Reply on RC1:

We thank the reviewer for the insightful comments and detailed suggestions on how to improve the manuscript. We found the comments to be very helpful and have incorporated them into the revised manuscript. In the following, the texts with blue font are the reviewer's original comments, the texts with normal font are authors' responses and the texts with italic font are authors' responses in the revised manuscript. Our detailed responses are as follows:

1) The structure and table of contents of the article need to be adjusted to make the overall presentation of the article clearer. For example, the first part is the introduction, and the second part is the data and methods. It is suggested to list 2.1 data sources, 2.2 methods and processing procedures. The third part is the result analysis. the fourth part is the discussion and the fifth part is the conclusion.

Response: We thank for the reviewer's comments. In the revised manuscript, we have adjusted the structure and table of the contents of the articles as the reviewer suggested. The first part is the introduction. The second is the data and method, 2.1 data sources, 2.2 methods and processing procedures. The third part is the result. The fourth part is the discussion, and the fifth part is the conclusion. Please refer to the structures and the content in the revised manuscript.

2) The conclusion of the article needs to be sorted out again to make it clearer. It is suggested to list 1,2,3 and so on.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have used the 1,2,3 to resort the conclusion. Please refer to lines 4-16, page 36, page 37 and lines 1-5, page 38:

5 Conclusion

The increased frequency of extreme flood events, coupled with interregional trade growth, requires national- and global-scale transportation networks to be more resilient to cope with disruptive events. Evaluation of system-level vulnerability and identification of risk hotspots is a first step to enhance the robustness of the transport system. This study presents a framework for performing system-level vulnerability and risk assessments of a railway system under flooding. The developed framework couples simulated flood events with state-of-the-art network analysis to measure system disruptions caused by floods to identify risk hotspots. The system vulnerability and *risk induced by the flooding are quantified in terms of the performance loss of the Chinese railway system. Results show that failure hotspots, system vulnerability and the risk of the Chinese railway system under floods are highly heterogeneous. The main conclusions are as follows:*

1. *High failure hotspots are mainly distributed in South China, i.e. Yangtze River, Pearl River and Southeast Basins. In addition, floods in the basins in central and eastern China have the highest impacts on the Chinese railway system. Floods in the Yangtze River Basin have the largest impact on the daily cancelled trains. At the same time, floods in the Huaihe and Haihe River Basins cause the largest number of detoured trains as well as associated increased time for the Chinese railway system compared with other basins.*
2. *At a national level, the average percentage of daily affected trains and passengers for the national system is approximately 2.7%. The mean average increased time for detoured trains reaches approximately 5 hours. At the provincial level, the provinces in Central China have the highest risks, estimated to be 4.5% relative to the number of the province's daily trains and more than 3.5% relative to the number of the province's daily passengers. The high risk in terms of the total increased time is mostly distributed in East China, whereas the highest average increased time is distributed in western provinces, such as Xinjiang and Tibet Provinces.*
3. *Using our current approach, the performance loss can be used as the start of the indirect risk assessment from the travel journey perspective. By combining the ticket prices and the operating cost per kilometre, the economic loss for the railway company can be calculated based on the affected trains and associated passengers (Lamb et al., 2019). As a key mode of transport for interregional trade, the failure of railway systems can produce large shocks for industries that depend on the supply that may come from flooded businesses. The risk values per province (such as expected daily cancelled trains) can be used as indicators to link with business disruptions. Future work can try to assess the shocks and indirect economic losses based on the Input and Output table and regional railway transportation performance decreased in our work.*

Reply on RC2:

We apologize for neglecting the comments that the reviewer made on the annotated paper. In the second round revision, these comments are carefully considered and incorporated in the revised paper. In addition, we thank the reviewer's thorough reading of the manuscript and valuable remarks that helped us to improve the manuscript. In the following, the texts with blue font are the reviewer's original comments, the texts with normal font are authors' responses, and the texts with italic red font are authors' responses in the revised manuscript. Our detailed responses are as follows:

1) Try to report only the main results in percentage terms (I would suggest to avoid to discuss results in absolute values in the abstract section)

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have described the main results in percentage terms in the abstract section. We have added the following description in lines 5-7, page 1:

At the national level, the average percentage of daily affected trains and passengers for the national system is approximately 2.7% of the total daily numbers of trips and passengers.

2) try to add some keywords not contained into the title

Response: We thank the reviewer for the suggestion. We have added " Monte Carlo method " and " network analysis" as the keywords. Please refer to lines 10-11, page 1:

KEYWORDS: railway system; flood; risk assessment; system vulnerability; Monte Carlo method; network analysis

3) please, explain what you exactly mean by "direct" economic loss

Response: "Direct economic loss" refers to "the costs for repairing the damaged railway infrastructure". We have added this description in lines 6-7, page 3:

In 2016, the direct economic loss (i.e., the costs for repairing the damaged railway infrastructure) of the Chinese railway system caused by floods was approximately 80 million USD (Editorial Board of China Railway Yearbook, 2001-2017).

4) Informal. I would suggest something like: as a result, consequently, therefore, etc.

Response: We thank the reviewer for the suggestion. We have changed the "such as" word into "as a result", "consequently", "therefore", etc throughout the manuscript to make the sentence more formal.

5) *seasonal variations and feedback dynamics problem*

- *I can image that you have assumed a timetable constant over time, without considering seasonal variations in the frequency of trips, new routes, etc? Is this correct? How realistic is this assumption and how potential seasonal variations can affect the results?*
- *also in this case, you didn't account for any potential seasonal variations? Could this affect the results? Of course your goal is to analyse the average number of affected trains and passengers over the year, but it could be usefull to analyse the typical period of occurrence of the main floods with respect to the seasonal variability of the train trips and number of passengers. Moreover, often the cancellation of one or more trains could imply an increment in the number of passengers in other trains that offer an alternative way to the cancelled trip. Did you account for any possible feedback dynamics on the number of passengers in the case of train cancellation?*
- *Probably the analysis of the results considering the seasonality in flood events occurrence in relation to potential seasonal variations in the timetable (number of trains) and passengers deserves to be explored in future and mentioned here.*

Response: We thank the reviewer for the comments and agree that the performance loss is affected by the seasonal variations of the trip timetable and feedback dynamics problem. Due to lacking timetable and passengers' capacity day by day, it is difficult to consider seasonal variations of the trip timetable and feedback dynamics problem. We have added the following description in the discussion part in the revised manuscript to address the problem. Please refer to lines 13-19, page 34:

Due to lacking timetable and passengers' capacity day by day, we have assumed a timetable constant over time, without considering potential seasonal variations as well any possible feedback dynamics on the number of passengers in the case of train cancellation. Since our goal is to analyse the average number of affected trains and passengers over the year, the assumption is reasonable. In future work, it is worth investigating the typical period of occurrence of the main floods concerning *the seasonal variability of the train trips and the number of passengers.*

6) This figure has been recalled in the manuscript also to show the main basins in which has been divided the China (page 9 - line 11). Please add this information also in the caption.

Response: We thank the reviewer for the suggestion. We have added main basins information in the caption: The spatial distribution of the railway network, average daily numbers of trains and the main river basin in China.

7) The authors are right but I think that they should provide some more details on what they mean by intensity. In some cases, the hazard is defined only on the base of the water depth, while in some other cases it can take into account also the combination of depth and velocity.

Response: We thank the reviewer for the suggestion. In this work, the hazard is defined on the base of the water depth (m). In the revised manuscript, we have added the following description in lines 16-17, page 9:

In this work, the hazard intensity is represented by the water depth (m).

8) Here you should clearly define the aspects related to the design standard under consideration (i.e. water level for cluverts, bridges, embankments, etc,) and the type of threshold selected (i.e. water level) with unit (m).

Response: We thank the reviewer for the suggestion. We have addressed this in Section 2.2.3. Please refer to lines 4-7, page 18:

The flood design standard of the culverts, bridges and embankments of the Chinese national railway system is designed for 100-year water depth, according to the standard for flood control (CRPH, 2016).

9) Much of the information in Section "2.1 National-scale flood event generation" are already reported in Section "2. Data and method". It is for this reason that I would suggest authors to remove this section and move (and integrate) information to section 2.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have removed most information of this part and moved (and integrated) information to section 2. Please refer to lines 6-21, page 10 and lines 6-11, page 11:

Figure 2 presents an overview of the framework used in this study. First, we generate a national- and river basin-scale flood event set. To do this, we use flood hazard maps for different return periods at the national scale, taken from a global flood hazard model (see Section 2.1.1). We then divide these into flood hazard maps for the major river basins and use a curve-fitting method to estimate the flood depth for any return period for any cell. We then apply a Monte Carlo sampling method (Metropolis 1987) to generate the flood events per river basin and aggregate these events to the national scale. Second, we define the railway system as a network using network theory (Newman, 2010). Third, we intersect the flood events with the railway network to identify the disrupted segments in the railway system based on a pre-defined failure threshold. In the last part of our analysis, we assess the system vulnerability and risk in terms of several performance loss metrics, including the daily total number and the total percentage of trains affected (i.e. cancelled or detoured) and involved passengers as well as the total increased time and the average increased time for the detoured trains. We also analyse the parameters sensitivity in the failure threshold and the related risk uncertainty.

2.2.1 National-scale flood event generation

To ensure the estimation is as accurate as possible for an event-based flood risk assessment based on the Monte Carlo sampling, a large number of independent flood events are required (Speight et al., 2017; Wu, 2019; Zhu et al., 2020). In the following subsections, we will describe the procedures to generate flood events, including input flood hazard maps, the function fitting procedure, and the Monte Carlo analysis in more detail.

10) Please, explain this sentence. Is this related to the adopted Monte Carlo sampling method or a more general statement? In this last case, why you should need of a large number of independent flood events to make an accurate flood risk assessment? Maybe, in order to characterise the return time of the flood events, or what else?

Response: The basic idea of the Monte Carlo sampling method is that when the number of simulations is sufficiently large, the frequency of an event approximates the probability of the occurrence of the event. Therefore, a large number of independent flood events are required to ensure the estimation is as accurate as possible for an event-based flood risk assessment based on the Monte Carlo sampling. To make it clearer, In the revised manuscript, we have added the following description in lines 6-8, page 11:

To ensure the estimation is as accurate as possible for an event-based flood risk assessment based on the Monte Carlo sampling, a large number of independent flood events are required (Speight et al., 2017; Wu, 2019; Zhu et al., 2020)

11) I guess that x, y are respectively the horizontal and vertical coordinate of any cell center. Is that correct? Please specify.

Response: It is correct, x, y are respectively the horizontal and vertical coordinate of any cell center. In the revised manuscript, we have added the following description in lines 15-17, page 12:

For each grid cell $g_{x,y}$ *(*x*,* y *are respectively the horizontal and vertical coordinate of grid cell centre), the annual exceedance probability flood depth* D_T *is calculated by Eq. 1:*

12) This is rather unclear to me. Why you need to define a T=1 year event, why you are associating a null corresponding depth (0 m) to a 1-year event? Why is it the same as that of 2-year event? Why did you decide to consider also 1-year return period? I add that talking about T=1 year event does not make much sense to me. It means P = 1 (see eq. 1) and (I think) that's why they impose the 1-year event equal to 0.

Response: (1) For the question:" Why you need to define a T=1 year event? Why did you decide to consider also a 1-year return period? I add that talking about T=1 year event does not make much sense to me. It means P = 1 (see eq. 1) and (I think) that's why they impose the 1-year event equal to 0." It is considering that the inundation depth-exceedance probability is from 0 to 1, when T=2 year event, the P=1/2; and when T=5, the P=1/5, when we impose the 1-year event, the P=1, which can make the inundation depth-exceedance probability is from 0 to 1. (2) For the question: "why you are associating a null corresponding depth (0 m) to a 1-year event? Why is it the same as that of 2-year event? why they impose the 1-year event equal to 0." It is because the depth of the 2-year event in GLOFRIS global fluvial flood hazard maps is equal to 0 m. So we assume that D_1 is equal to zero (i.e., 1-year event with a flood depth of 0 m). In the revised manuscript, we have added the following description into footnote in page 13:

It is considering that the inundation depth-exceedance probability is from 0 to 1, when T=2-year *event, the P=1/2; and when T=5-year event, the P=1/5, when we impose the 1-year event, the P=1, which can make the inundation depth-exceedance probability is from 0 to 1.*

As the depth of the 2-year event in GLOFRIS global fluvial flood hazard maps is equal to 0 m. We assume D_1 *is also equal to zero (i.e., 1-year event with a flood depth of 0 m).*

13) Please specify what i and j represent and their range.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have added the following description in lines 11-13, page 14:

"i" *is the sequence number of simulated flood event;* "j" *is the sequence number of basin number, which belongs to (1,9)*

14) I guess this is referred to the return period. Please rewrite.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have rewritten 100 year into 100-year event, etc, throughout the manuscript.

15) here, you should use 'Netot' instead of 'Nec+Ned'

Response: In the revised manuscript, we have changed $N_e^c + N_e^d$ into N_e^{tol}.

$$P_e^{tol} = \sum_{i}^{N_e^{tol}} CA_i * 0.8$$

16) what does the drainage capacity rate take into account? Why you have assumed the value 0.8? This didn't come from sensitivity analysis, right?

Response: The value and the concept of the drainage capacity rate are referred to Espinet et al., (2018), which is defined as the drainage capacity of embankment, bridge and culvert. In this work, the value is 0.7 for bridges and culverts in Mozambique. Considering China is more developed than Mozambique, we assume the infrastructure in China has a higher drainage capacity. In addition, we do also think that the parameter will lead to a large uncertainty to the performance loss. Therefore, we perform a sensitivity and uncertainty analysis in Section 3.4. In the revised manuscript, we have added the following description into footnote in page 18:

The value and the concept of the drainage capacity rate are referred to Espinet et al., (2018), which *is defined as the drainage capacity of embankment, bridge and culvert. In this work, the value is 0.7 for bridges and culverts in Mozambique. Considering China is more developed than Mozambique, we assume the infrastructure in China has a higher drainage capacity and a value of 0.8 is assigned.*

17) What values have you considered for CA. What is the source? Probably this information should stay in section 2.2.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have added the CA information in Section 2.1.3. Please refer to lines 4-12, page 8:

2.1.3 Chinese Railway data

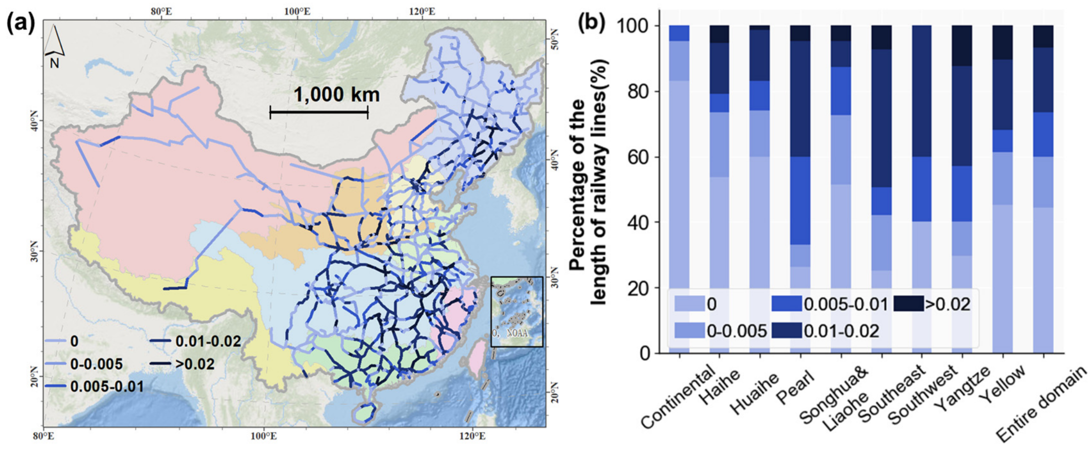
The geographic information, time table data, as well passenger's capacity data of Chinese railway are collected. The geographic information of railway system from OpenStreetMap (OSM), which provide the spatial distribution of the Chinese railway system (Fig. 1). The timetable data, which includes the daily number of trains and associated routes from the Railway Service Website, while the passenger's capacity data is obtained from https://www.china-emu.cn/ for China High-Speed Train (G Train, D Train, and C Train) and https://zh.wikipedia.org/wiki for others (Z Train, T Train, K Train, etc.).

18) Is this 'increased time' limit supported by any directive or something else?

Response: In China, the operation cycle of each train is once a day. Therefore, if the increased time for re-routing one trip is greater than 24 hours, it will stop and the new trip will be started the next day.

19) I would suggest to use the same background map as in the Figure 3. This would make possible to make a clearer distinction in terms of basins as well. Also, I suggest to specify that the bar named National is referred to the entire domain.

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have changed the background map for Fig. 4 and changed the National to the entire domain. Please refer to line 14, page 23:



20) This is not true for the Soutwest and Continenatal basins, where the percentage are pretty much constant. Does the author investigate this aspect?

Response: The low and constant impacts of daily affected trains observed in the southwest and continental basins are due to a lower railway line density and daily train flows. At the same time, the lower annual failure probability of the rail segments in these areas also leads to a lower probability of failed railway segments per flood event and results in lower impact. In the manuscript, we have added explanations in lines 9-21, page 29 and line 1, page 30:

For most basins, between the 25-year and 100-year flood events, the percentage of daily affected trains and daily cancelled trains relative to the total number of daily trains per flood event increases. The rule is not suitable for the Soutwest and Continenatal basins, where the percentage are pretty much constant and low. It is due to a lower railway line density and train trips in these two basins. A low impact is expected even all railway lines are disrupted. While the percentage of daily detoured trains relative to total daily trains and the total and average increased time, increases between the 25-year and 50-year flood events, and sharply decreases between the 50-year and 100-year events, especially for the Yangtze River, Yellow River and Pear River Basin floods. This is because most of the north-south rail lines in China, such as the Beijing-Guangzhou and Beijing-Jiulong lines, cross these basins. Most trains that are detoured under a 50-year event cannot be detoured under a 100-year event, as most of the north-south rail lines suffer failures at this hazard intensity.

21) Could you report this map in the SM?

Response: We thank the reviewer for the suggestion. In the revised manuscript, we have added this map in the SM. Please refer to lines 1-3, page 50

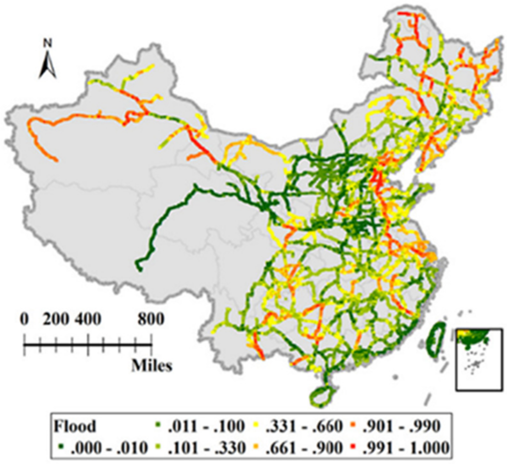


Fig. A8 Susceptibility map of the national railway network subjected to flood (source: Liu et al., 2018b)