Natural Hazards
and Earth System
Sciences
Discussions

Open Access

EGU

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]

5   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
10   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
15   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 becomes available, computing power increases and machine learning algorithms improve. Flood risk and impact assessments are also
20   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
25   improvements may require new data collection efforts, such as for the modelling of flood damages or defence failures. In other fields, machine learning may not 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. This paper presents some of the current developments on the application of machine learning for flood risk and impact assessment, and highlights some key needs and
30   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. These trends are expected to continue and will undoubtedly affect many

35   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 development and climate change (Kundzewicz

et al., 2014). It is therefore an opportunity for researchers and flood managers 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

40 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.

When assessing the interaction between floods and society, three different components can be recognized: hazard, exposure and impact. Hazards are the physical characteristics of a flood such as the extent, water depth, duration and flow velocity.

45 Exposure refers to the characteristics of the people and assets that can be affected by flooding, and impacts are the effects the hazard has on the exposure. 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 descriptive assessments in relation to the hazard,

50 exposure and impact components.

Flood risk and impact assessments have many different applications. A useful paradigm to look at these 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 prevent future events and

55 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. predictive model of people affected based on descriptive model of flood extent). The challenge in this phase is mostly data reliability.   In the

60 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 are used for emergency planning (e.g. Coughlan

65 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,

70 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'. An example of this is in flood damage modelling, where features such as water depth, flow velocity and building

75 materials can be used to predict a target variable such as monetary economic damage based on historical records. 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.

80

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

85 other advanced applications possible (Krizhevsky et al., 2012).

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

90 hasn't displaced the traditional process models. 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).

95 This invited perspective paper provides a review 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.

100

2.Exposure assessment

2.1 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,
105   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. It is already being used to label build-up areas based on nighttime lights (Goldblatt et al., 2018) or satellite images
110   (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.

115   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
120   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
125   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
130   but may be able to detect more complex features in 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 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.

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

3.Hazards assessment

3.1 Descriptive hazard assessments

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

Natural Hazards
and Earth System
Sciences
Discussions
EGU

165

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

170   "flood" (e.g. see Tkachenko et al. 2017, Bischke et al. 2017, Lopez-Fuentes et al. 2017). The method is currently used to 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

175   periods of cloud-cover or at night. Furthermore, the temporal resolution often prevents the observation of flash floods. SAR 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

180   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

185   trainers for automated machine learning methods for the future.

3.2 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

190   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

195   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 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,

200   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

205   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

210   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. For example, Lobbrecht & Solomatine (2002) used machine learning methods to optimise control decisions in the event of communication network

215   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

220   and design, classical reliability techniques such as First order reliability methods (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).

225

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.

230

4. Flood impact assessment

4.1 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

235　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

240　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

245　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

250　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.

4.2 Predictive impact assessments

Predictive flood impact assessments include models that translate hazard and exposure information into socio-economic

255　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

260　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;

265   Wagenaar et al., 2017). 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

270   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

275   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 &

280   Ryan, 2006). Machine learning models could take such complex processes better into account.

Machine learning can be used for structural health monitoring 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

285   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

290   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

295   (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

300 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.

5.Perspectives

5.1 Common challenges

305 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 in the future. Some data is simply not collected or there are measurement definition or quality issues. To fulfil the potential of machine

310 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 data collected for natural hazards, labelled flood extent data, social media hazard data and first floor elevation data.

5.1.2 Transferability of data

A critical assumption behind machine learning techniques is that the data being used to train a model is representative of the

315 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 domain adaptation or transfer learning (GFDRR, 2018) but applicable data is always required as well. Wagenaar et al. (2019) showed that sample selection bias correction, a form of domain adaptation, helps to improve machine learning impact models

320 in a transfer setting. Furthermore, it is important to work on efficient ways to communicate the applicability of data-driven models.

5.1.3 Ethics and Bias

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,

325 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 much attention. Such concerns include 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.

330

Bias problems can occur because of datasets that are not fully representative (e.g. people are accidentally excluded). In addition, we tend to measure what we have data for, rather than measure what matters most, potentially training datasets that reinforce past problems. For example, if certain settlements aren't detected in exposure maps, they may not receive emergency aid. These problems should be addressed by ensuring modelers understand the context of what they are attempting to model.

335 Ethical issues for machine learning 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. The people that are being modelled should be included in this decision making.

340

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 gets

345 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.

350

5.2 Future predictions

In the following section we draw some general conclusions about how machine learning will change flood risk and impact assessments.

5.2.1 Very likely changes

355 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. 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 through social media are also inevitable, given the amount of data available to social media companies.

360

5.2.2 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 application. It is typically difficult to predict whether such obstacles can be truly removed in the future and how long that will

365 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 or the training data required for descriptive hazard assessments. Sometimes the obstacle is also technical feasibility, for example whether it will really be possible to extract first floor elevation levels from streetview. Innovations are also interdependent, for example, when building feature information

370 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.

5.2.3 Unlikely changes

For some processes, machine learning may not be the best solution from a theoretical perspective. For example, the processes of how water flows are very well known and can be well approximated with existing equations. It, therefore, does not always

375 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 learning may play a role in some components of the model, but process models will very likely remain crucial in simulating the never before

380 seen conditions. Especially for predictive hazard models, there are many elements that are unlikely to change with the advance of machine learning.

5.2.4 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 and more accurate models could yield new practices in flood risk and impact assessments. Cheaper models would make flood

385 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 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
390 before the event to the most vulnerable people (e.g. Reuters, 2019), prioritization of emergency measures in buildings, targeted
disease outbreak prevention, early shipping of the right emergency goods and prioritization of early harvesting of crops.

References

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.
405

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

410 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.

415 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.

420

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.

425 Chojaczyk, A., Teixeira, A. P., Neves, L. C., Cardoso, J. B., and Guedes Soares C., 2015. "Review and application of 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-
430 based flood forecasting for triggering humanitarian action, Hydrol. Earth Syst. Sci., 20, 3549-3560, https://doi.org/10.5194/hess-20-3549-2016, 2016.

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.

435
De Waal, J.P., 2016. Basisrapport WBI 2017. Deltares 1230086-002

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.

440
Eilander, D., Trambauer, P., Wagemaker, J., Van Loenen, A., Harvesting social media for generation of near real-time flood maps. 12th International Conference on Hydroinformatics, HIC 2016

Gao, X., Klaiber, C., Patel, D., Underwood, J., 2019. AI is supercharging the creation of maps around the world.
445 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

450 GFDRR. 2018. Machine Learning for Disaster Risk Management. Washington, DC: GFDRR. License: Creative Commons Attribution CC BY 3.0

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

455 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 pixel-based image classification in google earth engine. Remote Sensing, Volume 8, Issue 8

460 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.

465

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.

Khan, A., Khan, H., & Vasilescu, L. (2008). Disaster Management CYCLE – a theoretical approach, Management and

470 Marketing Journal, 6, issue 1, p. 43-50, https://EconPapers.repec.org/RePEc:aio:manmar:v:6:y:2008:i:1:p:43-50.

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

475 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. 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.

480

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
485   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., 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.
490
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks.
Advances in Neural Information Processing Systems 25 (NIPS 2012)

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.L., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell,
495   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
500   Méthode des moindres quarrés"

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

505   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

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., 2011. Big data: The next frontier for
510   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.
515

Natural Hazards
and Earth System
Sciences
Discussions

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

520   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.

525

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.

530   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

535

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.

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

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

545   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.

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

Natural Hazards
and Earth System
Sciences
Discussions

550

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

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

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.

560

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.

Reuters, 2019. Bangladesh tries new way to aid flood-hit families: cash up front.
565 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)

570 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.

575

Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., Nagel, C., and Kreibich, H.: Flood loss estimation 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.

580 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.

Natural Hazards
and Earth System
Sciences
Discussions

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.

585

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
590   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.

595

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
600   urban growth from time series of Landsat-based estimates of impervious cover. Remote Sensing of Environment

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

605   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.
610

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

615

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.

620 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.

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.
625 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, 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,
630 101162.

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

635 Wagenaar, D.J., Hermawan, T., Van den Homberg, M., Aerts, J.C.J.H., Kreibich, H., De Moel, H., Bouwer, L.M., 2019. Improved transferability of multi-variable damage models through sample selection bias correction. Manuscript submitted for publication.

Watmough, G., R., Marcinko, C.L.J., Sullivan, C., Tschirhart, K., Mutuo, P.K., Palm, C.A., Svenning, J., 2019.
640 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

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.
645 (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 during flooding using Radarsat-2 imagery and a back propagation neural network algorithm. Geomatics, Natural Hazards and Risk, 6(4), pp.289-307.
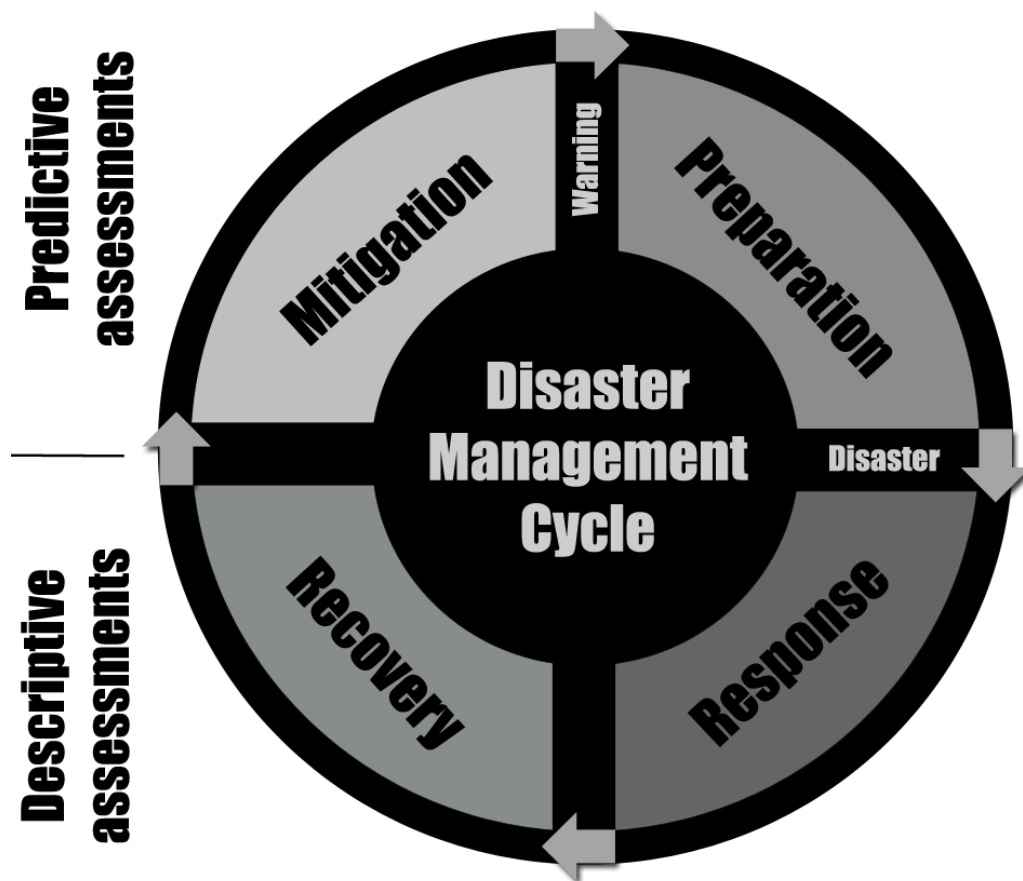
650

Figure 1: Disaster management cycle, a common paradigm tool.

655

660

Natural Hazards
and Earth System
Sciences
Discussions

665   **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

670

675

680

685

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