We thank the reviewer's thorough reading of the manuscript and valuable remarks that helped us to improve the manuscript. The comments are very helpful and we 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) Abstract: the discussion on the results is too large and detailed for an abstract while a brief description on the adopted methodology is totally missing;

Response: We thank for reviewer's comments and agree with the reviewer that the abstract needs a brief description of methodology and a more streamlined description of the results. In the revised manuscript, we have rewritten the abstract parts to address the concern of the reviewer. Please refer to lines 12-21, page 1 and lines 1-10, page 2:

Floods have negative effects on the reliable operation of transportation systems. In China alone, floods cause an average of ~1125 hours of railway service disruptions per year. In this study, we present a simulation framework to analyse the system vulnerability and risk of the railway system to floods. To do so, first, we developed a novel methodology for generating flood events at both the national and river basin scale. Based on flood hazard maps of different return periods, independent flood events are generated using the Monte Carlo sampling method. Combined with network theory and spatial analysis methods, the resulting event set provides the basis for national- and provincial-level railway risk assessments, focusing in particular on train performance loss. Applying this framework to the Chinese railway system, we show that the system vulnerability of the Chinese railway system to floods in different basins is highly heterogeneous as a result of spatial variations in the railway topology and traffic flows. Flood events in the Yangtze River Basin show the largest impact on the national railway system, with approximately 40% of the national daily trains being affected by a 100-year flood event in that basin. At the national level, the average number of daily affected trains and passengers for the national system are approximately 200 trips and 165,000 people (2.7% and 2.8% of the total daily numbers of trips and passengers), respectively. The event-based approach presented in this study shows how we can identify critical hotspots within a complex network, taking the first steps in developing climate-resilient infrastructure.

2) Introduction: the introduction should provide also some further details on both the adopted

methodology and metrics. An anticipation of the analyses that will be carried out is essential to encourage potential readers to go through the paper. Novelty of the proposed approach should be better stressed.

Response: We thank for the reviewer's suggestions. In the revised manuscript, we have added more details on both adopted methodology and metrics in the introduction section in lines 14-21, page 5 and lines 1-7, page 6; and the novelty of the proposed approach have been stressed in lines.

This study aims to develop a framework to quantify the system vulnerability and risk to transportation systems in terms of operational performance loss under large-scale flood hazards. System vulnerability in this study is represented as the system performance loss with different flood intensities. When assessing possible cascading effects, the use of independent flood events is necessary (Nones and Pescaroli, 2016), as the presented floods in regional-or national-scale flood footprints, which show the flood depth for a given return period in that area, may not all happen at the same time. To overcome the shortcomings in existing studies, we develop a simplified practicable and novel method for generating a set of independent flood events at the national and river basin scale. The independent floods are generated using a curve fitting method and Monte Carlo sampling method based on global flood hazard model maps and river basins. By coupling simulated flood events with the railway network using the spatial analysis method, we identify the railway failure hotspots caused by floods. At the same time, the potential performance loss is assessed using network theory. We illustrate our methodology by applying it to the Chinese railway system.

3) In Data and Method section and sect 2.1.1, the global flood hazard model should be better described (also providing some examples in the SM). All the adopted metrics should be defined much more carefully, with a more precise and effective use of terms. For instance, only the trains where passenger travel can be cancelled or detoured, while passengers cannot be cancelled or detoured; so the metrics named 'passenger cancelled induce' or 'passenger detoured' in my opinion should be renamed (and better defined at their first appearance in the text).

Response: We thank the reviewer for their suggestion. In the revised manuscript, we have added more detail on the global flood hazard model and provide the flood maps for 50 and 500-year events in supplement materials in lines 9-20, page 9 and lines 1-3, page 10. Adopted metrics have been redefined and please refer to our response to Question 4.

Our flood hazard data are extracted from the GLOFRIS global fluvial flood hazard maps of Winsemius et al. (2013), which are developed using the GLOFRIS modelling cascade provided in Ward et al. (2013) and Winsemius et al. (2013). The GLOFRIS modelling cascade first simulates daily discharge using the PCRaster GlobalWater Balance (PCR-GLOBWB) global hydrological model (Beek et al., 2008, 2011). Based on daily discharge, daily flood volumes are simulated using the PCR-GLOBWB extension for dynamic routing, DynRout (PCR-GLOBWB-DynRout) (Ward et al., 2013; Winsemius et al., 2013). In the next step, flood volumes, for different return periods: 2, 5, 10, 25, 50, 100, 250, 500 and 1000 years, are obtained using the annual time series for maximum flood volumes by fitting a Gumbel distribution. These flood volumes are then converted into inundation maps (30-arcsecond, ca.1-km) using the inundation downscaling model of GLOFRIS (Winsemius et al., 2013). In the appendix materials, we provide flood maps for the 50 and 500-year return periods. The maps show that the inundation depth highly varies in China. Railway lines in eastern coastal China and South China are faced with the most severe floods.

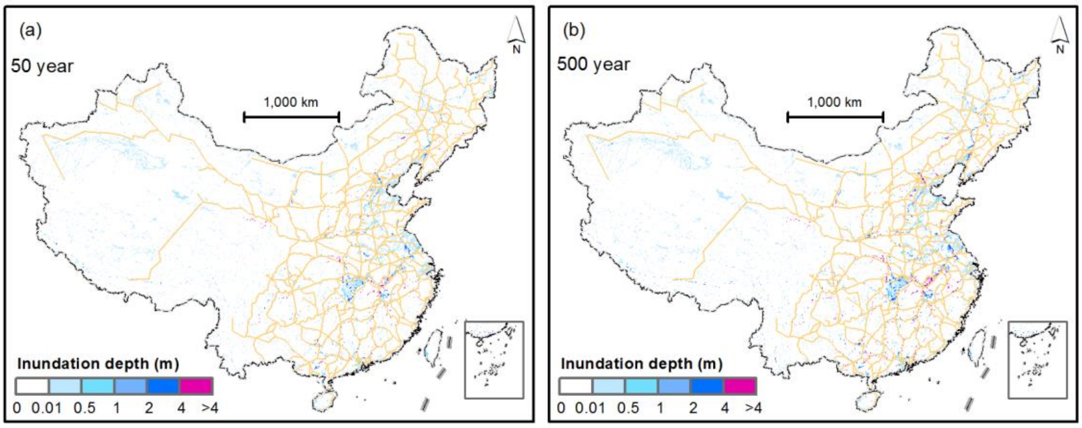


Fig. A1 (a) the 50-year flood, (b) the 500- year flood

4) The description of the fitting procedure (Sect.2.1.2) must be improved. Figure 2a is rather unclear to me, and the caption does not help the readers. Moreover, its size is too small and the inset legend cannot be read (similar problems are present also in Figures 5, 6 and 7). I would suggest to place the four graph in figure 2 in a 2 x 2 grid, enlarging each graph. Caption must be more clear for figure 2a and more concise for figs 2b, 2c and 2d.

Response: We apologize for the unclear description and thanks for your suggestions. In the revised manuscript, we have rewritten Sect 2.1.2 in lines 12-20, page 10, page 11, and lines 1-3, page 12. At the same time, Figure 2a as well as Figures 5, 6, 7 have been improved as followed.

2.1.2 Fitting procedure

For each grid cell, the GLOFRIS maps estimate the flood depth for the nine aforementioned return periods (2, 5, 10, 25, 50, 100, 250, 500 and 1000 years). To estimate the flood depth for any return period, we fit a quadratic spline function to develop an inundation depth-exceedance probability function (P) for each return period interval for each grid cell (Marsden, 1974; Vandebogert, 2017; Meshram et al., 2018). The quadratic spline is a method that uses a piecewise quadratic function to obtain the best-fitting curves. This interpolation method allows us to obtain a smooth continuous curve through the provided flood depths for the different return periods.

The method is applied as follows, and examples of the inundation depth-exceedance probability function of grid cells are shown in Fig. 2a:

For each grid cell $\mathrm{g_{x,y}}$*, the annual exceedance probability flood depth* D_T *is calculated by Eq. 1:*

$$P(D_T) = \frac{1}{T} \tag{1}$$

where D_T *is the magnitude of a flood depth with a return period of T- year,* $P(D_T)$ *is the exceedance probability of* D_T*.*

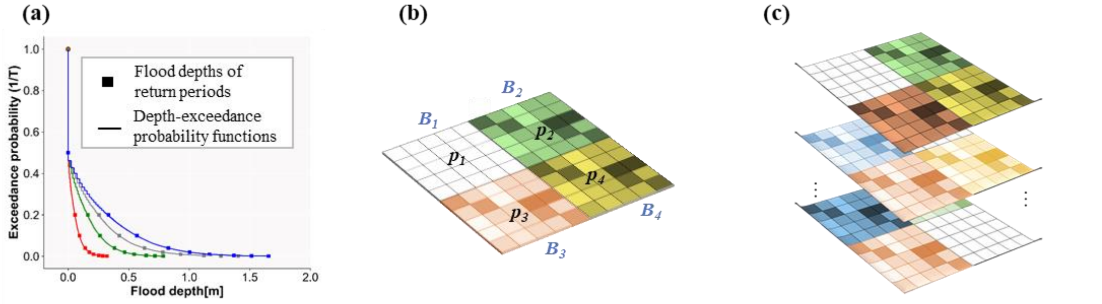
Let $Pr(D_T)$ *denotes a quadratic, continuously differentiable function of* $P(D_T)$*. Then, by definition:*

$$Pr(D_T) = aD_T{}^2 + bD_T + c \tag{2}$$

For each return period interval of grid cell $g_{x,y}$*, we can obtain its piecewise quadratic function by Eq. 3:*

$$Pr_{x,y}(D_T) = \begin{cases} Pr_{x,y}^1(D_T) = a_1 D_T{}^2 + b_1 D_T + c_1 \quad D_T \epsilon [D_2, D_5] \\ Pr_{x,y}^2(D_T) = a_2 D_T{}^2 + b_2 D_T + c_2 \quad D_T \epsilon [D_5, D_{10}] \\ \cdots \\ Pr_{x,y}^8(D_T) = a_8 D_T{}^2 + b_8 D_T + c_8 \quad D_T \epsilon [D_{500}, D_{1000}] \end{cases} \tag{3}$$

where $Pr_{x,y}(D_T)$ *is a set of continuous inundation depth-exceedance probability functions consisting of 8 continuous quadratic functions for* $g_{x,y}$ *and shows in Fig. 2a with curves. For* $a(a_1, a_2, \dots, a_8), b(b_1, b_2, \dots, b_8), c(c_1, c_2, \dots, c_8) \epsilon R$ *, we can calculate these constants by bracketing the critical point of* $\mathrm{P(D_T)}$ *and derivative of the function* $Pr_{x,y}(D_T)$*; details on the* *interpolation methods can be found in a previous study by Sun and Yuan (2006). In this work, we assume that only one event occurs per year in each basin since we assume the intensity of events is equal to or larger than 1-year. When the return period is lower than 2, the flood depth is set to zero which is the same as that of a 2-year event.*



(a) Fit inundation depth-exceedance probability functions

(b) Generate random number P_i for basin i to get independent flood event for each basin and combine them to obtain a national-scale event

(c) Repeat process (b) and generate 10000 national-scale flood events

Fig. 2 An example of generating national-scale flood events. In (b), p1, p2 , p3 , and p4 are the random number between 0 and 1 generated for basin B_1, B_2, B_3 and B_4, which are used to generate basin-scale events based on the functions in (a). The layers of basin-scale floods in (b) are combined into a national-scale flood event. The layers in (c) are the 10000 national-scale events using the process in (b).

2.1.3 Simulation procedure

To produce a time-series of flood events based on the created inundation depth-exceedance probability functions (Section 2.1.2), we use a Monte Carlo sampling method. 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 (Baker, 2008; Speight et al., 2017). The flood event generation procedure is presented in Fig. 2 and Appendix Fig. A1 and can be summarized in two steps. First, we generate independent events at each basin and combine them into a national event. For an event E_j^i, and for each basin B_j, a random number P_j^i between 0 and 1 is generated from a uniform distribution. The flood depth of the cells in basin B_j for event E_j^i can be calculated using P_j^i and the inundation depth-exceedance probability function based on the assumption that a flood event in one basin will produce a flood with the same intensity. For a national-scale flood event, basin-specific floods of nine basins can be randomly combined into a national-scale flood by assuming independence between the flood events among different *basins, this concept is presented in Fig. 2b. Second, we repeat this process 10000 times to generate a set of national-scale independent flood events as presented in Fig. 2c.*

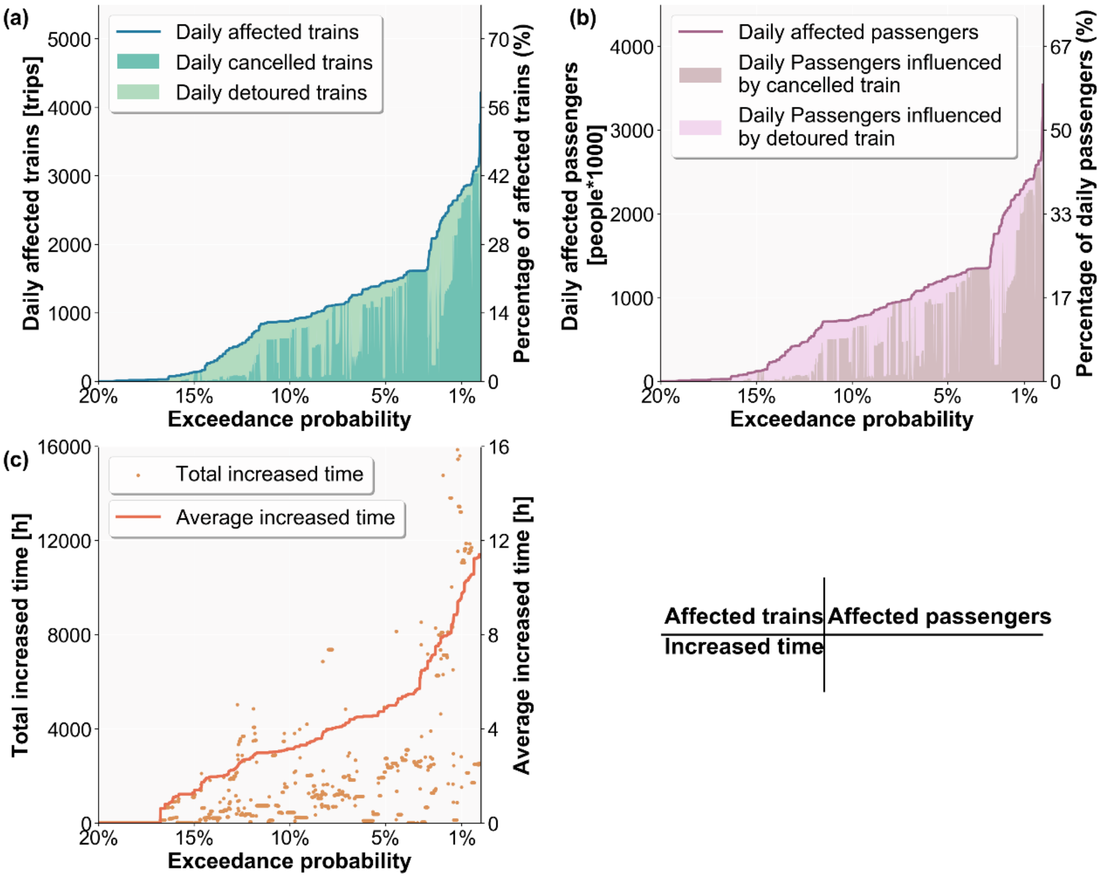


Fig. 5 Exceedance probability-performance loss curves

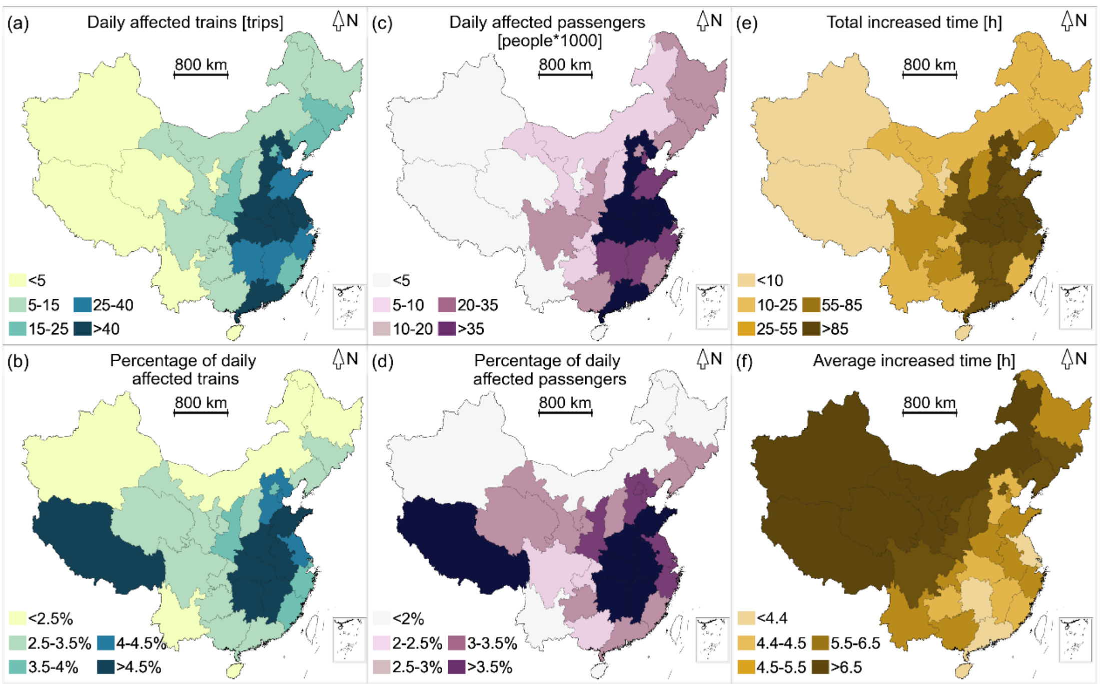


Fig. 6 Performance loss of the railway system per province.

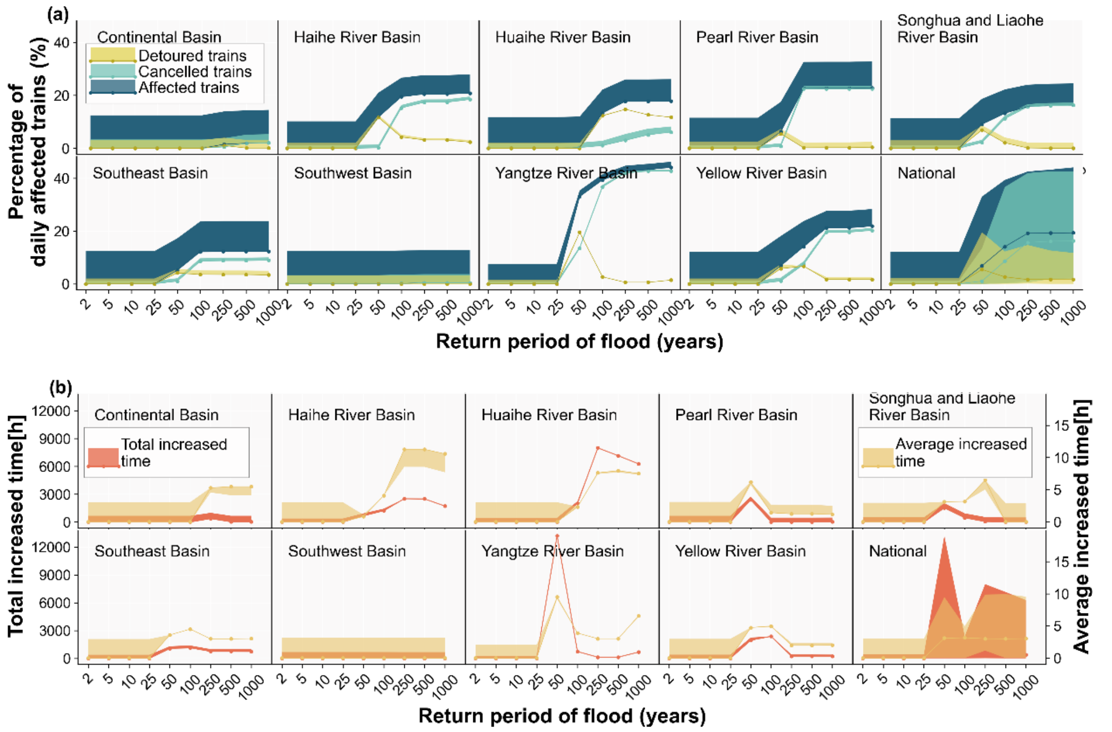


Fig. 7 System vulnerability curves induced by river floods from the national flood event set

5) 2.4 could be renamed “performance loss metrics” and restructured with a separate subsection for each metric. Subsection 2.4.2 could become sect. 2.5. All the assumptions made for the metrics definition must be better clarified.

Response: We thank for the reviewer’s suggestions. In the revised manuscript, we renamed “performance loss metrics” and restructured them in a separate subsection. For the assumptions of the metrics clearer, we added some descriptions in lines 2-4, page 18:

2.4 Performance loss metrics

2.4.1 Daily affected trains and passengers

2.4.2 Daily detoured trains and passengers influenced by detoured train

2.4.3 Total increased time for the detoured trains

2.4.4 Average increased time for the detoured trains

2.4.5 Daily cancelled trains and passengers influenced by cancelled train

2.5 Calculating system vulnerability and risk

2.6 2.6 Uncertainty and sensitivity analysis

We assume that the average number of passengers is 80% of the train’s capacity (Wei et al., 2017;

Rezvani et al., 2015). As such, the number of affected passengers P_e^{tol} *can be defined by Eq. 9:*

6) Results section presents a quite good description of the results while comments on the potential implications of the various results are almost totally missing or present only in the discussion section; this aspect could be improved. The discussion on the results of the sensitivity and uncertainty analysis in Setc.3.4 should be considerably improved; for instance, pie charts in Fig.8 should be explained and commented.

Response: We thank for the reviewer's suggestions. In the revised manuscript, we add the comments on the potential implications of the various results in results parts in lines 14-17, page 22, lines 1-4, page 27, and lines 7-10, page 30. Meantime, the results of the sensitivity and uncertainty also improved in and lines 7-16, page 31.

3.1 Failure hotspots of railway segments

Figure 4b shows the percentage of the length of railway lines that fall into each failure probability category for the national- and basin-level analyses. Nationally, the failure probability is greater than 0 for more than 55% of the total length of the railway lines. This percentage is heterogeneous across different river basins: it is highest in the Southeast Basin, followed by the Pearl River Basin and the Yangtze River Basin. Nationally, 6.8% of the length of the railway lines has a failure probability greater than 0.02, with the highest proportions in the Yangtze River, Yellow River, and Southeast Basins, with 12.5%, 10% and 7.2%, respectively. The results for the Failure hotspots indicate that the railways located in Yangtze River, Southeast and Pearl River Basins need more attention and planned prevention measures to reduce the failure probability induced by floods.

3.2 Risk analysis of the Chinese railway system

Several provinces appear at the highest level of the three metrics presented in Fig. 6 and can be classified as particularly vulnerable provinces. Anhui Province, for example, has one of the highest absolute and relative levels of risk to trains and passengers in Fig. 6a-d but also has the highest total increased time in Fig. 6e. Hubei Province shows one of the highest absolute and relative levels of risk to trains and passengers in Fig. 6a-d. Jiangsu Province has the highest absolute levels of risk to trains and passengers in Fig. 6a and c and one of the highest total increased time in Fig. 6e.

These provinces are at the highest risk compared to the other provinces. This information can help researchers and local authorities to determine high-risk areas and prioritize risk management interventions to reduce risk. These can be used in the first steps of developing climate-resilient infrastructure.

3.3 System vulnerability of the Chinese railway system

When comparing the results between the nine river basins, we find that, in general, floods in the basins in central and eastern China have the highest impacts on the Chinese national railway system. The percentage of daily affected trains (cancelled and detoured trains) of the total number of trains is the largest for the Yangtze River Basin, followed by the Pearl River Basin and the Yellow River Basin. In the Yangtze River Basin, the median percentage of daily affected trains (cancelled and detoured trains) to the total number of trains is close to 40% for a 100-year flood event. For the Continental and Southwest Basins, the value is close to zero. The high impacts of daily affected trains observed in the central and eastern area are due to a significantly higher railway line density and daily train flows compared to the more inland river basins (see Fig. 3). The higher annual failure probability of the rail segments in the central and eastern regions shown in Fig. 4 also leads to a higher probability of failed railway segments per flood event and results in higher impact. The daily detoured trains in the Huaihe and Haihe River Basins in eastern China are higher compared to other basins, which leads to a large total increased time when one flood occurs. The reason is that the Huaihe and Haihe River Basins are located in eastern China and only cross railway lines in the eastern coastal area. Therefore, the affected trains have more detour options through the lines of the Yangtze and Yellow River Basins, which lead to more detoured trains and associated total increased time. For each basin, based on the vulnerability curve, once we know the intensity of flooding that would occur, we can estimate the affected trains and passengers. Based on this kind of information, local authorities could prepare dispatch plans in advance of floods.

3.4 Risk uncertainty and parameters sensitivity

Figure 8 and Appendix Fig. A. 7 present the sensitivity of the results to the assumed parameters and the range of performance metric uncertainty. Overall, from the uncertainty histograms, we can see that all the performance metrics are right-skewed, especially for the average daily affected and affected passengers shown in Fig. 8a and c, and average daily cancelled trains and cancelled *passengers shown in Appendix Fig. A 7b and d, have a long right tail for high performance loss estimates. This seems a little bit less for the average daily detoured trains and passengers shown in Appendix Fig. A 7a and c, and average increased time for detoured trains shown in Fig. 8e, which is probably the result of the assumption that detouring is impossible when the increased time for re-routing is greater than 24 hours, resulting in a smaller range of detoured options and thus a smaller range in resulting performance loss estimates. The average number of daily affected trains ranges from 100 to 500 trips. For daily affected passengers, it ranges between 100,000 and 450,000 people, and the average increased time ranges between 3.5 hours and 5.5 hours with the change in the parameters.*

In Fig. 8b, d, f and Fig. A 7f, the pie-charts show how much the uncertainty in each input parameter contributes to the variance of the performance loss estimates. The results show that the performance loss estimates are particularly sensitive to the values used for the design standards. Using the different parameter settings, we see a variation in the design standards of approximately 43%. The variation in the drainage capacity rate and water level threshold produces similar uncertainty as the capacity loss, which is approximately 28%. Reducing uncertainty in risk assessment is particularly challenging as it would require location-specific parameters. Despite the difficulties, these geographically varying design standards should be developed in the future to reduce uncertainty and improve the performance loss estimates.

7) Some practical examples of the utility of the proposed approach should be reported in the conclusion to highlight the importance of the work.

Response: We thank for the reviewer's suggestions. In the revised manuscript, we have added some practical examples of the utility of the proposed approach in the conclusion part on page 20 and lines 8-21, page 37 and lines 1-2, page 38.

The developed system vulnerability curves and flood risk maps can provide information for decisions on safety and effectiveness of operation and maintenance. Various performance metrics can be considered by management departments based on their particular problems. 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 and passengers) can be used as indicators to link with business disruptions. Future work can try to assess the interregional trade based on Input and Output tables and regional railway transportation performance decreased in our work. The assessment of shocks and indirect economic losses induced by railway system failures is essential for policymakers to design railway infrastructures and to measure indirect economic losses.*

8) Please double-check your References. I have found out some inconsistencies. For example, in your manuscript you refer to Liu 2018a and 2018b, but in the References I have found Liu 2009, Liu 2018 and Lyu 2018.

Response: We thank you for the reviewer's suggestions. We have checked the references and revised them in the revised manuscript.