On the vertical structure and propagation of marine heatwaves in the Eastern Pacific

Eike Eduard Köhn¹, Meike Vogt², Matthias Münnich¹, and Nicolas Gruber¹

¹ETH Zürich ²ETH Zuerich

June 1, 2023

Abstract

Marine heatwaves (MHWs) have been recognized as a serious threat to marine life, yet, most studies so far have focused on the surface only. Here, we investigate the vertical dimension and propagation of surface MHWs in the Eastern Pacific using results from a high-resolution hindcast simulation (1979 to 2019), performed with the Regional Ocean Modeling System. We detect MHWs using a seasonally varying percentile threshold on a fixed baseline and track their vertical propagation across the upper 500 m. We find that nearly a third (29%) of the MHWs extend beyond the surface mixed layer depth (MLD). On average, these deep-reaching MHWs (dMHWs) extend to 110 m below the MLD and last five times longer than MHWs that are confined to the mixed layer (184 vs. 36 days). The dMHWs can cause stronger temperature anomalies at depth than at the surface (maximum intensity of 5.0°C vs. 1.9°C). This general subsurface MHW intensification even holds when scaling the temperatures with the respective local variability. A clustering of dMHWs reveals that 41% of them are block-like, i.e., continually remain in contact with the sea surface, 24% propagate downward, 20% propagate upward, while 15% appear at the surface multiple times. Although the water column MHW duration, intensity and severity are only moderately correlated with their corresponding surface-based MHW characteristics, dMHWs have the potential to be detected from the surface. Our study can help to augment the remote sensing-based monitoring of upper ocean exposure to MHWs.

On the vertical structure and propagation of marine heatwaves in the Eastern Pacific

1

2

3

Eike E. Köhn¹, Meike Vogt¹, Matthias Münnich¹, and Nicolas Gruber^{1,2}

 ⁴ ¹Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zürich, Switzerland
 ⁶ ²Center for Climate Systems Modeling, ETH Zurich, 8092 Zürich, Switzerland

7	Key Points:
8	• About a third of marine heatwaves (MHW) in the Eastern Pacific extend below the mixed layer 10% of MHWs extend on average deeper than 150m
9 10	 59% of deep-reaching MHWs (dMHWs) are partly invisible at the surface, as they either deepen sheel or exhibit multi-surfacing behaviour.
11 12	 As initial tests indicate that the surface-based detection of dMHWs seems feasi-
13	ble, dMHWs might be detectable using remote sensing data.

Corresponding author: E. E. Köhn, eike.koehn@usys.ethz.ch

14 Abstract

Marine heatwaves (MHWs) have been recognized as a serious threat to marine life, yet, 15 most studies so far have focused on the surface only. Here, we investigate the vertical 16 dimension and propagation of surface MHWs in the Eastern Pacific using results from 17 a high-resolution hindcast simulation (1979 to 2019), performed with the Regional Ocean 18 Modeling System. We detect MHWs using a seasonally varying percentile threshold on 19 a fixed baseline and track their vertical propagation across the upper 500 m. We find that 20 nearly a third ($\sim 29\%$) of the MHWs extend beyond the surface mixed layer depth (MLD). 21 On average, these deep-reaching MHWs (dMHWs) extend to 110 m below the MLD and 22 last five times longer than MHWs that are confined to the mixed layer (184 vs. 36 days). 23 The dMHWs can cause stronger temperature anomalies at depth than at the surface (max-24 imum intensity of 5.0 °C vs. 1.9 °C). This general subsurface MHW intensification even 25 holds when scaling the temperatures with the respective local variability. A clustering 26 of dMHWs reveals that 41 % of them are block-like, i.e., continually remain in contact 27 with the sea surface, 24% propagate downward, 20% propagate upward, while 15% ap-28 pear at the surface multiple times. Although the water column MHW duration, inten-29 sity and severity are only moderately correlated with their corresponding surface-based 30 MHW characteristics, dMHWs have the potential to be detected from the surface. Our 31 study can help to augment the remote sensing-based monitoring of upper ocean expo-32 33 sure to MHWs.

³⁴ Plain Language Summary

Periods of extremely warm water temperatures, referred to as marine heatwaves 35 (MHWs), pose a serious threat to marine life. While many studies have analyzed MHW 36 properties at the sea surface, little is known about their subsurface nature. We here use 37 a new approach to define MHWs throughout the water column and study the vertical 38 structure and propagation behaviour of MHWs in the Eastern Pacific based on a high-39 resolution numerical hindcast simulation (1979 to 2019). We find that nearly a third of 40 MHWs that are discernible at the sea surface reach below the layer of well-mixed wa-41 ters near the sea surface, the so-called mixed layer. These deep-reaching MHWs last on 42 average longer than mixed layer confined MHWs (184 vs. 36 days) and are marked by 43 larger temperature threshold exceedances (maximum intensity of 5.0 °C vs. 1.9 °C). We 44 furthermore identify four distinct vertical propagation patterns associated with the deep-45 reaching MHWs: 41% continually affect the sea surface, 24% propagate downwards, 20%46 propagate upward, while 15% appear at the surface multiple times. Overall, these anal-47 yses show that MHWs can affect the water column to a larger degree than diagnosed based 48 on the sea surface only. 49

50 1 Introduction

The observed warming of the global ocean has increased already markedly the fre-51 quency, duration, and intensity of marine heatwaves (MHWs) (Oliver et al., 2018), com-52 monly defined as prolonged periods of unusually high temperatures (IPCC SROCC, 2022). 53 MHWs affect marine organisms and ecosystems profoundly (Wernberg et al., 2013; Cav-54 ole et al., 2016; Smale et al., 2019; Sen Gupta et al., 2020), as well as the economies de-55 pending on services provided by these systems (Cheung et al., 2021). Further ocean warm-56 ing will increase the key impact metrics of MHW manifold (Frölicher et al., 2018), thus 57 increasing the threats to the future health of the oceans. It is thus not surprising that 58 the study of MHW has experienced a rapid surge in attention by the scientific commu-59 nity, policymakers, and the public (Collins et al., 2019). 60

So far, most MHW studies have focused on the surface only. This is largely a con sequence of the easy accessibility of high resolution satellite sea surface temperature (SST)
 data, while no corresponding observation-based product is available to study subsurface

extremes (Oliver et al., 2021). But even model-based studies have rarely analyzed MHWs 64 beyond the surface (e.g. Amaya et al., 2023), such that the vertical structure of surface 65 MHWs has remained largely elusive (Gruber et al., 2021). Gaining more insights into 66 the vertical extent and structure of MHWs is crucial to better assess the MHW drivers 67 and the mechanisms sustaining them (Oliver et al., 2021). A look beyond the surface is 68 also critically important to assess and understand the impacts of MHWs on upper ocean 69 ecosystems. The vertical MHW structure can be particularly relevant for animals at higher 70 trophic levels whose vertical habitats extend beyond the mixed layer (ML) or euphotic 71 zone and which often perform diurnal vertical migration (Steinberg & Landry, 2017; Bianchi 72 & Mislan, 2016). Deep extending MHWs may limit organisms in their ability to use ver-73 tical migration as a strategy to avoid the impact of extremely high surface ocean tem-74 peratures. 75

While a basin scale assessment of the depth structure of MHWs has not been con-76 ducted to date, a few studies have analyzed the vertical extent of individual events and 77 linked them to different drivers. For example, Elzahaby and Schaeffer (2019) analyzed 78 vertical temperature anomaly profiles from Argo floats during surface MHWs off East-79 ern Australia. The authors identify and link upper ocean MHWs (shallower than 150 m 80 depth) to physical driving processes at the ocean surface, while associating deep-reaching 81 MHWs (deeper than 800 m) to warm core eddies. Similarly, Scannell et al. (2020) and 82 Johnson et al. (2022) used Argo float data to investigate the vertical extent and prop-83 agation of the Northeast Pacific Blob heatwave, i.e., the strong surface MHW in the North-84 eastern Pacific that lasted for multiple years (2013-2016) Bond et al. (2015); Di Lorenzo 85 and Mantua (2016) and was likely even a compound extreme (Gruber et al., 2021). Scannell 86 et al. (2020) demonstrated that the heat anomalies from this event were lingering for years 87 at depth affected ecosystems for much longer periods than diagnosed from the surface 88 only (see also (Jackson et al., 2018; Freeland & Ross, 2019; Holser et al., 2022). Due to 89 the wide spacing (\mathcal{O} of 100 km) of the Argo float profiles and their 10 day repeat cycle, 90 these authors had to average the data considerably in time and space, thereby losing much 91 detail. The low temporal resolution was overcome in the studies by Schaeffer and Roughan 92 (2017) and Hu et al. (2021) by using high-frequency mooring time series throughout the 93 upper ocean off Eastern Australia and in the western tropical Pacific, respectively to demon-94 strate that MHWs can reach well below the ML in conjunction with downwelling winds 95 and wind-driven mixing. But the single point nature of the mooring data and the very 96 limited number of mooring sites strongly limits this approach. 97

Results from ocean models can overcome these observational limitations, but the number of studies having investigated the vertical structure and propagation of MHWs is rather small. Notable exceptions are the studies of Ryan et al. (2021) and Großelindemann et al. (2022) who used regional models to elucidate the depth structure of MHWs off Western Australia and on the Northwest Atlantic Shelf, respectively. They highlighted the important role of thermocline depth variability and (eddy driven) current anomalies in driving deep-reaching MHWs.

Extending the study of MHWs beyond the surface poses some challenges. For ex-105 ample, the classical "Eulerian" perspective on MHWs is not well suited to track any ver-106 tical propagation behaviour, as it identifies and groups adjacent days of extremely warm 107 temperatures solely in the temporal dimension (Gruber et al., 2021). A few studies have 108 therefore extended the definition of marine extremes by grouping extreme states in time 109 and space (Di Biagio et al., 2020; Desmet et al., 2022). We adopt here a so far unexplored 110 one-dimensional water column perspective to study the vertical propagation of MHWs 111 that affect the sea surface (Scannell et al., 2020). In this perspective, extremes are de-112 fined by grouping adjacent extreme states in the temporal and vertical dimension, which 113 allows extremes to have a vertical extent and to propagate vertically while staying fixed 114 at one horizontal location. This perspective permits to elucidate how deep MHWs reach 115 and for how long they affect the upper ocean water column at a particular horizontal 116

location. Furthermore, the ability to track the vertical propagation of MHWs allows us
to assess whether the vertical propagation of the extreme signal, as found by Scannell
et al. (2020) during the *Blob*, is a common behaviour of MHWs. This view can be instructive to better understand the biological impacts of MHWs, particularly regarding
resident species that are exposed to extreme temperatures across many depth levels and
species that perform diurnal vertical migration (DVM, e.g., Bianchi and Mislan (2016)).

In this study we aim to increase our understanding of the vertical structure of MHWs 123 and thus learn more about their driving mechanisms (Holbrook et al., 2019; Sen Gupta 124 125 et al., 2020; Vogt et al., 2022). The Eastern Pacific (EP) is used as a pilot region since it harboured some of the strongest MHWs in the last decades, including the Blob MHW 126 from 2013 to 2016 (Di Lorenzo & Mantua, 2016), the "Blob2.0" MHW in 2019 (Amaya 127 et al., 2020) and many MHWs associated with El Niño conditions (Holbrook et al., 2019). 128 We hence analyze the vertical structure of MHWs in a hindcast simulation of the EP. 129 First, we want to understand how deep MHWs extend into the ocean interior below the 130 ML. Our second goal is to understand how the subsurface structure of MHWs influences 131 their general characteristics, that are commonly studied from the satellite perspective 132 at the sea surface only. In this context, we aim to understand how representative surface-133 derived MHW characteristics such as duration or intensity are for MHWs that occupy 134 the water column beyond the surface ML depth (MLD). Finally, our here introduced method-135 ology to define MHWs in the water column further allows us to study the vertical prop-136 agation behaviour of MHWs across the EP basin. 137

¹³⁸ 2 Data and Methods

139

2.1 Model hindcast

We used the UCLA-ETH version of the Regional Ocean Modeling System (ROMS, 140 Marchesiello et al. (2003); Shchepetkin and McWilliams (2005)) to perform the model 141 hindcast. We employed a telescopic model grid focused on the tropical Southeast Pacific, 142 but covering the entire Pacific basin (Fig. 1a). In this grid, the resolution is finest along 143 the Peruvian coast ($\sim 4 \,\mathrm{km}$) and decreases towards the western Pacific, with the coars-144 est resolution of $\sim 40 \,\mathrm{km}$ occurring south of Australia (Fig. S1). This model setup, the 145 initialization, and much of the simulation protocols follows the work of Köhn et al. (2022). 146 The most important change is that we have improved the numerical representation of 147 advection in the model by using an isoneutral advection scheme, and by employing a $3^{\rm rd}$ 148 order WENO advection scheme in the horizontal direction and a 5th order scheme for 149 the vertical dimension (Liu et al., 1994; Shu, 1998)). We also updated the atmospheric 150 forcing from ERA interim (Dee et al., 2011) to ERA5 (Hersbach et al., 2020). For fur-151 ther details the reader is referred to Köhn et al. (2022). Here we provide a summary of 152 the simulation protocol. 153

After a 20-year model spin-up from World Ocean Atlas 2018 (Boyer et al., 2018), 154 we performed a hindcast simulation from 1979 to 2019 forced with ERA5 reanalysis data 155 (Hersbach et al., 2020), to which we applied the Drakkar Forcing Set (DFS5.2) correc-156 tion (Dussin et al., 2016) for the incoming shortwave and outgoing longwave radiation 157 (following Desmet et al. (2022)). The model spin-up was undertaken with a normal-year 158 forcing (Large & Yeager, 2004), as described in Köhn et al. (2022). Along the open bound-159 aries in the Southern Ocean, we use SODA3.4.2 reanalysis (Carton et al., 2018) as time-160 varying boundary conditions. We integrated the model from 1979 onward with a timestep 161 of 10 min and wrote out daily averages of the state variables across all depths. This yielded 162 over the 1979 to 2019 period a total of 14975 days of model output. We then regridded 163 the model data from the bathymetry following vertical coordinate (i.e., s-levels, Song and 164 Haidvogel (1994)) to 37 fixed depth levels (i.e., z-levels) in the upper 500 m, while main-165 taining the increasing resolution towards the surface (Fig. S2 in the Supplementary In-166 formation). 167



Figure 1. a) Model grid with SST averaged over full model hindcast. Grey lines show every 20th grid point of the native model grid. b) Taylor Diagram of model SST, MLD and SSH based on monthly climatologies computed over the entire hindcast and compared to monthly observational climatologies interpolated onto the ROMS grid. The data feeding into the Taylor Diagrams is limited to the focus region of the study, i.e., the Eastern Pacific east of 160°W and north of 20°S, delimited by the green lines in panel a.

To reduce the data size, we subsequently downsampled the model output by av-168 eraging the regridded data to 3x3 horizontal grid boxes. This way we reduced the out-169 put's nominal horizontal resolution, i.e., to $\sim 12 \,\mathrm{km}$ off Peru, but maintained the tele-170 scopic grid structure (Fig. 1a). We limited all following analyses to the downsampled 171 data in the upper 500 m of the water column. From the regridded and downsampled model 172 output, we calculated the MLD using the density threshold criterion of Holte et al. (2017), 173 that is a density change relative to the surface density value corresponding to a temper-174 ature change of 0.2 °C. 175

We limit the study domain to the EP region east of 160°W and north of 20°S (Fig. 176 1a) where simulated local SST biases are mostly smaller than 0.5 °C (Fig. S3a). In this 177 area monthly climatologies of the modeled temperature, MLD and sea surface height (SSH) 178 fields (calculated from 1979 to 2019) distribution (Fig. 1b) show a high spatio-temporal 179 correlation and comparable standard deviations with observed monthly climatologies (Fig. 180 1b). A more in-depth evaluation of model biases, variability and trends can be found in 181 Section 3 and in the Supplementary Information. These analyses reveal overall a high 182 fidelity of the model in capturing the mean state of the Pacific and its main pattern of 183 variability, giving us high confidence in the robustness of the results generated by this 184 model (Fig. 1b). 185

186

2.2 Detecting MHWs across the upper ocean

In order to study the vertical dimension of MHWs, we take a vertical one-dimensional perspective to identify MHW events across the upper 500 m. In this water column perspective, extreme temperatures are grouped in time and in depth to MHWs that can hence propagate vertically throughout the upper water column. We limit our analyses to those MHWs that were in contact with the surface at least once in their lifetime, and among those we mostly focus on the deep-reaching MHWs that extend below the MLD. In the following, we explain the individual steps, starting from the detection of extreme temperatures (Sec. 2.2.1) over the joining of individual extreme cells to form MHW events (Sec. 2.2.2), to the definition of the key metrics (Sec. 2.3).

2.2.1 Detecting extreme temperatures

From the daily hindcast temperature data, we calculated at each horizontal and 197 vertical grid point the seasonally varying MHW threshold temperature $T_{\rm thresh}$ based on 198 the local temperature time series following the methodology of Hobday et al. (2016) with-199 out de-trending the temperature data. To this end, we calculated the 90th percentile thresh-200 old based on the 30-year reference period 1982 to 2011. We then detected extreme tem-201 peratures (T) by requiring $T > T_{\text{thresh}}$ for each grid point and day. We obtain a four-202 dimensional Boolean array \mathbf{B} spanning across all 14975 hindcast days (daily output from 203 1979 to 2019), 37 depth levels and 335×232 downsampled horizontal grid points. 204

To discard short extreme "heat spikes", we applied morphological operations to the Boolean array **B**, consisting of a binary closing followed by a binary opening in the temporal dimension (see also Scannell et al. (2021)). Both operations were executed with a 5-day box-kernel. This temporal filtering of **B** ensures a minimum MHW duration of 5 days (congruent with previously defined minimum MHW durations, Hobday et al. (2016)) and guarantees that MHWs are at least 5 days apart from each other. Temporally isolated extreme spikes that last shorter than 5 consecutive days were thus removed (set to *False*), while extremes that are separated by less than 5 non-extreme days were merged.

213

196

2.2.2 Defining MHWs in the water column

We then grouped all extreme grid cells that are adjacent to each other in both time 214 and depth to form a coherent MHW in the water column (Fig. 2b). In order to be ad-215 jacent, extreme grid cells had to share either a corner or an edge. In this one-dimensional 216 water column detection, MHWs have a vertical extent and can grow, shrink and move 217 vertically over time. This differs fundamentally from the usual (Eulerian) perspective, 218 where extreme grid cells are only grouped in time at a given depth. We additionally iden-219 tify MHWs using this classical approach at the sea surface (surface-only detection) and 220 refer to them hereafter as "surface-only MHWs" (Fig. 2a). It is important to note that 221 in our study here, all MHWs (and surface-only MHWs) are fixed to one horizontal lo-222 cation and cannot propagate laterally, as we do not connect adjacent extreme grid cells 223 horizontally. 224

In the final step, we excluded all MHWs that never touch the sea surface, i.e., that 225 do not contain any extreme surface grid cells. This excluded about 78 % of the detected 226 one-dimensional MHWs in the upper 500 m water column (total MHW count 6 339 969) 227 and 67% of all extreme temperature grid cells. By retaining only the MHWs that can 228 be identified at the sea surface, we can relate our results to the many previous studies 229 that focused on MHWs at the sea surface (Oliver et al., 2018). We further separated the 230 retained MHWs into a subset that remains within the ML (referred to as surface MHWs, 231 sMHWs) and those MHWs that reach deeper into the ocean interior, referred to as deep-232 reaching MHWs (dMHWs, Fig. 3). In order to be classified as a dMHW, while exclud-233 ing MHWs that move only shortly and slightly beneath the MLD (see Sec. 2.1), we re-234 quired the MHWs to extend on average below a threshold defined by the instantaneous 235 MLD plus its climatological standard deviation, which is seasonally varying and calcu-236 lated over the 1982 to 2011 period (see dashed blue line in Fig. 2). These dMHWs are 237 the main focus of our study. 238

239 **2.3** MHW characteristics

To characterize the MHWs, a set of properties was calculated for each event (Tab. 1 and Fig. 2). The introduction of the vertical dimension allows for the definition of many



Figure 2. Detecting MHWs by aggregating extreme grid cells in two different ways. Panel a shows the classical surface-only detection (Eulerian perspective), while panel b) shows the one-dimensional water column detection of MHWs. Each box indicates a grid point in the time (panel a) and time-depth (panel b) dimension, i.e. the *event grid*. Colored boxes indicate extreme conditions (yellow to brown colors indicating increasing intensity), while white boxes indicate non-extreme conditions. Temperatures are expressed in a variability scaled form (\mathfrak{T} , see Sec. 2.3), with $\mathfrak{T} > 1$ indicating the presence of extreme temperatures. Panel a shows two individual surface MHWs, while panel b shows one single coherent and vertically propagating MHW. The solid and dashed blue line indicate the instantaneous MLD and the instantaneous MLD plus the climatological MLD standard deviation, respectively. The green line marks the upper boundary of the MHW. At the bottom of panel b the definition of the "vertical propagation" and the "propagation flexure" metrics are visualized, based on the upper MHW boundary, i.e., the green line. Grey text boxes contain labels and MHW characteristics as listed in Table 1. Text boxes outlined in orange highlight MHW metrics that are used in the clustering of deep-reaching MHWs.



Figure 3. Overview of MHW definitions and distinctions based on the water column detection of MHWs. In a first step, MHWs are distinguished on the basis of their attained mean depth, especially with regard to the mixed layer depth. The deep reaching MHWs (dMHWs), i.e., those extending significantly below the mixed layer, are further split up into four different clusters based on their vertical propagation behaviour.

more characteristics than in the classical surface-only perspective (e.g. Hobday et al.,

²⁴³ 2016), which can be grouped into two classes: a) additional surface characteristics and

244

b) column characteristics (Tab. 1).

MHW surface characteristics	Units	MHW column characteristics	Units
surface start day	date	column start day	date
surface end day	date	column end day	date
surface duration $({}^{s}D)$	days	column duration (^{c}D)	days
surface intensity $({}^{s}I \text{ or } {}^{s}\mathfrak{I})^{\dagger}$	°C or -	column intensity $({}^{c}I \text{ or } {}^{c}\mathfrak{I})^{\dagger}$	°C or -
surf. severity (${}^{s}S = {}^{s}I_{mean} \times {}^{s}D$)	°C×days	col. severity (${}^{c}S = {}^{c}\mathfrak{I}_{mean} \times {}^{c}D$)	days
number of interruptions	-	depth below $\operatorname{surface}/\operatorname{MLD}^{\dagger}$	m
max. interruption duration	days	vertical propagation ^{\ddagger}	${\rm m~day^{-1}}$
max. consec. days at surf.	days	propagation flexure ^{‡‡}	${\rm m}^2~{\rm day}^{-2}$
		mean fraction in ML	-
		start delay at surface [*]	days
		early ending at $\operatorname{surface}^{**}$	days

Table 1. Overview over characteristics for the MHWs defined across the water column as shown in Figure 2b. For metrics marked by a [†] the mean and maximum across each event is calculated. [‡]Vertical propagation is calculated as linear regression slope through upper MHW boundary. ^{‡‡}Propagation flexure is calculated as coefficient of the squared term in quadratic fit through upper MHW boundary. *Start delay at surface is calculated as surface start day minus column start day. **Early ending at surface is calculated as column end day minus surface end day.

For the MHW characteristics at the sea surface, we mostly used the same extreme event characteristics as Hobday et al. (2016): start and end days, duration (^sD), mean and maximum intensity (^s $I_{mean} \& {}^{s}I_{max}$), as well as severity (^s $S = {}^{s}I_{mean} \times {}^{s}D$). For the calculation of these surface characteristics, we only considered the extreme surface grid boxes within a water column-spanning MHW. For instance, for the surface duration we summed the number of days the MHW affects the sea surface (Fig. 2b). In addition to expressing intensities in absolute temperature exceedances above the threshold (I in °C), we calculated variability scaled MHW intensities (\mathfrak{I} , unitless). The variability scaled intensity \mathfrak{I} is diagnosed on the basis of the unitless temperature index \mathfrak{T} , which corresponds to the continuous form of the MHW categories introduced by Hobday et al. (2018). We therefore computed \mathfrak{T} following Sen Gupta et al. (2020) as:

$$\mathfrak{T} = \frac{T - T_{\rm clim}}{T_{\rm thresh} - T_{\rm clim}},\tag{1}$$

where, $T_{\rm clim}$ is the climatological temperature and $T_{\rm thresh}$ is the seasonally varying 90th 245 percentile threshold temperature. $\mathfrak{T} > 1$ thus indicates that the instantaneous temper-246 ature T is larger than $T_{\rm thresh}$. Expressing temperatures during MHWs using the vari-247 ability scaled \mathfrak{T} allows for an unbiased comparison of MHW intensities across different 248 locations, as threshold exceedances are generally proportional to the underlying variance 249 of the temperature time series (Oliver et al., 2021). We calculated $I_{\text{mean/max}}$ and $\Im_{\text{mean/max}}$ 250 as the mean or maximum of absolute threshold exceedances and $\mathfrak T$ values throughout each 251 event, respectively. Additional surface characteristics are the number of interruptions 252 at the surface (while being connected in the subsurface), the maximum duration of a sur-253 face interruption and the maximum number of consecutive extreme days at the surface. 254

Column characteristics describe how MHWs extend through the water column and 255 how they move vertically with time (Fig. 2, Tab. 1). We defined the column duration 256 (^{c}D) as the total number of days that a MHW affects the water column, regardless of 257 its vertical position. We calculated the temporal mean and maximum MHW depth rel-258 ative to the surface ($\delta_{\text{mean}}^{\text{surf.}}$, $\delta_{\text{max}}^{\text{surf.}}$) and relative to the MLD ($\delta_{\text{mean}}^{\text{MLD}}$, $\delta_{\text{max}}^{\text{MLD}}$). We therefore 259 defined the instantaneous MHW depth to be the depth of the lowest extreme grid cell 260 below the surface and below the MLD, respectively. We characterized the vertical prop-261 agation of a MHW as the slope of a linear function fitted to the temporal evolution of 262 the depth of the upper MHW boundary (Fig. 2). The focus on the upper MHW bound-263 ary thereby relates the vertical MHW propagation directly to the sea surface. We fur-264 ther described temporal changes in the vertical propagation behaviour of the MHW with 265 the coefficient of the second order term in a quadratic function fitted to the MHW's up-266 per boundary (i.e., the flexure, Fig. 2). We further estimated for each day of the MHW, 267 what fraction of the vertical MHW extent is located in the surface ML and calculated 268 the temporal mean fraction over the full column duration. Based on the surface and col-269 umn characteristics of MHWs, we further derived secondary characteristics such as the 270 MHW's start delay or its early ending at the surface. The mean column intensity (${}^{c}\mathfrak{I}_{mean}$) 271 was computed by averaging all \mathfrak{T} values from all grid cells of an individual MHW. The 272 maximum intensity (${}^{c}\mathfrak{I}_{max}$) was taken to be the maximum \mathfrak{T} value at any grid point through-273 out the MHW, regardless of its depth. The column severity (^{c}S) was calculated as the 274 product of the mean column intensity (${}^{c}\mathfrak{I}_{mean}$) and the column duration (${}^{c}S = {}^{c}\mathfrak{I}_{mean} \times$ 275 $^{\rm c}D$). 276

To jointly characterize all MHWs occurring at one horizontal location, we computed "composite" statistics of the respective characteristic across all locally occurring MHWs. These composite statistics are either a composite mean (denoted by an overbar \overline{X}), a composite maximum (denoted by a tilde \widetilde{X}) or a composite minimum (denoted by a hat \widehat{X}).

282

2.4 Evaluation of modeled temperature extremes

Given the absence of observational constraints for evaluating MHWs in the subsurface at the basin scale, we are limited to the comparison of the model simulated surface MHWs with observations. In order to construct an observation-based set of surface MHWs that is directly comparable to the simulated set, we applied the same methodology as outlined in Section 2.2.1-2.2.2 to the daily satellite observations of SST from the AVHRR OISSTv2 dataset (Reynolds et al., 2007). This dataset is provided at 0.25° resolution and covers the period from 1982 to 2019.

In the tropical Pacific, we nonetheless compared simulated temperature time series to observed temperature time series during the extreme warming event of the 1997 to 1998 El Niño (see Supplementary Information). We therefore used temperature data from five moorings from the TAO/TRITON mooring array in the tropical Pacific between 155°W and 95°W (McPhaden et al., 2010). This permits us to extend the evaluation of the model to depth, with El Niño-related variability being a strong constraint given its strong impact on subsurface MHWs.

297 2.5 Clustering of deep-reaching MHWs

In order to group the dMHWs according to their vertical propagation behaviour, we used a k-means clustering algorithm (Pedregosa et al., 2011). As we are particularly interested in the vertical propagation behaviour of dMHWs, making them partially invisible at the ocean surface, we performed the clustering on MHW characteristics that reflect the temporal evolution of the upper MHW boundary (see Table 1 and the Supplementary Information for more detail on the clustering procedure).

These characteristics consist of a subset of the characteristics introduced in Table 304 1 and Figure 2, as well as their combinations, namely: a) the ratio of surface to column 305 duration, b) the ratio of surface start delay to the column duration, c) the ratio of sur-306 face early ending to the column duration, d) the vertical propagation, e) the vertical prop-307 agation flexure, f) the ratio between the duration of the longest interruption at the sur-308 face and the column duration. Hence, the chosen characteristics focus on the upper bound-309 ary of each identified MHW and how it relates to the sea surface. Between all charac-310 teristics the rank correlations are generally below (above) 0.6 (-0.6) (Fig. S11). In or-311 der to eliminate collinearity between individual clustering features (Dormann et al., 2013), 312 we performed a principal component analysis on the six characteristics, whereby each 313 of them was scaled across all events. We identified only the first three principal compo-314 nents (PCs) to have eigenvalues above 1, which together explain 77.5% of the variance 315 in MHW dynamics (Fig. S12). The fourth PC still explains 14.4% of the variance, but 316 its eigenvalue is below 1. We hence continued to only use the standardized first three PCs 317 as clustering features (Kaiser's rule, Kaiser (1960)). To find the optimal number of clus-318 ters, we performed the k-means clustering on the first three PCs with varying numbers 319 of clusters, ranging from two to 19. Using the Calinski-Harabasz score (Caliński & Harabasz, 320 1974), the Davies-Bouldin score (Davies & Bouldin, 1979) and the Elbow method (i.e., 321 the within-cluster sum of squared distances), we identified the optimal number of clus-322 ters to be four (Fig. S13). For the final analysis, we thus clustered the first three PCs 323 using four clusters. 324

325

2.6 Sensitivity analyses

To test the sensitivity of the detected MHWs and their characteristics to our method-326 ological choices, we conducted the following sensitivity analyses: We test for a) the ef-327 fect of the chosen vertical resolution (37 vs. 19 z-levels in the upper 500 m), b) the ef-328 fect of horizontal coarsening during downsampling (averaging 3x3 vs. 5x5 grid boxes), 329 c) the effect of the horizontal downsampling methodology (meanpooling vs. maxpool-330 ing), d) the effect of the employed MHW temperature threshold percentile (90th vs. 95th), 331 e) the effect of the employed threshold reference period (1982 to 2011 vs. 1990 to 2019), 332 f) the effect of the temporal filtering of the Boolean array \mathbf{B} (5 day kernel vs. no filter-333 ing) and g) the effect of the analysis period (1979 to 2019 vs. 1982 to 2019). In each sen-334 sitivity case, we solely altered the respective methodological choice, while keeping all other 335 choices as in the reference case, which is used throughout the main manuscript. For a 336 better overview, the different choices are summarized in Figure S18. We find that our 337 results are generally robust with regards to the methodological choices and only differ 338 in expected manners between sensitivity cases (see Supplementary Information). 339

We further tested the robustness of the MHW clustering with respect to the omis-340 sion of 10% to 99% of detected MHWs and to the omission of individual MHW char-341 acteristics feeding into the principal component analysis. We therefore assess the degree 342 of agreement in cluster labeling between the standard case and each sensitivity case, us-343 ing Cohen's Kappa coefficient (κ , Cohen (1960), see Section 3.3 of Supplementary In-344 formation). $\kappa = 1$ indicates perfect agreement, while $\kappa = 0$ suggests agreement based 345 on random labeling. We find that the clustering consistently produces the four clusters 346 even under the omission of 99% of the detected MHWs ($\kappa > 0.97$, Fig. S21a). The clus-347 tering is also robust to the elimination of individual MHW characteristics (Fig. S21b). 348 In most cases, $\kappa > 0.7$. Only for the omission of the "start delay" and the "early end" 349 at the surface (relative to the MHW column duration, see Tab. 1), κ drops to around 350 0.61 to 0.67, indicating the important role of these characteristics in the clustering pro-351 352 cedure. Applying the clustering methodology outlined in Section 2.5 to the detected MHWs in the other MHW detection sensitivity cases (Fig. S18), further produces structurally 353 similar clusters (not shown). Together, these results indicate that the clustering into four 354 different vertical propagation behaviours is robust with respect to our methodological 355 choices and the underlying data. 356



Figure 4. Comparison between simulated and observed MHW (surface-only detection, Fig. 2a) occurrences, durations and maximum intensities. The top row shows the results for simulated MHWs, the bottom row for MHWs detected in the OISSTv2 dataset (Reynolds et al., 2007). Panels a) and b) shows the number of MHWs per year. Panels c) and d) show the composite mean MHW duration. Panels e) and f) show the composite maximum MHW duration. Panels g) and h) shows the composite maximum of maximum MHW intensities (expressed in absolute terms, that is in °C).

357 **3** Model evaluation

We find that the model captures the spatial structure of the number of surface-only 358 MHW (Fig. 2a) occurrences per year relatively well (Pearson's correlation of $r_P = 0.51$), 359 but underestimates the average number of surface-only MHWs per year (n) with an av-360 erage value across the EP study area (spatial average denoted by $\langle \dots \rangle$) of $\langle n \rangle_{\text{Mod.}} =$ 361 1.35 ± 0.38 vs. $\langle n \rangle_{\rm Obs.} = 2.25\pm0.60$ events per year. In contrast, the model has a ten-362 dency to overestimate the composite mean surface-only MHW duration \overline{D} in the EP study 363 area $(\langle \overline{D} \rangle_{\text{Mod.}} = 30.3 \pm 7.4 \text{ vs. } \langle \overline{D} \rangle_{\text{Obs.}} = 23.4 \pm 5.5 \text{ days})$. This overestimation is most 364 prominent in the central tropical Pacific. However, the model reproduces the compos-365 ite maximum duration well $\langle D \rangle_{\text{Mod.}} = 208.5 \pm 78.5 \text{ vs. } \langle D \rangle_{\text{Obs.}} = 233.8 \pm 89.3 \text{ days}$ 366 and also reproduces the spatial distribution with the longest events generally occurring 367 in the tropical EP. The modeled composite maximum of the event maximum intensities 368 agree well with observations with $\langle I_{\rm max} \rangle_{\rm Mod.} = 2.16 \pm 0.87^{\circ} \text{C vs.} \langle I_{\rm max} \rangle_{\rm Obs.} = 2.48 \pm$ 369 0.75°C. The composite intensity maxima spatially correlate with $r_P = 0.71$ between 370 model and observations and are highest in the tropical EP (locally exceeding 4 °C). 371

These evaluation results give us confidence that the model reproduces the observed extreme temperature variability at the ocean surface relatively well. The underestimation of the number and overestimation of duration of surface-only MHWs is a common deficiency of ocean models and is commonly attributed to unresolved variability as well as potential interpolation artefacts in the high-resolution observational SST data sets (Frölicher et al., 2018; Pilo et al., 2019; Gruber et al., 2021). An evaluation of extreme



Figure 5. 2D histograms of MHW mean depth vs. MHW column duration. The colored squares show the binned number of occurrences of the respective MHWs. Panel a) shows the mean depth below the surface $(\delta_{\text{mean}}^{\text{surf.}})$ on the y-axis, while panel b) shows the mean depth below the MLD $(\delta_{\text{mean}}^{\text{MLD}})$. The solid lines show the distribution of MHWs only with respect to depth (using the same 20 m depth bins as in the 2D histogram). The orange line shows the occurrence distribution as a probability distribution function (PDF) in % of all events. The red line shows the summed occurrence distribution from the surface downwards, that is as a cumulative distribution function (CDF). The black horizontal lines indicate the reference depth in each case (i.e., the sea surface and MLD). The top right of each panel shows the Pearson correlation between the respective MHW characteristics.

warming events in the subsurface is challenging, due to the lack of observational high 378 resolution data at the basin scale. Nevertheless, the simulated subsurface temperature 379 evolution during the 2013 to 2016 Blob and the 1997 El Niño match with observations 380 in the respective EP locations (see Sec. 4.3 and Fig. S10). We thus deem the model sim-381 ulation suitable to extend the view on surface MHWs into the ocean interior. For fur-382 ther evaluation, e.g., of model temperature biases, trends and differences in simulated 383 and observed temperature variability, the reader is referred to the Supplementary Infor-384 mation. 385

386 4 Results

Over the 41 years of simulation (1979 to 2019) we detect a total of 1 400 170 (onedimensional) MHWs with at least one surface expression in the study area of the EP. A surface-only detection would find 27% more events, i.e., 1 773 029, since it would count a resurfacing event as two separate events, while our one-dimensional perspective counts it as one (as depicted in the example sketches in Fig. 2).

392

4.1 Vertical extent of MHWs

The majority of the MHWs are quite shallow, with 50% reaching a mean depth of 36 m or less ($\delta_{\text{mean}}^{\text{surf.}}$, Fig. 5a). 90% of MHWs are on average confined to the upper 150 m and less than 1% reach a depth of 450 m or more below the surface. However, given that extreme conditions at the surface usually tend to extend to the MLD, it is more insightful to analyze the MHW mean depth relative to the MLD ($\delta_{\text{mean}}^{\text{MLD}}$) (Fig. 5b). Indeed, we find that only 13% of all MHWs are confined to waters shallower than the MLD, with this fraction likely associated with conditions where the active mixing is substantially



Figure 6. Mapped depth characteristics of dMHWs, that is MHWs that reach on average deeper than one (seasonally varying) climatological MLD standard deviation below the instantaneous MLD. a) Map of local fraction of dMHWs; b) same as a) but weighted with the respective MHW surface durations (weighted fraction can be interpreted as likelihood that a randomly chosen extreme surface temperature is part of a dMHW); c) composite mean average depth below the MLD of all dMHWs. In all panels, the dashed black line indicates the Equator.

shallower than the MLD. Including these events, a total of 71.5% of all MHWs are on 400 average confined to the ML. In this classification as ML-confined MHWs (abbreviated 401 as sMHWs), we pardon short or low-amplitude excursions of the MHW below the MLD, 402 by allowing them to reach on average less than one (seasonally varying) standard devi-403 ation of MLD (on average about 10 m) below the varying MLD itself (see Fig. 2b). Thus, 404 as they do not substantially leave the ML, we find that 71.5% of all MHWs are funda-405 mentally accountable for by MHW analyses based on mixed layer or sea surface temper-406 ature data. The remaining 28.5% of MHWs reach on average one MLD standard devi-407 ation or more beneath the MLD. In the following we will classify those MHWs as deep-408 reaching MHWs (abbreviated as dMHWs). About a third of those dMHWs (11% of all 409 MHWs) fall into the mean depth range from 10 m to 30 m below the MLD. The remain-410 ing two thirds (18% of all MHWs) extend on average by more than $30 \,\mathrm{m}$ on average be-411 low the MLD. The dMHWs are of particular interest to us, since they are partially de-412 coupled from the surface. 413

⁴¹⁴ We find a general link between the MHW mean depth and the event column du-⁴¹⁵ ration ($r_P \approx -0.6$), with longer events tending to reach deeper below the surface and ⁴¹⁶ the MLD (Fig. 5). MHWs lasting shorter than $^{c}D = 10$ days are mostly limited to the ⁴¹⁷ upper 100 m and ML, while longer lasting events often reach below the ML. For exam-⁴¹⁸ ple, 35% of events that last between $^{c}D = 100$ and 125 days, have a mean depth of more ⁴¹⁹ than 110 m below the surface.

The deep-reaching MHWs occur primarily in the eastern tropical Pacific and along 420 a coastal strip extending northward from Southern California up to Alaska, with regions 421 where 40% of all MHWs are dMHWs (Fig. 6a). In contrast, in the Subtropics, only 10%422 to 30% of the MHWs are deep-reaching. When we take into consideration that dMHWs 423 tend to last longer (see Fig. 5), i.e., when we weight each MHW by its surface duration, 424 the regional pattern of the dMHW becomes even more pronounced (Fig. 6b). With this 425 weighting, in most of the tropical EP more than 80% of all extremely warm tempera-426 tures at the sea surface are associated with dMHWs and in the subpolar North Pacific 427



Figure 7. Comparison between deep-reaching (dMHW, upper row) and surface MHW (sMHW, lower row) occurrences, durations and intensities. Panels a) and b) shows the number of events per year. Panels c) and d) illustrate the composite mean event duration. Panels e) and f) illustrate the composite maximum event duration. Panels g) and h) show the composite maximum of the maximum intensity recorded for each event (in terms of \mathfrak{T}).

this fraction exceeds 70 % (Fig. 6b). This contrasts with the subtropical gyres, where the weighted probabilities are less than 30 %.

Averaged across the EP, we find that dMHWs reach on average 110 m below the 430 MLD ($\langle \overline{\delta_{\text{mean}}^{\text{MLD}}} \rangle$, Fig 6c). However, there are coherent regional differences. The dMHWs 431 tend to reach deepest along the West Coasts of the Americas, especially in the Califor-432 nia Current System, and in the low-latitude EP, in particular along $\sim 10^{\circ}$ latitude on both 433 hemispheres. In these regions, the dMHWs extend, on average, substantially deeper than 434 160 m below the MLD, i.e., $\overline{\delta_{\text{mean}}^{\text{MLD}}} < -160 \text{ m}$ (Fig 6c). However, along the Equator, where 435 we see a high occurrence of dMHWs, these dMHWs do not reach particularly deep on 436 average, but only to around $\overline{\delta_{\text{mean}}^{\text{MLD}}} \approx -120 \,\text{m}$. In the subtropical gyres, the few MHWs 437 that qualify as dMHWs, reach generally around 40 m to 80 m below the MLD. The dMHWs 438 also remain relatively shallow in the Alaska gyre, with depth of typically less than 50 m 439 below the MLD. 440

441

4.2 Core characteristics of deep-reaching MHWs

Averaged across the EP study area, we find $\langle n \rangle = 0.26 \pm 0.15$ dMHWs per year compared to 0.8 ± 0.27 sMHWs per year, i.e., deep-reaching MHW occur about three times less often than sMHWs, consistent with the ratio of their total counts (Fig. 7a,b). The highest number of dMHWs per year are found in the eastern tropical North Pacific and along the North American coast (up to 0.8 dMHWs per year, cf. Fig. 6a). While occurring much less often than sMHWs, dMHWs last five times longer, with a mean dMHW column duration of $\langle \overline{^{c}D} \rangle = 184 \pm 82$ days compared to $\langle \overline{^{c}D} \rangle = 36 \pm 17$ days for the



Figure 8. Mean profiles of MHW related temperature anomalies across depth. Left panel shows absolute temperature anomalies relative to the extreme threshold (i.e., absolute intensities), while right panel shows variability scaled temperature anomalies (i.e., in terms of the variability scaled temperature \mathfrak{T}). Mean profiles are calculated by averaging the (scaled) temperature anomalies across all extreme grid cells associated with MHWs throughout the hindcast. (Red) orange lines show the (grid cell size weighted) mean profile averaged over all extreme grid cells. The black (grey) lines show the $25^{\text{th}}/75^{\text{th}}$ ($10^{\text{th}}/90^{\text{th}}$, 99^{th}) percentile levels for each depth.

sMHWs. The dMHWs tend to last especially long in the tropical Pacific and the sub-449 polar North Pacific, where the composite mean column durations (^{c}D) often exceed 250 450 days (Fig. 7c,d). The longest dMHWs are found in the subpolar North Pacific, with com-451 posite maximum column durations (^{c}D) of more than 3 years, which exceed the max-452 imum sMHW duration in the same region by over 2 years (Fig. 7e,f). These extremely 453 long dMHWs are associated with the Blob heatwave event in the subpolar NP (see dis-454 cussions below). The maximum column intensities of dMHWs are substantially higher 455 than of sMHWs ($\langle \overline{I_{\text{max}}} \rangle = 5.0 \,^{\circ}\text{C}$ vs. $\langle \overline{I_{\text{max}}} \rangle = 1.9 \,^{\circ}\text{C}$). The maximum dMHW in-456 tensities scaled with the background variability locally exceed ${}^{c}\mathfrak{I}_{max} \geq 10$ in the trop-457 ical EP, with marked bands of extremely high intensities along the Equator and along 458 $\sim 10^{\circ}$ N (Fig. 7g). These extremely high intensities occur below the ML, as the same re-459 gions do not stand out when analyzing the intensities of sMHWs (Fig. 7h, Fig. 4). In 460 the subtropical and subpolar regions, the maximum dMHW column intensities are only 461 slightly higher than the maximum sMHW intensities (generally less than 2 \mathfrak{T} units dif-462 ference). 463

The MHWs tend to intensify in the subsurface. Aggregated across all MHWs (dMHWs 464 and sMHWs) we find the strongest temperature threshold exceedances at around $\sim 100 \,\mathrm{m}$ 465 depth (Fig. 8a). Below 175 m, temperature anomalies (in absolute terms) drop on av-466 erage below the surface anomaly amplitudes and decrease steadily down to 500 m. How-467 ever, expressing the temperature anomalies in terms of the variability scaled \mathfrak{T} (Fig. 8b) 468 shows that MHWs tend to be subsurface intensified throughout most of upper 500 m of 469 the water column, with \mathfrak{T} values of around 1.5, which is slightly higher than the aver-470 age surface intensities ($\mathfrak{T} \approx 1.4$). 471

In summary, incorporating the depth dimension in the MHW definition and allowing MHWs to propagate vertically reveals a set of intense, and long lasting MHWs that



Figure 9. The Blob in the hindcast simulation. Maps in panel a) show snapshots of the variability scaled temperature \mathfrak{T} taken in March of the years 2012 to 2019 at the surface (upper row) and at 200 m depth (lower row). Panels b) and c) show time-depth sections of the temperature anomaly relative to the climatology (panel b) and the corresponding \mathfrak{T} (panel c) at 145°W, 44°N (indicated by the cross in the maps of panel a). White (dashed) lines in panels b) and c) indicate the MLD (200 m depth).

extend well below the ML. Especially in the subpolar North Pacific and the tropical eastern Pacifc, the difference of MHW durations and intensities between the sMHWs and
dMHWs is striking. These results highlight that MHWs can be extremely intense and
long-lasting in the subsurface, while being invisible or inconspicuous at the ocean surface.

479

4.3 The Blob as prime example of a long-lasting, deep-reaching MHW

Given the rich literature on the long-lasting Blob in the Northeast Pacific (Bond 480 et al., 2015; Freeland & Ross, 2019; Gruber et al., 2021), our finding above that this event 481 was also associated with extremely long-lasting subsurface MHWs exceeding 1000 days 482 (Fig. 7e) warrants a more detailed investigation. This permits us also to illustrate our 483 one-dimensional water column concept with a concrete and well studied example. Our 484 approach is aided by the fact that our simulated *Blob* reproduces the observed vertical 485 propagation behaviour rather well (Scannell et al., 2020) (cf. Fig. 9b and their Fig. 4b). 486 This is illustrated by the time series from 145°W and 44°N, that is the grid point clos-487 est to the study location of Scannell et al. (2020) (Fig. 9b) and also the grid location where 488 we found one of the longest dMHWs. This MHW starts at the sea surface in early 2015. 489 The associated positive temperature anomaly, which is initially strongest at the sea sur-490 face, moves subsequently downward below the ML, where it lingers at around 200 m depth 491 well into 2018, that is for over three years. At 200 m depth, the associated coherent MHW 492 signal appears to slowly move eastward between 2013 and 2019 (lower panels in Fig. 9a). 493 While the subsurface temperature anomalies are generally weaker than those at the sea 494

⁴⁹⁵ surface (~0.5 °C vs. >1.1 °C), the intensities in terms of the variability scaled temper-⁴⁹⁶ ature \mathfrak{T} show much higher values at around 200 m depth ($\mathfrak{T} > 4$ vs. $\mathfrak{T} \approx 2$), owing to ⁴⁹⁷ the generally reduced temperature variability in the subsurface. This reduces the denom-⁴⁹⁸ inator in the definition of \mathfrak{T} , leading to higher values of \mathfrak{T} (Eq. 1). Hence, the simula-⁴⁹⁹ tion shows a subsurface intensification of the MHW, relative to the local variability.

500

4.4 Linking the vertical MHW structure to surface characteristics

As most observations of MHWs are restricted to the sea surface, the question arises to which degree the water column MHW characteristics can be inferred from their surface characteristics. To study this link, we associate each detected surface-only MHW with its corresponding one-dimensional MHW in the water column. By definition, each water column detected MHW comprises the extreme grid cells of at least one surfaceonly MHW. In this context, multiple surface-only MHWs may be associated with the same water column MHW.

Individual characteristics of surface-only MHWs such as duration, maximum in-508 tensity or severity are only moderately correlated with the same characteristics for one-509 dimensional MHWs in the water column (Fig. S22). Thus, a direct estimate of individ-510 ual MHW characteristics based on surface-only MHW properties is challenging. Yet, the 511 comparison between dMHW and sMHW characteristics shows that dMHWs show dif-512 ferent characteristics than the ML-confined sMHWs (Fig. 7). Thus, through the use of 513 a logistic regression model, we might be able to identify those characteristics that allow 514 us to state, with a certain probability whether a surface detected MHW is a deep-reaching 515 one. To this end, we first link the surface-only MHW characteristics (duration, maxi-516 mum intensity and severity) to the vertical characteristics of the associated one-dimensional 517 MHW in a binary form (i.e., 1 if the associated MHW is a dMHW, 0 if the associated 518 MHW is a sMHW) and then fit a logistic regression model to the respective data pairs 519 across all detected surface-only MHWs (Fig. 10). 520

As expected from sMHWs generally being much shorter than dMHWs (see Section 521 4.2), we find that short surface-only MHWs are more likely to be associated with sMHWs 522 than with dMHWs (Fig. 10a), while surface-only MHWs that last longer than 49 days 523 are more likely to be associated with dMHWs. If we used this simple (uni-variate) lo-524 gistic regression model across the entire EP, 68% of predicted dMHWs are true positives, 525 while 32% of surface-only MHWs are mistakenly predicted to be dMHWs. If we fit such 526 a model regionally, i.e., just for the California Current System or the equatorial EP, the 527 uni-variate logistic regression model could already predict 72% (76%) of all events cor-528 rectly (Fig. S23). 529

An even better prediction is possible if severity $(D \times \mathfrak{I}_{mean})$ was used as single vari-530 ate predictor (Fig. 10c). In this case MHWs detected at the surface with a severity (S =531 $\mathfrak{I}_{\text{mean}} \times D$) of 71 days have a more than 50 % probability of being a dMHW. This re-532 gression then permits to detect dMHW correctly 70% of the times. Again, a regional-533 ization of the logistic model further enhances this fraction to 74% (77%) for the Cal-534 ifornia Current System (Equatorial Pacific, Fig. S23). In contrast, maximum intensity 535 $(I_{\rm max})$ as predictor variable yields only true dMHW predictions in 59% of all cases, with 536 $I_{\text{mean}} = 1.17 \,^{\circ}\text{C}$ as critical distinction value (Fig. 10b). sMHWs, i.e., MHWs that are 537 confined to the surface ML, are correctly predicted with a similar skill as dMHWs. Us-538 ing any of the three surface-only MHW characteristics, sMHWs are correctly predicted 539 in approximately two thirds of all cases. 540

These uni-variate logistic regression models suggest that surface-only MHW characteristics have a good potential to detect the presence/absence of a deep-reaching MHW, especially if prior knowledge about the spatial distribution of the deep-reaching MHW is taken into account. Future studies are required to explore this link in more detail and build more complex, multivariate statistical models that could also include other dynam-



Figure 10. Linking surface-only detected MHW properties to the depth extent of the associated water column MHWs. Panels a, b and c show logistic regressions between the surface-only detected MHW characteristics (duration, maximum intensity and severity, respectively) and the associated binary distinction between sMHW and dMHW. Each panel contains a histogram of the shallow (blue) and deep-reaching (orange) MHWs. Thin histogram lines indicate in both cases the histogram of the totality of MHWs. Thick black line shows the logistic regression based on the linear logit function f(x). Red text indicates locations where probabilities of the logistic regression correspond to 0.5. In the middle of each panel, a confusion matrix shows for the statistical model fit which fraction of predicted sMHWs (dMHWs) are true sMHWs (dMHWs) (columns sum to 1).

	E MH	Bloci IWs (C	k-lik Cluste	9 er 1)	D MH)eep IWs ((enin Cluste	g r 2)	MН	Shoa IWs (C	aling Cluste	er 3)	Mu MH	lti-sı IWs (C	urfac Cluste	r 4)	Sur (i.e. N	face //L co	MH nf. Mi	Ws HWs)
# of events (% of MHWs, % of dMHWs)	165'111 (11.8%, 41.3%)			93'764 (6.7%, 23.5%)			81'167 (5.8%, 20.3%)			59'704 (4.3%, 14.9%)			1'000'424 (71.5% of MHWs)		/s)					
% of surface MHWs Composite 100 Composite 80 normalised 5 60 time-depth 2 0 structure 0 max depth	Start			End	Start			End	Start			End	Start			End	Start			End
Characteristics	p 10	mean	p 50	p 90	p 10	mean	p 50	p 90	p 10	mean	p 50	p 90	p 10	mean	p 50	P 90	p 10	mean	p 50	p 90
Mean depth bel. surface (δ ^{surf.} mean, m)	-310	-125	-80	-21	-377	-156	-108	-32	-412	-177	-117	-32	-339	-169	-130	-36	-76	-32	-25	0
Mean depth below MLD (δ ^{MLD} mean, m)	-267	-91	-37	-8	-321	-111	-54	-12	-366	-138	-68	-12	-297	-127	-79	-13	-7	2	-0.2	10
Column duration (°D, days)	17	142	76	376	26	116	73	246	23	113	69	244	54	191	152	354	6	28	16	63
Surface to column duration ratio (*D/°D), no unit)	0.62	0.80	0.81	0.97	0.10	0.32	0.31	0.54	0.10	0.32	0.30	0.56	0.21	0.44	0.45	0.65	0.45	0.81	0.90	1
Vertical propagation (m day-1)	-0.63	-0.03	0.0	0.49	-2.26	-1.00	-0.66	-0.10	0.04	1.22	0.72	3.00	-0.42	-0.03	-0.04	0.38	-1.43	-0.25	0.0	0.52
Mean fraction in ML (no unit)	0.12	0.36	0.37	0.58	0.04	0.23	0.21	0.43	0.03	0.18	0.16	0.38	0.05	0.26	0.25	0.47	0.49	0.76	0.77	1
Max intensity (3 _{max} , no unit)	1.4	2.79	2.09	5.46	1.40	2.25	1.97	3.39	1.37	2.39	1.97	3.99	1.57	3.02	2.53	4.96	1.13	1.49	1.39	1.97
Severity (3 _{mean} × D, days)	21	222	102	659	30	161	92	350	26	157	85	359	64	274	204	529	7	34	18	79

Figure 11. Vertical propagation clusters for all dMHWs. The image panels show the structure of the four different cluster types using heatmaps calculated based on all MHWs of the respective cluster normalized in depth and time. Below, statistics calculated based on the MHW characteristics for each cluster is shown. These statistics include the mean, 10th, 90th and 50th percentile (median). For comparison, the rightmost panel and statistics column show the analoguous results obtained for the sMHWs, which are not part of the four other clusters.

ical variables, such as sea surface height or wind speed. Such models could furthermore
 explore the surface-based prediction of specific dMHW characteristics, such as their in tensity or duration.

549

4.5 Vertical propagation behaviour of deep-reaching MHWs

The example of the North Pacific Blob (Sec. 4.3) shows that the extreme signal as-550 sociated with dMHWs can undergo a vertical propagation throughout the MHW life-551 time. To better understand this vertical propagation behaviour and to distinguish be-552 tween different vertical propagation patterns, we cluster all dMHWs as described in Sec-553 tion 2.5. We identify four clusters of dMHWs (Fig. 11), that can be described by their 554 overall time-depth structure as: a) block-like events, b) deepening events, c) shoaling events, 555 d) multi-surfacing events. While the clusters are well separated, the diversity of event 556 characteristics within each cluster is large (see statistics and violin plots in Fig 11, S14, 557 S15, S17). 558

The block-like cluster, making up 41.3% of all dMHWs, is characterized by rela-559 tively shallow events (cluster averages denoted by [...]) of $[\delta_{\text{mean}}^{\text{surf.}}] = -125 \text{ m}, [\delta_{\text{mean}}^{\text{MLD}}] =$ 560 -91 m but also by events that are relatively long-lasting $[^{c}D] = 142 \text{ days}$. Through-561 out their lifetime, these events are quasi-permanently visible at the ocean surface, with 562 the cluster average surface to column duration ratio being relatively high, i.e. $[{}^{s}D/{}^{c}D] =$ 563 0.80. In contrast, the deepening events, which make up 23.5% of dMHWs, affect the sur-564 face only during 32 % of their average column duration of $[^{\rm c}D]=116$ days ([$^{\rm s}D/^{\rm c}D]=$ 565 0.32). These events exhibit on average a downward propagation of -1.00 m day^{-1} and 566 reach on average down to $[\delta_{\text{mean}}^{\text{surf.}}] = -156 \text{ m} ([\delta_{\text{mean}}^{\text{MLD}}] = -111 \text{ m})$. Shoaling events (20.3% of all dMHWs) show an average upward propagation of 1.22m day⁻¹ and also affect the 567 568



Figure 12. Spatial distribution of the contribution of the different MHW clusters to dMHWs, i.e. MHWs that reach on average more than one climatological MLD standard deviation below the instantaneous MLD. The fields are smoothed with a 3×3 Gaussian kernel. The unsmoothed four panels add up to 100% of all dMHWs.

surface ocean only during 32% of their $[^{c}D] = 113$ days lifespan. On average, shoal-569 ing events reach down to $[\delta_{\text{mean}}^{\text{surf.}}] = -177 \text{ m}$ and substantially below the MLD ($[\delta_{\text{mean}}^{\text{MLD}}] =$ 570 $-138\,\mathrm{m}$). The multi-surfacing cluster contains $14.9\,\%$ of all dMHWs. These events are 571 the longest-lasting ($[^{c}D] = 191$ days) and reach on average 169 m below the surface (127 m 572 below the MLD). They affect the sea surface on average during 44% of their lifetime. 573 In summary, we find that nearly two thirds of the dMHWs ($\sim 59\%$) show a behaviour 574 that is not block-like, i.e., that are located for a substantial part of their lifetime purely 575 below the sea surface and are thus not visible in any surface-only based detection algo-576 rithm. 577

The different clusters show distinct spatial distributions (Fig. 12). The block-like 578 dMHWs occur throughout the EP and in particular in the regions of the dMHW hotspots 579 (see Fig. 7a). Deepening MHWs hardly occur in the tropical EP, but make up around 580 a third of dMHWs in the subtropical and subpolar EP (Fig. 12). In contrast, shoaling 581 MHWs are mostly present at low latitudes between 10°N and 10°S. Multi-surfacing MHWs 582 mainly occur in the tropical EP, but also account for more than 20% of dMHWs in many 583 regions of the subtropical and subpolar North Pacific. These identified regional differ-584 ences in the occurrence of the different vertical propagation patterns of dMHWs indi-585 cate that different drivers are at play in their generation. 586

587 5 Discussion

In this study, we analyzed the vertical structure and propagation behaviour of MHWs that affect the sea surface, by extending the classical MHW definition (e.g. Hobday et al. (2016)) to incorporate the vertical dimension and by thus furnishing MHWs with a vertical extent. As a proving ground, we applied this methodology to output from a model hindcast simulation in the EP (1979 to 2019). But we consider the findings and insights gained here to be of relevance also in other regions of the world's oceans.

594

5.1 The vertical structure and propagation of MHWs

⁵⁹⁵ Our findings show that while the majority of MHWs detected at the ocean surface ⁵⁹⁶ are limited to the ML (Fig. 5,7a,b), about one third ($\sim 29\%$) of these reach on aver-⁵⁹⁷ age more than one climatological MLD standard deviation below the MLD. Thus, one ⁵⁹⁸ out of three times the additional buoyancy and increased water column stratification as-⁵⁹⁹ sociated with a surface MHW, which hinders the downward mixing of the temperature anomaly signal below the ML (Oliver et al., 2021), can be overcome. When the MHW
signal is below and partially decoupled from the ML, the resulting deep-reaching MHWs
also tend to last longer. This matches the findings of Elzahaby and Schaeffer (2019), who
also found an increase in MHW durations with increasing penetration depths in the Tasman Sea.

Our results show that MHWs reaching below the ML occur mostly in the eastern 605 tropical (North) Pacific as well as along the American west coast (Fig. 6a,b). So far, no 606 analysis of MHW depths exists across the full EP to compare with our findings shown 607 in Fig. 6. Yet, regional observations and analyses of MHWs appear to support our re-608 sults. For instance, tropical EP MHWs are often driven by El Niño events (Holbrook et 609 al., 2019). Since El Niño events go along with a strong thermocline warming (e.g., En-610 field, 2001), it is reasonable that the here detected MHWs show a propensity to reach 611 below the ML. This is further corroborated by the results of Hu et al. (2021), who find 612 in mooring array data that MHWs in the thermocline of the equatorial Western Pacific 613 often occur in conjunction with surface MHWs. Similary, the increased proportion of dMHWs 614 along the American coastline is perhaps associated with coastally trapped planetary waves 615 that lead to local warming and downwelling conditions (Frischknecht et al., 2015; Wei 616 et al., 2021). Lastly, the here detected deep extent of the North Pacific Blob down to \sim 617 400 m (Fig. 9) agrees with the previously documented vertical extent of strong subsur-618 face warm anomalies (Freeland & Ross, 2019; Scannell et al., 2020). 619

Next to revealing the MHW depth, our here employed methodology allows us to 620 analyse the subsurface characteristics of MHWs. We find that deeper-extending MHWs 621 tend to last longer and to show higher maximum intensities (Fig. 5, 7, 8, 11). Compared 622 to the ML-confined sMHWs, the dMHWs last substantially longer ($\Delta \langle \overline{^{c}D} \rangle = 148$ days) 623 and show stronger intensities $(\Delta \langle \bar{c}I_{\max} \rangle = 3.1 \,^{\circ}\text{C}$, Fig. 7). As an example of a dMHW, 624 we find that the extremely long subsurface *Blob* persistence described by Scannell et al. 625 (2020) was unique in the EP. Nowhere else do we identify such long dMHW durations 626 of over 3 years (Fig. 7e, 9), which are furthermore substantially longer than the max-627 imum surface-only MHW durations of less than one year in the same region (Fig. 4e). 628 In the tropical EP, the difference between the maximum surface-only MHW and dMHW 629 duration is less pronounced, but still around 200 days (comparing Fig. 4f and Fig. 7e). 630 There, maximum dMHW column durations of around 500 days are at the upper bound 631 of time scales associated with El Niño events (Okumura & Deser, 2010), which are known 632 to be the main driver of MHWs in this region (Holbrook et al., 2019). These examples 633 demonstrate that surface-only descriptions of MHWs can dramatically underestimate 634 for how long and with which intensity the upper ocean water column is subject to ex-635 treme conditions. 636

The surface-only perspective on MHWs cannot per se distinguish whether the as-637 sociated extreme signal is confined to the ML or whether it extends also below the ML. 638 The here chosen approach to define MHWs in the water column appears as a useful tool 639 to study this link. Based on simple uni-variate logistic regression models, we find that 640 the characteristics of surface-only MHWs carry some predictive capacity with regards 641 to the vertical extent of the associated (water column) MHW. For instance, a prediction 642 of a dMHW, using surface-only MHW severity as predictor, is correct in 70% of all cases. 643 These results encourage further studies to explore the link between surface-only MHW 644 characteristics and the vertical structure of the associated MHW. Such links can be used 645 to predict the vertical structure of MHWs that are detected with remote sensing data 646 of sea surface temperature. Additional predictor variables that are available from remote sensing at high spatio-temporal resolution, such as sea surface height, salinity or heat 648 and momentum fluxes might support the development of (multivariate) statistical mod-649 els (Su et al., 2018) that predict the vertical extent of MHWs when detected at the sea 650 surface. Such statistical models could furthermore be improved by a better dynamical 651 understanding of the MHW drivers. The coherent spatial patterns in the dMHWs oc-652



Figure 13. Summary of different vertical propagation patterns and their potential driving processes. Colors are associated with clusters as color-coded in Fig. 11, i.e. *block-like*, *deepening*, *shoaling* and *multi-surfacing* from left to right.

currence likelihood, as well as the mean depth of those dMHWs (Sec. 4.1, Fig. 6) sug-653 gest that similar driving mechanisms lead to such deep reaching MHWs. For instance, 654 the deep reaching dMHWs along 10°N, which reach on average $\sim 200 \,\mathrm{m}$ below the ML 655 (Fig. 6c), and which spatially coincide with the very high intensities of dMHWs (Fig. 656 7g), are in the alleyway of the westward propagating anticyclonic Tehuantepec eddies 657 (Palacios & Bograd, 2005), which are associated with strong positive temperature anoma-658 lies mainly in the thermocline (Purkiani et al., 2022). Hence, the identification of (an-659 ticyclonic) mesoscale eddies might further support correct predictions of deep reaching 660 MHWs (Elzahaby & Schaeffer, 2019). Eventually, the established statistical relationships 661 need to be put to test using observational data, i.e., using high-resolution (satellite-derived) 662 sea surface records of MHWs and associated temperature measurements across the wa-663 ter column. 664

Lastly, the fact that MHWs can affect the water column much longer than diag-665 nosed at the sea surface (see examples given above), highlights the important role of ver-666 tical propagation of the MHW signal (Scannell et al., 2020). Using the water column def-667 inition of MHWs, we find four clusters with distinct vertical propagation behaviour for 668 the dMHWs (Sec. 4.5). While block-like events remain quasi-permanently at the sea sur-669 face, the other three identified dMHW types ($\sim 59\%$ of dMHWs, $\sim 17\%$ of all MHWs) 670 affect the sea surface on average only 32% to 44% of their lifetime (Fig. 11). Hence, for 671 17% of all MHWs, up to two thirds of the MHW duration can be missed when relying 672 solely on surface-only MHW characteristics. 673

674

5.2 Potential drivers shaping the different vertical MHW structures

While a full driver analysis of the different vertical MHW structures goes beyond the scope of this study, we briefly discuss how the different propagation patterns in the extreme signal could be generated. Figure 13 provides therefore a non-exhaustive overview of potential driving mechanisms.

MHWs driven by atmospheric heat fluxes can lead to dMHWs, if the heat anomaly is transferred below the MLD. This can either occur through downward mixing (Fig. 13a) or through the detrainment process associated with variability in the MLD (Fig. 13d,

Scannell et al. (2020)). In this process, anomalous warm water detrains from a deep (win-682 ter) ML and is incorporated into the thermocline during the (seasonal) shoaling of the 683 MLD (Alexander & Deser, 1995). Thus, the MHW signal lingers in the thermocline, while 684 temperatures can return to normal levels in the ML above, forming a deepening MHW. The finding that deepening MHWs occur primarily in spring/early summer (not shown) 686 and only in the subtropical and subpolar EP, where the MLD undergoes a marked sea-687 sonal cycle (Fig. 12b), points towards a potential role of this detrainment process. In 688 the opposite sense, a thermocline extreme warm anomaly could be entrained into the ML 689 during the (fall time) MLD deepening, leading to a shoaling MHW (Fig. 13g), an infer-690 ence also supported by these events occurring predominantly in late summer/fall (not 691 shown). A combination of springtime MHW detrainment out of the ML and fall time 692 MHW re-entrainment into the ML would describe a full re-emergence cycle (and qual-693 ify as a multi-surfacing MHW, Fig. 13j), which has long been noted for its role in affect-694 ing wintertime ML temperatures of the Northeast Pacific (Namias & Born, 1970; Alexan-695 der & Deser, 1995). Hence, in extra-tropical regions the seasonal variability in MLD and 696 the associated detrainment (and (re-)entrainment) processes, are potential drivers be-697 hind vertically propagating MHWs. 698

MHWs driven by lateral advection are not necessarily restricted to the ML, but 699 can span across the water column (Elzahaby & Schaeffer, 2019; Oliver et al., 2021; Großelin-700 demann et al., 2022). Hence, depending on the vertical nature of the lateral cross-gradient 701 temperature advection, the local MHW signal can start simultaneously across the wa-702 ter column (Fig. 13c) or initiate at the surface or at depth (Fig. 13f and i, respectively). 703 Downward isotherm displacement associated with adiabatic heaving during the passage 704 of a warm core eddy can similarly lead to deep reaching MHWs, that likely qualify as 705 block-like MHWs (Fig. 13b). The dynamic (sub-surface) equatorial current system (Kessler, 706 2006) and the trajectories of warm-core eddies are thus potential hotspots for advection 707 driven MHWs with different vertical propagation signatures. 708

The presence of subsurface MHW signals might furthermore precondition the for-709 mation of a MHW in the ML above, by reducing the downward mixing of any excess heat 710 introduced into the ML. Depending on whether the MHW extends only once or multi-711 ple times in the ML, the overall MHW could take on a shoaling pattern (Fig. 13h) or 712 a multi-surfacing pattern (Fig. 13k). A potential region, where this process plays a role 713 714 is the equatorial Pacific, where we find elevated occurrences of shoaling events (Fig. 12c). In the tropical EP, where MHWs are tightly connected to El Niño events (Holbrook et 715 al., 2019), we discern a shoaling behaviour of the warm anomaly signal which initiates 716 in the subsurface during the 1997 to 1998 El Niño event (Fig. S10). Further support-717 ing the role of initial subsurface warming in driving shoaling events, Vogt et al. (2022) 718 find an important contribution of reduced vertical mixing of heat during the onset phase 719 of MHWs in the tropical EP. 720

5.3 Biological implications

721

The here presented extended view on MHWs provides new insights into the haz-722 ard MHWs represent to marine life. While we find that the majority of MHWs repre-723 sent an elevated heat stress purely in the ML, 29% of MHWs can directly impact or-724 ganisms in the thermocline below. The pronounced and deep reaching MHWs can ad-725 versely impact species' fitness due to limited nutrient/food availability or increased metabolic 726 demands (Smith et al., 2023). This effect could be aggravated by lowered oxygen avail-727 ability, causing even a compound extreme event (Gruber et al., 2021). At the same time, 728 the anomalous deep warming of temperature stratified water columns can open thermal 729 windows for vertically migrating marine species that are otherwise limited in their range 730 by cold subsurface temperatures (Seibel & Birk, 2022), creating potential winners. But 731 to fully evaluate the exposure of marine organisms to such extremes, it is important to 732 also consider the (active and passive) displacement of the organisms through their up-733

per ocean habitat (Hofmann Elizondo & Vogt, 2022). For instance, the (diurnal) ver tical migration performed by planktonic and nektonic species strongly influences at which
 times and depths the organisms are exposed to extreme conditions.

Considering the vertical extent of MHWs also has implications for the MHW driven 737 displacement of species. Based on sea surface temperature data only, Jacox et al. (2020) 738 estimated that organisms exposed to MHWs need to horizontally move tens to thousands 739 of km, in order to recover the original climatological conditions. This surface confined 740 view on species displacement neglects the possibility of vertical displacements. In response 741 742 to MHW conditions at the sea surface, organisms could transiently shift their vertical position to cooler subsurface waters, under the condition that the other habitat-forming 743 biotic and abiotic factors such as oxygen concentrations, food availability, light, etc. are 744 favourable (Reygondeau et al., 2013; Wishner et al., 2013; Seibel & Birk, 2022). In this 745 context, the vertical MHW extent as well as other (biogeochemical) extremes in the wa-746 ter column (Pörtner & Farrell, 2008; Bednaršek et al., 2018; Gruber et al., 2021; Burger 747 et al., 2022; Köhn et al., 2022) can influence the feasibility of this vertical displacement. 748

5.4 Caveats

749

In this study, we analyze vertical structures and propagation behaviours of MHWs, 750 using output from a hindcast simulation performed with the regional ocean model ROMS. 751 Similar to other models (e.g., Pilo et al. (2019)), our simulation shows biases in the du-752 ration and frequency of surface-only MHWs (Sec. 3, Fig. 4), despite the model's gen-753 eral skill to reproduce the observed temperature field (Fig. 1b, Supplementary Informa-754 tion). Beyond the sea surface, our evaluation of simulated temperature extremes is re-755 stricted to individual MHWs, such as the Blob (Fig. 9) and the 1997 to 1998 El Niño 756 event (Fig. S10), due to scarce high resolution observational data. The realistic repro-757 duction of the subsurface structure of these MHWs gives us confidence in the model's 758 skill in reproducing the subsurface extreme temperature variability. 759

The purely one-dimensional approach to define MHWs does not capture the full 760 three-dimensional evolution of MHWs over time. As such, we here analyze MHWs in-761 dependently from one another, even if they occur in horizontally neighbouring locations. 762 Furthermore, next to moving vertically through the water column, MHW signals can move 763 (be advected) horizontally. Depending on the nature of the lateral movement, the lat-764 eral displacement of the MHW signals might rectify locally as vertically propagating MHWs 765 (see Fig. 13f,i, Sec. 5.2). Similarly, a lateral movement of the extreme signal in the sea-766 sonally developing thermocline might lead to the break-up of a re-emergence cycle into 767 a local deepening MHW and a shoaling MHW downstream. As such, Tak et al. (2021) 768 found that the combination of the re-emergence phenomenon and lateral advection can 769 affect sea surface MHW statistics downstream of the North Pacific subtropical mode wa-770 ter formation site. To additionally account for the lateral coherence and movements of 771 MHWs, a full three-dimensional quasi-Lagrangian tracking of extremes would be required 772 (Desmet et al., 2022). 773

Lastly, it is important to note, that the propagation of the statistical MHW signal is not necessarily bound to the movement of a particular water mass, but can move and connect with ease across isopycnals. Hence, extreme conditions that occur simultaneously in the ML and below due to different drivers, can appear as one coherent MHW (e.g. as in Fig. 13e). An analysis of heat fluxes during the MHWs, could shed more light on the mechanisms driving the vertical structure and propagation of MHWs.

780 6 Conclusion

As the commonly used surface-only perspective on MHWs has so far not addressed the vertical structure of MHWs, we here extended the classical MHW definition (e.g. Hobday et al. (2016)) to incorporate the vertical dimension. This new perspective furnishes MHWs
with a vertical extent and furthermore allows for the study of the vertical propagation
behaviour of MHWs. We explored this new approach to study the vertical structure of
MHWs by using daily output from a high-resolution numerical hindcast simulation (19792019) in the Eastern Pacific.

We find that one third of the MHWs extend below the ML and are partially (up 788 to two thirds of their lifetime) undetectable at the sea surface. On average, the dMHWs 789 (deep-reaching MHWs) last five times longer than their ML-confined counterparts (sMHWs) 790 and are subsurface intensified, likely due to spatial displacements of sharp temperature 791 gradients within the thermocline. Initial tests further show that the characteristics of 792 MHWs diagnosed at the sea surface carry predictive skill regarding the presence/absence 793 of a deep-reaching MHW. These results suggest that there is potential for the detection 794 of deep-reaching MHWs from remote sensing. Nevertheless, model-derived relationships 795 between MHW surface characteristics and the subsurface MHW structure need to be put 796 to test with observations, by matching (satellite-derived) sea surface temperatures with 797 subsurface hydrographic data, for instance from Argo floats (Su et al., 2018). 798

The here used approach to define MHWs allowed for an explicit one-dimensional 799 tracking of the MHW signal. We find a variety in different vertical propagation behaviours. 800 While the ML-confined sMHWs and block-like dMHWs generally dominate, we find a 801 substantial fraction of dMHWs that shows a net upward or downward propagation or 802 even a multi-surfacing behaviour. The existence of shoaling and multi-surfacing MHWs 803 suggests, that studying subsurface warm anomalies can help to anticipate the develop-804 ment of surface MHWs, particularly in the tropical EP. However, further analyses will 805 be needed in order to understand under what conditions MHW signals can move between 806 the surface ML and the ocean interior. 807

808 Acknowledgments

This work has been funded by the Swiss National Science Foundation under grant agree-809 ment No. 175787 (Project X-EBUS: "Extreme Ocean Weather Events and their Role for 810 Ocean Biogeochemistry and Ecosystems in Eastern Boundary Upwelling Systems"), by 811 the Swiss Federal Institute of Technology Zurich (ETH Zurich), and by the European 812 Union's Horizon 2020 research and innovation programme under grant agreement No. 813 820989 (project COMFORT). We thank Fabio Benedetti for helpful inputs regarding the 814 clustering. We thank the Swiss Supercomputing Center (CSCS) and the ETH ITS Sci-815 entific IT Services for providing the infrastructure and support to run the model sim-816 ulations on Euler, the HPC cluster of ETH Zurich. All analyses are performed in python 817 and make use of the numpy, scipy, scikit-learn, pandas, astropy, xarray, sparse and mat-818 plotlib packages. The TAO/TRITON data was retrieved on Feb 23, 2022 from the Global 819 Tropical Moored Buoy Array Program (https://www.pmel.noaa.gov/tao/drupal/disdel/). 820 All data underlying the plots can be found under the following repository: https://doi 821 .org/10.3929/ethz-b-000612228 which can be accessed under the additional handle 822 http://hdl.handle.net/20.500.11850/612228. 823

824 References

- Alexander, M. A., & Deser, C. (1995). A Mechanism for the Recurrence of Win tertime Midlatitude SST Anomalies. Journal of Physical Oceanography, 25(1),
 122–137. doi: 10.1175/1520-0485(1995)025(0122:AMFTRO)2.0.CO;2
- Amaya, D. J., Jacox, M. G., Alexander, M. A., Scott, J. D., Deser, C., Capotondi,
 A., & Phillips, A. S. (2023). Bottom marine heatwaves along the continen tal shelves of North America. Nature Communications, 14(1), 1038. doi:
 10.1038/s41467-023-36567-0
- Amaya, D. J., Miller, A. J., Xie, S.-P., & Kosaka, Y. (2020). Physical drivers of

	the summer 2010 North Desific marine heatways. Nature Communications 11
833	1903 doi: 10.1038/s41467-020-15820-w
835	Bednaršek N Feelv B A Beck M W Glippa O Kanerva M & Engström-
836	Öst. J. (2018). El Niño-Related Thermal Stress Coupled With Upwelling-
837	Related Ocean Acidification Negatively Impacts Cellular to Population-Level
838	Responses in Pteropods Along the California Current System With Implica-
839	tions for Increased Bioenergetic Costs. Frontiers in Marine Science, 5, 486.
840	doi: 10.3389/fmars.2018.00486
841	Bianchi, D., & Mislan, K. A. S. (2016). Global patterns of diel vertical migration
842	times and velocities from acoustic data. Limnology and Oceanography, $61(1)$,
843	353–364. doi: 10.1002/lno.10219
844	Bond, N. A., Cronin, M. F., Freeland, H., & Mantua, N. (2015). Causes and im-
845	pacts of the 2014 warm anomaly in the NE Pacific. Geophysical Research Let-
846	ters, $42(9)$, $3414-3420$. doi: $10.1002/2015$ GL063306
847	Boyer, T. P., Garcia, H. E., Locarnini, R. A., Zweng, M. M., Mishonov,
848	A. V., Reagan, J. R., Smolyar, I. V. (2018). World Ocean Atlas
849	2018. NOAA National Centers for Environmental information. Dataset.
850	Burger F A Terbaar I & Frölicher T I (2022) Compound marine heatwayes
851	and ocean acidity extremes Nature Communications 13(1) 4722 doi: 10
853	.1038/s41467-022-32120-7
854	Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. <i>Com</i> -
855	munications in Statistics, $3(1)$, 1–27. doi: 10.1080/03610927408827101
856	Carton, J. A., Chepurin, G. A., & Chen, L. (2018). SODA3: A New Ocean Climate
857	Reanalysis. Journal of Climate, 31(17), 6967–6983. doi: 10.1175/JCLI-D-18
858	-0149.1
859	Cavole, L. M., Demko, A. M., Diner, R. E., Giddings, A., Koester, I., Pagniello,
860	C. M. L. S., Franks, P. J. S. (2016). Biological Impacts of the 2013–2015
861	Warm-Water Anomaly in the Northeast Pacific: Winners, Losers, and the
862	Future. Oceanography, 29(2), 273–285. doi: 10.5670/oceanog.2016.32
863	Cheung, W. W. L., Frolicher, T. L., Lam, V. W. Y., Oyinlola, M. A., Reygondeau,
864	G., Sumana, U. R., Wadnitz, C. C. C. (2021). Marine nigh temperature
865	Advances $7(40)$ esb0805 doi: 10.1126/sciedy.abb0805
800	Cohen I (1960) A Coefficient of Agreement for Nominal Scales Educational and
868	Psychological Measurement 20(1) 37–46 doi: 10.1177/001316446002000104
869	Collins, M., Sutherland, M., Bouwer, L., Cheong, SM., Frölicher, T., Jacot
870	Des Combes, H., Tibig, L. (2019). Extremes, Abrupt Changes and
871	Managing Risk. In HO. Pörtner et al. (Eds.), <i>IPCC Special Report on</i>
872	the Ocean and Cryosphere in a Changing Climate (First ed., pp. 589-655).
873	Cambridge, UK and New York, NY, USA: Cambridge University Press. doi:
874	10.1017/9781009157964
875	Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. <i>IEEE</i>
876	Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–
877	227. doi: 10.1109/TPAMI.1979.4766909
878	Dee, D. P., Uppala, S. M., Simmons, A. J., Berristord, P., Poli, P., Kobayashi, S.,
879	of the data assimilation system — <i>Quantarily</i> Journal of the Bould Meteorelasical
880	Society 137(656) 553-597 doi: 10.1002/gi 828
001	Desmet F Gruber N Köhn F F Münnich M & Voot M (2022) Tracking the
883	Space-Time Evolution of Ocean Acidification Extremes in the California Cur-
884	rent System and Northeast Pacific. Journal of Geophysical Research: Oceans.
885	127(5). doi: 10.1029/2021JC018159
886	Di Biagio, V., Cossarini, G., Salon, S., & Solidoro, C. (2020). Extreme event waves
887	in marine ecosystems: An application to Mediterranean Sea surface chloro-

000	phyll Biogeosciences 17(23) 5967-5988 doi: 10 5194/bg-17-5967-2020
888	Di Lorongo E di Montue N (2016) Multi vere possistence ef the $2014/15$ North
889	Di Lorenzo, E., & Mantua, N. (2010). Multi-year persistence of the 2014/15 North
890	Pacific marine neatwave. Nature Cumate Change, 0, 1042–1047. doi: 10.1058/
891	nclimate3082
892	Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Laut-
893	enbach, S. (2013). Collinearity: A review of methods to deal with it and a
894	simulation study evaluating their performance. Ecography, $36(1)$, 27–46. doi:
895	10.1111/j.1600-0587.2012.07348.x
896	Dussin, R., Barnier, B., Brodeau, L., & Molines, J. M. (2016). The Making
897	Of the Drakkar forcing set DFS5 (Tech. Rep.). Grenoble, France: LGGE.
898	https://www.drakkar-ocean.eu/publications/reports.
899	Elzahaby, Y., & Schaeffer, A. (2019). Observational Insight Into the Subsurface
900	Anomalies of Marine Heatwaves. Frontiers in Marine Science, 6, 745. doi: 10
901	.3389/fmars.2019.00745
902	Enfield, D. B. (2001). Evolution and historical perspective of the 1997–1998 El
903	Niño-Southern Oscillation event, Bulletin of Marine Science, 69(1), 7–25.
004	Freeland H & Ross T (2019) 'The Blob' - or how unusual were ocean temper-
904	atures in the Northeast Pacific during 2014-2018? Deen Sea Research Part I:
905	Ocean ographic Research Papers 150 103061 doi: 10.1016/j.dsr.2019.06.007
900	Frischknocht M. Münnich M. & Cruber N. (2015). Romote versus local influence
907	of FNSO on the California Current System — <i>Journal of Coonbusical Research</i> :
908	O_{course} 190(2) 1353-1374 doi: 10.1002/2014 IC010521
909	$\Gamma_{120}(2), 1555-1574.$ doi: 10.1002/20145C010551
910	richel marine Nature 560, 260, 264, dai: 10,1029/s41596,018,0292,0
911	giobal warming. Nature, 500, 500–504. doi: 10.1056/841560-016-0565-9
912	Grobelindemann, H., Ryan, S., Ummennoier, C. C., Martin, I., & Biastoch, A.
913	(2022). Marine Heatwayes and Their Depth Structures on the Northeast $U \subseteq G$ is a line fraction of G is line fraction o
914	U.S. Continental Shelf. Frontiers in Climate, $4, 857937$. doi: 10.3389/
015	
515	Control N Devel D W Devilation T I to Vent M (2021) Discuss characteristics
916	Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical ex-
916 917	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. <i>Nature</i>, 600, 395–407. doi: 10
916 917 918	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7
916 917 918 919	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
916 917 918 919 920	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Point of the 146(720). 1000-2010. Line 10.000/. i 2000.
916 917 918 919 920 921	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803
916 917 918 919 920 921 922	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C.,
916 917 918 919 920 921 922 923	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to back the DEC.
916 917 918 919 920 921 922 922 923 924	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi:
916 917 918 919 920 921 922 923 924 925	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014
916 917 918 919 920 921 922 923 924 925 926	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T.,
915 916 917 918 919 920 921 922 923 924 925 925 926 927	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine
915 916 917 918 919 920 921 922 923 924 925 926 926 927 928	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205
915 916 917 918 919 920 921 922 923 924 925 926 927 928 929	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled
915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022
915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022
915 916 917 918 919 920 921 922 923 924 925 926 925 926 927 928 929 930 931 932	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395-407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999-2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227-238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31 (2), 162-173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M.,
915 916 917 918 919 920 921 922 923 924 925 926 925 926 927 928 929 930 931 931 932	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022.109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of ma-
915 916 917 918 919 920 921 922 923 924 925 926 925 926 927 928 929 930 931 931 932 933	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022.109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi:
915 916 917 918 920 921 922 923 924 925 926 927 926 927 928 929 930 931 931 932 933 934	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Hobrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z
915 916 917 918 919 920 921 922 923 924 925 926 927 928 926 927 928 929 930 931 931 932 933 933 934 935 936	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and
915 916 917 918 920 921 922 923 924 925 926 927 928 927 928 929 930 931 932 933 934 935 936 937	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31 (2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and Magnitude of Subsurface Anomalies During the Northeast Pacific Blob as Mea-
915 916 917 918 920 921 922 923 924 925 926 927 928 927 928 929 930 931 933 933 934 933 934 935 936 937 938	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395-407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999-2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227-238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162-173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and Magnitude of Subsurface Anomalies During the Northeast Pacific Blob as Measured by Animal-Borne Sensors. Journal of Geophysical Research: Oceans,
915 916 917 918 920 921 922 923 924 925 926 925 926 925 926 927 928 929 930 931 930 931 932 933 934 935 936 937 938	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and Magnitude of Subsurface Anomalies During the Northeast Pacific Blob as Measured by Animal-Borne Sensors. Journal of Geophysical Research: Oceans, 127(7), e2021JC018356. doi: 10.1029/2021JC018356
915 916 917 918 920 921 922 923 924 925 926 926 927 928 929 930 931 932 933 931 932 933 933 934 935 936 937 938 939	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395-407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227-238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31(2), 162-173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and Magnitude of Subsurface Anomalies During the Northeast Pacific Blob as Measured by Animal-Borne Sensors. Journal of Geophysical Research: Oceans, 127(7), e2021JC018356. doi: 10.1029/2021JC018356 Holte, J., Talley, L. D., Gilson, J., & Roeminch, D. (2017). An Argo mixed layer cli-
915 916 917 918 919 920 921 922 923 924 925 926 927 926 927 928 929 930 931 931 932 933 933 934 935 935 936 937 938 939 939	 Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. Nature, 600, 395–407. doi: 10.1038/s41586-021-03981-7 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, JN. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999–2049. doi: 10.1002/qj.3803 Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. Progress in Oceanography, 141, 227–238. doi: 10.1016/j.pocean.2015.12.014 Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Smale, D. A. (2018). Categorizing and Naming Marine Heatwaves. Oceanography, 31 (2), 162–173. doi: 10.5670/oceanog.2018.205 Hofmann Elizondo, U., & Vogt, M. (2022). Individual-based modeling of shelled pteropods. Ecological Modelling, 468, 109944. doi: 10.1016/j.ecolmodel.2022 .109944 Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., Wernberg, T. (2019). A global assessment of marine heatwaves and their drivers. Nature Communications, 10, 2624. doi: 10.1038/s41467-019-10206-z Holser, R. R., Keates, T. R., Costa, D. P., & Edwards, C. A. (2022). Extent and Magnitude of Subsurface Anomalies During the Northeast Pacific Blob as Measured by Animal-Borne Sensors. Journal of Geophysical Research: Oceans, 127(7), e2021JC018356. doi: 10.1029/2021JC018356 Holte, J., Talley, L. D., Gilson, J., & Roemnich, D. (2017). An Argo mixed layer climatology and database. Geophysical Research Letters, 44 (11), 5618–5626. doi:

943	Hu, S., Li, S., Zhang, Y., Guan, C., Du, Y., Feng, M., Hu, D. (2021). Observed
944	vironmental Research Letters 16(10) 104024 doi: 10.1088/1748-9326/ac26f2
945	IPCC SBOCC (2022) The Ocean and Cryosphere in a Changing Climate: Special
940	Report of the Intergovernmental Panel on Climate Change (First ed)
947	bridge University Press doi: 10.1017/9781009157964
940	Jackson I M Johnson G C Dosser H V & Ross T (2018) Warming From
949	Recent Marine Heatwave Lingers in Deep British Columbia Fiord Geonbusical
950	Research Letters, $45(18)$, 9757–9764, doi: 10.1029/2018GL078971
952	Jacox, M. G., Alexander, M. A., Bograd, S. J., & Scott, J. D. (2020). Thermal dis-
952	placement by marine heatwayes. <i>Nature</i> , 584, 82–86, doi: 10.1038/s41586-020
954	-2534-z
955	Johnson, G. C., Hosoda, S., Javne, S. R., Oke, P. R., Riser, S. C., Roemmich,
956	D., Xu, J. (2022). Argo—Two Decades: Global Oceanography, Rev-
957	olutionized. Annual Review of Marine Science, 14(1), 379–403. doi:
958	10.1146/annurev-marine-022521-102008
959	Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analy-
960	sis. Educational and Psychological Measurement, 20(1), 141–151. doi: 10.1177/
961	001316446002000116
962	Kessler, W. S. (2006). The circulation of the eastern tropical Pacific: A review.
963	Progress in Oceanography, 69(2-4), 181–217. doi: 10.1016/j.pocean.2006.03
964	.009
965	Köhn, E. E., Münnich, M., Vogt, M., Desmet, F., & Gruber, N. (2022). Strong
966	Habitat Compression by Extreme Shoaling Events of Hypoxic Waters in
967	the Eastern Pacific. Journal of Geophysical Research: Oceans, 127(6),
968	e2022JC018429. doi: $10.1029/2022JC018429$
969	Large, W. G., & Yeager, S. (2004). Diurnal to decadal global forcing for ocean
970	and sea-ice models: The data sets and flux climatologies (Tech. Rep.). (No.
971	NCAR/TN-460+STR) University Corporation for Atmospheric Research. doi:
972	10.5065/D6KK98Q6
973	Liu, XD., Osher, S., & Chan, T. (1994). Weighted Essentially Non-oscillatory
974	Schemes. Journal of Computational Physics, 115(1), 200–212. doi: 10.1006/
975	Jcph.1994.1187
976	Marchesiello, P., McWilliams, J. C., & Shchepetkin, A. (2003). Equilibrium Struc-
977	ture and Dynamics of the California Current System. Journal of Physi-
978	cal Oceanography, $33(4)$, $753-783$. doi: $10.1175/1520-0485(2003)33(753)$.
979	ESADOI/2.0.00;2 McDhadan M. I. Busalasahi A. I. & Anderson D. I. T. (2010) A TOCA Batra
980	mcPhaden, M. J., Busalacchi, A. J., & Anderson, D. L. I. (2010). A TOGA Retro-
981	Spectree. Occurring uping, $25(5)$, 80–105. doi: 10.3070/occurring.2010.20
982	tomporature patterns Lowrad of Coophysical Research 75(20) 5052 5055
983	doi: $10.1029/IC075i030p05952$
984	Okumura V M & Deser C (2010) Asymmetry in the Duration of El Niño and
985	La Niña Lournal of Climate $23(21)$ 5826–5843 doi: 10.1175/2010 ICLI3502
980	1
088	Oliver F. C. J. Benthuvsen, J. A. Darmaraki, S. Donat, M. G. Hobday, A. J.
989	Holbrook, N. J Sen Gupta, A. (2021). Marine Heatwayes. Annual Re-
990	view of Marine Science, 13(1), 313–342. doi: 10.1146/annurev-marine-032720
991	-095144
992	Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexan-
993	der, L. V., Wernberg, T. (2018). Longer and more frequent marine
994	heatwaves over the past century. <i>Nature Communications</i> , 9, 1324. doi:
995	10.1038/s41467-018-03732-9
996	Palacios, D. M., & Bograd, S. J. (2005). A census of Tehuantepec and Papagayo
997	eddies in the northeastern tropical Pacific. Geophysical Research Letters, 32,

998	L23606. doi: 10.1029/2005GL024324
999	Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,
1000	Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. the Journal of
1001	machine Learning research, 12, 2825–2830. doi: 10.48550/ARXIV.1201.0490
1002	Pilo, G. S., Holbrook, N. J., Kiss, A. E., & Hogg, A. M. (2019). Sensitivity of Ma-
1003	rine Heatwave Metrics to Ocean Model Resolution. Geophysical Research Let-
1004	ters, 46(24), 14604–14612. doi: 10.1029/2019GL084928
1005	Pörtner, H. O., & Farrell, A. P. (2008). Physiology and Climate Change. Science,
1006	322(5902), 690–692. doi: 10.1126/science.1163156
1007	Purkiani, K., Haeckel, M., Haalboom, S., Schmidt, K., Urban, P., Gazis, IZ.,
1008	Vink, A. (2022). Impact of a long-lived anticyclonic mesoscale eddy
1009	on seawater anomalies in the northeastern tropical Pacific Ocean: A com-
1010	posite analysis from hydrographic measurements, sea level altimetry data,
1011	and reanalysis model products. Ocean Science, 18(4), 1163–1181. doi:
1012	10.5194 / os- 18 - 1163 - 2022
1013	Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury,
1014	O. (2013). Dynamic biogeochemical provinces in the global ocean. Global
1015	Biogeochemical Cycles, 27(4), 1046–1058. doi: 10.1002/gbc.20089
1016	Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax,
1017	M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Sur-
1018	face Temperature. Journal of Climate, 20(22), 5473–5496. doi: 10.1175/
1019	2007JCLI1824.1
1020	Ryan, S., Ummenhofer, C. C., Gawarkiewicz, G., Wagner, P., Scheinert, M., Bias-
1021	toch, A., & Böning, C. W. (2021). Depth Structure of Ningaloo Niño/Niña
1022	Events and Associated Drivers. Journal of Climate, 34(5), 1767–1788. doi:
1023	10.1175/JCLI-D-19-1020.1
1024	Scannell, H. A., Abernathey, R., Busecke, J., Gagne, D. J., Thompson, L., & Whitt,
1025	D. (2021). Ocetrac. Zenodo. doi: 10.5281/ZENODO.5102928
1026	Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M., & Riser, S. C. (2020).
1027	Subsurface Evolution and Persistence of Marine Heatwaves in the North-
1028	east Pacific. Geophysical Research Letters, $47(23)$, e2020GL090548. doi:
1029	10.1029/2020GL090548
1030	Schaeffer, A., & Roughan, M. (2017). Subsurface intensification of marine heatwaves
1031	off southeastern Australia: The role of stratification and local winds. Geophysi-
1032	cal Research Letters, 44(10), 5025–5033. doi: 10.1002/2017GL073714
1033	Seibel, B. A., & Birk, M. A. (2022). Unique thermal sensitivity imposes a cold-
1034	water energetic barrier for vertical migrators. Nature Climate Change, 12,
1035	1052-1058. doi: $10.1038/s41558-022-01491-6$
1036	Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E.,
1037	Alexander, L. V., Smale, D. A. (2020). Drivers and impacts of the
1038	most extreme marine heatwave events. Scientific Reports, 10, 19359. doi:
1039	10.1038/s41598-020-75445-3
1040	Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic model-
1041	ing system (ROMS): A split-explicit, free-surface, topography-following-
1042	coordinate oceanic model. Ocean Modelling, $9(4)$, $347-404$. doi: 10.1016/
1043	j.ocemod.2004.08.002
1044	Shu, CW. (1998). Essentially non-oscillatory and weighted essentially non-
1045	oscillatory schemes for hyperbolic conservation laws. In Advanced Numerical
1046	Approximation of Nonlinear Hyperbolic Equations (Vol. 1697, pp. 325–432).
1047	Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/BFb0096355
1048	Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub,
1049	S. C., Moore, P. J. (2019). Marine heatwaves threaten global biodiversity
1050	and the provision of ecosystem services. Nature Climate Change, 9, 306–312.
1051	doi: 10.1038/s41558-019-0412-1
1052	Smith, K. E., Burrows, M. T., Hobday, A. J., King, N. G., Moore, P. J., Sen Gupta,

1053	A., Smale, D. A. (2023). Biological Impacts of Marine Heatwaves. An-
1054	nual Review of Marine Science, 15(1), 119–145. doi: 10.1146/annurev-marine
1055	-032122-121437
1056	Song, Y., & Haidvogel, D. (1994). A Semi-implicit Ocean Circulation Model Using
1057	a Generalized Topography-Following Coordinate System. Journal of Computa-
1058	tional Physics, 115(1), 228–244. doi: 10.1006/jcph.1994.1189
1059	Steinberg, D. K., & Landry, M. R. (2017). Zooplankton and the Ocean Carbon Cy-
1060	cle. Annual Review of Marine Science, 9(1), 413–444. doi: 10.1146/annurev
1061	-marine-010814-015924
1062	Su, H., Huang, L., Li, W., Yang, X., & Yan, XH. (2018). Retrieving Ocean Subsur-
1063	face Temperature Using a Satellite-Based Geographically Weighted Regression
1064	Model. Journal of Geophysical Research: Oceans, 123(8), 5180–5193. doi:
1065	10.1029/2018JC014246
1066	Tak, YJ., Song, H., & Cho, YK. (2021). Impact of the reemergence of North Pa-
1067	cific subtropical mode water on the multi-year modulation of marine heatwaves
1068	in the North Pacific Ocean during winter and early spring. <i>Environmental</i>
1069	Research Letters, $16(7)$, 074036. doi: $10.1088/1748-9326/ac0cad$
1070	Vogt, L., Burger, F. A., Griffies, S. M., & Frölicher, T. L. (2022). Local Drivers of
1071	Marine Heatwaves: A Global Analysis With an Earth System Model. Frontiers
1072	in Climate, 4, 847995. doi: 10.3389/fclim.2022.847995
1073	Wei, X., Li, KY., Kilpatrick, T., Wang, M., & Xie, SP. (2021). Large-Scale Con-
1074	ditions for the Record-Setting Southern California Marine Heatwave of August
1075	2018. Geophysical Research Letters, 48(7). doi: 10.1029/2020GL091803
1076	Wernberg, T., Smale, D. A., Tuya, F., Thomsen, M. S., Langlois, T. J., de Bettig-
1077	nies, T., Rousseaux, C. S. (2013). An extreme climatic event alters marine
1078	ecosystem structure in a global biodiversity hotspot. Nature Climate Change,
1079	3, 78-82. doi: $10.1038/nclimate1627$
1080	Wishner, K. F., Outram, D. M., Seibel, B. A., Daly, K. L., & Williams, R. L. (2013).
1081	Zooplankton in the eastern tropical north Pacific: Boundary effects of oxygen
1082	minimum zone expansion. Deep Sea Research Part I: Oceanographic Research

¹⁰⁸³ Papers, 79, 122–140. doi: 10.1016/j.dsr.2013.05.012

Supporting Information for "On the vertical structure and propagation of marine heatwaves in the Eastern Pacific"

Eike E. Köhn¹, Meike Vogt¹, Matthias Münnich¹, Nicolas Gruber^{1,2}

¹Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland

 $^2\mathrm{Center}$ for Climate Systems Modeling, ETH Zurich, 8092 Zurich, Switzerland

Corresponding author: E. E. Köhn, Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland (eike.koehn@usys.ethz.ch)

Contents of this file

- 1. Model data description and evaluation
 - 1.1 Vertical regridding of model output
 - 1.2 Evaluation of the mean state
 - 1.3 Evaluation of the temperature variability
 - 1.4 Evaluation of temperature trends
 - 1.5 Evaluation of subsurface temperature extremes
- 2. Clustering of MHWs
 - 2.1 Clustering methodology
 - 2.2 Clustering results
- 3. Sensitivity Analyses
 - 3.1 MHW characteristics for different MHW detection cases
 - 3.2 Sensitivity of Boolean arrays to morphological operations
 - 3.3 Sensitivity of MHW clustering
- 4. Linking surface-only MHWs to their associated subsurface structure
- 5. Figures S1 to S23

1. Model data description and evaluation

1.1. Vertical regridding of model output

To account for the bathymetry following native model grid, we perform vertical regridding before calculating horizontal averages when horizontally coarsening/downsampling the model output (Sec. 2.1 of main text). We therefore regrid the model output from the bathymetry following s-level coordinates to z-levels, that is fixed depth levels, within the upper 500 m. We thereby choose the the z-levels to closely follow the vertical spacing in the maximally stretched s-level coordinate system. We therefore let the z-levels closely follow the s-levels at a location of the deepest model bathymetry, which is set to 6500 m (Fig. S2). This leads to 37 z-levels with a 5 m resolution in the upper 100 m and a gradual increases to a 50 m resolution below 350 m. By fitting the z-levels to the maximally stretched s-level coordinates, the regridding to z-levels does not add additional intermediate depth levels. Vertical temperature profiles from locations where the ocean bottom is shallower than 6500 m have a higher resolution in s-levels than in z-levels in the upper 500 m.

To test for the sensitivity of our results regarding the vertical resolution, we additionally coarsen the vertical z-level grid, by using only every second z-level for the MHW analysis, leading to 19 z-levels (Fig. S2, Section 2.6 of main text).

1.2. Evaluation of the mean state

The hindcast mean sea surface temperature (SST, 1979-2019) has a small bias of -0.02 °C compared to World Ocean Atlas 2018 (WOA2018, Boyer et al. (2018)) averaged across the full study area (Fig. S3). In the equatorial Eastern Pacific (EP) and Peruvian upwelling system, model SST values are locally up to 1.5 °C warmer than observed. The temperature biases are stronger in the subsurface with pronounced regional differences (Fig. S3). As such, at 200 m (400 m) depth, the tropical EP between 20°S and 20°N is on average too cold by 1.47 °C (0.99 °C), while the subpolar North Pacific north of 30°N is too warm by 0.41 °C (1.17 °C). At 100 m, we find additional warm biases in the eastern tropical North Pacific and the subtropical South Pacific with mean anomalies of up to 1.5 °C. This warm bias is linked to a slightly too deep mixed layer (and thermocline) in the subtropical gyres ($\sim 10-30$ m, Fig. S4a). Across the entire EP, the modeled mixed layer depth (MLD) is on average 12.4 m deeper than observed (Holte et al., 2017). In the subpolar NP, ROMS simulates a MLD that is 4.5 m too shallow. ROMS accurately reproduces the sea surface height (SSH) field structure (spatial correlation of the temporal mean SSH fields of 0.98, Fig. S4b), indicating that the mean geostrophic currents are well reproduced by the model.

1.3. Evaluation of the temperature variability

We evaluate the variability of the SST anomalies, by analysing their amplitudes and persistence in ROMS. The anomalies are thereby calculated relative to the seasonally varying SST climatology calculated for the period 1982–2011 (Fig. S5). For the evaluation, we compare the standard deviation and the autocorrelation e-decay time scale of the daily SST anomalies and compare the results to observational OISSTv2 data (Reynolds et al., 2007). The model reproduces the SST anomaly amplitudes reasonably well with maximum standard deviations of up to 1.5 °C in the tropical EP (Fig. S5a,b). In the Peruvian upwelling system the model overestimates the generally high standard deviation by about 0.5 °C. The persistence of SST anomalies, or the time scale of SST anomaly variability is also well reproduced by the model with generally longer e-decay times in the tropical EP, the subtropical NP gyre and the subpolar NP (Fig. S5c,d). However, the model has a tendency to overestimate the e-decay times, indicating longer persistence of SST anomalies. This is in agreement with the detection of generally longer but fewer marine heatwaves (MHWs, Sec. 3 of the main text, Oliver et al. (2021)).

1.4. Evaluation of temperature trends

We calculate trends in simulated the temperature field at the surface, at 100 m, 200 m, 310 m and 400 m depth over the full hindcast period (1979-2019, Fig. S6).

We find a general simulated cooling trend of around -0.03 °C yr⁻¹ throughout the tropical EP across all depths (Fig. S6a-e). The cooling trend is most pronounced at 100 m depth (up to -0.05 °C yr⁻¹), that is in the strong thermocline of the tropical EP. This cooling trend is stronger than comparable temperature trends calculated from the monthly resolved EN4 data set from 1981–2019 (Good, Martin, and Rayner (2013), Fig. S6f-j). The analysis of temperature time series shows that the calculation of long-term trends in the tropical EP is influenced by interannual temperature variability (Fig. S7). For instance, in the northern Humboldt Current System (location *c* indicated in Figure S6), the occurrence of fewer strong El Niño related warming events after the year 2000 compared to the preceding years, manifests in the diagnosis of a general cooling of the EP.

Below 200 m depth the model shows a pronounced cooling between $25^{\circ}-40^{\circ}$ N and $160^{\circ}-120^{\circ}$ W, which is not found in the EN4 data set from 1981–2019 (Good et al. (2013), Fig. S6c-e,h-j). The temperature time series within this region (chosen location *a*, Figure S6), shows that this cooling trend is relatively independent from interannual variability, but manifests mostly at the beginning of the hindcast with the strongest temperature decrease in the first few years (Fig. S7a). This suggests some model adjustment occurring when switching from the spin-up period to the beginning of the model hindcast. Such spurious trends in the simulated temperature field have the potential to affect the detection of extreme temperatures. We however limit the impact of this spurious model trend by calculating the temperature thresholds based on the years 1982–2011 (Sec. 2.2.1 of main

text) and thus by not taking the first three hindcast years into account. Nevertheless, as the MHW detection is performed over the full analysis time period, the spurious model drift leads locally to higher numbers of subsurface extreme temperatures at the beginning of the hindcast (Fig. S8, Fig. S9). To assess the influence of the first few hindcast years on the composite MHW characteristics, we conduct a sensitivity analysis and calculate the MHW characteristics only for the time period 1982–2019 (case G in Section 4 on sensitivities analyses). The results of the sensitivity analysis however suggest only limited impacts on the statistics of the one-dimensional MHW properties (see Fig. S19).

1.5. Evaluation of subsurface temperature extremes

Only few observations exist, documenting the structure and evolution of MHWs in the subsurface. This complicates an evaluation of the model's capability to realistically simulate subsurface temperature extremes. As we study MHWs in the EP between 1979– 2019 we are able to compare our model results to the subsurface evolution of the "Blob" as described by Scannell, Johnson, Thompson, Lyman, and Riser (2020) (see Sec. 4.3 of main text). Furthermore, we can make use of the TAO/TRITON mooring array in the tropical Pacific (McPhaden et al., 2010), which has gathered temperature time series at multiple locations within the upper 500–750 m of the water column. This type of temperature time series have already been used to calculate MHW characteristics in the western tropical Pacific (Hu et al., 2021).

Here we use the available temperature time series from five equatorial TAO/TRITON moorings east of 155°W, to evaluate the models performance in reproducing the vertical structure of temperature anomalies in the equatorial EP. As a test case we therefore focus on the 1997–1998 El Niño event (Fig. S10), which was an intense and long lasting MHW in the tropical EP (Sen Gupta et al., 2020). For the hindcast simulation and the TAO/TRITON array, we calculate temperature anomalies relative to the climatology, which we calculate following Hobday et al. (2016). For the hindcast, the climatology is calculated over the 1982–2011 period, as throughout the main study. For the TAO/TRITON data, we use all available data to calculate the climatologies, as the mooring data generally covers varying time spans between 1980 and 2022, but generally amounts to available data of around 20–25 years. In Figure S10, we show the temperature anomalies at all five mooring locations from TAO/TRITON and for the corresponding horizontal grid point

in the hindcast simulation. Despite some gaps in the observational data, a comparison between model and observations is feasible. In the observations and the model simulation, the strongest temperature anomalies during the 1997–1998 El Niño are generally found in the subsurface within the upper 200 m (Fig. S10). The model is able to reproduce the initial warming anomalies in the subsurface of the equatorial Pacific and shows an upward migration of the warm anomalies that is similar to observations. Similarly, the model reproduces how warm anomalies subside again first in the subsurface and turn to cold anomalies with the onset of the subsequent La Niña. While the model slightly underestimates the anomalous warming, especially towards the eastern mooring locations, this comparison gives us confidence, that the model realistically reproduces subsurface warming events in the equatorial Pacific.

2. Clustering of MHWs

2.1. Clustering methodology

We cluster the MHWs using a k-means clustering algorithm (Pedregosa et al., 2011), based on a set of their characteristics (see Sec. 2.5 of main text.). To eliminate collinearity between the clustering features (Fig. S11), we perform a principal component analysis (PCA). We find only the first three principal components (PCs) to have eigenvalues above one, which together explain 77.5% of the variance (Fig. S12). Following Kaiser's rule (Kaiser, 1960), we use the standardized first three PCs for the k-means clustering. We find the optimal number of clusters to be 4 (Fig. S13). This number of clusters maximizes the Calinski-Habarasz score (Caliński & Harabasz, 1974), minimizes the Davies-Bouldin score (Davies & Bouldin, 1979) and marks a clear transition in the within-cluster sum of squared distances (i.e., the Elbow method).

2.2. Clustering results

As described in the main text, we identify four different clusters of deep-reaching MHW (dMHW) vertical propagation types: a) *block-like*, b) *deepening*, c) *shoaling*, and d) *multi-surfacing* MHWs (Fig. 11 of main text). Despite the overall common shape for all extremes, the within-cluster variability of dMHWs is still substantial. Figure S14 shows 4 randomly selected MHW examples for each cluster. Figure S15 shows the clustering results for the MHW characteristics that were selected to be used in the principal component analysis prior to clustering. Both figures highlight the high variability of MHW characteristics for the MHWs (sMHWs), which were not considered for the clustering.

Fig. S16 shows the clustering results, but for the principal components on which the the clustering was finally performed.

Lastly, Fig. S17 shows the clustering results for secondary MHW characteristics, that is characteristics that are not directly used for clustering, such as the simple column duration, surface duration, the mean depth relative to the MLD, the mean depth relative to the surface, and the mean fraction of the MHW being present in the mixed layer (ML). Again for comparison, the ML-confined sMHW properties are also shown in Figure S17. We find that the distinction between the different clusters is also reflected in these secondary characteristics.

3. Sensitivity Analyses

3.1. MHW characteristics for different MHW detection cases

As outlined in Section 2.6 of the main text, we test for the sensitivity of the detection and characteristics of MHWs regarding seven different methodological choices (see Fig. S18 for an overview). In each sensitivity case we solely alter one methodological choice, while keeping all other choices as in the reference case, which is used throughout the main manuscript.

Figure S19 shows mapped (composite) MHW characteristics for the reference case and all seven sensitivity cases. All characteristics (number of surface-only MHWs per year, number of all MHWs (dMHWs and sMHWs) per year, percentage of MHWs that are dMHWs, mean dMHW duration, mean dMHW depth below the MLD, and the composite maximum of the maximum dMHW intensity) show only little variation between the reference and sensitivity cases. The most striking difference occurs for the sensitivity case D, with fewer, slightly shorter and slightly shallower MHWs than in the reference case (Fig. S19e,m,C,L). However, this is to be expected, as the threshold for the extreme temperature detection is elevated to the 95th percentile in this sensitivity case, leading to substantially less (\sim 50 %) grid cells harboring extreme conditions.

3.2. Sensitivity of Boolean array B to morphological operations

Applying the morphological operations to smooth the Boolean array **B**, as outlined in Section 2.2.1 of the main text, has the potential to affect the overall number of detected extreme days. Figure S20 therefore compares the number of extreme days in the unsmoothed and smoothed Boolean array **B** at different depth levels (surface, 50 m, 250 m and 500 m depth). For the surface, we perform the same comparison also for the obser-

vational SST data set. We find that in the upper ocean, the effect of the morphological operations on the total number of MHW days is relatively weak. Below 250 m depth, the tropical EP is however marked by local increases of up to more than 20% in the total number of extreme days (Fig. S20). This implies that the unsmoothed extreme signals in the subsurface tropical EP are often interrupted by short non-extreme periods (shorter than five days). The morphological operations fill these gaps. Still, sensitivity case F (Fig. S18) shows only minor differences in the analyzed MHW characteristics between the smoothed and unsmoothed case (Fig. S19).

3.3. Sensitivity of MHW clustering

We test the robustness of the MHW clusters which we obtained from the k-means clustering described in Section 2.5 of the main text. We therefore analyze how sensitive the clusters are with regards to a) an omission of 10–99 % of MHWs and b) the omission of individual MHW characteristics feeding into the principal component anlysis prior to clustering (Fig. S21). For each case we compare the cluster agreement for the labeled MHWs between the standard case and the sensitivity case using Cohen's Kappa coefficient (κ , Cohen (1960)). $\kappa = 1$ indicates perfect agreement, while $\kappa = 0$ suggests agreement based on random labeling.

As the MHW omission is random, we repeat each sensitivity case (10 % to 99 % of MHWs omitted) ten times, to avoid accidental high agreement between the reference and the respective sensitivity case. As we detect in total 1 400 170 MHWs, an omission of 99 % of MHWs still leaves ~14 000 MHWs for clustering. For all analyzed MHW omission cases we find on average very high agreements between the cluster assignment for the reference and sensitivity cases ($\kappa \geq 0.99$), indicating high robustness of the clustering. The clustering

consistently assigns the MHWs to the same clusters even under the omission of 99 % of the detected MHWs.

The respective elimination of one of the six MHW characteristics that feed into the principal component analysis, has a somewhat bigger impact on the cluster assignment agreement. In most cases, $\kappa > 0.7$. Only for the omission of the "start delay at the surface" and the "early end at the surface" (normalized with the MHW column duration, see Fig. 2), κ drops to around 0.65, indicating the important role of these characteristics in the clustering procedure. Nevertheless, as all sensitivity cases yield $\kappa > 0.6$, we deem the results of the clustering to be also robust to the general choice of MHW characteristics.

4. Linking surface-only MHWs to their associated subsurface structure

Figure S22 shows two-dimensional histograms between surface-only MHW characteristics and the corresponding characteristics diagnosed for the associated water column MHWs. Overall, the surface-only MHW characteristics are only moderately correlated with the corresponding characteristics of the MHWs in the water column. For instance, the surface-only MHW duration is correlated with a Pearson (Spearman) correlation of 0.46 (0.63) with the associated MHW column duration (Fig. S22a). Correlations for the severity are similar (Fig. S22c) and somewhat lower for the maximum intensity with $r_{\text{Pearson}} = 0.36$ ($r_{\text{Spearman}} = 0.55$, Fig. S22b). Hence, despite these weak to moderate correlations, longer lasting, more intense and more severe surface-only MHWs are generally associated with longer lasting, more intense and more severe MHWs in the water column, respectively. Yet, the two-dimensional histograms show that the longest lasting MHWs in the water column with durations of more than 1000 days are associated with relatively short-lived surface-only MHWs (less than 100 days, Fig. S22a). Similarly,

the most intense and severe MHWs in the water column are associated with relatively moderate surface-only MHWs. These results show the challenges in estimating individual subsurface MHW characteristics based on the surface MHW signature.

In the main text (Section 4.4), we explore the possibility of identifying dMHWs based on the properties of their associated surface-only MHWs. We therefore fit logistic regression models between the binary distinction of the associated MHWs into sMHWs and dMHWs and the associated surface-only MHW properties, such as duration, maximum intensity and severity. Figure 10 of the main text shows the model fit and the predictive capacities of the statistical models fitted for all detected MHWs in the EP. Figure S23 shows the same analysis but for individual subregions of the EP. As such, we separate the EP into eight different subregions by drawing separation lines along 5°S, 5°N, 23°N and 40°N. We distinguish between open ocean regions and coastal regions by using the 700 km distance from the coast isoline. Only in the equatorial Pacific and the subpolar North Pacific (west of 135°W) we do not consider a separate coastal region (Fig. S23). We find that individual regions show higher predictive capacity for the subsurface MHW structure than across the entire EP (compare with Fig. 10 of the main text), indicated by the confusion matrices. While all regions show correct dMHW predictions in around 60–75%, highest predictive capacity exists in the Equatorial Pacific. There, 77 % (76 %) of MHWs are correctly predicted based on surface-only MHW severity (duration).

References

Boyer, T. P., Garcia, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., ... Smolyar, I. V. (2018). World Ocean Atlas 2018. NOAA National Centers for Environmental Information. Dataset. https://accession.nodc.noaa.gov/NCEI-

- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communi*cations in Statistics, 3(1), 1–27. doi: 10.1080/03610927408827101
- CMEMS. (2019). Global ocean gridded L4 sea surface heights and derived variables reprocessed (1993-ongoing). Copernicus Marine Environment Monitoring Service. doi: 10.48670/moi-00148
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37–46. doi: 10.1177/001316446002000104
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1(2), 224–227. doi: 10.1109/TPAMI.1979.4766909
- Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12), 6704–6716. doi: 10 .1002/2013JC009067
- Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver,
 E. C. J., ... Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, 141, 227–238. doi: 10.1016/j.pocean.2015.12
 .014
- Holte, J., Talley, L. D., Gilson, J., & Roemmich, D. (2017). An Argo mixed layer climatology and database. *Geophysical Research Letters*, 44(11), 5618–5626. doi: 10.1002/2017GL073426
- Hu, S., Li, S., Zhang, Y., Guan, C., Du, Y., Feng, M., ... Hu, D. (2021). Observed strong

subsurface marine heatwaves in the tropical western Pacific Ocean. *Environmental Research Letters*, 16(10), 104024. doi: 10.1088/1748-9326/ac26f2

- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement, 20(1), 141–151. doi: 10.1177/ 001316446002000116
- McPhaden, M. J., Busalacchi, A. J., & Anderson, D. L. T. (2010). A TOGA Retrospective. Oceanography, 23(3), 86–103. doi: 10.5670/oceanog.2010.26
- Oliver, E. C. J., Benthuysen, J. A., Darmaraki, S., Donat, M. G., Hobday, A. J., Holbrook, N. J., ... Sen Gupta, A. (2021). Marine Heatwaves. Annual Review of Marine Science, 13(1), 313–342. doi: 10.1146/annurev-marine-032720-095144
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825–2830. doi: 10.48550/ARXIV.1201.0490
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily High-Resolution-Blended Analyses for Sea Surface Temperature. *Jour*nal of Climate, 20(22), 5473–5496. doi: 10.1175/2007JCLI1824.1
- Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M., & Riser, S. C. (2020). Subsurface Evolution and Persistence of Marine Heatwaves in the Northeast Pacific. *Geophysical Research Letters*, 47(23), e2020GL090548. doi: 10.1029/2020GL090548
- Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E., Alexander,
 L. V., ... Smale, D. A. (2020). Drivers and impacts of the most extreme marine heatwave events. *Scientific Reports*, 10, 19359. doi: 10.1038/s41598-020-75445-3



Figure S1. Telescopic grid of the humpac15 Pacific basin setup. Panel a shows the model grid dimension (outlined by black lines) and indicates the grid structure by representing every 50thgrid point by the grey lines. The color shows the grid cell size (Δx) , with finest resolution off Peru and coarsest resolution off Australia. Panel b shows the ratio of the model grid cell size and the first baroclinic Rossby radius of deformation (L_{Ro}) , calculated using a gravity wave speed of $c = 2.4 \text{ m s}^{-1}$. Values smaller (larger) than 1 indicate finer (coarser) model resolution than the deformation radius. Orange, green, and magenta lines indicate value of 1 for c = 2.0, 2.4, and 3.0 m s^{-1} , respectively.



Figure S2. Vertical regridding of model output from terrain following coordinates (s-levels) to fixed z-levels. Panel a shows the vertical levels as a function of depth. The black line shows the model native vertical grid (s-levels) at a location of maximum water depth, i.e. at a maximally stretched vertical grid. The blue and orange lines show the "full" and "coarse" z-level grid, respectively. In the "full" grid, 37 z-levels are chosen to closely follow the maximally stretched s-level depths. The "coarse" z-level grid takes only every second z-level of the "full" grid, leading to 19 z-levels. The different vertical grids are visualized in panel b.



Figure S3. Mean temperature biases of the ROMS model hindcast (1979-2019) at a) the surface, b) 100 m, c) 200 m and d) 400 m depth. Biases are calculated as ROMS minus World Ocean Atlas 2018 (Boyer et al., 2018).



Figure S4. Model evaluation of the mixed layer depth (MLD) and sea surface height (SSH). Panel a shows the hindcast averaged MLD in black contour lines as well as the hindcast averaged MLD bias (MLD') in color (compared to Holte et al. (2017). Panel b compares the hindcast averaged model SSH to satellite altimetry observations (CMEMS, 2019).



Figure S5. Maps showing the standard deviation (upper) and the auto-correlation edecay time (lower) of the de-trended SST anomaly (SST') field at the ocean surface. Left panels show the metrics based on observations (OISSTv2 data, Reynolds et al. (2007)), right panels for the ROMS hindcast. Temperature anomalies are calculated relative to the daily climatology of sea surface temperatures calculated over the time period 1982-2011 following Hobday et al. (2016).



Figure S6. Temperature trends in the ROMS hindcast. Upper row (panels a-e) shows mapped temperature trends calculated over the full hindcast (1979–2019) at different depth levels (surface, 100 m, 200 m, 310 m, 400 m depth). Lower row (panels f-j) shows the same, but derived from the monthly EN4 data set between 1981 and 2019. Shown depths are not identical as in ROMS, but chosen as close as possible. Letters a-d in all panels indicate locations for which temperature time series are shown in Figure S7.



Figure S7. Temperature time series from the daily ROMS hindcast (panel a, 1979-2019) and the monthly EN4 data set (panel b, 1981-2019) at the four locations a-d indicated in Figure S6. For both data sets the time series are shown at for five depths corresponding to the depths shown in Figure S6.



Figure S8. Panels show where extreme temperatures are detected at each depth level on the first day of the hindcast, i.e. Jan 1st, 1979.

X - 24



Figure S9. Time series of EP area fraction covered by a) all extreme grid cells, b) extreme grid cells associated with MHWs with a surface imprint, c) extreme grid cells associated with MHWs without a surface imprint (which are thus discarded in this study).



Figure S10. Temperature anomalies relative to the climatology during the 1997 El Niño event as simulated in the ROMS hindcast (top row) and as recorded with the TAO/TRITON array (bottom row) at five different mooring locations along the equator east of 155° W. The panels are arranged by their geographical locations from west to east. Temperature anomalies are calculated relative to the climatology following Hobday et al. (2016). For the hindcast, the climatology is calculated over the 1982–2011 period (as throughout the main study). For the TAO/TRITON data, we use all available data to calculate the climatologies, as the mooring data generally covers only ~20–25 years.

ratio: surf. to col. duration -	1.00	-0.50	-0.59	0.03	-0.10	-0.05
ratio: start delay to col. duration -	-0.50	1.00	-0.10	0.60	-0.21	-0.16
ratio: early end to col. duration -	-0.59	-0.10	1.00	-0.60	-0.18	-0.20
vert. propagation (lin. slope)-	0.03	0.60	-0.60	1.00	-0.01	0.03
vert. prop. curvature (quadratic term)-	-0.10	-0.21	-0.18	-0.01	1.00	0.37
ratio: max. gap at surf. to col. duration -	-0.05	-0.16	-0.20	0.03	0.37	1.00
-1 0 1 rank correlation	ratio: surf. to col. duration -	ratio: start delay to col. duration -	ratio: early end to col. duration	vert. propagation (lin. slope) -	vert. prop. curvature (quadratic term) -	ratio: max. gap at surf. to col. duration -

Figure S11. Correlation matrix of primary features, showing rank correlations between all MHW characteristics feeding into the principal component analysis.



Figure S12. Results of the principal component analysis. Panel a) shows the eigenvalues associated with each individual principal component (Scree plot). Panel b) shows the explained variance by each principal component (PC).



Figure S13. Analysis of optimal number of clusters. Clustering is repeatedly performed using 2–19 clusters. For each chosen number of clusters, the Calinski-Harabasz score (panel a), the Davies-Bouldin score (panel b) and the within-cluster sum of squared distances (Elbow method, panel c) is calculated.



Figure S14. Four examples of the time-depth MHW structures for each cluster, i.e., *block-like* (first column), *deepening* (second column), *shoaling* (third column), and *multi-surfacing* (fourth column) MHWs. Colored areas indicate extreme grid cells. Dashed black line shows linear fit to the temporal evolution of the upper MHW boundary.



Figure S15. Cluster results. Violinplots show the six MHW characteristics that were used in the principal component analysis based on which the four different clusters were identified (colorcoding/cluster numbering as in Figure 11). Black dots and white lines indicate the median and the interquartile range of the distribution, respectively. For comparison, the corresponding distribution and statistics for the mixed layer-confined sMHWs is shown (denoted by X), even though they are not considered in the clustering.



Figure S16. Cluster results. Violinplots show the three clustered PCs (colorcoding/cluster numbering as in Figure 11). Black dots and white lines indicate the median and the interquartile range of the distribution, respectively.



Figure S17. Cluster results. Violinplots show the clustering results for five further MHW characteristics that were not used in the principal component analysis feeding into the clustering (colorcoding/cluster numbering as in Figure 11). Black dots and white lines indicate the median and the interquartile range of the distribution, respectively. For comparison, the corresponding distribution and statistics for the mixed layer-confined sMHWs is shown (denoted by X), even though they are not considered in the clustering.

		Regridded model output			Thresh./c	lim. array	Bool. array	All arrays
	CASE ID	Vertical depth levels	rtical Horizontal Downsampling h levels downsampling method		Threshold percentile	Threshold period	Opening/ closing	Analysis period
Reference case	0	37-zlevels (full)	3x3	meanpool	90th percentile	1982-2011	5days	1979-2019
	А	19-zlevels (coarse)	3x3	meanpool	90th	1982-2011	5days	1979-2019
Regridding cases	В	37-zlevels	5x5	meanpool	90th	1982-2011	5days	1979-2019
	С	37-zlevels	3x3	maxpool	90th	1982-2011	5days	1979-2019
Threshold	D	37-zlevels	3x3	meanpool	95th	1982-2011	5days	1979-2019
cases	E	37-zlevels	3x3	meanpool	90th	1990-2019	5days	1979-2019
Smoothing case	F	37-zlevels	3x3	meanpool	90th	1982-2011	None	1979-2019
Period case	G	37-zlevels	3x3	meanpool	90th	1982-2011	5days	1982-2019

Figure S18. The different sensitivity cases regarding the MHW detection. The top row shows the reference case used throughout the main manuscript. In each sensitivity case A-G, only one of the methodological choices outlined in 2.6 is altered (highlighted in red).



Figure S19. Composite MHW characteristics in different sensitivity cases (see Fig. S18). For comparability, a minimum duration criterion is applied to case F, in which the Boolean array **B** is not filtered using the morphological operations (see Sec. 2.2.1 of main text). It requires that the MHWs have a surface duration of at least five days.



Figure S20. Effect of morphological operations on the number of extreme days in sea surface observations between 1982–2019 (top row, OISSTv2, (Reynolds et al., 2007)) and below in the 1979–2019 ROMS hindcast (surface, 50 m, 250 m and 500 m depth). Left (middle) column shows number of extremes days in smoothed (unsmoothed) Boolean array **B**. Right column shows the relative difference (smoothed minus unsmoothed) in %.



Omitted feature prior to principle component analysis

Figure S21. Sensitivity analysis of MHW clustering using Cohen's Kappa coefficient. In the upper panel, we test for the robustness of the clustering with respect to the omission of 10 % to 99 % of omitted MHWs. In the lower panel, we test for the robustness of the clusters with respect to the omission of individual MHW characteristics feeding into the principal component analysis. In both panels we conduct for each case 10 sensitivity clusters. We indicate the mean across all 10 cases by the bold black number and the minimum and maximum value by the numbers below and above, respectively.

X - 36



Figure S22. Association between surface-only MHW characteristics and the associated water column MHW characteristics. Panels a, b and c shows each a two-dimensional histogram for the duration (D), maximum intensity $(I_{\text{max}} \text{ in }^{\circ}\text{C})$ and severity $(S = \Im_{\text{mean}} \times D$ in days) for surface-only MHWs and their associated water column MHWs, respectively. Each panel shows the 1:1 (dashed black) line as well as the Pearson and Spearman correlation (r) between the surface-only and associated water column MHWs.



Figure S23. Confusion matrices for regionalized logistic regression model based predictions of MHWs that are either deep-reaching (dMHWs) and ML-confined (sMHWs). Purple, blue and orange matrices show correct/false predictions based on surface only duration, maximum intensity and severity, respectively. Underlying map shows the different regions, demarcated by the black lines. Coastal boxes extend to 700 km offshore. Latitudinal regional boundaries are at 5°S, 5°N, 23°N, and 40°N.