We would like to thank the reviewer for taking the time to review our manuscript and for the valuable comments and suggestions to improve our manuscript. We respond to the comments by referring to the page and line numbers in the original manuscript.

Although this study is interesting and relevant, I agree with the first reviewer that the manuscript should be proof-red by a native speaker and many commas are missing (I listed some in the minor comments below). In addition, how some sentences are structured and some terminology used makes the reading difficult. Specifically, I recommend to state first what it was done, and then give the reasons why it was done in that way, instead of the opposite order. The paper is long and it addresses several different scenarios and analyses (climate changes scenarios, changes in the tidal phase...), so the use of generic terms in some cases (e.g. L218 “data set”) makes difficult the reading because it is not clear to which simulation/data set the authors are referring to. In addition, I would use “event” instead “ensemble member” e.g. L239, L247.

We went again though the text clarifying when possible and addressing the specific comments of the reviewers. The “data set” definition was explained in the Line 164 or it was explicitly stated in the text when some other data were meant. We think it is important to keep the “ensemble member” term, although we agree that the usage of this term is unconventional in this context. We want to underline the deference between an “event” – a particular atmospheric situation and corresponding surge, which is unique for a given atmospheric situation and an “ensemble member” – a particular constellation of tidal component and surge component and there are many of them for a given atmospheric situation.

In the introduction and discussion, I am also missing more references of other studies using a similar approach than in this paper as well as other studies made for the German Bight region. For example, Arns et al. (2015) analyzed the non-linear effects of different SLR scenarios on the peaks of storm tides at the German Bight and Santamaria-Aguilar et al. (2017) assessed the effects of these scenarios on the storm surge hydrographs. In addition, Wahl et al. (2011) developed a statistical approach for generating a large number of storm surge events.

Thank you for pointing out the necessity of additional references also including other methods of analysis. There was a considerable amount of studies during the past decades investigating storm tides in the German Bight, so we wanted to limit references only to the relevant methodology. Now the introduction is partly reformulated and more references are added.

“This information is usually assessed and provided in form of high percentiles or return values obtained from frequency distribution estimates. There is a spectrum of methods used to construct such estimates (e.g.Debernard and Røed(2008), Arns et al. (2015b), Santamaria-Aguilar et al. (2017) for dynamical modeling approach, Wahl et al. (2011) for stochastic modeling approach or Dangendorf et al. (2013) for processing of tide gauge observations). In the present study we are interested in the40 spatial and temporal evolution of particular very severe storm tide events in coastal areas and estuaries and, thus, diverge from statistical approach. So far, more detailed information and assessment of particular events that are extremely severe and rare are uncommon. Potential sources of such events comprise historical data as well as modelled data for past, present and future.”

I find the section of data, methods and experiments difficult to read and follow due to the large number of datasets, models and simulations made/used. However, the summary and discussion is well structured and clear. I recommend to rephrase first sentence of section 2.2 and to re-structure the section ordering the different data sets e.g. Start with the hydrodynamic model used, hindcast forcing, and climate change scenarios and models. (However, it is very clear in the diagram of Figure 2). In some cases, the reading would be easier if the type/variables is specified e.g. “multi-decadal hindcast” or “climate realizations”, which can refer to atmospheric forcing or water levels. It would also be interesting to know the length of the hindcast and climate scenarios period i.e. specify the years.

Chapter 2 is rearranged and reformulated according to the suggestions. The description of the area under investigation (2.1) is followed by the description of the ''North Sea'' (2.2) and the ''German Bight'' (2.3) models used in our investigation and by the description of the data set (2.4). Finally, the selection of events and amplification experiments is specified (2.5).

Chapters 2.4 and 2.5 are reformulated as follow:

"2.4 Data set

A set of numerical simulations for which atmospheric as well as marine data are available is required for the detection and ranking of extreme storm tides and subsequent modifications. Furthermore, a desired homogeneity and comparability of resulting water level fields suggests that the local water level data should be simulated with the same hydrodynamic model for the North Sea. However, the global and regional atmospheric conditions may and should vary in their origins to ensure diversity of possible storm and storm tide events. Thus, the set of underlying atmospheric conditions comprises a multi-decadal hindcast (Geyer (2014)) for the period 1948-2016 based on downscaled NCEP-NCAR global reanalysis (Kalnay et al. (1996)) and six downscaled climate change realizations. In details, the global climate realizations include four CMIP3 members for the SRES A1B and B1 scenarios (e.g. Nakicenovic and Swart (2000), Houghton et al. (2001)) covering the period 2001-2100 and corresponding present-day conditions for 1960-2000. Other realizations include two CMIP5 members for the AR5 RCP8.5 scenario (e.g. Stocker et al. (2013), Taylor et al. (2010)) for the periods 2006-2100 or 2071-2100 and corresponding present-day conditions for 1971-2005 or 1971-2000. The climate simulations were obtained with different global models (ECHAM5-MPIOM (e.g. Röckner et al. (2003), Marsland et al. (2003)), EC-EARTH as part of EURO-CORDEX (e.g. Hazeleger and Coauthors (2010)) and CMCC (Scoccimarro et al. (2011)) using different initial conditions. The global atmospheric realizations from these simulations as well as the hindcast were downscaled with different regional circulation models (different versions of CCLM (e.g. Rockel et al. (2008), Hollweg et al. (2008))), RCA4 (e.g. Samuelsson et al. (2011)) providing regional atmospheric climate realizations for the Northeast Atlantic. These regional atmospheric data from the hindcast and climate projections were used to force the hydrodynamic model TRIM-NP ("North Sea" model) and to obtain water levels in the North Sea and the Northeast Atlantic (e.g. Gaslikova et al. (2013), Weisse et al. (2014), Weisse et al. (2015)). The resulting set of water level data is used for further analysis in this study and is referred to as "data set" further on. For the entire data and model flow see also Figure 2.

The climate realizations do not include any rise in mean sea level.Water level changes are due to changes in the atmospheric forcing only. Furthermore, possible changes in bathymetry within the course of the time are neglected in the hindcast as well as in the climate realizations.

2.5 Selection of events and amplification experiments

Different classifications of storm tides exist using e.g. water levels above a reference height or the probability of water levels. Here, the classification of the Bundesamt für Seeschifffahrt und Hydrographie (Federal Maritime and Hydrographic Agency ,see Müller-Navarra et al. (2003)) is used: A storm tide is an event with water levels exceeding mean tidal high water at least by 1.5 m, a severe and a very severe storm tide denote events exceeding MHW by 2.5 m and 3.5 m, respectively.

The analysis of extreme storm tides is mainly focused on the East-Frisian coast in particular on Borkum and the Ems estuary. However, the impact of storms in the North Sea varies along the coasts depending on the wind direction and the resulting wind set up. Therefore, from the data set, time series of water levels were extracted for a location seaward of the island of Borkum (in the following labeled as "Borkum") and two other locations in the German Bight (Figure 1): one location in the outer Elbe estuary (labeled as "Elbe Mouth") and one location seaward of the North-Frisian island of Amrum (labeled as "Amrum").

Figure 2 describes the workflow for the simulation of the original water levels included in the data set and for the construction of the amplified water levels. A potential amplification due to tidal variations is tested for selected events at Borkum, whereas Elbe Mouth and Amrum are used to compare the effects at Borkum with those at other coasts of the German Bight. The methodology used to investigate the potential amplification of the selected storm tide events comprises four steps.

In step 1, extreme storm events are selected from the corresponding time series using three criteria:

- height of water levels,
- duration of water levels continuously exceeding NHN+1.15 m (MHW at Borkum, DGJa (2014))
- series of storm tides with high water levels exceeding MHW + 1.5 m within one week

Water levels are considered with respect to NHN. The storm tide events for Borkum are ranked with respect to their water levels and their durations. For the further analysis of a possible amplification, the event with the highest high water was defined as "EH". The event with the maximum duration was defined as "EL". The strongest event chain from the selected events was defined as "EC", where "strongest" describes the combination of the maximum number of storm tides within a week and the maximum intensity. The intensity is given by the area between the water level curve and a threshold.

In step 2, possible amplification of the selected extreme events due to different combinations between wind field and astronomical tide was tested. Maximum water levels may be increased by variations of relative propagation and arrival time of tidal high water and atmospheric storm. They may also become higher if the specific storm occurs around spring tides rather than around neap tides.

Thus, ensembles of large-scale North Sea water level simulations for each selected event were generated. For ensemble one, the astronomical tide given at the open model boundaries was shifted hourly within +/-6 h around the wind speed maximum near Borkum. For ensemble two, the highest astronomical spring tide found in the tidal simulations for the period 1948-2100 was used instead of the original tide and the astronomical tides were shifted again hourly. For each member of ensemble one and two, water level time series were extracted for the three locations. The time series were analysed and members were selected focusing on the strongest amplification for Borkum. Comparing the time series for the three locations, it is estimated how the amplification for Borkum affected the water levels at Elbe Mouth and Amrum.

Respective data from the ensemble members with the highest amplified water levels near Borkum (in the following identified by "_a") for each event were used for further fine-grid simulations of the German Bight and the Ems estuary in steps 3 and 4.

In step 3, high resolution water level simulations for the German Bight and the attached estuaries for the ensemble member with the highest amplified water levels near Borkum for the selected events derived from step 2 were performed.

In step 4, the events from step 3 were further amplified by applying an increased river runoff to examine the impact of runoff variations and a sea level rise to place the results in the context of future climate change. For these amplification simulations the highest observed river runoff for the Ems of 1200 m^3s^{-1} (1946, DGJb (2018)) was assumed. This extreme river runoff was measured in February 1946, i.e. in a season where storm tides are probable. Furthermore, simulations with two sea level rise scenarios of 50 cm and of 100 cm were investigated. These values cover the likely range of median values for the global sea level rise as well as the bandwidth of the local sea level rise for the North Sea until 2100 as reported by Stocker et al. (2013). The sea level rise was applied at the open boundary of the German Bight model by shifting the boundary values for water level by the selected amount of sea level rise.

In order to investigate the impact of the storm surge barrier in the Ems on water levels, the storm tides were simulated with open and with operated barrier in steps 3 and 4."

Minor comments:
Title: I suggest to change "very severe" for extreme, which is the term commonly used in the literature and actually, it is also used in the manuscript e.g. L138 (Here and along the manuscript). In addition, enhancement can also be changed to "amplification", which is the term used more often along the manuscript.

''enhancement'' is changed to ''amplification'' and ''very severe'' to ''extreme''

L1. Change "essential" for major

changed

L25. Environmental threat-> Natural hazard/threat

changed to ''natural hazard''

L27.Inflicted heavy losses-> caused large damages

changed

L29-30. Rephrase. The use of commas is not correct in this sentence.

The sentence is changed

''Mainly due to these measures more recent storms, e.g. 1976 or 2013, caused no severe damages although water levels higher than those of 1962 have been observed at various coastal sections …''

L30. Risk of what?

''of flooding'' is inserted

L31. Remove associated with anthropogenic climate change.

changed

L32. Storm surge-> If storm tides is the term used, please be consistent along the manuscript.

All ''storm surge'' terms are changed into ''storm tide''.

L32-33. Link this sentence with the previous one. In addition, references can be added as e.g. Arns et al. 2015

The text is reformulated and the reference Arns et al. 2015 is added.

"In modern times, two major storm tide disasters that caused large damages at the North Sea coasts occurred in the years 1953 and 1962. Since then coastal defenses have been significantly improved throughout the coastline. Mainly due to these measures more recent storms, e.g. 1976 or 2013, caused no severe damages although water levels higher than those of 1962 have been observed at various coastal sections (NLWKNa(2010), NLWKNb(2007)). Nevertheless, risk of flooding is still present and may increase due to expected climate change. Thus, the rise of the mean sea level may lead not only to an increase in the height of the storm tides and longer duration of water levels exceeding certain thresholds (e.g. Idier et al. (2019) and references therein) but also to shorter arrival times of the storm tide at the coast and in the estuaries (e.g. Arns et al. (2015a)). These effects, among others, may aggravate risks related to storm tides and may have consequences for coastal protection e.g. for the dike heights or the warning times, but also for such issues as the drainage of low-lying coastal areas."

L61. : : :forcing, a possible amplification can occur or possible amplifications
L63-64. Add comma after variations and considered.
L65. Comma after study.

Commas are inserted.

L68. The climate realizations used, comprising CMIP3 and CMIP5 scenarios, reflect only.....and local bathymetric changes or changes in the local bathymetry.
L75. Comma after set. Remove distinct

changed

L76. Simulations of what?

''Water level'' is inserted.

L79. Comma after surges
L82. Comma after estuary.
L84. Comma after Emden
L94. Comma after Bight

Commas are inserted.

L99. The Ems estuary is situated in the southern German Bight, at the border.... (Remove North Sea because the location of the German Bight was already specified).

changed

L136. Rephrase. For instance, "The methodology used to investigate the potential amplification of the storm tide events comprises four steps"

changed

L141. It is not clear here how an event is defined, which is explained in L212-215. These lines should be moved to this section as they are part of the methods and not of the results.

Chapter 2 is rearranged and reformulated according to the suggestions. Events are explained more precisely and these lines are now at the beginning of chapter 2.5 ''Selection of events and amplification experiments'', see above.

L152. If the SLR is not included in the simulations of climate scenarios of the North Sea model, why the largest spring tide of each climate scenario is used and why it would change between them? Are the climate scenarios for different periods? How is the tide extracted from the simulated water levels?

The procedure description is reformulated.

'' For ensemble two, the highest astronomical spring tide found in the tidal simulations for the period 1948-2100 was used instead of the original tide and the astronomical tides were shifted again hourly.''

L153. Comma after two.

inserted

L154-155. Rephrase this sentence.

This sentence is reformulated. See Chapter 2.5 above.

L166. Remove "To the North Sea" and add "ocean" boundary of the German Bight model.

Changed according to the suggestions.

L204. Comma after conservation.
L217. Comma after 3.
L219. Remove comma.
L238. Comma after EH.

All changed.

L243-245. Divide the sentence in two and add commas.

changed

L251. Comma after EH.
L252. Change “except”-> “with exception of”
L254. Comma after water and members.

All changed.

L258. How much was the increase? These lines are too vague: “few centimeters”, “not a substantial increase”, “nearly no changes”......
L262. Rephrase. Single high waters?

The text is changed and values are added.

“In case of the longest event EL (included in EC Figure 4), both amplification procedures - shifting of the astronomical tide against the wind and replacement of the original astronomical tide with the highest spring tide together with shifting - result in an increase of the highest high water by only a few centimeters. In the original event EL the highest high water already coincides with an astronomical spring tide about 7 cm lower than the highest one. Thus, both applied procedures lead to relative changes of the three highest water level peaks, however not to a substantial absolute increase of the maximum water level during EL. Furthermore, the length of EL shows nearly no changes. Possible amplification was also tested for the entire EC event including EL. The storm tides following EL experience an increase of some single high waters up to 20 to 30 cm together with a decrease of other high waters for some ensemble members. Thus, there was no general amplification regarding the intensity (see chapter 2.5) of the event chain EL/EC. Therefore, the amplification procedures for EL/EC were discarded.”

L264-266. Move to section 2.5.

The text is slightly changed, but we think it is useful at this place to clarify again which events are transferred to be simulated with the German Bight model.

L278-279. There is no need of explaining again where Elbe mouth and Amrum are located.

The phrase is removed.

L285. Change differing to different.
L285. Comma after Amrum.

changed

L287. Rephrase: “The olive curves of both Elbe mouth and Amrum correspond to the same simulation, which incorporates both the largest spring tide and the phase shift of the tide”.

The sentence is changed

''The olive curves of both Elbe Mouth and Amrum correspond to the same ensemble member, which incorporates both the largest spring tide and a phase shift of the tide.''

L289-291. Use duration above MHW instead of time period
L299. Comma after amplification.
L300. Comma after the parenthesis. Is EH_a instead of EH?

All changed.

Figures 7 & 8. Dashed lines cannot be clearly differentiated. I also recommend to add a line showing the MHW level in figure 8 as the changes of the duration above this level are discussed.

A line showing MHW is include in Figure 8 and a sentence concerning the similarity of the dashed and the solid lines is added.

L318. Was a simulation with a SLR of 0.5m also performed? This was not mentioned before.

It was mentioned on L165

L325. The highest
Figures 9, 10 & 11. Font size of legends, axes and labels is too small.
L326. Add parenthesis (HW).
L330. Rises-> raises
L338. Comma after addition. Is decreasing-> decreases.

All changed

L350-357. I do not understand these lines and why are in this section. Rephrase them and move them to the discussion. (Or simply remove them, because it is repeated in lines 435-438)

We added the reason for the analysis to clarify the aim of the investigation.
Here, we address the point, that a chain of events is not only important with respect to coastal protection but also for the drainage of the low lying hinterlands. We explain, which water level in one sewer exemplarily is important for draining and we use this threshold in table 2. We think this text is useful here. In the discussion it is only mentioned that EC would hinder natural drainage.

"During storm tides not only questions concerning coastal protection are important, but the draining of the protected areas during storm tides must be ensured, too. In the lowlands close to the mouth of the Ems draining of urban (e.g. Emden) and agricultural areas (e.g. Knock) is of major interest. The aim of the sewer at Knock is to drain the low lying hinterland (with a ground level of about NHN + 0 m) and keep the inland water level at Knock lower than NHN - 1.40 m (KLEVER (2018)). At Knock the mean low water MLW is NHN - 1.58 m so that draining without pumping is only possible for a short time even during mean tides. Caused by long lasting high water levels during storm tides draining is even more restricted. For the chain of storm tides EC (Figure 8) even without amplification pumping is needed nearly during the whole period of 176 hours (Table 2). The water must be pumped against a water level in the Ems higher than MHW for about 90 hours. This period will increase by about 40 hours in case of a sea level rise of 100 cm."

L380. Increases o causes an increase of.
L382. Highest-> High or an increase of the highest
L412. There-> They are from

All changed

L413. Clarify this line. The absence of considerable increase of storm surges correspond to the magnitude or frequency? Because this study is focused only on 3 types of events, but it does not include any analysis of changes in the trends/ variability of storm surges.

That is correct, we do not investigate long-term trends in this study, however, the used met-ocean data sets were analyzed earlier. The text is changed accordingly.

‘’These events originate from the first half of the emission scenario period of two different climate realizations. Gaslikova et al. (2013) showed that the annual maximum water levels of these climate realizations displayed strong multi-decadal variability but no significant long-term trends from 1961 to 2100. Thus, the found highest water levels exceeding the water levels measured since the beginning of the 20th century at Borkum (Figure 3) could be possible already under present-day conditions as no sea level rise is included in the original realizations.’’

L433. Rephrase.

The sentence is changed:

‘’Against the background of climate change and the need to develop future coastal protection strategies it is not only important to know the possible height of an extreme event but also its duration.’’

L460. Particular

changed