Response to Anonymous Referee #2

Reply by the authors on 13 august 2020.

The paper shows a methodology to calculate direct flood risk for roads on the basis of developed damage curves. This is a relevant topic and the methodology that is presented in the paper provides very useful information and contributes to advance risk knowledge particularly in the European context.

We thank the referee for the in-depth review work and for the nice comments on usefulness of the methodology and results.

My comments are as follows:

The paper indicates that a new object-based approach and new damage curves are proposed. However, I recommend to clarify what the novel contributions are *[22]*, since object-based approaches have been used before (e.g. Hackl et al, 2017). In paragraph 65 it is mentioned that Koks et al. (2019) used OMS data for a global multihazard analysis but that several assumptions were made due to data scarce regions of the world, but in the European context data is more complete and a more detailed analysis can be carried out. Does this mean that the approach in the paper is the same as in Koks et al. (2019) but with more detail?. What are the differences with the approach in Koks et al. (2019) *[22A]*? What are the differences with the approach in Bubeck et al. (2019) *[22B]*?

The novel contributions of our study are as follows:

- *Indeed, the computational core of the model is similar as what was used by Koks et al. (2019). However, we have developed new vulnerability curves (which were very stylized in Koks' study) and did an extensive literature review on road reconstruction and damage costs to support these. Also, we use road metadata/attributes on the road type, number of lanes, presence of bridges and street lighting to improve the damage estimate, which was not done in Koks study. Finally, we make an extensive comparison to grid-based approaches using the same hazard data, to gain an understanding about the differences between both approaches. Also, we present our results with high spatial detail, rather than aggregated per country. All these elements are new.*
- *The spatial extent of our study is much broader than for example the work of Hackl et al., which had a study area of 20*20 km, whereas we cover entire Europe. We will make a reference to this in the introduction.*
- *Bubeck studies the railway network, whereas we study the road network which contains much more elements and is more diverse in terms of road types.*

In the introduction, we will more clearly indicate the above novelties.

The paper will gain clarity if a clear differentiation between direct and indirect damage is established *[23]*. This will set the context to clarify the difference between an approach that analyses the physical damage to the road and one that provides a network analysis.

We thank the referee for pointing this out, because one of the largest benefits of taking an object-approach for direct damage, is that the same objects can be used to construct a network on which an analysis of indirect damage can be done. Hence, the object-based approaches enables the study of direct and indirect damage in a coherent way.

In the last lines of the first paragraph of the introduction, we will explicitly make the distinction between direct and indirect damage and indicate that the remainder of the paper will focus on improving the estimates of the direct, physical damage. In the conclusion, we readdress this issue when we illustrate the potential of the object-based approach for studying indirect damage as well.

The main objective of the paper, as stated in line 65 is to compare the results and performance of an object-based approach to the more traditional grid-based approaches. It seems to me that such comparison is problematic. Results and performance of a grid-based approach depend on scale. If the grid is too coarse results will be coarse and then limited to, for example, strategic decisions in flood risk management. If the scale is detailed the performance should be better and the decisions informed by a more detailed flood risk assessment, therefore, will be more local. Object-based approaches are intended for detailed assessment, therefore they should provide a better level of information than e.g. mesoscale analysis. I recommend to clarify in the paper the differences in scale of the methods and how they can be compared *[24A]*. For example, what is the purpose of the continental-scale analysis (what decisions can be informed? It is for hotspot identification? are hotpots better identified by the object-based approach?) is this purpose better fulfilled by the object-based approach? *[24B]*. According to the explanation on section 2.2 and figure S2 it is clear that Corine 2012 and LUISA 2018 have a coarse resolution (100 m) then it is impossible that an object with only several meters of width can be appropriately represented by this information, besides Corine 2012 even lacks the representation of the roads in some cases. Then, I recommend to elaborate the discussion on the limitations of this data sets and their purpose *[24C]*.

In line with what was also suggested by referee 1 [comment 3A], we will expand our description of how roads are represented in the grid-based based methods with CORINE and LUISA, and what the implications for our comparison are [24A, 24C].

We can partly understand the referees objections against comparing the object-based and grid-based approaches, because there are indeed large differences between both methods in terms of scale and intended purpose. However, we reasoned that these differences instead would be an interesting starting point for a comparison of the 'new' continental-scale object-based approach, to investigate if it made a significant difference on the model results. Actually, rather than interpreting our results as a 'criticism' of the grid-based approach, one could also see it is as rather strong confirmation of the validity of the grid-based approach, because we show that LUISA with the Huizinga damage curves can give reasonable estimates of the order-of-magnitude of the continental-wide damage, [as you also point out in comment 34].

There are also arguments pleading for a comparison between the both approaches. Both aim at gaining continental-scale estimates of the damage to road infrastructure. Despite the limitations of the grid-based approach, it used to support EU-scale policies (possibly even beyond the original intentions of the academia developing these models). We think the object-based approach better fulfills this purpose, because the results are directly presented at the level of road segments which more directly meets the needs of road owner or operator. Therefore, we think it is useful to compare the 'new' approach to the more common approach of assessing flood risk at the continental scale.

*We do not agree that object-based approaches are necessarily more suitable for high-resolution approaches while grid-based approaches are more suitable for low-resolution approaches. Firstly, because there are also examples of very high-resolution studies using grid-based approaches, for which it can be questioned if an object-based approach would add value. Secondly, because we show how an object-based approach can be used with a rather coarse (100*100 m) continental scale inundation model, which can help to identify flood hotspots on the continental level.*

With regard to [24B], indeed the purpose of hotspot identification indeed is better fulfilled by the object-based approach, because:

- *We have more detailed information about the location of the road (because as we earlier argued, the precise location of the road cannot be easily captured in the rasterized grid)*
- *We have more detailed information about the characteristics of the road (so we can distinguish a major highway in the European E-road network from a local motorway), which is not directly clear from a mere land use raster*
- *Because the road objects can also be studied as a network graph, the object-based approach cannot only identify flood hotspots in terms of direct damage, but also in terms of indirect damage. Arguably, this indirect perspective is very important for the EU-policy maker or road owner. After all, a road flood hotspot is not a spot with only large damage, but also with high criticality for overall network performance.*

Line 91 mentions the Huizinga infrastructure damage curves. However, no introduction of the Huizinga et al. (2017) reference is made. Since these curves are very important in the paper, consider including an introduction of these and why you chose those curves before they are mentioned in line 91 *[25]*.

We thank the referee for the insight that we introduce the Huizinga curves too late in the manuscript, whereas they are indeed important for the comparison in the paper. Therefore, in the updated manuscript, we will already introduce them in the introduction. With regard to the question why we choose the Huizinga curves the answer is simple: this is the most comprehensive set of damage curves available for Europe, and are therefore used in virtually all continental-scale studies (for examples see the enumeration in the old section 2.3.4), making them a fitting benchmark for our new curves.

Line 109 indicates that flood hazard is represented with a set of inundation maps taken from Alfieri et al. (2015), with a recent update by Dottori et al. (2020) which cover most of the European domain at a grid resolution of 100 m. The scale of the hazard maps could be considered coarse. Consider including in the paper a discussion about the impacts of the scale of the hazard maps particularly when using the object-based approach *[26]*. Are the scales commensurated?

*For continental-scale hazard modelling, a 100m*100m arguably is state-of-the-art, because comparable models (such as GLOFRIS) operate on a 1*1 km scale. The highest resolution currently available is from the Fathom model (90*90 m), Sampson et al., 2015. But we fully agree that on a lower spatial scale, much higher resolution hazard models are available, and that it would be very interesting to test the performance of the object-based models for these models. With the very high exposure data of OSM, the resolution of the hazard data on this scale indeed is currently the bottleneck in the modelling, and in that sense the scales in our models are not fully commensurated because an improvement can be made on the hazard-side now. This we already discussed in the second-last paragraph of the conclusion, in a recommendation which we will rephrase as: "The flood hazard data can be easily substituted with (small-scale) high-resolution data, to fully exploit the level of detail offered by the OSM exposure dataset. Moreover, more local scale case studies are required to validate the proposed damage curves."*

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model. Water Resources Research, 51(9), 7358–7381. https://doi.org/10.1002/2015WR016954

Paragraph starting at line 133. Please refer to the appropriate table at the beginning of the paragraph otherwise it is difficult to understand from where the percentage of 27% is obtained. How were these percentages obtained by Blending of Huizinga (2007)? and why damage was calculated to the percentage of infrastructure in each land use class? Could it be any type of infrastructure? *[27]*

We will make reference to the table in the beginning of the paragraph. The use of the percentages "originates from Huizinga's suggestion to map the damage functions to CORINE using a land cover cross-tabulation from the EEA)" (section 3.1).

We rewrote the entire paragraph as: ""To correct for the underrepresentation of infrastructural land use (and other land use types) in CORINE, for each land use class some percentage of infrastructure is assumed (Table S1,2). This follows from Huizinga's suggestion (2007, p. 2-22) to map his damage functions to CORINE using a cross-tabulation by the EEA (2006), which became their default implementation method. To enable a comparison with the object-based approach, we maintained the percentage of infrastructure per land cover class, whereas the contributions of the other damage curves were ignored (Table S1, 2). An implication of these percentages is that, although motorways and trunk are mostly missing in CORINE, damage for (local) roads in urban and industrial areas (amongst others) is still calculated, albeit without any explicit spatial reference to the actual road position, but based on their average presence in these urban and industrial land use types. Also, note that in the land cover category 'road and rail networks and associated land', a uniform share of 27% infrastructural land use is assumed, which to some extent corrects the overestimation of the actual road widths in LUISA.".

Huizinga (2007) mentions that the curve is explicit for roads, although in practice, many studies also use them in other land use categories such as airports and railways where the curve seems to cover a share of the runways, railways etc., besides the roads that it represents in these land use categories. But the curve is intended for roads, which enables a fair comparison.

Line 148 implies that the approach in the paper was proposed by Koks et al. 2019 and that the difference with the method in the paper is the use of damage curves tailored to the European context and uses additional metadata. I recommend that this is clearly stated, particularly in the introduction and the conclusions *[28]*.

This point was also raised in comment [22]. We will stress the novelty of our approach in the introduction and conclusion as elaborated under [22].

Line 205 states that only flow depth will be used for the damage curves. However, velocity is indeed considered at least to develop two groups of curves. I recommend to review the way in which the parameters to be considered are presented *[29A]*, because it is confusing that velocity appears afterwards in the analysis. Furthermore, what consideration should be given to duration *[29B]*? Is this factor important in damage? Should it be considered somehow?

We thank the reviewer for this suggestion, the original order in which we present the parameters are presented is indeed confusing. We removed '… an approach also adopted in this study' (lines 205-206) and rewrote and restructured section 2.2 as follows.

"In general, damage curves derive flood damage from flood parameters such as water depth and flow velocity. Continental-scale models typically work with functions relating damage to water depth only (Alfieri et al., 2016a; de Moel et al., 2015; Winsemius et al., 2013). Flow velocity, however, is at least as important as water depth for explaining damage to roads (Kreibich et al., 2009; Merz et al., 2010;

Thieken et al., 2009). Under low flow velocities (< 0.2 m/s), there is hardly any structural damage to pavements, whereas under high flow velocities (> 2.0 m/s) there is most likely severe structural damage (Kreibich et al., 2009). Indeed, pictures of floods in Europe show that under very quiet flow conditions, a road can remain almost undamaged whereas under flash floods with strong currents, complete reconstruction may be required. The flood hazard maps used in this study represent floods in rivers with an upstream area >500 km^2 whereas large flow velocities are typically found in smaller water courses in steep upstream areas and locally close to dike-break locations (De Moel et al., 2009). Therefore, we assume that the predicted floods have relatively low flow velocities. We deal with the remaining uncertainty by estimating two depth-damage curves that span the uncertainty of this typical slow flow velocity; one for the low-flow estimate and one for high-flow estimate that can be reasonably expected for large river floods."

Kreibich et al. (2009) show that structural damage of road infrastructure is best explained from flow velocity/flow force and an 'intensity' indicator which is a combination of water depth and flow velocity, so we removed the reference to 'duration of the flood' (since it distracts from the narrative).

Line 217 mentions for the first time in the paper a Lisflood model. It is very unclear what was Lisflood used for. In the previous sections the source of the flood hazards is explained without any mention to Lisflood and in the methodology the use of Lisflood is not explained. I recommend to complement the methodology section explaining the use of Lisflood, what type of model was developed, what area was modeled and what information was used for the model *[30]*.

We apologize for the confusion generated here. We reformulated this sentence (lines 217-220) as well as a paragraph in lines 465-468 to better explain our reasoning. Furthermore, we now explain in Section 2.1 that the flood hazard maps used in this work were calculated with the hydrodynamic model LISFLOOD-FP, while the hydrological input was calculated by the hydrological model LISFLOOD. Alfieri et al (2014) described the complete procedure including a general description of LISFLOOD and LISFLOOD-FP, and how they were applied. Since in the present work the maps were not modified, we opted to make reference to the original papers, rather than providing again a description.

Line 220 indicates that it was assumed that the predicted floods have relatively low flow velocities. Please elaborate your explanation here *[31]*: Your study area does not cover any area where flash floods occur?, how did you carry out that identification?

Our study area covers entire Europe. In Europe flashfloods do occur. However, our study is focusing on the large scale river floods occurring in Europe. River floods typically originate from much larger catchments than flashfloods, and also have much larger timescales (inundation takes days rather than hours). The hazard maps only includes floods originating from catchments large than <500 km^2, and therefore simulate river floods rather than flash floods. Large scale river floods typically have lower flow velocities than flashfloods.

We rephrased this in the section where we also addressed the comment on flow velocity, see comment [29].

Line 225 explains that an overview of the damage curves and supporting narratives can be seen in the supplemental document. In that document, there are figures of the damage curves with a description (narrative). However, it remains unclear how the curves were constructed. Commonly, damage curves are constructed from damage data, structural models or expert judgement. What method was used here *[32]*? In case, the method is expert judgement, an explanation of procedure, experts, validation should be given.

The damage curves were constructed by comparison of the costs of road construction and repairs (as tabulated in the SI) with the repair activities required after floods of different intensities. So they are initially based on a combination of expert judgment and actual road (re)construction data. The narratives and resulting curves were discussed with flood risk and transport modelling experts and discussed with the Dutch road operator.

We will add to 2.3.3 "To construct the curves, we compiled an overview of road construction, maintenance and repair costs. These costs were compared to the road repairs needed following a river flood, derived from literature and photo imagery of river floods in Europe. The damage narratives and resulting curves were validated during an expert workshop with flood risk and transport modelling experts and with the Dutch road operator, see acknowledgements."

Line 266 mentions the use of a model is it the Lisflood model? *[33]*

Indeed, the hazard maps originate from the Lisflood and Lisflood-FP model as discussed under [30]. Here, this is of minor importance for our argument, so we will rephrase as: "We masked the 1:10 year flood hazard map (which better resembled the reported inundation than the 1:100 year flood map) over the extent of the observed inundation and calculated the damage using the object-based model."

Section 3.1 emphasizes the need of a more in depth discussion about the suitability of the comparison. Again, the question arises about what the percentage of infrastructure refers to in the CORINE and LUISA data. *[34]*

Please see our reply to comment [27]: "To correct for the underrepresentation of infrastructural land use (and other land use types) in CORINE, for each land use class some percentage of infrastructure is assumed (Table S1,2). This follows from Huizinga's suggestion (2007, p. 2-22) to map his damage functions to CORINE using a cross-tabulation by the EEA (2006), which became their default implementation method. To enable a comparison with the object-based approach, we maintained the percentage of infrastructure per land cover class, whereas the contributions of the other damage curves were ignored (Table S1, 2). An implication of these percentages is that, although motorways and trunk are mostly missing in CORINE, damage for (local) roads in urban and industrial areas (amongst others) is still calculated, albeit without any explicit spatial reference to the actual road position, but based on their average presence in these urban and industrial land use types. Also, note that in the land cover category 'road and rail networks and associated land', a uniform share of 27% infrastructural land use is assumed, which to some extent corrects the overestimation of the actual road widths in LUISA.".

Figure 4. The value of the LUISA flood risk is the upper limit of the interquartile range. I believe that this result should be discussed, considering the uncertainty is LUISA less useful? In terms of hotspots is LUISA providing significantly different results? *[35]*

We were also surprised by this result, so we also think it is interesting and worth discussing. We will add to line 420: "indicating that using LUISA, a fair proxy of the total damage to road infrastructure can be obtained". The LUISA results cannot be easily visualized in a way that allows for hotspot identification for systematic comparison to our object-based approach. In general, one would always need a map of road network to compare the rasters to, for identification of flood hotspots.

Line 328. From this line on, the results refer to the object-based approach? Please clarify. *[36]*

As the caption above this subsection indicates, all results in this section refer to the object-based approach combined with the new damage curves (unless where otherwise stated, for comparison *purposes). We will rephrase line 238: "Figure 5 shows how the risk (deterministic estimate) is geographically spread over Europe." which refers back to the earlier mentioned deterministic estimate of the object-based approach with the new damage curves.*

Line 425 states that the Huizinga infrastructure function is a fair proxy for the average damage to road assets but is unsuitable for assessing damage at the individual road level. Is this to be expected? Are the Huizinga functions intended for damage at the individual road level? Here scale issues may have a role. *[37]*

This could indeed be what one expects from the original purpose of the Huizinga's damage function, which was developed for the JRC's inundation model combined with the CORINE land cover classification. It nevertheless is an interesting insight, because we have shown that the CORINE – infrastructure combination has several pitfalls (such as attributing infra damage to the wrong land use types). Moreover, CORINE is now complemented by LUISA, which is constructed in a fundamentally different way (with vector-based information data added to the grid, rather than just remote-sensed data). Also, many have used Huizinga's damage functions on scale levels beyond the original use, for example Jongman et al., 2012. Jongman suggested that the infrastructure function highly underestimated the damage, an observation that we have questioned on the basis of the results of our study. So although some might think that this is what was to be expected from Huizinga's curves, other's might have held a different view.