Accepted Manuscript

Assessing time, cost and quality trade-offs in forecast-based action for floods

Konstantinos Bischiniotis, Bart van den Hurk, Erin Coughlan de Perez, Ted Veldkamp, Gabriela Guimarães Nobre, Jeroen Aerts

PII: S2212-4209(18)31427-4

DOI: https://doi.org/10.1016/j.ijdrr.2019.101252

Article Number: 101252

Reference: IJDRR 101252

To appear in: *International Journal of Disaster Risk Reduction*

Received Date: 10 December 2018

Revised Date: 11 May 2019

Accepted Date: 20 July 2019

Please cite this article as: K. Bischiniotis, B. van den Hurk, E. Coughlan de Perez, T. Veldkamp, Gabriela.Guimarã. Nobre, J. Aerts, Assessing time, cost and quality trade-offs in forecast-based action for floods, *International Journal of Disaster Risk Reduction* (2019), doi: https://doi.org/10.1016/j.ijdrr.2019.101252.

Assessing time, cost and quality trade-offs in forecast-based action for floods

Konstantinos Bischiniotis[1], Bart van den Hurk[1,2], Erin Coughlan de Perez[1,3,4], Ted Veldkamp[1,5], Gabriela Guimarães Nobre[1], and Jeroen Aerts[1]

[1]Institute for Environmental Studies, Vrije Universiteit (VU), Amsterdam, the Netherlands

Deltares, Delft, the Netherlands

[3]International Research Institute for Climate and Society, Columbia University, New York, USA

[4]Red Cross Red Crescent Climate Centre, The Hague, the Netherlands

[5]Water Department, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Abstract

Forecast-based actions are increasingly receiving attention in flood risk management. However, uncertainties and constraints in forecast skill highlight the need to carefully assess the costs and benefits of the actions in relation to the limitations of the forecast information. Forecast skill decreases with increasing lead time, and therefore, an inherent trade-off between timely and effective decisions and accurate information exists. In our paper, we present a methodology to assess the potential added value of early warning early action systems (EWEAS), and we explore the decision-makers' dilemma between acting upon limited-quality forecast information and taking less effective actions. The assessment is carried out for one- and a two-stage action systems, in which a first action that is based on a lower skill and longer lead time forecast may be followed up by a second action that is based on a short-term, higher-confidence forecast. Through an idealized case study, we demonstrate that a) that the optimal lead time to trigger action is a function of the forecast quality, the local geographic conditions and the operational characteristics of the forecast-based actions and b) even low-certainty, long lead time forecasts can become valuable when paired with short-term, higher quality ones in a two-stage action approach.

Keywords: early warning early action system, relative economic value, forecast-based financing, flood risk, decision-making

1. Introduction

Flood risk management aims to reduce the impacts that floods pose to humans and the environment. To achieve this, flood risk mitigation strategies have traditionally focused on long-term protective strategies, using hard infrastructure. However, no matter how high a protection level is, a residual risk always remains. To further reduce this risk 'softer' emergency actions (e.g. temporary flood protection measures, evacuation) (Kabat et al., 2005) that are triggered by forecasts are applied during the time window between the flood alert and the actual event. A system in which warnings are translated into anticipatory actions is called an early warning early action system (EWEAS). EWEAS increase resilience and reduce mortality in low-income countries with recurrent disasters, where limited budgets for structural measures lead to high residual risk (Golnaraghi, 2012). Therefore, EWEAS are considered important components in flood risk management strategies (UNISDR, 2004) and their success is primarily associated with their ability to issue reliable flood alerts at lead times (LT) that are sufficiently long to implement risk reduction measures (UNICEF, 2015).

In flood risk management, EWEAS are usually triggered by hydrological forecast models. These models are subject to different types of uncertainty that are associated with the model itself, the available hydro-meteorological data, the geographical characteristics and the quality of the meteorological forecasts (e.g. Verkade and Werner, 2011; Zappa et al., 2011). To quantify and express this uncertainty probabilistically, ensemble streamflow prediction systems are used. This is achieved by producing multiple forecast simulations by an ensemble of numerical weather prediction and/or with perturbed initial conditions (e.g., Cloke and Pappenberger, 2009; Wetterhall et al., 2013). Probabilistic forecasts are preferred rather than deterministic ones since they give the opportunity to the users to select triggering action probability thresholds based on their minimization or maximization objectives (Roulin, 2007; Krzysztofowicz, 2001; Cloke and Pappenberger, 2009; Jaun et al., 2008; Velázquez et al., 2010; Buizza, 2008).

Similarly to most forecast systems, hydrological probabilistic forecast models exhibit a decrease in skill with increasing LT, revealing an inherent trade-off in the implementation of the EWEAS between timely decisions and accurate information. Recent advances in flood forecasting have led to more informative forecasts, with better skills and longer LTs (Golding, 2009). This has provided the opportunity to take actions that require longer implementation time but may have a larger risk-reducing impact than actions with shorter implementation time. However, in cases where potential consequences of acting in vain are high, postponing actions can be preferred, even if the action effectiveness decreases. Alternatively, decision-makers may decide to follow proactive, no-regret strategies to increase the portfolio of options at a later stage (Heltberg et al., 2009; UNDP, 2010).

In most cases, the basic rationale of EWEAS assumes an essentially linear sequence of actions, starting with the definition of the discharge thresholds that correspond to floods and of the forecast probabilities required to trigger action, the issue of the forecast and the final decision. At a later stage, the evaluation of these systems is usually carried out through cost-benefit analyses (e.g., Murphy, 1977; Katz and Murphy, 1997; Richardson, 2000(Priest et al., 2011)(Priest et al., 2011)(Priest et al., 2011)(Priest et al., 2011)), that is tailored to the needs and requirements of each end-user. Although it is not possible to create an objective measure that quantifies the EWEAS performance for all end-users, the basic rationale is that the EWEAS provide added benefit to the risk mitigation strategies when the benefits (reducing the risk) of taking action outweigh the overall costs (e.g. costs of forecast and other management activities, cost of acting in vain). In the flood risk management context, the cost-benefit analysis has been extensively used to assess the value of different forecast types. For example, Wilks (2001) estimated the economic value of seasonal and weather precipitation forecasts, taking into account their limited reliability. Roulin (2007) assessed the relative economic value of a hydrological ensemble prediction system in two Belgian catchments. Verkade and Werner (2011) compared the benefits of single value and probabilistic forecasts for a range of LTs and Matte et al. (2017) incorporated risk aversion into the cost-loss decision model. While these studies have assessed the value of EWEAS for a single action-forecast combination, they have not examined the potential benefits of preparatory measures that are triggered by forecasts with longer lead times. In addition, they have used discrete values for the ratio between residual and potential damage over time, while budget and implementation time constraints are not taken into account.

In this study, we build on existing valuation approaches to present a methodology that assesses the economic value of EWEAS, taking into account trade-offs concerning forecast quality, restrictions in the implementation of actions, and time-varying costs and losses. The assessment is carried out for an one- and a two-stage action system, in which a first action that is based on a lower skill and longer lead time forecast is followed up by a second action that is based on a short-term, higher-confidence forecast. We demonstrate the EWEAS added value in an idealized case study, using forecast data from the global flood awareness (GloFAS) in Akokoro, Uganda. We must note that the scope of our paper is not to profoundly analyse the model's forecast skill for this case study, but rather to demonstrate how an operational forecast and its skill assessment can be incorporated into the decision-making process.

The paper is organised as follows: In section 2, we present the necessary background information for the evaluation of EWEAS. In section 3, we outline the basic components of the EWEAS we have used in our idealized case study, and in section 4, we present the results. In section 5, we discuss the limitations of this study and outline options for further research. In section 6, we summarize the main conclusions.

2. Methods: evaluation of a flood Early Warning Early Action System (EWEAS)

In this section, we present the necessary components to consider for the evaluation of EWEAS (Figure 1):

- the forecast model that provides the early warnings, which in our study is GloFAS (section 2.1);
- the discharge thresholds that correspond to floods of different magnitudes, the probabilistic thresholds that trigger action, and the forecast skill assessment at different lead times(section 2.2);
- the forecast-based actions and the differences in taking action at one- and at two-time steps.(sections 2.3 and 2.4).

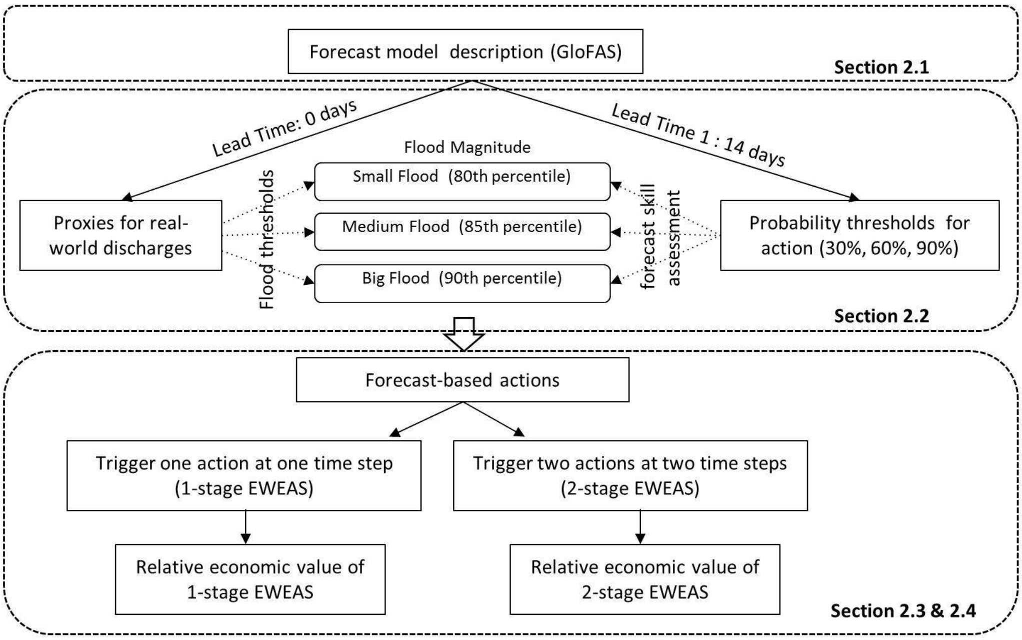

Figure 1 Flowchart that outlines the steps taken towards the configuration and evaluation of EWEAS

2.1 Forecast model description: GloFAS

Every flood risk mitigation decision-making process starts with the application of a forecast model. In this study, we use the Global Flood Awareness System (GloFAS) (Alfieri et al., 2013), a global model that produces daily forecasts to issue flood alerts at a 0.1^{o} spatial resolution by using 51-ensemble member streamflow forecasts, each driven by meteorological forecasts 15 days ahead. Its forecast probabilities are based on the fraction of the ensemble members exceeding a predefined discharge threshold. For example, if 10 out of 51 members exceed a threshold, the probability of its exceedance is 0.19. GloFAS is being used operationally by the forecast-based financing project of the Red Cross (Coughan de Perez et al., 2015) in several developing countries around the world such as Peru, Bangladesh, Nepal, and Uganda. For a more detailed discussion on GloFAS, we refer to Alfieri et al. (2013).

In our study, we used GloFAS forecasts for the river cell of the Victoria Nile that exhibits the highest annual mean discharge in the Akokoro subcounty in Apac district, Uganda (1.55N, 32.55E). This area has experienced catastrophic flood events in the past (e.g. August 2007, October 2012) and has been used as a case study of the partners for resilience project (https://partnersforresilience.nl/).

2.2 Thresholds for triggering action and forecast skill assessment

To evaluate forecast skill it is first needed to define discharge thresholds that are representative for flood events. In operational EWEAS, when the forecasted discharges exceed these thresholds at pre-agreed probabilities, flood risk mitigation actions are triggered. Regarding the skill of the forecast model, decision-makers are mostly interested in the event-based metrics, namely the correct hits (CH), the misses (MS), the false alarms (FA) and the correct negatives (CN), since these are necessary for the actual valuation of losses and benefits. A forecasting model that systematically underestimates the probability of floods leads to a high likelihood of missed events, while overestimations lead to frequent false alarms. Given the absence of perfect forecasts, decision-makers aim to set the action-triggering forecast probabilities in such a way that they meet their requirements, while at the same time maximize the potential benefits of using the forecast model. For instance, Coughlan de Perez et al. (2016) identified the forecast probabilities of GloFAS that should trigger action in two districts in

Uganda, using as basic criterion that the FA ratio, which is the verification score of interest to humanitarians (Hogan and Mason, 2012) and is defined as the number of false alarms per total number of alarms, is lower than 0.5. On the other hand, under other circumstances (e.g. budget restrictions),decision-makers prefer not to take action unless they are absolutely certain that an upcoming hazard will occur (Demeritt et al., 2007; Suarez and Patt, 2004).

These event-based metrics are usually calculated over aggregated large spatial scales, such as a country or a continent (Thiemig et al., 2015; Bischiniotis et al., 2019), given the limited availability of sufficient information on rare flood events at specific locations. However, EWEAS are usually applied to smaller spatial scales (e.g., a village, town or province) and consequently, end users are interested in the local forecast skills.

To be in line with this need, we used daily flood forecasts from GloFAS over a period of approximately 8 years (between May 1st 2008 and December 31st 2015) for a specific location with lead times from 0 to 14 days (LT0 to LT14) to a) set the discharge thresholds above which a flood occurs, and b) evaluate different forecast probability thresholds that trigger action. We used the LT0 discharges, which refer to the initial conditions that forecasts were issued, as a proxy for the real-world discharge. From this time series, we calculated the 80th, 85th and 90th percentile, considering that they represent the thresholds of small-, medium- and big-magnitude floods, respectively, similarly to Coughlan de Perez et al. (2016). In the real world, we would expect much higher discharge percentiles to trigger flood events, but given the limited available forecast time series, we used relatively low ones to generate sufficient statistics and demonstrate the concept of our methodology. We distinguished different flood magnitudes to illustrate the diversity of the model skill in predicting different floods, as well as to address how the budget, time constraints, costs and damage have an effect on different flood outcomes. We used three probability thresholds for triggering action (30%, 60% and 90%) to demonstrate that this can also affect the overall usefulness of the EWEAS. The probabilities are estimated using the different members of the ensemble of GloFAS forecasts as indicated in 2.1.

In our study, the forecast skill assessment is carried out using the forecasts of each LT separately for all three probability thresholds and for all three flood thresholds (Table 1), taking also into account the period that the action can provide protection, following Coughlan de Perez et al. (2016). This means that as soon as an action is triggered after a forecast warning, it has a lifetime period, within which the action is not re-triggered and can provide protection effectively. Taking action’s lifetime into account is a consideration that potentially increases the forecast skills since in case a flood does not occur exactly on the forecasted date but within the lifetime period, the flood signal is counted as correct hit (CH). If there is no flood during this period, the flood signal is counted as false alarm (FA), while if a flood occurs but no flood signal was issued, it is a Miss (MS). The flood conditions (i.e. discharge higher than the threshold) can remain after the expiration of the action’s lifetime. In this case, if there is a flood signal, the action is re-triggered, while flood conditions are ongoing. In our analysis, we considered this case a new event (we further discuss this in section 2.4). Furthermore, each flood magnitude is treated separately and thus, successive exceedance of different flood magnitude thresholds (e.g. first a small and later medium flood) are regarded as two individual events, i.e. one small and one medium flood.

Table 1 Event-based metrics such as Correct Negatives (CN), Misses (MS), False Alarms (FA), and Correct Hits (CH)) are calculated for each flood magnitude (FM_Q), probability threshold (PT_i) and lead time (LT_j).

Flood Magnitude(FM_Q)	Small (Q80)/Medium (Q85)/Big (Q90)	
Probability Threshold (PT_i)	i=30%,60%,90%	
Lead Time (LT_j)	j=1:14	
Event-based metrics	$CN(FM_Q,PT_i,LT_j)$	$MS(FM_Q,PT_i,LT_j)$
	$FA(FM_Q,PT_i,LT_j)$	$CH(FM_Q,PT_i,LT_j)$

2.3 Forecast-based actions

A wide range of potential forecast-based actions exists in early action protocols, all having different features: cost, implementation time requirements, lifetime, tangible and intangible benefits. For example, temporary flood measures such as sandbags can be installed or put in place to protect dwellings and critical infrastructure; evacuation can be applied to reduce fatalities and chlorine tablets can be distributed to provide clean water and prevent the spread of disease. In some cases, the actions can be complementary. To demonstrate this relationship, we use two decision-making approaches: a static (one-stage action) and a dynamic (two-stage action) one. In the first, a decision for action is taken at one point in time. In the second, decisions are taken at two time points; initially a preliminary action at longer LT and subsequently a main action. In our case, the preliminary action is not a prerequisite for triggering the main action but is used to facilitate it, as it is explained in sections 2.4.2 and 3), if this is triggered at a later LT. In this way, we assess the added value of sequential decision-making, similar to the 'ready-set-go' approach, a methodology applied within the humanitarian sector allowing the progressive staging of actions (Goddard et al., 2014).

2.4 Relative economic value of EWEAS

To evaluate the EWEAS, we use its relative economic value (V_{ew}) (e.g. Katz & Murphy, 1997, Verkade and Werner, 2011, Lopez, et al., 2018). This is defined as the relative reduction in total losses from disaster responses when using early warnings by a forecast model (TL_{ew}) compared to the total losses when no forecast model is available and only climatological probability information is used (TL_{no_ew}) (Eq. 1):

$$V_{ew} = (TL_{no\_ew} - TL_{ew})/TL_{no\_ew} \quad \text{(Eq.1)}$$

where,

V_{ew}: Relative economic value of the EWEAS

TL_{no_ew}: Total losses incurred when there is no forecast

TL_{ew}: Total losses incurred when action is taken based on a forecast

When $V_{ew} > 0$, the EWEAS provides added value in flood risk mitigation, since losses are lower when appropriate forecast-based actions are implemented compared to not taking action at all.

2.4.1 Evaluation of an one-stage action EWEAS

In an one-stage action system, decision-makers have to choose between two options at each time step: to take action or to wait for further forecast information that comes with shorter LTs. Therefore, this choice can be seen as a repetitive problem, in which decision-makers face the same dilemma at each LT, until action is taken (Figure 2 left).

To compute the relative economic value of the EWEAS (V_{ew}), the event-based skill metrics (CH, MS, FA and CN) are required. As mentioned in section 2.2, in our study, we a) calculated these metrics for each flood magnitude, for all three probability thresholds (i.e. 30%, 60% and 90%) and for each forecast LT(Figure 2, right) and b) the forecast-based action is triggered if the forecast issues a warning that exceeds the predefined threshold, while no action is taken when no warning is issued. The forecast-observation pairs are illustrated in the contingency table (Table 2).

Table 3 shows the consequences of these pairs; when no action is taken and a flood occurs (MS), the losses are equal to the damage (D) that corresponds to the observed flood magnitude. When action is taken in vain in case of a FA, the losses are just the implementation costs of the action taken (C). When action is correctly taken (CH), the total losses are the sum of the action costs (C) and the residual damage that has been partly or entirely mitigated thanks to this action (RD). Therefore RD <= D. When no warning is issued and no flood occurs (CN), there is no action and no damage. In case of an FA, there is often a change to the original cost, ΔC that may account for e.g. the reputational risk

(Coughlan de Perez et al., 2015). Although this can be significant in some cases, we assume that it is 0.

The forecast-based actions are not instantly carried out. For this reason, we consider that a longer LT allows more implementation and the actions are more effective in damage reduction. Hence, the cost of the action is a function of time and implementation requirements and therefore, the action's effectiveness and consequently the residual damage are also dependent on the available budget, the implementation costs and requirements. This is illustrated with an example in section 3.

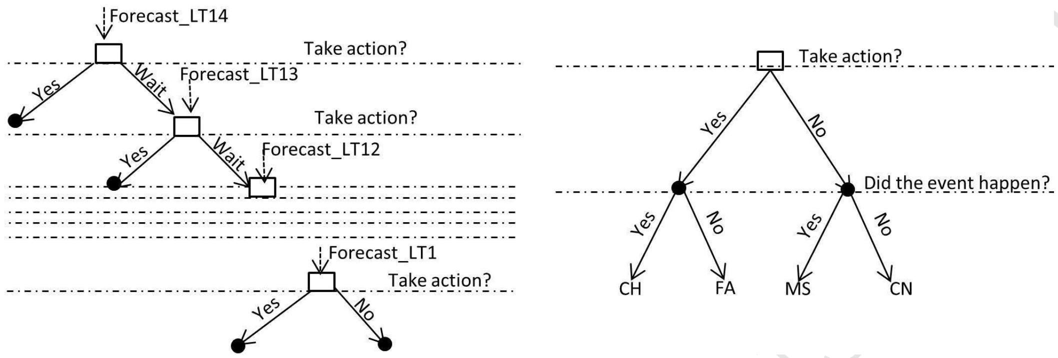

Figure 2 One-stage Action: the repetitive dilemma of whether or not to trigger action (left), and the event tree (right) used to calculate the event-based skill metrics (i.e. Correct Hit (CH), Miss (MS), False Alarm (FA) and Correct Negative (CN)). The dashed lines demonstrate the different time steps, the squares the time points that decisions need to be made and the black dots the time points of a final decision.

Table 2 Contingency table illustrating the evaluation metrics (CN: Correct Negatives, MS: Misses, FA: False Alarms, CH: Correct Hits) based on the forecast probability that a certain discharge will be exceeded in relation to the probability threshold to trigger action.

	Flood	No Flood
Forecast probability > probability threshold	CH	FA
Forecast probability < probability threshold	MS	CN

Table 3 Contingency table that illustrates the cost of action (C), damage (D) and residual damage (RD) when forecast-based action is taken.

	Flood	No Flood
Forecast probability > probability threshold	C+ RD	C
Forecast probability < probability threshold	D	0

The total losses of having no EWEAS (TL_{no_ew}) are equivalent to using the total number of flood events (i.e. MS + CH) multiplied by the damage (D) corresponding to each flood magnitude (Eq.2).

$$TL_{no\_ew} = (CH + MS) \cdot D \qquad \text{(Eq.2)}$$

The total losses (TL_{ew}) when taking action based on a one-stage EWEAS over a finite time period is calculated by aggregating the product of the losses of each forecast and observation pair (Table 3) and their corresponding occurrences (Table 2; Eq.3).

$$TL_{ew} = (CH) \cdot (C+RD) + (FA) \cdot (C) + (MS) \cdot D \qquad \text{(Eq. 3)}$$

In reality, a failure of the measure can have the same consequences as a miss and cannot be neglected. To avoid this level of complexity, however, we assumed in this analysis that the failure probability of the action taken is 0. In the supplementary material, we present the equation when accounting for the failure probability (Eq. S1).

2.4.2 Evaluation of a two-stage action EWEAS

As discussed in 2.3, in a two-stage action system, decision-makers have the option to take preliminary actions triggered at longer LTs (e.g. at LT14), followed by a main action triggered at shorter LT (e.g. between LT13 and LT1). The preliminary action facilitates the implementation of the main action, increasing its effectiveness. Similarly to the one-stage action, decision-makers face the dilemma to wait or act (Figure 3, left). This procedure can be more complicated if the decision-maker is granted a range of days to trigger preliminary action (e.g., anytime between LT14 and LT7). However, for the sake of simplicity, we assume that preliminary action can be triggered only at LT14 and is implemented within one day, as it will be discussed in section 3. In result, the estimation of the relative economic value (V_{ew}) of the EWEAS requires the joint performance of the two lead time forecasts in relation to the outcome (i.e. flood or no flood) (see Table 4) (e.g. forecast at LT14 – CH and forecast at LT1- CH, forecast at LT14 – CH and forecast at LT1- MS). In this way, for each LT triggering action, our contingency table has eight entries (Figure 3, right). The probability thresholds used to trigger the preliminary and the main actions are not necessarily the same. Therefore, the skill metrics of the entire system are different for each threshold combination used. In our case, there are 9 combinations possible (i.e. 30%, 60%, 90% for LT14 (threshold 1) times 30%, 60%, 90% for the later LTs (threshold 2)).

The total losses from taking action are calculated by the aggregation of the actions' implementation costs and the residual damage that accrue from the joint system of two forecasts (Table 5) multiplied by their corresponding occurrences (Table 4). In practice, given the restricted budget that is usually allocated to forecast-based measures, decision-makers are requested to determine in advance the budget fraction that is allocated to the first and second stages; in our study this budget allocation is fixed (see example in section 3). However, the aggregation of the cost of the preliminary (C_1) and the main actions (C_2) cannot exceed the available budget. Although we consider that preliminary action has implementation costs, it is only used to facilitate the main action rather than providing protection against floods itself. Thus, when only preliminary action is taken, damage is not mitigated. On the other hand, when the main action is triggered, damage is mitigated regardless if preliminary action is taken (RD_{12}) or not taken (RD_2). However, since the preliminary action increases the effectiveness of the main action, $RD_{12} <= RD_2$.

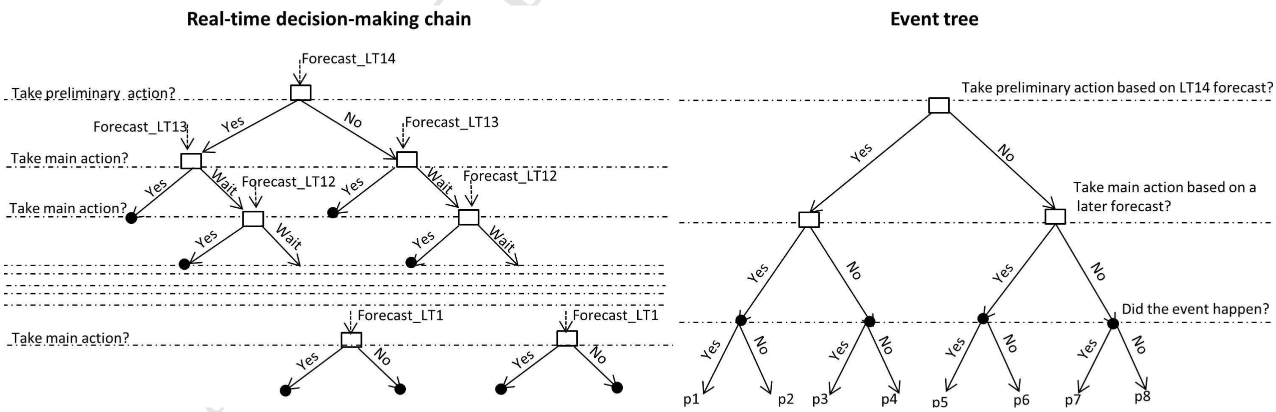

Figure 3 Real-time decision-making chain that illustrates the decision-makers' dilemma of whether and when to take preliminary and main actions (left), and the event tree used to calculate the evaluation metrics of the joint forecast system in the two-stage action system. The dashed lines demonstrate the different time steps, the squares the time points that decisions need to be made and the black dots the time points of a final decision.

Table 4 Contingency table that outlines the evaluation metrics (p1:p8, see Figure 3 right) for the two-stage action system based on the forecast probabilities in relation to different triggering action thresholds for the preliminary action (triggered by forecast 1 [F1] at LT14) and the main action (triggered by forecast 2 [F2] between LT13 and LT1).

	F_1 probability > probability threshold_1		F_1 probability < probability threshold_1	
	Flood	No Flood	Flood	No Flood

F2 probability > probability threshold_2	$p_1=CH_{F1}\cap CH_{F2}$	$p_2=FA_{F1}\cap FA_{F2}$	$p_5=MS_{F1}\cap CH_{F2}$	$p_6=CN_{F1}\cap FA_{F2}$
F2 probability < probability threshold_2	$p_3=CH_{F1}\cap MS_{F2}$	$p_4=FA_{F1}\cap CN_{F2}$	$p_7=MS_{F1}\cap MS_{F2}$	$p_8=CN_{F1}\cap CN_{F2}$

Table 5 Contingency table that presents the costs and damage of taking action at two stages. Preliminary action is triggered by forecast 1 (F1) at LT14 and main action is triggered by forecast 2 (F2) between LT13 and LT1.

	F_1: LT14 > threshold_1		F_1: LT14 < threshold_1	
	Flood	No Flood	Flood	No Flood
F_2 probability > threshold_2	$C_1 + C_2 + RD_{12}$	$C_1 + C_2$	$C_2 + RD_2$	C_2
F_2 probability < threshold_2	$C_1 + D$	C_1	D	0

Similar to a one-stage system, the V_{ew} is calculated using the total losses when there is no EWEAS (Eq.4) and when EWEAS is used (Eq.5);

$$TL_{no\_ew} = (p_1 + p_3 + p_5 + p_7) \cdot D \qquad \text{(Eq.4)}$$

$$TL_{ew} = p_1 \cdot (C_1+C_2+RD_{12}) + p_2 \cdot (C_2+C_2) + p_3 \cdot (C_1+D) + p_4 \cdot (C_1) + p_5 \cdot (C_2+RD_2) + p_6 \cdot (C_2) + p_7 \cdot D \qquad \text{(Eq.5)}$$

As in 2.4.1, the equations used hereby do not take into account the failure probability of the risk mitigation measures. Equation S2 in the supplementary material presents the total losses in case the failure probabilities of both the main and preliminary actions are taken into account.

3. Configuration of the EWEAS used in our case study

In addition to the generic methods and parameters described in Section 2, EWEAS should be configured based on the needs, requirements and risk mitigation capabilities of the study areas. To facilitate the reader’s understanding and demonstrate some of the key features that are important in operational flood risk decision-making, in our study, we use volunteer training and sandbag dike construction as examples of preliminary and main forecast-based actions, respectively. Based on these actions, we show a) how the financial, temporal and location parameters interact with each other and b) how they lead to the calculation of the residual damage after the implementation of the EWEAS that is necessary for its evaluation (Figure 4).

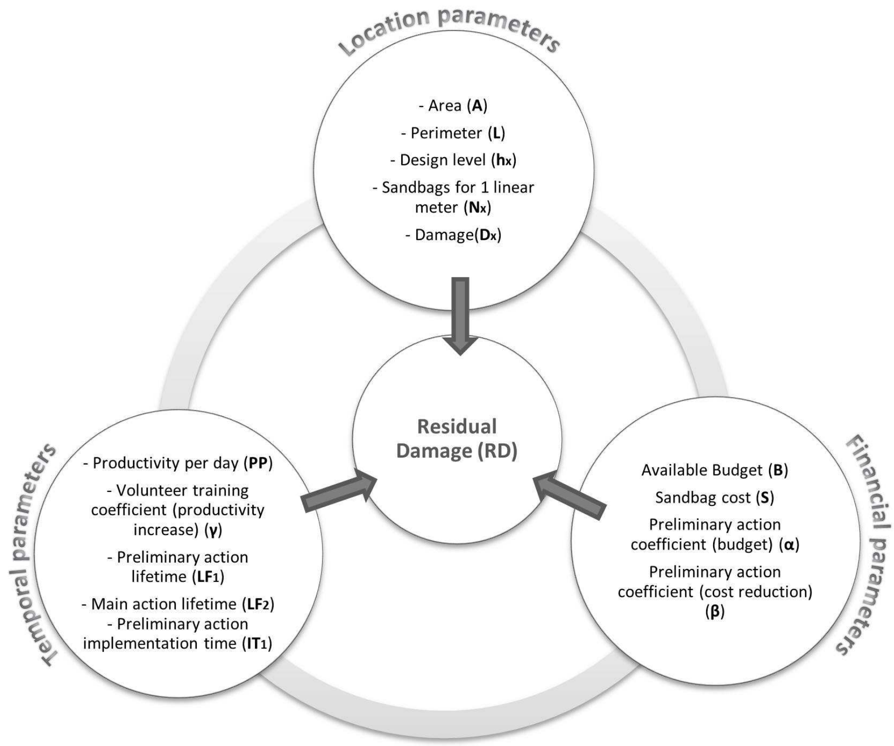

Figure 4 Scheme showing the parameters that are taken into account in our case study example.

In our example, the decision-makers use the EWEAS to provide protection at a fictitious area with size A and perimeter L during the time period that GloFAS forecasts are available. Although a lot of flood adaptations are available, for the sake of simplicity, we here assume only one forecast-based action: to construct a sandbag dike ring around the area every time a flood warning is issued. Sandbags are often readily available in developing countries such as Uganda, at relatively low cost and are effective in preventing flooding with water levels of up to one meter in height (Kelman and Spence, 2003; Botzen et al., 2009). To achieve greater effectiveness, we assume that sandbags are prepositioned in the location (Rawls & Turnquist, 2010). Although forecast LT and mitigation time can be different (following the forecast issue, time is required to disseminate it and take action (Carsell et al., 2004), we consider these two to be identical similarly to Verkade and Werner (2011). The use of other measures would require some adaptations, but the basic rationale would remain the same.

As discussed in section 2, we treat each lead time separately. Action is triggered (i.e. the sandbag dike construction starts) as soon as a flood forecast warning is issued and is not interrupted by successive forecasts that may 'recall' the flood signal. The design height depends on the threshold above which a flood is defined (h_s, h_m or h_b, with the subscripts s, m and b referring to small-, medium- and big-magnitude floods, respectively) and we assume that protects against all floods. To reach this height for one linear meter, N sandbags are needed (N_s for small-, N_m for medium- and N_b for big-magnitude floods, respectively). Given the trapezoidal sandbag dike cross-section, these numbers are not linearly proportional to the water level. The total dike length that can be constructed L_d depends on the design dike height, the placement productivity rate PP (sandbags placed per day) that the available manpower allows (i.e. with one day LT (LT1), we can place 1·PP sandbags, with two days LT (LT2), 2·PP, etc.), and consequently on the forecast LT of triggering action (i.e. the longer the LT, the more time available). In our example, the sandbag dike ring has a square shape, and therefore, the area that can be protected is calculated in Eq. 6.

$$\text{Area Protected} = \left(\frac{\frac{\text{LT}\cdot\text{PP}}{\text{N}_\text{x}}}{4}\right)^2 \qquad \text{(Eq.6)}$$

Therefore, the cost of the main action is not only subject to the flood magnitude, which determines the height and the number of sandbags that should be placed, but it is also a function of the LT, at which action is triggered, and of the PP, which determines how many of them can be placed.

In addition, as it happens in reality, the budget B (USD) that is allocated to the forecast-based actions is restricted and therefore, the maximum total costs and protected area are subject to this restriction. In the one-stage action system (see section 2.4.1), the entire budget is used for the sandbag dike construction (main action), which involves the purchase and placement cost S (USD/bag) by employed personnel. In the two-stage action (see section 2.4.2), a fraction α of the total budget is allocated to the preliminary action, leaving $(1-\alpha)\cdot B$ available for the main action. When the initial forecast at LT14 does not issue a flood warning signal, preliminary action is not triggered. Hence, the entire budget can be used for the main action.
In our study, we use as an example of preliminary action volunteer training, whose potential in disaster impact mitigation is increasingly recognized worldwide (Whittaker et al., 2015). This facilitates the main action, both monetarily and temporally, by a) reducing the cost S per sandbag with a factor β, since no placement by employed personnel is needed and b) increasing the placement productivity rate PP by a factor γ. The preliminary action has a lifetime LF_1 days and the main action LF_2 days. We assume that the preliminary action has a fixed implementation time IT_1, which lasts one day (see section 2) and its LF_1 lasts as many days as main action is being implemented, if it is triggered by the following forecasts so as the main action is constantly facilitated. As described in section 2.2, LF_2, which is involved in the calculation of the event-based metrics, is fixed and exceeds the forecast range so no extra action is needed during this period. When the flood duration exceeds LF_2, we consider that action as triggered anew, if the forecast continues to predict high discharge levels. In the real world, effort would be exerted to expand the action's lifetime through maintenance activities that require less cost and implementation time. However, to avoid this level of complexity, we treat the two actions equally, using the same costs and implementation time as if no sandbag dike is present. The potential damage D, when no mitigation action is taken, depends on the flood magnitude (D_s for small-, D_m for medium- and D_b for big-magnitude floods).

Financial and temporal constraints lead to restrictions on the total area A that is protected. This partial protection is a metaphor for real situations, in which authorities prioritize the areas to protect. In our case, when the main action is triggered, the residual damage RD is the fraction of the area that is protected per total area multiplied by the potential damage (Eq.7). This implies that potential damage is homogeneously distributed in the area and that residual damage is only a function of the protected area, which stays completely dry, whereas the unprotected area is flooded. This is a result of the assumption that sandbags can only reduce water level entirely in the protected area and not partly. Therefore, decision-makers of our EWEAS aim to create a sandbag dike ring with sufficient height for a smaller area rather than protecting a larger area with lower dike. In case the action is able to partly reduce the water column in the protected area, then Equation 7 would be multiplied by an effectiveness ε that would be function of the inundation level.

$$\mathrm{RD} = \frac{\text{Area protected}}{\mathrm{A}} \cdot \mathrm{D} \qquad \text{(Eq.7)}$$

Figure S1 (supplementary) show schematically the steps taken to calculate the protected area. The numerical values of all parameters presented are given in the Table S1 (supplementary).

For the one-stage EWEAS, we calculate the relative economic value V_{ew} for the time and budget restrictions that we presented, and we carry out a sensitivity analysis to examine how the V_{ew} of each flood magnitude is affected by the absence of restrictions on budget or time. Subsequently, we calculate the V_{ew} for the two-stage EWEAS. The sensitivity analysis was not carried out for the two-stage EWEAS, since the budget and the implementation time of the preliminary action are considered to be fixed and hence, they do not depend on budget and time changes. We must also note that our model is different from the 2-stage system described in Katz and Murphy's (1997). In their work, the budget is used all at once (to take actions that completely eliminate risk), damage can accrue at various points in time and an early action does not serve as a facilitator of a later one.

4. Results

4.1 Forecast skill

Figure 5 displays the daily discharge produced by the GloFAS simulations at LT0 for the period between 1 May 2008 and 31 December 2015. The wet season in that area is from April until November, with a principal peak between April and August, and the dry season is from December until March. The daily discharge time series values are used as a baseline for observed flood occurrences (small flood [80th percentile-blue line], medium flood [85th percentile-red line] and big flood [90th percentile-green line]). The main action lifetime LF_2 is 30 days (see Table S1 in the supplementary material). As described in sections 2.2 and 3, if a flood lasts longer than this period, a new event is considered to have occurred. If the discharge exceeds a higher threshold, we also count the number of lower threshold events (e.g. if the 90th percentile is exceeded, we count one event for big-, one for medium- and one for small-magnitude events). So, the number of independent events against which action can be taken is 21 for small-, 16 for medium- and 12 for big-magnitude floods.

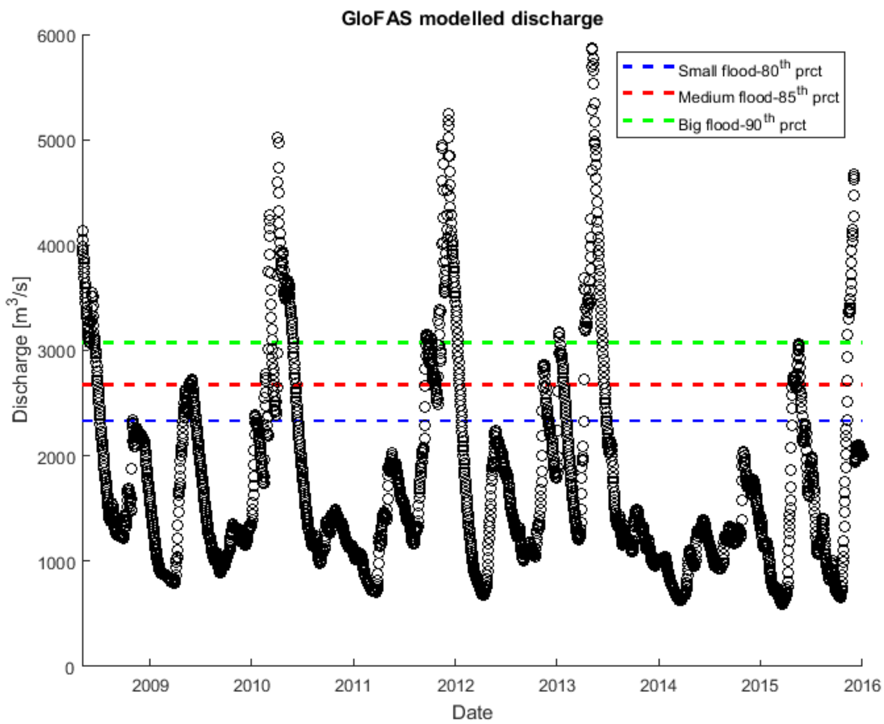

Figure 5 The GloFAS modelled daily discharge at LT0 from 1 May 2008 until 31 December 2015 for Akokoro, Uganda. Blue, red and green lines denote the triggering action thresholds for small (80th percentile), medium (85th percentile) and big (90th percentile) floods, respectively.

Figure 6 presents the CH and FA as functions of the forecast LT for the three flood magnitudes and the three triggering action probability thresholds (30%, 60% and 90%). The MS rates are implicitly indicated, since they are equal to the difference between the number of events of each flood magnitude and the CH. We observe that up to LT4, the number of CH usually remains the same and it decreases with longer LTs; as a consequence, MS increases. The relationship between FA and LT is not as straightforward, but in general, the number of FA is higher for smaller magnitude floods and lower probability thresholds. Furthermore, we can observe that both the number of CH and FA is not strongly sensitive to the selected probability threshold. This can be attributed to a) the fact that in this river cell, the model tends to forecast high discharges using high probabilities, b) the limited number of events and c) the fact there are some cases where flood events last longer than the action's lifetime and therefore, forecasts predict with high certainty that the discharge remains above the flood thresholds during the flood period.

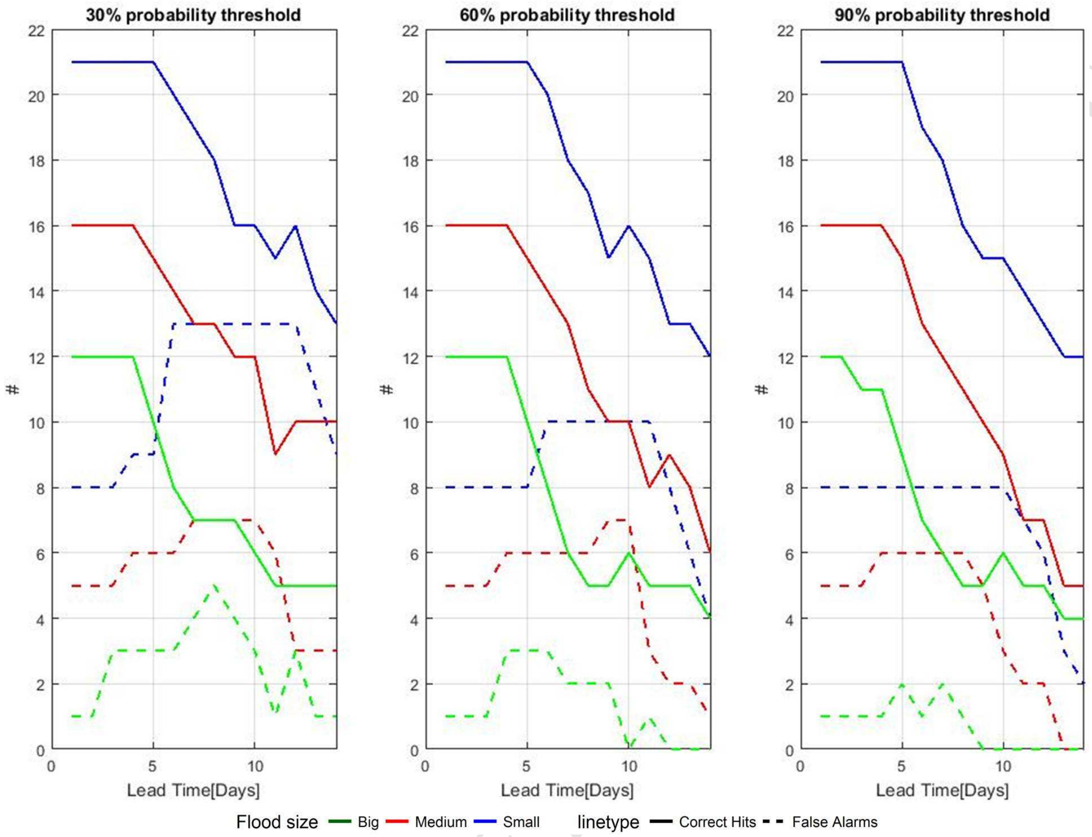

Figure 6 Forecast skill expressed in number of Correct Hits (CH) (solid lines) and False Alarms (FA) (dashed lines) as functions of lead time (x axis) for all three flood magnitudes (small flood: blue line, medium flood: red line, big flood: green line) when using 30% (left), 60% (medium) and 90% (right) threshold probabilities of detecting a flood.

4.2 Added value of EWEAS in one-stage approach

Figure 7 presents the ability of the EWEAS to provide protection to the entire study area by creating a sandbag dike around it. This is demonstrated for the different flood magnitudes and for each LT that an action can be triggered, taking into consideration budget (B) and placement productivity (PP) constraints, which determine whether there is sufficient implementation time (IT) for the action. So, using the parameters from Table S1, when the protected area (Equation 6) is larger than the actual study area, it means that there is both sufficient time to protect the entire area and budget to finance the action costs (Figure 6, green box). Similarly, we demonstrate the result for the other IT/B combinations. For small floods, the budget requirements are low, and given the available sandbag placement productivity rate, there is a temporal cut-off point only at LT4. At shorter LTs, there is not sufficient time to construct a sandbag dike around the entire area. For medium floods, this point shifts to LT7, since the increased water levels require a higher dike crest and therefore, longer implementation times. Finally, for big floods, there is neither sufficient time nor budget to protect the entire area, when action is triggered at the LT of our forecast range (LT1-LT14). There is sufficient time to do so from LT15 backwards. However, B is still insufficient.

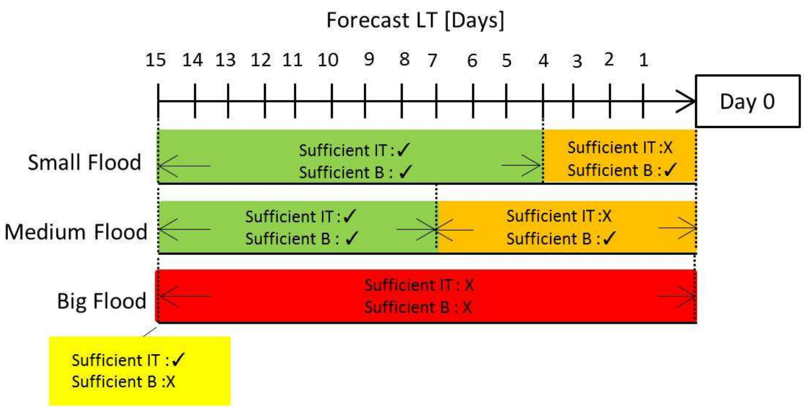

Figure 7 Qualitative demonstration of the EWEAS's ability to protect the entire study area A as a function of LT and flood magnitude, given the restrictions on the budget (B) and action implementation time requirements (IT). The time intervals in colour exhibit whether there is sufficient B and IT to protect the entire area; in green, both B and IT_1 are sufficient, in orange only B is sufficient, in yellow only IT is sufficient and in red neither B nor IT are sufficient.

As we discussed in section 3, the damage reduction is only proportional to the percentage of the total area that is surrounded by the sandbag dike ring. This percentage is listed in Figure 8 at each LT that action is triggered for each flood magnitude (blue line-small flood, red line-medium flood and green line-big flood), which determines the height of the sandbag dike and consequently, the number of sandbags needed. As qualitatively presented in Figure 7, full protection is achieved when actions are triggered at LTs longer than LT4, and LT7 for small and medium floods, respectively, while for big floods the maximum protection percentage is 30% from LT8 onwards.

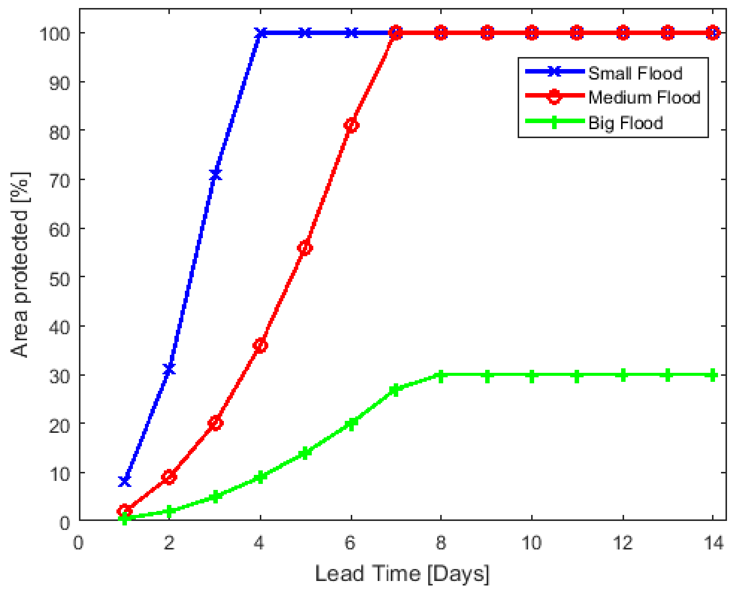

Figure 8 Percentage of the area protected as a function of the triggering action at each LT for the three flood magnitudes (small flood: blue line, medium flood: red line and big flood: green line).

Figure 9 presents the V_{ew} as a function of the LT at which action is triggered for different probability thresholds and flood magnitudes. In small floods, an optimum V_{ew} is reached at LT4 to LT5. At these LTs, the full protection of the area is feasible in terms of time limitations; the budgets are sufficient and the forecast skill is better than that of longer ones, in the sense that the CH number decreases over time and number of FA usually either remains the same or increases. In few cases at longer LTs, we observe that the FA number is lower. Nevertheless, the high MS level keeps the V_{ew} relatively low. In addition, at shorter LTs, the V_{ew} is identical for all the probability thresholds. As already discussed in

4.1, this can be attributed to the model's tendency to yield high probabilities for this discharge threshold at these LTs in this river cell.

Medium floods demonstrate an optimum value at LT7, when using a threshold probability of 60%. The sudden drop of V_{ew} at LT11 using 30% and 60% probability thresholds can be attributed to the erratic forecast skills at this LT, as a result of the small dataset. Similarly, the forecast value is higher at LT12 than at LT9 to LT11 when using the 60% probability threshold, which is a result of non-monotonous trends of MS, CH and FA over time and their resulting costs. At the long LTs, we observe that the V_{ew} is slightly higher when using the 30% threshold compared to the others. Despite the already described limitations of the forecast dataset, this is an indication that the optimal triggering action probability threshold can differ from LT to LT. A low forecast threshold at longer LTs may result in more FA; however, when action is correctly triggered, it can provide the additional time needed for the extra protection of the area, outweighing the unnecessary costs of acting in vain. Hence, since the action triggering is a repetitive dilemma faced by the decision-maker (Figure 2), the selection of the optimal probability thresholds should be carefully selected at each decision time point.

Finally, the low V_{ew} for big floods, often below 0, demonstrate that the EWEAS does not provide any added value on the long-term, despite the fact that the forecast skill in the shorter lead times is high (e.g. LT1). The highest V_{ew} for big floods of our EWEAS is achieved at LT10, using a 90% threshold probability, but is still quite low compared to the other flood magnitudes. The main reasons are that a miss by the forecast leads to extremely high economic consequences and that the measures that are within our set of options, given the available budget and placement productivity rate, cannot provide effective protection.

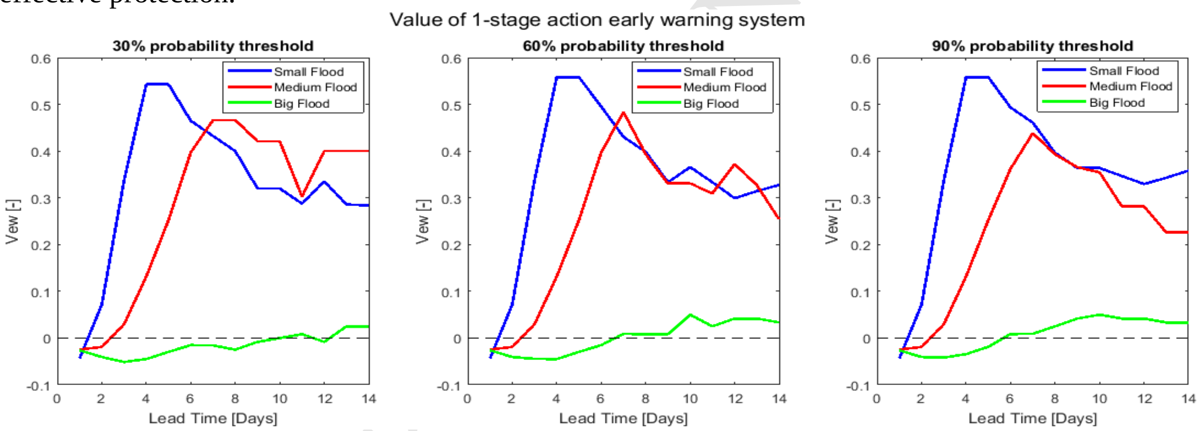

Figure 9 Value of the EWEAS (V_{ew}) for triggering action at each LT, using the 30% (left), 60% (middle) and 90% (right) probability thresholds, for flood events of different magnitude (small flood-blue line, medium flood-red line, big flood-green line).

4.2.1 Sensitivity analysis of one-stage action

The evaluation of the EWEAS involves numerous parameters that interrelate with each other and affect the overall outcome. A sensitivity analysis was performed to highlight the role of the two major boundary conditions for the application of the EWEAS: the available budget (B) and placement productivity (PP). Results of this analysis are shown in Figure 10. We use three combinations: a) restricted B and unlimited PP (i.e. infinite sandbags can be placed in one day; solid lines), b) unlimited B and restricted PP (dashed lines) and c) unlimited B and unlimited PP (dotted lines).

When B is restricted and PP unlimited, the relative economic value V_{ew} of all flood magnitudes reaches the highest value at LT1, where the forecast skill is highest while decreasing at longer LTs. At LT1, V_{ew} for medium flood exceeds that of small floods, while for big floods it is the lowest. This order varies when taking action at other LTs, reflecting that V_{ew} is not always linearly related to the flood magnitude or LT. This variation illustrates the difficulties that decision-makers face when, given the limited budget they have at their disposal during a finite time period, they have to choose when and at which flood magnitude they will initiate action (e.g., a small and frequent flood, but with relatively low potential damage and relatively inexpensive measures; or a big and rare flood with high potential damage and expensive measures).

When B is unlimited and PP is restricted, the lowest relative economic value V_{ew} for all flood magnitudes is at LT1. This indicates that even an excellent forecast skill and a sufficient budget are not enough for EWEAS to provide added value, since an increase in V_{ew} is also dependent on the temporal parameters (i.e. available time, implementation requirements and the coping capacity PP of the system). For small and medium floods, the V_{ew} increases up to the point that it meets the line representing restricted PP and unlimited B. After this point, the dashed and solid lines coincide, demonstrating that the added value of the system is subject only to the forecast skill. On the contrary, in big floods, the V_{ew} keeps increasing until LT14, indicating that a larger budget would provide extra value if action is taken at long LTs, even with poor forecast skill (four correct hits, eight misses), since not taking action has large economic consequences.

Finally, when both B and PP are unlimited, the highest values are found at LT1, decreasing over longer LTs. The small and medium flood actions are insensitive to budget increases. Therefore, an increase in V_{ew} at short LTs (LT4 and LT7 respectively) can result from a PP increase or forecast skill improvement, while at longer LTs, V_{ew} is only dependent on the forecast skill. For this reason, at these flood magnitudes, the three lines coincide. Contrastingly, for big floods, any increase in B or PP positively affects the relative economic value of the system.

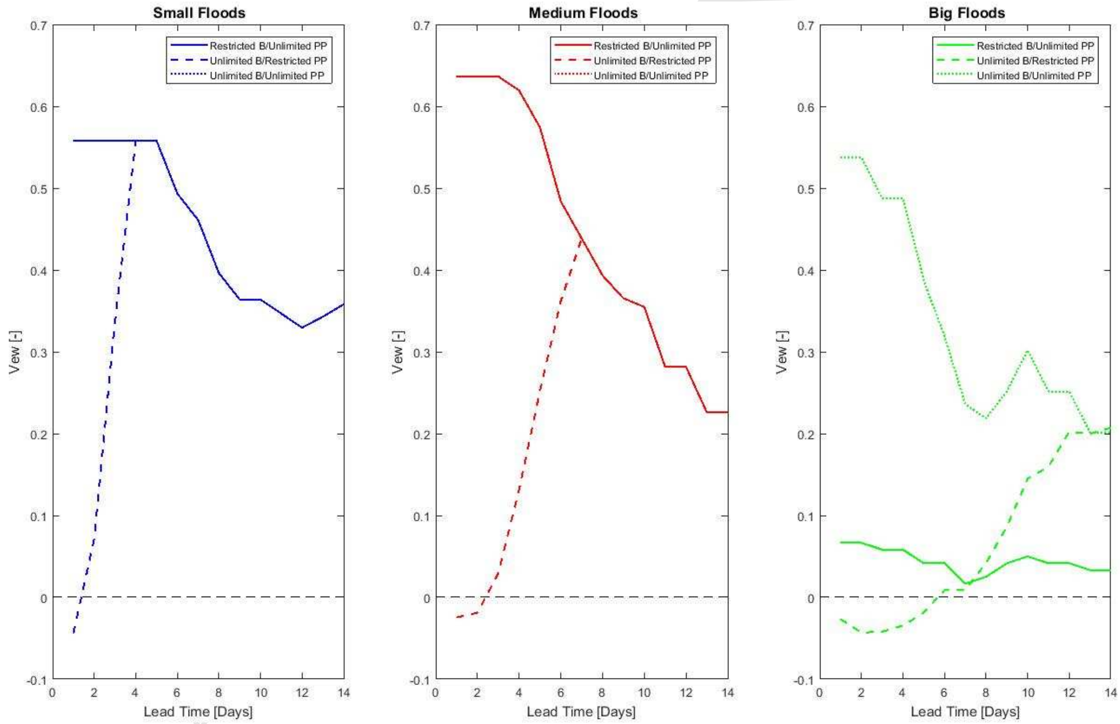

Figure 10 V_{ew} as a function of LT for small (left panel), medium (middle panel) and big floods (right panel) under a 90% probability threshold as trigger for action, when a) the budget B is restricted and placement productivity PP is unlimited (solid lines), b) B is unlimited and PP restricted (dashed lines) and c) both B and PP are unlimited (dotted lines). For small- and medium-size floods, an unlimited B and PP (dotted lines) overlap with a restricted B and an unlimited PP (solid lines) at LTs shorter than LT4 and LT7 respectively, whereas all lines coincide at longer LTs.

4.3 Added value of EWEAS in two-stage approach

In a two-stage decision-making system, the event-based metrics (CH, MS and FA) of the two triggering action LTs are jointly calculated (see Table 4). This is likely to lead to different optimal probability thresholds that trigger the two actions (i.e. there are three thresholds for early and three thresholds for late action, which results in nine combinations). In Figure 11, we demonstrate the lowest and the highest relative economic values V_{ew} from this set of thresholds (solid lines), together with V_{ew} for the one-stage action (dashed lines) of a 90% probability threshold for each of the three flood magnitudes at each LT. Although decision-makers are interested in the highest V_{ew}, we also include the lowest V_{ew} to indicate that sometimes even the worst combination of the two-stage approach is better than the optimal value of the one-stage approach. This is observed mainly at the short LT of small and medium floods, where the forecast tends to yield high probabilities and therefore, the low and the high thresholds produce identical results. In addition, at these LTs, an increase in V_{ew} is predominantly affected by an increase in placement productivity PP that is provided by the preliminary action, indicating that the preliminary action does provide added value.

The difference between the minimum and the maximum values of the two-stage approach increases over time, reflecting the variations in forecast skill and demonstrating the need for the careful selection of the optimal thresholds at each LT that action is taken.

In small floods, the highest V_{ew} of the two-stage approach exceeds that of the one-stage approach for all LTs, while the optimal LT to trigger action remains unchanged (LT4 and LT5), mainly indicating that the preliminary action leads to lower implementation costs for the same protection level. In medium floods, the maximum V_{ew} in the two-stage approach is always higher, and the minimum V_{ew} is lower than that of the one-stage approach for all LTs from LT7 onwards. In this case, the optimal V_{ew} is shifted by one day (LT6, instead of LT7), compared to the one-stage approach, demonstrating that the decision-maker is able to postpone the decision and wait for new forecast information. This delay generates a higher relative economic value, since the preliminary action provides the extra time needed for procuring a more accurate forecast and maintaining the same safety level. For big floods, for which the existing budget and time constraints make the protection of the entire area unfeasible, the optimal time point to trigger the main action is at LT10 for the two-stage approach. This is consistently more cost-effective than the one-stage approach, indicating that having the possibility to trigger preliminary action is a risk-free option, since this engenders lower construction costs (hence, more available funds) and higher placement productivity (hence, lower implementation time). However, in these events V_{ew} is still much lower than in the other two scenarios, demonstrating that, in practice, a reduction in the number of misses at long LT that is accompanied with a budget increase is needed to achieve higher EWEAS performance. Table S2 (supplementary material) outlines the combinations of probability thresholds that produce the minimum and maximum V_{ew} for all LTs and flood magnitudes.

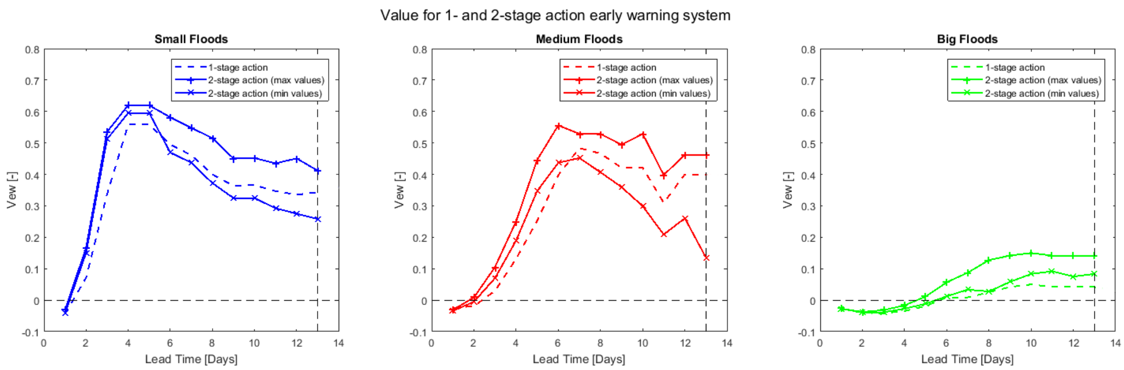

Figure 11 Minimum and maximum V_{ew} derived from the different combinations of forecast probability thresholds for the two-stage action approach (solid lines) compared to the one-stage action (dashed lines) for small- (blue lines), medium- (red lines) and big-magnitude floods (green lines). Vertical dashed line and right boundary shows the time period during which preliminary action is carried out.

5. Discussion and Recommendations

Assessing the performance and the accuracy of a hydrological model is a challenge globally (Veldkamp et al., 2018), and particularly in developing countries, where observations for calibration or evaluation of these models are sparse. In many of these countries, global models are often used as a primary source of information (McNulty et al., 2016) to trigger humanitarian action (Coughlan de Perez et al., 2016), in spite of a lack of consistently good performance and high resolution forecasts. Usually, the assessment of the quality of a forecast model for a given river basin is carried out by comparing its output for each section to the observed discharge (e.g. Bartholmes et al., 2008). However, the short period for which forecasts were available in our study (approximately 8 years) and the rare nature of flood events hamper a thorough forecast skill assessment. This is the reason that we used relatively low discharge thresholds. Alternative ways to allow a statistically robust assessment would be to pool together observed flood events in large regions. For instance, Thiemig et al. (2015) calculated the skill metrics of the African flood forecasting system for entire Africa and Bischiniotis et al. (2019) computed the skill of GloFAS in Peru. However, both forecast skill and risk mitigation actions are highly location-dependent which restricts the use of large spatial aggregates of the forecasting systems. Therefore, we chose to focus on one location, using relatively low percentiles from the modelled discharge as flood proxies. Forecast with longer time series is a prerequisite for a more thorough evaluation that will lead to more accurate results.

The evaluation of the operational forecast system skill is different than its evaluation from a hydrological point of view. For this reason, we incorporated operational characteristics such as the lifetime of the forecast-based actions in the skill assessment, which is particularly relevant for end-users of the humanitarian sector (Coughlan de Perez et al. 2016). The actions' lifetime duration has an impact on the skill assessment and consequently on the overall benefits of the EWEAS; for example, a hypothetical measure with short implementation time and very long lifetime (e.g. 2 year) would lead to a lower number of event-based metrics, while a measure with a very short lifetime (e.g. 1 days) would require higher accuracy regarding the onset time of the event and would lead to higher number of event-based metrics.

In our study area, we observed that the model tends to forecast high discharges using high probabilities, which was also noted by Coughlan de Perez et al. (2016) in 2 similar river cells in Magoro and Kapelebyong, Uganda. This led to similar results among the three triggering action probability thresholds used. To improve forecast skill, various bias-correction methods exist (e.g. Atger, 1999; Eckel and Walters, 1998; Krzysztofowicz, 1992; Krzysztofowicz and Long, 1990). Post-processing GloFAS output instead of using raw forecasts may have affected our results (e.g., Wilks, 2001), but the overall concept of our methodology is not critically dependent on these bias-adjustments. However, such post-processing is recommended to the end users of this model for this area, before triggering flood risk mitigation actions.

Changes in discharge at rivers with high water volumes, like the one used in this research, occur at slow rates (Alfieri et al., 2013). Therefore, it is expected that hydrological forecasts will not differ substantially between lead times that are only a few days apart. This makes the application of multi-stage actions that are based on hydrological forecasts more likely, in contrast to decision-making systems that solely use forecasts with lower autocorrelation, such as precipitation forecasts, to trigger action. Hence, following the assessment of the 2-stage decision-making system that was illustrated in this research, end users should work with forecasters to explore where and which forecasts to use so as the 'ready-set-go' approach is worthy.

To facilitate the understanding of our concept, we used as an example of forecast-based action that mitigates flood damage by the placement of sandbags around the study area. We acknowledge that this action may not be the most suitable measure for every study area, but it acts as a measure metaphor with dynamic effectivity, implementation time and cost/benefit ratio. A thorough analysis that meets the local needs, characteristics and physical boundary conditions must precede the selection of forecast-based actions. For example, we assumed that the water levels will not exceed a level for which sandbags cannot provide protection. Higher water levels would require other types of measures to mitigate flood risk (e.g. removable flood barriers). Also, we assumed that the sandbag dike ring will be uniform, which in reality will depend on local characteristics and flow conditions. Finally, we assumed that the sandbags are prepositioned in the study location and that therefore no transportation time and costs is required. In case sandbag transportation was considered the preliminary action that was triggered by an earlier forecast, then this action would be a prerequisite for the implementation of the main action and Eq. 4 would be substituted by Eq.S3 (supplementary). Hence, before implementing a 'Ready-Set-Go' approach, the interrelationships between the actions should be quantified. Although the incorporation of these details is very important for practical applications, we consider that the simplifications made allow us to demonstrate in a more clear way the paper's scope.

We distinguished between three flood event magnitudes, intending to show how these affect our system, considering that as soon as a flood threshold is exceeded, damage will be deterministic. In reality, this will not be the case, since damage will depend on the inundation level and therefore water level/damage curves are needed. The distinction between different flood levels can raise several questions to a practitioner. For example, at the time that a big flood is forecasted by the model, the area could possibly already experience a small flood. Identifying the optimal way to act and the actions that can be adapted is a major challenge for end-users. These are required to give answers to the questions on whether it is worthier to start building a short sandbag dike that can later turn into a higher one, build a very high one as soon as the first forecast is issued, or is it worthier to take action against small and frequent floods rather than big and rare ones, given the budget restrictions. This illustrates the large number of degrees of freedom in the real world's decision context, and can be studied in future research.

Another source of uncertainty in the evaluation of the EWEAS is the paucity of data regarding the costs and benefits of forecast-based mitigation actions. In our study, we only considered simplified, tangible costs of the mitigation actions. In operational flood risk management, however, other intangible costs can strongly affect the EWEAS value. For instance, a system may lose its credibility when action is taken in vain due to frequent false alarms, leading to reduced responses for future alerts (LeClerc and Joslyn, 2015), a phenomenon known as the 'crying wolf effect' (Breznitz, S., 1984). Although other tangible costs can be easily added into our evaluation system, the quantification of intangible costs is complex, and to the best of our knowledge no extensive record exists.

Similarly, in our example we have used simple representations of the early action benefits. In reality, multiple sets of measures with different targets and levels of suitability are at decision-makers' disposal for each occasion. For example, evacuation prevents the loss of lives, chlorine tablets prevent the spread of diseases, training raises public awareness, and temporary flood barriers protect critical infrastructure. All these have different characteristics and for a complete evaluation of the benefits of EWEAS the entire range of actions should be considered (Pappenberger et al., 2015). Furthermore, different actors have different goals (e.g. maximize the number of prevented events or minimise the total expected losses) and thus, there is not a truly objective measure of the EWEAS benefit. In the humanitarian sector, for instance, maximising prevention is usually more appropriate for decision-makers with fixed budgets in specific locations, while minimising cost is more suitable for decision-makers who aim to reach larger geographical areas (Lopez et al., 2018). Finally, preliminary actions that can be considered 'no-regret' options, owing to negligible costs or because they provide a risk-free benefit, are usually carried out to facilitate other actions, without a directly quantifiable benefit. Aggregating and estimating the overall effectiveness of these measures is complex, and thus a comparison of flood damage between an event with ex-ante risk mitigation measures and an event for which no measures are taken is not easily made. Further research and operational data on the effectiveness of these measures would be highly valuable. More elaborated cost/benefit analysis would provide more insights on the EWEAS evaluation and may alter the optimal time point to trigger action, but the elementary trade-off between rapid action and waiting for higher quality forecasts will remain present under all circumstances.

6. Conclusions

In this study, we adapted existing approaches to present a methodology that assesses the added value of early warning early action systems (EWEAS) in flood risk mitigation, when action can be taken at different time points. In doing so, we used a configuration of an EWEAS, taking into account forecast uncertainty, limited budgets, constraints on actions' implementation time, and time-varying costs, damage and benefits. We used forecasts from a global flood forecast model (GloFAS) in Akokoro, Uganda and the lifetime of the forecast-based actions to evaluate the forecast skill from operational point of view and we explored two scenarios of taking action; a) at one point in time (one-stage action) b) at two points in time (two-stage action), where initially a preliminary action, based on a lower skill and longer lead time forecast, and subsequently, a main action, triggered by a shorter-term and higher confidence forecast, are taken. Using an idealized case study we showed that a two-stage system can provide added value to the overall effectiveness of EWEAS; in small floods, the preliminary action actually helps by decreasing the costs of the main action. in medium floods it allows the decision-makers to postpone the decision to take action while waiting for a higher quality forecast. In big floods, where the available budget and time requirements are not sufficient for the protection of the entire study area, the preliminary action always leads to a higher economic value than when taking only the main action. This shows that low-certainty and long lead time forecasts can be useful when paired with high-certainty and short lead time information. Finally, we demonstrated that even if the forecast skill is high, the relative economic value of EWEAS can be small or non-existent, which is subject to the capability to act upon a forecast. This shows that the preparation time needed for the forecast-based actions should not be neglected when early action protocols are formed, as the optimal lead time to trigger action is a function of forecast quality and operational characteristics of the forecast-based actions. Therefore, investments should focus on both extending the forecast range and accuracy and increasing adaptation capabilities, either by providing sufficiently large budgets for effective measures or by reducing their implementation time. Otherwise, even an excellent forecast system will have a limited benefit.

Acknowledgements: We thank the Copernicus Emergency Management Service (CEMS) and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the GloFAS forecasts. The project was funded by NWO-VICI grant nr. 453-13-006, and NWO New Delta grant nr. 869.15.001.

7. References

1. Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J. and Pappenberger, F.: GloFAS-global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17(3), 1161–1175, doi:10.5194/hess-17-1161-2013, 2013.
2. Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D. and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, J. Hydrol., 517, 913–922, doi:10.1016/j.jhydrol.2014.06.035, 2014.
3. Atger, F.: The Skill of Ensemble Prediction Systems, Mon. Weather Rev., 127(9), 1941–1953, doi:10.1175/1520-0493(1999)127<1941:TSOEPS>2.0.CO;2, 1999.
4. Bartholmes, J. C., Thielen, J., Ramos, M. H. and Gentilini, S. , 2008. The European Flood Alert System EFAS; Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrology and Earth System Sciences Discussions, 5, 289–322, doi:10.5194/hessd-5-289-2008.
5. Bischiniotis, K., Van Den Hurk, B., Coughlan De Perez, E., Zsoter E., Grillakis, M., and Aerts, J.: Evaluation of a global ensemble flood prediction system in Peru, *submitted in Hydrological Sciences Journal*, 2019
6. Breznitz, S.:. Cry Wolf: The psychology of false alarms. New York: Psychology Press, 1984
7. Buizza, R.: The value of probabilistic prediction, Atmos. Sci. Lett., 9(2), 36–42, doi:10.1002/asl.170, 2008.
8. Carsell, K. M., Pingel, N. D. and Ford, D. T.: Quantifying the Benefit of a Flood Warning System, Nat. Hazards Rev., 5(3), 131–140, doi:10.1061/(ASCE)1527-6988(2004)5:3(131), 2004.
9. Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375(3–4), 613–626, doi:10.1016/j.jhydrol.2009.06.005, 2009.
10. 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(4), 895–904, doi:10.5194/nhess-15-895-2015, 2015.
11. Coughlan de Perez, E., Van Den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., De Suarez, J. M., 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(9), 3549–3560, doi:10.5194/hess-20-3549-2016, 2016.
12. Eckel, F. A. and Walters, M. K.: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble, Weather Forecast., 13(4), 1132–1147, doi:Doi 10.1175/1520-0434(1998)013<1132:Cpqpfb>2.0.Co;2, 1998.

13. Goddard, L., Baethgen, W. E., Bhojwani, H. and Robertson, A. W.: The International Research Institute for Climate & Society: why, what and how, Earth Perspect., 1(1), 10, doi:10.1186/2194-6434-1-10, 2014.
14. Golding, B. W.: Long lead time flood warnings: Reality or fantasy?, Meteorol. Appl., 16(1), 3–12, doi:10.1002/met.123, 2009.
15. Golnaraghi M.: Institutional partnerships in multi-hazard early warning systems: A compilation of seven national good practices and guiding principles Springer, New York, 2012.
16. Heltberg, R., Siegel, P. B. and Jorgensen, S. L.: Addressing human vulnerability to climate change: Toward a "no-regrets" approach, Glob. Environ. Chang., 19(1), 89–99, doi:10.1016/j.gloenvcha.2008.11.003, 2009.
17. Hogan, R. J. and Mason, I. B.: 3 Deterministic forecasts of binary events, edited by: Jolliffe, I. T. and Stephenson, D. B., available at: https://eprint.iacr.org/2015/1164.pdf (last access: 31 August 2016), 2012.
18. Jaun, S., Ahrens, B., Walser, A., Ewen, T. and Schär, C.: A probabilistic view on the August 2005 floods in the upper Rhine catchment, Nat. Hazards Earth Syst. Sci., 8(2), 281–291, doi:10.5194/nhess-8-281-2008, 2008.
19. Kabat, P., van Vierssen, W., Veraart, J., Vellinga, P. and Aerts, J.: Climate proofing the Netherlands., Nature, 438(7066), 283–284, doi:10.1038/438283a, 2005.
20. Katz, R. W. and Murphy, A. H.: Economic value of weather and climate forecasts., 1997.
21. Kelman, I. and Spence, R.:. A limit analysis of unreinforced masonry failing under flood water pressures. Masonry International, 16 (2): 51-61, 2003.
22. Krzysztofowicz, R. and Long, D.: Fusion of Detection Probabilities and Comparison of Multisensor Systems, IEEE Trans. Syst. Man Cybern., 20(3), 665–677, doi:10.1109/21.57281, 1990.
23. Krzysztofowicz, R.:. Bayesian correlation score: a utilitarian measure of forecast skill. Mon. Wea. Rev., 120:208–219, 1992.
24. Krzysztofowicz, R.: The case for probabilistic forecasting in hydrology, J. Hydrol., 249(1–4), 2–9, doi:10.1016/S0022-1694(01)00420-6, 2001.
25. LeClerc, J. and Joslyn, S.: The cry wolf effect and weather-related decision making, Risk Anal., 35(3), 385–395, doi:10.1111/risa.12336, 2015.
26. Lopez, A., Coughlan de Perez, E., Bazo, J., Suarez, P., van den Hurk, B. and van Aalst, M.: Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings, Weather Clim. Extrem., doi:10.1016/j.wace.2018.03.006, 2018.
27. Matte, S., Boucher, M. A., Boucher, V. and Fortier Filion, T. C.: Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-Averse decision maker, Hydrol.

Earth Syst. Sci., 21(6), 2967–2986, doi:10.5194/hess-21-2967-2017, 2017.

28. McNulty, S., Cohen Mack, E., Sun, G., Caldwell, P: Hydrologic modeling for water resource assessment in a developing country: the Rwanda case study , Lachassagne, P. and M. Lafforgue (eds). Forest and the Water Cycle: Quantity, Quality, Management. Cambridge Scholars Publishing, 2016.
29. Murphy, A. H.: The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation, Mon. Weather Rev., 105(7), 803–816, doi:10.1175/1520-0493(1977)105<0803:TVOCCA>2.0.CO;2, 1977.
30. Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S. and Thielen, J.: The monetary benefit of early flood warnings in Europe, Environ. Sci. Policy, 51, 278–291, doi:10.1016/j.envsci.2015.04.016, 2015.
31. Rawls, C. G. and Turnquist, M. A.: Pre-positioning of emergency supplies for disaster response, Transp. Res. Part B Methodol., 44(4), 521–534, doi:10.1016/j.trb.2009.08.003, 2010.
32. Richardson, D. S.: Skill and relative economic value of the ECMWF ensemble prediction system, Q. J. R. Meteorol. Soc., 126(563), 649–667, doi:10.1002/qj.49712656313, 2000.
33. Roulin, E.: Skill and relative economic value of medium-range hydrological ensemble predictions, Hydrol. Earth Syst. Sci., 11(2), 725–737, doi:10.5194/hess-11-725-2007, 2007.
34. Thiemig, V., Pappenberger, F., Thielen, J., Gadain, H., de Roo, A., Bodis, K., Del Medico, M. and Muthusi, F.: Ensemble flood forecasting in Africa: A feasibility study in the Juba-Shabelle river basin, Atmos. Sci. Lett., 11(2), 123–131, doi:10.1002/asl.266, 2010.
35. Thiemig, V., Bisselink, B., Pappenberger, F. and Thielen, J.: A pan-African medium-range ensemble flood forecast system, Hydrol. Earth Syst. Sci., 19(8), 3365–3385, doi:10.5194/hess-19-3365-2015, 2015. UNICEF (2015). UNICEF/WFP Return on Investment for Emergency Preparedness Study. January 2015
36. UNDP (United Nations Development Program). A 'No-Regrets' Risk-Based Approach to Climate-Proofing of Public Infrastructure: Improved National and Sub-National Planning for Resilience and Sustainable Growth. July 2010.
37. UNICEF: Return on Investment for Emergency Preparedness Study. January 2015
38. UNISDR: Guidelines for Reducing Flood Losses, United Nations International Strategy for Disaster Reduction, DRR7639. UNISDR. http://unisdr.org/we/inform/publications/558, 2004.
39. Velázquez, J. A., Anctil, F. and Perrin, C.: Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments, Hydrol. Earth Syst. Sci., 14(11), 2303–2317, doi:10.5194/hess-14-2303-2010, 2010.
40. Veldkamp, T. I. E., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N.,

Zaherpour, J. and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: A multi-model validation study, Environ. Res. Lett., 13(5), doi:10.1088/1748-9326/aab96f, 2018.

41. Verkade, J. S. and Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning, Hydrol. Earth Syst. Sci., 15, 3751-3765, https://doi.org/10.5194/hess-15-3751-2011, 2011.
42. Wetterhall, F., Pappenberger, F., Alfieri, L., Cloke, H. L., Thielen-Del Pozo, J., Balabanova, S., Daňhelka, J., Vogelbacher, A., Salamon, P., Carrasco, I., Cabrera-Tordera, A. J., Corzo-Toscano, M., Garcia-Padilla, M., Garcia-Sanchez, R. J., Ardilouze, C., Jurela, S., Terek, B., Csik, A., Casey, J., Stankunavičius, G., Ceres, V., Sprokkereef, E., Stam, J., Anghel, E., Vladikovic, D., Alionte Eklund, C., Hjerdt, N., Djerv, H., Holmberg, F., Nilsson, J., Nyström, K., Sušnik, M., Hazlinger, M. and Holubecka, M.: HESS Opinions "forecaster priorities for improving probabilistic flood forecasts," Hydrol. Earth Syst. Sci., 17(11), 4389–4399, doi:10.5194/hess-17-4389-2013, 2013.
43. Whittaker, J., McLennan, B., Handmer, J.: A review of informal volunteerism in emergencies and disasters: Definition, opportunities and challenges, International Journal of Disaster Risk Reduction, 13, pp. 358-368, 2015
44. Wilks, D. S.: A skill score based on economic value for probability forecasts, Meteorol. Appl., 8(2), 209–219, doi:10.1017/S1350482701002092, 2001.
45. Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, Atmos. Res., doi:10.1016/j.atmosres.2010.12.005, 2011.