To the Reviewers

We are very thankful to the reviewers for their very fast handling of our manuscript "Operational hydrodynamic service as a tool for coastal flood assessment". We highly appreciate their comments and suggestions concerning our work. Below please find our point-to-point response to the reviewers' comments as well as explanations of the resulting changes in the manuscript. In particular, we include all modified fragments of the manuscript in this reply letter below. All changes are highlighted by bold font. Along with the changes we give some surrounding non-modified sentences to provide the context.

Author's reply to Reviewer 2:

Paper is generally well written and clear and describes an interesting local flooding tool

Author's reply:
We sincerely thank the reviewer for the positive feedback and for undertaking a thorough reading of the manuscript. Their valuable comments and suggestions are greatly appreciated and will contribute significantly to enhancing the quality of both the study and the manuscript.

1. Line 61: Can you elaborate on who are the stakeholders you are aiming this at? Eg Local government agencies, emergency responders, business owners, beach users?

Author's reply:

We completely agree with the reviewer that specifying the stakeholders targeted by our operational strategy is crucial for understanding its overall goals. It also helps readers envision how the strategy can contribute to reducing impacts. In this context, our primary focus is on local government authorities and those directly responsible for beach safety. These stakeholders are in the best position to mitigate potential overall effects and, if necessary, to warn beach users. We have clarified this in the manuscript to enhance the understanding of the intended end-users of our approach. Please, find the new updated sentence in P.3, Introduction, L. 62 and here:

> Modified manuscript text
> "...This methodology provides end-users, **such as local government authorities and those responsible for beach safety**, with forecasting outcomes..."

2. Line 92: I don't believe it is available for Windows, but for linux and MacOS I would recommend looking at the cylc workflow engine (https://cylc.github.io/) which is more sophisticated, for example would allow the initial server request to be automatically retried if it fails (and the following tasks would then wait until the request succeeded).

Author's reply:
We sincerely thank the reviewer for the effort in suggesting such an interesting tool to enhance the strategy's efficiency and stability. After taking a look on it, and as the reviewer points out, the tool appears to be designed primarily for Linux and macOS. Since our study is conducted on Windows, we have decided not to implemented it at this moment on this presented manuscript. However, after evaluating the tool, we have added a sentence in the

revised manuscript to recommend its use for MacOS and Linux. Additionally, if we use our strategy to those operating systems in the future, we will certainly consider using the tool to improve overall stability, as we recognize its potential as a valuable option. Please, find the new updated sentence in P.5, Operational architecture, L. 96-97 and here:

> Modified manuscript text
> "...A similar approach for Linux could be implemented with equivalent settings through "at" package commands **or cron tab**, **while for MacOS using** Automator and iCalendar. **Additionally, the Cylc Workflow Engine (Oliver et al., 2018) could be utilized for both operating systems as well**. In our application...

3. Line 111: What was the reason for picking the IBI dataset over the MEDSEA one? What are the key differences between the two?

Author's reply:

This is an interesting point. As explained in the manuscript, both IBI and MEDSEA datasets are available for the study area. They are quite similar, with the main key difference being the resolution: IBI has a resolution of 1/36º, while MEDSEA has a resolution of 1/24º. Although both datasets have additional features that make them slightly different, these do not significantly affect the overall performance or make one superior to the other. While MEDSEA's higher resolution might suggest it is the better option, the IBI dataset has already been widely used as boundary conditions in building an operational service. This point was the mean reason to chose it over MEDSEA. Even so, we have added some additional references in order to enhance and highlight this fact. Please, find the new updated sentence in P.5, Operational architecture, L. 117 and here:

> Modified manuscript text
> "...IBI has been **widely** used as a boundary condition to build operational downstream services through high-resolution hydrodynamic models (e.g. Sotillo et al. (2020,2021); Lorente et al.(2019)), offering a daily updated high-resolution ocean analysis and forecast product."

4. Line 116: What is the grid resolution of the CMEMS dataset? How near is the closest point?

Author's reply:

We fully acknowledge that the previous version of the manuscript lacked the position of the CMEMS data point. To address this, we have now incorporated it into Figure 1, ensuring its correct placement in relation to the computational domain. Regarding the resolution, as explained in response to Question 3, it is 1/36º. However, we have not included this information in the text to avoid potential confusion. Since we are using only a single point from this dataset (R1), mentioning the resolution might suggest that we are utilizing additional points. Please, find the new updated sentence in P.5, Operational architecture, L. 121 and here:

> Modified manuscript text
> "...level is requested from the closest CMEMS point to the XBeach computational grid, hereinafter referred to as R1 **(Figure 1)**. The information..."

and in P.4, Figure 1 and here:

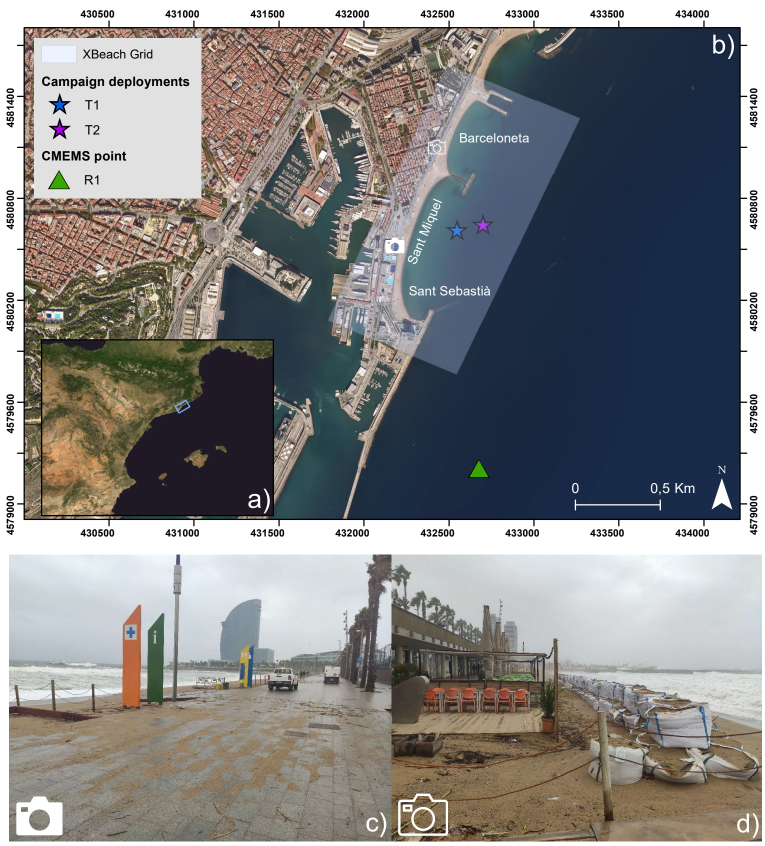


Figure 7: Location map of the study area. a) NW Mediterranean Sea. Blue rectangle marks Barcelona. b) Close-up of southern region of Barcelona beaches. Stars **(T1 and T2)** show the location of the field campaign deployments. **Triangle presents the location of the CMEMS reference point (R1)**. Gray rectangle represents the extent of the computational domain. c) show the ...

Additionally, we have created a figure for the reviewer (below) to clarify the positions of the two other closest points. This should help illustrate why we selected this specific point and why we are not using multiple points. While we did not include this figure in the main text, we have considered adding it as an appendix. However, we believe it might introduce unnecessary complexity and length to the manuscript without providing essential information.

Figure 8: Locations of various CMEMS points, with the green point representing the selected one (R1) and the red being the other closest to the computational domain.

5. Line 133: Is the magnitude/severity of the flooding impact always proportional to the area that is flooded? Could there be events where only a small proportion of cells are flooded but there is still a severe impact on infrastructure in that part of the beach?

Author's reply:

Strictly speaking, it is possible for an event to occur where a medium proportion of flooded cells could cause severe impacts on infrastructure (with small proportion, it is really very unlikely). However, such an event would be very rare due to the homogeneous distribution of the topographic contours in the study area. The scenario proposed by the reviewer would require the presence of a very narrow, locally emerged beach, which is not currently the case. In the future, the beach could evolve and potentially create such a situation, but we believe it is so unlikely that it does not warrant a dedicated study or analysis at this time. Nevertheless, we recognize the relevance of this point and have considered it due to its potential importance. For this reason, we decided to include the inundation contour (as shown in Figure 4) in the results provided to end-users. By doing so, local impacts can be assessed through these forecasted inundation contours. While our color-based categorization system may not detect rare events involving a small proportion of flooded cells with severe impacts, the inundation contours provided to end-users enable them to identify such unusual scenarios and respond accordingly. As explained, we did not include this specific point in the discussion because we consider such occurrences to be extremely rare. In the vast majority of cases, there is a consistent relationship between the magnitude of an event and the flooded area. Nonetheless, by including the flooding line as an output, we ensure that even these rare cases can be accounted for.

6. Line 143: Prioritising graphical information is a sensible choice. I would also suggest reviewing the format and type of data provided with the stakeholders to ensure that it meets their needs and expectations.

Author's reply:

We agree that prioritizing graphical information is a sensible choice. This decision is based on our experience working with stakeholders for whom this strategy is primarily intended. In these cases, presenting statistical information does not fully convey the magnitude and spatial distribution of the potential impact. For this reason, we opted to present the information in a graphical format and if needed or asked, provide a statistical or numerical report. Additionally, the color-based alert system enables stakeholders to easily understand the severity of the event. Nevertheless, we acknowledge the value of keeping open the possibility of reviewing and adapting the format and type of data provided to better meet users' needs. Furthermore, we believe that the original version of the graphical information provided lacked an adequate legend to help stakeholders accurately interpret the magnitude and alert levels. In response to the reviewer's suggestions, we have added a sentence reflecting these ideas and updated Figure 3 to strengthen the graphical information. Please, find the new updated sentence in P.7, Operational architecture, L. 152-153 and here:

> Modified manuscript text
>
> "...to understand than statistical analysis of the event. Even so, additional metrics (e.g. total percentage of flooding or maximum flood extent in meters) could be easily included in the e-mail, **and the format and type of data provided could be reviewed with the end-users to ensure that it meets their needs and expectations.**"

and also find the new Figure 3 in P.7, Figure 3 and here:

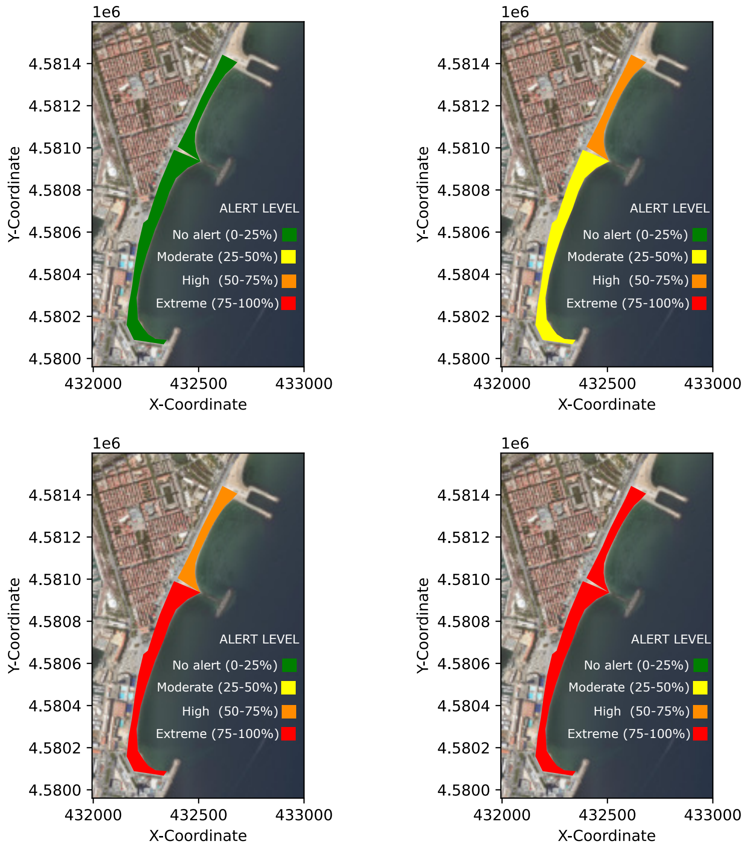


Figure 9: Examples for different alert levels provided within the e-mail. **Percentages represent total flooded area**. From least to most impact: top left, top right, bottom left, and bottom right. Ortophoto obtained from the ICGC WMS

7. Line 158: “Typically the hydrodynamic component of the model is not re-validated. . . ” – I am not sure why this line is included since you do validate the hydrodynamics, would suggest removing as it is confusing.

Author’s reply:

We agree with the reviewer that this sentence could be confusing, as we do validate the hydrodynamic component of the model. Our intention with this statement was to emphasize the importance of this validation, especially since many studies do not validate it and instead rely solely on model performance and previous studies. However, we acknowledge that the sentence may cause misunderstandings and agree that there is no need to highlight the hydrodynamic validation, as it should be performed whenever possible. Therefore, we have removed the sentence in the revised version of the manuscript. Additionally, we have taken the opportunity to revise the placement of the next sentence regarding the model’s usual underestimation of wave height. We believe this sentence disrupted the coherence and readability of the section, especially after removing the one proposed by the reviewer, so we have moved it to the results section, where the underestimation is more clearly demonstrated. We have also made modifications in the results section to further improve readability. Please, find the new updated sentence in P.9, Validation strategy, L. 166-167 and here:

> Modified manuscript text
> ”...First, the correct representation of the hydrodynamics based on wave height was verified using data from a field campaign conducted from 9th March to 27th April, 2022, as part of the MARLIT project (MARLIT, 2021) In...”

and in P.13, Hydrodynamic validation, L. 244-246 and here:

> Modified manuscript text
> ”... the wave height. **This underestimation pattern has been observed in previous studies (De Beer et al., 2021; Oliveira et al., 2020; Bucley et al., 2014)**. Despite this...”

8. Line 193: What is z? Can you explain a bit what the ULISES codes are/do?

Author's reply:

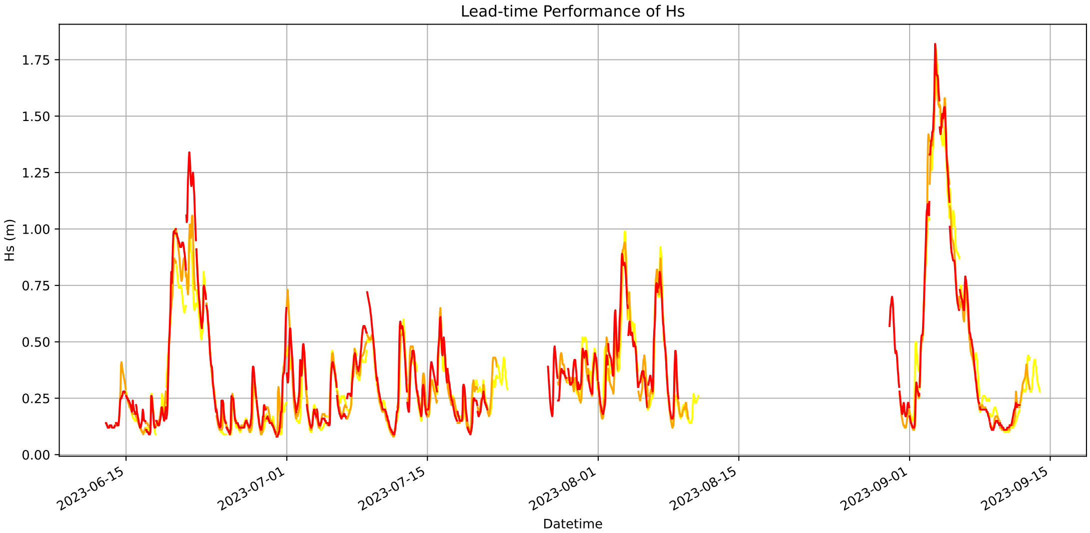
We thank the reviewer for these thoughtful questions, as we believe they will help provide additional context for understanding how this tool works. Regarding the question about z, it represents the water level at the time of the analysis, accounting for both astronomical and meteorological tides. ULISES is the software used to facilitate the transformation of image coordinates into real-world coordinates. We have added a sentence to clarify this concept and provide more context about ULISES software, and also to further explain the definition of z. Please, find the new updated sentence in P.11, Validation strategy, L. 204-207 and here:

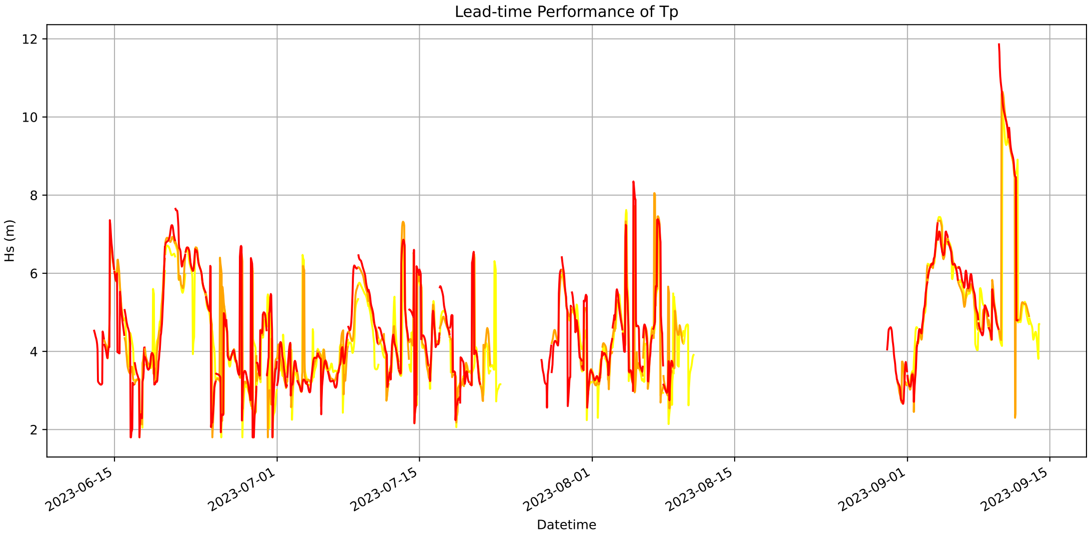
> Modified manuscript text
>
> "The calibration of the cameras allows to transform these pixel coordinates of the shoreline to real world coordinates, provided that z, **defined as the water level encompassing both meteorological and astronomical tides,** is known at the shoreline. **To facilitate this transformation**, ULISES (Simarro et al., 2017)**, an open-source software developed for extrinsic calibration and the generation of plan views in coastal monitoring systems using videos,** and relatives **codes** (https://github.com/Ulises-ICM-UPC) were used."

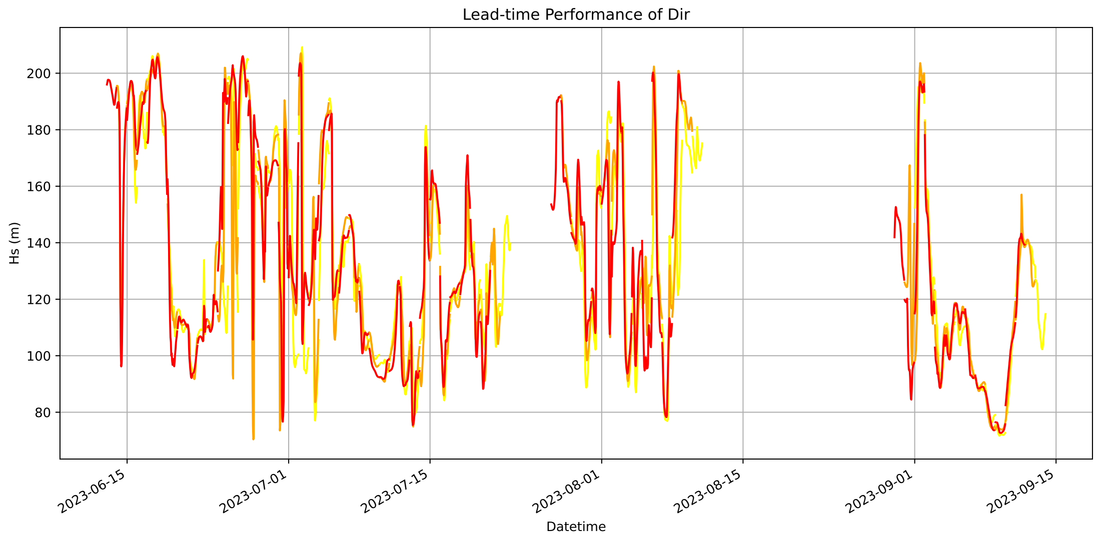
9. Have you looked at all at how the model performance changes with forecast leadtime?

Author's reply:

We appreciate the valuable suggestion provided by the reviewer. The model's lead time performance is directly correlated with the quality of the forecast itself. XBeach is highly dependent on the input data, so when the inputs are consistent and accurate, the resulting outputs are also reliable and consistent. We analyzed the forecast lead time using data gathered since the operational model was implemented. This data was stored for further analysis. However, we chose not to include it in the manuscript, as we felt that adding this information might introduce unnecessary complexity and potentially confuse readers. Nevertheless, to address the reviewer's comment, we have revisited this analysis and are providing additional figures that illustrate how Hs, Tp, and Dir perform over a three-day forecast period with good accuracy. In the images below, the red lines represent forecasts made on the first day, orange lines correspond to second-day forecasts, and yellow lines indicate third-day forecasts (similar to the gray scale in Figure 5 of the manuscript). It is clear that the performance of the third-day forecasts is very similar to that of the first-day forecasts, which are the most reliable. For this reason, we believe that the model's performance does not vary significantly over the three-day forecast lead time. This observation aligns with our initial hypothesis and supports our decision not to extend the forecast period to five days in order to maintain this good lead-time performance, as discussed in the manuscript.







Lead-time Performance of Dir
Hs (m)
200
180
160
140
120
100
80
2023-06-15
2023-07-01
2023-07-15
2023-08-01
2023-08-15
2023-09-01
2023-09-15
Datetime

10. Can the % of flooded cells be calculated from the camera to compare to the model prediction? Were the correct colour alerts issued when compared against the observed flooding?

Author's reply:

We appreciate the reviewer's suggestion. In our initial study, we did not consider this exercise for comparing model predictions. After extensive discussion among the co-authors, we concluded that while it is possible, it would be highly challenging and ultimately not directly comparable with the model results. The main difficulty lies in converting the two extracted lines (baseline and flooding) into a mesh with the same resolution as the model. This process is complex because each studied storm starts from a different baseline, making the generation of comparable cells for percentage computation non-trivial. However, this challenge alone was not the reason we decided not to implement this approach. As discussed throughout the manuscript, one of our main points of concern is that we only have a single topobathymetry, while the studied storms occurred at different times. Since the beach evolves over time, the baseline also changes. Comparing the model with storms that are distant from the time of topobathymetry acquisition, such as those from 2019, is not advisable, as the baselines differ, leading to variations in the total flooded area. The flooding line, which serves as the key metric in our study, does not have this issue since it represents a spatial reference, and the back-promenade remains unchanged. However, as explained in the manuscript, discrepancies arise due to baseline variations, affecting the results. For this reason, we believe that evaluating the comparison between the model and the camera using Euclidean distance is appropriate for our case. We acknowledge that in cases where updated topobathymetries are available for each study scenario, the analysis proposed could enhance the evaluation of model performance. However, given the baseline discrepancies in our study, such an approach could yield results that are difficult to interpret, regardless of whether they appear favorable or not. We hope this explanation clarifies our decision not to pursue this analysis, despite recognizing its potential value for other studies with similar methodologies.

11. Line 319: Reference to “significant resolution difference” between the CMEMS and Xbeach models – what actually are the resolutions for both?

Author’s reply:

In response to Question 4, we have explicitly stated the resolution of the CMEMS grid in the text, as well as the resolution of the XBeach grid (5m) in the Methodology section. This is why we highlight the significant resolution differences between the two. In the Discussion section, we also state why we selected this approach instead of alternative strategies, such as downscaling using SWAN nested runs. To summarize, we used the CMEMS point despite the resolution differences because we only required a single point, positioned close to our computational domain, as input. Additionally, we put significant effort into validating both the hydrodynamics and flooding processes to ensure confidence in this selected point. The resolution-related changes and their implications have been addressed in other questions and parts of the text. We hope that this revision, along with the explanation of the resolution differences and our strategy selection, aligns with the reviewer’s expectations.

12. Line 351: Is the goal for the end users to run the system themselves on their network, or for it to be run eg by your institute and they just receive the output?

Author’s reply:

We would like to thank the reviewer for highlighting this question, as it is an important and insightful point. Our strategy has been designed to be easily implementable on any standard desktop computer, operating automatically once installed. This allows end-users to run the system after installation and a brief explanation of the approach. However, we recognize that some stakeholders, particularly those who may not be familiar with the implementation, might prefer to receive only the final output. To accommodate this, we have incorporated the e-mail sending feature to facilitate communication with such stakeholders. Ultimately, the system is designed to be flexible, allowing either the end-users or the institute to operate it, depending on the specific needs or preferences of the users. We have added a clarifying sentence to address this aspect, which we believe is a valuable point. Please, find the new updated sentence in P.20, Operationality, L. 379-380 and here:

> Modified manuscript text
> ”...open and accessible manner, **allowing them to run it by their own or to receive the outputs via e-mail, depending on preferences and needs**.”

13. Throughout – for a named storm the name should come first ie "storm Celia" and not "Celia storm"

Author's reply:

We fully agree with the reviewer that the name should come first. We have made this correction throughout the document. Please, find the new updated sentence here and along the new version of the manuscript:

Modified manuscript text
"...**storm Celia**..."

Modified manuscript text
"...**storm Gloria**..."

Modified manuscript text
"...**Storms OCT19 and DEC19** triggered..."

Modified manuscript text
"...for the **storm DEC19**, they..."

Modified manuscript text
"...**Storms Celia and Isaak** resulted..."

Modified manuscript text
"...**storms Celia and Isaak**, which had..."

14. Line 77: "being the median grain size" → "with median grain size"

Author's reply:

Please, find the new updated sentence in P.3, Study area, L. 80 and here:

Modified manuscript text
"...45 m, **with** median grain size..."

15. Line 166: “being the signification wave height” → “with a significant wave height”

Author’s reply:

Please, find the new updated sentence in P.9, Validation strategy, L. 172-173 and here:

Modified manuscript text

"...This event exhibited a significant wave height (H_s) at the peak of...

16. Line 281: beaches → beaches’

Author’s reply:

Please, find the new updated sentence in P.18, Discussion, L. 306 and here:

Modified manuscript text

"XBeach was used to simulate the **beaches’** response...”

Author's reply to Reviewer 1:

This paper presents the setup, validation, and application of an operational service for forecasting flood impacts on three urban beaches in Barcelona, Spain, using the XBeach model. Researchers highlight that the operational tool is designed for standard desktop computers to offer a user-friendly, high-resolution system with a three-day early-warning timescale. Validation of the tool is presented by comparing modelled results to a field campaign in March 2022. Video analysis is utilised to further validate model ability to predict flooding. The approach is applicable to other areas where such data is available. The paper is upfront in the drawbacks of the model and approach. Some recommendations for future research are made.
This research is a novel, local-scale application of XBeach to support decision making and flood preparedness in the coastal zone. The paper is well written, and contributes to the field of research on tools to improve coastal resilience. Additional attention is needed to clarify the justification for this tool, how flooding is interpreted and defined, and further validate the accuracy of the warning system. I list these key points below, in addition to numerous suggestions to improve readability.

Author's reply:
We sincerely thank the reviewer for their thorough reading and insightful analysis of our work. The comments provided are highly valuable and will significantly contribute to improving both the manuscript and the study itself. In the following pages, we offer detailed responses to each comment to enhance the quality of the manuscript overall.

Clarify the justification for this tool

The conclusion that bathymetry is important for accurate representation of shallow water wave processes at the coast is not groundbreaking, and would have been known from the outset. So the introduction needs to better set the justification and context for this type of tool, and why an operational model (which is reliant on accurate, up to date bathymetry) is applied here when that bathymetric data isn't available or used.
Consider how other technologies could support delivery of more up to date bathymetries, or how morphological updates / outputs from XBeach could be used to feed into the next days simulation.

Author's reply:
We have worked to improve the overall structure of the manuscript to better clarify the context of the tool. Additionally, in our responses to the reviewers' specific questions, we have addressed these key points in detail. From the outset, as the reviewer state, we hypothesized about the bathymetry influence. Although we understood its importance, we wanted to assess its impact in an operational context, as its magnitude was not entirely clear to us. With this understanding, we

focused on validating it in our approach as part of the analysis. Then, the validation demonstrates that the method performs well at present, but we also aimed to highlight how its accuracy will degrade over time due to these changes in topobathymetry. The camera-based validation has been instrumental in confirming this trend. Our primary objective was to develop and present a methodology for building an operational service that is computationally efficient and easy to implement. Once established, it was crucial to assess whether this strategy was sufficiently accurate for application in the study area and for presentation to stakeholders. As the reviewer rightly pointed out, bathymetry plays a critical role. For this reason, we have proposed various approaches to address this issue in the discussion section. In summary, our main goal was first to develop this operational service and show how to do replicate it and then, verify and demonstrate its performance. To more details, we have addressed the reviewer's suggestions regarding bathymetry updates in our responses to Questions 41 and 42; and for specific discussions about the service and bathymetry in Questions 33, 38 and 39.

How flooding is interpreted and defined

It is not clear what metric is being used to define flooding, so it is difficult to grasp what you are actually presenting in respect to flooding results.

- Is it a maximum water level, where water level exceeds a critical threshold? Flooding occurs where maximum water levels surpass the ground surface elevation.
- Is it maximum landward extent of wave runup over the simulation? Flooding is defined when run-up exceeds coastal defences.
- Number of wet pixels averaged over the simulation? XBeach outputs maps of inundation depth at each grid cell. Is flooding defined as areas where water depth is above a specific threshold?

Would your 'flooding' results vary if a different approach is taken to define extent?

Author's reply:
We thank the reviewer for pointing this out. We agree that the first version did not clearly specify the metric used to define flooding. We have now improved this section to better address this issue. We have provided the responses in Questions 14–17, but in summary, we defined flooding based on the number of wet pixels or cells, as determined in XBeach, using a threshold depth (third option of the ones provided by the reviewer). For a more detailed explanation of our flooding definition, please refer to the mentioned responses (Q. 14-17). We acknowledge that using a different metric, such as the other ones suggested by the reviewer, could yield varying results. However, we believe that our chosen approach was the most appropriate based on the model's outputs definition, the dynamic of the beach and the data available for validation. The model provided the flooded cells, and we applied a threshold condition to enhance the confidence in

flooding detection, considering both wave and water level contributions. We hope that in the revised version of the document, our explanation of flooding definition is now clearer and helps the reader better understand how we approached this aspect in our study and analysis.

Further validate the accuracy of the warning system

I don't think the discussion goes far enough to show how useful the operational tool is. Table 3 shows the warning system that would be assigned to each event that was simulated. It would be nice to see evidence of flooding presented e.g. did Storm Gloria (2020) cause substantial flood impacts (e.g. damage, disruption, need for evacuation)? Would this tool have averted disaster if a red warning was issued to residents 3 days in advance?
Table 3 would also benefit from a column showing maximum significant wave height or maximum storm tide height (tide + surge + max wave height) across the event, for easier comparison. Could you add column 3 and 5 from table 1 together?
Can you comment on whether it is a linear relationship between increasing storm tide and the warning level? What are the thresholds in forcing conditions that cause the warning level to change (i.e. become more severe)?

Author's reply:
We agree with the reviewer that the initial version of the manuscript did not adequately demonstrate the usefulness of the operational tool. To address this, we have revised the document based on the reviewer's insightful questions and suggestions. Specifically, in response to this section's remarks, we have incorporated the return period in Table 3 (Question 20) to clarify the magnitudes of the studied storms. Additionally, we have specified the flooding impacts of these storms throughout the manuscript, as reflected in our response to Question 36. We have also elaborated on how our strategy could help mitigate damages, emphasizing the benefits of a three-day advance warning for implementing mitigation actions (Question 39). Regarding the final point, we do not observe a clear linear relationship between increasing storm tide and warning levels. However, for the same wave magnitudes, a rise in tide levels does result in greater impacts. Flooding in this context depends on multiple variables, including storm tide, wave height, wave period, storm direction, and storm duration. The interplay of these factors determines the final flooding level, making it difficult to establish a threshold for only one variable that directly corresponds to a specific warning level. The only threshold explicitly stated in the manuscript concerns cases where Hs is below 2 m, which we do not classify as a storm, in accordance with standard regional classifications. We hope that the changes made to the manuscript, along with this explanation and our responses to the specified questions, adequately address the reviewer's concerns.

Specific comments

ABSTRACT

1. L1: Begin the abstract with a sentence to introduce the context and need for this tool e.g. no sufficient tools are currently available, needed to avert storm impacts, improve preparedness of coastal communities.

Author's reply:

We fully agree with the reviewer that adding a contextual sentence to highlight the need for such tools enhances the overall quality of the abstract and helps the reader better understand the aims of the study. Accordingly, we have included a sentence following the reviewer's recommendations. Please, find the new updated sentence in P.1, Abstract, L. 1-2 and here:

> Modified manuscript text
> "**Coastal communities are increasingly vulnerable to storm impacts, highlighting the urgent need for predictive tools and enhance preparedness. In this work,** a comprehensive, high-resolution hydrodynamic operational service using XBeach model is presented ..."

2. L2: Explain what the operational tool will do. Move this sentence to earlier in the abstract: 'The operational system is designed to provide early-warning coastal flooding at three-days horizon.'

Author's reply:

We appreciate the reviewer's suggestion and agree that moving this sentence earlier in the abstract will more effectively emphasize the strategy. Please, find the new updated sentence in P.1, Abstract, L. 3-4 and here:

> Modified manuscript text
> "...NW Mediterranean Sea. **The operational system is designed to provide early-warning coastal flooding at three-days horizon**. The operational...

3. L10: Explain if the warning system is tested / validated to demonstrate its usefulness.

Author's reply:

We agree with the reviewer that while the abstract discusses the hydrodynamic and flooding line validation, it does not clarify whether the service has successfully tested or validated to demonstrate its usefulness. To address this, we have added a sentence as suggested. Please, find the new updated sentence in P.1, Abstract, L. 11-12 and here:

> Modified manuscript text
>
> "...The service provides a warning system with a specific categorisation of the event, enabling the end-users to prepare for a possible flooding. **The strategy is currently running in operational mode, issuing alert warnings at the correct severity level**. The...

INTRODUCTION

4. L22: The tool is useful to more than just stakeholders in the coastal zone. It would be useful for coastal managers, emergency services, landowners, businesses, residents. Surely something broader would be better here e.g. 'Decision makers must have suitable tools to. . . '

Author's reply:

We completely agree with the reviewer that these types of tools are useful not only for stakeholders but also for coastal managers, emergency services and residents, among others. We believe that adopting a broader sentence, as suggested by the reviewer, is a much better approach to encompass all these groups. Please, find the new updated sentence in P.2, Introduction, L. 24 and here:

> Modified manuscript text
>
> "...events (Chondros et al., 2021). **Decision makers** must have **suitable** tools to mitigate...

5. Line 29: ‘Among others’ too vague. What other processes are important?

Author’s reply:

It is true that the sentence was not well-written and did not convey the intended message. Our aim was not to include other processes but to highlight that information is needed from the open sea to the beach, including shallow waters, accurate topobathymetric data and beach characteristics for example. We have rephrased this to clarify the point. Please, find the new updated sentences in P.2, Introduction, L. 31 and here:

> Modified manuscript text
> ”... To monitor and forecast coastal areas accurately, it is crucial to use different levels of information, from the open sea related to the beach**, including wave, topobathymetric and morphology data**...”

6. L33: ‘Various models’ too vague. Give examples e.g. numerical or machine learning models, and hindcast or forecast models?

Author’s reply:

We concur with the reviewer that this part of the text was unclear. Our intention was to convey that, when using these types of systems, different numerical models and products are typically integrated to provide comprehensive and advanced information about the area of interest. For example, in our case, we combine data from IBI products, which is derived from numerical models, and integrate it into XBeach, another numerical model. We have rephrased this section in the revised manuscript to clarify that such integration of different numerical models and products is common in these cases. Please, find the new updated sentences in P.2, Introduction, L. 35-36 and here:

> Modified manuscript text
> ”Coastal Ocean Forecasting Systems (COFS) enhance this by focusing on local areas, **usually** integrating data from **different numerical** models, **products** and observations to provide advanced...”

7. L39: 'Leeway' – can this be quantified? How much time is needed realistically? What is the current lead time of forecast systems, and why does this need to be improved?

Author's reply:

We believe that quantifying the leeway needed is difficult and inaccurate, as it depends on both the specific end-user and the forecasting results. In this part of the introduction, we focus on stating how technological advancements have reduced computational times, enabling the implementation of these strategies. This encompasses various forecasting results, not just the one presented in our study.

Nonetheless, for the specific case of mitigating flooding impacts, we believe that such quantification would also rely more on subjective perspectives than on objective procedures. For example, it will vary depending on the end-users, as well as among end-users themselves, based on their decision-making efficiency and their capacity to mobilize mitigation actions.

Regarding lead time, we propose that a three-day is sufficient for the end-users to act if necessary, although, as noted in the discussion, we could extend our forecast to five days within the same framework of delivering results each morning if necessary. However, extending the forecast inherently reduces the accuracy of predictions for the fourth and fifth days. Importantly, improving the lead time of forecasting systems is not the aim of our study. Instead, our objective has been to develop a strategy that reduces computational demands without compromising lead time or the timely delivery of results. Besides, our goal was to work in this field since the number of such kind of forecasting systems that include flooding analysis and simulation of the emerged beach areas is very limited. Finally, based on discussions with potential end-users, we have been informed that this feature significantly enhances the ability to mitigate impacts. For example, as explained in the discussion, it could improve the efficiency of beach flag selection, which is not the initial intended use.

We hope this explanation satisfies the reviewer by clarifying why we leave the quantification of leeway open-ended, while proposing that three days is a sufficient margin for end-users to act and mitigate potential impacts.

8. L53: 'Challenges' – it's not entirely clear from the previous paragraph what these challenges actually are. Would be worth summarising here.

Author's reply:

We agree with the reviewer that, upon reaching this part, it may be unclear what specific challenges are being referred to, as they are not explicitly defined as 'Challenges' in the previous paragraph. We wanted to state that our aim is to highlight that our operational service addresses the need to deliver results in time, with accurate and reliable information, all without requiring a complex HPC environment. To clarify this concept, we have rephrased this part accordingly. Please, find the new updated sentences in P.2, Introduction, L. 54-56 and here:

> Modified manuscript text
> "...The proposed operational service in this work, created to forecast flooding impacts, **is designed to deliver forecasts to end-users with sufficient lead time to act, featuring** a solution that does not require a complex HPC environment. **The EWS's** most important feature is **this** rapidness in obtaining practical results, so the whole..."

9. L57: The strategy uses hydrodynamic information previously computed by other models. I think you are referring to the CMEMs data, but it's not clear here. A reference is needed to explain what model results you are referring to. If it is not the CMEMs data then what is it? What if a user doesn't have access to previously completed modelling results? Does it make the whole approach unfeasible? If so, then details of these previously completed model runs must be included in the methodology too.

Author's reply:

We agree with the reviewer that the original sentence in the introduction, where the overall architecture and strategy have not yet been explained, could lead to confusion regarding which model is being referenced. As suggested by the reviewer, we are indeed referring to CMEMS products. To avoid misunderstanding, we have clarified this point by rephrasing the sentence to explicitly state that we are using CMEMS. Additionally, to cover the other requests of the reviewer, as explained later in the document, if the server request fails, it follows the approach outlined in Table 4. Please, find the new updated sentences in P.3, Introduction, L. 59-60 and here:

Modified manuscript text

"...The developed strategy in this **contribution** does not compromise the resolution of the model, as it utilizes hydrodynamic information **from CMEMS products (https://marine.copernicus.eu/)**. This data..."

METHODOLOGY

10. The methodology needs a data framework diagram to show how the boundary conditions link up. The workflow in Fig 2 is great, but I'm still left unsure how all the boundary conditions and validation data link together. Could sit as a subplot in Fig 2.

Author's reply:

We are a bit unsure on how to deal with this suggestion. The workflow presented in Figure 2 shows the diagram of the operational architecture, which is explained in detail in Section 2.2, as noted in the caption. In this figure, we present this workflow as the standard to build the operational service, following the approach presented in this contribution, to run in an operational environment. Then, the validation data is not included in this diagram, as it was only used to verify that the XBeach model achieves good accuracy and can be integrated into the operational chain with strong performance. Without this validation step, we would not have been able to demonstrate this. It is possible that the explanation and connection between different sections of the manuscript were not clear enough. We hope that the updated version, along with the added explanation, now facilitates a better understanding of this part. Nevertheless, we have provided an example below to illustrate what we believe the reviewer is suggesting. However, we think that incorporating this diagram as a subplot could confuse the reader and overshadow the emphasis on the operational architecture in Figure 2.

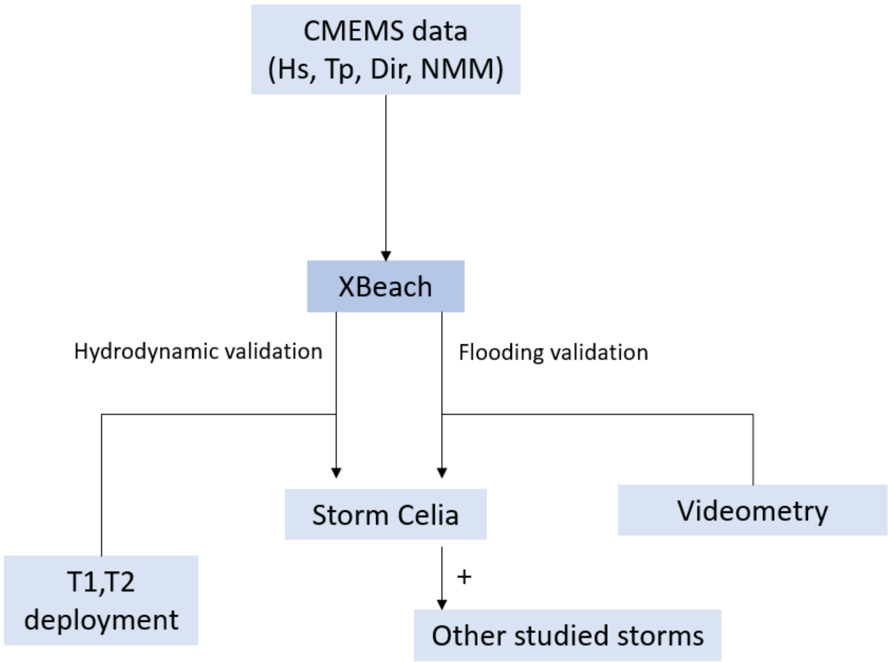


Figure 2: Workflow of the validation strategy

However, it is true that the diagram lacked visual clarity regarding how the boundary conditions are connected, especially the relationship between the CMEMS point and the XBeach boundary conditions. To address this, we have improved Figure 1 to clearly indicate the location of the CMEMS point. Please, find the new updated Figure in P.4, Figure 2 and here:

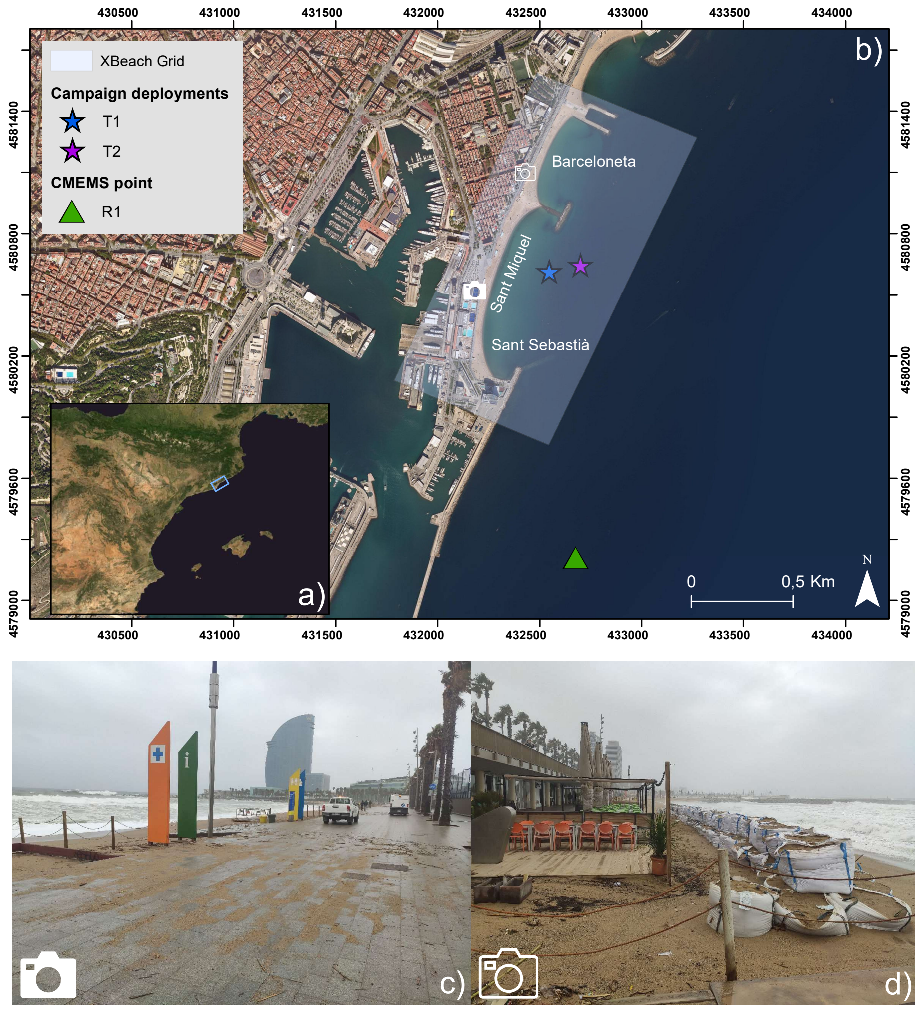


Figure 3: Location map of the study area. a) NW Mediterranean Sea. Blue rectangle marks Barcelona. b) Close-up of southern region of Barcelona beaches. Stars **(T1 and T2)** show the location of the field campaign deployments. **Triangle presents the location of the CMEMS reference point (R1)**. Gray rectangle represents the extent of the computational domain. c) show the ...

11. L93: 'Triggers are selected' – what kind of triggers can be used? Water level, beach level?

Author's reply:

We sincerely thank the reviewer for highlighting this issue. It is absolutely correct that the previous version of the paragraph did not clearly define what "trigger" was referring to, especially since later references to thresholds, such as wave height, could cause further confusion. In this case, we were specifically referring to "initialisation times" which denote the times selected to start the chain and deliver the results, as stated in the subsequent sentence. To address this, we have moved the sentence to the place after the definition of the times used in the presented approach. Additionally, we have explicitly used the term "initialisation time" to enhance clarity and improve comprehension. Please, find the new updated sentences in P.5, Operational architecture, L. 99-100 and here:

> Modified manuscript text
> "...Additionally, the Cylc Workflow Engine (Oliver et al., 2018) could be utilized for both operating systems as well. In our application, the chain starts at 02:00h local time and shares the result of the operational chain at 09:00h through the sending of an e-mail to the competent person. **These initialisation times are arbitrarily selected and subject to changes based on demands.** Data request... "

12. L101: More detail needed on the XBeach set up. Appendix A gives details of the parameterisations, but not enough information is provided on the model domain. What is the size, extent, distance offshore? How many CMEMs data points are used to the force the model and where do the CMEM point(s) sit in respect to the offshore boundary?

Author's reply:

We completely agree with the reviewer that the previous version of the manuscript lacked sufficient detail on the XBeach setup. To address this, we have expanded Appendix A to provide additional information and to better contextualize the computational domain. The previous version also had the resolution of the grid and the configuration placed in the main text (L. 105) and a link to Figure 1 to comprehend the extend of the domain. Regarding the CMEMS data points, it is specified in the text that only one point (R1) is used to force the model. Nevertheless, in response to Question 10, we have included this point in the new map within the new Figure 1 to clearly illustrate its location relative to the offshore boundary and the quantity of points we use in this study. Please, find the new updated sentences in P.5, Operational architecture, L. 105 and here:

Modified manuscript text

"Appendix A provides the parameters **and the setup** used in the simulations conducted for this study. "

and in P.23, Appendix A and here:

Modified manuscript text

"The XBeach initial and boundary conditions, along with the parameters used in the simulation presented in this study, are summarized in Table A1. **The domain extends approximately 1 km perpendicular to the beach, reaching about 600 m offshore, and 1.5 km parallel to the shoreline. The computational domain encompasses all the beaches, which are analyzed separately during the flooding assessment to issue individual alerts.**"

13. L123: Does the three day model time include spin up? How is this accounted for?

Author's reply:

We believe this is a very interesting point. In this case, the three-day model runtime includes the spin-up process. We chose this approach because, as shown in Figure 8 (Hydrodynamic validation), the model quickly reaches a stable state with an accurate representation of the hydrodynamic conditions once initialized (accounting the discussed underestimation). Therefore, the spin-up we considered corresponds to the time it takes for the waves to travel across the domain and reach the coast to initiate flooding since in XBeach, the waves enter from the off-shore part of the domain. Based on our expertise with the model and the size of the domain, the time required for the waves to reach the coast is no more than a couple of hours at most (and it is, in fact, shorter). Given that the forecast period is three days, we consider this runtime sufficient for the model to perform correctly. We did not include this explanation in the manuscript as we believe it could cause some confusion, particularly because the spin-up time in this case is much shorter than in other types of models.

14. L130: How is a flooded cell identified?

Author's reply:

We agree with the reviewer that, although the previous version of the manuscript specifies that the number of flooded cells is determined using the output variable wetz, this may not provide enough clarity for readers unfamiliar with this variable or the method used to compute flooding. Additionally, the manuscript lacked details about whether a depth threshold was applied or considered in the analysis. We do not believe this needs a separate section, as suggested in Question 17, since the amount of information available does not justify an entire section. However, we strongly agree that this information should be properly integrated into the methodology section. To address this, we have reorganized and expanded parts of the text to provide a clearer explanation of the flooding definition and related concepts. In this response, we also address Questions 15–17, as they are closely related. Please, find the new updated sentences in P.5, Operational architecture, L. 127-131 and here:

> Modified manuscript text
> "...during the chain. **Once the simulation is complete, XBeach provides the number of flooded cells using an output variable called *wetz*. This variable identifies wet and dry cells across the entire domain, with flooded cells defined as those that start dry and become wet during the simulation. The total flooding is then calculated as the average number of wet cells over the simulation. To minimize model errors and exclude areas with very thin flooding layers, a cell is only considered wet when it surpasses a threshold of 5 cm. These XBeach flooding results**, which..."

15. Flooding isn't just based on area, but also the depth of the water. How is this accounted for?

Author's reply:

As explained in the revised manuscript (refer to Question 14), to avoid accounting for very thin water layers, we have defined flooded cells as those with a water depth equal to or greater than 5 cm. Conducting a more detailed analysis would require a site-specific definition of vulnerable elements in each study area. For example, a building might be considered "at risk" with a different water depth (and velocity) compared to a person or a public shower. Such an analysis would require an extensive assessment, similar to those conducted by Fukuda for river flow impacts. We have not pursued this detailed approach because our goal was to present a strategy adaptable to any location. Incorporating a more refined assessment of flooding impacts based on water depth would necessitate a tailored definition for each area, which would vary significantly. Therefore, following discussions with stakeholders, we chose to present a scalable definition of flooding based on the area, while incorporating a 5 cm depth threshold to account for this aspect.

16. Is there a minimum depth threshold?

As explained in more detail in Questions 14 and 15, we implemented a minimum depth threshold of 5 cm to minimize model errors and exclude areas with excessively thin flooding layers.

17. A separate section on how flooding is defined, calculated, and identified would be a useful.

Author's reply:

As explained in Question 14, we believe there is not enough content to justify creating a separate section on how flooding is defined. However, we have reorganized some parts of the text and added more information to clarify this point and link it to the definition of the flooded area percentage. We hope that this updated structure and additional details now meet the reviewer's requirements.

18. L134: 25% seems to be quite a high threshold for the green warning. Is there evidence to show that past evidence that would have triggered a green warning did not cause any flooding? How do you know this RAG rating is accurate, and these level of events cause a hazard?

Author's reply:

We appreciate the reviewer's observation, as it highlights an important and interesting point. The selection of percentage levels has been agreed with meetings with stakeholders aimed at establishing a gradual escalation of flooding levels, expressed in whole numbers. This approach delegates the responsibility for interventions and emergency responses to decision-makers, who ultimately determine whether to activate mitigation strategies. To further improve the quality of the provided output and reduce reliance solely on the color-coded system, we have included the flooding line (as shown in Figure 4). This addition enables decision-makers to better visualize the extent of flooding and identify affected areas. Lastly, the validation of our approach has confirmed that the rating system is sufficiently accurate to be considered successful. All studied storm events known to have caused impacts on beaches were correctly categorized with an appropriate level of alert. This outcome underscores the effectiveness of our strategy and the RAG rating system in detecting varying levels of hazard.

19. L144: Can you suggest numerical impacts that could be included here?

Author's reply:

We believe that the term we originally used was not the most appropriate. In this case, "additional metrics" would be a better choice than "numerical impacts," as the latter might be confusing given our use of numerical models. Additionally, as suggested by the reviewer, we have included some of these additional metrics that can be provided to the end-users. Please, find the new updated sentences in <u>P.8, Operational architecture, L. 152-153</u> and here:

<u>Modified manuscript text</u>
"...Even so, **additional metrics (e.g. total percentage of flooding or maximum flood extent in meters)** could be easily included in the e-mail,..."

20. L157: How does the storminess seen in the field campaign period compare relative to the longer term record e.g. within the last 30 years. State the return period of the events simulated, or the percentile each event represents when compared to model or observation data. E.g. does storm Celia represent a 90th percentile event, or a 50th percentile event? This would provide more context for the validation.

Author's reply:

We agree with the reviewer that providing context for the validation and fully understanding the magnitude of the studied storms requires a clear definition of storminess. Since the CMEMS reference point lack long-term records and represent modeled data, we have determined the return period using the nearest buoy, the Barcelona II buoy, managed by Puertos del Estado. To accurately interpret storminess, we have classified some storms with a return period of ≤ 5, indicating they occur, on average, once every 1–5 years. Isaak may seem surprising because, at the CMEMS point, it appears similar to Celia, despite having a 20-year return period at the buoy. This discrepancy arises because the CMEMS and buoy locations differ, and additional factors such as wave directionality can lead to a reduction in wave height at the CMEMS reference point. In response to the reviewer's comment, since only the Celia storm was initially presented in the referenced line, we have updated some sentences after this where all the studied storms are presented, referencing storminess to provide better context and help the reader understand the magnitude of the studied events. Please, find the new updated sentences in P. 12, Validation strategy, L. 233-236 and here:

> Modified manuscript text
>
> "In addition to the storm Celia, which was used for both hydrodynamic and flooding validation, other storms were tested to enhance confidence in the model's flood results. **These storms were** documented by newspapers and the flooding magnitude captured by the camera **was** examined. **Table 1 provides a summary of the storm characteristics, including their storminess expressed in terms of return period (RP) using data from the "Barcelona II" buoy managed by Puertos del Estado, which provides a longer time series....**"

and the updated Table 1 in P. 13 and here:

Table 1: Characteristics for the storm and calm events studied at the CMEMS reference point (R1).

Storm event	Date	Maximum H_s (m)	Maximum T_p(s)	Maximum WL (m)	RP (yrs)
OCT19	October 2019	2.2	9.1	0.89	≤ 5
DEC19	December 2019	3.7	10.5	0.93	≤ 5
Gloria	January 2020	4.7	11.9	0.69	60
Celia	March 2022	4.0	10.0	0.50	≤ 5
Isaak	February 2023	4.0	10.0	0.55	20
Calm periods					
Calm 1	March 2022	0.19	7.9	0.75	-
Calm 2	June 2023	0.5	5.6	0.60	-

21. L164: Similar to above, quantify 'major storm'

Author's reply:

We agree with the reviewer that the term "major storm" is unclear and adds confusion rather than providing meaningful context about the magnitude of the event. As a result, we have revised the sentences to emphasize wave height first, giving greater prominence to the magnitude of the storms, and rephrased this section accordingly. Our aim was to present the characteristics of the storm here, followed by new sentences highlighting the significant impacts on the studied beaches (Question 22). In this section, we have opted not to include the return period (RP) directly, as we believe it is more effective to contextualize it later with the other studied storms. Please, find the new updated sentences in P.9-10, Validation strategy, L. 172-174 and here:

Modified manuscript text

"...**During the field campaign, data from a storm named Celia was recorded** (Figure 6). This event **exhibited a significant wave height (H_s) at the peak of about 4 m and an associated peak period (T_p) of 10 s, with an almost constant wave direction of about 120º at both deployments**...

22. L167: Quantify 'substantial coastal flooding' – how much flooding, what were the impacts?

Author's reply:

We agree with the reviewer that the term "substantial coastal flooding" is too vague and that providing a more specific description of the impacts could improve the reader's understanding of the magnitude of the event. Previously, we referenced Figure 1 without sufficient context. To address this, we have rephrased this section to clarify that the event caused damage to beach bars and equipment such as showers, while also displacing sand onto the promenade. Additionally, we now mention that sandbags were deployed in an attempt to mitigate the damage. Please, find the new updated sentences in P.10, Validation strategy, L. 174-176 and here:

Modified manuscript text

"...The 15-day storm caused **significant coastal flooding on the beaches, damaging infrastructure such as beach bars and showers, displacing sand onto the promenade, and prompting the placement of sandbags to mitigate further damage, as illustrated in Figure 1c,d**.....

23. L168: Provide the exact dates

Author's reply:

We have included the exact dates for these minor storms, which are also shown in Figure 6. Also, to complement and address Question 24, we have clarified that these storms are classified as minor and were therefore excluded from the study. The wave heights during these events were less than or close to 2 meters, which is the threshold used to define a storm. Additionally, some co-authors monitorised these events and they observed no significant flooding. During this period, we prioritized studying the Celia storm and then selected other storms known to have caused flooding impacts to provide greater context and variety in the study. As mentioned, we have revised the sentences describing these storms accordingly to make these points clear. Please, find the new updated sentences in

P.10, Validation strategy, L. 176-178 and here:

> Modified manuscript text
> "...Additionally, two **minor storms, occurring from April 11–14 and April 20–22, originating from the East and East-Northeast, respectively,** were recorded at the end of the campaign. **However, these were excluded from the study due to their low wave heights.**...."

24. Figure 6: Annotate the storms, including aforementioned return period or percentiles, and exact dates.

Author's reply:

Continuing with Question 23, we believe that including the return period for these minor storms would not be appropriate for the reasons previously mentioned. In this case, their return period would be less than one year, which is not meaningful to report. As also explained in Question 23, we have now clarified that these storms were excluded from the study due to their low wave heights. For the Celia storm case, as noted in earlier responses, the return period has been included in Table 1, as we agree with the reviewer that it provides valuable context regarding the magnitude of the event.

25. L179: Amend this sentence to: 'Therefore, for calm periods, both approaches are good' just to clarify what you mean.

Author's reply:

We have amended this sentence to clarify the meaning , as suggested by the reviewer. Please, find the new updated sentences in P.10, Validation strategy, L.189 and here:

> Modified manuscript text
> "...monitoring. Therefore, for calm periods**, both approaches** are **appropriate**, but..."

26. L181: Clarify what an ‘Argus-like station’ is.

Author’s reply:

We agree with the reviewer that clarifying the term ”argus-like station” is necessary. It refers to a fixed video station that captures images hourly. The term was originally used to honor ”ARGUS,” the first project to deploy this type of station. However, we recognize that this phrasing might be unclear or confusing to readers unfamiliar with the context. Therefore, we have revised the sentence to explicitly state that we are referring to a fixed video station. Please, find the new updated sentences in P.10, Validation strategy, L. 191 and here:

> Modified manuscript text
> ”Images of one of the cameras from the **fixed coastal video monitoring station** in Barcelona city are ”

27. L183: You selected one out of six available cameras. Would the results have been different if you used a different camera? Have you tested this? Clarify why you picked the one you did.

Author’s reply:

We thank the reviewer for pointing this out. The camera used in this work is the only one, among the six available, that overlooks the beaches of interest. The other cameras are positioned in different areas, allowing to monitor the whole 4 kms of Barcelona beaches. The previous version of the manuscript did not clarify this, which may have led to the misunderstanding highlighted by the reviewer. We have included this information in the revised version. Please, find the new updated sentences in P.11, Validation strategy, L.194 and here:

> Modified manuscript text
> ”...resolution. The **only** camera over-viewing the beaches”

28. L186: Define a flooding line and how this is identified.

Author's reply:

In this study, we defined the flooding line as the maximum flooding observed in each timex image. For selecting the used images, we conducted a quality assessment to identify those where the maximum flooding was clearly visible within the studied time period, which were then manually identified using Python scripts. Since the timex images were primarily used for validation, our focus was on specific time points rather than averaging flooding over the entire event. This approach enabled us to select the most accurate images for determining the flooding lines. The paragraph has been restructured to provide a clearer definition of the flooding line and to enhance understanding of the methodology. Please, find the new updated sentences in P.11, Validation strategy, L. 201-207 and here:

> Modified manuscript text
> "...where the cameras are placed) or due to undesired sudden changes. **For this study, these timex images were used to determine flooding lines, defined as the maximum flooding observed in each image. Lines were manually digitised using Python scripts, with points selected and then converted to pixel coordinates. The calibration of the cameras allows to transform these pixel coordinates of the shoreline to real world coordinates, provided that z, defined as the water level encompassing both meteorological and astronomical tides, is known at the shoreline. To facilitate this transformation, ULISES (Simarro et al., 2017), an open-source software developed for extrinsic calibration and the generation of plan views in coastal monitoring systems using videos, and relatives codes (https://github.com/Ulises-ICM-UPC) were used.** "

RESULTS

29. Validation isn't really a result. I would move this to its own section.

Author's reply:

We sincerely appreciate the reviewer's suggestion. However, in this case, we believe that this part is best suited to remain in the Results section for a couple of reasons. First, validation is one of our primary parallel objectives. Our manuscript presents the operational strategy, followed by the validation strategy, and then reports the results for both. Specifically, we include the metrics that demonstrate the performance of the validation, which we believe qualifies as results. Finally, we showcase how the operational strategy performs across several known storms. In our view, the validation results are indeed results because they contain quantitative metrics and insights derived from our methodology. What does not belong in the Results section is the explanation of the validation process and the motivation behind it, these are appropriately included in the methodology section. Furthermore, if we were to remove the validation results from the Results section, it would leave this section with only the analysis of the known storms and the corresponding color-alert outcomes, which we feel would result in an poor and short section. Ideally, we might have preferred a different organizational structure for the manuscript, one that does not strictly adhere to the conventional Introduction-Methodology-Results-Discussion-Conclusion format. However, since the journal put a preference in this structure, we have opted to include the validation results in the Results section to provide a more comprehensive view of our findings. We hope that this explanation clarifies our reasoning and is acceptable to the reviewer. Nevertheless, we greatly value the suggestion, which we discussed thoroughly among the co-authors.

30. L226: remind the reader what M1 and M2 are here. Also consider giving them more obvious names. These make me think of tidal constituents which confused me. Use T1 or T2, standing for Tripod instead.

Author's reply:

We agree with the reviewer that the previous labels "M1" and "M2" could lead to confusion. In line with the reviewer's suggestion, we have renamed them "T1" and "T2," where "T" stands for tripod. This change has been implemented consistently throughout the text, replacing all instances of "M1" and "M2" with **T1** and **T2**. We have also reminded the reader what T1 and T2 are in this section. Please, find the new Figure 1 in P.4, Figure 1 and here:

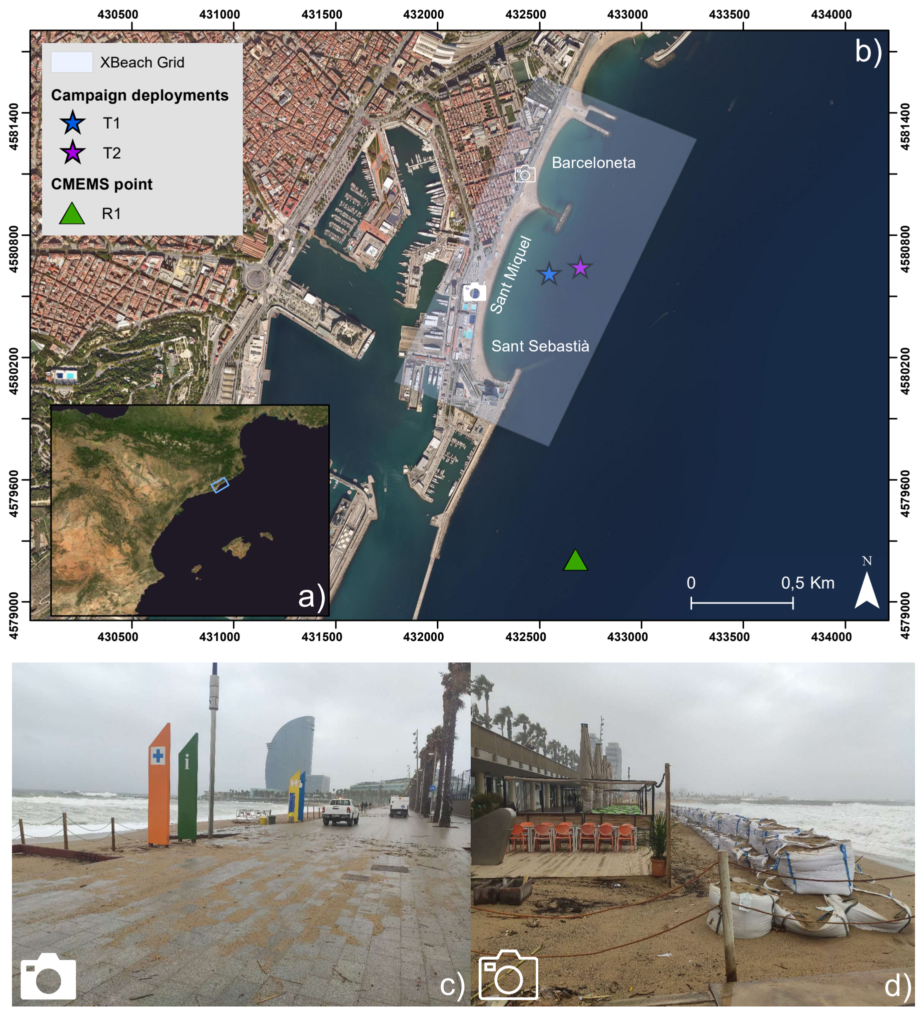


Figure 4: Location map of the study area. a) NW Mediterranean Sea. Blue rectangle marks Barcelona. b) Close-up of southern region of Barcelona beaches. Stars **(T1 and T2)** show the ...

and the new sentence in P.13, Hydrodynamic validation, L. 241-244 and here:

Modified manuscript text

"Figure 8 illustrates the comparison between the model results and observed values from both **field campaign deployments** T1 and T2 **(Figure 1)**, during storm Celia. At T2**, the one placed deeper,** the model shows a closer match with the observed data at the storm's peak, whereas for T1, **the shallower**, it slightly..."

31. L229: could underestimation be because you are only using one CMEMs point to force the model. Would a space – varying boundary condition be more accurate? Consider giving a reason for this.

Author's reply:

The reviewer's suggestion regarding the reason for the underestimation is very insightful. However, in this case, we believe it is not due to using a single CMEMS point to force the model. This underestimation is clearly observed in the shallower deployment, but not in the deeper one during the storm peak, typically corresponding to the maximum flooding moment, where this issue does not appear. Therefore, we concur with the literature's findings that the model has limitations in accurately reproducing this type of propagation especially for enclosed beaches. Additionally, the other available CMEMS point is located much farther from the computational domain compared to the one we selected, making its use less reasonable in this context. While we did not implement space-varying boundary conditions due to this significant distance to the next available point, other studies and works inside the academia not published involving co-authors applying space-varying conditions have shown that for domains of this size—typical for XBeach, the differences in wave height are minimal, although differences in wave direction may be more pronounced. Thus, while the underestimation could be partially attributed to directional spreading differences, we do not consider this the main factor, as the model performs better for the shallower deployment, as previously mentioned. We hope this explanation addresses the reviewer's concerns regarding our choice of using one CMEMS point and provides clarity on why we believe space-varying boundary conditions would not resolve the observed underestimation.

32. L231: 0.49 m is a big RMSE for a wave height range < 4 m. You should also present a mean bias to see if this error is consistent.

Author's reply:

We acknowledge that an RMSE of 0.49 m may seem large for the shallow deployment, especially when compared to the wave range. To address this, we also analyzed the SK, which provides a normalized measure of model performance relative to a baseline. With SK values ranging from 0 to 1, we believe that achieving an SK of 0.8 represents a strong result that underscores the model's capability. As discussed in the text, the underestimation observed with XBeach is a well-documented challenge and is difficult to resolve given the limitations of our current data. The goal of this exercise was to demonstrate that the model successfully captures the storm dynamics. While there is some underestimation, the model follows the observed patterns with good accuracy. However, we agree with the reviewer's suggestion to include the mean bias, as it can provide additional clarity about whether the error is consistently an underestimation or more variable. In the initial version, we opted not to include it because we believed the underestimation appeared clear. However, we recognize that incorporating mean bias could enhance the analysis and offer further value to the reader. Please, find the new updated sentences in P.11, Validation strategy, L. 209-210 and here:

Modified manuscript text

"For each type of validation, different error metrics were considered based on data characteristics. The hydrodynamic model's performance was calculated through the Root Mean Square Error (RMSE)**, the mean bias error (MBE)** and the skill score assessment index (SK) (Equations **2,3 and 4**). RMSE is a widely-used and reliable metric for assessing the relationship between two data series, making it an appropriate choice for evaluating these types of results and models..."

and in P.12, Validation strategy, L. 211-212 and here:

Modified manuscript text

"...RMSE is a widely-used and reliable metric for assessing the relationship between two data series, making it an appropriate choice for evaluating these types of results and models**, MBE provides additional insight by quantifying the systematic tendency of the model to overestimate or underestimate observations...**"

and in P.12, Validation strategy, Equation 3 and here:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i) \tag{0.1}$$

and in P.13, Hydrodynamic validation, L. 244-245 and here:

Modified manuscript text

"...This underestimation pattern **remains consistent throughout the entire event, as evidenced by the negative MBE of -0.43 m for T1 and -0.25 m for T2, and** has also been observed in previous studies.."

and in P.14, Hydrodynamic validation, Figure 8 and here:

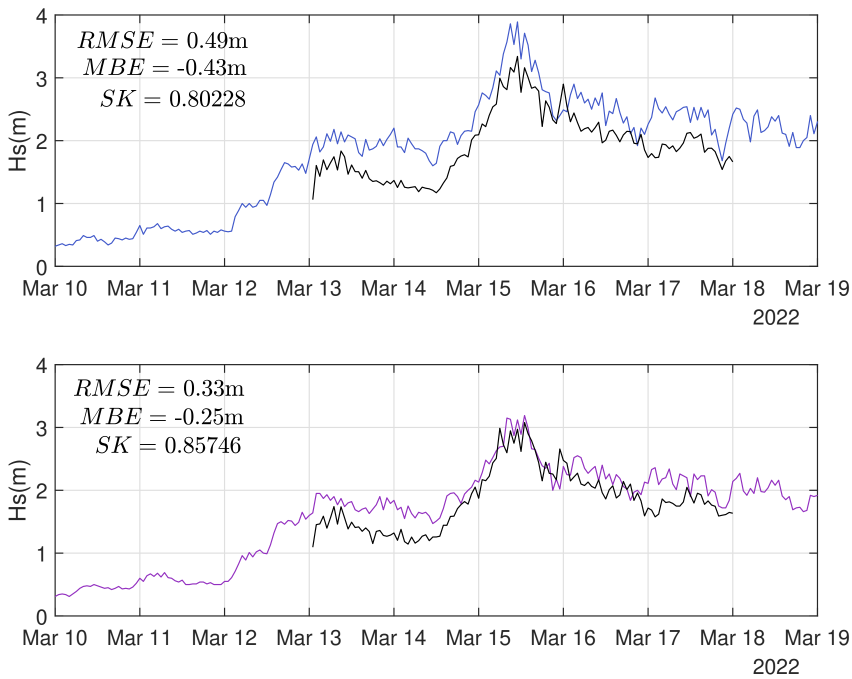


Figure 5: Hydrodynamic response validation of the model. Top panel and the light blue line corresponds to T1 while bottom panel and purple line to T2. Black lines show the model output for each point.

33. Section 3.2: The main results about the important of up-to-date bathymetry gets lost here. I have read this section several times and still can't quite get my head around what you are trying to show. I think this is where you show that duration of time after last bathymetric survey reduces the accuracy of the model results. Can you plot how RMSE in model results increases with duration after last bathymetric survey? More detail is needed in L244 – 254 to explain when the last survey dates were relative to the model results.

Author's reply:

We thank the reviewer for this insightful comment. In this section, we aim to demonstrate two key aspects, as indicated in the title. First, the flooding validation, which provides confidence in the good performance of the model. To achieve this, we compare the flooding lines obtained from camera observations with those generated by the model. The methodology for this comparison is explained in Line 222, starting with the sentence, "To validate flooding...". We believe that our previous version of the manuscript did not correctly specify the metric used to represent model performance. Instead of RMSE, we employed the averaged Euclidean distance for a similar purpose. To clarify this, we have updated the text and equation to explicitly state that we used the averaged Euclidean distance as the performance evaluation metric. Please, find the new updated sentences in P.12, Validation strategy, L.224-230 and here:

> Modified manuscript text
> "...the Euclidean distance was used to analyse the differences between **both observed and modelled. Based on these values, the averaged Euclidean distance was computed as Equation 4. In this equation,** n **represents the total number of points**, p_1 and q_1 represent the x-coordinates, while p_2 and q_2 represent the y-coordinates of the points (p) and (q), respectively. **The averaged distance represents an error metric of the model performance**. "

and in P.12, Equation 5 and here:

$$\overline{d(p,q)} = \frac{1}{n}\sqrt{\sum((p_1 - q_1)^2 + (p_2 - q_2)^2)} \quad (0.2)$$

We have attempted to generate the plot suggested by the reviewer using the averaged Euclidean distance, which serves as the performance evaluation metric in this case and acts similarly to RMSE. Below, the reviewer can observe this plot, where the evolution is visible.

However, due to the limited number of data points and the challenges related to Gloria (as discussed in the Results and Discussion section), it is not entirely straightforward to interpret. We did not include the calm period in 2022 because the plot represent data on a monthly scale rather than daily, and this calm period overlapped with the solid line, causing potential confusion. Ultimately, we decided not to incorporate this plot into the revised manuscript, as we believe it could confuse the readers. However, we sincerely appreciate the reviewer's suggestion, as it prompted us to generate and try the suggestion and improve the overall definition of Averaged euclidean distance.

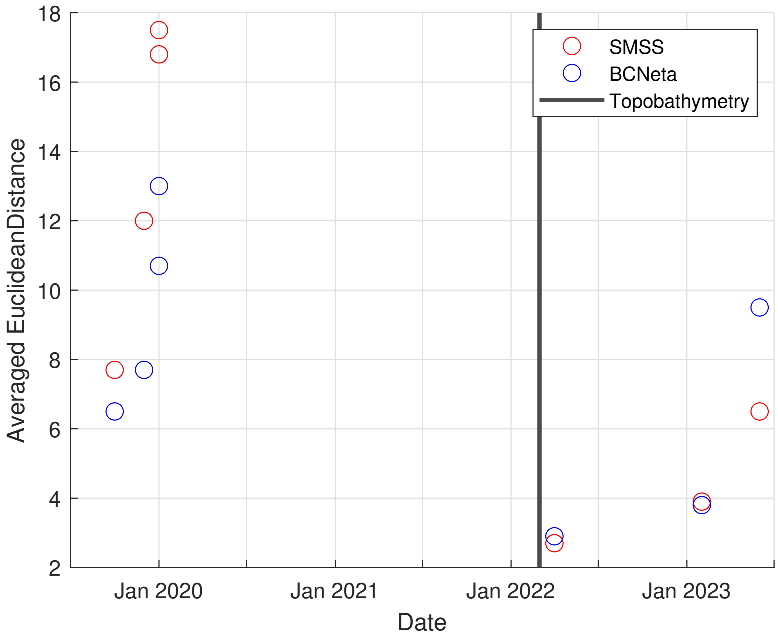


Figure 6: Averaged Euclidean distance evolution.

The second aspect we aim to demonstrate, which the reviewer has correctly identified, is how the duration of time since the last bathymetric survey affects the accuracy of the model results, this is referred to as our sensitivity analysis. We fully agree with the reviewer that there was a huge lack of contextualization in the previous version to help readers understand the time gaps between the survey and the studied storms. Previously, we only mentioned the timing of the bathymetric survey once in the methodology section. In this revised version, we have made several updates to improve clarity: we now reiterate when the topobathymetric data was gathered and provide more detailed comparisons to highlight how the accuracy of the model decreases as the studied storms occur farther from the survey date. We hope that the revisions in this updated manuscript, along with this detailed explanation, address the reviewer's concerns and align with their expectations. Please, find the new updated sentences in P.14, Flooding validation and sensitivity analysis, L.253-256 and here:

> Modified manuscript text
>
> "Figure 9 **presents the flooding lines obtained from both the videometry and the model** for SMSS and Barceloneta beaches during the studied storm events. Figure 10 presents **these** results for both SMSS and Barceloneta beaches under the calm condition scenarios. Additionally, Table 2 summarizes the **averaged** Euclidean distances obtained **from the comparison between videometry and model using Equation 5 for all the studied cases**. "

and in P.14, Flooding validation and sensitivity analysis, L.257-258 and here:

> Modified manuscript text
>
> "In the context of the storm Celia, which **hit the coast only 10 days after the collection of the** topobathymetric data **on March 4th**, the comparison depicted in Figure 9 reveals a strong correlation between the model and the camera across SMSS beaches, as well as with Barceloneta beach. Supporting this observation, Table 2 provides quantitative evidence with obtained values from the Euclidean distance analysis of 2.7 m for SMSS beaches and 2.9 m for Barceloneta. "

and in P.15, Flooding validation and sensitivity analysis, L.266-267 and here:

> Modified manuscript text
>
> "Specifically, the biggest differences are observed for the storm Gloria, **which happened two years before the topobathymetry gathering**. **Additional** causes..."

and in P.15, Flooding validation and sensitivity analysis, L.272-274 and here:

> Modified manuscript text
>
> "As shown in Table 2, storm Celia had the best approximation, followed by Isaak **(that occurred one year after the topobathymetry collection)**, both achieving subpixel accuracy. Conversely, the accuracy decreased for storms that were farther away, with OCT19 storm of 2019, **three years before the gathering,** showing the highest euclidean distance value, of 12 m, excluding storm Gloria"

34. L253: Add in date of: 'with smaller distance discrepancies for storms closer to the date of topobathymetry gathering'.

Author's reply:

As addressed in the response to Question 33, we have revised the text to provide proper contextualization of the dates. With these changes, we believe the sentence is now clear and adds coherence to the overall structure.

35. L257: This is where a clearer definition of flooding would help. It's difficult for the reader to visualise what you are describing here. I it depth or extent of flooding?

Author's reply:

We agree with the reviewer that the lack of explanation in the previous version regarding how flooding is computed made it difficult to fully understand what we were trying to convey. In the revised version, we hope this issue has been solved. Additionally, we have incorporated the term "extent" in this part, as suggested by the reviewer, since we believe it provides valuable context and improves the readability of the sentence. Please, find the new updated sentences in P.15, Flooding validation and sensitivity analysis, L.278 and here:

> Modified manuscript text
> "...Comparing with SMSS, Barceloneta better catches the flooding **extent** for all the studied storm scenarios except for Celia..."

36. Section 3.3: See comments at start of review: you need to better say if the flood impacts did occur during Celia / Gloria. Did severe flooding happen when you predicted it would? Would this tool have averted flooding?

Author's reply:

We agree with the reviewer's suggestion that this section needs to better address whether flooding impacts occurred during the storms and provide context regarding the resulting damage. As now specified in the revised version of the manuscript, Storm Gloria was the largest storm ever recorded in Catalonia, causing significant damage along the coast, including casualties. In Barcelona, its beaches suffered severe impacts, with flooding affecting almost the entire beach and huge causing damages to infrastructures. For Storm Celia, the damages were also substantial, prompting the implementation of quick defense measures, such as the placement of sandbags, in an attempt to mitigate its effects, as it can be observed in Figure 1c,d. Our proposed strategy could have provided advance warnings of such flooding, enabling more effective action to reduce its impacts. For example, sandbags could have been positioned in a more strategic and effective manner, or the promenade could have been closed with sufficient lead time to ensure public safety. We have added information to Section 3.3 to clarify the real-world effects of these storms. Additionally, we have included a sentence at the end of the "Modeling Uncertainties" subsection in the Discussion to explain how our approach could have helped reduce the associated impacts. Please, find the new updated sentences in P.17, Operational analysis, L. 293-295 and here:

> Modified manuscript text
> "...Storm Gloria is the only one reaching a red alert level, indicating an extremely high alert, with flooding areas of 81.3 % at SMSS beaches and 83.2 % at Barceloneta. **This storm broke historical records for significant wave height and caused widespread damage along the entire Spanish Mediterranean coast, including severe beach erosion, extensive damage to coastal infrastructure, material losses due to flooding, and even some casualties (Sotillo et al., 2021)**..."

and in P.17, Operational analysis, L. 300-301 and here:

> Modified manuscript text
> "... The flooding areas for Celia and Isaak at Barceloneta were 33.9 % and 45.8 %, respectively. **Each of these storms also caused significant flooding impacts and infrastructure damages to the studied beaches. In particular, some impacts of Storm Celia are illustrated in Figure 1c,d**..."

and in P.18, Modelling uncertainties, L. 364-366 and here:

> Modified manuscript text
> "...and demonstrates that even though the accuracy concerning flooding lines is reduced over time, the approach remains capable of detecting the storms and provide alerts. **The service enables accurate forecasting and timely warnings, allowing for more effective interventions to mitigate the associated impacts. For instance, sandbags or other quick defense measures could be strategically positioned, or the promenade closed with adequate lead time to ensure public safety.**"

37. Table 3: What is a flooding percentage? Colour code the table rows with the alert warning colour. Add a column showing the written alert warning too. At the moment we have to flick back to page 6 to understand it all. make it easy for the reader to understand this, it's your main result!

Author's reply:

We completely agree that a more specific caption is necessary to help readers better understand the table. We have addressed this in the revised version. Regarding the colour code, we are very grateful to the reviewer for pointing this out. In the initial version of the manuscript, we had indeed used a color code for the table rows, as we shared the same thought as the reviewer. However, we had to remove it due to the submission guidelines, which specify that "colored table cells should be avoided." Thanks to the reviewer's suggestion, we revisited this issue and found a solution: we have now included colored squares within the table cells as text to represent the alert levels for each storm and beach. We believe this adjustment provides significant additional context to the table while adhering to the journal's guidelines. Please, find the new table in P.18, Table 3 and here:

Table 2: **Percentage of beach flooded area, calculated as the ratio between the number of flooded cells and the total number of cells within the model. Colored squares indicate alert levels: green - no alert, yellow - moderate alert, orange - high alert, and red - extremely high alert level.**

Storm	SMSS	Barceloneta
OCT19	51.3 %	53.4 %
DEC19	68.9 %	60.0 %
Gloria 2020	83.2 %	81.3 %
Celia 2022	51.0 %	33.9 %
Isaak February 2023	60.0 %	45.8 %
Calm March 2022	<0.1 %	<0.1 %
Calm June 2023	<0.1 %	<0.1 %

DISCUSSION

38. Reiterate again at the start the overall aims of the research

Author's reply:

We agree with the reviewer that the aims of the research should be included in the Discussion section. These aims are, in fact, specified at the beginning of the second subsection, which focuses on operationality. We structured it this way because we first wanted to discuss the methodology and the model's performance, which ultimately led us to achieve the study's objectives. The discussion of these aims is located in Section 4.2 (Operationality), where we describe how we developed this comprehensive methodology and evaluated its effectiveness on a standard desktop computer, ensuring it provides timely forecasts. Importantly, these goals are not directly related to the model's performance itself. Therefore, we chose to reiterate the study's aims in the second subsection rather than the first. While we could position this subsection earlier in the Discussion, we believe its current placement draws greater attention from the reader by being closer to the conclusion, enhancing the coherence and flow of the manuscript. We hope the reviewer finds our strategy valid, as reiterating the aims in this way allows us to maintain logical consistency while keeping the reader engaged.

39. Discussion focuses on the model performance, rather than on the usefulness and application of the tool. Can you infer the wave and water level conditions that would cause a red warning alert? Consider how these results are helpful for a decision maker and what information could be taken directly from this paper and applied to improve preparedness now.

Author's reply:

We appreciate the reviewer's suggestion regarding the inference of wave and water level conditions that could trigger a red warning alert. We agree that this approach is highly valid and could serve as a parallel strategy to our current method. In our study, we based the forecasting outputs and results on a more deterministic approach, relying on a well-validated model. The approach suggested by the reviewer, as we understand it, adopts a more probabilistic perspective. By analyzing all possible combinations of wave and water level conditions, it seeks to determine whether specific values of these parameters would result in an alert. This is an extensive and complex analysis that could complement our results or even stand as an independent method for defining alerts. However, we did not pursue this approach in our study because we believe that alerts are influenced by a wide range of variables, including wave and water levels, wave direction, topobathymetry distribution, storm duration, and storm peak duration, among others. To perform such an analysis effectively, a more extensive dataset would be required. We agree that this approach would be valuable for future research. Regarding the reviewer's request to enhance the discussion and clarify how our results are helpful for decision-makers, we have revised the manuscript to address this throughout the text, answering previous reviewers' comments. For instance, in the Discussion section, we added the following to emphasize the practical applications of our system:

- "The system has been developed to provide advance notice of potential flooding impacts during high-intensity storm conditions, allowing for the implementation of quick defence measures to mitigate associated damages as for example in the promenade, showers or beach bars."
- "the system can help to prepare in advance the decision for a beach closing to recreational use with red or yellow flags."

We believe these points highlight the value of our strategy for end-users and decision-makers. We hope this explanation clarifies why we did not implement the probabilistic inference analysis, despite recognizing it as an interesting and promising approach, and that the revised discussion meets the reviewer's expectations.

40. What would you need to do to change this into a forecasting too? you use hindcast data, so how different would the setup be for forecasting? Would the same CMEMs data be suitable?

Author's reply:

We are not entirely sure what the reviewer intends to express in this question. Although we validated the performance of the model using past storms and the corresponding CMEMS data, the strategy presented is fully designed to operate in an operational forecasting mode. The entire operational architecture enables us to provide end-users with a three-day forecast, as demonstrated in Figure 5. It was not possible to validate the model using actual forecasts because this would require future events, which are inherently unpredictable. However, as stated in the Operationality section of the Discussion, the system has achieved a Technology Readiness Level (TRL) of 6. This means that we have conducted operational forecasting tests in collaboration with some end-users, who received daily emails containing beach condition updates and graphical information, as shown in Figures 3 and 4. The main objectives of this paper was to present the proposed strategy, highlighting its ability to deliver a low-computational-cost operational service, and to demonstrate that the system has been thoroughly validated, ensuring high confidence in its performance. With this approach, people can reproduce it in other areas or similar studies. Furthermore, the CMEMS data specified in the Methodology section (IBI) has been thoroughly tested in forecasting mode and is proven to provide accurate inputs for the modeling chain. We hope this explanation clarifies that our strategy is fully capable of delivering reliable forecasting results to end-users.

41. Have you considered how higher resolution mapping of intertidal areas, e.g. using a standard marine X-band radar could provide updated morphology (https://www.sciencedirect.com/science/article/abs/pii/S0169555X16306493; https://www.tandfonline.com/doi/full/10.1080/1755876X.2018.1526462)?

Author's reply:

We thank the reviewer for providing such interesting references. We had considered using a similar approach, but we believe we face two key challenges. First, to calibrate and validate such products with sufficient confidence, access to local bathymetric data is highly advisable, which we currently lack. Second, as these products reach greater depths, their reliability decreases. In such cases, we would need to merge these updates with existing deep bathymetries, which could introduce grid inconsistencies and, consequently, errors in the model results. Additionally, this updated section would also need to be integrated with the topography, creating additional challenges in ensuring a seamless connection between datasets. For these reasons, we have not incorporated such strategies into the manuscript at this time. However, we are keenly aware of these challenges and plan to explore them further to enhance the approach presented in this study.

42. XBeach can evolve morphology over a simulation – would it be feasible to update the morphology in the model each day and then use this as input for subsequent runs? It would be challenging to validate without observation, but could be an additional use of the model.

Author's reply:

We agree that this is an interesting discussion. In terms of computational power, it would be perfectly feasible, as the addition of this feature does not significantly increase simulation time. However, some key challenges arise when considering transferring the morphology evolution from one simulation to the next, instead of reverting to the initial state as per our current strategy. First, based on our expertise with the model and insights from the literature, we have observed that XBeach performs very well under storm conditions but struggles with mid-term evolution and recovery processes. The issue with carrying forward the same topography is that after a storm impacts the coast and alters the beach equilibrium, the model may fail to properly recover to the new state. Additionally, in the case of Barcelona beaches, and similarly for other beaches where our approach could be applicable, human interventions such as beach nourishments or other maintenance actions are often implemented to maintain the beach for tourism purposes. Transferring the topobathymetry between simulations does not account for these interventions and would ultimately lead to the same problematic. Second, as the reviewer pointed out, it would be extremely difficult to validate over time whether this strategy of passing the topobathymetry between simulations works accurately. This is because long-term validation would require months of data to adequately assess and demonstrate that the evolution remains consistent across different seasons, which is very costly. For these reasons, we suggest in the discussion that a more effective approach would involve periodically updating the topobathymetry, focusing on critical areas. To achieve this, we propose the use of videometry as a potential solution. We hope this explanation clarifies our reasoning for not incorporating the transfer of topobathymetry from one simulation to the next into our current strategy.

CONCLUSIONS

43. is fine as it is, but some of the additional analysis including mean bias and defining red alert conditions could be included.

Author's reply:

We appreciate the reviewer's comment regarding the conclusions. We also find the suggestion of incorporating mean bias analysis and defining red alerts to be valuable. Initially, we planned our conclusions to be structured as a synthesis of the key findings from the study, without delving into specific details. For instance, we discussed the hydrodynamic validation but do not explicitly state the obtained RMSE, SK, or mean bias. Similarly, we mentioned the color-based alert definition and its use in storm detection validation without elaborating further. We believe that incorporating a bit of information in the conclusions with a concise format could help to effectively summarize and highlight the novelty of our work and the key aspects addressed in the manuscript. However, incorporating these details on hydrodynamic validation would require too much additional information, including mean bias, RMSE, and SK, to be sufficiently comprehensive. For this reason, we have added text discussing the red alert definition but have decided not to include mean bias, as doing so would necessitate the inclusion of other additional details. Please, find the new updated sentences in P.22, Conclusions, L.419-420 and here:

Modified manuscript text

"...The defined color-based alert system has been used to characterize the validated storms, **ranging from red alert conditions, indicating extremely high flooding, to green alert, signifying no hazard. The methodology has demonstrated the ability of the service to detect flooding hazards**..."