Response to RC2

This study presents the past evolution of marine heatwaves in the North Sea, and in particular the German bight. It first compares different observation-based gridded products and different model-based simulations, including regional and global reanalyses. The different products show reasonable agreements, though the reanalyses and satellite-driven are more in line with each-other. All datasets show a significant warming trend everywhere in the North Sea, and a trend towards an increase frequency of marine heatwaves, most of them are significant. The trends disappear with detrending, and natural variability acted to reduce slightly the variability in the last decade. The study then picks one model, for which the authors have access to full output and can re-run, although not the best amongst the other datasets, still represents MHW and trends relatively correctly. With this (global) model, they take a look at the different drivers of MHWs. They conclude that MHWs in the German Bight require several conditions to happen: 1. Some pre-conditioning in the previous season, and 2.a. Atmospheric drivers in the summer acting to enhance solar radiation and reduce latent heat cooling, or 2.b. oceanic advection driven by atmospheric low-pressure systems in winter months. They also examine the atmospheric conditions for advective MHW to happen. Finally, they show that anomalies in the North Atlantic don't advect into the North Sea, and therefore anomalies in the southern North Sea are mainly linked with atmospheric conditions. They also show that longer term variability of temperature in the German Bight is a subtle equilibrium between surface heat fluxes and ocean advection.

The article shows new Insights into the drivers and tendencies of MHWs in the North Sea, and in particular highlights new process-understanding of cold-season MHWs in the southern North Sea, which are linked with advection of warmer water from the English Channel into the North Sea, which, when the background anomalies are already high, can lead to MHW triggering. They also show that weather-regimes alone are a necessary but not sufficient condition to develop marine heatwaves.

I enjoyed reading this article and I gained new insight into these processes influencing different timescales of surface temperature variability in the North Sea. The article is very long, but it is well separated into sizeable chunks to read. The methods are robust and statistical methods are well used to assess significance. The figures are of excellent quality.

Thank you very much for your encouraging feedback and we are pleased that you enjoyed reading the article. Please find our response to your comments below. Line numbers in our response refer to the revised manuscript.

I have a few comments which I would like the authors to address before I recommend the manuscript for publication:

- MHW intensity – you take the maximum MHW intensity in your article, showing no significant trend. However, maximum MHW intensity i would the conclusion still be true with mean MHW intensity?

We selected the maximum intensity here, because we are interested in the most extreme temperature anomalies that can be reached in the North Sea (the upper end of the temperature distribution). From an ecosystem perspective, all measures of intensity can be relevant, but the maximum intensity is of particular importance for impacts related to the thermal limit of marine species.

To test whether our results depend on the chosen measure of intensity, we have repeated the analysis with the mean intensity, but do not find significantly different results. The trends in maximum and mean intensity have an overall very similar pattern in all datasets. None of the datasets show a significant increase, but most of them a significant decrease in the central and south-eastern North Sea.

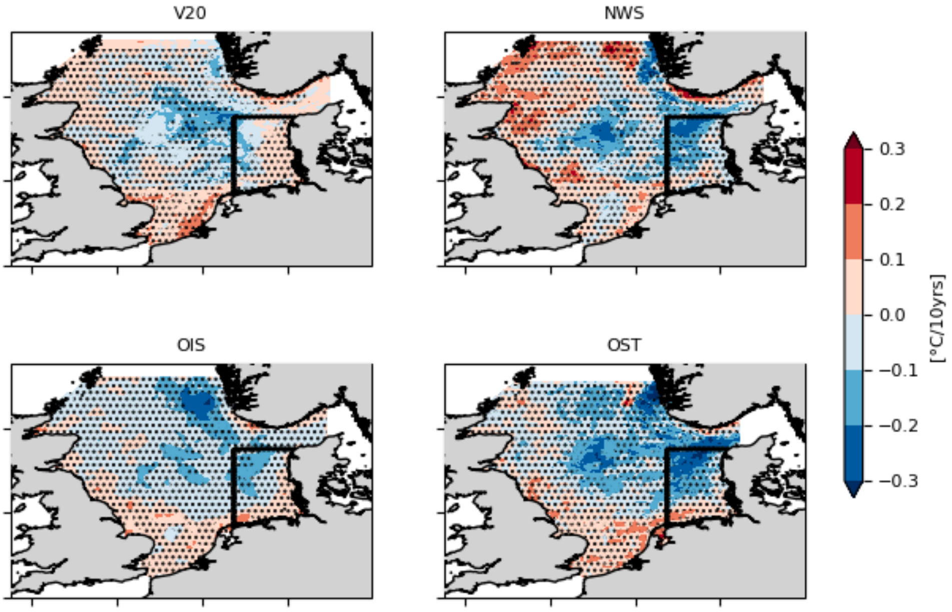

Figure: Linear trends (1993-2023) in mean intensity in VIKING20X, the European Northwest Shelf reanalysis, OISST and OSTIA. Non-significant slopes based on a significance level of 5% are indicated by dots.

- Your result that OIS has a larger SST trend than OIS or any other dataset in the North Sea is important, in particular as it’s the dataset used in Mahamed et al. 2025 – please highlight it in the conclusion.

We have added the following sentences to the discussion section to highlight this result: “This is an important result, because the OIS dataset is used for example by *Mohamed et al. (2025)* as the only dataset to assess past trends in North Sea MHWs. Although we don't expect their results to fundamentally change with other datasets, their trend estimates are likely at the upper limit.“ (line 777)

- Please do a general re-read, as there are a few typos here and there, I highlight a few below in the minor comments

Thanks for pointing us to these mistakes. We made our best to check the entire manuscript again.

- Regarding MHW drivers (Figure 4): I’m surprised LW cooling suppression (e.g. by low-level cloud) doesn’t come up as a driver in the autumn/winter – can the authors comment on this? (LW+SE+LA only significant in one case in the detrended baseline)

The reason for this result is that the LW heat flux contributes to most events in winter, but the contribution is often below 30%, which is our threshold to separate the “main drivers”. The mean contribution to the development of surface heat flux driven MHWs in the cold season (December to April) is around 20%. For advective MHWs the LW contribution is usually positive, too. Therefore, changes in the longwave heat flux reinforce advective driven MHWs. As noted in the manuscript the choice of using 30% to separate the main drivers is a subjective choice and lowering it to 20% for example would cause LW to appear more often (especially in winter). Nevertheless, the LW contribution to most MHWs is still smaller than the contribution of the other heat flux components (at least in VIKING20X).

- I understand that the authors use a common period to compare all the datasets, but their conclusions Line 739 may not hold if taking the full 1982-2025 dataset – please comment on this – does the trend emerge in OST when the full available period taken?

The trends are indeed significantly different from zero for MHW days, duration and frequency when the full satellite period (1982-2023) is considered in datasets that cover this period (V20, OST, OIS, DMI). Maximum (and also mean) intensity trends are still not significantly different from zero in any of the datasets and mostly negative. We fully agree that this point should be mentioned and was added to the discussion (line 771).

- On the variability on seasonal timescales (6.1), I am surprised by the major influence of Ti on mean season temperature anomaly in MJJJ found in Figure 8, but this does not translate into presence of at least 1 MHW/Ti conditional probability in Fig. 7d, despite a strong MHW/mean seasonal temperature conditional probability. Can the authors explore this further?

The p-value (0.06) is just barely above the significance threshold. The probability p_{11} is slightly lower than in winter and early fall, but actually higher than in late fall. p_{11} being smaller than one shows that years occurred in which T_i was anomalously high, but no

MHWs have occurred (applies to 3 years). The main reason for the non-significant p-value is the combination with a slightly lower p_{00}. p_{00} being smaller shows that in summer MHWs can occur more often even with an average (or below average) initial temperature (applies to 6 years). By itself this does not contradict a very high dependency of the season mean temperature anomaly on the initial temperature anomaly.

However, the dependency of MHWs on the season mean temperature anomaly itself is significant (figure 7a). Since the correlation between season mean temperature anomaly and T_i is high, but not equal to one, there are still years in which the season mean temperature anomaly is high, despite the initial temperature being neutral or low. The fact that the conditional probabilities are significant for the season mean temperature (figure 7a), but not for the initial temperature (figure 7d) reflects that in these particular years MHWs have occurred (applies to 4 years). This is reasonable, because years that had a below average initial temperature anomaly, but still a higher season mean temperature anomaly were likely years with exceptionally strong surface heat flux forcing that also triggered MHWs.

Overall, the data does not allow us to conclude with high certainty that MHW occurrence and initial temperature are dependent in MJJ, but the opposite cannot be concluded either. The results do not suggest they are unrelated. It is also likely that with a longer timeseries, the dependency would be statistically significant. Furthermore, the fact that the conditional probabilities are significant for the mean, but not initial temperature anomaly is physically reasonable. We have added the upper explanation to the section about conditional probabilities (line 646).

- I would invite the authors to be clearer in section titles and in figure captions about which graph relates to the whole of the North Sea, and which to the German Bight only. I kept being confused between the two regions. I think only section 4 and figure 1-3 relates to the whole North Sea, but the rest of the article is just for the German Bight, please make it clearer

We fully agree that it was not always obvious which section referred to the German Bight or the entire North Sea. Indeed, only sections 4.1 and 4.2 (figure 1 & 2) show results for the entire North Sea to put changes in the German Bight into context. Later sections consider only the German Bight. We have added this information in the section titles and captions (we did not repeat it in every subsection title, but only added the information to the main title of the sections).

The section titles now read:
“4 Past evolution of MHWs in the North Sea”
“5 Atmospheric and oceanic conditions during MHWs in the German Bight”
“6 Ocean preconditioning in the German Bight”
“7 Local and remote forcing of German Bight MHWs”

- Regarding MHW drivers, the fact that you are detrending or not seems to have a large impact on results, in particular the detrending brings new MHWs in the 1980s associated with heat advection. Your event set shown in Figure 4 c and f seems to be almost entirely different for every year depending on detrending or not, and seems to have more MHWs in the detrended baseline than the fixed one. This makes me think that you removed an all-year trend, rather than a month-by-month or season-by-season trend. The trend is usually larger in summer than winter, so if you removed a yearly-averaged trend, then more winter events will emerge in the past by construction, but this is incorrect (Smith et al. 2025).

This is a good point and we have subtracted the trend calculated from the entire timeseries. However, this seems to be the default and is used in any publication we are aware of. Even *Smith et al*. (2025) only mention that this could lead to problems, such as the one you describe, but their (and that of the publications they cite) definition of “detrended baseline” does not incorporate a seasonal detrending to our understanding. A paper they cite in this context is *Wang et al. (2022)* who remove the “change in seasonal cycle” by using the mean of a reference period, together with the seasonal cycle of a future period. Because they analyse climate scenarios, the “analysis” and “reference” periods are not the same in their study. For us this is not possible, because we only have a single 30-year period. They do not describe a suitable method to remove changes in the seasonal cycle within a single 30-year period.

Nevertheless, we still think your point is absolutely right and should be considered. We have performed a detrending that takes seasonal variations of the trend into account. The only suitable method seems to be calculating a daily trend over 30-years following a similar method that is also used to define the MHW climatology (using an 11-day window for smoothing daily trends). This trend is then removed for each day of the timeseries. Other options we can think of introduce jumps from one month/season to the next and could cause MHWs to be discontinuous across these time windows.

The resulting trend for each day of the year is shown in the figure below. Interestingly, the trend is generally larger in the cold season than in the warm season, with a short peak in June and two minima in April and August. Since the trend does not show a clear seasonal cycle, but rather several maxima and minima, changes in the trend mostly average out when a period of about 2 month or longer is considered (trends for the seasons as defined in figure 2 for example are almost identical). Therefore, the effect on the detected MHWs is also very small. About 90% of the MHWs days are the same with both methods. Furthermore, because the daily trends are slightly stronger in the cold season, MHWs that are only detected with daily trends removed occur predominantly in winter and MHWs that only occur with the full timeseries trend removed occur mostly in summer. As seen in the figure below the effect of removing daily trends is to slightly amplify the seasonal cycle in MHW days. Therefore, the higher number of winter MHWs with the detrended baseline is not caused by disregarding the seasonal variations of the trend, but even more winter MHWs are detected when doing so. In any case the timing and drivers of MHWs in the German Bight in a particular year are very different between the detrended and fixed baselines, independent of how exactly the detrending is performed.

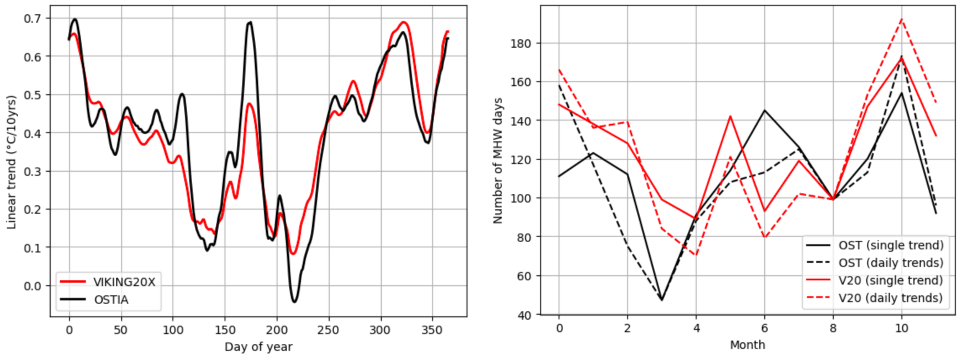

Figure: Linear trend (regression slope) of the German Bight temperature dependent on the day of year (1993-2023) in the VIKING20X model and OSTIA satellite-based dataset (left). Number of MHW days per month when a single trend or daily trends are removed.

Especially because we do not find a clear seasonal dependency of the trend, we think the best option is to rely on the well-established method to remove a single trend from the timeseries. We still address this issue in the method section when introducing the baseline definitions (line 234).

Minor comments:

- European Northwest Shelf -> Northwest European shelf

We have changed this when using the term as a geographical reference. However, the reanalysis product is listed under the name “European North West Shelf – Ocean reanalysis” on the Copernicus website. Therefore, we didn’t change it when referring to the dataset.

- Abstract: line 12: replace “Instead” by “In this region,”

Changed.

- Line 39: Hu et al. 2021 can be complemented by Okajima et al. (2025), who showed wet bulb globe temperature enhanced by a MHW in Japan, leading to human discomfort.

Thanks for pointing us to this interesting publication. We have cited it in the respective line.

- Lines 111-115: please add that VIKING is a NEMO-based ocean configuration

Of course, this should have been mentioned. Thank you for noting it. We have added the model details to the respective section.

- Line 158: It it based -> it is based

Changed.

- Please check GLORYS12 reference: should be Lellouche et al (first names/last names mixed up in the reference)

Changed.

- Figure 1: please use the same y-axis scale between 1a and 1b.

Changed.

- Lines 380-381: the overestimation of MHW days by models & reanalyses: this may be linked with the satellite datasets having no diurnal cycle, and instead representing a night-time, foundation SST (Good et al. 2020) – please comment on this in the article.

Yes, the satellite data represents the night-time temperature while in the model a diurnal cycle can exist. If the diurnal cycle grows, which one might expect during shortwave driven MHWs, this could lead to an increased difference between model and satellite-based SST. The model SST would then grow faster and higher than the foundation SST. Nevertheless, none of the models have a vertical resolution that allows a skin-layer to develop and especially VIKING20X (6 m surface layer) only represent a very smoothed diurnal cycle. Since we do not have hourly output from the model it is hard to quantify the exact impact of that difference, but it is possible that it contributes to the difference between models and observations. We have added the following statement to the respective section:

"This results in 3-4 more MHW days per year for both baselines. In part, this difference could be explained by the satellite datasets representing a foundation temperature (without a diurnal cycle; *Good et al. 2020*), while the model-based datasets usually simulate the diurnal cycle. This will cause the temperature to rise stronger in the model datasets because one would expect an amplification of the diurnal cycle during surface forced MHWs (at least for shortwave driven heatwaves). Nevertheless, the diurnal cycle is strongly smoothed, because of the limited vertical model resolution." (line 397)

- Line 404: define Southern Bight

We refer to the region located near the entrance of the English Channel as the Southern Bight. This was added to the text.

- Line 404: and frequency, have increased -> and frequency have increased**,**

Changed.

- Line 440: less -> fewer

Changed.

- Figure 2: the dots overimposed on the temperature trend maps make the colour scale hard to read – I suggest to use fewer colour levels (e.g. 0, 0.1, 0.2, 0.3, 0.4), so that it's easier to read the figure

RC1 suggested to use the stippling to indicate non-significant values, which helps to better read the color scale in the important (significant) regions. We have also increased the contour interval for the temperature trend and also MHW days and intensity.

- Line 44: less -> fewer

Changed.

- Line 445: the difference between minimum and maximum -> missing which quantity you are talking about: SST?

It referred to MHW days, stating that the seasonal cycle is more pronounced in the satellite datasets. We have changed it to:

"The seasonal cycle in MHW days exhibits a larger amplitude in the satellite datasets compared to the reanalysis datasets and especially compared to the VIKING20X model." (line 467)

- Paragraph Line 460: is the decrease trend in standard deviation statistically significant? It looks more like interannual variability

The sentence was meant to state that for standard deviation and mean the changes are identical when comparing VIKING20X and OSTIA, not that the change in mean and the change in standard deviation themselves are similar. This may have been unclear. The mean shows a clear (and significant) trend. As you suggested, the standard deviation does not show a significant trend over the period shown in figure 2, but variability on shorter timescales dominates. The sentence explaining the comparison between VIKING20X and OSTIA now reads:

"The temporal evolution of the mean is similar between the two datasets, and the same holds for the standard deviation." (line 482)

- Line 544: These result**s**

Changed.

- Line 540: strong input of heat at the ocean surface -> please rephrase, as LA-driven MHWs are rather a lack of cooling rather than a warming (so there is no “input of heat”)

This is true, the latent heat flux is cooling the ocean in the mean and during a MHW there is a latent heat flux anomaly that warms the ocean by less evaporative cooling. The sentence now reads:
“This is again consistent with relatively weak winds and a strong heat flux anomaly at the ocean's surface (in this case due to reduced evaporative cooling).” (line 564)

- Line 569: Northeast of Scotland -> look North**west** of Scotland to me (upper **left** boundary of the map)

Sorry for the confusion, you are right and we have changed it.

- Line 637: various different weather patterns -> various weather patterns

Changed.

- Lines 650: 1988/89 -> 87/88 if I interpret the figure correctly

Yes, you are right. It should have been 1987/88 and we have changed the years.

- Line 651: 1987 -> 1988

Changed.

- Figure A2: c, f and i have a negative depth scale – please correct to be consistent with other panels

Thank you for spotting this mistake. We have adjusted the depth labels.

References:

Okajima, S., Kosaka, Y., Miyasaka, T., & Ito, R. (2025). Unprecedented marine heatwave significantly exacerbated the record-breaking 2023 East Asian summer heatwave. AGU Advances, 6, e2025AV001673. https://doi.org/10.1029/2025AV001673

Smith, K. E., sen Gupta, A., Amaya, D., Benthuysen, J. A., Burrows, M. T., Capotondi, A., Filbee-Dexter, K., Frölicher, T. L., Hobday, A. J., Holbrook, N. J., Malan, N., Moore, P. J., Oliver, E. C. J., Richaud, B., Salcedo-Castro, J., Smale, D. A., Thomsen, M., & Wernberg, T. (2025). Baseline matters: Challenges and implications of different marine heatwave baselines. *Progress in Oceanography*, *231*, 103404. https://doi.org/https://doi.org/10.1016/j.pocean.2024.103404