Response to Reviewer 1

The revised version of the manuscript shows substantial improvement over the previous version. The methods section now includes a complete and transparent description of the statistical models. An appendix has been added that includes a sensitivity analysis comparing statistical models of different complexity. This helps to interpret the results in the main paper.

The appendix shows that some features of the statistical model are robust and not an artefact of overfitting (e.g. heavy precipitation increases the odds ratios, precipitation close to the event is more important than precipitation prior to the event). The absolute numbers of the odds ratios are however not always stable. In these cases the absolute values of the odds ratios should not be interpreted.
We have removed the interpretation of specific numerical values of the odds ratios for the meteorological controls.

I agree with the authors that a balance should be found between the omitted variable bias and overfitting. However, the sensitivity study in the appendix does not allow to decide if/at which stage of complexity overfitting is present/starts. That the most complex model has the lowest AIC is not sufficient to prove that the most complex model is not affected by overfitting. (an examples for this can be found e.g. here https://stats.stackexchange.com/questions/524258/why-does-the-akaike-information-criterion-aic-sometimes-favor-an-overfitted-mo). Some of the relationships between predicant and predictor shown in the manuscript are unexpected and contradict physical reasoning. In these cases I strongly assume that this is a consequence of overfitting. Examples for surprising results for which no physical explanation is offered in the discussion are:
We cannot rule out overfitting for our results, especially for gravitational mass movements where we have only a few hundred data points. However, many of the succeeding points can be explained either through technical, operational or hydrological reasons.

Line 329: "For gravitational mass movements, the coefficient is not statistically different from one, meaning there is no evidence of a statistically significant difference between the odds of a gravitational mass movement with and without heavy rainfall." The values in table 4 at day of heavy rainfall and 1 day after heavy rainfall are statistically significant. How can this be?
This is due to the fact that the model in Table 2 contains interaction terms (equation (3.1)), and the model in Table 4 contains no interaction terms (equation (3.2)). We have refrained from including interaction terms in the second model to maintain simplicity, as interactions were not relevant for the question being tackled in the second model.

Line 377: "For tree fall events, the odds ratio on the day of a heavy rainfall is 0.296 and statistically significant, meaning that the odds of a tree fall event occurring on the same day as a heavy rainfall is less than a third that of a situation when no heavy rainfall occurs. In contrast, one day after a heavy rainfall event, a tree fall event is 2.4 times more likely to occur than in days with no heavy rainfall." Is there any physical reason to explain this. I assume this result is an artefact caused by the fact that the complex model includes two predictors for the event day that have to "share" the odds ratios (day of heavy rain and precipitation at route segment (at the event day)).
The explanation for this is an operational one and lies in the way how the data for tree fall is collectedby Deutsche Bahn. As mentioned in Secion 4.2, the data collection for the DB damage database is done solely for the purpose of ensuring the safe and swift resumption of railroad operations. Tree fall events are reported by train operators only upon encountering them en route. When a heavy storm or rainfall is expected, often train journeys are cancelled in advanced to ensure

the safety of passengers and employees. Therefore, less trains travel on days of heavy rain, making it less likely to encounter tree fall events on the same day. Only after the storm has settled does the DB deploy their workers to survey the damage, and then most of the tree fall disruptions along the rails are encountered and reported.

Line 429: "For c) and i), the curve “with heavy rainfall event” has a U-shape. Thus, the probability of a natural hazard occurring during a heavy rainfall event is higher when the soil moisture takes on extreme values than when it takes on average values. In the case of tree fall, this is particularly the case for low soil moisture values, and in the case of floods for high soil moisture values. The arc shape in f) indicates that the probability of occurrence is highest at medium soil moisture values."
If there is a physical reason for the shape of the curves in c,f,i, please explain. Why do medium values show the opposite signal compared to both extremes? I think this relationship is not related to physics but probably caused by overfitting.
For c): There is evidence that extremely dry and extremely wet soil are determinants of floods (Vichta et al. 2024), mainly due to the hydrophobic properties of soil, and oversaturation, respectively. Heavy rainfall in an environment with very dry and hydrophobic soil or very wet and oversaturated soil can therefore easily trigger a flood event, and this is reflected clearly in the results. In moderate soil moisture cases, where the soil can still absorb water brought about by heavy rains, the effect of heavy rain is then less pronounced.
For i): It has been shown that drought stress can cause tree mortality (Grote, et al. 2016), meanwhile, soil oversaturation can cause waterlogging stress in trees (Gill, 1970; Kreuzweiser & Rennenberg, 2014). With trees that are already under stress and vulnerable in very dry or very wet soil conditions, a heavy rainfall event could cause additional stress and be more likely to trigger tree fall.
These explanations have been added to the text of the paper.
For f): We have added in the paper that we refrain from interpreting this result because of the insignificance of the coefficient of heavy rain on gravitational mass movement.

Line 452: "the mean elevation within the heavy rainfall area reduces the odds of gravitational mass movement events". More events at higher elevation might be interpreted as an indication for the existence of a hill (with a slope). I cannot think of any physical reason that explains less events at higher elevations.
There are several possible aspects to this result. First the number of gravitational mass movement events is very small in the available data set, so more data is needed for a robust validation of this finding. Second, train lines in Germany are more often located in lower-elevated areas, and for train lines built in higher elevated areas, more attention is likely given to landslide prevention. However, a deeper analysis of the distribution of railway lines among different elevations, not to mention better (and more) gravitational mass movement data, is necesary to validate these possible explanations.

I still think that unphysical predictors such as the year of the observation, lessen the informative value of the statistical model. As long as only the robust results are interpreted in the article, this is acceptable.
I therefore suggest that the authors add a sentence stating that they will only describe and interpret robust results. All other "results" should be removed from the manuscript. Then the manuscript can be published.8
We have added in line 310 a sentence saying we will only interpret robust results.

Further comments:

Line 401: "using the results of the interaction terms" Is this based solely on the interaction term or on the full model including the interaction term? In case this for the interaction term only: The odds ratio for the interaction terms is not statistically significant in many of the combinations. In line 352 you argue that you don't discuss results which are not statistically significant. Here you even include a figure+discussion. I would remove figures and discussion for the not-significant combinations. In case this is for the full model: What are the values for the variables that are held constant (mean/median?)?

We used the full model to generate the graphs in Figure 4. The figure is generated by using the predictive margins approach (Williams, 2012): for each observation in the dataset, the actual observed values of the control variables are used to calculate the predicted probability of a hazard event if there had been heavy rain and if there had been no heavy rain. These observations were then ordered according to relevant meteorological variables, say precipitation, and the predicted probabilities with and without heavy rain for observations with the same precipitation levels were averaged. Therefore, for each value of precipitation, we get two points: the average probability with heavy rain (points on the dashed line) and the average probability without heavy rain (points on the solid line).

We have added an short explanation of the approach in line 370.

Line 490: is V3 really small scale?

All variables in the CatRaRE data set are calculated for the event zone. As most of the heavy rainfall events are of small extent, we assumed that the data describes the local climate conditions. For better understanding, we replaced the formulation „small scale“ by „local“.