Dear Professor Kienast, dear referees,

Thank you very much for the helpful comments and suggestion for our manuscript. We can follow most of the arguments of all referees and will thus incorporate almost all suggestions made. In the following, we have listed responses to the different points made. There are only a few cases in which we would prefer not to follow the suggestions made, but we offer compromise solutions in most of these cases.

Kind regards also on behalf of the co-authors

Ulrich Kotthoff

Comments/replies to the Review by F. Naughton (Referee)

Referee 1: The manuscript "Reconstructing Holocene temperature and salinity variations in the western Baltic Sea region: A multi-proxy comparison from the Little Belt (IODP Expedition 347, Site M0059)" by Kotthoff et al., presents new and a large array of proxy data from IODP Expedition Site M0059 covering most of the Holocene period. The study is methodologically sound, the results are significant, and the interpretation is well justified. The manuscript is very clearly written and supported by good quality figures. The manuscript makes several very valuable contributions to the palaeoenvironmental understanding of the last 8 kyr in the western Baltic Sea by using coupled ocean and atmospheric multi-proxy data. Overall, I would strongly recommend this manuscript for publication in Biogeosciences journal with very minor modifications in response to my comments on the pdf.

Yours sincerely, Filipa Naughton

Please also note the supplement to this comment: http://www.biogeosciences-discuss.net/bg-2017-101/bg-2017-101-RC1-supplement.pdf

Authors: We are thankful for this review of our manuscript. In the following, we respond to the different comments in the text. We are particularly thankful for the detailed checking of our references. In the following we reply in detail to the comments in the supplements.

Page 3

Referee 1: "replace Frenzel et al 2005 by Frenzel and Boomer 2005"

Authors: Will be replaced.

Referee 1: "is this reference Filipsson et al., 2016? if not introduce this reference in the references section."

Authors: The reference will be corrected.

Page 5

Referee 1: "this reference is not mentioned in the reference section"

Authors: Will be added.

Referee 1: "references should be in chronological order"

Authors: We will reorganize the references.

Page 6

Referee 1: "Even if this paper is accepted how can we see the age model? In the references we can't see where this mentioned paper will be published. You should have a table with core depth and the

radiocarbon ages as well as a figure with the resulted age model. You can cite the Van Helmond et al accepted paper as well."

Authors: The paper is now published, but it is probably reasonable to add such a table nevertheless.

Referee 1: "How? you must say how they were extrapolated: e.g. using clam program? by hand using the last sedimentary rate value?"

Authors: The ages were indeed extrapolated linearly, which will be explained in more detail.

Referee 1: "Although this correlation is very interesting and a good point to reinforce the reliability of your age model, it would be nice to see the age model of site M0059. The temporal resolution of pollen data from Site M0059 is much lower than that of Lake Belau, therefore the last 2500 years of correlation have some slight discrepancies. However it does not seem to affect your interpretation."

Authors: See above. We will add more information on the age model and the Helmond et al. paper is now accepted.

Referee 1: "need a reference of the method"

Authors: We will add a reference.

Referee 1: "replace to: ....for diatom analysis were...."

Authors: Will be done.

Referee 1: "include "the" to be consistent with what said in the diatom methodology."

Authors: We will generally see that the methodologies become more consistent (see below).

Page 11
Referee 1 makes several suggestions concerning Fig. 2, we will follow these suggestions.

Page 12
Referee 1: "you should/(could if you have) include the pollen concentration curve next to the TOC curve."

Authors: The pollen concentration could be calculated for most of the samples, we will add a related curve, and also add a dinocysts concentration curve (see below).

Page 14:
Referee 1: "since you have just isolated points with ammonia beccari i wonder if this is representative"

Authors: *Ammonia beccarii* is present in the surrounding samples too (<3%). We agree that it should be mentioned that the particular high abundance (~15%) is reflected in one sample. We will change the text accordingly and also write "higher abundance compared to…" instead of "relatively high abundance"

Referee 1 makes several suggestions for section 3.3.3 which we will all follow.

Page 15:
Referee 1 mentions a missing reference which we will add.

Referee 1 (correction): "which can be found"
Authors: We think that there is a misunderstanding here, which can be avoided by rephrasing the related sentences.

Referee 1: "include non-heterotrophic taxa in the text here"
Authors: We will check for consistency between this sentence and the following sections.

Page 16: Referee 1 made 4 comments on this page which all shall be incorporated.

Page 17: The wrong/missing reference mentioned by Referee 1 will be corrected/added.

Page 18:
Referee 1: "in figure is until 5.5 ka"
Authors: We will rephrase the sentence.

Referee 1: "As previously mentioned in my comments the single peaks cannot be representative of major changes. "

Authors: We will rephrase the sentence in accordance with the comment Referee 1 made concerning the related aspect on page 14.

Referee 1: "you could probably also mentioned based in the pollen based temperatures estimates that evaporation was not strong because MTWA did not increase during this interval."

Authors: True, we will add that aspect.

A missing "(" will be added.

Page 19:
Referee 1: "again one single data point is not enough to extrapolate a rapid change in the salinity"

Authors: Similarly to the comments above, we agree to some degree and will rephrase the sentence. But we still regard these increases in percentages as important. Even if it is only a single point (which will be clearly mentioned), the counting sums for the related sample are so high that the signal is significant and a change in salinity is a logical inference. Besides, a salinity signal is in congruence with the other proxies.

Referee 1: "rephrase the end of this sentence"
Authors: Will be rephrased.

Referee 1: "since you do not have a peak in pp at 1700 but rather an increase along the EZ3 it would

be better to just mention what happened during the entire interval."

Authors: We will rephrase the related sentences.

Referee 1: "If pp increase (and runoff) why there is no signal of that in the TOC and BTI curves? well maybe a slight increase of the TOC could also support this idea?"

Authors: Indeed, diatom based pp often fluctuates in an inverse relationship with soft bodied algae and other forms. Especially where eutrophication is concerned as soft algae will replace diatoms. The lack of signal in TOC curves likely reflects trade-offs between diatom and soft algae productivity (i.e. diatom productivity replaces or is replaced by other pp groups, so a change in TOC may not always be seen). We will discuss this in the related section.

Page 20:
Referee 1: "i think that you can also do not have enough resolution in your record to show that linkage. so you can also propose that."
Authors: We will mention that the temporal resolution may hamper the related interpretation.

Page 21:
Referee 1: "at a long-term sst seems to decrease gradually as the LDI and even more similar to the MTWA decrease trend, even if the signal of TEX is noisy!"

Authors: We are not sure what the question is in this case.

Referee 1: "I do not see any similarities between the TEX and the clumped isotope record for this period. Maybe you should look to the data at long term and forget the rapid reversal episodes."

Authors: We have to agree with referee 1, we cannot really suggest a common trend. We will only use the absolute values concerning the clumped isotope record.

Referee 1: "the pattern is from 19º to 17.5 and not from 17.5 to 19. please change"

Authors: Will be corrected.

Page 23:
Referee 1: "this reference is in press in the reference section"

Authors: The Krossa et al. paper has now been published, we will correct the references (see below).

References:
Referee 1 has marked several corrections to be made in the references, we will incorporate them.

Comments/replies to the review by R. Limoges.

Referee 2: "This paper compares and uses the signals provided by multiple biogenic proxies to interpret the hydrologic evolution in the western Baltic Sea region during the past 8000 years. Inter-comparison of the proxy-data and quantitative reconstructions allow for robust reconstructions and time constraint of the major regional hydrological transitions (salinity and temperature). This paper is a good contribution and I suggest it is accepted with minor modifications. I kindly refer to the

attached pdf (supplement).”

Authors: We are thankful for this review of our manuscript. In the following, we respond to the different comments in the additional text file provided.

Referee 2: “General comments
The information provided in this paper is highly relevant and the interpretations are well justified. However, although I understand it is difficult to be concise when reporting on so many proxies, the manuscript is long and should be thoroughly reviewed to remove repetitive sections, and to restructure certain sentences and paragraph sections (particularly lines 532-556 present one long section in which the reasoning is jumping around a bit between proxies and from one time interval to another). Furthermore, there are a lot of abbreviations and this makes the reading unnecessarily challenging.”

Authors: We agree. The MS already went through several “shortening cycles”, but we try to remove redundancies. It is true that the mentioned section of the text can be shortened in particular. We will also discuss removing some of the abbreviations.

Referee 2: “I would be cautious with the discussions concerning the performance of the different proxies. Although many proxies are used, only a very few amongst them are used in an optimal way (e.g. not statistically robust dinocyst counts; a noncalibrated foraminifer species used for Mg/Ca-based estimates, contamination of foraminifer test used for Mg/Ca ratios, and assemblage and geochemical analyses on foraminifera that are suspected to be affected by dissolution - which probably also is the same for the ostracodes, etc) – this is okay if trying to answer a scientific question, but I would have some reservations when the authors evaluate the performance of these proxies. Their applicability can be evaluated, but not their performance (question of using the right term).”

Authors: This aspect is indeed very important. It is true that we cannot evaluate the “performance” of some of the proxies. We will carefully rephrase the related sections. We should also add a sentence in the abstract clarifying that it is the complex setting which could make the applicability difficult.

Referee 2: “Specific comments
Figure 1. The « Little Belt », « Great Belt » and « Öresund » should be added to this figure so the readers that are not familiar with the region can refer to this map. Perhaps also « Lake Flarken », also mentioned in the text.”

Authors: Will be done.

Referee 2: “Figure 3. What species of diatoms were considered F, BF, BM, etc?”

Authors: “We have already prepared a table with all encountered species, and which category they were included in. This table (comprising more than 170 species) should be added to the supplementary data.”

Referee 2: “Figure 3. Add dinocyst concentrations.”

Authors: Dinocysts concentrations will be added.

Referee 2: “Replace “non-heterotrophic dinocysts” by “phototrophic”.

Authors: We will do so.

Referee 2: "Introduction
Lines 54-59. These sentences are somewhat unrelated to the remainder of the manuscript and the objectives of this study."

Authors: We will reconsider these sentences. While oxygen concentration is an important factor for our study, it is true that the introduction should not start with this aspect. The first sentence could be removed in any case.

Referee 2: "Line 77. Can you be more specific here? What conditions can influence the application of Mg/Ca in the Baltic Sea?"

Authors: We will be more specific in an updated version of the manuscript, and mention e.g. low salinity and large gradients in carbonate system parameters as important factors in the Baltic.

Referee 2: "Lines 78-83. Include references here."

Authors: References will be added.

Referee 2: "Lines 86-87. You introduce the TEX86 proxy, but the text further down mentions the TEXL86 proxy; please clarify what the TEXL86 is."

Authors: We will add more detailed information on the TEX^{L}_{86} , which is a derivation of the originally proposed TEX_{86} lipid paleothermometer and particularly applicable in subpolar and temperate settings, to the introduction section.

Referee 2: "Line 89. What does LDI stand for?"

Authors: "LDI" (long chain diol index) is explained in the abstract, but this should be indeed explained in the main text. In addition, "LDI" is among the abbreviations which we might avoid in context with the comment made by Referee 1.

Referee 2: "Line 94. The term biogenic should be used instead of "biotic"."

Authors: We agree.

Referee 2: "Line 99. The length of the core is not important, it is the sedimentation rates and temporal resolution (it is not because it is long, that the sedimentation rates and temporal resolution are high)."

Authors: True, we will rephrase the related section, though the length information might be interesting, too.

Referee 2: "Line 103. […] moreover to Greenland ice core records and marine records from the North Atlantic □ this sentence seems out of place, since this is not done later in the manuscript."

Authors: In context with comments by Referee 3, we might add comparisons with the Greenland ice cores. We will change the sentence accordingly.

Referee 2: "Line 103. replace "records" with "cores" (drilling provides cores, analyses provide

records)"

Authors: We agree.

Referee 2: "Methods
- The thickness of the samples is only given for forams, but not for other proxies.
- For many proxies, the same amount of samples was analyzed (36). Presumably, these represent the same samples/levels in the core (although this is not that obvious from the manuscript). Instead of repeating this information every time, perhaps a short introducing paragraph can be added to provide this information (same-sample analyses for which levels, which ones in higher resolution, etc)"

Authors: It is, unfortunately, not the case that all authors could use samples from the same levels due to limited core material. We used closely neighboring samples, in most cases and aimed at at least 35 samples per proxy over the time interval analysed. This information and the sample thickness will be added.

Referee 2: "2.1. Lines 114-115: is the latitudinal span (numbers) relevant? So why not the longitudinal span, too? But both would not seem relevant to the study, which focuses on a specific, smaller region. The coordinates of the core would seem to suffice."

Authors: We will remove this, and generally, we see options to shorten this section.

Referee 2: "2.2.3. What mesh sizes were used for sieving?
● Did you add marker grains for calculating the concentrations?"

Authors: The mesh size (7µm) and information on pollen concentration will be added.

Referee 2: "● Zonneveld and Pospelova, 2015 is not an appropriate reference here – it is a determination key. Perhaps you can refer to de Vernal and Marret, 2007 [de Vernal, A., Marret, F., 2007. Organic-walled dinoflagellates : tracers of seasurface conditions, In Hillaire-Marcel and de Vernal (eds.) Proxies in Late Cenozoic Paleoceanography, Elsevier,pp. 371-408.]"

Authors: We will exchange the reference.

Referee 2: "● What does "rarity of counted types" mean?"

Authors: We will rephrase the related sentence.

Referee 2: "2.2.4. Line 201: "selected depths": do you mean the same 36 depths, or a selection of these 36?"

Authors: We will rephrase this sentence (the 36 samples are meant).

Referee 2: "2.2.5. The use of heavy liquid separation is understandable, but is this common practice? Could you provide a reference that illustrates the influence - or, ideally, the lack of it – on assemblage composition (selective removing of certain species because of sediment infilling, fragmentation of fragile species, etc.). Ideally, in order to really test the power of the proxy, all samples should have undergone the same preparation method..."

Authors: The residual fraction after heavy liquid separation was checked for foraminifera; only very

few specimens were found. The heavy liquid separation thus did not influence the analyses. We will rephrase parts of section 2.2.5 and add this aspect.

Referee 2: "2.2.6. Line 230. A total of 75 30cm3 sediment samples were processed for ostracod analysis (confusing otherwise)."

Authors: We will rephrase this sentence.

Referee 2: "Results
Line 455. Operculodinium centrocarpum (?), Spiniferites spp., Lingulodinium machaerophorum (?)"

Authors: We will add the species names/spp.

Referee 2: "Line 460. When using "Gymnodinium cf. nolleri", it is implied that a cyst type was found with a morphology that looks like G. nolleri, but at the same time very clearly is not G. nolleri. Is this what the authors mean: cysts whose morphology cannot be attributed to a known species? Or, if different species of Gymnodinium are meant (i.e. nolleri, catenatum, microreticulatum), then "Gymnodinium spp." should be used."

Authors: In this case, we cannot completely agree with referee 2 – "cf." does not indicate clearly that the specimens are not *G. nolleri*. Still, we should rephrase this sentence to explain what is meant. What we wanted to imply is that we are sure we found *Gymnodinium nolleri*, but that we cannot exclude that some of the *Gymnodinium* specimen we found belong to other species. We analyzed a few specimens in detail (also discussed the identification with S. Ribeiro as mentioned in the acknowledgements), but some specimens were only identified via their size and structure, not via a detailed analysis.

Referee 2: "Discussion
Transition from line 492 to line 493. Somehow a circular reasoning, as the variations themselves are inferred from the proxies. The first two sentences of this paragraph could be removed."

Authors: We agree that this section contains redundant phrases. We will rephrase this section.

Referee 2: "Line 502. How do the ostracodes indicate low primary production?"

Authors: We will rephrase this sentence (it was meant that they indicate freshwater conditions)

Referee 2: "Lines 501, 504, 506. These sentences could be restructured to avoid repetition. In addition, the sentence "these factors indicate that EZ1 presents a low productivity freshwater environment" should be moved downward, as (the more convincing) arguments are given following this sentence; the low concentrations of marine palynomorphs does not indicate that the setting was one of low total productivity."

Authors: We agree and will rephrase these sentences.

Referee 2: "Line 526. Add reference after "[…] may indicate more saline conditions"."

Authors: We will add a reference.

Referee 2: "Lines 543-556. This entire section needs a bit of restructuring and rephrasing; now it is sometimes confusing and unclear what periods and water masses (bottom, surface) are compared and discussed."

Authors: We will carefully rephrase this section.

Referee 2: "- bottom water = increasing salinity (Line 545-546); surface water = decreasing salinity (Line 549-553), correct? This contrasting evolution (if I understood correctly) is worth stressing and discussing further?"

Authors: Indeed, we already discussed this internally and will follow the suggestion.

Referee 2: "- L553: ...regard this as the most… what period exactly (early Littorina, entire Littorina,...)? "most marine" seems in contradiction with decreasing salinity..."

Authors: True, we will rephrase this sentence.

Referee 2: "- L554: increased with respect to what? Modern? EZ1?"

Authors: It is true that this should be explained. (We mean the preceding lower salinity.)

Referee 2: "Lines 559-601. please rephrase, there seems to be something not entirely correct about this sentence."

Authors: In any case, the sentence is very long, we will rephrase it to make it clearer.

Referee 2: "Line 600. The use of "juvenile (percentages)" comes out of the blue here. Also, given the low ostracod counts, how significant are such percentage/relative changes? How many specimens are we talking about here?"

Authors: It is true that the number of specimens is low in some samples. We will reconsider if we need this information at all.

Referee 2: "Line 611. Inorganic and inorganic proxies are not the best terms here – perhaps "geochemical analyses" and "biomarker""

Authors: The best solution may be to write "inorganic- and organic-based" (since biomarkers" are also geochemical).

Referee 2: "Line 637. Not everybody is familiar with the Boreal and Atlantic terminology and timescale; please introduce properly or replace with/add actual dates."

Authors: We will do so.

Referee 2: "Line 651. When do these eustigmatophyte algae bloom?"

Authors: Currently no information on the annual growth cycle of eustigmatophytes in the Baltic Sea or

other brackish-marine settings is available. Previous studies indicate that the LDI-derived surface water temperatures match best with summer surface water temperatures in various oceanographic settings such as the Southern Ocean (Lopes dos Santos 2012). Likewise, the core top sample at Site M0059 shows LDI-derived temperatures that are most similar to surface water temperatures in July, providing additional evidence that blooms of eustigmatophytes occur during summer in the Little Belt region. We will add a short discussion on the timing of eustigmatophyte blooms and LDI-derived temperatures in the Baltic Sea.

Referee 2: "Line 656. LCD would seem a good candidate to be left out as abbreviation and be written out in full instead (as is done just a few lines above). Other good candidates, since used only very rarely, would surely be BIT (line 679), MWP and MHP (685; the latter only used once, i.e. where the abbreviation is given!?), HTM (line 691), and BWT (line 697)."

Authors: We will follow at least some of these suggestions, it is true that we use quite a high number of abbreviations.

Referee 2: "Technical corrections
Line 43. […] changes in salinity, but often do not allow quantitative
Line 44. […] is associated with particularly large uncertainties […]
Line 248. Due to poor preservation
Line 290. A total of 40 sediment samples collected […]
Line 351. […] was divided into four overall environmental zones.
Line 397. A. beccarii
Line 456. G. nolleri
Line 465-467. This information belongs to the method section.
Line 501. "biological" should be replaced by "biogenic".
Line 647. Same as above. Change different inorganic and organic for "biogenic proxies"."

Authors: We will incorporate these corrections.

Comments/replies to the review by Referee 3.

Referee 3: "The manuscript by Kotthoff et al. based on a multi-proxy approach is very interesting. However, while combining that much different proxies is not an easy task, some more thorough discussion is needed concerning (1) the climatic forcing possibly explaining the different salinity and temperature trends observed over the Holocene, and (2) the high discrepancies between proxies for the same parameter (temperature). A graphical comparison with previously published records from the study area is also missing. The text suffers from some imprecision in the Results and the Discussion, some parts of the text should be reorganized, some figures should be modified and some new figures should be provided (as supplements). Finally, a calibration issue related to some organic proxies should be addressed. Therefore, I recommend the publication of the present study, but only after major revisions."

Authors: We are thankful for this detailed review, particularly of aspects related to the biogeochemical proxies. We agree that it would make sense to discuss the climate forcing and some of the discrepancies in more detail (see below). Concerning graphical comparison with other records, we are not sure if this should be done in the framework of this manuscript, because on the one hand, there are no nearby records covering a similarly long time interval, and on the other hand, the sheer amount of proxies we use would result in numerous additional figures. At least a few of the authors plan to examine certain intervals from Site M0059 in higher resolution, so that maybe a graphical comparison with other records should be reserved for this? But, as discussed below, a compromise might be to add some other datasets to a supplementary figure.

Referee 3: "Major comments:
The introduction should be reworked partly. After the second paragraph, it sounds like an enumeration and description of the proxies that will be applied in the study. It is not necessary and belongs rather to the Discussion part."

Authors: We had a long discussion how to handle this. In our opinion, we found a reasonable approach – in an earlier version, we had shifted the proxy description to the subsequent sections, and most co-authors did not like this approach. We would thus prefer not to move these descriptions. This does not mean that we cannot follow referee 3's arguments, but we think in this case it is also a matter of different tastes. We could, however, shorten the related sections.

Referee 3: "Instead, previously published Holocene records from the Baltic Sea and the Skagerrak region should be mentioned and the main results should be described as in lines 97 to 104, but in more details.
In my opinion, at least the following studies on Holocene temperature and salinity changes should be mentioned: Emeis et al., 2003, The Holocene (salinity and temperature); Warden et al., 2016, Organic Geochemistry (salinity); Krossa et al., 2015, Boreas, and Krossa et al., 2017, The Holocene (salinity and temperature); Ning et al., 2015, Boreas (salinity); Butruille et al., 2016, The Holocene (temperature); Zillen et al., 2008, Earth-Science Reviews (climate and hypoxia); Widerlund and Andersson, 2011, Geology (salinity). Based on these previous results, the necessity of a long and continuous record from the Belt Sea as intermediate location linking the Baltic Sea and the Skagerrak region can be introduced (in lines 97-104)."

Authors: We agree that it would make sense to incorporate more information from these papers in the introduction and will do so.

Referee 3: "Some other parts of the text should be reorganized. As some results are discussed/presented in Section 2.2.1, I would suggest merging it into Section 3.1. The problem related to Mg/Ca contamination is mentioned at least three times in the manuscript (Methods, Results, Discussion). Because of repeating this issue again and again, one could consider removing completely this record from the study as this proxy is not really reliable. It would be a pity however. Therefore, I recommend shortening and grouping the different parts about this contamination issue somewhere in part 4.2, and discussing this issue in more details in the supplements (if necessary). "

Authors: These are reasonable suggestions. We will in particularly follow the suggestions concerning the Mg/Ca contamination aspect. Concerning section 2.2.1, it is true that results concerning the age model are mentioned, but these are not new results but rather necessary information from other publications (with the van Helmond-Paper presenting the age model now published, see above) which in our opinion should be given before the proxy methods are introduced, so we would rather prefer to leave this here.

Referee 3: "Concerning Section 4.2, while it is possible to discuss the records based on the different environmental zones for the salinity and productivity proxies, it is confusing for the temperature proxies as these latter present completely different long-term and short-term trends. I suggest reworking/reorganizing this part. First, the differences in the temperature proxy records (trends, absolute values, amplitudes, etc) should be discussed, then the temperature trends should be summarized as a function of the environmental zones, and finally the potential forcing behind the temperature records should be discussed (see below)."

Authors: We agree that section 4.2 should be reorganized and will follow the concept suggested by Referee 3.

Referee 3: "I have a few comments concerning the TEXL86 temperature proxy. First of all, I was wondering if a standard was used for GDGTs quantification. If yes, I suggest using the absolute concentration of the branched GDGTs rather than the BIT index as the BIT index is often mostly function of variability in crenarchaeol (usually the dominant GDGT)."

Authors: We agree with Referee 3 that quantifying absolute concentrations of GDGTs is desirable. Unfortunately, no standard was available for quantification of isoprenoid and branched GDGTs and thus we used the BIT index to provide information on the relative changes of the aquatic and terrestrial derived GDGTs.

Referee 3: "Generally speaking, the use of TEXL86 should be avoided because the crenarchaeol regioisomer plays a role in the temperature predictability (relatively more of the regioisomer is observed at higher temperatures)."

Authors : We are somewhat surprised by this comment as the TEX^{L}_{86} has specifically been designed for low temperature environments such as the Baltic Sea. Furthermore, the TEX_{86} has been established as a marine temperature proxy. It is of limited use in freshwater and brackish settings though due to the absence or low abundance of the crenarchaeol regioisomer in many lacustrine and brackish systems. In fact, the crenarchaeol regioisomer could only be identified with confidence in about 60% of the samples that have been analyzed for isoprenoid GDGTs. Therefore, we consider the TEX^{L}_{86} most suited to reconstruct the temperature variability of the Little Belt. However, we will add an additional graph to the supplements showing the TEX^{H}_{86}-temperature record and explain differences in TEX^{L}_{86} and TEX^{H}_{86}-temperature profiles in the discussion section.

Referee 3: "Furthermore, the TEXL86 calibration from Kabel et al. (2012) is only based on the highest correlation with summer SST, but has no "biological grounds". When looking at the supplementary information in Kabel et al. (2012), it appears that the correlation is high (r2 > 0.7) for all months from May to November (i.e. not only for the summer months) and not only for TEXL86, but also relatively high (r2 > 6 from June to October) for TEXH86. Moreover, the IODP M0059 site location is out of the area covered by Kabel et al. (2012) calibration, what may play a role considering a possible influence of strong salinity gradient on Thaumarchaeota distribution in the western Baltic Sea."

Authors: Referee 3 is of course right that a comparatively high correlation with summer to autumn temperature is also observed for the TEX^{H}_{86}. But it is significantly lower than for the TEX^{L}_{86} and therefore we deem the TEX^{L}_{86}-reconstructed temperatures more reliable. Although the Kabel et al. (2012) calibration does not include the Little Belt region, the vicinity of the studied site to the Baltic Proper and the strong environmental gradients that are observed in the western Baltic region suggest that other calibrations may not be as applicable to reconstruct SSTs. However, we have now also calculated SST based on the TEX^{H}_{86} and will discuss difference between both temperature profiles in the text.

Referee 3: "Another factor potentially complicating the TEXL86 record is the presence of a redoxcline and hypoxic to anoxic conditions. It is known that in the modern, Thaumarchaeota are most abundant at depth near the redoxcline in the Baltic Sea (e.g. Labrenz et al., 2010, ISME Journal; Berg et al., 2014, ISME Journal). Therefore, on the one hand the recorded temperature may rather be from the subsurface, or even the near bottom if anoxic conditions are present near the bottom as in the Bornholm Basin."

Authors: We agree with the reviewer that the TEX^{L}_{86} (as the TEX_{86} in general) may record subsurface water in certain oceanographic regions. However, comparing our TEX^{L}_{86}-reconstructed water temperature from the core top sample to measured water column temperatures strongly argues for a synthesis of isoprenoid GDGTs in surface waters of the Little Belt region as ambient subsurface temperatures are much too low and do not explain the proxy-based temperatures. Therefore, we consider

it unlikely that Thaumarchaeota growing at the redoxcline were a substantial source of GDGTs. However, we will extend our discussion on the spatiotemporal water column signal of archaeal GDGTs in the Baltic Sea.

Referee 3: "On the other hand, culture experiments have shown that increased O2 limitation may result in increased TEX86 SST estimates (Qin et al., 2015, PNAS). Indeed, van Helmond et al. (2017) have shown that seasonal hypoxia occurred over the last 8,000 years at Site M0059 and intensified during the HTM and, more especially, during the MCA, i.e. when the TEXL86 temperatures are highest. This aspect should be shortly discussed. It would be very interesting to plot also a TEXH86-based temperature record using a global calibration, e.g. Schouten et al. (2013, Organic Geochemistry) or Kim et al. (2012, EPSL) subsurface calibrations on Fig. 4 (or as supplement) as well, and to discuss potential differences."

Authors: O_2 limitation may indeed result in increased TEX_{86}-SST estimates but as GDGTs synthesizing Thaumarchaeota likely inhabited the surface waters of the Little Belt, they were likely little affected by the spread of seasonal hypoxia during the HTM and MCA. Both periods, on the contrary, are well known to be characterized by optimum climate conditions and we thus consider the trends to warmer surface water temperatures reasonable. However, we will include a more detailed discussion on how environmental parameters (including redox conditions) may affect the TEX_{86} and its interpretation to the "discussion" section. Although in lower resolution we will now also include a figure showing TEX^{H}_{86} temperature trends using a set of different calibrations and a short discussion in the supplements.

Referee 3: "Moreover, as methanogenic and, more especially, methanotrophic archaea produce GDGTs involved in TEX86 in substantial amounts, it would be interesting to test their potential influence by plotting e.g. the Methane Index (Zhang et al., 2011, EPSL) as supplementary information."

Authors: We thank the reviewer for the comment. To determine the possible effect of GDGTs derived from methanotrophic archaea on the calculation of the TEX^{L}_{86}, the Methane Index (MI) was calculated and will be added as supplementary information to the manuscript. MI values are below those considered diagnostic for methanotrophy and thus suggest that the TEX^{L}_{86} was not confounded by the addition of GDGTs from methanotrophic archaea.

Referee 3: Some imprecision are apparent in the text. Compared to e.g. the LDI record (same samples as for GDGT analysis), the oldest three samples of the TEXL86 record are "missing" (not plotted), without explanation, what makes a true comparison difficult (see e.g. lines 667-668)."

Authors: The missing TEX^{L}_{86}–based temperatures from EZ1have now been measured and will be added to the manuscript.

Referee 3: "Furthermore, to be as correct as possible, a global TEX86 calibration for lakes (e.g. Powers et al., 2010, Organic Geochemistry) should be applied for the samples from EZ1, as this latter is characterized by freshwater conditions (lines 510-511)."

Authors: We absolutely agree with the reviewer but did not use the lake calibration for the one sample from EZ1 because the sample is characterized by elevated sulfur concentrations, indicating a brackish-marine influence on the uppermost sample from EZ1. The remaining samples (that have only now been added to dataset and shall be added to the manuscript), however, have been deposited exclusively under freshwater conditions and we used the summer lake surface water temperature calibration of Powers et al. (2010) to reconstruct TEX_{86}-based temperatures.

Referee 3: "Obviously the TEXL86 record is NOT "... to some degree similar to the clumped isotope

record" as stated by the authors (lines 674-676). For examples, the temperatures are equally high in the HTM and the modern in the clumped isotope record, but not in the TEXL86 record. The absolute values are different as well."

Authors: This comment is similar to one of referee 1. We admit that the cited statement cannot be made, we will rephrase the sentence.

Referee 3: "Moreover, if plotting the TEXL86 temperature data of Kabel et al. together with Site M0059 on Fig. 4 (what I suggest to do), I suspect that these two records are not that similar (lines 685-688) concerning both the temperature amplitudes and the trends."

Author: We are surprised by the comment as we did not state that the Kabel et al. record and our record are similar. The Gotland Basin is significantly deeper than the Little Belt region and elevated temperatures at the latter site may simply be due to a shallower production depth and consequently higher water temperature. Trends in both records, however, are indeed similar.

Referee 3: "As for the TEXL86, a calibration based on lake sediments (Rampen et al., 2014) should be used for the samples from EZ1. The strong temperature increase (10 ◦C) at the transition between EZ1 and EZ2 may be an artefact due to the different calibrations as discussed in lines 659-661."

Authors: Unfortunately, this comment is not clearly phrased. We assume that Referee 3 refers to the LDI? In any case, we now applied the lake specific calibration by Rampen et al. (2014) to sediments deposited in the freshwater environment of EZ1 (the putative Ancylus Lake phase) to reconstruct surface water temperatures.

Referee 3: "The Diol Index is not convincing as surface salinity proxy. It suggests similar conditions/salinity during the freshwater lake, as well as in the mid-Littorina Stage (ca. 4,500- 4000 cal. yr BP) and maybe the late Littorina Stage (ca. 1,000-500 cal. yr BP), although the Littorina Stage was marine-to-brackish. While salinity may affect this index together with temperature, I suggest removing this proxy from the study, or discussing it in more details."

Authors: We cannot completely agree with the reviewer. Obviously, it is true that the index suggests that the salinity was similar during the freshwater lake stage and specific intervals during the Littorina Sea stage. Yet, however, the diol index is an empirical measure that does not gradually change with salinity but allows a rough separation of freshwater, brackish-marine and marine conditions with some overlap. The diol index, however, shows a similar trend as *Gymnodinium* suggesting that there was an intermediate interval of "fresher" conditions in the Little Belt region. Based on this observation, we consider the diol index an valuable addition to our proxy records but discuss it potential constraints in more detail.

Referee 3: "Based on the proxy results, it is in my opinion difficult to separate between surface and deep salinity changes, especially at 37 meter water depth. The salinity history reconstructed here concerns probably rather the complete water column. While a precipitation increase (pollen-based record) may explain a salinity decrease between 8,000 and 4,000 cal. yr BP, why is the salinity increasing/high between 4,000 and 1,000 cal. yr BP, while the precipitations are highest? Please, discuss the potential mechanisms for this salinity increase, as well as for the salinity decrease over the last 1,000 years. Considering the high heterogeneity in the different temperature proxies from Site M0059, some previously published, marine-based and pollen-based temperature records as mentioned in lines 626-636 should be plotted in Fig. 4 for comparison."

Authors: We agree that we should discuss the potential mechanisms for the salinity increases after 4,000 cal. yr BP. Concerning additional temperature records, see below (e.g. comments concerning

Figure 4).

Referee 3: "A discussion concerning the forcing and mechanisms behind the temperature records and the difference in the temperature trends of the proxies is missing, or not thorough enough. Why are the LDI and TEXL86 records that much different although both should reflect summer temperature? Why are the pollen-based and TEXL86-based summer temperature records that much different? Why are the trends in MTCO and MTWA opposite? What are the expected/modeled evolutions of winter and summer temperature in northern Europe during the Holocene? How does seasonality change over the Holocene? What about insolation? Etc . . ."

Authors:
This is also in context with the question if other records should be discussed. While we are reluctant to add to many additional records (see comments concerning Fig. 4), we agree that the discussion on these aspects should be broadened.

Referee 3: „Minor comments (some redundancy with the major comments is possible):

Lines 124-125: Are those surface or deep currents?

Authors: We are going give more details here. The outflowing low salinity water surely flows at the surface and remains traceable there to some degree. The saltier, denser water flows along the bottom, sinks when it crosses a sill and only becomes shallower when it mixes with other, lower saline water masses.

Referee 3: "Lines 145-147: What are the time intervals for "a transitional low salinity phase" and "the Littorina Stage"? Please, add."

Authors: We will add information on the time.

Referee 3: "Line 330: Is now published in Marine Geology."

Authors: We will update the reference.

Referee 3: "Lines 340-343: For consistency, these lines should be moved to Section 3.2. What is meant with "... and between Holes (Fig. 2)."?"

Authors: That sentence needs rephrasing, it was meant that the TOC data of all Holes from Site M0059 is consistent.

Referee 3: "Chapter 3.3.2 (font size too small): A reference to Fig. 3 is missing."

Authors: Will be added.

Referee 3: "Line 491: Why not mentioning that this is the Ancylus Lake Stage as in van Helmond et al. (2017)?"

Authors: In opposite to van Helmond et al., there is no agreement among us that this is unequivocally the Ancylus Lake Stage. But the possibility will be discussed.

Referee 3: "Line 502: Why "lowermost part"? The complete EZ1 suggests freshwater conditions."

Authors: True, this shall be rephrased.

Referee 3: "Lines 504-505: But this is in disagreement with van Helmond et al. (2017) suggesting a eutrophic freshwater environment with high productivity..."

Authors: True, the discrepancy should be mentioned and discussed.

Referee 3: "Lines 510-511: To be as correct as possible, a TEX86 global calibration for lakes (e.g. Powers et al., 2010, Organic Geochemistry) should be applied for the samples from EZ1, as this latter is characterized by freshwater conditions."

Authors: We agree with the reviewer and now apply a freshwater TEX_{86} calibration to the samples from unit EZ1.

Referee 3: "Lines 537-538 and 541: Could you develop/discuss these sentences about foraminiferal 18O? It could also be a temperature effect ..."

Authors: Change in $d1^8O$ indeed could also be a temperature change, but temperatures based on the other proxies mainly decrease during this interval, and the change in salinity would fit with the assemblage based changes. Therefore it seems more likely that the change in $d^{18}O$ is mainly salinity. We will discuss this in more detail.

Referee 3: "Lines 546-547: On which proxy (proxies) are these salinity values based?"

Authors: We will add that they are based on ostracods.

Referee 3: "Line 549: The values of the Diol Index are as low as in freshwater water conditions. This is not so realistic."

Authors: As previously mentioned the diol index does not gradually change with salinity and absolute values of this proxy should be interpreted with caution. Changes in the diol index, however, indicate a freshening of the Baltic Sea during the Littorina Sea phase, which is in agreement with variations in *Gymnodinium*. However, we agree that absolute diol index values might be somewhat misleading and now discuss this and possible pitfalls in the interpretation of this index in the text.

Referee 3: "Line 552: I'm not convinced about this decreasing surface salinity as (1) the results based on the Diol Index are not realistic and (2) there is no trend in the diatom assemblages."

Authors: But marine diatoms are decreasing after ca. 6500 yr BP. That this decrease starts later than the decrease in the Diol Index might indeed need some additional discussion, which we will add. Concerning the "realism" of the Diol Index, see comment above.

Referee 3: "Lines 562-565 and 576: The peak (one sample ...) in marine diatoms is not synchronous with the high values of the Diol Index that occurred after the transition. These are two distinct events. And the Diol Index values are not "particularly high" compared to the rest of the record."

Authors: We admit that the peak in the Diol Index is around 900 yr BP, not between 1.000 and 1.200 yr BP, but a sample for the Diol Index at around 1150, if not "particularly high", still shows high values compared to the average value of the whole record. Furthermore, there are three samples between 1.000 and 1.200 yr BP with high marine diatom percentages. So we would still not exclude that the signals are related, but we admit that the related sentences should be rephrased to be more precise.

Referee 3: "Lines 569-570: Why is "1,700 cal. yr BP" written here in brackets? While the high productivity together with the high precipitation during EZ3 is apparent, there is nothing particular at ca. 1,700 cal. yr BP."
Authors: There are peaks in ADA and CRS, but we agree that this remains unclear and should be explained.

Referee 3: "Lines 571-572: This sentence is not clear. Please, explain."

Authors: The sentence shall be rephrased, we admit it is confusing.

Referee 3: "Line 590: Replace with "between 2,000 and 300 cal. yr BP"."

Authors: The two peaks are at ca. 2,000 and 300 yr BP (so writing "between" would be imprecise), but we will rephrase the sentence to make that clear, it can obviously lead to misunderstandings the way it is written.

Referee 3: "Line 593: For consistency with van Helmond et al. (2017), please use Medieval Climate Anomaly (MCA)."

Authors: While it is reasonable to be consistent with Helmond et al., we prefer MWP since the term "anomaly" is rather unfortunate in the opinion of several of our co-authors.

Referee 3: "Line 599: This sentence is long. Suggestion: "... ostracods. As the assemblage ..."."

Authors: We agree and will rephrase the sentence.

Referee 3: "Line 611: Why not mentioning the pollen-based transfer function as organic temperature proxy?"

Authors: We agree.

Referee 3: "Line 618: "... are feasible ..." sounds strange, wouldn't e.g. "... were obtained ..." be better?"

Authors: We will rephrase this sentence.

Referee 3: "Line 622: Change "comprising" to e.g. "representing"."

Authors: It would probably even be better to change it to "from the interval between", since these two samples neither comprise nor can represent 100 yrs.

Referee 3: "Line 623: Replace "... fits well with ..." with "... are close to ...". "

Authors: Will be done.

Referee 3: "Lines 624-626: If possible, a calibration based on lake sediments (Rampen et al., 2014) should be used here. The strong temperature increase (10 ◦C) at the transition between EZ1 and EZ2 may be an artefact due to the different calibrations as discussed in lines 659-661."

Authors: As mentioned above, we now applied the lake specific calibration by Rampen et al. (2014) to sediments deposited in the freshwater environment of EZ1 to reconstruct surface water temperatures of the putative Ancylus Lake phase.

Referee 3: "Lines 626-636: Some of these records (marine-based and pollen-based) should be shown here for comparison, especially considering the high heterogeneity in the different temperature proxies from Site M0059. A reference to Krossa et al. (2017) alkenone-based records form the Skagerrak is missing."

Authors: We agree only to some degree here – for example, the pollen-based records including temperature data are quite far away and from a terrestrial archive. Furthermore, we want to avoid too many or too complex figure (see above/below). However, a few other temperature records could be added. Compare our answers to the comments to Fig. 4 – maybe a good compromise would be to add another supplementary figure, which could also comprise, e.g., the Greenland ice record and other temperature records.
The reference will be added.

Referee 3: "Lines 632-633: Because of the extremely low and inconstant sample resolution of the clumped isotope record, no trend can be seen. Remove this part of the sentence. For the same reason, I would further suggest to remove the lines between the dots in Fig. 4. Such an extrapolation is not realistic."

Authors: We agree that this should be rephrased and we will used dashed lines in Fig. 4.

Referee 3: "Line 641: Change MWP into MCA."

Authors: We prefer MWP (e.g. because the term "anomaly" may be misleading), but we will see that all texts (also figure captions, see below) use the same terms consistently.

Referee 3: "Lines 655-657: This sentence is not necessary. It could be removed.
Lines 661-663: This sentence is not necessary. It could be removed."

Authors: Here, we do not agree. There are only few studies that have used the LDI in paleotemperature studies yet and the study in referred to in the sentence demonstrates that the LDI reflects summer SSTs in other oceanographic regions which is in agreement with our observation. We, therefore, would prefer to leave the sentence as is.

Referee 3: "Lines 667-668: However, the oldest three samples of the TEXL86 record are missing, what makes a true comparison difficult. I suggest removing "...absolute temperatures based on the TEXL86 lipid paleothermometer and ...". "

Authors: To increase comprehensibility we will shortened the sentence and refer to the TEX_L^{86}-

reconstructed temperatures in the following sentence.

Referee 3: "Lines 667-668: If not done, add a line break here. The text is much too dense."

Authors: See reply to previous comment.

Referee 3: "Lines 672-674: Remove this sentence. If a summer calibration (Kabel et al., 2012) is used, than the reconstructed SST should be close to summer SST."

Authors: We will now state that the calibration of Kabel et al. reflects a late summer to autumn temperature in the Baltic Sea.

Referee 3: "Lines 674-676: No, the TEXL86 record is NOT "... to some degree similar to the clumped isotope record". The temperatures are equally high in the HTM and the modern in the clumped isotope record, but not in the TEXL86 record. The absolute values are different as well. Remove this part of the sentence. And change the end of the sentence in " ... as well as the temperature records based on pollen and Mg/Ca ratios of benthic foraminifera.""

Authors: This is in accordance with earlier comments of Referees 1 and 3. We will rephrase the sentence following these suggestions.

Referee 3: "Lines 685-688: Please, plot the TEXL86 temperature data of Kabel et al. together with Site M0059 on Fig. 4. I suspect that these records are that similar concerning both the temperature amplitudes and the trends. Remove "the" before "TEXL86"."

Authors: As mentioned previously, both settings are highly different with the shallow Little Belt site and the deep Gotland Basin. Temperatures are obviously different between both sites, which may be attributed e.g. to a deeper production depth in the Gotland Basin. However, trends between both sites are similar with comparatively high water temperatures during the MCA and MHP.
"the" will be removed.

Referee 3: Lines 713-714: But this concerns only a very little aspect/part of the records ... This is not really convincing. Same comment for the Abstract.

Authors: Here, we do not agree with Referee 3. Significant changes in the lithology and depositional environment occur at the transition from EZ1 to EZ2 and our aim was to document how these changes affect and/or impact the different salinity and temperature proxies used in this study. Although the transition comprises only a small part of the sediment profile it is a highly significant aspect of the presented research.

Referee 3: "Line 718: NO ! This temperature increase is very probably an artefact."

Authors: Considering that we now have also TEX^{L}_{86} datapoints for this interval and that the lake-specific calibration by Rampen et al. (2014) was applied (see above) for the LDI-based temperature reconstructions, this sentence will be rewritten. Using the lake-specific calibration, the reconstructions for the four lowermost datapoints (EZ1) imply temperatures which are ca. 4 °C higher compared to the calibration for marine conditions. This means that there is still a rapid increase between EZ1 and EZ2, but not as rapid as before.

Referee 3: Line 726: But no quantitative record is shown in this study...

Authors: This might be a misunderstanding: Are the temperature reconstructions not qualitative?

Referee 3: "Lines 727-730: These results are based on a figure from the supplements..."

Authors: We are not sure if this is a problem: We assume that our approach to discuss this aspect but to put the related figure in the supplement is a good compromise to save space.

Referee 3: "Figures:
Fig. 2: Please change MWP to MCA and HCO to HTM (for consistency with the text) and explain the acronyms (LIA as well)."

Authors: We have discussed this, most of us prefer "MWP" over "MCA", but we agree that there should be consistency to the text and will thus change the text accordingly.

Referee 3: "Fig. 3: For consistency with the text, please rename "Diatom abs. Abundance" into "Abs. Diatom Abundance (ADA)". And add "(CRS)" after "Chaetoceros resting spores"."

Authors: We will do so.

Referee 3: "Fig. 4: Why no GDGT-based data exist for the three deepest/oldest samples although LDI-based data are present? Please change MWP to MCA and HCO to HTM (for consistency with the text) and explain the acronyms (LIA as well). Plot the TEXL86 temperature data of Kabel et al. together with Site M0059. The scale of the BIT index is not readable. The BIT index should be removed and should be plotted correctly (e.g. with a break in the Y axis) with the TEXL86 temperature record as supplementary figure. Add 2 previously published temperature records from the region (1 pollen-based and 1 marine-based record). The text in the topmost part should be turned over."

Authors: Three additional samples have been measured and will be added the mansucript and the related sections of the text. We will optimize Figure 4 according to several of Referee 3's suggestions and reorganize the figure text. However, in order to avoid to make the figure even more complex, we suggest that we use a supplementary figure to show the Kabel et al data vs. our data (the BIT index can be moved to this figure, as suggested). Concerning pollen-based data, we are not aware that the modern analogues technique has yet been used on a record close to Site M0059. One reasonable approach would be to apply the MAT to the record from Lake Belau, but this should rather be done in the framework of a separate publication which also features a higher resolution for Site M0059. Similarly, a comparison with pollen-based temperature reconstruction based on other methods and from sites which are far away from Site M0059 might be too complex for the scope of our MS, though we agree that this would be a good thing to do. Adding a pollen-based temperature record from another site, even if it is not directly adjacent, to a supplementary figure may be a good compromise. In order show a "validation" of the pollen-based data, we will also add a curve to the suggested supplementary figure which shows the similarity of the used pollen analogues to the fossil samples.

Referee 3: Supplementary information:
Where are the captions for Tables S1 to S4?

Authors: Captions will be added.