Reviewer 1

The manuscript “Stochastic properties of coastal flooding events – Part 2: Probabilistic Analysis” describes the development of time series of beach flooding from pictures using the Convolutional Neural Network (CNN) model-based semantic segmentation methods described in Part 1 (this manuscript is a two-part submission). The authors QA/QC the output and find the statistics of the measured “flooding” match up well with those from predicted “flooding” using the total water level (still water level + wave runup (R2% as defined by Stockdon et al., 2006)) over a beach elevation based threshold (and as previously analyzed in Rinaldo et al., 2021).

The manuscript is well-written and concise. I think the methods are neat, and I like the characterizing “how much water in the frame” as a more detailed method for flood detection than a binary Y/N, but also less complicated than a full georectification of the water line elevation. While the authors display interesting findings, the main contribution of the research isn’t clear. The paper reads as a validation of the work presented by Rinaldo et al., 2021, however its broader importance to the field could be better explained and tied into other literature. The authors suggest in the second to last line of the manuscript that “Our results formalize … the first probabilistic model of coastal flooding events driven by wave runup.” But I’m not entirely sure what the probabilistic model is and/or how it might be used more broadly. This paper could be strengthened with more explanation about what the authors mean by probabilistic model and tightening up the key terminology. My specific suggestions are found below.

Reply) We thank the reviewer by the very helpful comments and suggestions. Our main goal was to describe the probabilistic characteristics of flooding events at that particular site, as needed to quantify the impact of e.g. nuisance flooding. The comparison to the findings of Rinaldo et al. 2021 is a secondary, albeit important, element in our work that adds generality to our findings as that study involved several locations and larger time scales. We now discuss this important distinction, among other things, in the introduction. We also introduced a new figure (Fig.1) comparing our study to Rinaldo et al. 2021’s and added a new section 5 summarizing the probabilistic model we were referring to in the conclusions.

The new introduction includes the text:

The primary goal of the present study is to describe the probabilistic structure of flooding events measured at a recently eroded site in northern Texas. Flooding events were defined applying the peak-over-threshold method to a high-resolution time series of water area fraction, obtained from coastal images using Convolution Neural Network (CNN)-based image segmentation as explained in a companion contribution to this journal volume (Kang et al., in review). A central outcome of our research is the validation of the results of Rinaldo et al. (2021). As shown in Fig. 1, although our study complements the spatial and temporal range investigated by Rinaldo et al. (2021), it is limited to a single site and roughly half-year data. However, we can use our results to establish the validity of Rinaldo et al. (2021)’s more general predictions.

General comments:

The motivation of the study could be clearer. I understand the authors are more or less validating the Rinaldo et al., 2021 paper, but the details of why this is important and how this adds to the field are less clear.

Reply) We thank the reviewer for pointing out this important aspect as we now realize it was not clear in the original manuscript. Rinaldo et al. used a phenomenological formulation of total water levels to estimate events overtopping a given elevation and evaluate their probabilistic characteristics. They found a relatively simple relation between the size of an overtopping event and its frequency, or the inverse return period, for multiple locations. That type of relation is fundamental when trying to evaluate the occurrence and size of flooding events as it answers the question of what is the overtopping threshold of a 1/10 yr event, 1/100 yr, etc. Therefore, validating their results with actual flooding data (instead of relying on deep-water data) adds credibility to their predictions. Furthermore, our results set the limits of application and interpretation of their findings. We add more details in the revised introduction, which include the texts:

High-frequency and low-intensity coastal flooding (…) is very difficult to predict in detail and has to be described statistically. This probabilistic description can ideally lead to the estimation of both, the overtopping frequency $\lambda(Z)$ (or return period $T = \lambda^{-1}$) of a given elevation Z; and the average size $\bar{S}(Z)$ of events overtopping Z. This information can then be used to assess the vulnerability of coastal features and coastal infrastructure and plan accordingly.

Recently, Rinaldo et al. (2021) investigated the stochastic properties of high-water events (HWEs), which are associated with coastal flooding, on several locations along the US and the world. (…) These findings can be summarized in an equation for the overtopping frequency of a threshold elevation Z : $\lambda(Z) = \lambda_b \exp[-(Z - Z_r)/\bar{S}]$, where $\lambda_b = 18\mathrm{yr}^{-1}$, $\bar{S}$ is the site-dependent average size of HWEs ($\bar{S} \approx 0.3\mathrm{m}$) and Z_r is a reference elevation that depends on the tidal amplitude and average wave runup and can be interpreted as a characteristic beach elevation (Rinaldo et al., 2021).

The authors could better explain the comparisons between the Rinaldo et al., 2021 work and their results, specifically when it comes to the "beach elevation" threshold comparisons. In the introduction (when referring to the Rinaldo et al., 2021 paper) and later (section 4), the authors discuss beach elevations. I find this discussion/comparison a bit confusing, as there's no context as to what the beach elevations mean or are related to and/or what the characteristic beach elevation is. First, there's no datum to any of the elevation measurements, which is important when discussing height. Second, it seems the whole beach elevation discussion is a threshold analysis, where the authors are potentially deciding on an appropriate beach elevation in which to evaluate flooding over. I wonder if it makes sense to call these "beach thresholds" rather than "beach elevations."

Reply) We agree that those elevations are actually thresholds for overtopping that affects the overtopping frequency. However, the relation to the 'effective' or 'characteristic' beach elevation (as defined in Rinaldo et al., 2021) is important because it provides a way to compare the overtopping frequency of Rinaldo et al. (which depends on the variable beach threshold) with ours (which is fixed by the actual beach elevation). We now properly define the beach elevation the first time Rinaldo et al. (2021)'s work in presented (see red text above) and the first time we use it in our HWEs analysis (section 4), see text below:

(…) Finally, we defined a high-water event (HWE) as the set of consecutive daily total water levels exceeding a given elevation Z_c relative to MSL (Rinaldo et al., 2021). Here Z_c is interpreted as a characteristic beach elevation in which case HWEs represent potential flooding events. (…)

The authors use a peak over threshold approach for threshold selection to define flooding and choose the 2% threshold from the distribution of water area fraction. Was this threshold varied with each change in the time window? E.g., the 2% value with a 5-minute time step is not going to be the same as the 2% value of the daily time step. In general, did the authors test different threshold values (even for the same timestep)? The authors also test a few different "given beach elevations" and the HWE statistics over those thresholds, so it would seem natural to test different thresholds of the flood measurement model too.

Reply) There is a subtle but important distinction between flooding conditions and flooding events. Flooding conditions is when water pixel area goes beyond what is typical during normal conditions in the 5-minute time series. The 2% threshold is chosen to clearly separate the extremes (flooding conditions) from the normal in the 5-minute time series (see the **new figure 5** illustrating that point). On the other hand, flooding events are defined by clusters of consecutive points in the time series above the 2% threshold defining flooding conditions in the first place. By changing the time window we are redefining the extent of flooding events (i.e. do we have flooding conditions in consecutive minutes, hours, or days?) but are not changing the underlying reality of having a flooding condition or not at any given time (defined by the 2% threshold). Using say a 4% threshold would just filter some small size flooding conditions (<4%). Instead, what we did is to characterize the frequency and size (%) of the flooding events for a given timescale such that it is possible to estimate the return period of events above a given size. We added the probabilistic model (Eq. 3) derived from our results in the new section 5.

Furthermore, it is important to note that that 2% threshold reflect the actual flooding conditions in the data and thus the actual beach elevation at our site. In contrast, the threshold in the HWE formulation can be associated to any overtopped elevation, e.g. beach, dune, dike. In some way, our data represents the realization of the HWE formulation when the overtopping elevation corresponds to the beach. We added more details about the differences between both formulations in the revised version.

There is a lot of focus on wave runup and the flood variability wave runup produces and the development of a flooding driven by wave runup. However, the flooding the authors are tracking is based on the total water level, not just the wave runup, as I don't believe they are removing the tidal and/or SWL signal. So, it's not accurate to say their model provides "runup predictions" (Line 11, also section 4 title). Yes, wave runup is a contributor to the beach flooding the authors are tracking, however there are examples from their results that to me suggest there's more at play than just wave runup. One example of this is that the correlated events for timescales less than 10 hours. If you look at Figure 8C, the number of events really stabilizes >12 hours and certainly by 24 hours, which is also the duration of semi-diurnal tidal cycles. This hints to the relationship being related to the tide + weather (waves, anomalies) rather than related to timescale of local weather alone affected by the daily cycle as suggested by the last line on page 8. The authors should clearly distinguish between the total water level and the wave runup.

Reply) We thank the reviewer for this comment. Yes, the flooding we measure is consequence of the wave runup and setup superimposed to the SWL, which contains the tides (see below).

The new introduction now includes the text:

High-frequency and low-intensity coastal flooding is mostly driven by extreme values of wave runup superimposed to the tidal signal (Serafin and Ruggiero, 2014; Serafin et al., 2017), as the elevation of natural beaches typically adjusts to the average 25 wave runup (Rinaldo et al., 2021).

We agree about the potential effect of the tidal signal in event correlations for less than 12h. The discussion and conclusions section now includes the text:

(…) This change in temporal correlation for timescales around 10 hours could be related to the tidal period (which is about 12h at this location) and the day-night cycle potentially disrupting any local weather pattern behind the flooding event.

In relation to the wave runup, the authors discuss beach slope a few times. It's unclear to me how the authors determined the beach slope, as the paper cited doesn't have much information about beach slope for their study site. I also wonder if some of the mismatch between predicted and measured flooding is due to the beach slope and/or slope variability at their site, since it's a broad amount of coastline they view. Some assessment of the local beach slope, especially over the period of data collection would be important, especially since they also state in their conclusions that " and our measurements take into account local beach erosion due to hurricane Harvey." On this note, I don't doubt that the Stockdon formulation may overpredict TWLs at time, especially on the hourly/daily scale, since it's the wave runup, which is the 2% highest R values. I wonder if you used wave setup + SWL, more of the mean wave-driven water level component, that you'd see better agreement with your flood hits/misses. I also wonder if there are more limitations to this comparison besides just the variability of waves. For example, what about the fact that the authors are looking at a water surface over a large stretch of beach and the % coverage rather than just a transect like the HWE prediction methodology is doing? Because of site-to-site variability, you could have some areas that would consistently "flood" but that wouldn't be the same for the whole stretch of the beach.

Reply) For the analysis of the HWEs in the area, we used a representative slope estimated in Rinaldo et al. using a single cross-shore profile in the DEM. We agree the lack of detail of the slope spatio-temporal variability adds to the uncertainty of the predictions relative to the measurements and might account for the discrepancy. Although in this study we didn't try subtracting the total swash excursion factor from the Stockdon formulation of wave runup (to only keep setup and SWL), we agree that would be a good next step moving forward. However, our measurements are not suited to test the parametrization of the Stockdon model as we don't measure the same thing. The last point regarding the site-to-site variability is also important and highlights the uncertainties in translating a predicted flooding, using the calculated HWEs, into the actual flooding at a given beach location. Below are a few examples of how we addressed this comment in the text.

In section 4, when explaining the calculation of HWEs:

(…) Since we did not perform measurements of the beach profile at the study site and for the observation period, we assumed the beach slope needed to calculate the wave runup, was constant and equal to 0.02, as found by Rinaldo et al. (2021).

In the discussion and conclusions section:

(…) When focusing on the daily correlation of predicted and measured flooding, the predictions from the analysis of HWEs (Rinaldo et al., 2021) captured most of the occurrence of daily flooding, although it noticeably overpredicts them. The large fraction of false positives in the predicted flooded days (particularly at the end of the measurement period), even after correcting for a different beach elevation, could result from the assumption of a constant beach slope along the whole beach section covered by the camera and for the whole observation period. Since run-up predictions using off-shore data (Stockdon et al., 2006, 2014) are essentially valid for a single transect and thus neglects the alongshore variability of the bathymetry or the details of wave shoaling (García-Medina et al., 2017; Atkinson et al., 2017), it would be difficult to capture the complexity of the site-to-site variability of flooding over a relatively large beach section. On the other hand, it could be that the predicted flooding was taking place somewhere else along the beach and was not captured by our local observations. A final possibility is that our sampling frequency of one picture every 5 minutes is not high enough to capture all possible large runup events (as predicted by the HWEs formulation), in which case the false positive rate could be lower. This is supported by the fact that the distribution function of the duration of flooding events has a lower-limit of 3 minutes.

Regardless of these sources of potential errors, and more in line with the statistical nature of wave runup data and the uncertainty in the calibration of the model parameters in the first place (García-Medina et al., 2017; Atkinson et al., 2017), one can argue that the prediction only indicates conditions favorable to flooding events somewhere along the shoreline and not necessarily the actual occurrence of a flooding event at a precise location. This statistical interpretation would agree with our findings.

Terminology: The authors use many different terms to describe the same phenomena throughout the article, for example, coastal flooding, high-water events, flooding events, high-water events overtopping the beach, … I think it would be good to define what you're measuring and stay consistent with the terminology throughout the article. I find the use of the terminology "coastal flooding" (especially in the title) a little broad and misleading, as it seems this is really looking at "beach flooding."

Reply) We agree, we now clarify the relation to coastal flooding whenever we introduce high-water events and the overtopping of a characteristic beach elevation. As for the term coastal flooding vs. beach flooding, we now clarify in several places in the manuscript that we are referring to high-frequency and low-intensity flooding (also called nuisance flooding). Of course, in a relatively high elevation beach that would only lead to a minor beach flooding, however, the same event in a lower-elevation coastal area could lead to a relatively large flooding.

The authors also use terminology of 24 h or 1 day for their daily signal, and I suggest keeping it consistent and choosing one to use throughout.

Reply) We kept the 24h.

Line by Line:

Line 22: "In this work…" I believe the authors are referring to the Rinaldo et al., 2021 paper, but the way it is written it could be referring to this present manuscript.

Reply) Corrected.

Line 25: The term overtopping here is odd to me. Flooding the beach, exceeding the shoreline, etc makes more sense for the types of events being discussed here, as overtopping typically refers to total water levels exceeding barriers such as dunes, seawalls, etc.

Reply) We like the term overtopping because it clearly compares two elevations: the water level and a given threshold that is exceeded, which naturally leads to flooding. For the point of view of HWEs, the threshold elevation is arbitrary, could be the shoreline elevation, or a dune, or a seawall. In our data, though, the term overtopping is not necessary as it is clear we are measuring beach and back-beach flooding conditions (the dunes were eroded during hurricane Harvey). We now clarify that coastal flooding naturally follows from the overtopping of a characteristic beach elevation. From the new introduction:

High-frequency and low-intensity coastal flooding is mostly driven by extreme values of wave runup superimposed to the tidal signal (Serafin and Ruggiero, 2014; Serafin et al., 2017) overtopping a characteristic beach elevation, or any other feature close to the shoreline. As the characteristic elevation of natural beaches typically adjusts to the average wave runup during high tide, they are only flooded during extreme events (Rinaldo et al., 2021). (…)

Line 26-27: "They also found that the size and intensity of an event, defined by the maximum total water level during the event, does not vary with increasing elevation" I think the authors are referring to some beach elevation threshold, but it's not clear from this sentence if readers are not familiar with the Rinaldo et al., 2021 work.

Reply) We agree the phrasing is confusing. We now explain the results of Rinaldo et al. in more detail and more clearly in the introduction.

Line 43: Why was the observation period chosen to be 6 months?

Reply) After that time we unfortunately lost the camera system during an unexpected storm.

Line 74: worded weirdly, “seems to follow” Or “followed” or “follows” but not “seems followed”

Reply) Corrected.

Line 77 – 78: The authors suggest here that no event in their time series lasts more than 3 hours, which means there’s a physical upper limit to sustained flooding when there’s not a large storm. Do the authors have a long enough record to make this statement? Where there storms over this time period? There are also other factors that increase water levels – to me, this statement suggests tides are a big contributor to “when”/”how long” flooding occurs, if they’re only occurring over a few hours.

Reply) Although there were a couple of winter storm during that period, there were no hurricanes, so that statement is only valid for our measurement period. Also, yes, tides could be a contributing factor. However, in this region astronomical tides are relatively minor and water levels are mainly affected by waves. The new discussion and conclusions section includes the text:

(…) The lack of events longer than 3 hours in our nearly six-month period, during which there was no large storms, seems to suggest a physical upper limit for sustained flooding conditions perhaps related to high tides. However, in this region astronomical tides are relatively small and water levels are mainly affected by waves, which would again point to wave runup driving the observed flooding, as suggested by the high-water event analysis. (…)

Line 89- 90: Do the authors have beach slope measurements during the time period? Did the beach slope vary a lot? How spatially and temporally varying is beach slope in this location? I’m not sure the authors can say much about robustness with respect to beach slope without any measurements (this is mentioned above too).

Reply) We agree with this point. Unfortunately we don’t have slope data for that period. We added this when discussing the sources of uncertainty in the comparison to the HWE predictions in the discussion section (see above).

Line 118: Did the authors use the data developed in Rinaldo et al., 2021 or create a new time series using the same methodology? It’s unclear from this section. Either way, I think the authors should provide details about the tide gauge and buoy data used in this analysis. I’m interested also in what the water depth of the buoy used was? Stockdon et al., 2006 was developed from waves at the 20m contour that were linearly backshoaled to be “deep water waves” and they recommend linearly backshoaling waves if nearshore buoys are used rather than deep-water buoys.

Reply) We created a new time series using the same methodology. We added the following text in the data acquisition and method section:

Following the methodology from Rinaldo et al. (2021), which involved calculating the hourly time series of total water elevation for the same site using a beach slope of 0.02, we generated a new time series of daily total water levels relative to MSL. This required summing the still water level as measured by a tidal gauge, and a semi-empirical estimation of the 2%- exceedance wave run-up. The latter relied on off-shore values of the significant wave height and peak wave frequency, and the local beach slope (Stockdon et al., 2006, 2014).

Our data sources included the tidal gauge at Galveston Pier 21 (29.31° N, 94.793W), and wave buoy station 42035 (29.236°N, 94.403°W). Both located in Galveston, Texas, they provided hourly measurements of water levels and significant wave heights and peak period. While the water depth of the wave buoy was 15m, we did not consider reverse shoaling to deeper water, as recommended by Stockdon et al. (2006), to maintain the simplicity of our analysis and directly compare to the results of

Rinaldo et al. (2021). Since we did not perform measurements of the beach profile at the study site in the observation period, we assumed the beach slope, which is needed to calculate wave runup, was constant and equal to 0.02 (Rinaldo et al., 2021).

Figures:

Even though there is a companion paper, I think this paper still needs a map to show where the study site/camera is located. Since they're two separate papers I don't think we can assume the readers will read both…(or should have to read both)!

Reply) Done (new Fig. 2).

Figure 1: Plotting the 2% threshold on the time series panel would be helpful!

Reply) Note that the 2% threshold for flooding conditions is imposed to the excess water area fraction time series (Fig. 4B) and not to the original, pre-processed data, shown in Fig.3 and Fig.4A. We now show the 2% threshold in Fig.2B.

Figure 4: Why are the two figures on different x-axes? It feels misleading as my eyes are trying to draw similarities between the two. Please put on the same axis by extending A or decreasing the time window of B. It's important to know that (it looks like) everything in A is one event in B at the daily scale…!

Reply) We agree and now include a shaded region in the new Fig.6B showing the time window of panel A.

Figure 6: I would remind readers here how you are defining event size, either in the caption or on the y-axis, e.g., max % water area over the 2% threshold

Reply) Done.

Figure 7B: is never cited in text – can authors put rejection range/plot? Legend overrides data on plot Figure 7C is not cited until much later. Good practice is to describe figures in order of appearance.

Reply) Done.

Figure 10: You note that this represents when the predicted flood frequencies are equal, but it doesn't look like that from the figure. Is it that the duration of the predicted water levels are much longer?

Reply) Yes, it is exactly that. Although the frequency is the same, i.e. the number of events is the same in both data and HWE prediction, the duration of the events in the latter is much longer. We added this implication to the figure's caption and the main text.

Figure 11: what does the shading represent?

Reply) We substituted the shaded area by a line at the relevant elevation 0.7m used in Fig.12.

I didn't see any data availability statement or a location where codes could be found – please see the journal's Data Policy.

Reply) Data is available in the Texas Data Repository (TDR), doi.org/10.18738/T8/FLGDS0.

Reviewer 2

General comments

The authors derive statistical information about the frequency and intensity of the 'flooding of the beach' (whether the beach is submerged). This is derived from the segmented coastal imagery data presented in part 1 of the paper. The statistics from the data are compared to the model by Rinaldo et al. (2021), where significant differences are found.

In part 2 of the paper, the choice to look at the % of pixels identified as water in the coastal imagery shows some negative consequences, in my opinion. Because no wave runup height (or more accurately total water level elevation) is being distilled from the images, quite some steps need to be undertaken to actually interpret the physical meaning behind the data, as described in Section 4 of the manuscript. This results in a fairly roundabout way of comparing the results from the area-fraction approach to data from tidal gauges and offshore wave data, which in turn results in a seemingly significant difference (from Figures 9 through 11). All of this makes it difficult to draw clear conclusions. Hence, it stays unclear to the reader what the added value of the method presented is and how generically applicable it is.

Reply) We thank the reviewer by the helpful comments and suggestions. Our main goal was to describe the probabilistic characteristics of flooding events at that particular site, and to compare them to the findings of Rinaldo et al. 2021. Since the distributions functions are all normalized by the average flooded size, the only thing we need to assume to compare both results is that the our flooded area, defined by % of pixels, is linearly correlated to the total water elevation, which certainly provides a physical interpretation to our results (see below). Indeed, we find an exponential distribution of event sizes as in Rinaldo et al. 2021, which confirms our very simple way to quantify flooding for statistical applications. In Figs. 11-13 we show how well the HWEs method can predict the actual occurrence of flooding in our site at a given time, which only rely on the fact that water crossed an elevation threshold at the beach in both the model and our data, not on the way to measure it.

As for the value of our study: Rinaldo et al. used a phenomenological formulation of total water levels to estimate overtopping events and evaluate their probabilistic characteristics. They found a relatively simple relation between the size of an overtopping event and its frequency, or the inverse return period, for multiple locations. That type of relation is fundamental when trying to evaluate the occurrence and size of flooding events as it answers the question of what is the overtopping threshold of a 1/10 yr event, 1/100 yr, etc. Therefore, validating their results with actual flooding data (instead of relying on deep-water data) adds credibility to their predictions. Furthermore, our results set the limits of application and interpretation of their findings.

We now clarify the central purpose and context of our study in the introduction (see below). We also introduced a new figure (Fig.1) comparing our study to Rinaldo et al. 2021's and added a new section 5 summarizing the probabilistic model.

The new introduction includes the text:

High-frequency and low-intensity coastal flooding (…) is very difficult to predict in detail and has to be described statistically. This probabilistic description can ideally lead to the estimation of both, the overtopping frequency $\lambda(Z)$ (or return period $T = \lambda^{-1}$) of a given elevation Z; and the average size $\bar{S}(Z)$ of events overtopping Z. This information can then be used to assess the vulnerability of coastal features and coastal infrastructure and plan accordingly.

Recently, Rinaldo et al. (2021) investigated the stochastic properties of high-water events (HWEs), which are associated with coastal flooding, on several locations along the US and the world. (…) These findings can be summarized in an equation for the overtopping frequency of a threshold elevation Z : $\lambda(Z) = \lambda_b \exp[-(Z - Z_r)/\bar{S}]$, where $\lambda_b = 18 \mathrm{yr}^{-1}$, $\bar{S}$ is the site-dependent average size of HWEs ($\bar{S} \approx 0.3$m) and Z_r is a reference elevation that depends on the tidal amplitude and average wave runup and can be interpreted as a characteristic beach elevation (Rinaldo et al., 2021).

…

The primary goal of the present study is to describe the probabilistic structure of flooding events measured at a recently eroded site in northern Texas. Flooding events were defined applying the peak-over-threshold method to a high-resolution time series of water area fraction, obtained from coastal images using Convolution Neural Network (CNN)-based image segmentation as explained in a companion contribution to this journal volume (Kang et al., in review). A central outcome of our research is the validation of the results of Rinaldo et al. (2021). As shown in Fig. 1, although our study complements the spatial and temporal range investigated by Rinaldo et al. (2021), it is limited to a single site and roughly half-year data. However, we can use our results to establish the validity of Rinaldo et al. (2021)'s more general predictions.

Also, the discussion and conclusions section includes the text:

When focused on the daily timescale, we found that flooding events can be modeled as a Poisson process with exponentially distributed sizes, in agreement with recent findings using a run-up model to predict coastal flooding (Rinaldo et al., 2021). The main probabilistic properties of measured and predicted flooding events can thus be described by Eqs. 3 and 4, respectively. One way to understand the similar form of both equations is through the relation between flooded area and water depth at the shoreline. Assuming the beach slope in our field site is relatively constant then we would expect both to be proportional, in which case the fraction of water pixels would also correlate with water depth at the shoreline. Therefore, our agreement with Rinaldo et al. (2021) suggests that the exponential distribution is robust with respect to potential variations of the local beach slope during the measurement period and alongshore variations of the flooded area at the spatial scale defined by the camera field-of-view.

Specific comments

1. L017: "the importance of…has recently become clear". I would have expected either an accompanying reference that illustrates this point or a more expansive argumentation for this claim in the introduction of the current paper.

Reply) Done.

L019: The Kriebel & Dean (1993) reference formatting is incorrect, including the first names of both authors.

Reply) We updated the references.

2. L025: "…high-water events overtopping the beach (i.e. flooding events)…" is a bit ambiguous, as 'the beach' is not a single identifiable height/elevation that is being overtopped (in contrast to e.g. a dune or dike crest being overtopped). Please make clear what is actually meant here. This issue is also present in Line 115 (also see the commentary on the ambiguity in the use of the term flooding in my commentary on part 1 of this paper.)

Reply) Yes, we agree with the reviewer, and we thank her for highlighting this inconsistency. We define flooding conditions, in contrast to normal conditions, statistically from our time series. Because normal conditions (characterized by a relatively Normal distribution of the water fraction) are related to the existence of a shoreline, we can roughly interpret flooding conditions as the overtopping of the shoreline elevation. We have included a proper definition of what we measure, as well as a consistent physical interpretation and consistent terminology, in the revised version, see below:

We defined a flooding event as the set of consecutive values of the water area fraction $A|_\tau (t)$ that exceeded the 2% threshold (Fig. 6). This threshold allowed a clear separation between typical fluctuations in water area and the extreme values that characterize flooding conditions (Fig. 5B) and can be associated to a characteristic beach elevation above the shoreline.

Furthermore, we now clarify that coastal flooding naturally follows from the overtopping of a characteristic beach elevation. From the new introduction:

High-frequency and low-intensity coastal flooding is mostly driven by extreme values of wave runup superimposed to the tidal signal (Serafin and Ruggiero, 2014; Serafin et al., 2017) overtopping a characteristic beach elevation, or any other feature close to the shoreline. As the characteristic elevation of natural beaches typically adjusts to the average wave runup during high tide, they are only flooded during extreme events (Rinaldo et al., 2021). (…)

3. L059: The fact that the pictures are taken every 5 minutes (which I understand from a data storage/transfer capacity standpoint) means that not all wave runup events are captured, and the data (regardless of whether timescale is chosen to calculate the maximum over) will underestimate the actually occurred runup heights. This should at least be reflected upon, for instance in the discussion. Especially when comparing to a method using Stockdon et al. (2006), which calculates the 2% exceedance runup height which will almost certainly be underestimated in measurements taken every 5 minutes.

Reply) We agree and we added a clarification in the revised version (see below). However, note that for the comparison between our data and the predictions from TWL by Rinaldo et al. using Stockdon's method, and the related 2% exceedance criteria, we divide the time series of flooding magnitude (water % in our data and actual water elevation in Rinaldo et al.) by their mean, which in principle should correct for any multiplicative correction factor.

The text below was included in the section about the duration of flooding events:

(…) lower limit d_{min} = 3 min. The fact that this lower limit is below the 5-minutes temporal resolution of our data suggests that we are missing many relatively short flooding events. Interestingly, from Fig. 8, short flooding events are not necessarily of small size.

We also added this text to the discussion and conclusions section:

(…) A final possibility is that our sampling frequency of one picture every 5 minutes is not high enough to capture all possible large runup events (as predicted by the HWEs formulation), in which case the false positive rate could be lower. This is supported by the fact that the distribution function of the duration of flooding events has a lower-limit of 3 minutes.

4. L066: "the time water area increased above 2%" this seems like a very arbitrary threshold, and one the depends quite a bit on the camera view and in turn the positioning and orientation. Also, a reference is made to Rinaldo et al. (2021) suggesting that this is in line with what they do, but they describe a "…2% exceedance wave runup…", which is something completely different. Namely, that is the vertical wave runup level that is exceeded by 2% of the events.

Reply) The 2% threshold is chosen to clearly separate the extremes (flooding conditions) from normal conditions in the 5-minute time series (we added a new figure 5 illustrating that point). It has nothing to do with the 2% exceedance criteria of Stockdon et al. Furthermore, the 2% threshold is applied only to the time series of excess water area fraction, which was corrected to minimize the effect of changes in the camera position and orientation (see new Fig. 4). Of course, the 2% value itself, instead of say 1.5% or 2.5%, and the fact that we keep the same value for all time windows, is more or less arbitrary. This is a drawback of the way we fixed the camera poles. However, the extreme nature of most flooding events makes the threshold selection less relevant as shown by the consistent exponential distribution of the water % data (the exponential distribution of the marks in the peak-over-threshold method don't change with the threshold value).

5. L087: The Cramér (1928) reference is missing the year of publication.

Reply) Done.