ARTICLE

Check for updates

Ninety years of coastal monitoring reveals baseline and extreme ocean temperatures are increasing off the Finnish coast

Norman Goebeler [1✉], Alf Norkko[1,2] & Joanna Norkko [1]

Global marine heatwave assessments often rely on satellite-derived sea surface temperature. However, these data have low accuracy in coastal areas, are unable to measure sub-surface temperatures and have only been available since the 1980s. Here, we analyse 90 years of in situ surface and bottom (30 m) water temperature data from a Finnish coastal monitoring site. Water temperatures were significantly higher between 1991–2020 than 1931–1960 and 1961–1990. We find strong differences between satellite-derived and in situ temperatures, with in situ temperatures being lower in autumn and winter and higher in spring. Measurements at the seafloor indicate marine heatwaves occurred during all seasons between 2016 and 2020, with intensities and durations exceeding previous records. Since the 1990s, we find an upward shift of the baseline temperature and increasingly frequent occurrence of temperatures previously considered as an extreme. Our findings highlight the importance of long-term in situ data and choice of climatological reference periods for assessing change.

[1] Tvärminne Zoological Station, University of Helsinki, Hanko, Finland. [2] Baltic Sea Centre, Stockholm University, Stockholm, Sweden.
✉email: norman.gobeler@helsinki.fi

As predicted by the IPCC[1] (2014), warming of the oceans continues and the frequency of extreme events is increasing[2], and the most recent years (2015–2020) have been the warmest years in the oceans since 1955[3]. The upper layers of the ocean are absorbing most of the atmospheric heat and shallow coastal ecosystems experience the highest (relative) impact of the warming. The most palpable consequences in these productive and diverse systems are seen as range shifts and local extinctions of many species and losses of ecosystem goods and services, following periods of abnormally warm water temperatures[4]. Such discrete, prolonged and anomalously warm events can be defined as marine heatwaves (MHWs) when compared to a local, 30-year climatological baseline[5].

The disturbance of marine ecosystems caused by MHWs has been widely documented around the globe over the last two decades[4,6–10]. This has propelled a wide range of research investigating short-term extreme events, in contrast to the long-term gradual increases in average temperature[11]. Recently a framework for investigating the physical drivers, detection, predictability characteristics and various ecological impacts of MHWs has been proposed[12]. A common framework for defining MHWs is necessary in order to be able to analyse impacts and to compare trends across time and space, locally as well as globally.

While the effects of MHWs are increasingly being reported, there is a general lack of accurate temperature data to quantitatively link the cause and effects (in time and space, incl. depth). Most studies on the impacts of MHWs rely solely on remotely-sensed sea surface temperature (SST) data. Remotely-sensed data have a great advantage in monitoring and analysing SSTs and MHWs over a large spatial extent[13], but (1) lack ability to describe sub-surface temperatures, (2) are temporally limited since satellites were not launched until 1980 and (3) demonstrate inaccuracies in coastal waters compared to in situ data[14]. These issues are of considerable importance in a shallow water body like the Baltic Sea with highly complex, large coastal archipelago systems.

Most parts of the Baltic Sea have experienced a larger temperature increase over the last decades compared to global oceans[15,16]. This gives the Baltic Sea a time machine character of how global coastal systems might respond to a warming climate[17]. The warming trend is not limited to surface waters but is also observed in sub-surface layers thus affecting the entire water column and the seafloor. Due to the complex bathymetry and hydrology, the warming trend in deeper layers is heterogeneous as e.g. the layer between 35 and 75 m depth in the Bothnian Sea shows a similar signal as the Gulf of Finland at 15–35 m[16]. Some of the first studies on sub-surface MHWs in the global oceans indicate their independence from surface MHW[18] and suggest the stability of the stratification and local mechanical forces (e.g. Ekman downwelling) causing the heat transport to deeper layers[19,20]. Key to such studies is the long-term monitoring of deeper layers, which is rare due to technical challenges. Most current data rely on CTD-casts by research vessels or by long-term monitoring stations, although increasingly sub-surface water conditions are monitored using Argo-floats (fleet-monitoring.euro-argo.eu) and permanent sensors at specific sites, e.g. mooring buoys.

The assessment of anomalies and extreme events like MHWs requires knowledge of a norm-state and defining its threshold. The World Meteorological Organization recommends using the 30-year time period from 1961–1990 as a climatological standard reference period (RP) for long-term climate variability[21]. However, since SST satellites only became available in the 1980s, a commonly used reference period is from 1983–2012[8,22]. This time period, however, refers to a time when the global SST was already on a steep increase[23] causing an elevated baseline that might underestimate the magnitude of current extreme events.

Coastal zones are often spatially heterogeneous and highly complex in bathymetry, creating a variety of microhabitats with individual temperature characteristics poorly captured by remotely-sensed techniques. The spatial resolution of the commonly used satellite product NOAA Optimally Interpolated SST[24] (OISST) refers to a $0.25 \times 0.25°$ grid, which corresponds to approximately $400 \, km^2$ in the Baltic Sea area. Although the spatial resolution has been greatly improved via the EU Copernicus Marine Service Product ("Baltic Sea—Sea surface temperature reprocessed"), with a $0.02 \times 0.02°$ grid which corresponds to an area of about $2.5 \, km^2$, it is still crude for complex archipelago systems with a multitude of habitats. Satellite data have been useful for providing new insights into open water patterns and broad-scale ecological studies, but are only partly applicable on smaller scales and in particular habitats[13]. Moreover, satellite-derived SST seemingly underestimate the intensity of MHWs[14].

In this study we examine in situ temperature data from the surface (0–2 m) and sub-surface (30 m deep at the seafloor) layers from 1931 until 2020 collected at the monitoring site Storfjärden. Storfjärden is located in the channel east of Hanko Peninsula (south-west Finland; Supplementary Fig. 1) connecting the river mouth and the open Baltic Sea. It can therefore be classified as middle archipelago region with highly dynamic water masses influenced by riverine freshwater run-off and saline water from the open sea, and is prone to upwelling and highly susceptible to wind forcing[25]. The thermocline is typically at 5–10 m depth between January and August, and strongest in summer[26]. The aim of this study was to (1) compare historic with current climatological periods, (2) investigate historical extreme temperatures, (3) point out inaccuracies of remotely-sensed SST and (4) combine long-term with recent high-resolution temperature data to identify MHWs in sub-surface layers between 2016 and 2020. The remotely-sensed SST was retrieved from E.U. Copernicus Marine Service Information (CMEMS) from the Storfjärden pixel with a spatial resolution of $0.02 \times 0.02°$ (see panel b of Supplementary Fig. 1 for pixel area). Our results highlight the urgent need for in situ temperature monitoring networks in coastal zones to understand the characteristics of MHWs and their subsequent ecological impacts.

We demonstrate an ongoing warming trend for the period 1927–2020 in the surface and bottom layer of the monitoring site Storfjärden, south-west Finland. Consequently, the ongoing warming caused the climatological baseline and its thresholds to be significantly warmer on almost every day of the year for the 30-year reference period of 1991–2020 compared to previous periods from 1931–1960 and 1961–1990, and in the surface and bottom layer, respectively. This results in a severe change of perspective when evaluating extreme temperatures based on different reference periods. Satellite-derived data sets are not able to demonstrate this shift in the baseline as they have only been initiated in the 1980s. Further, we demonstrate significant differences in the climatological baseline and its thresholds derived from remotely-sensed SST and in situ SST for the periods from 1982–2011. Satellites are not able to capture sub-surface conditions, where we detect 18 marine heatwaves on the seafloor in just four years of high-resolution recordings, which are important for investigating the effects of MHW on benthic ecosystems.

Results

Long-term trend. Figure 1 shows the annually averaged long-term water temperature in the surface and bottom layer and the percent of extremely warm observations (RP31-20) of Storfjärden from 1927 until 2020. The temperature in the surface layer increased by 1.8 °C (linear trend= 0.019 °C per year; y-intercept= 6.5 °C) and in
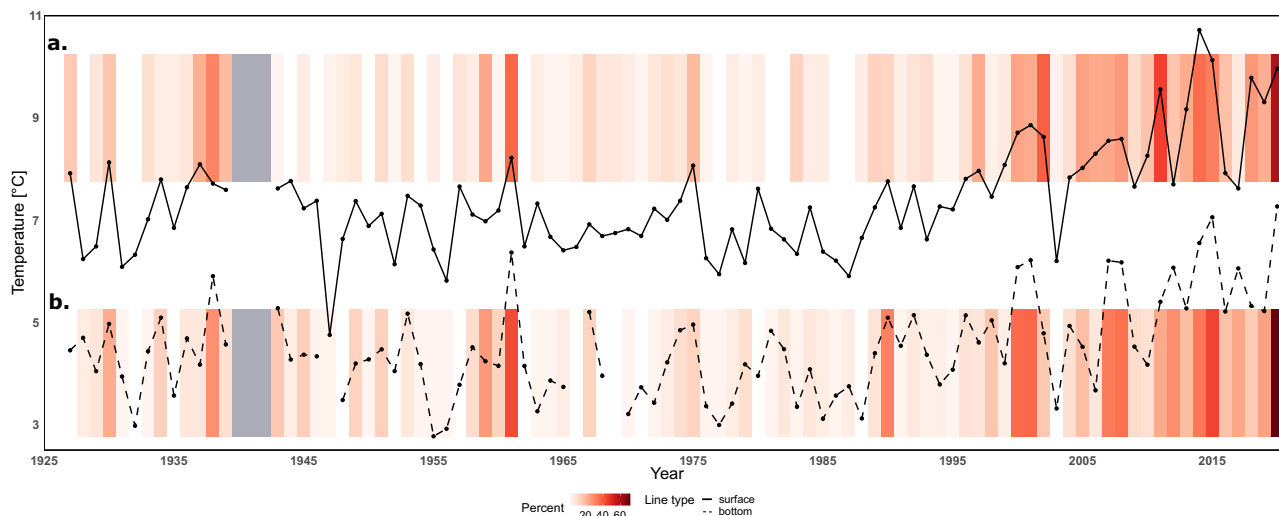
Fig. 1 Yearly average temperature and abnormally warm extremes at Storfjärden. Yearly average temperature of Storfjärden from 1927 until 2020 in the surface layer (**a**, solid line) and bottom layer (**b**, dashed line). The stripes represent the occurrence of abnormally warm observations (RP31-20) as percent of all measurements per year in a. the surface and b. bottom layer. Greyed areas are years with less than 30 observations.
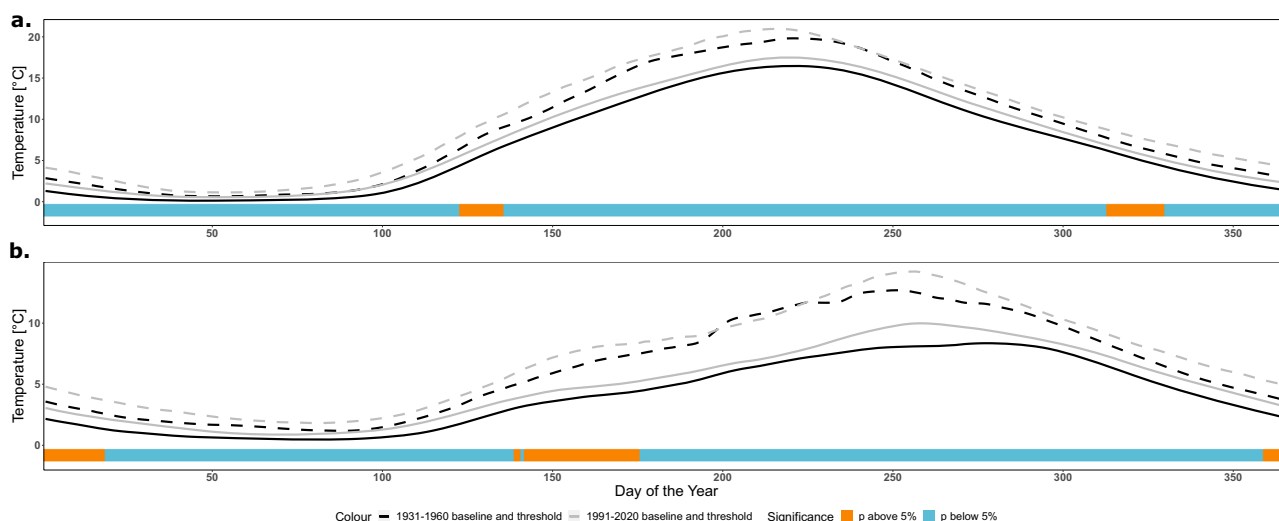

Fig. 2 Comparison of climatological baselines of reference periods 1931–1960 and 1991–2020. Climatological baseline (solid lines) and 90th percentile (dashed lines) of RP31-60 (black lines) and RP91-20 (grey lines) of the surface (**a**) and bottom (**b**) layer derived from the interpolated long-term time series from Storfjärden. The bottom ribbon of each panel refers to the days of the year, where the two probability density functions of each day are significantly different ($p < 0.05$, light blue) or not different ($p > 0.05$, orange).

the bottom layer by 1.3 °C (linear trend= 0.014 °C per year; y-intercept= 3.7 °C). The warming appears to be accelerating, as the six warmest years in the surface ever recorded were after 2012 (Supplementary Table 1). Further, in a ranking of the 20 warmest years, in the surface only 5 years were before the year 2000, 6 years for the bottom. The number of warm extremes is increasing in both layers, peaking in 2020 with 58 % of the measurements above normal in the surface and 71 % in the bottom.

Comparison of Reference Periods. The long-term time series starting in August 1926 allowed defining multiple 30-year reference periods. A comparison of annual temperature distributions for RP31-60 and RP91-20 is shown in Fig. 2. The mean climatological baseline and its thresholds for the surface and bottom layer of RP91-20 are higher than of RP31-60 over the entire year. Averaged over the entire year the climatological baseline of RP91-

20 has been elevated by 0.88 °C (sd = 0.29), the 90th percentile threshold by 0.996 °C (sd = 0.53) and the 10th percentile threshold by 0.78 °C (sd = 0.63) in the surface layer compared to RP31-60. Analogously for the bottom layer, the climatological baseline has been elevated by 0.83 °C (sd = 0.32), the 90th percentile threshold by 0.83 °C (sd = 0.46) and the 10th percentile by 0.63 °C (sd = 0.37). Minor time periods in each layer (30 days of the year in the surface layer spread across spring and autumn, 62 days of the year in the bottom layer spread across spring and winter) were not significantly different, but most days of the year had two significantly different probability density functions (see Supplementary Fig. 2 for a figure of Dmax and p-values resulting from Kolmogorov-Smirnov (KS)-test). This highlights the importance of declaring which reference period is used in an assessment of extreme events, as the resulting MHW-parameters will differ markedly.
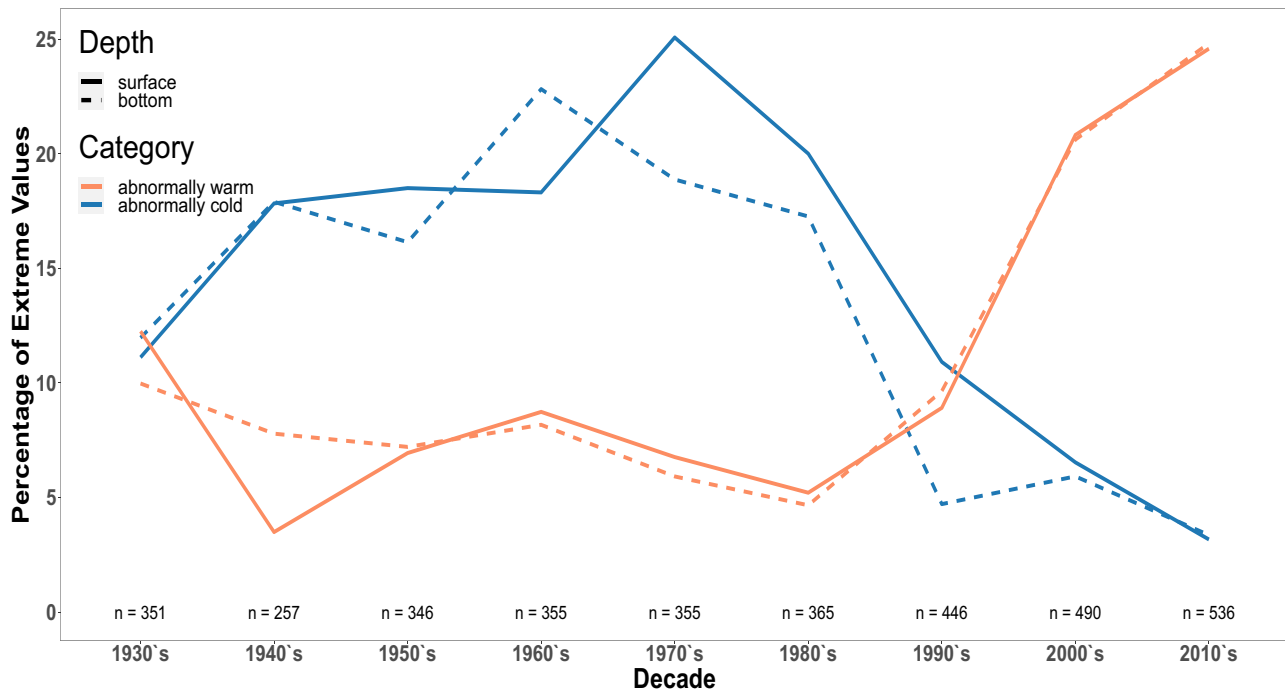
Fig. 3 Decadal temperature extremes according to 1931–2020 reference period. Percent of extreme observations per decade in the surface and bottom layer of Storfjärden. Orange lines are of "abnormally warm" category exceeding the 90th percentile of RP31-20, whereas blue lines are of "abnormally cold" category negatively exceeding the 10th percentile. Dashed lines refer to bottom layer, solid lines refer to surface layer. n refers to the total number of measurements per decade.

In the following assessments of this study (except in the comparison with remotely-sensed SST) we applied a 90-year reference period from 1931–2020 as this summarizes past colder and current warmer periods creating "conservative" climatological values.

Quantifying extreme Observations. All vertical temperature profiles from Storfjärden, starting 1931-01-01 until 2020-12-16, were evaluated against the 90-year reference period (RP31-20). "Near normal" observations, none of which exceed the 10th or 90th percentile, were excluded for clarity. Until the 1990's "abnormally warm" observations were mainly less than 10% of all measurements, whereas "abnormally cold" categories increased until the 1970's to almost 25% of total measurements in the surface layer, 1960's with 22.8% for the bottom layer, respectively (Fig. 3). After the 1990's, "abnormally warm" observations steeply increased to about 25% of all observations, whereas "abnormally cold" observations declined to 3% of observations.

Both surface and bottom layers demonstrated a steep increase in "abnormally warm" observations in all months of the year and a decline with partial disappearance of "abnormally cold" observations (Fig. 4). The summer months June, July and August had the smallest increase of "abnormally warm" observations in both layers, whereas the winter and spring months (November—May) had the highest increase, i.e., the largest change was observed in the winter and spring months. During the 2000's the month with the highest increase of "abnormally warm" observations was January for both layers. During the latest decade (2010s) most "abnormally warm" observations were in September for the surface layer and in December for the bottom layer. This analysis demonstrates that warm extremes presently dominate in the surface and bottom layer at Storfjärden, but lacks in specifying the duration of extreme events. In the following section we will therefore introduce daily temperature data from the seafloor of Storfjärden to identify sub-surface MHWs.

Sub-surface MHWs at Storfjärden analysed from high-resolution data between 2016 and 2020. Since the initiation 2016-08-02 until the latest available datapoint 2020-12-09 (1591 days) measurements were recorded on a total of 1335 days, of which 432 days were part of a MHW with a total cumulative intensity (above the 90th percentile) of 511.1 °C*days (Supplementary Table 2), when comparing to the 90-year reference period from 1931–2020. MHWs of moderate and strong categories[22] were observed across all four seasons with varying durations and intensities (Fig. 5). The notably longest MHW started 2019-12-17 and lasted until 2020-06-30 (with a 9-day gap in recordings at the beginning of April 2020 due to instrument maintenance) with a duration of 198 days, a cumulative intensity above the threshold of 252.2 °C*days and a maximum intensity of 6.9 °C above the threshold on the 2020-06-22 (RP31-20: climatological mean 4.5 °C; threshold 7.9 °C). Abrupt changes in temperature, e.g., the rapid onset and decline of the MHW 2018 at the end of July, are common for Storfjärden as this area is prone to upwellings. The highest daily average temperature recorded was during the summer MHW of 2018 with 17.8 °C and a daily maximum of 21.29 °C on 2018-07-31 (RP31-20: climatological mean 6.6 °C; threshold 10 °C).

Comparison of remotely-sensed and in situ surface long-term data. In general, remotely-sensed and in situ surface temperatures correlated positively for the tested period (1982-01-01 – 2019-08-31; R = 0.96; p < 0.01; RMSE = 1.85; Fig. 6). Despite the strongly correlating trend, care should be taken as single measurements deviated markedly from each other by up to 11.2 °C (2004-06-21; RS-SST 20.7 °C, in situ SST 9.5 °C).

The baseline climatology is essential for defining MHWs. Here we compared the probability density function of each day of the year (doy) for the exact 30-year period of 1982-01-01—2011-12-31 on the remotely-sensed and interpolated long-term SST data by applying a two-sided Kolmogorov-Smirnov test (see
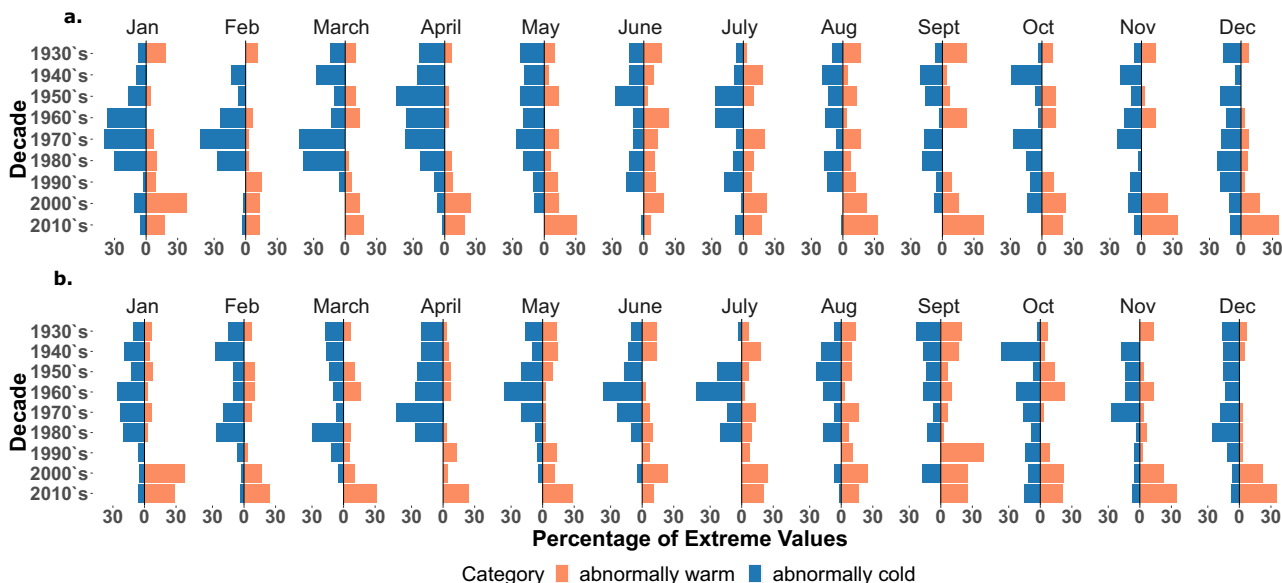
Fig. 4 Decadal temperature extrema per month according to RP31-20. Monthly percent of extreme observations per decade in the surface (**a**) and bottom (**b**) layer of Storfjärden. Orange columns are of "abnormally warm" category exceeding the 90th percentile of RP31-20, whereas blue columns are of "abnormally cold" category negatively exceeding the 10th percentile.
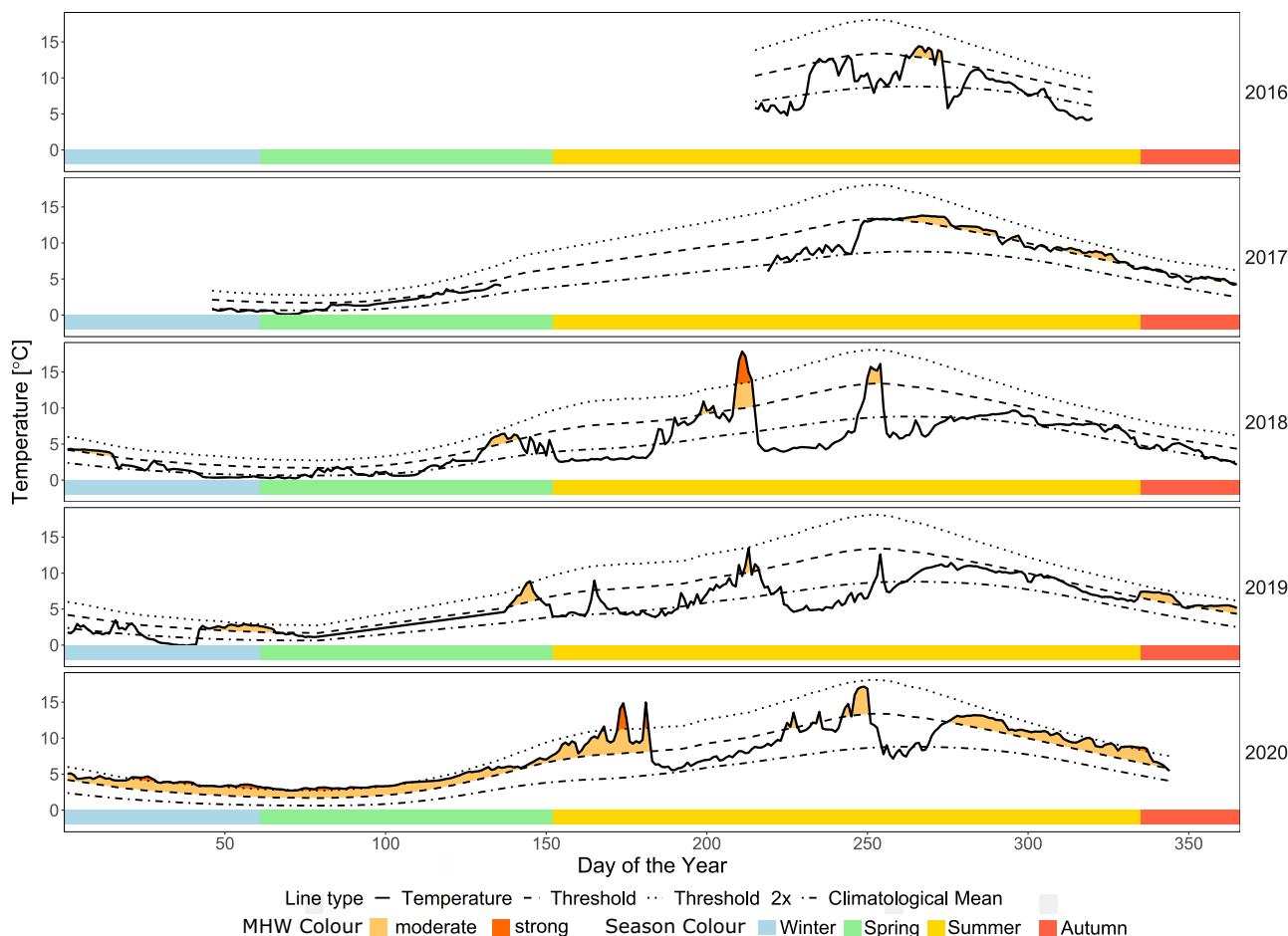
Fig. 5 Marine Heatwaves at the seafloor (30 m) of Storfjärden. Marine Heatwaves between 2016 and 2020 at Storfjärden at 30m depth according to RP31-20. Linetype: solid—Daily average temperature; (following linetypes in order from bottom to top for each panel:) dotdash—mean climatology; dashed —threshold climatology; dotted—threshold_2x climatology. Coloured ribbon refers to the seasons: light blue—winter; green—spring, yellow—summer, light brown—autumn. MHWs of moderate intensity (Category I) are of light orange colour and strong MHWs (Category II) of dark orange colour.
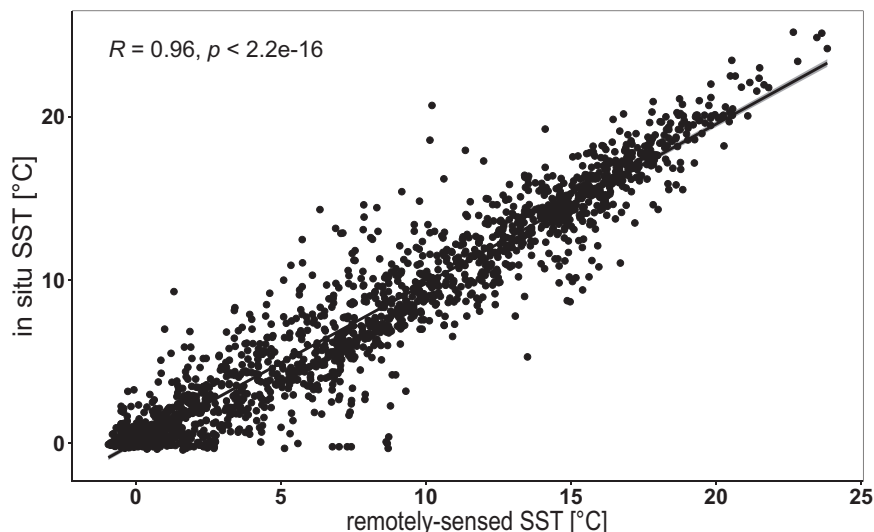
Fig. 6 Correlation of satellite-derived SST and in situ SST. Correlation of in situ surface water temperatures with satellite-derived SST from Storfjärden for the period from 1982-01-01 – 2019-08-31. RMSE = 1.85.
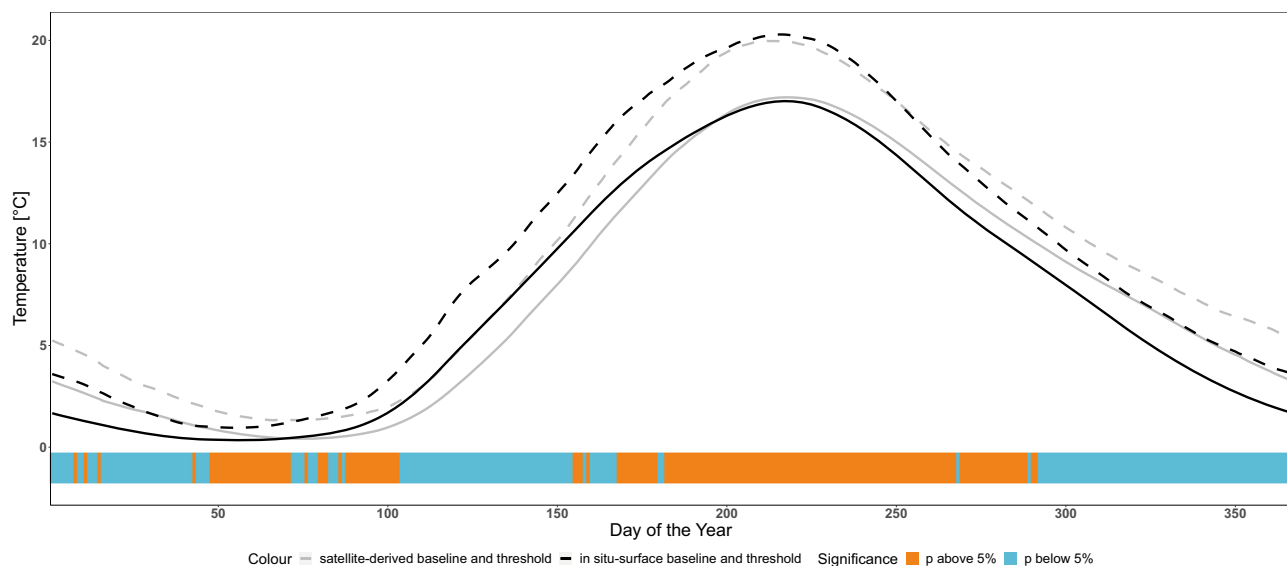

Fig. 7 Comparison climatological baselines of in situ and satellite-derived SST for 1982–2011. Reference Periods from 1982-01-01 – 2011-12-31 derived from remotely-sensed (grey lines) and interpolated in situ (black lines) SST from Storfjärden. Solid lines refer to the mean climatological baseline and dashed lines to the 90th percentile threshold, respectively. The bottom ribbon refers to the days of the year, where the two probability density functions of each day are significantly different ($p < 0.05$, light blue) or not significantly different ($p > 0.05$, orange).

Supplementary Fig. 3 for a figure of Dmax and p values resulting from KS-test). Significant dissimilarities were identified on 193 days, particularly for autumn until the onset of spring (265th doy to 56th doy) and late spring until beginning of summer (120th doy to 169th doy) (Fig. 7). The climatological mean derived from the long-term dataset is lower from autumn until winter and higher during spring than the mean climatology derived from the remotely-sensed dataset.

Discussion

In the last decades most parts of the oceans[4] and lakes[27] around the globe have experienced heatwaves and the frequency, duration and intensity are projected to increase even under an optimistic scenario of 1.5 °C global warming[28]. The entire Baltic Sea has experienced an average temperature increase from 1982 to

2016 of 1.09 °C (derived from extrapolated vertical profiles from eight sub-basins), which is two times higher than the upper 100 m of the Atlantic Ocean[16]. More specifically, for the monitoring site at Storfjärden, at the south-west coast of Finland, we demonstrate an increase of the trend for the period from 1927 until 2020 by 1.8 °C SST (1.3 °C bottom). In a previous study on the same time series between 1927 and 2012 an increase of 1.0 °C SST (0.8 °C at the bottom) was reported[26]. This means that the overall trend estimate has almost doubled when including the years after 2012. In the current study, we have in more detail investigated the underlying temperature data forming the trend, with particular focus on the fluctuations, and analysed normal and extreme temperatures between 1931 and 2020. We analysed the occurrence of MHWs only from the recent (2016–2020), high-resolution dataset. For the 90-year long-term dataset, we only analysed the occurrence of extremes (can be single data points)
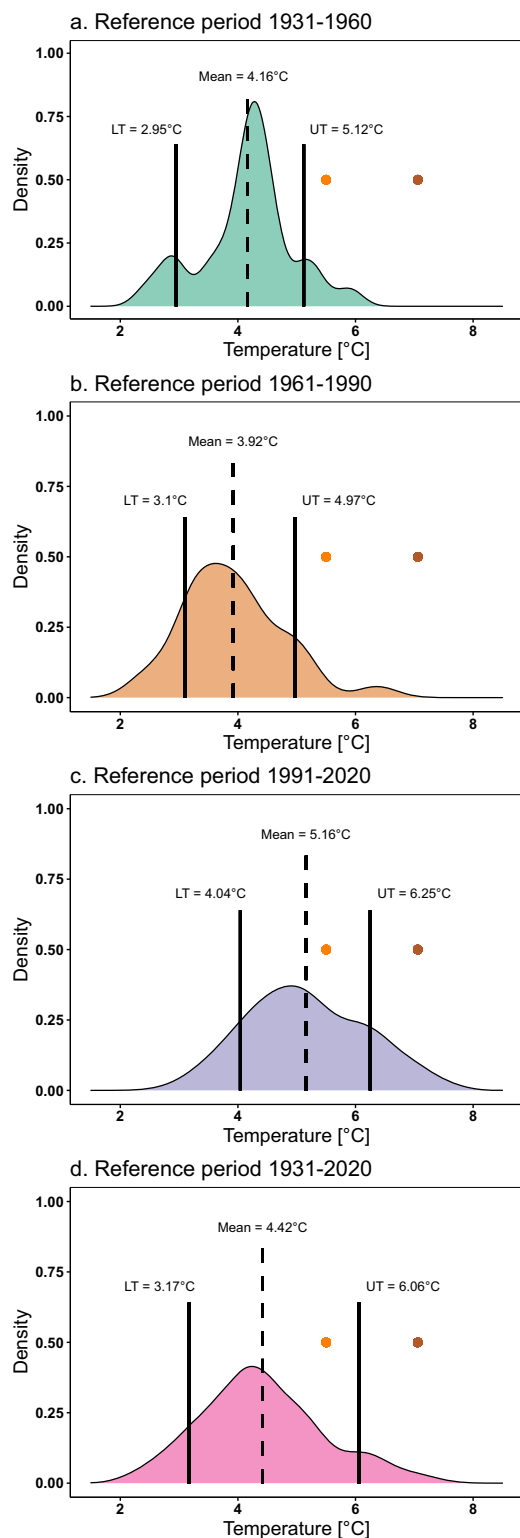
Fig. 8 Changing the perspective. Probability densities of four different reference periods (**a** Reference period 1931–1960, **b** 1961–1990, **c** 1991–2020, **d** 1931–2020) and their respective mean temperature (dashed line), lower (10th percentile, LT) and upper thresholds (90th percentile, UT) (solid lines). Yearly bottom average temperature of Storfjärden for 2019 (orange dot) and 2020 (brown dot).

and there is no higher-resolution data available from the early days to analyse MHWs the way they are defined nowadays with a minimum of 5 consecutive days.

The daily probability density functions of the time periods 1931–1960 and the latest 1991–2020, from the surface and bottom layer of Storfjärden, are for most days of the year significantly different (Supplementary Fig. 2). This means that the climatological baseline and its thresholds from RP31-60 has been elevated by approximately 1 °C over the entire year in RP 91-20. The climatological baseline is representative for 30 years of data and not only highlights the persistent warming trend over decades but also changes our perspective on extreme temperatures. When applying RP31-60 to the entire time series 633 abnormally warm and 450 abnormally cold observations can be detected in the surface, and 648 and 397 in the bottom, respectively. By evaluating the time series referring to RP91-20 only 273 abnormally warm but 952 abnormally cold observations can be detected in the surface, and 268 and 897 in the bottom, respectively. This shift is schematically illustrated, by comparing the yearly bottom average temperature at Storfjärden for 2019 (5.5 °C) and 2020 (7.06 °C) with the PDFs and respective thresholds of different reference periods, in Fig. 8. 2019 can be considered an extreme year compared to the reference periods 1931–1960 and 1961–1990, but is considered normal for the 90-year reference period from 1931–2020 and also the latest reference period 1991–2020. The yearly average temperature of 2020 can be considered as an extreme year compared to any reference period. These differences in classification, whether a temperature is "normal" or "extreme", is of theoretical, climatological nature and more research is required to understand their ecological relevance. Yet, for the evaluation of extreme events, the choice of reference period changes their estimated intensity, magnitude and duration due to the relative character of the PDF. Instead of evaluating temperature data in relation to one historic climatological period, they could also be evaluated by separate, moving climatological periods. Then an increasing trend and its extremes would shift at a similar pace[29] or in other words: a moving reference period would assign the long-term warming to the climatological baseline[30]. Undoubtedly, the high frequency of extreme temperatures in the latest decades of this study is tightly related to long-term warming. Detrending or applying a moving reference period is yet of little ecological relevance, but mostly useful when investigating the temperature variability and forecasting. A moving reference period might be appropriate for organisms that adapt over short timescales[27], but adaptation particularly for the upper thermal limit can be rather slow[31]. However, the specific baseline chosen depends on the question addressed and more research is required to elucidate which reference strategy to apply.

Reference periods that span 30 years include interannual oscillations that drive climate variability. Nevertheless, particular MHW events can be directly attributed to these oscillations[32]. Of particular importance for the climate of the Baltic Sea are the North Atlantic Oscillation (NAO) and the Atlantic Multidecadal Oscillation (AMO). The winter NAO index is dominant in explaining the mean SST on an annual scale, whereas the AMO explains almost 60% of the variability on an interannual scale[33]. The 90-year reference period used in this study, includes a warm phase of the AMO peaking in the 1930s with an approximate even occurrence of cold and warm extremes (at around 12% of all observations during this decade, Fig. 3). Thereafter a cold phase with a negative maximum occurred in the 1970s which corresponded to the maximum of cold extremes in this study at about 25% (Fig. 3), and another warm phase starting in the 1990s, where we demonstrate the disappearance of cold extremes and dominance of warm extremes at about 25% (Fig. 3). Despite this

close link between temperature extremes and the phases of the AMO, the latest decades after the 2000s demonstrate unprecedently warm temperatures at an exceptional frequency. Kniebusch et al. (2019)[33] suggest an enhanced effect of the warm phase through global warming and propose a weakening once the AMO index engages into its cold phase again. It can therefore be expected that the next phase will be a cold phase moderated by the ongoing global warming and therefore possibly warmer than the previous cold phase. A discussion on the contribution of the NAO to annual extremes, and the identification of regional drivers, would require additional investigations that are beyond the scope of this study. Further, on a monthly scale a shift in the bottom layer happened from the 2000's to the 2010's; previously most excessive warm observations were recorded from June to January and in the following decade between November and May. Changes in the surface layer from the 2000's to 2010's are less pronounced, except for a steep decrease in extreme warm observations in January and a steep increase of warm observations in September. The long-term temperature increase in the bottom and surface layer of Storfjärden and the tendency towards milder winters with lesser sea ice extent until 2012[26,34] has continued between 2010 and 2020. The ongoing trend and high frequency of warm extremes emphasize their tight connection and suggest that current extremes will become the new normal.

Sub-surface water temperatures are essential for understanding biological and chemical processes on the seafloor and coastal ecosystems in general, but current remote sensing techniques cannot resolve them and have limited capacity for capturing SST in spatially heterogenous coastal areas. While there are large deviations between in situ and remotely-sensed SST also on an oceanic coast[14], this deviation can be even more pronounced in the highly complex archipelago systems along the Baltic Sea coast. In this study we show that while in situ and remotely-sensed SST are highly correlated ($R = 0.96$; $p < 0.01$; $RMSE = 1.85$), the resulting long-term climatological values differ significantly. Deviations accumulate when considering the long-term period and occur, in our case, mostly during the warming and cooling periods of the year (spring and autumn). The impact of this mismatch is further enhanced when considering local ecosystem effects of episodic events such as MHWs. Despite the high overall correlation, deviations of up to 11.2 °C (2004-06-21; RS-SST 20.7 °C, in situ SST 9.5 °C) were recorded, which highlights the difficulty when assessing ecosystem consequences of episodic events[13]. Few locations have datasets spanning 30 years with daily measurements for defining climatological baselines, which in turn makes the use of satellite-derived climatological values necessary. Nevertheless, such mismatches in accuracy should be considered when evaluating extreme events by mixing in situ and remotely-sensed data. Importantly, however, it should be recognised that remotely-sensed estimates of SSTs are of very limited use for interpreting trends in sub-surface temperatures. The coastal zone can be divided into 5 temperature classes from innermost to outermost areas with depth and exposure as the dominating explanatory variables[25]. The water mass at Storfjärden is highly dynamic and is characterized by stratification between May and September causing a sharp vertical temperature gradient[34], is prone to horizontal exchanges of water masses due to up- and downwelling[35] and can thus experience large temperature changes over short times (2018-07-31; 13 °C in 24 h). In this study we used the climatological baseline derived from 90-years of monitoring, and high-resolution data from an automatic logger to investigate recent occurrences of sub-surface MHWs. In only a few years of high-resolution monitoring, 18 MHWs have been detected across all seasons ranging from short durations of 5 days up to 198 days and intensities of moderate to strong categories[22] even peaking in all-time record temperatures during the July

MHW in 2018. Summer MHWs have diverse effects on the seafloor and benthic communities[36–38]. For example, observations related to the two summer MHWs in 2018 on the southern coast of Finland, both in terms of ecological impacts and also a potentially less considered effect on coastal greenhouse gas emissions highlight potential MHW impacts. The summer MHWs in July 2018 (6-day duration; 21.29 °C maximum temperature 2018-07-31) and September 2018 (5-day duration; 16.75 °C maximum temperature 2018-09-11) were at the minimum duration of an MHW-event and might have been caused by elevated air temperatures and easterly winds resulting in a downwelling (Supplementary Fig. 6). Following the July 2018 MHW, diverse temperature related signals in long-term projects and a large mortality (>50%) in Mytilus *edulis* beds were observed (Westerbom, personal communication), raising concerns for the fate of multiple other key habitats in the area, such as seagrass meadows. Humborg et al. (2019)[39] recorded elevated greenhouse gas emissions during a research cruise directly after the September MHW close to Storfjärden, suggesting a combination of a MHW and a storm event triggered the outgassing. Particularly shallow areas demonstrated elevated CH_4 emissions, due to organic-rich sediment and high temperatures. Consequently, it is implied that with ongoing warming and increased frequency of MHWs, organic-rich coastal sediments of the Baltic Sea could accelerate their out-gassing of greenhouse gases[39]. The coincidence of the timing of the cruise and the combined MHW-storm event highlight the need to combine long-time observations, high-resolution in situ monitoring and event-based preparedness. Without the appropriate data to define the reference conditions and baselines, we are not able to assess climatological change or to predict subsequent ecosystem effects.

Materials & methods

Long-term time series. We make use of data collected at the monitoring site Storfjärden (59°51'33"N, 23°15'35"E), close to Tvärminne Zoological Station (University of Helsinki, Supplementary Fig. 1). The monitoring site has a depth of about 30 m and is located in the transition zone between the open Baltic Sea and the mouth of the Pohjanpitäjänlahti fjord-like estuary east of the Hanko Peninsula. The bight area is typical for the Finnish coast with a large number of skerries and islands creating a geomorphologically complex area. It is open towards the south and facing typically southwesterly winds making Storfjärden, located in one of the main channels, well connected to the open Baltic Sea.

The recording of water temperature and salinity started in August 1926 and has been sampled manually on the first, tenth and twenty-first day of the month, making it one of the oldest time series in the Baltic Sea next to the time series of Utö in the outermost, southwestern part of the Archipelago Sea[40]. Unlike the time series from Utö, the recordings at Storfjärden have been conducted regularly over the decades with only minor gaps, particularly between 1940 and 1942 during Soviet occupation. Over time the methods for measuring temperature at different depths has changed. Between 1926 and 1989 a water sampler with a reversing thermometer was used to record temperatures at 0 m, 5 m, 10 m, 20 m and 30 m depth (additionally 15 m starting in 1956). Starting in 1989 the frequency of measurements increased (from about every 10 days to about every 7 days, Supplementary Table 3) and a higher vertical resolution was achieved using a SIS CTD plus 100 and from 2009 onwards more readings were obtained by different CTDs (RBR-Concerto, FSI NXIC CTD, Valeport miniCTD, CastAway CTD) partly replacing the water sampler. The Storfjärden monitoring site is one of the most intensely sampled sites in this area, and the dataset used in the

Table 1 Time periods (starting 1st January and ending 31st December) used for climatological baselines and thresholds.

Reference Period	Abbreviation	Comment
1931–1960	RP31-60	Overall, 30 years, but without 1940–1942
1961–1990	RP61-90	Recommended by the World Meteorological Organization
1982–2011	RP82-11	First complete 30-year period of satellite-derived SST
1991–2020	RP91-20	Latest period with data available until 16.12.2020
1931–2020	RP31-20	90-year Reference Period

current study is a compilation of data provided by the Finnish Meteorological Institute (FMI), the Tvärminne Zoological Station (TZS) and multiple projects that have obtained CTD-readings at Storfjärden (Supplementary Table 3).

High-resolution seafloor temperature Data. In August 2016 a sensor (YSI Exo2) for high-resolution temporal monitoring of multiple parameters 30 cm above the seafloor was deployed at the Storfjärden monitoring site (30 m depth). The sensor measures temperature, salinity, oxygen, pH, and turbidity every 30 min throughout the year (except when retrieved for maintenance).

Remotely-sensed dataset. The remotely-sensed SST temperature data product "Baltic Sea – Sea Surface Temperature Reprocessed" (Product identifier: SST_BAL_SST_L4_REP_OBSERVA-TIONS_010_016; Location: https://resources.marine.copernicus. eu/product-detail/SST_BAL_SST_L4_REP_OBSERVATIONS_ 010_016/INFORMATION, accessed 2021-10-14), referred to as L4 product, used in this study derived from E.U. Copernicus Marine Service Information (CMEMS)[41]. The temporal coverage of this product ranges from 1982-01-01 until 2019-08-31 and has a spatial resolution of $0.02 \times 0.02°$. The dataset contains the SST of the closest pixel to Storfjärden, which refers to about 2,5 km². The climatological mean and the 90th percentile for the period from 1982-01-01 until 2011-12-31 were estimated according to the definitions by Hobday et al., 2016[5].

Dataset quality and preparation. First, we performed a quality control of the long-term time series and the high-resolution data removing values with markedly decreasing salinity (e.g., ~2 psu within the last 1 m) towards the bottom, fluctuating or false depth signals and obviously false outliers in visual inspections. After quality control, the time series data were averaged over depths between 0–2 m (surface) and below 29 m (bottom), respectively. This resulted in the long-term time series data containing 3657 unique surface datapoints (3653 for bottom) between 1926-08-23 and 2020-12-16 (~10.6% of all days within this time frame are represented). The high-resolution data was averaged across time to produce a single value for each day.

In order to create a continuous daily record and remove the effect of different sampling frequencies over time, a daily record of temperatures in the surface and bottom was created by interpolating between datapoints when the gap was less than 31days. Linear interpolation is an accepted method of gap filling in time series as temperatures typically don't change very abruptly[21]. While we apply this to a dynamic, coastal site, we feel the approach is justified for creating daily datapoints that are further averaged over at least 30 years for each day to describe overall climatic conditions. Five gaps larger than 31 days were identified. Three of these gaps were between December and January (1950 with gap = 33 days, 1962 with gap = 41 days, 1979 with gap = 41 days). The gaps were linearly filled as winters typically vary very little and they followed a prevailing pattern of warmer temperatures in December and colder temperatures in

January. The two remaining gaps, between 1940 and 1942 and between July and September 1947, were not filled.

Reference periods, thresholds and analyses. A MHW is defined as an anomalously warm, prolonged and discrete event. This means that the water temperature has to exceed the 90th percentile of a preferably 30-year climatological period for at least 5 consecutive days[5]. MHWs can further be classified based on the multiples of difference between the mean climatological baseline and its corresponding 90th percentile[22] (Category I "moderate" 1-2x, Category II "strong" 2-3x). A critical factor in this assessment is the time period it is being compared to. Historically, a 30-year period was used merely because only 30 years of data were available at that time[21]. The current general recommendation is to apply a 30-year span, preferably from 1961–1990, but modifications depending on data availability and application might also be used. Due to the availability of a particularly long in situ time series, we created five reference periods based on the linearly interpolated long-term time series (Table 1, schematic overview Supplementary Fig. 4).

An overview of where in this study which reference period has been applied and additional detailed information can be found in the supplementary material (Supplementary Table 3).

We defined a probability density function (PDF) of the temperature data in a day-of-the-year manner for the respective reference period, the climatological mean and the thresholds according to the 10th and 90th percentile[5,14]. Afterwards, the climatological mean and the thresholds were smoothed by a 31-day moving average with the respective day centred. In the analyses we compared the applied dataset to the thresholds derived from the reference period based on the linearly interpolated long-term time series. To be named an extreme event, a MHW needs to persist over at least 5 consecutive days[5]. We therefore calculated MHWs for the high-resolution and the remotely-sensed data, using the software module heatwaveR[42] in R-studio. For long-term time series, we did not feel this was appropriate as the time period between measurements was generally greater than 5-days[43]. Instead, we categorized and counted single measurements exceeding the threshold, in total or monthly per decade, as "abnormally warm", exceeding the 90th percentile, "abnormally cold", negatively exceeding the 10th percentile, and "near normal" for values between those thresholds. As the frequency of measurements has increased since the 1980's, the counts of extreme values were standardized to the number of total measurements and expressed in percentages.

Long-term trends from 1927 until 2020 were estimated by linear regression (least squares) for the surface and the bottom layer separate. Years with less than 30 measurements were excluded.

In order to compare the difference between any two reference periods a two-sided Kolmogorov-Smirnov-test was applied to both probability density functions at every day of the year separately for the surface and bottom layer. The two-sided KS-test is a standard, non-parametric test to test if two samples come from the same underlying distributions and its null hypothesis is

that both probability functions come from the same distribution. This results that a p value below 5% rejects the null hypothesis and demonstrates a significant difference between the probability density functions. As the number of measurements in the latest period between 1991 and 2020 is markedly higher (surface, $n = 1470$) than in previous reference periods (1931–1960, surface, $n = 956$; 1961–1990, surface, $n = 1083$), a KS-test was conducted to test if the increased frequency biased the results. Therefore, a separate dataset was created where measurements causing a gap smaller than 10 days were omitted for the RP91-20. This reduced the sampling frequency (from 1470 measurements to 901 per 30 years) to the level of previous reference periods. Thereafter, the gaps were linearly interpolated and the temperature distributions of every day of the year between the applied RP91-20 with 1470 measurements and the RP91-20 with forced gaps ($n = 901$) were tested with a KS-test and showed no significant difference (surface, lowest $p = 0.586$ at doy 171, overall mean $p = 0.982$; bottom, lowest $p = 0.586$ at doy 165, overall mean $p = 0.977$; Supplementary Fig. 5). This showed that the largely increased sampling frequency during the latest 30-year period does not influence the resulting climatological baseline and its thresholds.

Comparing in situ and remotely-sensed SST and reference periods. In situ SST data from Storfjärden is representative for an area classified as outer, exposed archipelago[25]. Here we investigated the correlation of in situ surface data from the long-term time series with remotely-sensed SST and compared the respective climatological means and thresholds of RP82-11 by applying a KS-test to their daily temperature distributions.

Data availability

Code availability

References

1. IPCC, 2014, Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
2. Bindoff, N. L. et al. Changing Ocean, Marine Ecosystems, and Dependent Communities. *IPCC Spec. Rep. Ocean Cryosph. a Chang. Clim.* [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. *Press* 447–588 (2019).
3. Cheng, L. et al. Upper Ocean Temperatures Hit Record High in 2020. *Adv. Atmos. Sci.* **38**, 523–530 (2021).
4. Smale, D. A. et al. Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nat. Clim. Chang.* **9**, 306–312 (2019).
5. Hobday, A. J. et al. A hierarchical approach to defining marine heatwaves. *Prog. Oceanogr.* **141**, 227–238 (2016).
6. Garrabou, J. et al. Mass mortality in Northwestern Mediterranean rocky benthic communities: Effects of the 2003 heat wave. *Glob. Chang. Biol.* **15**, 1090–1103 (2009).
7. Frölicher, T. L. & Laufkötter, C. Emerging risks from marine heat waves. *Nat. Commun.* **9**, 2015–2018 (2018).
8. Oliver, E. C. J. et al. Longer and more frequent marine heatwaves over the past century. *Nat. Commun.* https://doi.org/10.1038/s41467-018-03732-9 (2018).
9. Garcia-Herrera, R., Díaz, J., Trigo, R. M., Luterbacher, J. & Fischer, E. M. A review of the european summer heat wave of 2003. *Crit. Rev. Environ. Sci. Technol.* **40**, 267–306 (2010).
10. Marbà, N., Jordà, G., Agustí, S., Girard, C. & Duarte, C. M. Footprints of climate change on Mediterranean Sea biota. *Front. Mar. Sci.* **2**, 56 (2015).
11. Holbrook, N. J. et al. Keeping pace with marine heatwaves. *Nat. Rev. Earth Environ.* https://doi.org/10.1038/s43017-020-0068-4 (2020). in press.
12. Oliver, E. C. J., Wernberg, T., Benthuysen, J., Chen, K. & Eds. *Advances in Understanding Marine Heatwaves and Their Impacts.* Lausanne: Frontiers Media SA. vol. 7 (2020).
13. Smale, D. A. & Wernberg, T. Satellite-derived SST data as a proxy for water temperature in nearshore benthic ecology. *Mar. Ecol. Prog. Ser.* **387**, 27–37 (2009).
14. Schlegel, R. W., Oliver, E. C. J., Wernberg, T. & Smit, A. J. Nearshore and offshore co-occurrence of marine heatwaves and cold-spells. *Prog. Oceanogr.* **151**, 189–205 (2017).
15. Rutgersson, A., Jaagus, J., Schenk, F. & Stendel, M. Observed changes and variability of atmospheric parameters in the Baltic Sea region during the last 200 years. *Clim Res.* **61**, 177–190 (2014).
16. Liblik, T. & Lips, U. Stratification has strengthened in the baltic sea – an analysis of 35 years of observational data. *Front. Earth Sci.* **7**, 1–15 (2019).
17. Reusch, T. B. H. et al. The Baltic Sea as a time machine for the future coastal ocean. *Sci. Adv.* **4**, eaar8195 (2018).
18. Hu, S. et al. Observed strong subsurface marine heatwaves in the tropical western Pacific Ocean. *Environ. Res. Lett.* **16**, 104024 (2021).
19. Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M. & Riser, S. C. Subsurface Evolution and Persistence of Marine Heatwaves in the Northeast Pacific. *Geophys. Res. Lett.* **47**, 1–10 (2020).
20. Schaeffer, A. & Roughan, M. Subsurface intensification of marine heatwaves off southeastern Australia: The role of stratification and local winds. *Geophys. Res. Lett.* **44**, 5025–5033 (2017).
21. WMO, *Guide to Climatological Practices.* (2018).
22. Hobday, A. J. et al. Categorizing and naming marine heatwaves. *Oceanography* **31**, 162–173 (2018).
23. Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J. & Heimbach, P. Global reconstruction of historical ocean heat storage and transport. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1126–1131 (2019).
24. Reynolds, R. W. et al. Daily high-resolution-blended analyses for sea surface temperature. *J. Clim.* **1**, 5473–5496 (2007).
25. Veneranta, L., Vanhatalo, J. & Urho, L. Detailed temperature mapping–Warming characterizes archipelago zones. *Estuar. Coast. Shelf Sci.* **182**, 123–135 (2016).
26. Merkouriadi, I. & Leppäranta, M. Long-term analysis of hydrography and sea-ice data in Tvärminne, Gulf of Finland, Baltic Sea. *Clim. Change* **124**, 849–859 (2014).
27. Woolway, R. I. et al. Lake heatwaves under climate change. *Nature* **589**, 402–407 (2021).
28. Frölicher, T. L., Fischer, E. M. & Gruber, N. Marine heatwaves under global warming. *Nature* **560**, 360–364 (2018).
29. Rey, J., Rohat, G., Perroud, M., Goyette, S. & Kasparian, J. Shifting velocity of temperature extremes under climate change. *Environ. Res. Lett.* **15**, 034027 (2020).
30. Oliver, E. C. J. et al. Marine Heatwaves. *Ann. Rev. Mar. Sci.* **13**, 313–342 (2021).
31. Bennett, J. M. et al. The evolution of critical thermal limits of life on Earth. *Nat. Commun.* 1–9 (2021) https://doi.org/10.1038/s41467-021-21263-8.
32. Holbrook, N. J. et al. A global assessment of marine heatwaves and their drivers. *Nat. Commun.* **10**, 1–13 (2019).
33. Kniebusch, M., Meier, H. E. M., Neumann, T. & Börgel, F. Temperature variability of the baltic sea since 1850 and attribution to atmospheric forcing variables. *J. Geophys. Res. Ocean.* **124**, 4168–4187 (2019).
34. Merkouriadi, I. & Leppäranta, M. Influence of sea ice on the seasonal variability of hydrography and heat content in Tvärminne, Gulf of Finland. *Ann. Glaciol.* **56**, 274–284 (2015).
35. Haapala, J. Upwelling and its influence on nutrient concentration in the coastal area of the Hanko Peninsula, entrance of the Gulf of Finland. *Estuarine, Coastal and Shelf Science* **38**, 507–521 (1994).
36. Sorte, C. J. B., Fuller, A. & Bracken, M. E. S. Impacts of a simulated heat wave on composition of a marine community. *Oikos* **119**, 1909–1918 (2010).
37. Pansch, C. et al. Heat waves and their significance for a temperate benthic community: A near-natural experimental approach. *Glob. Chang. Biol.* **24**, 4357–4367 (2018).
38. Morón Lugo, S. C. et al. Warming and temperature variability determine the performance of two invertebrate predators. *Sci. Rep.* **10**, 1–14 (2020).
39. Humborg, C. et al. High emissions of carbon dioxide and methane from the coastal Baltic Sea at the end of a summer heat wave. *Front. Mar. Sci.* **6**, 1–14 (2019).
40. Laakso, L. et al. 100 Years of atmospheric and marine observations at the Finnish Utö Island in the Baltic Sea. *Ocean Sci.* **14**, 617–632 (2018).
41. Høyer, J. L. & Karagali, I. Sea surface temperature climate data record for the North Sea and Baltic Sea. *J. Clim.* **29**, 2529–2541 (2016).

42. Schlegel, R. W. & Smit, A. J. heatwaveR: A central algorithm for the detection of heatwaves and cold-spells. *J. Open Source Softw.* **3**, 821 (2018).
43. Schlegel, R. W., Oliver, E. C. J., Hobday, A. J. & Smit, A. J. Detecting Marine Heatwaves With Sub-Optimal Data. *Front. Mar. Sci.* **6**, 1–14 (2019).

Author contributions

N.G., A.N. and J.N. conceived the conceptualization for this synthesis study. N.G. performed the data curation, analysis and wrote the first draft of the manuscript. All authors contributed to the writing and editing of the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43247-022-00545-z.

Correspondence and requests for materials should be addressed to Norman Goebeler.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Regina Rodrigues, Clare Davis and Heike Langenberg. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.