REVIEW 3

We thank the reviewer for their thoughtful comments. In the responses below, the reviewer comments are in bold and italic, and our responses are in normal font.

This study presents a comprehensive analysis of the leading dynamical processes of global MHWs during the onset and decline periods, based on ECCO and OISST datasets. While the paper is generally clear and the results could contribute to our understanding of MHW evolution, significant shortcomings preclude recommending it for publication in its current form.

1. ***Firstly, the authors used daily and monthly datasets to calculate MHWs after removing linear trends. The spatial patterns of the MHW metrics differ significantly between the OISST and ECCO datasets (see Figures 3 and 4). This suggests that ECCO does not capture MHW characteristics in the same way as observations. Therefore, I suggest that the authors use more ocean reanalysis datasets, such as GLORYS or BRAN.***

Thanks for your comment. We acknowledge the differences between ECCO and OISST in MHW metrics, as shown in Figures 3–4. Some of these differences may be related to ECCO's resolution: Pilo et al. (2019) show that while model configurations with resolutions ranging from 1° to 1/10° degree, can all qualitatively represent broad-scale global patterns of MHWs, modeled MHWs are weaker, longer-lasting, and less frequent than in observations, especially for models with lower resolution.

This is due in part to smoother SST time series and longer autocorrelation times (Cooper et al., 2017) in models compared to observations, which can suppress some of the short-lived variability and artificially extend the duration of events. High-resolution, eddy-permitting models perform generally better, particularly in dynamic regions like western boundary currents, but still exhibit biases (Pilo et al., 2019; please also refer to Fig. R1 here below). For example, Fig. R1 here below shows that while the higher-resolution ocean reanalysis GLORYS, one of the products suggested by the reviewer, does have a better comparison with OISST for MHW duration relative to ECCO (bottom row in Fig. R1), but exhibits an overall large positive bias in MHW frequency (top row in Fig. R1), a metric that is better captured by ECCO.

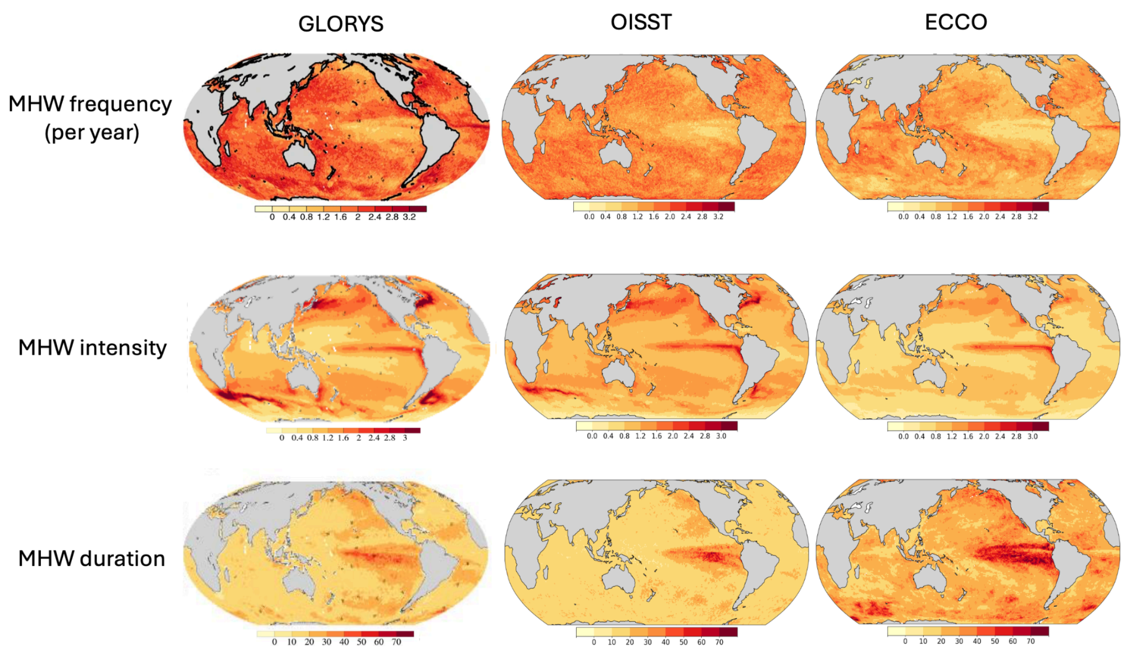


Fig. R1. MHW metrics (see labels on the left) compared across different products (see labels on the top). GLORYS panels are from Guo et al. 2024.

For these reasons, the differences we see between ECCO and OISST do not seem to diminish the value of our analysis. Despite its limitations, which are common to most modeling products, ECCO offers unique advantages for MHW research. In particular, ECCO's ability to close the heat budget exactly (Forget et al., 2015) makes it uniquely suited for budget analyses. This provides a critical perspective on MHW dynamics that observational products alone cannot offer, thereby justifying the broader scope of the manuscript. Also, ECCO compares well with observational estimates of MHW based on upper ocean heat content (Figure 6a-c based on ECCO, versus Figure 6d-f based on Argo observations).

While higher resolution data assimilating systems are available (e.g., GLORYS, BRAN), they are non-conservative and do not allow for heat budget closure, which is key to our study.

This text (bold font embedded in the text from the original submission for context) was included in Section 4.1 of the revised manuscript:

ECCO provides an overall good representation of **the spatial patterns** of both the long term linear trend in upper ocean temperature (e.g., Figure 1) and the 90th percentile anomalies (Figure S1 in the Supplement); also, spatial patterns of MHWs frequency, average duration, and average intensity in ECCO are consistent with observations (Figures 3-6). Yet, a smaller number of near surface MHW events shorter than a month (i. e. duration between 5 and 29 days) is seen in ECCO compared to observations (Figure 3a d), with only some of these events showing a signature in upper (5-55m) ocean heat content (Figure 3 g).

Some of the differences between ECCO and observations may be related to ECCO's resolution: Pilo et al. (2019) show that while model configurations with resolutions ranging from 1° to 1/10° degree, can all qualitatively represent broad-scale global patterns of MHWs, modeled MHWs tend to be weaker, longer-lasting, and less frequent than in observations, especially for models with lower resolution. High-resolution, eddy-permitting models perform generally better in representing MHW characteristics, particularly in dynamic regions like western boundary currents, but still exhibit biases (Pilo et al., 2019).

Discrepancies between models and observations are due in part to smoother SST time series and longer autocorrelation times (Cooper et al., 2017) in models (compared to observations), which can reduce short-lived variability and emphasize events of longer duration.

Capotondi, Antonietta, et al. "A global overview of marine heatwaves in a changing climate." Communications Earth & Environment 5.1 (2024): 701.

Cooper, Fenwick C. "Optimisation of an idealised primitive equation ocean model using stochastic parameterization." Ocean Modelling 113 (2017): 187-200.

Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation. Geoscientific Model Development, 8, 3071-3104, http://dx.doi.org/10.5194/gmd-8-3071-2015, http://www.geosci-model-dev.net/8/3071/2015/

Guo, X., Gao, Y., Zhang, S., Cai, W., Chen, D., Leung, L. R., ... & Wu, L. (2024). Intensification of future subsurface marine heatwaves in an eddy-resolving model. Nature Communications, 15(1), 10777.

Pilo, Gabriela S., et al. "Sensitivity of marine heatwave metrics to ocean model resolution." Geophysical Research Letters 46.24 (2019): 14604-14612.

2. ***Secondly, in the section discussing the mechanism of global MHWs, the authors conducted a heat budget analysis, but did not present the residual term. Does this mean that ECCO can perform a close heat budget analysis?***

We appreciate the reviewer's comment and the opportunity to clarify this important point.

Although not visible (as orders of magnitude smaller), the residual term is displayed in Figures 11-13. Figure R2 below also shows the residual term corresponding to Figure 7 in the manuscript (which is orders of magnitude smaller than budget terms).

Finally, we included this sentence in Section 3.2 (ECCO ocean heat budget):

"We note that ECCO ocean heat budget closes practically exactly, as residuals are orders of magnitude smaller than the other budget terms."

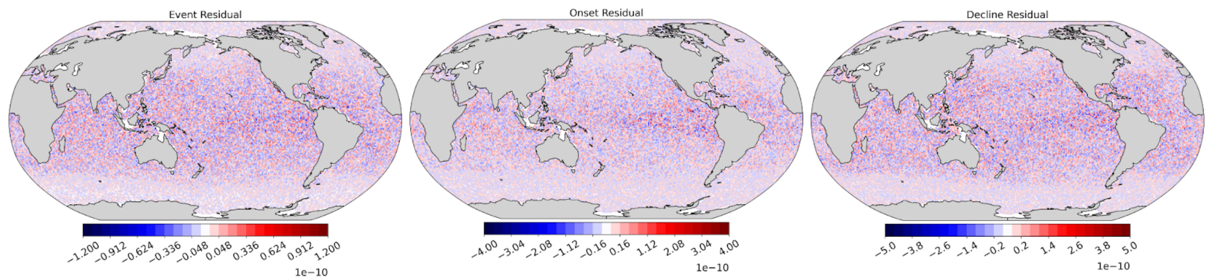


Fig. R2. Maps representing the heat budget residual for the events average (left), the onset phase (middle) and the decline phase (right), corresponding to Figure 7 in the manuscript.

3. ***Finally, I agree with RC1 that the identification of leading terms is based on predefined thresholds (30%, 16%) that are not explained. The authors should demonstrate this more clearly.***

We appreciate the reviewer's comment regarding the use of predefined thresholds. In the revised manuscript, we implemented a simplified criterion for the identification of leading terms (described later in this response): while the selected threshold is somewhat arbitrary, it helps to summarize our results. Also, we clarify that our results are robust to ∓5% changes in the selected threshold (see figures at the end of this response).

While Figures 9, 11-13 (shown at the end of this response) were updated to reflect the new (simplified) method, results are consistent with the previous version of the plots. We also now show a range for the values in Table 1, to indicate how the percentages change with a threshold value of 25% and 35% instead of 30% (see table at the end of this message). As discussed for the overall classification of leading terms above, the purpose of the table is to summarize which processes are most often leading terms, and the focus is not the exact percentage value reported.

In the revised manuscript we included this text (in the methods section) to introduce the simplified criteria for the identification of leading terms:

"**Towards summarizing our global and regional findings for the leading dynamical processes driving MHWs, we introduce the terms "the" leading term and "a" leading term, defined as follows. For each phase, we identify the budget terms that contribute to it (among forcing, advective convergence, diffusive convergence), then we sort them by the magnitude of the contribution. A term provides "the" leading contribution if it exceeds the next largest term by at least 30%. A term also provides the leading contribution if it is the only process contributing to the phase of interest. If the largest two contributors are both greater than the third by at least 30%, but neither is larger than the other by 30%, then each of the two terms provides "a" leading contribution. The same happens if these two terms are the only contributors and neither is larger than the other by 30%. If none of the terms provides a contribution that is 30% larger than other contributions, all three terms contribute comparably. We note that while the 30% threshold is arbitrary, it serves the purpose of summarizing our results, and visually identifying which processes are most often**

leading terms. Our findings are robust to ∓5% change in the (30%) threshold percentage used (not shown).”

Updated Table 1:

	Terms	NEP (20 events)		SWP (38 events)		TASMAN (37 events)	
		Onset	**Decline**	**Onset**	**Decline**	**Onset**	**Decline**
"the" leading term	F	40-45%	25-35%	71%	29-32%	38-49%	38-41%
	A	15-25%	10-15%	13%	18%	22%	5%
	D	10%	20%	5%	3%	0-3%	11-16%
"a" leading term	F	10-25%	20-30%	11%	37-39%	24-32%	32-35%
	A	15-25%	15%	0%	13%	14-19%	5-8%
	D	5-10%	25-35%	11%	29-32%	11-19%	32-35%

How often each budget term is "the" leading term vs "a" leading term (excluding overlap with "the") across NEP, SWP, and TASMAN regions. A range of percentage values is shown for each case, indicating how percentages change with a threshold value of 25% and 35% instead of 30%, in the definition of “the” vs “a” leading term.

Updated Figure 9:

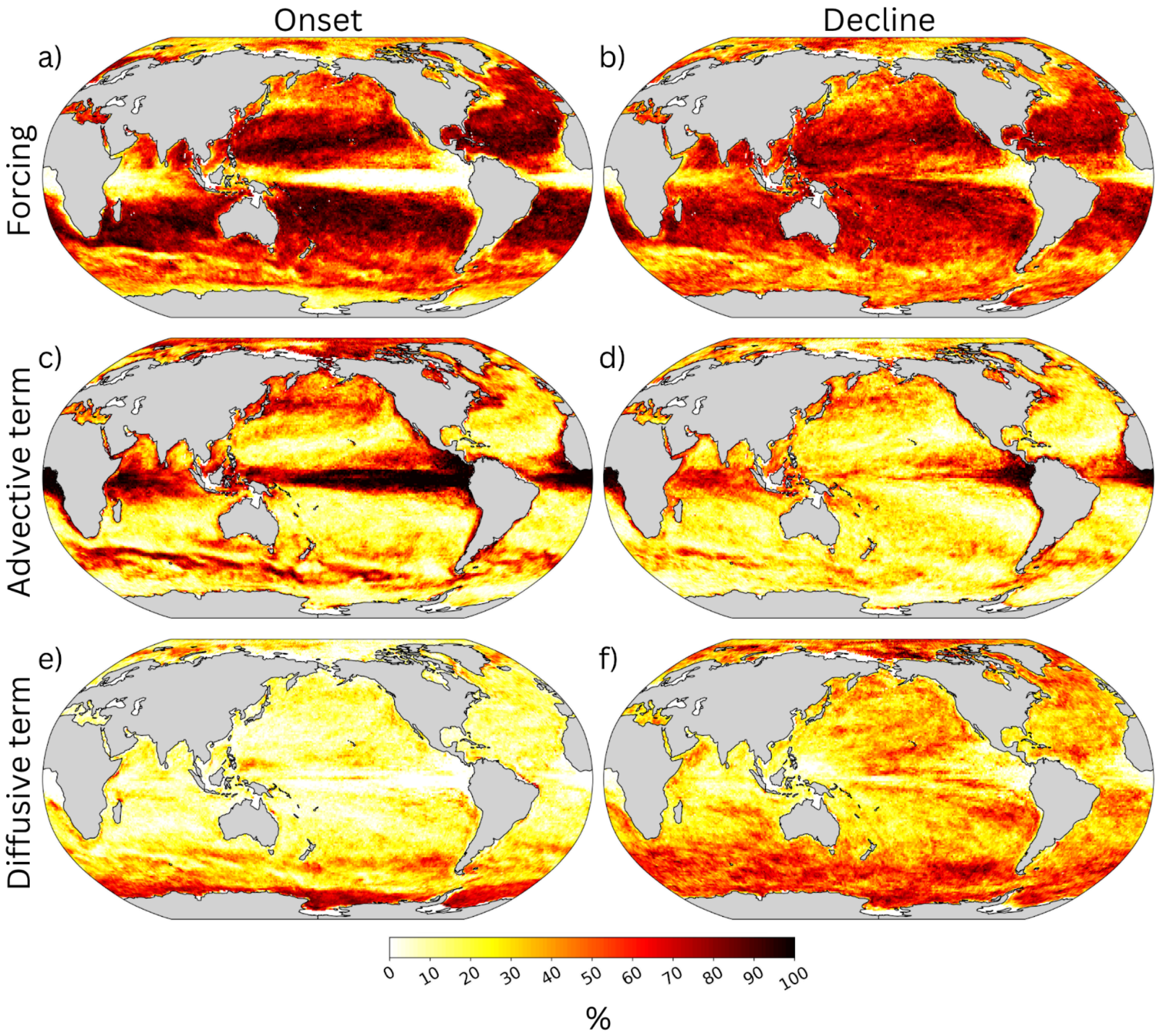
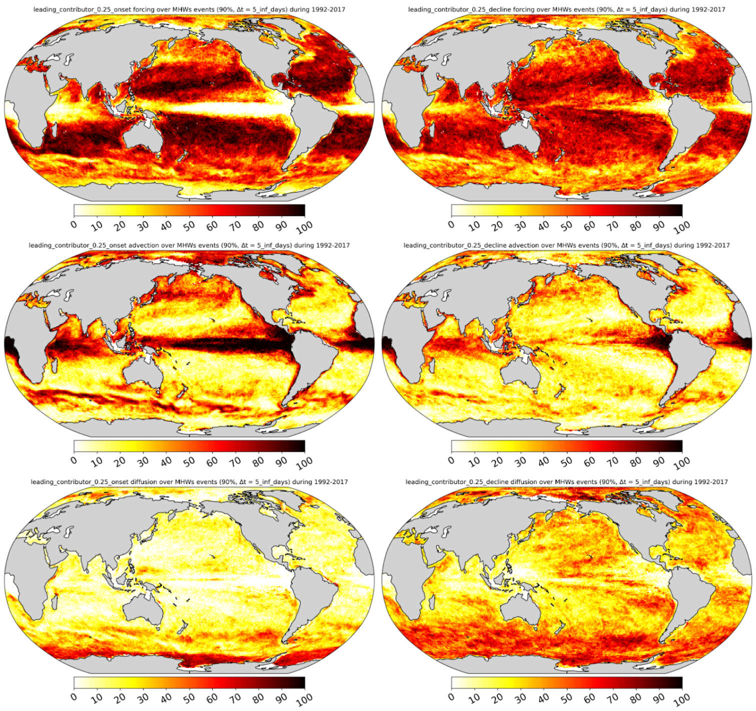


Figure 9. Same as Figure 8, now for the percentage of times each term is a leading contributor to the MHW (a, c, e) onset and (b, d, f) decline phase. As an example, we include in the count for panel (a) both a case where atmospheric forcing is the only contributor to the onset phase and a case where atmospheric forcing is a leading contributor together with advective and/or diffusive convergence of heat.

Versions of Figure 9 using 25% and 35% (instead of 30%) in the criteria to define “the” versus “a” leading term:

25%



35%

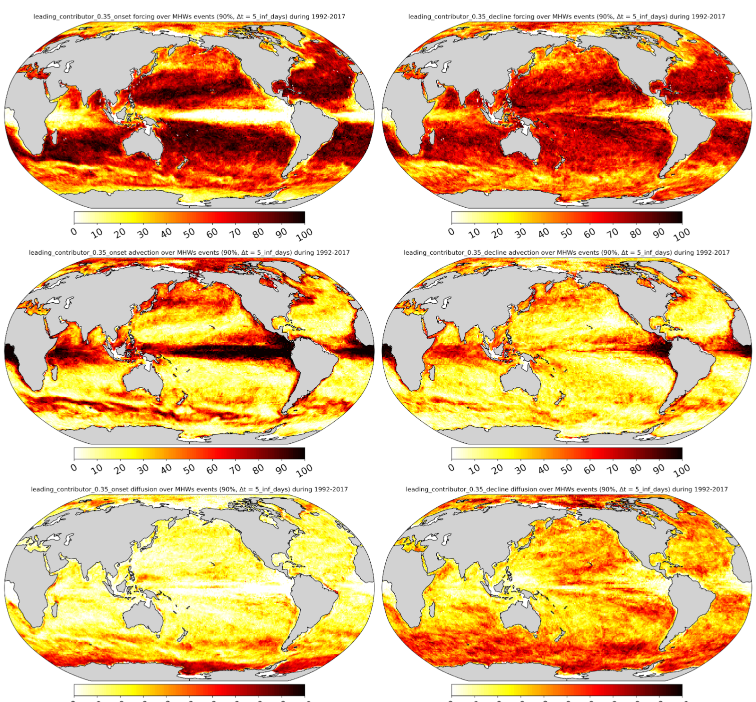
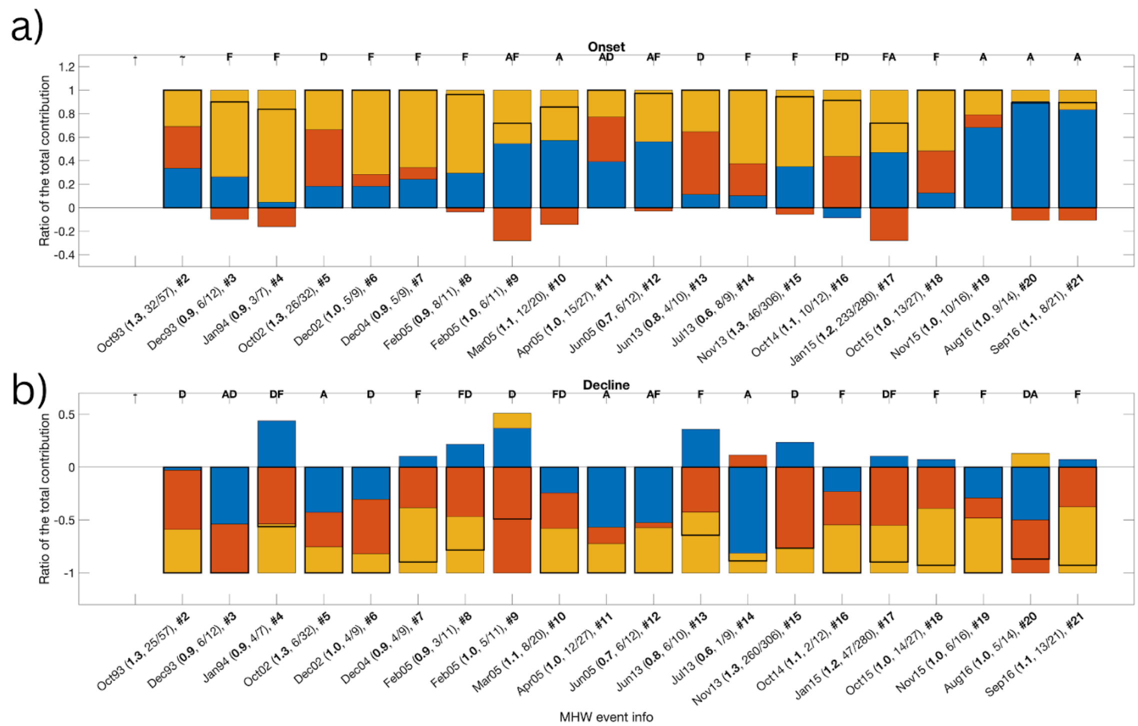


Figure 11 (NEP)



The figure above (and similarly the ones that follow, for other regions) shows the ratio of the total contribution for each budget term during each phase (onset/decline), as well as the tendency (scaled too by the total contribution). In each figure, the top panel is for the onset phase, the bottom one for the decline phase. Each stacked bar represents the relative contribution of each term during that phase, with the total contribution (i.e., the sum of the terms that contribute to that phase) normalized to 1. The black outline over each bar indicates the total temperature tendency during that phase (scaled by the total contribution to that phase), showing agreement with the sum of the individual terms and confirming budget closure. The x-axis labels denote the start date of each MHW event, followed by values in parentheses indicating the event intensity (as average temperature (degC) in the layer used for the OHC estimate), onset/decline duration (in days), and total duration. The letter codes above each bar indicate which term(s) dominated the total temperature tendency during that phase (A = advection, D = diffusion, F = forcing), e.g., "F" is used for cases when the forcing provides the leading contribution, "FA" is used when both forcing and advective convergence provide a leading contribution, "AFD" characterizes cases where advective convergence is larger than forcing and diffusive convergence and forcing is larger than diffusive convergence, yet the difference does not meet the 30% criteria as in the previous two cases. Finally, "~" corresponds to cases where all terms contribute comparably.

Figure 12 (SWP)

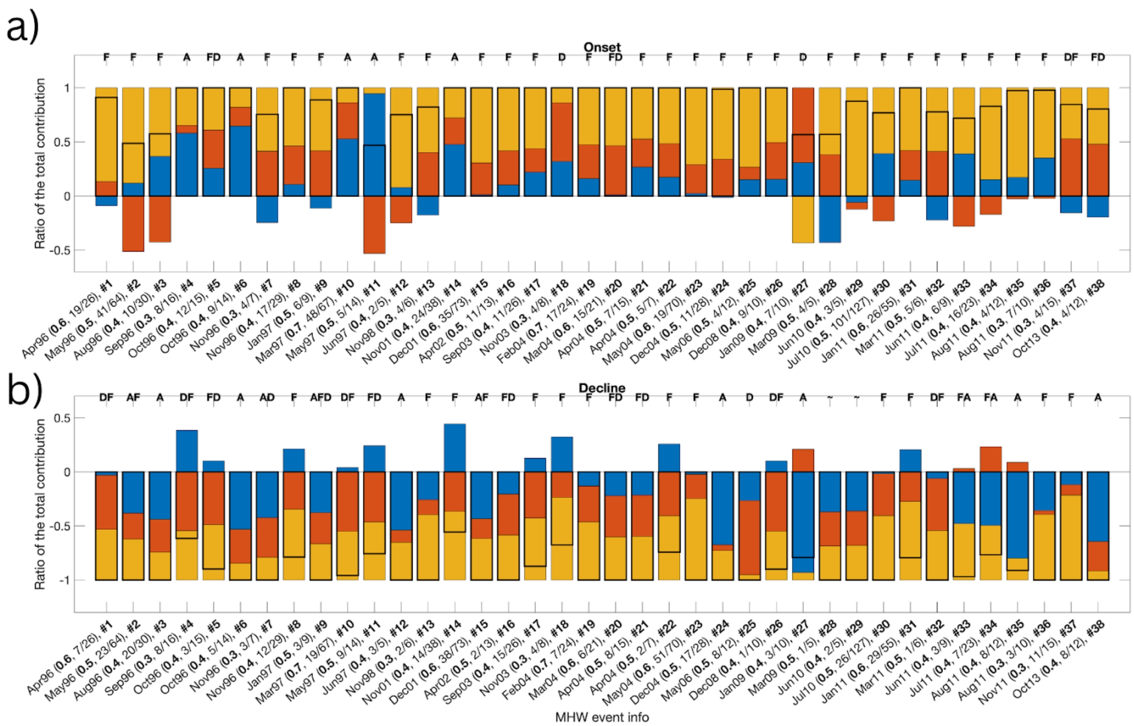


Figure 13 (TASMAN)