Response letter of hess-2021-497

Dear Anonymous Referee #1,

Please find the responses to the comments.

Comments made by the reviewer were highly insightful. They allowed us to greatly improve the quality of the manuscript. We described the response to the comments.

Each comment made by the reviewers is written in *italic* font. We numbered each comment as (n.m) in which n is the reviewer number and m is the comment number. In the revised manuscript, changes are highlighted in yellow.

We trust that the revisions and responses are sufficient for our manuscript to be published in *Hydrology and Earth System Sciences*

Sincerely
Yohei Sawada, Rin Kanai, Hitomu Kotani

Responses to the comments of Referee #1

This paper presents the improvement of an existing socio-hydrological model on the interactions between flood forecasting and flood loss by including social collective trust. The manuscript uses the model to investigate the cry wolf effect (where individuals may be less likely to implement protective measures if they have experienced false alarms). I believe including trust and investigating its role is a relevant contribution to the socio-hydrological literature. The manuscript shows an interesting analysis of the (potential) role of social collective trust and its implications for early warning systems.

(1.1) However, I believe a major limitation of the work is the lack of comparison between model results and data or empirical evidence. I appreciate that there may not be enough data available to actually compare the model results to data, but given this limitation I believe the model equations and parameter choices should be much better substantiated with evidence from the literature. In addition, one could, in a descriptive way, compare the results with findings in the literature related to the cry wolf effect rather than only compare the results to the results of another model. In the current state, the manuscript does not provide enough evidence for the model assumptions and their relevance. This means that it is impossible to draw any useful conclusions from the results of the analysis, since it is unclear how well the model represents reality.

→ First, our model and findings are qualitatively consistent to empirical evidence found in previous works. This point has been clarified in the revised version of the paper. See also our responses to the comment of Referee #2 (2.11).

> "Our stylized model and findings are consistent to the previous works. In our model, the subjective perceptions of warning system's accuracy controls social collective trust in a weather agency and preparedness actions, which is consistent to Ripberger et al. (2015). Our simulation results reveal that more actual false alarms hamper preparedness actions and induce more damages, which is consistent to the findings of Simmons and Sutter (2009) and Trainor et al. (2015). The behavior of the optimal warning threshold is similar to Roulston and Smith (2003)."

Second, the comparison between our SKK model and the GL model in Figure 2 actually shows that our SKK model is more consistent to the published literature at least qualitatively. Figure 2 indicates that in the original GL model, it is necessary to reduce the number of false alarms to minimize the total loss only when the cost of mitigation and protection actions responding to issued warning (C in equation 6) is large. On the other hand, the reduction of false alarms is always necessary to minimize the loss in the SKK model. We found that previous works revealed that this cost is negligibly small compared with the total loss of flood disasters. Based on the fact that the mitigation cost is negligible, and forecasters take care of reducing false alarms, our extension of the GL model improves the consistency of the simulation to reality. This point was indeed unclear in the original version of the paper. We have firstly mentioned that C is negligibly small according to the previous literature in the revised version of the paper.

> "Note that this cost has been found to be negligibly small compared with avoidable damage. For instance, Schroter et al. (2008) showed that the cost C is approximately 2 % of avoidable damage. In previous works, this cost was often neglected (e.g., Pappenberger et al. 2015; Hallegatte 2012). Although Gironz Lopez et al (2017) assumed that there are non-negligible costs of mitigation and protection actions, we will discuss how differently their model and our newly proposed model works with no mitigation costs (i.e. $\eta = 0$) as well as the original settings of Gironz Lopez et al (2017)."
>
> "Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., Thielen, J.: The monetary benefit of early flood warnings in Europe. Environmental Science & Policy, 51, 278-291, https://doi.org/10.1016/j.envsci.2015.04.016, 2015"
>
> "Schroter, K., et al: Effectiveness and efficiency of early warning systems for flash-floods (EWASE). First CRUE ERA-Net Common Call – Effectiveness and efficiency of non-structural flood risk management measures, 132pp. available from www.crue-eranet.net, 2008"
>
> "Hallegatte, S.: A cost effective solution to reduce disaster losses in developing countries Hydro-meteorological services, early warning, and evacuation, The World Bank Policy Research Working Paper, 6058, available from https://openknowledge.worldbank.org/bitstream/handle/10986/9359/WPS6058.pdf?s, 2012"

Then, we have clarified that the SKK model can simulate the behavior of forecasters and the relationship between warning thresholds and total losses more realistically than the GL model in the results section of the revised paper.

> "Note that the costs of mitigation and protection actions with $\eta = 0.1$ in the experiment 2.3 is comparable to the flood damages. As discussed above, this high cost of mitigation and protection actions was not supported by previous works although Girons Lopez et al. (2017) used this parameter."
>
> "Considering that the previous works indicated that the cast of mitigation and protection actions is negligibly small (i.e. it is realistic to assume $\eta = 0$), the SKK model reproduces the relationship between warning thresholds and total losses more realistically than the GL model."

We also briefly mentioned this point in the discussion section of the revised paper.

> "While the GL model realistically simulate the behavior of the optimal warning threshold only when unrealistically high costs of mitigation and protection actions are assumed, our stylized model needs no costs of mitigation and protection actions to realistically simulate the behavior

of the optimal warning threshold. Our stylized model is more consistent to the previous works in which the costs of mitigation and protection actions responding warnings were found to be negligibly small (e.g., Schroter et al. 2008; Hallegatte 2012; Pappenberger et al. 2015)."

Some other remarks:

(1.2) The authors use socio-meteorology in their title and in the final paragraph of the discussion and conclusion they call for a new field called socio-meteorology. However, it is not clear to me why this work is so different that it does not fit within the field of socio-hydrology (the authors are only using discharge and forecasts of discharge in their model, to me this is hydrology, not meteorology). I would suggest to choose a different title, and stick to using socio-hydrology, as the authors do throughout the entire manuscript (the socio-meteorology is in fact only mentioned as an afterthought in the final paragraph of the manuscript).

→ We propose a new title "Impact of cry wolf effects on social preparedness and efficiency of flood early warning systems". We believe this version of the title directly show what we investigated in this paper.

The reviewer mentioned that using discharge and forecasts of discharge is hydrology and is not meteorology. We disagree with this comment. To provide "forecasts" of discharge, weather forecasting is absolutely necessary, which we believe is in the field of meteorology (or it is often called as hydro-meteorology). We essentially intended to investigate the function of "forecasts" of discharge in the dynamics of social preparedness, so that we are now going to the interdisciplinary field which includes hydrology, meteorology, and social sciences.

(1.3) Introduction, lines 43 to 80: after reading the introduction for the first time I had the impression that there was actually no evidence for the cry wolf effect and for a relationship between the false alarm ratio and the implementation of measures. This made me wonder what the relevance of the presented model and manuscript is. However, after re-reading I see that I misinterpreted and there are studies that do find evidence in support of the cry wolf effect, but also some that do not. I would suggest the authors rewrite this part of the introduction to better present the argument for why their study is important.

→ We believe that many previous works found and quantified the cry wolf effects in meteorological disasters. In addition, many forecasters may believe the cry wolf effects when they design early warning systems. Therefore, it is crucially important to consider them. These points were indeed unclear in the original version of the paper, and we have clarified them in the revised version of the paper by performing many modifications in this paragraph. Note that we carefully reviewed previous works which contradicts with each other in the balanced way, which we believe is still appropriate and essentially unchanged in the revised version of the paper.

"Many previous studies have found and quantified the cry wolf effects in meteorological disasters."

"Roulston and Smith (2003) found that the warning threshold of the actual weather warning systems can be justified only if the cry wolf effect is considered. This finding implies that many forecasters believe the existence of the cry wolf effects and the design of early warning systems was substantially affected by how the cry wolf effects are considered."

"It should be noted that while these previous works supported the cry wolf effect as an important factor to be considered for the design of warning systems, some studies discussed the myth of cry wolf effects implying that they do not exist."

"Although Trainor et al. (2015) supported the existence of the cry wolf effects, they also found that there is a wide variation in public definition of false alarms and actual false alarm ratio does not predict perception of false alarm ratio."

" Although the existence of the cry wolf effect is still debatable due mainly to the lack of field data and the ambiguity of the quantification of public perception of false alarms, the current evidence suggests the importance of understanding the effect of false alarms on behavioral responses to warning to design efficient flood early warning systems."

(1.4) In the model description in line 148 (and after) the authors talk about preparedness actions (and mitigation and protection actions), please elaborate and explain what these actions are. There are many preparedness actions that do not depend on a flood warning to be implemented, what about those actions? These kind of measures may actually be implemented when experience of damage is high and trust in flood warning is low (which is the opposite of the cry wolf effect).

→ Here we modelled preparedness actions which were done by responding issued warnings such as evacuation and safekeeping of assets. We fully agree with the reviewer's comment that many preparedness actions are unrelated to early flood warning. These preparedness actions are not included in our model to focus only on the impact of social preparedness on the efficiency of early flood warning. This point was indeed unclear in the original version of the paper. We have clarified this point in the revised version of the paper.

"If a flood event is successfully forecasted and a warning is issued (i.e. $P \geq \pi$), this damage is mitigated by preparedness actions such as evacuation and safekeeping of assets. Note that preparedness actions which are not triggered by FEWS were not considered in this stylized model to focus only on the impact of social preparedness on the efficiency of FEWS"

(1.5) Equation 6 models the cost of mitigation and protection actions, why is this relevant? Please discuss why you calculate this. Later, in section 3.1, I see that the total loss is calculated as D + C. I suggest to move this to section 2, since it is quite important and now it is a bit hidden away, which means the importance of C is unclear. Also how are the costs of protection actions determined? What is this based on? Also, why is the loss calculated as D+C, please explain this.

→ Although Table 1, which appears in section 2, shows how to calculate the total loss, this point should have been emphasized when C is introduced. We have clarified this point in the revised version of the paper.

"Whenever a warning is issued, C is included in the total loss."

As we discussed in our responses to the comment (1.1), we think that this cost is essentially unnecessary and should be neglected. We need to describe it just because the original GL model used it and it is important in their model. Please see our responses to the comment (1.1). We have not further modified the paper responding to this comment.

(1.6) In lines 177 to 179, the authors state that it is reasonable to assume that trust in FEWS increases (decreases) when prediction succeeds (fails). Please elaborate, this is the main contribution of the manuscript and this claim should be substantiated more. (The authors reference Wachinger et al. (2013), but Wachiger et al. (2013) actually hypothesise that the cry wolf effect may be an explanation for the risk perception paradox and do not provide the evidence to support this hypothesis.)

→ In the original version of the paper, the previous sentence explain why we decided to choose this simple model.

"Previous studies pointed out that the recent forecast accuracy and false alarm ratio affected the performance of preparedness actions (Simmons and Sutter 2009; Trainor et al. 2015; Ripberger et al. 2015; Jauernic and van den Broeke 2017)."

We fully agree that Wachiger et al. pointed out that the behavior of individual risk perception is more complicated than expected. We did not originally want to cite this paper here. In the revised version of the paper, we have simply deleted this reference. In addition, we clarified that our stylized model is consistent to the published literature related to cry wolf effects in the discussion section of the revised paper. See also our responses to (1.1).

"Our stylized model and findings are consistent to the previous works. In our model, the subjective perceptions of warning system's accuracy controls social collective trust in a weather agency and preparedness actions, which is consistent to Ripberger et al. (2015). Our simulation results reveal that more actual false alarms hamper preparedness actions and induce more damages, which is consistent to the findings of Simmons and Sutter (2009) and Trainor et al. (2015). The behavior of the optimal warning threshold is similar to Roulston and Smith (2003)."

(1.7) In lines 200 to 202 the authors state: "In our proposed model, high social collective trust in FEWS can maintain the high level of social preparedness even if a community completely loses past flood experiences (equation (7))." To me it seems unlikely that preparedness stays high solely based on trust while people have forgotten about floods. Is there any evidence from the literature that supports this assumption?

→ We agree with this reviewer's comment. Here we intended to demonstrate how the new model works using some extreme cases, which we believe was misleading. Theoretically, people take preparedness actions when E goes to 0 but T is high (see equation 7). However, this situation rarely happens in our model. Social collective trust increases when disasters are predicted. In this case, social collective memory also increases because disasters happen. Therefore, E and T are somewhat correlated in many cases, and the combination of zero E and high T may not happen. What we would like to say is that the negative effect of small E can be partially mitigated by high T. We have deleted the description of unrealistic extreme cases and relaxed this sentence in the revised version of the paper.

"If social preparedness is determined only by social collective memory as Girons Lopez et al (2017) proposed, small social collective memory directly results in insufficient social preparedness actions. In our proposed model, high social collective trust in FEWS can induce social preparedness actions even if a community loses past flood experiences to some extent (equation (7))."

(1.8) For all variables and parameters: what are the units?

→ Units were indeed unclear in the original version of the paper. We have clarified them in the tables of the revised paper.

(1.9) For all equations and values the authors choose: please provide more evidence from the literature as to why this is a good representation of reality. This is especially important given the lack of data for comparison with model results, as mentioned in my main point.

→ All equations but equations 7 and 9 come from Girons Lopez et al. (2017). One can find more evidence which supports each equation in Girons Lopez et al. (2017) and references therein. We would like to avoid repeating the detailed explanation in this paper for brevity. Although this point was already mentioned in the original version of the paper, we have slightly modified the sentence to emphasize this point.

> "For brevity, the detailed explanation of equations shared with Girons Lopez et al. (2017) is omitted in this paper. See Gironz Lopez et al. (2017) and references therein for the complete description including the empirical evidence which supports each equation."

The explanation of the other equations has been strengthened by responding to the other comment. Currently no empirical evidence can support the equation 7 with $\gamma = 0.5$ very well although this chose of the equation and parameter is useful to analyze the essential behavior of our proposed model. This specific issue was pointed out by Referee #2. We attached the comment and our responses below.

(2.6) Line 239: Why did you set gamma = 0.5? Why exactly 0.5? What does it mean?

→ We believe that the meaning of $\gamma = 0.5$ was clearly explained in this sentence.

> Lines: In this paper, this original model is hereafter called the GL model. On the other hand, when we set $\gamma = 0.5$ in equation (7), our model considers both social collective memory and social collective trust in FEWS with same weights to calculate social preparedness.

Although this choice of γ is somewhat arbitrary because there is no knowledge about the importance of social collective trust to induce preparedness actions compared to social collective memory. Assuming the same weights give us the most straightforward interpretation of the contributions of both factors to social preparedness and the total loss by floods since we do not have to consider the asymmetric contributions of two factors in equation (7). This point was indeed unclear in the original version of the paper, and we have clarified this point in the revised version of the paper.

> Lines: There is no existing knowledge about the relative importance of social collective memory and social collective trust. Assuming the same weights give us the most straightforward interpretation of the contributions of social collective trust and memory to social preparedness and the total loss by floods since we do not need to consider asymmetric contributions of the two factors in equation (7). Therefore, $\gamma = 0.5$ is appropriate to analyze the essential behavior of our proposed model. This new model with $\gamma = 0.5$ is hereafter called the SKK model.

(1.10) Table 2 and lines 207- 208: why are those parameters fixed and why do they have those values? Are they based on anything?

→ The fixed parameters are not important in our analyses. We simply choose the values which are consistent to the previous work. This point was indeed unclear in the original version of the paper, and we have clarified it in the revised version of the paper.

> "These parameters are not focused on our analysis, and we chose their values from the previous works."

The values of κ_c, θ_c, α_0, and χ are same as Girons Lopez et al. (2017). We have clarified this point in the revised version of the paper.

> "The values of κ_c, θ_c, α_0, and χ are same as Girons Lopez et al. (2017)."

$\mu_m = 0$ means the forecast is unbiased, which was mentioned in the original version of the paper.

> "While Girons Lopez et al. (2017) changes μ_m in their simulation, we set $\mu_m = 0$ assuming the forecast is unbiased."

We have mentioned once again here in the revised version of the paper.

> "We set $\mu_m = 0$ assuming the forecast is unbiased (see also equation 2 and its description)."

Although the value of β was chosen somewhat arbitrary, it was in range specified by the original model (Girons Lopez et al. 2017). Also, the results of Girons Lopez et al. (2017) indicated that this parameter is not very sensitive to relative loss. This point was indeed unclear in the original version of the paper. We have clarified this point in the revised version of the paper.

> "Our specified β is within the range proposed by Girons Lopez et al. (2017). In addition, the results of Girons Lopez et al. (2017) indicated that this parameter is not sensitive to relative loss."

We set λ assuming social collective memory has 25-year half-life, which is within the range of previous works which quantified this half-life by empirical data. This point was indeed unclear in the original version of the paper, and we have clarified it in the revised version of the paper.

> "We set λ assuming that social collective memory has 25-year half-life which is within the range of previously quantified values (e.g., Fanta et al. 2019; Barendrecht et al. 2019)."

(1.11) For the parameters that are varied, why those values?

→ In this study, we did not intend to mimic the real-world phenomena. Our purpose of the numerical experiments is to understand the behavior of our newly proposed stylized model. The effect of changes in parameters on the optimal warning threshold is more important than their values themselves. However, we realized that the strategy of changing parameters to understand the model's behavior has not been clear enough in the original version of the paper. We have addressed this issue in the revised version of the paper. In the experiment 2, the prediction skill was controlled by σ_m, μ_v, and σ_v. We prepared two sets of the parameters for relatively accurate and inaccurate prediction systems. We have explained this point in the revised paper.

"The prediction skill is controlled by σ_m■, μ_v, and σ_v■. The greater values of these parameter provide inaccurate prediction. We prepared two sets of the parameter for relatively accurate and inaccurate prediction system (see Table 4)"

Please see our responses to the comment (1.1) for the discussion of the cost parameter η in the experiment 2. We used $\eta = 0.1$ which was used in the original GL model as well as $\eta = 0$ which we believe is more consistent to the published literature. We have clarified this point in the revised paper.

"Following the settings of Girons Lopez et al. (2017), we set $\eta = 0.1$. In addition, we also performed the numerical simulation with $\eta = 0$ (i.e. negligible costs of mitigation and protection actions) which is more consistent to the published literature than the original settings (see section 2)."

In the experiment 3, we mimic the hypothetical "green" and "technological" societies by changing δ. From the original value in Girons Lopez et al. (2017), we decreased and increased δ to mimic the green and technological societies, respectively. This point has been clarified in the revised paper.

"From the original value of the damage threshold proposed by Girons Lopez et al. (2017) (i.e. $\delta = 0.35$), we decreased and increased δ to simulate the green and technological societies, respectively (see Table 5)."

In the experiment 4, we focused on the responses of our proposed model to the parameters in the dynamics of social collective trust (τ_{TP}, τ_{FN}, and, τ_{FP} in equation (9)). We added a sentence to clarify this point in the revised version of the paper.

"We analyze how the optimal warning threshold is changed by changing τ_{FN} and, τ_{FP} (see Table 6)."

(1.12) Figure 1: what does half of social collective trust and social collective memory mean? Why half?

→ This is because total social preparedness is calculated as 0.5E+0.5T in our settings. If purple and pink lines are summed up, one can obtain black line. This point was mentioned in the original version of the paper.

"In this paper, this original model is hereafter called the GL model. On the other hand, when we set $\gamma = 0.5$ in equation (7), our model considers both social collective memory and social collective trust in FEWS with same weights to calculate social preparedness."

However, we realized that this point needs to be clarified in the caption of Figure 1. We have clarified this point in the revised version of the paper.

" Note that the sum of half of social collective memory and half of social collective trust in FEWS is social preparedness in (b)."

(1.13) In line 289 it is stated that figure 2 shows predefined warning threshold, but the figure axis title is predefined probability threshold. Same for figures 3 and 4.

→ We have fixed this point in the revised version of the paper.