The manuscript presents an interesting framework for setting thresholds for estuarine compound flooding using combined hydrodynamic statistical techniques. The approach is applied to Conwy Estuary, North Wales, a particularly vulnerable area to compound flooding. It represents a current thematic area and can be particularly useful in improving compound flooding assessment in estuaries worldwide. The study's main objective is to identify the coastal and fluvial conditions that lead to flooding in an estuarine system.

The manuscript is well-written (English level) and presents high-quality scientific content with interesting results and discussions. No doubt, a lot of work has been done, and the research is highly relevant. However, the connection between chapters (structure) needs improvement. In addition, the authors make some general assumptions throughout the text about specific terms that can confuse the reader, especially if he/she is not familiar with the United Kingdom or the field of coastal/estuarine flooding (especially in the Introduction and Methods Section). Further explanations need to be given in these specific points to improve the reader's comprehension for a broad scientific audience.

Therefore, I do not recommend publishing this manuscript before major revisions.

My main critics are the following:

1. In Section 1 (Introduction), the importance of the topic is well developed, although some key elements need some clarification. For example, the authors talk about thresholds being fundamental to assessing compound flooding; however, they do not explain what a threshold is or why they are important from a technical point of view.

2. In section 2 (Methods), much information is given; however, some details and the connection between the subsections must be improved. For instance, why the authors used the recorded data (subsections 2.2 and 2.3) is not well explained. What is the objective of collecting that information? Is it for validation of the hydrodynamic simulations? A clear statement (short paragraph) should be written right at the beginning of the subsection, as it was done for subsections 2.4 and 2.5. Subsection 2.6 needs the same explanation at the beginning of the paragraph.

3. The authors use specific terms not defined anywhere in the text—for example, skew-surge, storm surge, total water level, flood event. Although familiar to the coastal flooding research public, they must be defined the first time they are mentioned in the

manuscript.

4. In sections 2.1 and 2.2, several flooding events are described; however, a general description of storm surge climate and tidal regime in Conwy Estuary is missing.

5. It is hard to follow section 2.2. and the reasoning behind the selection of flooding events. The authors mentioned that only six flood events were recorded in TWL and river discharge; however, further in the text, they mentioned the top 50 TWL and Qmax events. The authors need to make this explanation more clear. In addition, the authors also need to explain better what was considered a flood event. This general information should be shown at the beginning of the section, followed by a detailed description of single events already presented in the text.

6. Section 2.4. Is the model solving the equations of movement in 2D or 3D? Please clarify it. In addition, if it is solved in 2D, please justify why it is not solved in 3D and vice-versa.

7. In subsection 2.5, a general description should be added to the first paragraph of the subsection, explaining the number of scenarios, which are the drivers tested (total water level, storm surge, river discharge, time lag) and the simulation period (72 hours). I know this information is given through the section; however, it is scattered and confuses the reading.

8. Section 2.8 does not clearly explain why joint probabilities were used. Why is it important? This is partially explained in the introduction and should be remembered here. Similarly, a further explanation of why Copulas is applied from a technical point of view should be given. For instance, are Copulas used to generate synthetic samples of extreme sea levels and river discharges, thus making their respective probability distribution more robust to apply joint probability methods? A general explanation (short and from a practical point of view) should be given about how statistical marginal distributions, copulas, and Bayesian methods are linked to each other.

9. In Section 2.8, there are references to several tables from another paper. I do not believe this is a good way to mention it. If applied, this content should be added to supplementary data or referred to the work without the table number.

10. Section 4 (Discussion) lacks a clear first paragraph, generally stating the manuscript's main findings. The first paragraph is too vague and similar to the

conclusion section's first paragraph. I suggest making a general statement on the main findings and then detailing each in the following subsections.

11. In Section 4.1, I do not understand why the documented flooding records are discussed if they are not even in the section Results (they are shown in Methods). I suggest shortening it to conclusions or moving it to the end of subsection 4.2, "Thresholds for flooding", where you could link the importance of historical records of flooding events to validate numerical/statistical techniques and calculate optimal thresholds (for instance).

12. Section 4 shows several interesting discussion points; however, the authors did not discuss much about the novelty of the method. How innovative is the approach applied here? For instance, Section 4.2 shows a good discussion about setting thresholds from an end-user point of view (flooding mitigation planning); however, the authors do not discuss the hydrodynamic-statistical approach itself.

13. To my understanding, the approach presented here relies on having sufficient data (recorded flooding, water level and river discharge observations, good quality topo-bathymetry data), which is not the reality of several regions globally (perhaps also in the UK). I see the authors discussed that in section 4.2; however, the discussion is focused on the UK. I suggest adding a short discussion about the quality of the forecast and data availability on regional (e.g., Europe) and global scales.

14. In section 5 (Conclusions), the authors wrote a good introductory first paragraph; however, just after, they talk about historical floods, which is not the paper's main goal (I suggest deleting this paragraph or shortening and moving it to the end of the section). I would have expected that they answered the main objective straight after the introductory paragraph (to identify the coastal and fluvial conditions that lead to flooding in an estuarine system).

SPECIFIC COMMENTS

15. Lines 6 – 11. Please replace "UK" with "United Kingdom".

16. Line 14. Please replace "UK" with "United Kingdom".

17. Line 19. Please replace "N-Wales" with "North Wales".

18. Lines 20–22. It is not clear what was amplified. What does sensitivity 7% mean?

19. Line 30. Replace "–" (en-dash) with "—" (em-dash).

20. Line 34. I understand the UK refers to the United Kingdom; however, the authors should define all the acronyms. This acronym has not been previously defined in the text. Suggestion: "United Kingdom (UK)". Then, after that, you could only use the UK.

21. Line 80. "Modelling statistical and probabilistic methods," wouldn't they be the same?

22. Line 89. "N-Wales". Please homogenise the use of "North Wales". If "N-Wales" is used, then it needs to be defined, for instance: "North Wales (N-Wales)….". Then, use only N-Wales after that.

23. Line 110. In "November 1980 - February 2023", a dash is inappropriate. En-dash "– " should be used for ranges of dates and numbers. Please replace throughout the document whenever applicable.

24. Lines 126–133. The terms "total water level (TWLmax)", "predicted tide level", "skew surge", and "storm surge" are not defined. For instance, does the "total water level" include the sea-level anomaly/trend? If so, the sea level trend could interfere with your results. This should be well clarified.

25. Line 130. Please define the acronym "NRW" and homogenise the use of Natural Resource Wales. Use "Natural Resource Wales" or NRW throughout the document (e.g., line 138).

26. Figure 1. The labels of the x-axis in panels (e) and (f) do not follow the same pattern. Please homogenise them.

27. Line 150. What is "event hydrographs"? Please define it.

28. Line 177. Please change "however," to "; however, …" or ". However, …"

29. Line 188. I am not used to the term "Web scraping approaches". Is that a proper term to be used in a scientific paper? Maybe change it to "Web searching"?

30. Table 1. It seems to be cut at the bottom of the preprint. Please check if this is indeed the case. Bellow a screenshot:

Label	Code
0	None
1	River discharge
2	Storm surge
3	High tide
4	Storminess

31. Line 201. In the sentence "… with yellow dots indicating there is evidence of flooding and blue dots indicating there is no evidence of flooding." How can it not have evidence of flooding and be on the internet? I couldn't understand it.

32. Lines 209–210. "…. leave uncertainty in where to set driver thresholds and patterns for flooding, especially for less extreme Q_{max} and TWL_{max} that led to compound flooding.". The concept of threshold (quantiles, peaks-over-threshold, block maxima) and event definition (How long an event was considered to last? Was used any declustering schemes?) needs to be clearly described in the introduction or previously in the methods section.

33. Line 213. I do not fully understand what the authors considered a top 50 Q_{max} and TWL_{max}. The authors mentioned that only a few recorded flooding were identified (6 of NWR and 20+ in web search). However, in Figures 2 and 3, the authors show more events than that. Top 50 events mean that you have selected the top 50 events, or are you taking the events above the 50% percentile? Please make it more clear.

34. Line 224–226. The main sources for DTM, bathymetry, and flood defence locations should be mentioned in the main text.

35. Line 245–248. How exactly did the authors gradually adjust the channel bed elevations? Manually editing the bathymetry? The Neal et al. (2022) work should be better described. One or two short sentences should be enough. Also, what is a "stepwise manner"?

36. Line 249. Why did the authors use two scores (RMSE and Kling-Gupta Efficiency)? Is there any advantage to using that? Formulas should be added to the supplementary material.

37. Line 253. "in the upper estuary". Please add the names of the stations in parentheses.

38. Line 254: "tributaries". Please explain.

39. Line 270. "The M2 tidal constituent has an amplitude of 2.71 m and was used to produce a constant sinusoidal curve for 72 hours". Why only the M2? Are the shallow-water harmonics not important in this estuary? Why 72 hours?

40. Line 272. "scale factor of 25 cm…. thus creating 13 water level time series". Sorry, I could not understand the reasoning. Between 1.82 m (high neap tide) and 3.6m (high spring tide) and a scale factor of 25 cm (adding 25 cm to 25 cm), I could only count seven water level time series and not 13. Please clarify it.

41. Line 276. What is a representative surge shape?

42. Line 297. What does "spin-up" time mean?

43. Line 306. Where exactly does the 40 discharge time series come from? Please clarify.

44. Figure 5. Is the y-axis label of the panel (c) correct ("Number of events")? Shouldn't It be "Total water level OD (m)"?

45. Line 343. Please replace "(ROI, see Figure 1a)," for "(ROI), see Figure 1a, .."

46. Lines 348–358. Please consider joining this paragraph with the previous one.

47. Line 357. Please clarify why 520-simulation parameter space Scenarios 1–3. I understood there were 1560 simulations.

48. Lines 360–366. Please refer to what section will show and discuss these results. It is confusing to the reader to know if you are talking about the simulation scenarios

previously described or the spatial analysis of the flooding area. In addition, what does lateral flood extent mean? Please clarify and define it.

49. Line 368. Why do the authors want to use joint probabilities? At which part of the method the authors are applying this? Is there anything to do with hydrodynamic modelling? Please make it more clear.

50. Line 371. "to the data", which data?

51. Line 378. "Table 6 of Moradian et al. (2023)". Please see comment 9. Apply the comment throughout the text.

52. Line 380. Why so many metrics? Please justify.

53. Line 403. "dependence measures". I suggest changing this term to "dependence metrics" or "correlation coefficients" and then removing "Correlation Coefficient" after the name of each coefficient you mentioned just after. Also, the authors need to explain further why dependence metrics are important. What is it used for?

54. Line 407. The authors need to explain better why they want to use Bayesian methods and how they did it.

55. Line 415. "is the probability of A being true and". The subject of the sentence is missing.

56. Lines 423–432. Please join these two paragraphs. The second one gets confusing when not directly linked to Figure 6.

57. Lines 446–448. Couldn't it be a question of scale in the graphics? Is the figure showing a normalised plot? If not, you compare different units (m) and (m3/s). I would plot a normalised plot to double-check this question. Please correct it if applicable.

58. Figures 6, 7, and 8. What is OD in the y-axis label? Is it the vertical datum? It should be explained in the figure caption.

59. Line 491. Sometimes, the authors use FloodArea in italics, and sometimes they do not use it. Please homogenise it throughout the manuscript.

60. Line 519. Please replace "9d" with "Figure 9d".

61. Line 524. Please explain the terms Q1Twl13, Q40 TWL1, Q20TWL7 and Q40TWL13. I could not understand why the authors used them.

62. Line 526. What is "TWLmax: Qmax parameter space"? Is that figure 6? Please refer to the figure or provide further explanation.

63. Line 530. What does "lateral extent of flooding" mean?

64. Figure 10. Same as Figure 9. A better explanation of the numbers following Q and TWL is needed. Also, the use of the icon is not clear.

65. Lines 569–575. Please see comment n° 10.

66. Line 578. "piecemeal fashion". Is that a scientific term? Please replace it.

67. Lines 609–611. You can also mention that earth observation records can supplement estuarine topo-bathymetry and geometry data for multiple purposes, including hydrodynamic modelling. Reference suggestions:

Valentin Heimhuber, Kilian Vos, Wanru Fu, William Glamore, InletTracker: An open-source Python toolkit for historic and near real-time monitoring of coastal inlets from Landsat and Sentinel-2, Geomorphology, Volume 389, 2021. https://doi.org/10.1016/j.geomorph.2021.107830.

And

Costa, W. L. L., Bryan, K. R., and Coco, G.: Modelling extreme water levels using intertidal topography and bathymetry derived from multispectral satellite images, Nat. Hazards Earth Syst. Sci., 23, 3125–3146, https://doi.org/10.5194/nhess-23-3125-2023, 2023.

68. Section 4.2 is confusing. It seems the authors are introducing new results instead of discussing the current results. I could understand the relevance of the discussion; however, I suggest the authors re-write parts of the section to clarify that new results are not being shown.

69. Line 616. What is "web scraped tag(s)"? Was it explained anywhere in the manuscript?

70. Line 619. I do not follow the statement, "The coastal events (Figure 12c) occur across a range of river discharge combinations, and thresholds 620 may not need to consider this driver". Figures 12 a and b show that flooding events (time lag and river) occur in a similar range of river discharge to coastal events. Please make it more clear.

71. Figure 12. Is Sea Level ODN the same as Total Water Level OD? Please explain why the axis labels are different. The panel indication (a), (b), and (c) are not shown.

72. Line 653. Please clarify which ranges are considered extreme in parenthesis.

73. Lines 654–656. "The volume of riverine freshwater is the dominant driver contributing to high water levels in the estuary. This could be evidence of the backwater effect, where high river discharge can push back low levels of tidal water, resulting in a temporary increase in water levels within the estuary". Please provide some references that corroborate it.

74. Lines 658–659. Please re-order the sentence "It is when the river discharge is between 450-550 m3/s in the Conwy Estuary that flood forecasts need to be particularly accurate. " to "Results shown that flood forecasts need to be particularly accurate for Conwy Estuary when the river discharge is between 450-550 m3/s". In addition, please say in parentheses if this range of values is mild or extreme.

75. Line 679. In the sentence: "The parameter space could be developed by considering additional hydrograph time lags and exploring the timing of the surge relative to tidal high water, which could influence the magnitude and volume of the total water level (Lyddon et al., 2018; Khanam et al., 2021)." I suggest two references:

Costa W, Bryan KR, Stephens SA and Coco G (2023) A regional analysis of tide-surge interactions during extreme water levels in complex coastal systems of Aotearoa New Zealand. Front. Mar. Sci. 10:1170756. doi: 10.3389/fmars.2023.1170756

And

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., et al. (2020). Non-linear interaction modulates global extreme sea levels, coastal flood exposure,

and impacts. Nat. Commun. 11, 1–9. doi: 10.1038/s41467-020-15752-5

76. Line 688. "Sea-level rise and geomorphic changes will lead to a new baseline for flooding and new driver-thresholds and interactions ". Reference suggestion:

"Khojasteh, D., Glamore, W., Heimhuber, V., and Felder, S. (2021). Sea level rise impacts Estuar. dynamics: A review. Sci. Total Environ. 780, 146470. doi: 10.1016/j.scitotenv.2021.146470"

77. Lines 713–716. "The research highlighted the incomplete nature of recorded flooding extents held by national agencies, which are important to build a database of past episodes of flooding (e.g., when and where has flooded, and under what conditions) and undertake further analyses such as temporal trends in flooding. Such a database is crucial for developing accurate and timely flood warnings. ". This passage is a bit unclear; maybe change it to

"The research highlights that the recorded flooding extents held by national agencies are incomplete. This database is important to build knowledge on past flooding episodes (e.g., when and where has flooded, and under what conditions), undertake further analyses such as temporal trends in flooding, and develop accurate and timely flood warnings."

78. Section 5. It is confusing that historic flooding records are included in the Conclusion section but not in Results. Instead, they are described in Methods. I suggest removing the historic events from the conclusion or moving the historic flooding records from methods to results.

79. Section 5. See comment n° 14. Suggestion: the third and fourth paragraphs should be joined together and placed as the second paragraph. The paragraph in lines 713–720 should be the third or last.