

Authors' responses to reviews of egusphere-2025-5161 submitted to *Natural Hazards and Earth System Sciences*

Manuscript title: “Combining hazard, exposure and vulnerability data to predict historical North Atlantic hurricane damage”

Authors: Alexander F. Vessey, Alexander J. Baker, Vernie Marcellin-Honore, and James Michelin

To the editor:

We are grateful to the editor and reviewers for the time they invested in our submission to provide a helpful critique of our study and its suitability to *Natural Hazards and Earth System Sciences*. Having considered the reviewers' feedback, we now submit a revised manuscript and present the reviewers' comments here, each followed by our response in **bold, green type**. We have accepted almost all the suggestions and recommendations made; there are only a few instances where we felt a clarification sufficed. All references cited here are also cited in our revised manuscript. We look forward to editorial and reviewer feedback on this resubmission.

Review #1

RC1: '[Comment on egusphere-2025-5161](#)', Anonymous Referee #1, 03 Dec 2025

Citation: <https://doi.org/10.5194/egusphere-2025-5161-RC1>

The objective of the study is very worthwhile. The Saffir-Simpson scale has long been recognized as lacking and less-than-optimal to fulfil its aims of hurricane warning. Therefore, an effort to substitute or accompany that scale with another metric could very much benefit society and is very welcome. This study summons a multi-pronged approach to address this, consulting and combining a vast range of datasets for an ambitious multi-variate analysis of the drivers of hurricane damages, finally motivating and proposing such an alternative scale. I think the study is overall well planned and executed; however, critical gaps in the presentation of its data and methods prevent it from being publishable at this stage. I recommend major revisions, and I would like to review a new version, if the authors will produce it, such that I can more fully evaluate the Results and Discussion sections, that I cannot fully evaluate now due to unclear Data and Methods.

We are grateful for the reviewer's positive assessment of our study's aims and execution, and the feedback provided about the clarity of our methodology has helped us improve our manuscript significantly.

Main points:

- The introduction is generally well-written and the reasoning and rationale are well-formulated.
- The study uses a very large number of datasets to explore the most informative variables. It seems to be quite exhaustive in this respect. The method for back-extending size data seems well devised and valid.

We are of course pleased to receive this feedback.

- The section on damage/loss data needs major re-working as it is very unclear. Other parts of Sections 2 and 3 also need major improvements. See my detailed comments below.

We have made a number of revisions to these sections to improve clarity. We thank you for your detailed comments. Please see our responses below to each of the recommended points.

- The data per each landfall should be included in a table, potentially in the appendix or in an appropriate, permanent online location associated with the manuscript. The Github link included in the Data availability section does not yet include data. This needs to be addressed in the rebuttal.

Yes, we recognize the value of sharing this data and will upload data and analysis code to the public Github repository.

- At some point in the manuscript it becomes indirectly clear that it is about hurricane landfalls and damages in the United States, not in the North Atlantic. This is a problem, and it should be made immediately clear, starting from the title and the abstract.

Yes, we agree with the reviewer here. We have revised the language here to make it clear that we focus on United States landfalls and not North Atlantic landfalls. The title has been changed to ‘Combining hazard, exposure and vulnerability data to predict historical United States hurricane losses. In addition, where we previously said ‘North Atlantic hurricanes’, we have replaced this with ‘U.S. landfalling hurricanes’.

- A point of reflection for the Discussion: the distinction between overall impacts and damages. A limitation of the new scale is that it does not say anything about non-economic impacts (e.g. lives). Did the S-S scale do better in that respect? Also, the new scale will weigh more heavily in favour of highly exposure areas, to the disadvantage of sparsely populated ones. In that respect the S-S scale is more fair. How do you deal with this? It would seem right to acknowledge that other hazard-only-based scales (such as some reviewed in the intro), could improve on the S-S scale in this respect. More in general, maybe a small discussion is warranted, of how different needs from different stakeholders can be addressed by either solution.

This is a very good point. We focused on financial loss, but we agree that a complete understanding of impacts should include fatalities. We have added additional discussion of this point (lines 540-542). The response of communities to early warnings (and whether they are able to receive these at all), and the factors that determine who is able to—and who does—evacuate prior to landfall or migrate post-landfall are complex (see, for example, discussion in Mustafa et al., 2023, <https://doi.org/10.1016/j.ijdrr.2023.103726> and Wong-Parodi and Garfin, 2022, <https://doi.org/10.1088/1748-9326/ac4858>). A full consideration of these complexities is beyond the scope of our study, and we designed our study to focus on economic loss. Within this analysis, we determined population density impacted by each hurricane by totalling gridded WorldPOP data within the hurricane’s R34 at each time step. Historical hurricane loss and impacted

population are correlated (Spearman's coefficient 0.57). It is possible in principle to focus on fatalities as the predictand, but that should be another study, requiring expertise in human behaviours and impacts.

Detailed points

Abstract

The passage “[the S-S scale is] an effective public communication tool” seems to purely pay lip service to the tool. This is repeated verbatim in the introduction. Nothing specifically about this scale seems to be particularly increase its effectiveness. Please consider rephrasing or skipping this – up to you.

We think it is necessary to distinguish between the Saffir-Simpson scale's use for public communication during a hurricane event and how well it reflects hurricane impacts and losses. We have revised this to be more clear and added a reference to Cass et al. (2023) to support this statement. We do not wish to give readers the impression that the scale has no use. Revision reads: “The widely used Saffir–Simpson scale is an effective public-communication tool, but is based on a single hazard quantity (wind speed) and therefore has little utility for predicting economic losses.”

“limiting its predictive value”: I would recommend adding “and early-warning value”.
This has been added.

The sentence starting with “The model significantly reduces” contains a seeming repetition, about comparisons with 1) a model “using landfall wind speed” and 2) a model of “single-parameter predictions, including landfall wind speed maxima”. What is the difference between the two?

Revised accordingly.

The closing sentence promises transferability to other regions. This does not seem supported in the article, and I suggest either making a convincing case for this in the Discussion, or eliminating.

We have removed “is transferable to other cyclone-prone regions”

Significance statement

Does the second sentence need to include both terms “impact” and “losses”? that's potentially confusing. There are more instances in the manuscript where loss and damage seem to be used interchangeably: please also address this.

Revised accordingly.

Section 1

The NOAA 2024 reference does not seem to inform about damages of year 2025.

Thank you—corrected to “Between 1980 and 2024”

Sentence starting with “The most important threats” lacks a coordinating verb.

Revised accordingly.

The explanation of why hurricane damages are more challenging or uncertain compared to TC activity is not clear or logical; please check this. Besides, please be explicit about the zoom in into hurricanes, from tropical cyclone. Besides still: this sentence is repeated two times, in this paragraph!

This has been revised. Damage assessments are difficult, as damage can vary greatly per building, and many buildings can be impacted by a single storm. The explanation of why hurricane damages are more challenging or uncertain has been revised to reflect this.

Please clarify “economic financial damage”.

We have made the terminology used simpler and more consistent, now using “(economic) loss” throughout.

Correct to: “is a critical step in mitigating...”

Revised accordingly.

Line 68: please spell out what you mean by “intensity”, as it needs to be clear in this context.

Yes, we agree. “Intensity” has been changed to “maximum wind speed” - as per Chavas et al., 2025.

The overview of previous efforts to improve on the Saffir Simpson scale is well written and accurate. Please consider contending with other potentially relevant efforts. To my knowledge: Tripathy et al. 2024 (<https://doi.org/10.1038/s43247-023-01198-2>) and the simple mean sea level pressure metric in the elsewhere cited paper of Klotzbach et al. 2022a. Further, it would be coherent to add a remark about the implications of Pilkington and Mahmoud (2016) and of Baldwin et al. (2023) for an alternative intensity scale (the focus of the paragraph).

Thank you—we have cited Tripathy et al. (2024) and added a statement to comment on these studies and link to the following paragraph that introduces our study’s aims.

The sentence “These recent studies add to a growing body of evidence that combining these factors is necessary to capture risk (Ward et al., 2020)” is almost repeated from one of the prior paragraph, including the same reference.

Agreed—we removed this sentence.

Line 95-on. Please revise syntax of this sentence. More in general, in this sentence and part, I think you could make the point more clear, that a ‘usable’ new scale should also fulfil the need for rapid implementation/computation, which is relevant for your methodology.

Revised accordingly.

Research questions are a bit redundant and not very useful. The third one, in particular, does not seem appropriate, since the reader knows nothing yet about this ‘Hurricane

Predictive Damage Scale'. In the second question, it is not clear what 'more' refers to. Please reconsider.

Our research questions have been revised accordingly. We have rephrased them to include more context as to why they could be important to address.

Section 2

“(section 2.2)” is repeated in short succession.

Thank you—corrected.

In this section, you start to deal with “losses”, whereas so far you had dealt with “damage”. Please harmonize this or clarify the difference.

We are grateful for this comment and now use “loss” consistently. See additional definition (line 97-98).

Line 115: the list of loss sources does not correspond 1-to-1 to that in figure 1 (very useful figure, by the way!). It seems to me that it should. Please also check for the other categories of data.

We are grateful for the reviewer’s positive assessment of the usefulness of this figure. This has been revised accordingly to ensure Figure 1 includes sources of data for all data categories.

In this part, you should explain, briefly, what “normalized” means, in general, and – if needed – in each different study. Normalization is further expanded from line 149, but there is no overall explanation of the purpose it serves and it comes too late. This is not trivial, as there seem to be different concepts of normalization, justifying different methods and supporting different results and conclusions. I do not specialize in this aspect, but it seems that the literature on the topic of trends of normalized hurricane losses is rich of discussions and fraught with implications. Data and methods across publications and datasets has fundamental differences, and it seems quite complicated to harmonize across them. I commend the authors for their effort here. But I am very confused about the criteria and methods for the selection of the data across sources. It is important that this is done transparently and clearly, and this needs strong improvement. For example: Blake et al seem to report disasters from 1851, why do you write that loss estimates are available from 1965?

The reviewer has raised an important point here. We agree that normalisation, and how this differs among studies, contributes to uncertainty in historical loss estimates. In our study, we do not treat any particular historical loss dataset preferentially, and we have revised the text (lines 136-146) to explain this. We instead calculated the average loss per hurricane and use a single approach to normalisation across datasets. The loss datasets used are the first aspect described in section 2, and we have shortened this section overall and made several revisions for clarity (from line 166 onwards). A summary of normalisation is given (lines 166-174) and we state that there is uncertainty due to normalisation differences between datasets. It is our view that evaluating the strengths and weaknesses of each normalisation methodology is beyond the scope of this

submission. However, our revisions were made to make the rationale for our unified approach clear.

What are the differences between that source and National Centers for Environmental Information (2025)? What are the implications that the “Billion dollar loss record” only include records past that arbitrary threshold (unlike, e.g., EM-DAT)?

There are several differences between the methodologies of each historical loss data source. However, for brevity and clarity, we do not go into this in detail, and instead provide a summary of each dataset, including the number of historical storms included since 1979 (see Table S1). The impact of having an arbitrary threshold is that some weaker storms may not be included, so including data (e.g., EM-DAT) provides data for weaker storms. Yes, the methodologies of each historical loss data source differ, but this manuscript needs to be shortened, as recommended by reviewers, so adding more detail is not ideal in this case.

What does the sentence at line 134-136 mean? Are EM-DAT and Delforge et al. (2025) the same source? If so, do not use both terms interchangeably. Why do you remain with a time series that starts at 1979 (table 1)?

This raises an important point of clarity. We have cited Delforge et al. (2025) in section 2.1 and now use the dataset name ‘EM-DAT’ consistently throughout the manuscript. We have also revised the sentence highlighted here. On the point about the time series beginning in 1979, our justification is that rainfall and building density data go back to 1979. If we omitted these variables, we could include storms back to 1950 in our training data. However, further back, fewer data are available for each storm, and uncertainties are higher. We chose 1979 as an optimal start year because this is the satellite era, which is considered a more consistent and reliable period for analysis. We have made this clear in our

Line 115: “Historical hurricane economic loss estimates were collated from various government agencies and published studies” and line 127: “We collated hurricane loss estimates from multiple sources”. Please avoid confusion and write this methods step in one place. Actually, the first 2 paragraphs of section 2.1 should be reorganized: as they are they report similar facts about each dataset in different places and in no particular order.

Thank you—these two paragraphs have been revised accordingly (first paragraph of section 2.1.1).

Please check if you can slightly improve the explanation of how you handle the complication that different sources treat differently data losses from hurricanes with multiple landfalls, lines 137-143. I think I understand it, but I wonder if clarity can be improved.

Yes, we agree. This has been revised accordingly. We have described the example of Hurricane Katrina (2005) (Figure 3), which made landfall in Florida, then travelled across the Gulf of Mexico to then make another landfall in Louisiana. See the revised text:

“If a hurricane makes multiple U.S. landfalls, Weinkle et al. (2018), EM-DAT, and Blake et al. (2011) report total losses aggregated across the storm track, whereas Grinsted et al. (2019) and Muller et al. (2025) report losses separately for each landfall. For example, Hurricane Katrina (2005), which made landfall in Florida and Louisiana (Figure 3), has two per-landfall loss estimates from Grinsted et al. (2019) and Muller et al. (2025), but has aggregated losses in Weinkle et al. (2018), EM-DAT, and Blake et al. (2011).”

I am confused about Table 1: where are the named storms in it, and why are they relevant? Same for “bypassing” hurricanes.

Revised accordingly.

At line 154 you explain that you normalize the damage data. But each of the source datasets already applies some sort of normalization, likely each in a different way. Why do you normalize again, are the differences across sources taken into account, or do you base your normalization on the raw data pre-normalization? This should be very clear, so that the reader can understand if your methods of normalization serves the scopes of this study.

Thanks—we see the confusion here. We have revised the text accordingly within this paragraph. Each dataset provides un-normalised and normalised historical loss estimates, and each follows its own normalisation approach, including differing reference periods depending on their publication year (e.g., Weinkle et al. (2018) is normalised to 2017 and Muller et al. (2025) is referenced to 2022). We have applied a single normalisation approach to the un-normalised loss estimates from each study, keeping consistency. It is necessary to apply our own normalisation approach to ensure that loss estimates are representative of the present-day and to minimise errors due to the reference year of each previous study.

Line 170: please specify the temporal duration of the maximum wind speed (1-min, 10-min, etc).

Revised accordingly.

Line 171: it shouldn't be necessary to specify “(i.e., beyond RMW)”, if the RMW and R34 etc are defined.

Revised accordingly.

Line 173: at the timestep before the storm centre crosses over land, the effect of the land can be already present in large sectors of the hurricane (depending also on the timestep size, which you could specify). Reword to, e.g., “atmospheric fields are minimally impacted”.

Revised accordingly.

Why are data from HURDAT2 prioritized over IBTrACS? What happened to hurricanes for which track data lack in both datasets (as implied by “if available”, line 177)? Did you check for, or are there reports on inconsistencies across the two datasets?

There are some data inconsistencies between the HURDAT2 and IBTrACS. We chose to use the former to allow us to incorporate the latest additions and

revisions to historical hurricanes from the HURDAT Reanalysis Project (https://www.aoml.noaa.gov/hrd/data_sub/re_anal.html), particularly the RMW data added in 2024.

Citation is missing for the Global Tide Surge Model.

We have added this.

Line 184: “simulated storm-tide level”, for consistency with the prior terminology. “Storm surge may be larger in the hours before or after a hurricane makes landfall, depending on antecedent tidal height”. This seems incorrect, according to general terminology (e.g., <https://oceanservice.noaa.gov/facts/stormsurge-stormtide.html>): storm surge only depends on meteorological forcing, not on tidal phasing (or only minorly and indirectly). Probably here you mean “storm tide”.

While for all other hazard-related variables you take instantaneous data, for rainfall you also take accumulations: why is that? Even if accumulation rather than instantaneous intensities is plausibly tightly related to damages, why using accumulation at locations far from the locations of damage, as done by integrating accumulations along the whole track? Why 500 km radii here?

Yes, ‘storm tide’ is the correct terminology here. This has been revised accordingly. As the reviewer points out, for some hazard variables, instantaneous occurrence is more important for damage, whereas, for other hazard variables, accumulation can be more important for damage. It is for this reason that we include both instantaneous and accumulation variables for hurricane rainfall (see Figure 8 includes both ‘Max MSWEP Rain Accumulation’ and ‘Max MSWEP Rain Rate’). The choice to use a 500 km radii is arbitrary, but we know that hurricane rainfall footprints are different to wind footprints, hence wind radii values of R34 are not so applicable here. Our choice of radius follows previous research (e.g., Stansfield and Reed, 2023, <https://doi.org/10.1038/s41612-023-00391-6>), but we acknowledge that this may lead to an overestimation of hurricane-related precipitation (Stansfield et al., 2020, <https://doi.org/https://doi.org/10.1175/JHM-D-19-0240.1>). We have added this clarification to the text (lines 207-208).

Line 195: revise syntax. Also, how is rainfall data integrated with MSWEP?

Revised accordingly. Rainfall data from NOAA (<https://www.wpc.ncep.noaa.gov/tropical/rain/tcrainfall.html>) were not integrated with MSWEP. They are used as independent datasets and the text has been revised to state this clearly.

Line 203: “vary between the two datasets”

Revised accordingly.

Line 210: it seems that population density data for a year 1979 landfall come from WorldPop of year 2000. The assumption of population stationarity across a 21-year period seems problematic. One would wonder whether it is not best to discard population density data altogether, and avoid artifacts introduced by this limitation – also considering that the study focuses on economic damages, not on human impacts. More in general, data used combine time-varying and time-invariant datasets. This

seems to be correctly stated for each dataset, and fig. 1 summarizes this visually. However, this combination can clearly introduce artifacts in the results. A short reflection on this aspect, maybe at the beginning of the Data section, could clarify the expected impacts of this; and this should be revisited in the Discussion section.

We appreciate these comments and agree that assuming population stationarity over a 21-year period is an important limitation to highlight (see added text at line 240-1). The absence of open-source, time-varying historical data necessitated this assumption. (There are data development activities ongoing, e.g. <https://doi.org/10.48550/arXiv.1712.05839>). Using both time-varying and time-invariant datasets may introduce artifacts in our results. As suggested, we have reflected on this in our discussion, drawing attention to the need for improved data availability for risk research (lines 589-99).

L 214: please define better the Hurricane risk score: vulnerability and resilience are indirectly correlated, so how are they together in the same score?

We have revised the text in section 2.1.4 to include these definitions. This risk score is produced within FEMA's National Risk Index Framework. "Community resilience" is defined by National Institute of Standards and Technology as the ability of a community to prepare for anticipated natural hazards, adapt to changing conditions, and withstand and recover rapidly from disruptions. "Social vulnerability" is defined as the susceptibility of social groups to the adverse impacts of natural hazards, including disproportionate death, injury, loss, or disruption of livelihood, and considers the social, economic, demographic, and housing characteristics of a community that influence its ability to prepare for, respond to, cope with, recover from, and adapt to environmental hazards. Overall, "Community resilience" is proportional to a community's ability to react, whereas, "Social vulnerability" is proportional to more chronic factors.

L 218: "and we averaged these two variables across...". Besides, what is the footprint here: R34, 500 km, or other?

This refers to each hurricane's R34 footprint, and has been revised accordingly.

L 226-on: what is V_n ? There is no V_{max} in eq 3. Why you use V_{half} of 140 knots, if Vickery et al. suggest lower values? Why V_{thresh} of 40 knots?

Yes, 'Vmax' has been removed from the formula description.

We have added additional text to describe why we chose these values for V_{half} and V_{thres} . Vickery et al. (2006) and Hazus et al. (2009) suggest that v_{half} can vary between 120 kts and 160 kts, depending on building characteristics. The lowest category of the Saffir-Simpson Scale (i.e., Tropical Depression) has a lower wind speed bound of ~35 knots, and past Tropical Depressions have caused damage in the past (e.g., Tropical Depression Allison in 2001). However, Hazus et al. (2009) suggest that V_{half} is more likely to be ~50 kts. In this study, damage estimates were computed for historical hurricanes using this single vulnerability function for all buildings, with $v_{thres} = 40$ kts and $v_{half} = 140$ kts. The text has been revised here to reflect this. (The additional Hazus reference is available to be viewed here: [Hazus wind damage functions — physrisk documentation](#))

L 235: “At each timestep, v_{max} is used with Eq. 2 and Eq. 3 (Emanuel, 2011) and the extracted exposure value and building density”. What does this sentence clarify, further than the prior explanation?

We have added “thereby allowing v_{max} to vary in time” to clarify this.

On the estimation of size for the older period. The method is mostly well documented, but some details are missing. From which datasets are v_{max} and the other physical variables taken? How do you obtain 4220 observations, if you have 134 hurricanes in table 1 (I imagine you took also non-landfalling, and you took multiple timesteps per track, please explain)? What are nm? From line 252 you move to R50 and R64, but seem to also discuss aspects that are relevant also for R34: e.g., that RMW is incomplete during 1979-2002: why? What is the difference between “estimates from HURDAT2” and “reconstructions from Gori et al.”?

This comment raises where clarifications are needed. Within IBTrACS, we omitted all timesteps (rows) with NaNs across all R34, V_{max} , RMW, C_p and latitude rows. This leaves 4,220 timesteps, which we used to train our model, and then test it with a leave-one-out validation method. This clarification has been added to the text. “nm” stands for nautical miles, which is the unit used in IBTrACS. This has been added to the figure caption. Out of the inputs to our R34-estimation model, RMW is the quantity less available within IBTrACS (than V_{max} , RMW, C_p and latitude). As RMW was found to be the most influential input variable to our model, it is necessary to either i) obtain it from HURDAT2 or ii) use the corresponding value at the same timestep from Gori et al., who reconstructed RMW by combining the TC wind model of Chavas et al. (2015) and ERA5 data. We have revised the text to make this all clear.

On fig. 2: how can you include observations from 1979 here (as per caption), if observations start in 2002? What does “model observed” mean? Why are the correlation and MAE different between legend and main text? Why did you include the trendline in blue and did not present it? It suggests that predictions systematically underestimate R34, maybe something to mention in the main text. Lastly: figures generally don’t need titles, and info should be in the caption.

Yes, the reviewer is correct. We have changed the figure caption from “1979-2023” to “2002-2023”. “Model observed” has been changed to “observed”. The main text has been updated to reflect the values in Figure 2. We have added a description in the main text of the trendline. Yes, there is a slight underestimation of R34, but this is minor.

“Where RMW is missing from IBTrACS, RMW is replaced by values from HURDAT2 or Gori et al. (2023)”: this seems redundant with the prior sentences.

Revised accordingly.

Line 257 “RMW values from the previous timestep were used” and line 261 “RMW observations from previous timesteps were used”. Please check this explanation, as there seem to be something redundant of wrong here.

Revised accordingly.

Across the manuscript TC and hurricane are often used arbitrarily for the same concept. Please check and harmonize to one term.

Revised accordingly.

L 281: “In this study, we used a weighted combined-rank framework, linear regression framework and the random forest decision-tree framework to combine input predictors across hazard, exposure and vulnerability to predict historical hurricane damage.”

Please improve this important sentence. What is combined here: frameworks, inputs? “Inputs” and “predictors” seem redundant, correct? More in general, please clarify in this section how three, very different approaches are combined in your model: this is missing. Suggestion: “Our target prediction variable is damage for each hurricane, averaged across the datasets presented in section X”.

Yes, we understand the confusion here. We test 3 independent statistical approaches using the same set of inputs to predict hurricane loss. This sentence has been revised.

What do you mean with “maximising the sample of hurricanes for which a loss estimate is available”? More in general, I don’t understand the difference between table 1 and 2: please clarify, and if possible, consolidate in one table, with one extra column. You seem to have 106 hurricanes for which all data are suitable, out of 134 hurricanes for which damage data are suitable. Are the 28 non-overlapping hurricanes entirely discarded from the analysis? If so, I suggest they shouldn’t feature in any table 1. Further, it is not clear why you deal with “named storms” in the caption of table 2, whereas the column header deals with TCs. Also the caption says damage, the column header says loss.

There are several helpful points here. The sentence referred to here simply refers to using multiple loss datasets to cover a larger number of historical landfalls. (If we only used loss estimates from Muller et al. (2025), for example, we would sample only 33 events.) This has been clarified (line 312). As we used loss information from multiple sources, our sample size is 106 unique historical hurricanes, thus ‘maximising’ the amount of data that can be used to train and test our statistical model. Additionally, as suggested, we have combined tables 1 and 2, corrected the column header, and “named storms” has been replaced with “hurricanes.” The 28 non-overlapping hurricanes are discarded from the analysis because we do not have all of the input variables. In most cases, we lack an R34 estimate, even lacking the ability to estimate R34 with our model (Figure 2), because either RMW, Vmax, Cp or Latitude is missing. In most cases RMW is missing, even when we try to find the corresponding values in HURDAT2 or Gori et al., or we even use the RMW value from the preceding timestep. Through our analysis, we have identified that there are numerous amounts of missing data, where even reconstructions or estimations aren’t available. We highlight data needs in the conclusion section.

Line 296: “To reiterate: a key aim of this study, to develop an approach to estimate expected damage for future forecast landfalling hurricanes.” This is superfluous, if you deem that the aim is sufficiently clear from the introduction – as it should be.

This has been removed.

Line 314: this sentence is unclear, in the light of the preceding explanation “Linear and normalised input variable ranks were derived”. In the sentence thereafter, what are “alike loss ranks”? please make sure that sentence is clear.

Revised accordingly.

Fig. 4. Please separate the normalized and non-normalized losses into two separate columns, instead of stacked. Or are red and blue proportions of a total? Unclear. The titles of sections 3 and 4 should be harmonized, as they are they don’t clarify how their content is organized: “Historical relationship between hurricane v_{max} and damage”, “Relationships between historical hurricane damage and risk-related variables”. What are risk-related variables? You have not introduced them. Isn’t v_{max} risk-related, and if so why is it presented in a separate section?

We appreciate the feedback on the subheadings and have reworded these for consistency. Fig. 4 has been revised as suggested.

Line 417: compared to what does the random forest improve skill? I imagine compared to the raking with single hazard-related variables. Please be explicit.

Thanks—this was unclear. We have clarified this (line 426) and the improvement to which we referred original is in line 427-429.

Fig. 6: the titles of both panels is the same. Correlation in the legend is also an indication of the goodness of fit.

Thank you—corrected.

Fig. 7: the titles of both panels is the same (please check across the manuscript this recurring problem). The description of panels in the caption does not seem to correspond to the axes titles.

Thank you—corrected.

Fig. 8: There needs to be a table that clearly explains each dataset, their reference and matches is with its abbreviation used here. This could be consolidated/integrated with/in fig. 1. Why do you only use the NOAA financial loss values here, instead of the multiple sources described in the data and methods sections? Maybe add a brief clarification in the caption, that IBTRACKS cp is strongly anticorrelated because its relationship to TC intensity and damage is inverse. Up to you.

We have added a supplementary material table, with acronyms and sources. ‘NOAA’ was a mistake in the caption and has been changed to ‘average loss’. A note has been added to the caption to explain that IBTrACS Cp is anti-correlated with damage, so it has a negative correlation.

Section 7 on Summary and conclusion should not include a section on “key results”, as per header of 7.1. True to the header, the first two paragraphs of section 7 are a

summary of the study. I question however how useful this is, given that there is an abstract already. I suggest to shrink this part to the minimum necessary to follow the main points argued in the rest of section 7.

As suggested, we have shortened this first part of section 7.

Line 521: If I am not mistaken, you do not use variables of inland flooding and coastal flooding. You use storm surge, which is much different.

Thanks—we have removed “inland and coastal” to avoid confusion about terminology here.

Review #2

RC2: 'Comment on egusphere-2025-5161', David N. Bresch, 06 Jan 2026

Citation: <https://doi.org/10.5194/egusphere-2025-5161-RC2>

The paper is well written, but would greatly benefit from substantial shortening. I do not understand why one undertakes such an effort for just one particular hazard in one particular region, while global high resolution catastrophe models exist and have been calibrated for many hazards and regions etc. (just e.g. for TC Eberenz et al. 2021, in addition to the cited Eberenz et al., 2020) and even take climate change into account (Meiler et al. 2023 and 2025).

The reviewer has highlighted two important things here. First, the need to be more concise, and we have made a number of revisions throughout to achieve this (see also comments from the other reviewers), including removing repetition and moving Fig. 7 to the supplement. Second, the reason for our focus on US landfalls is the greater availability of open-source data for statistical analysis. This allows us to explore relationships between losses and different predictors, sampled within observed / estimated hurricane footprints. We aimed to develop an approach that is in principle applicable globally. Our study is complementary to Eberenz et al. (2021)—we have now cited this in the introduction, along with the other papers referred to here (see 3rd paragraph). We have added discussion of Meiler et al., 2023; 2025 to the discussion (section 7.1.1.).

The value of the proposed Hurricane Predictive Damage Scale' (HPDS) is quite limited, as it can only complement the Saffir–Simpson category. Based on my more than decade-long working experience in risk management in reinsurance, I would not have seen value in such a HPDS score compared to an ensemble-based (Buizza, 2006) catastrophe models output back then. A cat model takes the bespoke portfolio of insured assets and their pertinent vulnerabilities (different asset classes etc) into account. Further to that, it is not the 'from ground up' loss that is of most relevance to industry players, as the financial cover conditions, i.e deductibles of primary insurance and attachment end exit points of (re)insurance programs do (obviously) matter quite a lot.

We appreciate the reviewer's sectoral experience here. Our scale is intended to complement the Saffir-Simpson scale. We have clarified this in the part of the introduction that discusses hurricane classification schemes. Our scale is intended to satisfy a need within insurance and related sectors for an easily interpretable classification of a hurricane's likely losses prior to landfall, which is computationally inexpensive. It should also be useful to responding agencies to place a forecast landfall within the context of historical losses in an open-source way, building on from a few studies that have attempted to improve upon the S-S scale's shortcomings in this regard.

The paper would greatly benefit from a comparison of the proposed HPDS with catastrophe model estimates of damage, e.g. using open-source models such as CLIMADA (Aznar-Siguan and Bresch, 2019) or OASIS LMF (<https://oasislmf.org>).

Thank you—a similar comment also made by another reviewer. We understand the reason for this suggestion, but it is difficult to compare with an industry standard,

as many insurers use proprietary models or those of, for example, RMS, which are not the same as the open-source models currently available. There is evidence that open-source models underestimate losses significantly for different storm types. We have included discussion of this in the revised introduction and outlined the need for a full inter-comparison of catastrophe models in the discussion section, which would be hugely valuable but is beyond the scope of our paper.

Page 2, line 61-63: You write “complex and high resolution catastrophe models (e.g., Florida Cat Model) are developed, but these are computationally expensive and unsuited to estimating losses for a forecast TC, as forecasts evolve on sub-daily timescales.” Really? I do not perceive standard impacts models (Aznar-Siguan and Bresch, 2019) as being computationally expensive nor unsuited to estimating losses, in particular not for tropical cyclone impacts based on e.g. ECMWF ensemble predictions (Davidson et al., 2020; Kam et al., 2024). Please explain. It would have been interesting to calibrate the Kam et al. (2024) model to US damage, which would only entail switching from a population to an asset exposure layer (Eberenz et al., 2020) and recalibrating the model (cf. Riedel et al., 2024).

We have revised this statement (see paragraph beginning line 82). Please also see our related comment above. Additionally, our method combines population and exposure factors, and it is the skill gained from doing so which we present in our manuscript. Comparison with the model suggested by the reviewer would be interesting, but involves substituting different predictor layers, and therefore a somewhat different approach.

Page 4, line 97: “...to make appropriate preparations and (re-)insurance...”. Please be more precise. Preparations and (re)insurance are quite different tools. (Re)insurance has to be agreed upon far ahead (usually annual terms, renewal annually). There is ample literature on the subject, including the incentivising effect of (re)insurance on prevention, as insurance premiums might be lower in case of demonstrated efforts in prevention and preparedness. And there isn't the risk-sharing/pplong effect of endurance which diversifies risk and lowers capital costs (e.g. Ciullo et al. 2023).

Thank you—this comment highlights an ambiguity. We have revised this to be more clear. We have also added brief statements on pooling to the discussion and cited Ciullo et al. (2023) (lines 600–603).

Page 3, line 590ff: You might consider to briefly reflect on Meiler et al., 2023 and Meiler et al., 2025.

These are very helpful reference suggestions. We now include both in our introduction.

References as mentioned:

- Aznar-Siguan, G., and Bresch, D. N., 2019: CLIMADA v1: a global weather and climate risk assessment platform, *Geosci. Model Dev.*, 12, 3085–3097. <https://doi.org/10.5194/gmd-12-3085-2019>

- Buizza, R. The ECMWF ensemble prediction system. in *Predictability of Weather and Climate* (eds Hagedorn, R. & Palmer, T.) 459–488 (Cambridge University Press, Cambridge, 2006).
- Ciullo, A., Strobl, E., Meiler, S., Martius, O., and Bresch, D. N., 2023: Increasing countries financial resilience through global catastrophe risk pooling. *Nature Communications*, 14, 922. <https://www.nature.com/articles/s41467-023-36539-4#Sec7>
- Davidson, R. A. et al. An integrated scenario ensemble-based framework for hurricane evacuation modeling: part 1—decision support system. *Risk Anal.* 40, 97-116 (2020).
- Eberenz, S., Lüthi, S., and Bresch, D. N., 2021: Regional tropical cyclone impact functions for globally consistent risk assessments, *Nat. Hazards Earth Syst. Sci.*, 21, 393-415, <https://doi.org/10.5194/nhess-21-393-2021>
- Kam, P. M., Ciccone, F., Kropf, C. M., Riedel, L., Fairless, C., and Bresch D. N., 2024: Impact-based forecasting of tropical cyclone-related human displacement to support anticipatory action. *Nature Communications*, 15:8795 . <https://doi.org/10.1038/s41467-024-53200-w>
- Meiler, S., Ciullo, A., Kropf, C. M., Emanuel, K., and Bresch, D. N., 2023: Uncertainties and sensitivities in the quantification of future tropical cyclone risk. *Nature Communications Earth & Environment*, 4, 371. <https://doi.org/10.1038/s43247-023-00998-w>
- Meiler, S., Kropf, C. M., McCaughey, J. W., Lee, C.-Y., Camargo, S. J., Sobel, A. H., Bloemendaal, N., Kerry Emanuel, K., and Bresch, D. N., 2025: Navigating and attributing uncertainty in future tropical cyclone risk estimates. *Sci. Adv.*, 11, eadn4607. <https://www.science.org/doi/10.1126/sciadv.adn4607>
- Riedel, L., Kropf, C. M., and Schmid, S., 2024: A Module for Calibrating Impact Functions in the Climate Risk Modeling Platform CLIMADA. *Journal of Open Source Software*, 9(99), 6755. <https://doi.org/10.21105/joss.06755>

Review #3

RC3: 'Comment on egusphere-2025-5161', Anonymous Referee #3, 15 Jan 2026 reply

Citation: <https://doi.org/10.5194/egusphere-2025-5161-RC3>

Overall-a worthy paper. I would recommend really trying to reduce the length and verbosity in many sections. My recommendation is accept with minor revisions, but it is close to major revisions needed (went back and forth).

We are grateful for the overall positive assessment here. The reviewer has highlighted the need to be more concise, and we have made a number of revisions to achieve this.

In the significance statement, the SS category is used to convey the wind impact, not the overall impacts, so just add the word wind before convey. I would also change the title to reflect on United States hurricane damage in the Atlantic basin, since it doesn't include international data.

**Revised title as recommended (and in line with another reviewer's comment).
Significance statement revised for clarity.**

Line 66- It is not used to predict how bad total damage will be. It is just wind.

Thank you—revised (please see also comment by reviewer #1).

Line 69 - Again the SS Scale isn't used to predict how bad a hurricane's damage in operations. It is used to give an indication of the expected wind damage. One has to look elsewhere for the others hazards, including storm surge and inland rainfall/flooding, which can be huge problems. It doesn't misrepresent anything, because that's not its intent to be a total all-encompassing scale. A hurricane could go from category 1 to 5 if it hits a large city using some of these total scales, with surge being very dependent on the exact track that isn't skillfully enough forecast (one county either side of the forecast track can change damage by tens of billions). This isn't useful information for the general public in preparation for a hurricane, but it is easy to see why insurance companies would like a way to predict the total damage using available information.

We agree that this statement was misleading. We have revised this to be clearer: “To characterise potential hurricane damage, there are calls to modify the Saffir–Simpson scale (Wehner and Kossin, 2024), develop multi-hazard equivalents (Tripathy et al., 2024), and adopt multidisciplinary (i.e., hazard, exposure and vulnerability) approaches to understand hurricane impacts (Ward et al., 2020; Camelo and Mayo, 2021).”

Line 115- The data sources in this section all provide a variety of ways to normalize the data. Did you take the raw from each dataset and then normalize it? There's also the normalization to 2024 dollars from raw (year of landfall) dollars. Anyways it is confusing to know how the various datasets were used because you can't exactly combine them verbatim without some work. There also an awful lot of overlap between those datasets, so they aren't as independent as suggested.

Additionally in this section - Blake (2011) was eventually superseded by data from the NCEI billion dollar disaster database in the official NHC data record in the tropical cyclone reports for a "one NOAA" estimate. Hopefully that is clear - the storms back to 1980 I believe were updated because they all had NFIP data included, though the overall table had data back to 1900. I will also note that the NCEI dataset is no longer being updated after 2024, so there is potentially some benefit in seeing your results without that dataset involved.

The review points out an important consideration here about this dataset. We think NCEI data remain a useful, even though it is currently discontinued. It is worthwhile including this dataset to increase the loss the number of historical events between 1979-2024 for which a loss estimate is available. Moreover, there are contributions to the literature on this dataset (<https://link.springer.com/article/10.1007/s11069-015-1678-x> and <https://link.springer.com/article/10.1007/s11069-013-0566-5>). We of course hope that data availability improves in future.

Line 137- Did you mean that the estimate is averaged from all of those sources if the source provided an estimate? It was a little unclear what you meant (note typo in that line as well at end). I will say some of the damage estimates have a pretty large discrepancy, Harvey being a good example of the difference in NCEI vs Weinkle- don't know how to account for that in your results.

The reviewer is correct here. We have clarified this in section 2.1.1.

Line 143- How about near-miss or near-shore hurricanes- bypass makes it sound like it missed, when it most certainly did not in the effects (which is what we are concerned with).

This is a good question and one we considered when designing our analysis. We have added the reason for excluding bypassing storms to line 182-184).

If the paper needs to be shortened (it does seem long), I'm not sure that Section 4 adds a whole lot. It does complement some of the work that Klotzbach did for landfall cp, but it doesn't seem as necessary as the rest of the paper.

We have made revisions in several places to shorten the paper, including moving Fig. 7 to the supplementary material, and so we have left section 4 (shortened) in for the reason the reviewer points out here: to complement published work, including that of Klotzbach et al. (2020).

Section 5 is interesting, but needs to be compared to other catastrophe models to test how reliable it is. Clearly many sources have shown that the SS scale isn't sufficient, but are we improving on anything that the (re)insurance industry and others don't already have? A comparison to an industry standard would strengthen the paper.

Thank you—a similar comment also made by another reviewer. We understand the reason for this suggestion, but it is difficult to compare with an industry standard, as many insurers use proprietary models or those of, for example, RMS, which are not the same as the open-source models currently available. There is evidence that open-source models underestimate losses significantly for different storm

types. We have included discussion of this in the revised introduction and outlined the need for a full inter-comparison of catastrophe models in the discussion section, which would be hugely valuable but is beyond the scope of our paper.

Line 491- Allison was a tropical storm, not a hurricane.

Thank you—corrected.