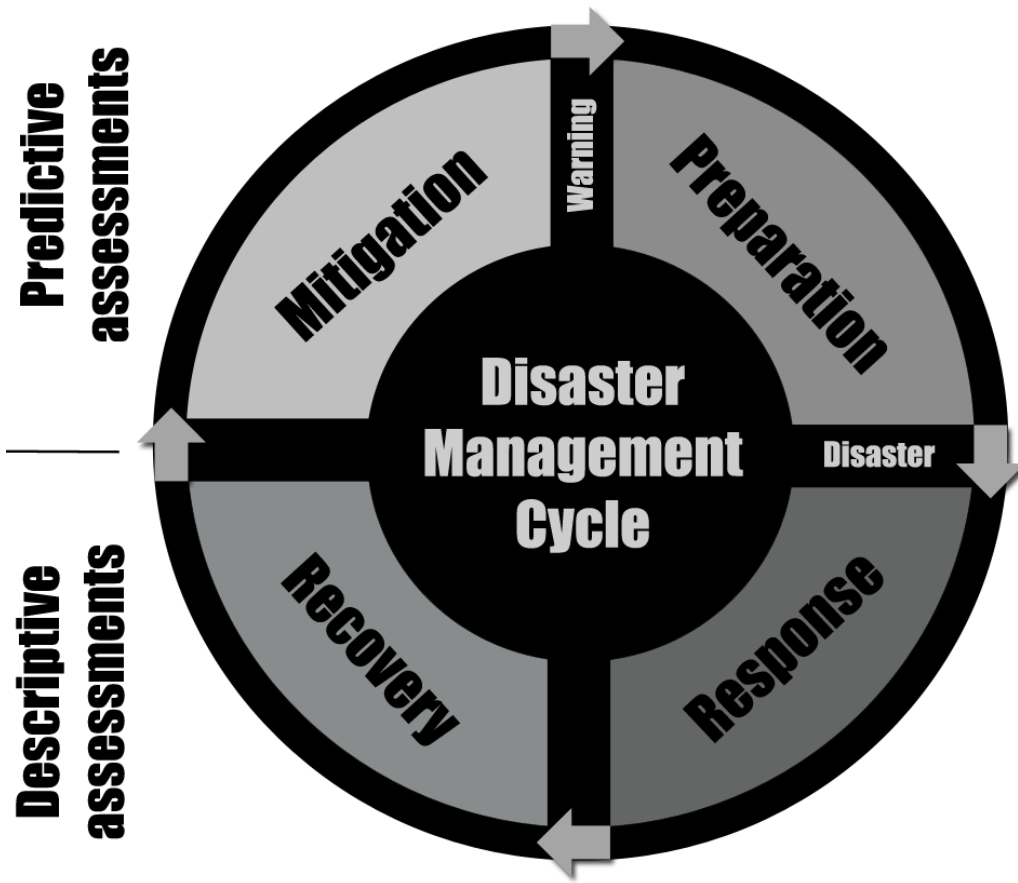
Figure 1: Disaster management cycle, a common paradigm tool.

Table 1: Overview of different types of flood risk and impact assessments

	Predictive	Descriptive
Exposure	Urban growth modelling	Identification of current built-up area
Hazard	Flood modelling	Mapping current and past floods
Impact	Forecasting impact (e.g. damage)	Assessing flood impacts (e.g. damage) after they have occurred

800

805

810

815

820

Table 2: Future predictions

	Predictive	Descriptive
Exposure	Likely incremental changes, e.g improved Cellular Automata transition rules	Very likely significant changes e.g. automatic exposure detection including building features
Hazard	Diverse field, changes are more likely to be complementary or to specific components of modelling	Likely changes in detection due to remote sensing and social media algorithms.
Impact	Potential for significant changes (i.e. multi-variable data-driven methods)	Significant changes likely for some elements others will probably remain the same