Response letter of hess-2021-497

Dear Anonymous Referee #2,

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 #2

This paper presents an advanced modelling approach to studying the relationship between society and natural extremes. The authors developed an older model by Girons Lopez et al. (2017) by adding the "cry wolf effect" phenomenon. I find this a stimulating extension.

As mathematical modelling is not my field of expertise, I cannot assess the modelling section (which is the core of the paper). Therefore, my review is more focused on the context and overall results.

The paper is well-written and readable. However, some aspects should be explained better (see below). I also pose several questions for the discussion.

Questions and comments:

(2.1) Title: Please consider changing the title to more reflect the paper's topic. E.g. "Possibilities of mathematical modelling in socio-meteorology: flood prediction etc." or "Mathematical modelling of the cry wolf effect".

→ 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.

(2.2) Line 68: You mention that some studies claim that the cry wolf effect does not exist. How do your results affect this debate? Why do you think some authors have found the cry wolf effect problematic? Please discuss these questions.

→ First, field data which are useful to analyze the cry wolf effects are still lacking. It is necessary to collect the wide range of data which have an enough variance of false alarm ratios. Second, although many researchers tried to verify that citizens tend to take preparedness actions such as evacuation when they believe that a false alarm ratio of their warning system is low, the public perception of false alarm ratios is complicated and often contradicts to the actual false alarm ratio. This ambiguity of the public definition of false alarms makes the analysis complicated. These points were indeed unclear in the original version of the paper. We have clarified this point in the revised version of the paper.

"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 to understand the effect of false alarms on behavioral responses to warning to design efficient flood early warning systems."

(2.3) Lines 78-80: I agree with this sentence. However, it would be nice to give a practical recommendation – what does your research imply? How should we consider the social aspects?

→ Although our practical implications are clear, we hesitated to explicitly mention them based on our preliminary modelling work in the original version of the paper. We have clarified them in the discussion and conclusion section of the revised version of the paper.

"Although our model is simple and stylized, we can provide practically useful implication to improve the design of FEWS. First, considering the dynamics of social collective trust in FEWS is more important in the technological society with infrequent flood events than in the green society with frequent flood events. It implies that weather agencies need more efforts to be trusted by general citizens to induce their preparedness actions when a community is protected by flood protection infrastructures such as levees and dams more heavily. Second, as the natural scientific skill to predict flood is improved, the efficiency of FEWS gets more sensitive to the behavior of social collective trust, so that forecasters need to determine their forecasting threshold by considering the social aspects. Considering the recent advances of the skill to predict extreme hydrometeorological events, this finding implies that it is becoming more important for forecasters to take social dynamics responding to weather forecasts into consideration."

(2.4) Line 139: You mention that the "trust in flood warnings" is based on the accuracy of warnings. Are there any other possibilities how to increase public trust? E.g. by social activities, education etc.? I understand that the mathematical model must be simple, but please discuss this.

→ We fully agree with this reviewer's comment. Trust in FEWS may be significantly impacted by social activities and education, which we neglected in our stylized model. This point was indeed unclear in the original version of the paper. We have clarified this point in the revised version of the paper:

"Note that although there are many other factors which affect social collective trust in FEWS such as social activities and education, we did not consider them in our stylized model."

(2.5) Lines 163-164: Please define the terms "social collective memory" and "social collective trust".

→ The definition of them in this work was indeed unclear in the original version of the paper. Social collective memory is the shared knowledge and information about past flood disasters occurred in a community. We have clarified it in the revised version of the paper.

> "Social collective memory is shared knowledge and information about past flood disasters occurred in a community."

The definition of social collective trust is analogous to social collective memory. It is shared knowledge and perception of the reliability of the information issued from authorities. We have clarified it in the revised version of the paper.

> "Social collective trust is defined as shared knowledge and perception of the reliability of the information issued from authorities. "

(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.

> "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.

> "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."

(2.7) Line 257: Please explain the terms "green society" and "technological society".

→ We believe that the definitions are explained right after this sentence. In the revised version of the paper, we have added more details.

> "In green society, risk is dealt with mainly by non-structural measures. In this society, the flood

protection level is so low that many flood events occur, which increases social collective memory of flood events. In technological society, the flood protection level is so high that risk can be dealt with by structural measures as well as non-structural measures. Since flood events occur less frequently in the technological society, the high level of social collective memory cannot be maintained."

(2.8) Tables 2, 3, 4, 5, 6: Please explain how you got the values of the parameters. Are these values based on empirical knowledge or literature review? Or are they just selected arbitrarily? What do these values mean?

→ Referee #1 also pointed out this point. How to specify the fixed parameters shown in Table 2 and how to change the other parameters in Tables 3-6 can be found in the revised version of the paper. We attached the comments of Referee #1 and our responses below, which we believe address this comment (2.8).

(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 $\sigma_m^{\blacksquare}$, μ_v, and $\sigma_v^{\blacksquare}$. 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 the original GL model also used 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).”

(2.9) Figure 1: Why do you show precisely the time range you show? Is it a random selection? What does mean the height of the colour bars? Is it the flood intensity or damage level? Would you please add a description of the y-axis?

→ It is nearly a random selection, but we chose the time range which clearly shows the difference between the SKK and GL models. The purpose of the experiment 1 and Figure 1 is to demonstrate how differently the SKK and GL models work. This purpose of the experiment 1 was mentioned in the original version of the paper.

> “In the experiment 1, the timeseries of state variables of the two models are compared to demonstrate how differently the SKK and GL models work.”

In the revised version of the paper, we have emphasized this purpose when we explained Figure 1.

> “The purpose of Figure 1 is to demonstrate how differently the SKK and GL models work by showing the timeseries.”

The height of the color bars shows total loss by the outcomes (see Table 2), which was mentioned in the original version of the paper. We refer to Table 2 to make it easier to understand this point in the revised version of the paper.

> “Blue, red, and green bars show total loss by the outcomes of false positive, false negative, and true positive, respectively (see Table 2).”

We have added a description of the y-axis in the revised version of the paper.

(2.10) Discussion and conclusion: According to your findings, it is possible to give a practical recommendation to FEWS strategy? I.e. do you suggest issuing fewer warnings (to reduce the cry wolf

effect but risk the damages of flooding) or more warnings (to be safer but risk the cry wolf effect)? Please discuss.

→ In the revised version of the paper, we have explicitly mentioned that our findings are practically useful, and the practical implications have been explained. See our response to the comment (2.3).

We believe that our stylized model has a potential to assess the appropriate warning threshold if it can be accurately localized. 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 by briefly describing:

> "Similar to Roulston and Smith (2003), our stylized model has a potential to help forecasters determine the optimal warning threshold if it can be appropriately calibrated by empirical data."

We avoid directly suggesting issuing fewer warnings or more warnings in this paper. It strongly depends on the local conditions. It is too risky to say it based on our current findings.

(2.11) Discussion and conclusion: Please also discuss your findings in the context of papers on the cry wolf effect you mentioned in the Introduction section.

→ Our stylized model and findings are consistent to the previous works cited in the Introduction section. This point was indeed unclear in the original version of the paper, and we have clarified it in the revised version of the paper. In the original paper, we explained that our stylized model cannot fully support the existing empirical data in one paragraph. We added the paragraph to comprehensively explain the similarity and differences from previous works.

> "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).

(2.12) Final remarks: Your paper is based on the modelling approach only. Would you please suggest how it would be possible to validate your findings on real data?

→ This comment was mostly identical to the comment of Referee #1. Although it is not currently possible to directly compare our results with empirical data, our proposed model is consistent to the published literature at least qualitatively. The response to the comment (1.1) is attached below.

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 evaculation, 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)."

One of the major obstacles for the validation of the model is the lack of data. Although the lack of data was mentioned in the original version of the paper, we did not explicitly say that the lack of data makes it difficult to validate the model. We have clarified this point in the revised version of the paper.

"It should be noted that most of previous studies related to the cry wolf effect focused on tornado disasters and the systematic econometric analyses have not been implemented for flood disasters, which makes it difficult to validate our proposed model."