Reply to RC1: 'Review' of nhess-2020-299

We thank the reviewer for their helpful comments. Our replies are inline below:

The references are accessible and sufficient in number, although Hersbach (2020) should be cited for ERA5.

Corrected, thank you.

Is it possible to remain consistent in the units? Figures 1, 3, and 5 use m/s, Figure 4 uses knots, and Table 1 uses km/h.

Figure labels have been changed to use m s^{-1} by default, with additional knot conversions in some cases. Table 1 has been converted to m s^{-1}.

The title could be made more relevant, mentioning return periods / exceedance probability, and the generalized additive model, e.g. The use of generalised additive models to examine extreme wind return periods from tropical cyclones in Bangladesh.

We have updated the title to read: "Extreme wind return periods from tropical cyclones in Bangladesh: insights from a high-resolution convection-permitting numerical model".

Space between: 4.4 km

Corrected.

Line 66: Add 9-member: "generate 9-member ensemble simulations of 12 historical C2 tropical cyclone"

Updated.

Line 40: Change 'course' to 'coarse'

Corrected.

Line 58: Change 'combe' to 'comb'

This should be 'combine' – now corrected.

Line 194: Include in brackets what the resolution of ERA5 is, as low for this study but not low in comparison to other reanalysis products. "Given the relatively low ERA5 resolution (x)"

'(31 km)' added to this sentence.

Lines 210-211: Add a citation following this statement: "current generation weather forecast models are capable of accurately predicting the landfall location and track of tropical cyclones in the BoB many days in advance."

We have added citations of Mohanty et al., (2020) and Singh and Bhaskaran (2020) as examples.

Line 141: Change 'of high wind speed' to 'to high wind speed'.

Corrected.

Note in the figures that the town labels are the 18 most populated. Town names difficult to see in some of the figures.

'…with the 18 most populated towns and cities marked by circles.' added to the figure captions. We have made some subtle alterations to the label plotting whilst trying not to obscure to much of the underlying data.

Figure 2 caption: Change from "of (Augustin et al., 2012)" to "of Augustin et al. (2012)"

Corrected.

Figure 3: Superscript -1 for the units. Also change 'outlines' to 'outlined'.

Corrected.

Table 1: Change 'define' to 'defined'.

Corrected.

Citations

Mohanty, S., Nadimpalli, R., Mohanty, U. C., Mohapatra, M., Sharma, A., Das, A. K. and Sil, S.: Quasi-operational forecast guidance of extremely severe cyclonic storm Fani over the Bay of Bengal using high-resolution mesoscale models, Meteorol. Atmos. Phys., doi:10.1007/s00703-020-00751-4, 2020.

Singh, K. S. and Bhaskaran, P. K.: Prediction of landfalling Bay of Bengal cyclones during 2013 using the high resolution Weather Research and Forecasting model, Meteorol. Appl., 27(1), e1850, doi:10.1002/met.1850, 2020.

Reply to RC2: 'Review of manuscript' of nhess-2020-299

We thank the reviewer for their comments.

In general we note that the accompanying data validation paper (Steptoe et al., n.d.) is currently still under review with a different journal. As we feel this paper goes some way to addressing some of the concerns of the reviewer, we have attached it as a supplement to this reply.

Our replies are inline below:

Major comments:

1. What is the study trying to achieve?

Fundamentally, the paper aims to improve our understanding of wind (gust) extremes in Bangladesh from two perspectives: (1) improved spatial understanding, in terms of the spatial extent of wind extremes; (2) improved temporal understanding, in terms exceedance probabilities. Section 1 highlights a number of problems encountered in the current literature for this region, including: (1) a focus on the limited historical record (L51-54); (2) a focus on parametric wind field models that are computational expensive and often specifically tuned to particular TC basins (L54-57); and (3) take a holistic hazard approach, but which are then limited to examining specific events (L57-61). In all cases, the limited observational record adds additional constraints.

Section 3 summarises the state of gust hazards, based on 12 events, in terms of exceedance likelihoods and return periods and highlights the spatial variability of the hazard. We expect this information would be of use to long-term mitigation planning (on the order of years or decades). Section 3.1 focuses on a robust, transparent way to make decisions when information is uncertain.

Why is the problem formulated this way — what is the value of wind hazard maps conditional on a TC making landfall somewhere (anywhere) in Bangladesh? In the introduction, the authors seem to be thinking about applications involving risk assessment, that is, knowing the long-term probabilities of a given hazard occurring. But for that application, one wants information that is not conditional on any short-term forecast; one wants the overall unconditional hazard. For this, 12 observed storms are not enough, I'd think, even with 9 ensemble members for each; one would want to span the whole space of possible storms. This normally involves looking at large numbers of synthetic storms, as in the catastrophe modeling done in the insurance industry.

This paper addresses the hazard specific to one country. To make the most of our computationally expensive numerical modelling, it is necessary to constrain the hazard in question. In this case, rather than say looking at TCs in the Bay of Bengal, of which many TCs don't impact Bangladesh, we focus on those TCs that do specifically impact Bangladesh. This imposes the condition of our results being dependent on a TC impacting Bangladesh. Most, if not all, statistical models will have some degree of conditionality to their results, even if they are not explicitly stated – climate stationarity being a common implicit assumption, in the case of synthetic event sets from the catastrophe modelling community.

Quantifying the chance of extreme events is fundamentally constrained by the limited length of the observational record, and there is increasing understanding that events can occur in the current climate that have no observed precedence (e.g. Thompson et al., 2017, 2019). This underlies our interest in using a full numerical approach to simulating TCs. Our ensemble modelling approach provides the opportunity to explore plausible, yet unseen hazards. We agree that basing these simulations on 12 events is limiting, but the number of events was determined by the availability of source data (ERA5) – see our later reply for further discussion of this. Exploring the whole space of possible storms is the ultimate goal, but no methods currently achieve this.

Then, in the conclusions, the authors seem to be thinking about applications to shortterm forecasting, rather different from the applications suggested earlier. But in that case, one would have an actual forecast of the specific storm that would be coming.

The aim of the paper is to outline a method of improving the hazard-to-decision making process. The timeframes over which decision might be made could be hours, or day, or years. We are not clear which aspect of the conclusion (ie. Section 4) the reviewer feels specifically apply only to short term forecasting. The conclusion does specifically reference 1-in-20 and 1-in-100 year return periods, that may facilitate long-term disaster risk reduction planning.

This method of Section 3.1 is applicable across all time-scales, but we recognise that in this particular case (ie. relating to TC hazards) at very short time scales, weather forecast information will probably supersede our generalised hazard assessment (as presented in Section 3). However, the decision making framework is still applicable even in a forecasting situation – it's just that the predictive hazard data (ie. $p(x|y)$) would be derived from a forecast rather than our hazard maps, which we note in L186-187.

Why would one want a general forecast only conditional on a TC anywhere in Bangladesh?

See our answer re. conditionality above.

Surely wind maps from a past storm whose center made landfall on the westernmost part of the coast, for example, are not relevant if the actual storm is heading to the easternmost part of the coast; but the maps produced here include both equally. It makes no sense to me. I think the authors need to explain up front what problem they are trying to solve and why they are trying to solve it this way. I cannot actually see what useful problem justifies this particular approach. I might be missing something, but they need to make the argument more clearly.

We have edited the final paragraph of the Introduction to clarify these points:

> "In this study we seek to improve our understanding of the latent extreme gust speed hazard associated with TCs. To address the lack of observation data in this region, we use the latest generation Met Office regional model over the BoB, to simulate 9 versions of 12 historical tropical cyclone cases representing 1979-2019. This generates spatially and temporally consistent, counterfactual simulations (relative to well defined and observed TC cases), albeit limited by the constraints of the model configuration and computational resources. This ensemble configuration enhances our understanding of how each cyclone may evolve if a similar event were to happen again. We combine the ensemble information in a spatially coherent manner to produce hazard maps at 4.4km resolution over Bangladesh for extreme wind (gust) hazards. Using Bayesian inference, we estimate gust speed exceedance intervals (return periods) across all of Bangladesh, and demonstrate how this information can be directly integrated into a decision making framework."

2. The simulations are not described nearly enough. We know only the grid resolution. We aren't told where the domain boundaries are, the length of the integrations, etc. The reader is referred to Steptoe et al. (n.d.) for this information. I don't know what "n.d." means, but this information is not available to the reader that I can see. At least the basics should be in the paper. Also we

should see some comparison of the simulations to observations for these storms, so we have some idea how good the simulations are.

Yes, we agree that validating the RA2 model output is important. Unfortunately, delays in the peer review process mean the accompanying paper Steptoe et al. (n.d.) is still under review, but perhaps we should have included it as a supplement to this submission. As noted above, we have now included it as a supplement to this reply. We have added some further information on the basic simulation set-up into Section 2 as follows:

> "The RAL2 4.4km domain avoids placing model boundaries over the Himalayas and covers Nepal, Bhutan, Myanmar, most of India, and parts of the Tibetan plateau. To ensure model stability over this mountainous terrain, the RAL2 model was run with a 30 second time-step."

3. The statement that the median peak gust speeds simulated exceed those in IBTRACS and ERA5 by 22-43 m/s is shocking — that is an ENORMOUS discrepancy. However the statement also makes no sense, or at least is not explained well enough. What is the definition of a gust in the model, in IBTRACS, and in ERA5, for example? To my knowledge actually neither IBTRACS nor ERA5 includes information on gusts, as normally defined, just "sustained", i.e., 1-min or 10-min average, wind speeds; a gust is usually defined as higher frequency, i.e., 3-second average or such. Is a gust in the model resolved or parameterized? Is it just the instantaneous wind speed at a time step on the model grid? If so, at what time resolution was it saved – every time step (and what is the model time step)? And once we have the answers to these questions, what level of agreement should one expect between the model and ERA5 or IBTRACS, given what I'm nearly certain are the very different natures of these data sets?

We feel the answers to these questions are all provided by the accompanying paper Steptoe et al. (submitted), which is currently going through the publication process with another journal. For transparency we have attached our submission as a supplement to this reply.

Yes, the differences between RA2 and ERA5 is large, but it is known that extreme gusts associated with vigorous convection in ERA5 are generally under-estimated, sometimes by a factor of two (Owens and Hewson, 2018). As you point out, IBTrACS doesn't explicitly provide gust speed information (although our accompanying papers also compares maximum wind speed), so we have amended the text to clarify this.

Addressing some of your specific concerns:

- Yes, IBTrACS only estimates wind speed. In Steptoe et al., we find IBTrACS maximum wind speed estimates are more similar to RA2 wind speed estimates, than ERA5 wind speed estimates (for details, please refer to the paper).
- ERA5 gusts are parametrised based on the 10m wind speed, friction velocity, atmospheric stability, roughness length and a convective contribution based on wind shear.
- RA2 model uses a gust parametrisation based on 10m wind speed with scaling proportional to the standard deviation of the horizontal wind that also accounts for friction velocity, atmospheric stability and roughness length

Given the lack of reliable observations for this region, it is difficult to definitively say which dataset is 'best', however we feel there is enough agreement between the IBTrACS and RA2 wind speed (discussed in Steptoe et al.) that analysis of RA2 gust speeds will provide useful results for understanding gust extremes in this region.

More minor, specific comments:

Line 75: How were these 12 storms chosen? Of what are they meant to be representative? What is the strategy here? See major comment 1 above.

We highlight the limitation of 12 storms in Section 3.2. Specifically, the number of events was determined by the availability of source data (ERA5) for driving the regional model (RA2), for TC events that specifically impact Bangladesh. As ERA5 provides the initial conditions for our RA2 model, we are limited to the recent historical period from 1979-present (in this case, roughly August 2019, at which point we ceased any further RA2 simulations). We feel that these 12 events are representative of the recent historical period of TCs in Bangladesh.

Section 2.1: The authors go right into describing this somewhat sophisticated statistical model, but the reader at this point doesn't have enough of a clue what the objective is. "Condense information from all 9 regional model ensemble member footprints into a coherent spatial summary of the tropical cyclone hazard" is not enough. What is the reason to think these 12 storms x 9 ensemble members are representative enough for the purpose here, whatever that is? Can the authors please explain in plainer English what is being done here and why? And how can we determine if the answer we get is good or bad?

We agree that ideally we would have access to more that 9 versions of 12 events, but observational records are poor for Bangladesh. IBTrACS data only provides information along a trajectory, and ERA5 data does not have sufficient resolution to capture TCs properly. We feel that our 108 simulations are the best possible source of data at the current time. Some elements of the 'good or bad' assessment of these simulations is done in Steptoe et al. (n.d.), which is now attached as a supplement for your information. We have added additional clarification around the purpose in the Introduction, as per our comments in reply to your Major comment 1.

Line 147: isn't it a bit disturbing that the results are so strongly influenced by just two storms? Don't we want something more statistically robust than that?

There are very few historical events that track over the northern part of Bangladesh, so we are not surprised at this result. Given the available information, this is the best estimate we can make of the potential for high gust speeds in these regions, but note that the posterior predictive distribution for grid cells in these locations still contains information derived from the other 10 events. Looking at the 95th and 99th percentile gust speeds will inevitably reflect the extreme events in any given location, but these events will be rare occurrences. Note that we have now included a view of the 1-in-2 event to better reflect more common TC wind extremes.

We allude to the statistical robustness (or otherwise) of the predictive distribution in L203-L208. Specifically note that under Bayesian inference, sampling a wider range of historical cases would reallocate the credibility of extreme gust speeds across possibilities.

Section 3.1: As in major comment 1, what is the problem being solved here? What can one use these results for that wouldn't be better served by a normal operational forecast simulation (or ensemble of same), which would have details of the specific storm?

The aim of this section is to demonstrate a method for making decisions under uncertainty. A 'normal operational forecast' simulation has uncertainty associated with it (irrespective of whether this is communicated to the user of the forecast information). The problem is: 'how to make the optimal decision given the uncertainty'. This is a concept addressed by Decision Theory that this Section explores in the context of our previously defined predictions of gust speeds, derived from our statistical (GAM) model. An operational forecast (or ensemble of forecasts) does play an important role, which is the point we were trying to make in L184 (as per your next comment below).

In the context of exposure to a hazard and being able to plan and mitigate against vulnerability, we think that it is an advantage to understand the optimal action given the historical hazard information. This is what Figure 6 shows. With new information (such as a specific storm forecast) this optimal action would (probably) be superseded, but the method of translating the (new) hazard information into action is exactly the same.

Line 184: "and in reality it would be updated..." How so, how much would this change the answer, and again why do it this way?

The warning level is a function of the posterior predictive distribution obtained from the GAM, and the loss function defined in Table 1. The extent to which new information (eg. a new event forecast) would change the 'answer' depends on how much the new forecast deviates from the posterior predictive distribution and the confidence you give this forecast. We have added words to clarify these points, but note that aim of this section is to demonstrate a transparent workflow that clearly translates hazard information into actions.

Line 194: "to be sure they would be identifiable within the low-resolution ERA5 data." I don't understand this at all. Why do the storms need to be identified in ERA5? How is ERA5 data being used, other than to initialize the model? This is totally unclear.

As ERA5 data is used to initialise the RA2 model, TC features must be present in order for the RA2 simulation to proceed. It is possible that for small scale, or weak cyclogenic features, the resolution of the ERA5 dataset will smear these out. Trying to feed these sorts of events into the RA2 model would result in nothing being simulated. Hence, we focus on known events that have peak wind speeds of at least 33 m s^{-1} (ie. a Category 1 TC as defined on the IBTrACS website). We have added some words to this section to clarify these points.

Citations

Owens, R. G. and Hewson, T. D.: ECMWF Forecast User Guide, ECMWF, Reading, UK., 2018.

Steptoe, H., Savage, N., Sadri, S., Salmon, K., Maalick, Z. and Webster, S.: Tropical cyclone simulations over Bangladesh at convection permiting 4.4km & 1.5km resolution, Sci. Data, (submitted), n.d.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Slingo, J. M., Brown, S. and Belcher, S. E.: High risk of unprecedented UK rainfall in the current climate, Nat. Commun., 8(1), 1–6, doi:10.1038/s41467-017-00275-3, 2017.

Thompson, V., Dunstone, N. J., Scaife, A. A., Smith, D. M., Hardiman, S. C., Ren, H.-L., Lu, B. and Belcher, S. E.: Risk and dynamics of unprecedented hot months in South East China, Clim. Dyn., 52(5–6), 2585–2596, doi:10.1007/s00382-018-4281-5, 2019.