EGUSPHERE-2025-270
Response to referee 1

We thank the referee for the useful and productive comments. Here, we reply to the raised issues. Referee comments are reported in blue italic font.

Referee 1

Main comments

1. *The manuscript does not provide any information on the validation of the numerical models used in the framework, specifically the wave model, storm surge model, river discharge model, or the hydrodynamic model. While the paper presents a generalized workflow, the quantitative conclusions (e.g., flood volumes, durations, extents) rely heavily on the reliability of the model outputs. Without presenting model validation or calibration results, it is difficult to assess the robustness of the findings. This is particularly critical for extreme hydrologic/hydrodynamic modeling, where prior studies have shown that numerical models often fail to estimate peak magnitudes accurately unless properly calibrated. I strongly recommend including a validation section for each model, either in the main text or supplementary materials, so that readers can better understand the limitations and reliability of the framework and derived results.*

We thank the reviewer for this important and constructive comment, which helped us improve the transparency and completeness of our modeling approach. In response, we have added a dedicated section to the Supplementary Material detailing the calibration and validation procedures adopted for the water level, wave climate, and hydrological models. While we could not find suitable observational data to directly validate the hydrodynamic model (HEC-RAS) in reproducing flooding events in the study areas, we acknowledge this limitation clearly in the revised text.
Nevertheless, the use of HEC-RAS in flood modeling is widely established in the literature, including applications to coastal and compound flooding scenarios. Its reliability has been demonstrated in several studies, including urban-scale and coupled modeling frameworks (e.g., Pandey et al., 2021; Bennett et al., 203), which supports its adoption in our case.
To enhance clarity, we have uploaded the cal_val.pdf document containing the newly added section to the Supplementary Material.
We also revised the main text (lines 261–263) as follows:
"*In this section, we describe the implementation of the three numerical models: WaveWatch III, SHYFEM and LISFLOOD, employed to perform the main part of the downscaling procedure. Each of the models has a particular setup on the basis of the analyzed coastal city. Furthermore, a calibration/validation procedure has been carried out for each of them to have an estimate of their skill to reproduce observed events. The detailed description of the different procedures is reported in the Supplementary Material.*"
For the sake of clarity, we use the expression "calibration/validation" in a broad sense, not strictly referring to the use of one dataset to tune the model parameters (calibration), and a separate

dataset to assess model skill (validation). Due to limitations in data availability, we show that the models are nonetheless able to reproduce observed quantities under different conditions: through a systematic trial-and-error calibration (for the water level model SHYFEM), by adopting a previously validated setup (in the case of the wave climate model WWIII), or by relying on community-driven configurations (as for the hydrological model LISFLOOD). This clarification is also included in the cal_val.pdf document provided in the Supplementary Material.
Additionally, we plan to revise the Results section (line 497) to include one key outcome of the calibration procedure:
"*As a result of the calibration and validation process, we observed a tendency of the evaluation run to underestimate the most extreme values measured by wave buoys (see Supplementary Material, Figure S12).*"

2. *To provide boundary conditions for the hydrodynamic model, the time series of Q and total water levels are obtained by averaging superimposed 49-hour time series around the annual maxima peaks. The time series surrounding the annual maxima may exhibit multiple peaks and varied shapes across a wide range, particularly for Q, depending on the upstream catchment and rainfall characteristics. Therefore, averaging at each timestep and treating it as the representative hydrograph may overlook temporal dependencies and could be unrealistic. A similar issue may arise for storm surge and/or total water levels (see Quinn et al., 2014; Santamaria-Aguilar et al., 2017). To support the approach used in this study, a figure showing all individual and averaged hydrographs (possibly in the supplementary material) or appropriate citations of relevant studies that have previously applied this method would strengthen the justification.*

We thank the reviewer for the valuable comment, which helped us improve the clarity and communication of the results. In response, we plan to add in the Supplementary material three additional figures, one for each city, depicting the time series of discharge Q and total water level for the Historical, RCP45 and RCP85 runs. Each subplot contains the superimposition of the time series around the yearly maximum (grey lines), the average (black line) and a red shaded area representing +/- one standard deviation.

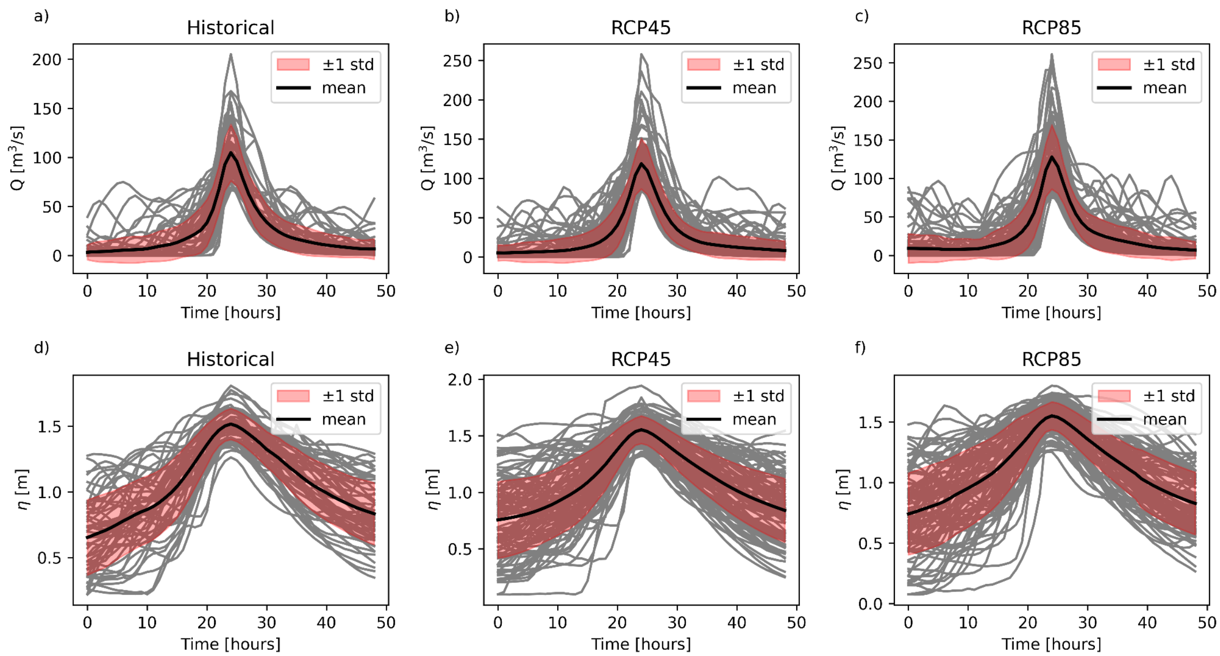

Fig S5 Shape of the synthetic hydrograph for the city of Massa. Centered yearly maximum time series are reported in grey, average in black, and +/- one standard deviation is reported as shaded area. a), b), c) refer to discharge for Historical, RCP45 and RCP85 runs, respectively. d), e), f) refer to total water level for Historical, RCP45 and RCP85 runs, respectively.

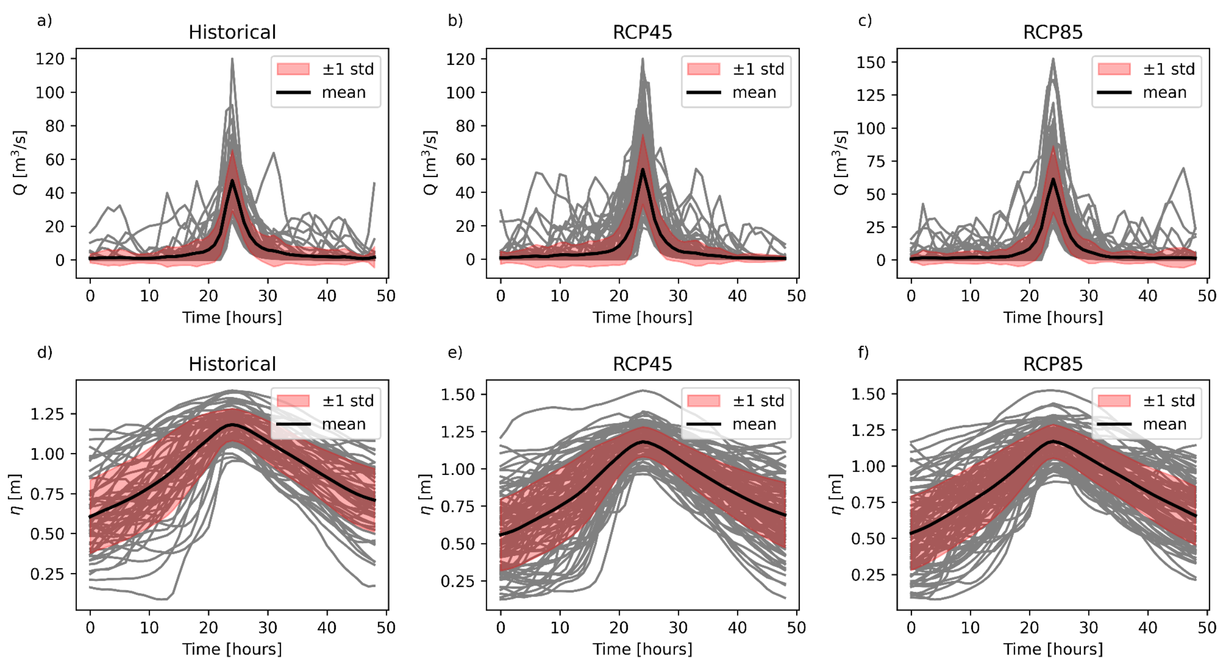

Fig S6 Shape of the synthetic hydrograph for the city of Villanova. Centered yearly maximum time series are reported in grey, average in black, and +/- one standard deviation is reported as shaded area. a), b), c) refer to discharge for Historical, RCP45 and RCP85 runs, respectively. d), e), f) refer to total water level for Historical, RCP45 and RCP85 runs, respectively.

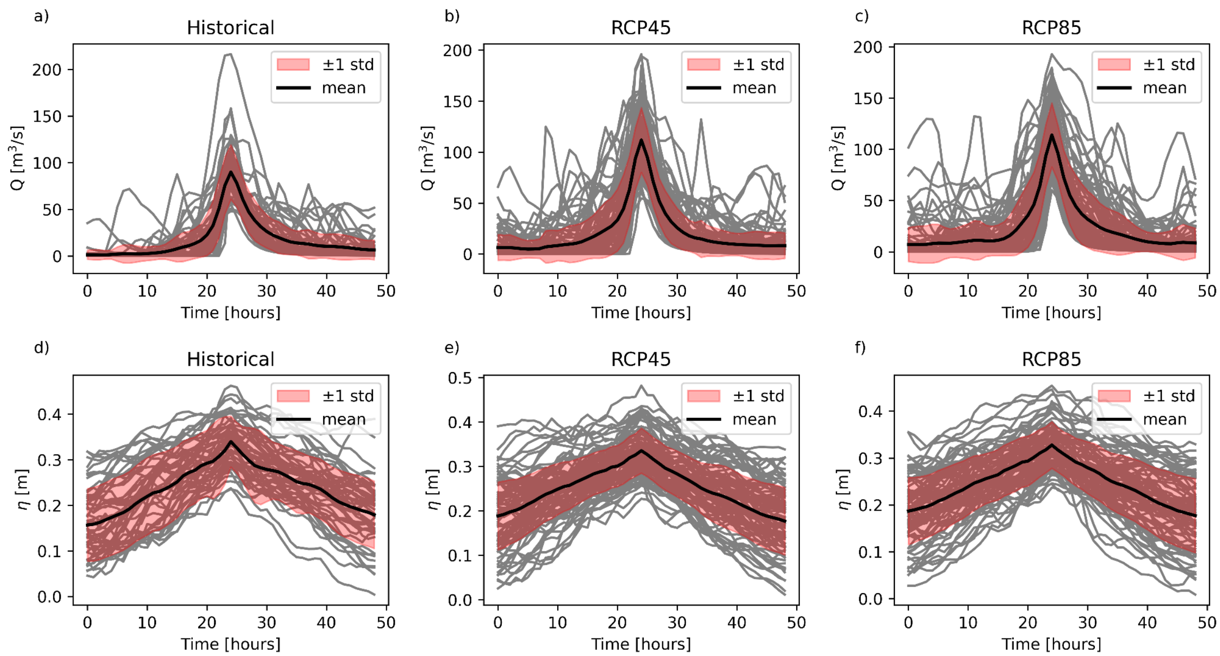

Fig S7 Shape of the synthetic hydrograph for the city of Oarsoaldea. Centered yearly maximum time series are reported in grey, average in black, and +/- one standard deviation is reported as shaded area. a), b), c) refer to discharge for Historical, RCP45 and RCP85 runs, respectively. d), e), f) refer to water level for Historical, RCP45 and RCP85 runs, respectively.

The reported figures show different variability ranges (red shaded area), depending on the city and the variable analyzed. Nonetheless, the spread among realizations remains moderate and does not compromise the representativeness of the synthetic hydrograph. Quinn et al. (2014) investigated the temporal variability of storm tides by aligning observed peaks and explicitly analyzing tide-surge interactions. In our study, we opted for a simpler and more precautionary approach: we phase-aligned only the residual components, assuming that the tidal level coincides with its peak during the extreme event.

Figures S5, S6, and S7 demonstrate that the variability of residuals is generally stable throughout the event duration. While this represents a worst-case scenario, it still allows for consistent comparisons across climate scenarios—particularly since climate change primarily affects residual components and mean sea level, rather than tidal dynamics.

Although our method does not capture the probabilistic variability arising from tide–surge interactions, it provides a robust and comparable framework across cities and scenarios. Moreover, direct coupling of projected tidal time series with predicted residuals would likely introduce further uncertainty, given the long timescales and limitations in forecasting tidal constituent changes. As Quinn et al. (2014) noted, this added variability would make it more difficult to clearly define floodable areas.

Therefore, while our method may appear simplified, it is appropriate and justified for the purpose and temporal scale of this study.

We acknowledge that trends in tidal constituents, as highlighted by Santamaria-Aguilar et al. (2017), were not explicitly included in this study. These long-term (century-scale) changes in tidal amplitudes are driven by meteorological, oceanographic, and hydrographic variability, and are known to be highly site-specific. Since such variations have been detected from past observations rather than derived from predictive models, and no consolidated numerical methods are currently available to project them, we consider it reasonable not to include them directly. Nevertheless, we mention this limitation in Section 5 and in the Discussion as a relevant aspect deserving further investigation.
We are aware that averaging extreme events around their peaks smooths out their temporal variability. Nevertheless, our intention was to derive a representative event shape that captures the typical rise and fall of the signal. The limitations associated with this approach, such as the possible omission of temporal dependencies and multi-peak structures, are unavoidable when only the maximum value (linked to a return period) is available for constructing synthetic hydrographs. To overcome this, an impact-based approach can be adopted. In such a framework, the statistical analysis is conducted *a posteriori*, based on simulated water depths resulting from multiple realizations of extreme events, using full time series of Q and total water level as boundary conditions. This methodology, already mentioned at lines 687–703, allows a more explicit consideration of variability in time.
To our knowledge, the derivation of a synthetic hydrograph starting from the maximum discharge was proposed by Brunner et al. (2017). Here we do not refer to methods where the input is rainfall with a given return period and duration, subsequently transformed into discharge using a rainfall-runoff model. In their work, Brunner et al. present a detailed procedure to construct synthetic design hydrographs (SDHs) based on “the fitting of probability density functions to observed flood hydrographs of a certain flood type taking into account the dependence between the design variables peak discharge and flood volume”. Their method also involves a normalization step, similar to what we adopted. However, we intentionally opted for a simpler procedure that can also be extended to total water levels, a case for which -to our knowledge- no comparable methodology is available.
We expanded the discussion accordingly, addressing this point and clarifying our methodological choices with reference to the available literature.

We plan to add the following from lines 444-445 onwards:
“*Such a procedure is applied to every run to get the appropriate BC. The figures showing the superimposition of the annual maxima events for river discharge and water level for the three cities of Massa, Villanova and Oarsoladea, are reported in the Supplementary Material.*”

In addition, we plan to add the following at line 457:
“*This approach does not take into account long-term trends potentially present in the tidal constituents as observed by Santamaria-Aguilar et al. (2017).*”

We finally plan to expand the discussion after line 694.
“*Additionally, another issue that can be overcome in case the impact-based approach is employed, is that related to the creation of adequate synthetic boundary conditions associated with specific return periods. The choice to average the extreme events superimposed in phase at*

the peak can smooth out their variability, given that they show different variability ranges based on the selected city and variable (Quinn et al., 2014). Nevertheless, the obtained variability ranges do not exceed the magnitude of the associated extreme event (see Figure S5, S6, S7 in the Supplementary Material). It was our intent to derive a shape for the time series that is representative of the main behaviour of the analyzed variable during its rise and fall around the maximum. The derivation of a synthetic hydrograph starting from the maximum discharge is proposed by Brunner et al., (2017), where they retrieve a synthetic design hydrograph based on "the fitting of probability density functions to observed flood hydrographs of a certain flood type taking into account the dependence between the design variables peak discharge and flood volume". They also pass through a normalization step, similar to what we carried out. However, we tried to keep a simpler approach that can be also extended to the total water levels, for which we did not find an analogous procedure.

Accounting for the tide by adding a fixed $\Delta\eta_{\mathrm{Tide}}$ to the extreme event hydrograph (Massa and Villanova), or by simulating a semidiurnal tide (Oarsoaldea) as a boundary condition can overlook the long-term (century-scale) modifications in tidal ranges. Santamaria et al. (2017), using site specific past observations, found they are driven by meteorological, oceanographic, and hydrographic variability. The difficulty to forecast them using numerical tools partly justifies the decision not to explicitly include this aspect in the present study."

Added references

Brunner, M. I., Viviroli, D., Sikorska, A. E., Vannier, O., Favre, A. C., & Seibert, J. (2017). Flood type specific construction of synthetic design hydrographs. Water Resources Research, 53(2), 1390-1406.

Quinn, N., Lewis, M., Wadey, M. P., & Haigh, I. D. (2014). Assessing the temporal variability in extreme storm-tide time series for coastal flood risk assessment. Journal of Geophysical Research: Oceans, 119(8), 4983-4998.

Santamaria-Aguilar, S., Schuerch, M., Vafeidis, A. T., & Carretero, S. C. (2017). Long-term trends and variability of water levels and tides in Buenos Aires and Mar del Plata, Argentina. Frontiers in Marine Science, 4, 380.

3. The purpose of comparing the historical and evaluation simulations is unclear. These runs differ in only 7 years of annual maxima, with the remaining 26 years overlapping. It's mentioned that this analysis is done to "test the ability of the model to reproduce observable extreme events." I would like to see some discussion on how this is achieved in the author's reply in the open discussion. Additionally, while the authors suggest that this comparison of quantile plots helps to estimate the degree of over- or under-estimation in projections (e.g., L 624-L625), the interpretation of these results and their contribution to the main analysis are not well explained.

We start from the assumption that the Evaluation run, being a reanalysis, closely reflects observed reality as it benefits from data assimilation. In contrast, the Historical run is a free simulation constrained only by the radiative forcing.

To assess the model's ability to reproduce observable extreme events, we compare the statistical distribution of annual maxima from the two runs using quantile–quantile (QQ) plots. If the values align along the 1:1 line, this suggests that the Historical run can statistically reproduce the distribution of observed extremes. Deviations from the 1:1 line indicate systematic under- or over-estimation. Since the RCP45 and RCP85 scenarios are based on the same model, we extend this diagnostic to those projections, assuming that similar biases may occur.

We plan to add the following after lines 624-625:
"*Indeed, we make the hypothesis that the Evaluation run, being a reanalysis, is close to reality given that It benefits from a data assimilation procedure, thus incorporating the information from observations.*"

Other comments

1. Figures 6 & 7: The term "envelope of the water depth" is unclear. Do the flood maps represent the maximum depth reached at each grid cell throughout the simulation period, or are they snapshots at a specific moment in time? Please clarify this in the text or figure captions.
The flood maps represent the maximum water depth reached at each cell during a simulated event. The following sentence (line 531):
"*The envelope of the water depth simulated through the hydrodynamic model for coastal and riverine floods with a 100 yr RP are reported in Figure 6 and Figure 7, respectively.*"
will be modified to:
"*The envelope of the water depth, that is the spatial distribution of the maximum water depth reached at each computational cell during the hydrodynamic simulation, are reported in Figure 6 and Figure 7 for coastal and riverine floods with a 100 yr RP, respectively.*"

2. The authors claim that the uncertainty, which was not accounted for due to the absence of a multimodel ensemble, was recovered in the statistical analysis by calculating confidence intervals through the bootstrap method. However, these represent fundamentally different types of uncertainty and should not be considered as "recovered" from one another.
We changed the text (lines 599-601):

"*However, such an estimate, associated with the extreme values from the analyzed time series, was recovered in the statistical analysis by calculating confidence intervals through the bootstrap method.*"
As follows:
"*We tried to partially compensate for the lack of such an uncertainty estimation, by calculating confidence intervals through the bootstrap method in the statistical analysis, although this is a different source of uncertainty.*"

3. Line 87: Typo: the word "and" is misspelled.
Corrected.

4. Line 343: The sentence implies that the shape, timing, and magnitude of streamflow are determined solely by upstream hydrological processes. However, downstream conditions such as tides and storm surge can also play a significant role. Please revise the sentence to reflect this more accurately.

The sentence at lines 342-343:

"*In order to model river floods, it is necessary to define the discharge hydrographs, i.e. the evolution in time of flow rate in given cross sections.*"

Is modified as follows:

"*In order to model river floods, it is necessary to define the inflow discharge hydrograph as a boundary condition, i.e. the evolution in time of flow rate in the upstream cross section.*"

We also added the following proposition after line 348:

"*Moreover, the downstream boundary condition defined by the water level at the outlet affects the evolution of the hydrograph while traveling along the river (see Section 5.2).*"

5. Lines 614–616: The phrase "associated with the highest waves" is unclear whether it refers to wave heights or wavelengths. Please specify.

We revised the sentence (lines 614-616):

"*Furthermore, the sensitivity of runup to wave height for Massa's beach slope, combined with the wavelengths associated with the highest waves (70-95 m) is modest, approximately 0.2-0.25 m.*"

As follows:

"*Furthermore, the sensitivity of runup to wave height for Massa's beach slope, calculated using wavelengths ranging between 70 and 95 m (those associated with the highest waves) is modest, approximately 0.2-0.25 m.*"

6. Lines 654–657: It is unclear why the analysis in this section was conducted only for one location instead of all three sites. The phrase "number of events per year higher than specific values" would benefit from clarification; please indicate what those specific threshold values are.

With "number of events per year higher than specific values" we intend those values reported in the y axis of the bar coloured subplots.

We modified the proposition (lines 656-657):

"*Furthermore, we determined the number of events per year higher than specific values, clustering the events by decade (Figure 8b and 9b).*"

as follows:

"*Furthermore, we determined the number of events per year (coloured patches) higher than specific values of η and Q (reported in the abscissa), clustering the events by decade (Figure 8b, c, d and 9b, c, d, for Historical, RCP4.5 and RCP8.5 run, respectively).*"

We correct the Caption of Figure 9:

"*a) Cumulative duration of water discharge above the 99.9 %-ile in days per year for HIST (green line), RCP4.5 (red line), RCP8.5 (black line), RCP4.5 and RCP8.5 plus the effect of RSLR (red dashed line and black dashed line, respectively), for the city of Massa. Number of events per year with peak values larger than specific values, grouped by decades for: HIST b), RCP4.5 c) and RCP8.5 d)*"

Removing the part related to RSLR (typing mistake).

In addition, we will include in the Supplementary material the analysis performed for the city of Villanova and Oarsoaldea. We mention it in the main text by adding the following sentence after line 654:
"*both* $\eta_{TOT}(t)$ *and* $Q(t)$*. (The same analysis for the city of Villanova and Oarsoaldea is reported in the Supplementary Material, Figures from S14 to S17).*"

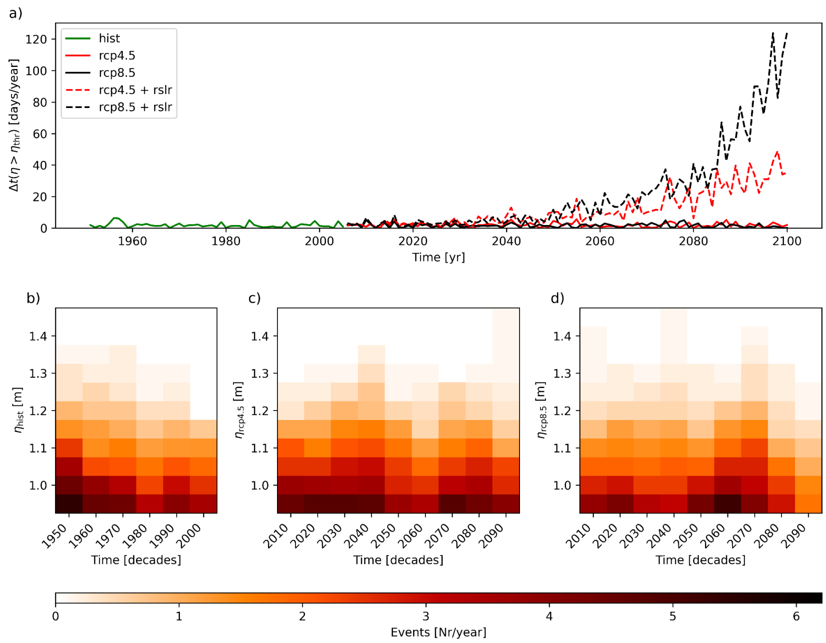

Fig S14 a) Cumulative duration of total water level above the 99.5 %-ile in days per year for HIST (green line), RCP4.5 (red line), RCP8.5 (black line), RCP4.5 and RCP8.5 plus the effect of RSLR (red dashed line and black dashed line, respectively), for the city of Villanova. Number of events per year with peak values larger than specific values, grouped by decades for: HIST b), RCP4.5 c) and RCP8.5 d)

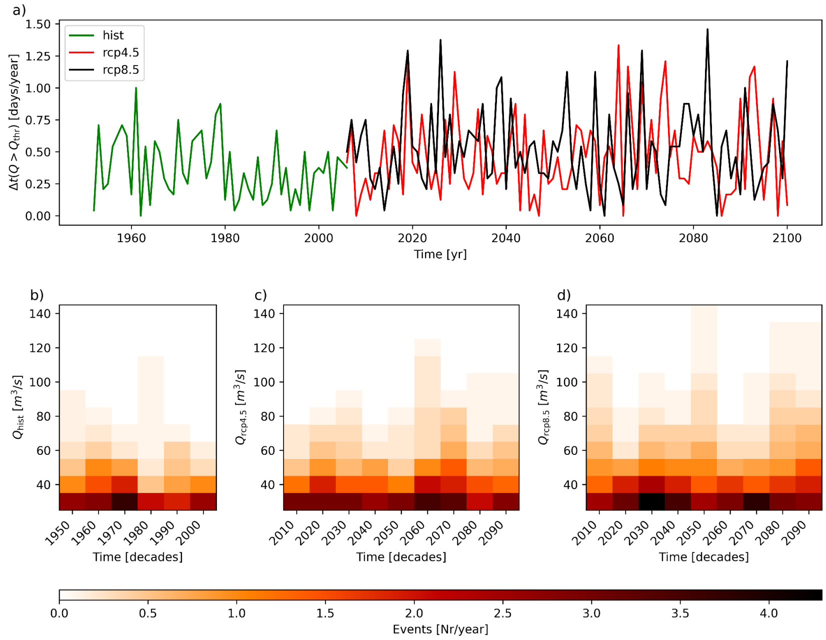

Fig S15 a) Cumulative duration of water discharge above the 99.9 %-ile in days per year for HIST (green line), RCP4.5 (red line), RCP8.5 (black line), for the city of Villanova. Number of events per year with peak values larger than specific values, grouped by decades for: HIST b), RCP4.5 c) and RCP8.5 d).

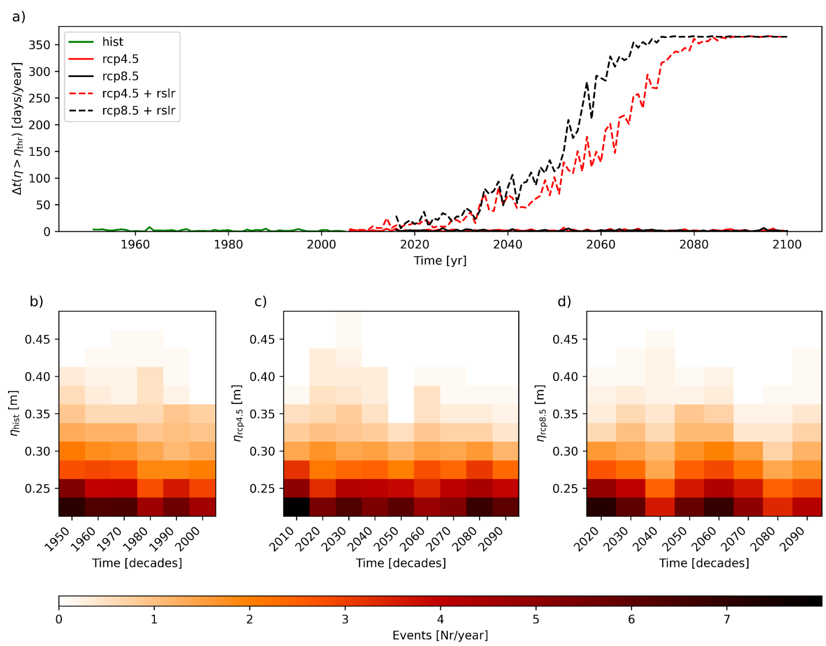

Fig S16 a) Cumulative duration of total water level above the 99.5 %-ile in days per year for HIST (green line), RCP4.5 (red line), RCP8.5 (black line), RCP4.5 and RCP8.5 plus the effect of RSLR (red dashed line and black dashed line, respectively), for the city of Oarsoladea. Number of events per year with peak values larger than specific values, grouped by decades for: HIST b), RCP4.5 c) and RCP8.5 d)

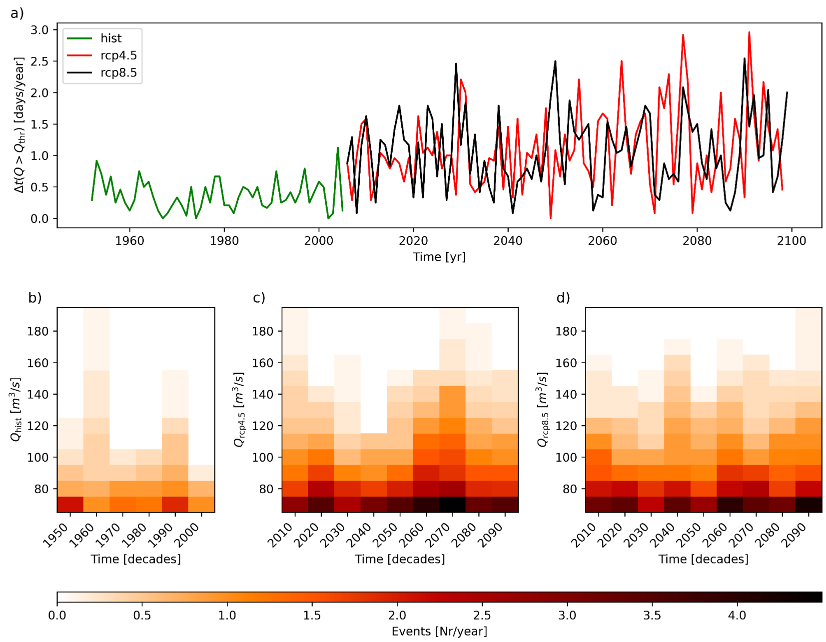

Fig S17 a) Cumulative duration of water discharge above the 99.9 %-ile in days per year for HIST (green line), RCP4.5 (red line), RCP8.5 (black line), for the city of Oarsoaldea. Number of events per year with peak values larger than specific values, grouped by decades for: HIST b), RCP4.5 c) and RCP8.5 d).