

Large-Scale Drivers of Marine Heatwaves Revealed by Archetype Analysis

Christopher Chapman (✉ chris.chapman@csiro.au)

CSIRO <https://orcid.org/0000-0002-0877-4778>

Didier Monselesan

CSIRO <https://orcid.org/0000-0002-0310-8995>

James Risbey

CSIRO Oceans and Atmosphere <https://orcid.org/0000-0003-3202-9142>

Ming Feng

CSIRO Oceans and Atmosphere <https://orcid.org/0000-0002-2855-7092>

Bernadette Sloyan

CSIRO <https://orcid.org/0000-0003-0250-0167>

Article

Keywords:

Posted Date: April 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1495572/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Large-Scale Drivers of Marine Heatwaves Revealed by Archetype Analysis

Christopher C. Chapman^{1,2}, Didier P. Monselesan¹, James S.
Risbey¹, Ming Feng^{2,3}, and Bernadette M. Sloyan^{1,2}

¹*CSIRO Oceans and Atmosphere, Hobart Marine Laboratories, Castray Esplanade,
Hobart, Tasmania, Australia*

²*Center for Southern Hemisphere Ocean Research, Hobart Marine Laboratories,
Castray Esplanade, Hobart, Tasmania, Australia*

³*CSIRO Oceans and Atmosphere, Indian Ocean Marine Research Center, Crawley,
Western Australia*

March 28, 2022

Introduction

Extreme oceanic events, such as marine heatwaves, can have disastrous impacts on ecosystems and marine industries. Given their potential consequences, it is important to understand how broad-scale climate variability influence the probability of local extreme marine events. Here, for the first time, we employ an advanced data-mining methodology, archetype analysis, to identify large scale climate drivers and teleconnections that lead to marine extremes in certain regions. This methodology is applied to the Australasian region, where it identifies instances of anomalous sea-surface temperatures, frequently associated with marine heatwaves, as well as the broadscale oceanic and atmospheric conditions associated with those extreme events. Additionally, we use archetype analysis to assess the ability of a low-resolution climate model to accurately represent the teleconnection patterns associated with extreme oceanic temperatures, and discuss the implications for the predictability of these impactful events.

In recent years a number of high-profile and devastating marine heatwaves have brought increased public awareness and scientific focus on these events.^{1,2} The growing recognition of their impacts has resulted in an intense effort to understand the physical drivers of these phenomena, with the ultimate goal of improving their prediction and providing information to enable adaptation and

mitigation measures.^{3–7}

A key tenet of modern climatology is that conditions at a particular location may be influenced by remote drivers, often many thousands of kilometers away, through so-called *teleconnections*⁸ and there is a strong desire to understand the role played by teleconnections in driving impactful regional events, such as marine heatwaves.^{1,2,9–11} This desire is motivated by the fact that large-scale variability typically has longer timescales, is better represented in coarse-resolution climate models, and is hence more predictable than smaller-scale local processes.^{12–14}

To date, the vast majority of studies that investigate marine heatwaves and cold-spells focus on detailed case-studies of events at a particular geographic region,^{10,15–19} although there are a number of studies investigating the connection between larger regions and remote drivers.^{9,11,20,21} To link the local extremes with remote drivers, the general approach taken to-date is to begin by defining extreme events at one or more distinct locations, then explore statistical or dynamical connections between those events and large-scale climate modes such as an El-Niño.^{9–11,18,19} As the analysis proceeds from local to global scales, we will call this approach the ‘inside-out’.

While this ‘inside-out’ approach has dramatically advanced the understanding of the characteristics and physical drivers of marine extremes, unambiguously separating local and remote influences is difficult due to the complex interconnection between components within the climate system. As stated in a recent review:

However, fixed-region budget approaches are limited to analysing the drivers of marine heatwaves locally, while remote forcing and atmospheric and oceanic teleconnections can also be important contributors to the development and decline of marine heatwaves. Hence, there is merit in considering large-scale dynamical frameworks that connect remote drivers to marine heatwaves events ...¹

In this study, we present a new ‘outside-in’ methodology that directly identifies large-scale patterns associated with extreme sea-surface temperatures.^{22–24} To do so, we employ a powerful data-mining methodology –Archetype Analysis (herein AA)– that seeks to represent a high dimensional spatiotemporal dataset as a mixture of a finite number of ‘archetypal’ spatial patterns along with a probabilistic time-series. The archetypal patterns are themselves constructed as a weighted average of a small number of snapshots of the original dataset that correspond to ‘extreme’ states.²² Thus, an approximation of the original dataset, x , written as \tilde{x} , is given by:

$$x(\text{space, time}) \approx \tilde{x}(\text{space, time}) = \sum_i^P s_i(\text{time}) z_i(\text{space}) \quad (1)$$

where z_i is the i th archetypal pattern, P is the number of archetypes, and $s_i(\text{time})$ is the *affiliation time-series* of the i th archetype, which can take values between 0 and 1 and describes. The archetypal patterns can be interpreted as *extreme modes of variability*, and the affiliation probability is the likelihood that one of these mode is expressed at any given time.²⁴ AA has been employed previously to identify the characteristics of extreme rainfall events²⁵ and long-lived atmospheric regimes.²⁶ To the best of our knowledge, this work is the first application of AA to extreme marine events.

We apply AA to satellite derived sea-surface temperature (SST) over the Australasian region to identify large-scale patterns that correspond to temperature extremes (i.e. marine heatwaves and marine cold spells) and show that AA unambiguously identifies teleconnection patterns associated with extreme events. What is more, AA is able to reveal subtleties - for example distinguishing between the influence of ‘classical’ El Niños and central Pacific (Modoki) El Niños. Once the large-scale archetypal patterns have been obtained, we then investigate their impact on specific regions - hence ‘outside-in’. Finally, we present a new application of AA by assessing the capacity of a modern climate model to represent the teleconnections associated with marine temperature extremes.

Broad-Scale Drivers and Teleconnections Leading to Marine Temperature Extremes

Archetypal Patterns Associated with SST extremes

To begin our analysis, we apply AA to 39 years of SST anomalies over the southwestern Pacific and southeastern Indian Ocean basins (Fig. 1)(see Methods). In our analysis we chose (after experimentation) a total of eight archetypes, of which four associated with marine heatwave conditions in Australasia are shown in Fig. 1. Although the AA methodology is applied only to the Australasian domain (black box in Fig. 1) we plot the resulting archetypes by compositing over the southern Indo-Pacific to show the broad-scale SST patterns. The labelling of the archetypes is arbitrary.

In Fig. 1, we immediately recognise familiar spatial patterns associated with well known climate modes, such as the classical and central Pacific (Modoki) La Niña (archetypes #1 and #2) and El Niño patterns (archetypes #3 and #4). These inferences are supported by considering the affiliation time-series (solid black line in the right-hand column of Fig. 1) and the temporal distribution of the weights that are used to construct the archetypal patterns (orange bars). For example, the weights used to construct the archetype #4 cluster around the years 1998 and 2016, years known to correspond to years with powerful El Niños.²⁷ In addition to the ENSO-like patterns in the equatorial Pacific, additional features are evident. For example, elevated SST anomalies are evident along the west coast of Australia (archetype #1, Fig. 1a), around New Zealand

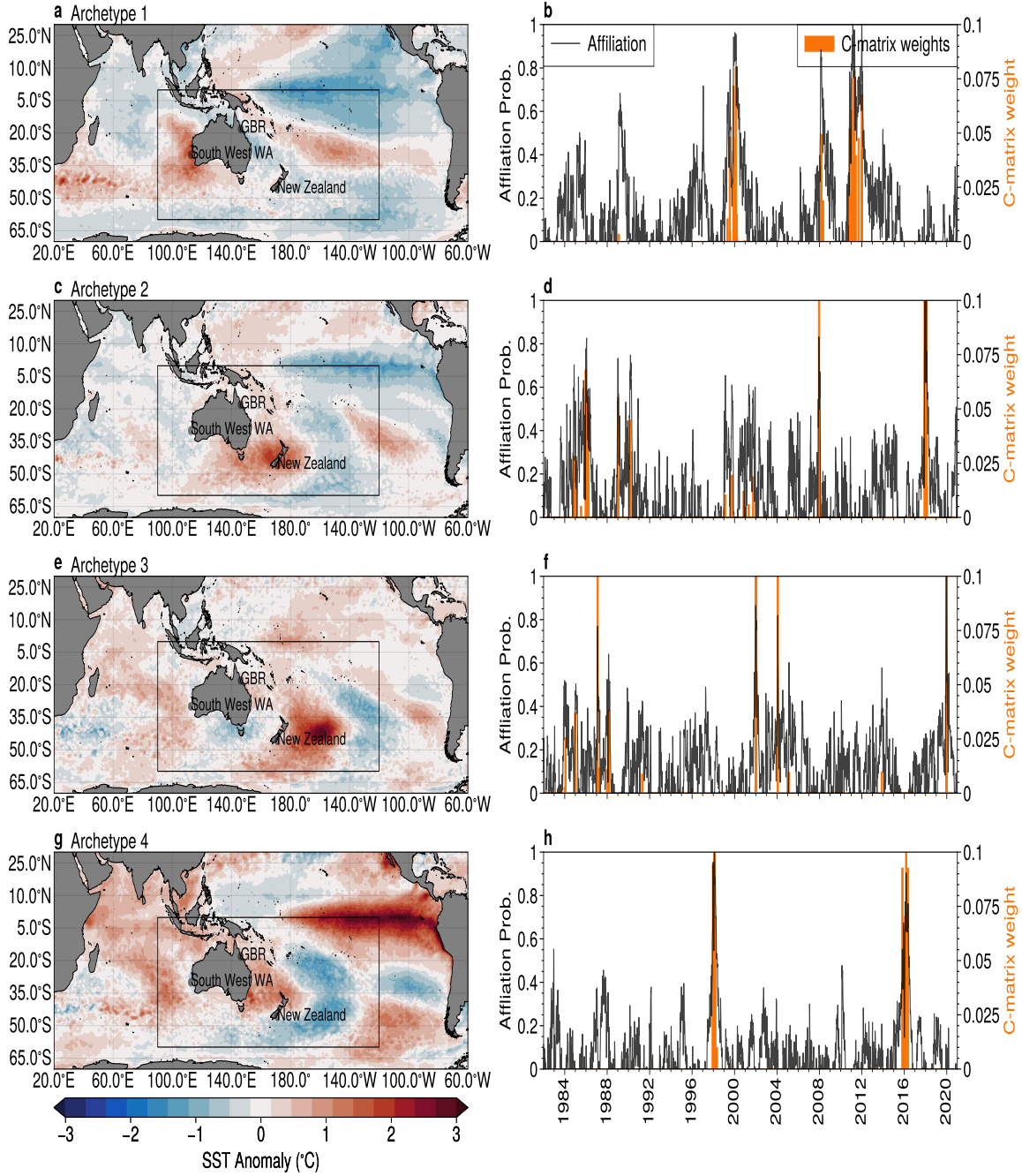


Figure 1: **Archetypal Patterns and Affiliation Time-Series:** (left) Detrended SST anomalies for four of the archetypal patterns computed over the Australasian region (indicated by the black box), and (right) associated affiliation time-series (black solid line) and the weights applied applied to each time snapshot to form the archetypes (known as *C*-matrix weights, orange bars, see methods). The four archetypes plotted here are selected based on their associated with well-known marine heatwaves (locations indicted in the text).

(archetype #2, Fig. 1c), and through Great Barrier Reef region (archetypes #3 and #4, Fig. 1e,g). These patterns show the large-scale conditions that are likely during periods with the affiliation time-series are close to 1, and their expression may bias the oceanic system in favour of extreme marine temperatures in particular regions.

Investigation of the affiliation time-series in Fig. 1 reveals periods of persistence and recurrence. We show this in further detail in Fig. 2a, where each coloured block corresponds to periods where a particular archetype is both dominant, and persists for at least 20 days. A number of persistent regimes can be identified, such as 16 month period from November 2010 until June 2012 period, when archetype #1 dominated, which manifested as an exceptionally strong La-Niña,²⁸ or the 6-8 months in 1998 and 2016 of archetype #4 that corresponded to powerful El-Niños. The inter-annual variability determined by summation of the number of event days for each dominant archetype in each year, shown in Fig. 2b, revealing substantial year-to-year variability, with certain regimes dominating in certain years while being completely absent in others. Hints of seasonality are indicated in Fig. 2c, which shows the total event days for each month. In particular, the archetypes that most clearly resemble El-Niño are (archetypes #3 and #4) show a clear expression in summer months.

The key result from the previous analysis is that AA reveals broad-scale, occasionally persistent modes of variability, and that these modes are associated with areas of warm and cold anomalies around Australia and New Zealand. We now follow the ‘outside-in’ methodology and investigate the links between regional extreme ocean temperatures and these broad-scale patterns through a series of case-studies.

South-Eastern Indian Ocean Marine Heatwaves

The southeastern Indian Ocean basin is a recognised global warming hot-spot, warming faster than the global average²⁹ was the location of one of the most intense and devastating known marine heatwaves, with temperatures of more than 3°C higher than the climatological average occurring in the summer of 2010-2011.^{16,30-32} We now investigate the relationship of broad-scale extreme patterns identified by AA to the occurrence of marine heatwaves in this region.

In Fig. 3a, we show the SST anomalies for the day of peak intensity of the 2010-2011 extreme marine heatwave event (1st of March 2011) at a representative location (30°S,112.5°E, indicated by the grey circle) as well as the composite average of SST anomalies for the day of peak intensity of all marine heatwaves at that location. The SST anomalies for both the single event snapshot (Fig. 3a) and the composite average (Fig. 3b) shows warm SSTs over a broad geographical range, from latitudes 20°S to 35°S, and longitude of 105°E and the western coastline of Australia, with the highest temperatures generally found

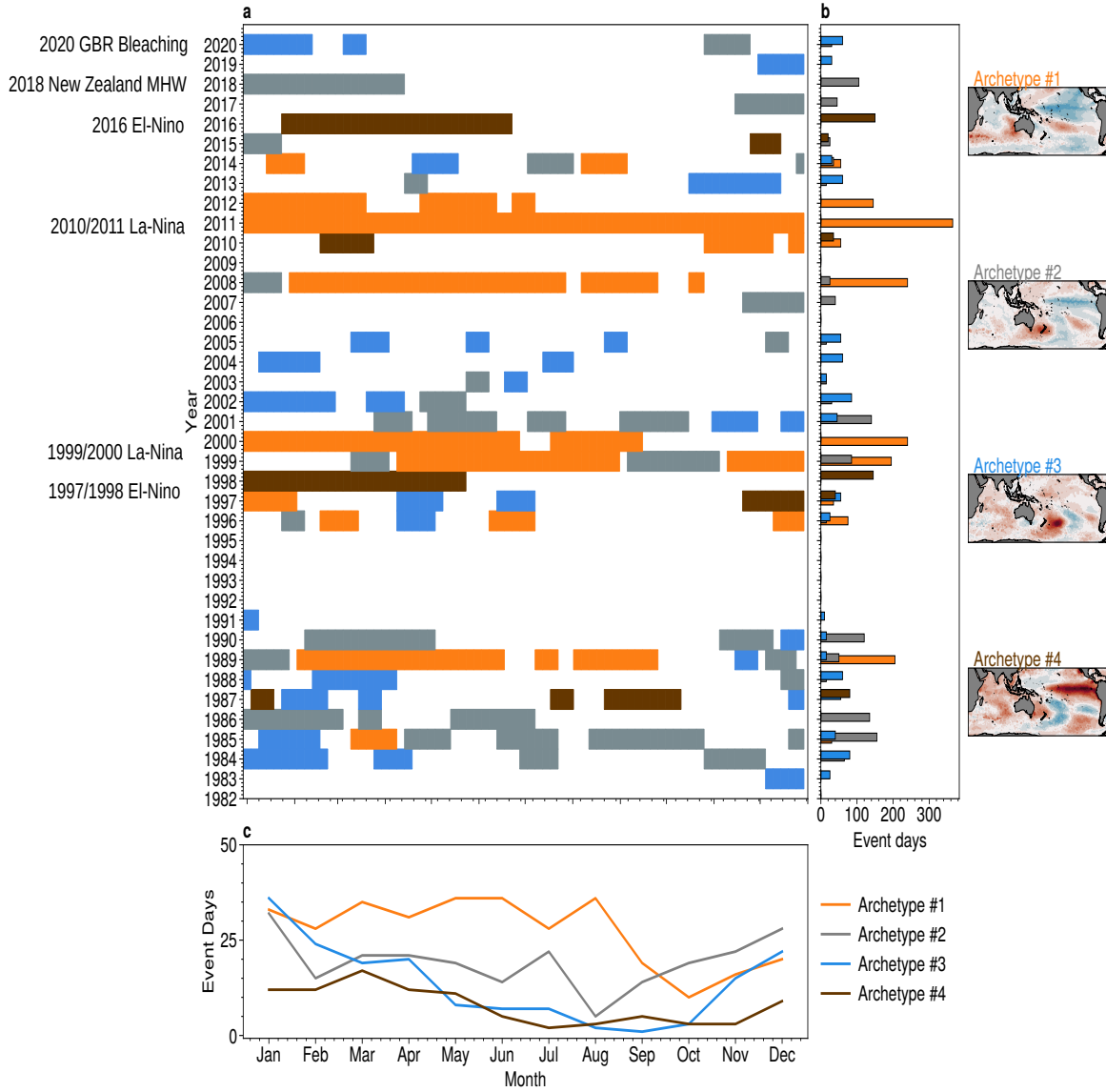


Figure 2: **Temporal Regimes of Archetypal Patterns:** **a** Coloured blocks indicate periods where a particular archetype was dominant for at least 20 days. The y -axis indicates the year, while the x -axis indicates the calendar day-of-year. Blanked periods show days where no qualifying event was found; the total number of archetype event days for each archetype that occur **b** for each year; and **c** for each month over the annual cycle. Maps to the right are to remind the reader of the spatial patterns of each archetype. Well known climatological events are highlighted by the in figure text.

closer to the continent. We identify a *best matching archetype* by comparing the spatial patterns shown in Fig. 3a,b with the archetypal patterns plotted in Fig. 1, in this case archetype #1 (Fig 4c). The archetype has an almost identical distribution of warm SSTs over the region to that of the SST composite average north of 35°S, and a strong similarity to the SSTs associated with the 2011 event.

The temporal relationship between the regional marine heatwaves and the archetypal extreme mode can be revealed by examining the anomalous SST time-series at our chosen representative location (Fig. 3d) and the affiliation time-series for the best matching archetype (Fig. 3e). The SST anomaly time-series shows low-frequency variation, with periods of above or below average temperatures that can persist for several months to several years. Marine heatwaves also show low-frequency behaviour, with periods of several occasionally high intensity events that occur clustered together, separated by longer periods with few, moderate intensity events.³³ Unsurprisingly, the periods of more frequent marine heatwaves are correlated with periods of higher than average SST.

Recall that the affiliation time-series, shown in Fig. 3e, can be interpreted as the probability that a particular archetype is expressed at a given time. In this case, the affiliation time-series is highly correlated with periods of above average temperatures at our representative location. For example, the affiliation time-series on the 1st of March 2011, the date of peak intensity of the extreme marine heatwave, is a maximum and approaches 1, indicating that the best-matching archetype is the dominant regime during this period. The affiliation probability is generally high during periods of high SST and frequent marine heatwaves and low during periods with few marine heatwaves (e.g. 1990–1996 and 2001–2008). The reconstruction of the representative SST anomaly using a single archetype (shown as the orange curve in Fig. 3d) also deviates from 0 only during periods with frequent marine heatwaves.

Having established the utility of a single archetype in capturing the broad-scale spatiotemporal variability associated with marine heatwaves in the south-east Indian Ocean, we now employ AA to identify the teleconnection patterns that accompany the extreme events in this region. The mean spatial-patterns for the satellite derived SST anomalies associated with archetype #1 are shown for the Pacific and Indian ocean basins in Fig. 4a, while the surface air temperature and the mid-tropospheric atmospheric circulation (represented by the 500hPa geopotential height anomaly and the 500hPa wind anomalies) are shown in Fig. 4b. In addition to the area of anomalously high temperatures along the western Australian coastline, we note anomalously cool SSTs in the equatorial Pacific, with a temperature minimum found at approximately 170°W, characteristic of the central Pacific La Niña (also known as La Niña Modoki).^{34,35} Investigation of the sub-surface temperature anomalies obtained from Argo floats along the equator associated with this archetype confirm this interpretation (Fig. 4c): a cool subsurface in the eastern Pacific down to around 300m depth, with a surface expression in the central Pacific, which co-occurs with a warm anomaly in the

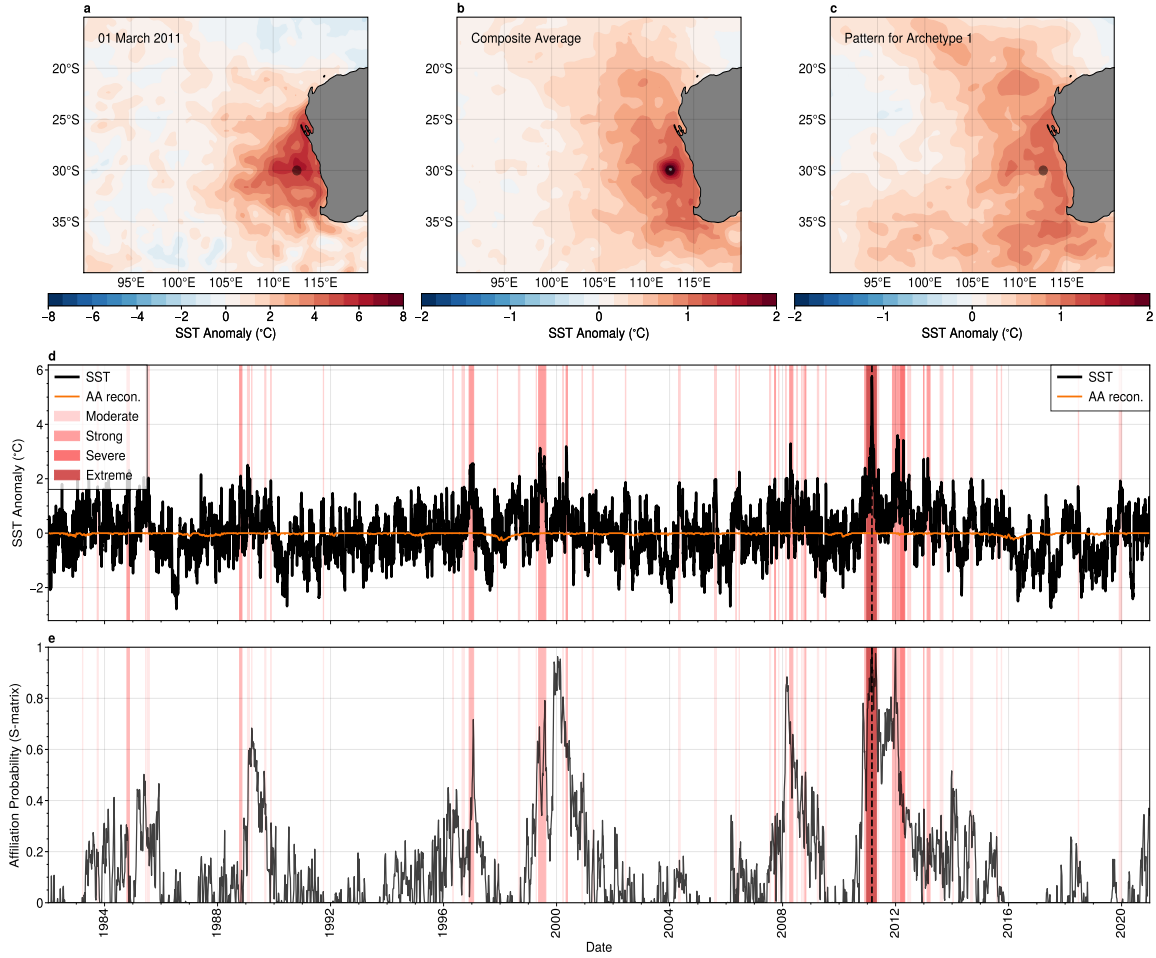


Figure 3: The relationship between Marine Heat Waves and Archetype #1 in the Southeastern Indian Ocean: **a** snapshot of SST anomalies for the peak of the 2010-2011 extreme marine heatwave event, which occurred on the 1st of March, 2011; and **b** SST composite average for all marine heatwave detected. Statistics are calculated at the representative location 30°S,112.5°E, indicated by the grey circle. **c** the SST anomalies for best matching archetypal pattern (archetype #1); **d** time-series of SST anomalies (black) and the reconstruction from archetype #1 (orange) at the representative location shown in panels **a–c**; **e** time-series of archetype affiliation probability for archetype 3. Colored bands in panels **d,e** indicate marine heatwaves occurrences, coded by the severity category described in *Hobday et al. 2018*³²

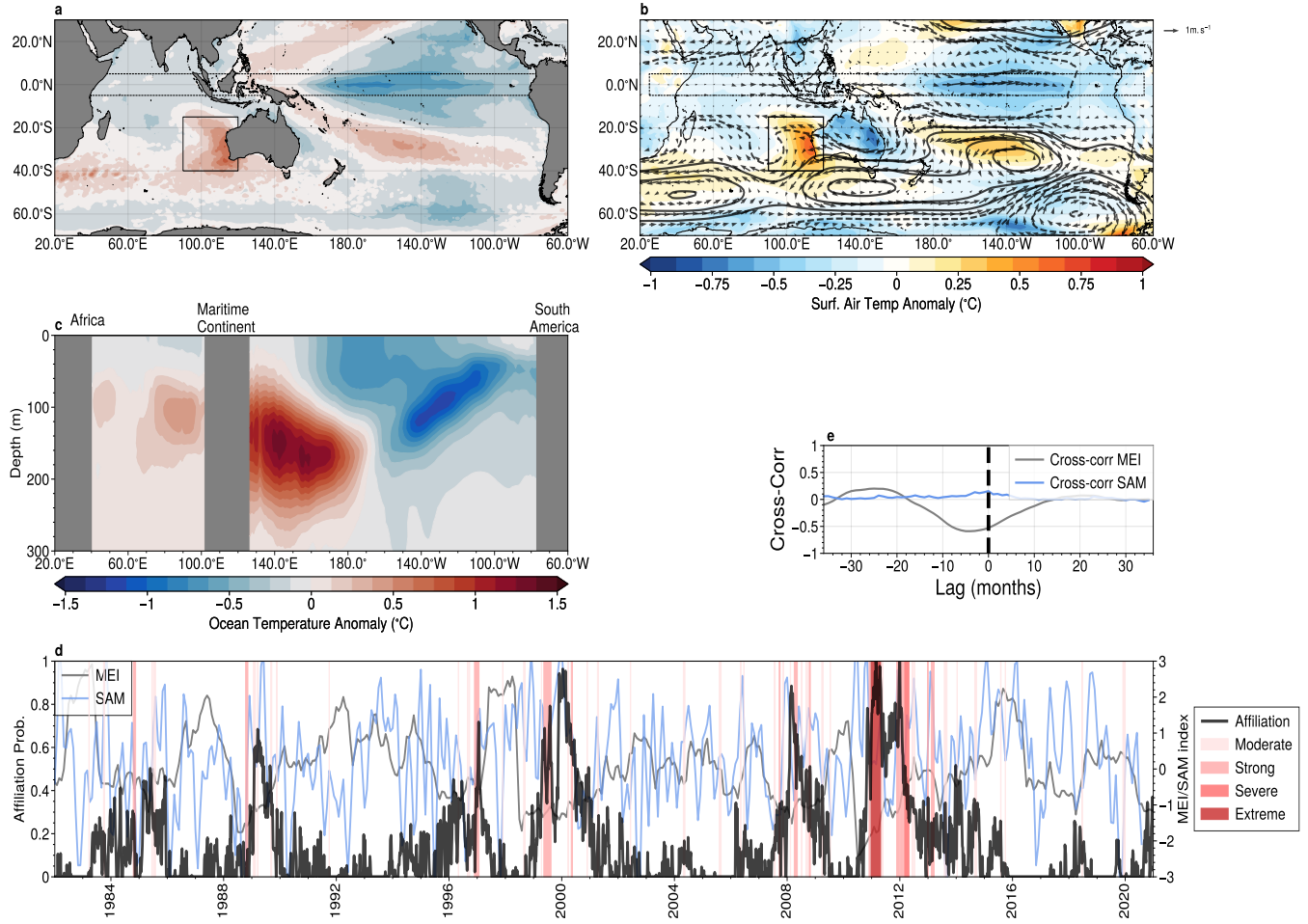


Figure 4: Teleconnections Associated with Marine Heat Waves in the Southeastern Indian Ocean: **a** SST anomaly; **b** surface temperature (colours) with anomalous mid-tropospheric (500hPa) geopotential height (contour lines, contour interval 5m) and winds (vectors); and **c** equatorial subsurface temperatures, associated with archetype #1, the best-matching archetype for marine heatwaves in the southeast Indian Ocean. **d** The affiliation time-series (solid black) together with the multivariate ENSO index (MEI, grey) and the Marshall SAM index (blue). Periods of marine heatwaves are indicated by red shading. **e** the lagged cross-correlation between the affiliation time-series and the MEI (gray) and the Marshall SAM index (blue). Negative lags correspond to the MEI/SAM index leading the affiliation.

western Pacific. The thermocline anomalies in the western Pacific are known to propagate via oceanic teleconnection to instigate the southeast Indian Ocean marine heatwave.³⁶ We note that there is much weaker sub-surface temperature expression in the equatorial Indian Ocean.

The anomalous mid-tropospheric atmospheric circulation associated with this archetype (Fig. 3b) shows warm surface air temperatures over the southeast Indian Ocean region, along with wind anomalies from the east, which are likely to contribute to the above average temperatures by bringing warmer air from the Australian continent over the ocean. Note that cyclonic wind anomalies in the southeast Indian Ocean are often associated with locally amplified marine heatwave events³⁷ which is more evident in near-surface wind field (not shown). Equatorial Pacific the surface air temperatures are suppressed and the mid-tropospheric winds show the characteristic divergence pattern associated with the central Pacific La Niñas.³⁴ A ridge of high pressure extends across the Southern Ocean and anomalously strong eastward winds associated with the polar jet stream are found south of 55°S, which is a feature characteristic of the positive phase of the Southern Annual Mode (SAM).³⁸

To further illustrate the relationship between the archetype and the climate modes, we plot the affiliation time-series in Fig 4d together with the Multivariate ENSO Index³⁹ (MEI) and the SAM index.⁴⁰ Significant anti-correlation between the affiliation time-series and the MEI can be seen in Fig. 4d. The apparent anti-correlation between these metrics is confirmed by a lagged cross-correlation analysis (Fig. 4e), which shows negative cross-correlation coefficient of ~ 0.5 with the MEI index leading the affiliation by ~ 6 months. We perform a similar lagged cross-correlation analysis between the the Marshall SAM index and the affiliation, finding only a weak correlation near zero lag, a maximum cross-correlation coefficient of 0.15.

We have clearly identified two different teleconnection patterns that contribute to the development of marine heatwaves in the southeast Indian ocean. The dominant driver is found to be central Pacific La Niña, with the SAM a potential secondary driver. We note that there could be other regional drivers of the marine heatwaves in the region, such as the Indian Ocean Dipole and regional air-sea coupling, which are not assessed in this study.

South Pacific and Tasman Sea Marine Heatwaves Near New Zealand

Our next case study concerns the South Pacific and Tasman Sea regions near New Zealand. This region was the site of a severe category marine heatwave in the Austral summer of 2017-2018, that co-occurred with extreme land temperatures.^{41,42} The impacts of this event were widespread, with the largest recorded

annual loss of glacier ice mass in New Zealand’s recorded history.⁴¹

We plot SST anomaly for the peak of the 2017-2018 marine heatwave event at a representative location (here 45.9°S, 171°E, 5th December 2017) and the composite average of all events at this location in Fig 5a,b. The spatial patterns in the single day snapshot and the composite average are very similar, albeit with different magnitudes, with warm SST centered near New Zealand’s south island (approximate longitude 170°E, latitude 45°S), extending west into the Tasman Sea.

The pattern of the best-matching archetype (archetype #2 of Fig. 1g-h), which we plot in Fig. 5c shows a remarkable visual similarity between the spatial patterns in Figs 5a,b to both the showing the same anomalously warm SST localised around New Zealand. The temporal evolution of the SST anomalies and the affiliation time series, shown in Figs 5d,e, showing that the extreme events are efficiently captured by a single archetype. As in the previous case study, marine heatwaves cluster, with a number of events occurring in a relatively short period of time, punctuated by longer periods with only a small number of isolated, weaker events. With only a single exception (between 1989 and 1991), the marine heatwave clusters occur during periods where the affiliation time series is persistently greater than 0.5. Examples of these periods are 1984–1987, 1999–2004, 2005, 2013–2015; and 2018–2020. We note also that the peak of the 2019-2020 severe marine heatwave (Fig. 5a) co-occurs with the absolute maximum of the affiliation time-series.

As before, we examine the broad-scale SST (Fig. 6a); mid-tropospheric atmospheric circulation Fig. 6b; and equatorial sub-surface temperatures (Fig. 6c) associated with the best-matching archetype. Concurrently with anomalously high SST centered on New Zealand (longitude $\sim 170^\circ\text{E}$, latitude $\sim 45^\circ$, indicated by the box in Figs. 6a–c), cooler SSTs are seen in the equatorial Pacific, extending from a longitude of 180° to South America, reminiscent of La Niña. In contrast to the previous case-study, both the anomalous atmospheric circulation, surface air temperatures (Fig. 6b) and the sub-surface ocean temperatures (Fig. 6c) are weak in the equatorial region. However, a strong *blocking* high pressure system can be seen in the atmospheric field to the east of New Zealand. The anomalous atmospheric circulation directs warm air from the north and reduces cloud cover over the region, consistent with previous work that attributes the severe 2017-2018 event to a similar persistent blocking system.^{41,42}

The spatial patterns shown in Figs. 6a,b suggest that marine heatwaves around New Zealand are associated with classical La Niña type patterns, as well as persistent atmospheric blocking high pressure systems. However, a lagged cross-correlation shows only weak correlation of the affiliation time-series with the multivariate ENSO index (Fig. 6d and inset panel), with a peak correlation coefficient of -0.25 at a lag of zero, and we note that the magnitude of the ocean

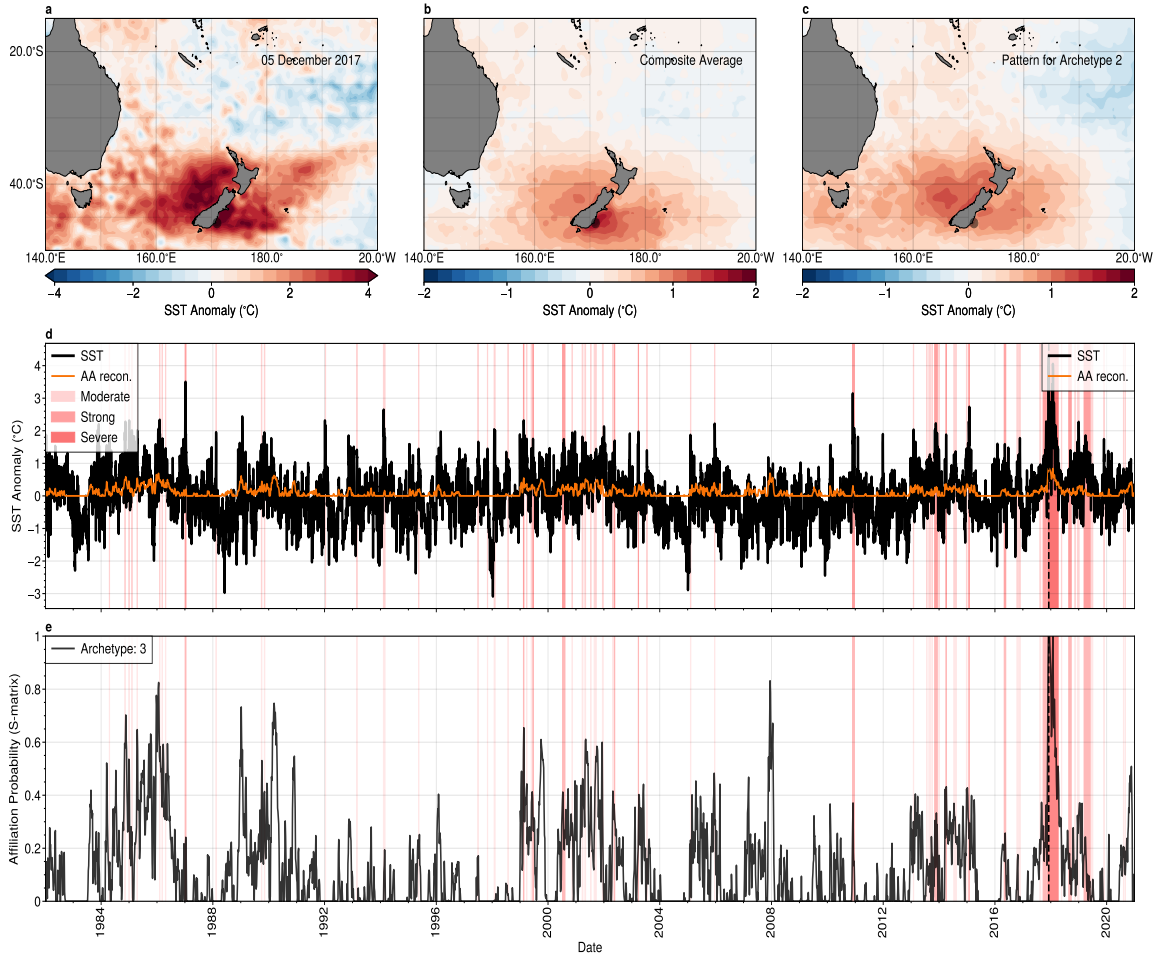


Figure 5: The relationship between Marine Heat Waves and Archetype #2 in the Tasman Sea: **a** snapshot of SST anomalies for the peak of the 2017-2018 severe marine heatwave event, which occurred on the 5th of December, 2017; and **b** SST composite average for all marine heatwave detected at a representative location. Statistics are calculated at the representative location 45.9°S, 171°E, indicated by the grey circle. **c** the SST anomalies for best matching archetypal pattern (archetype #2); **d** time-series of SST anomalies (black) and the reconstruction from archetype #2 (orange) at the representative location shown in panels **a-c**; **e** time-series of archetype affiliation probability for archetype #2. Colored bands in panels **d,e** indicate marine heatwaves occurrences.

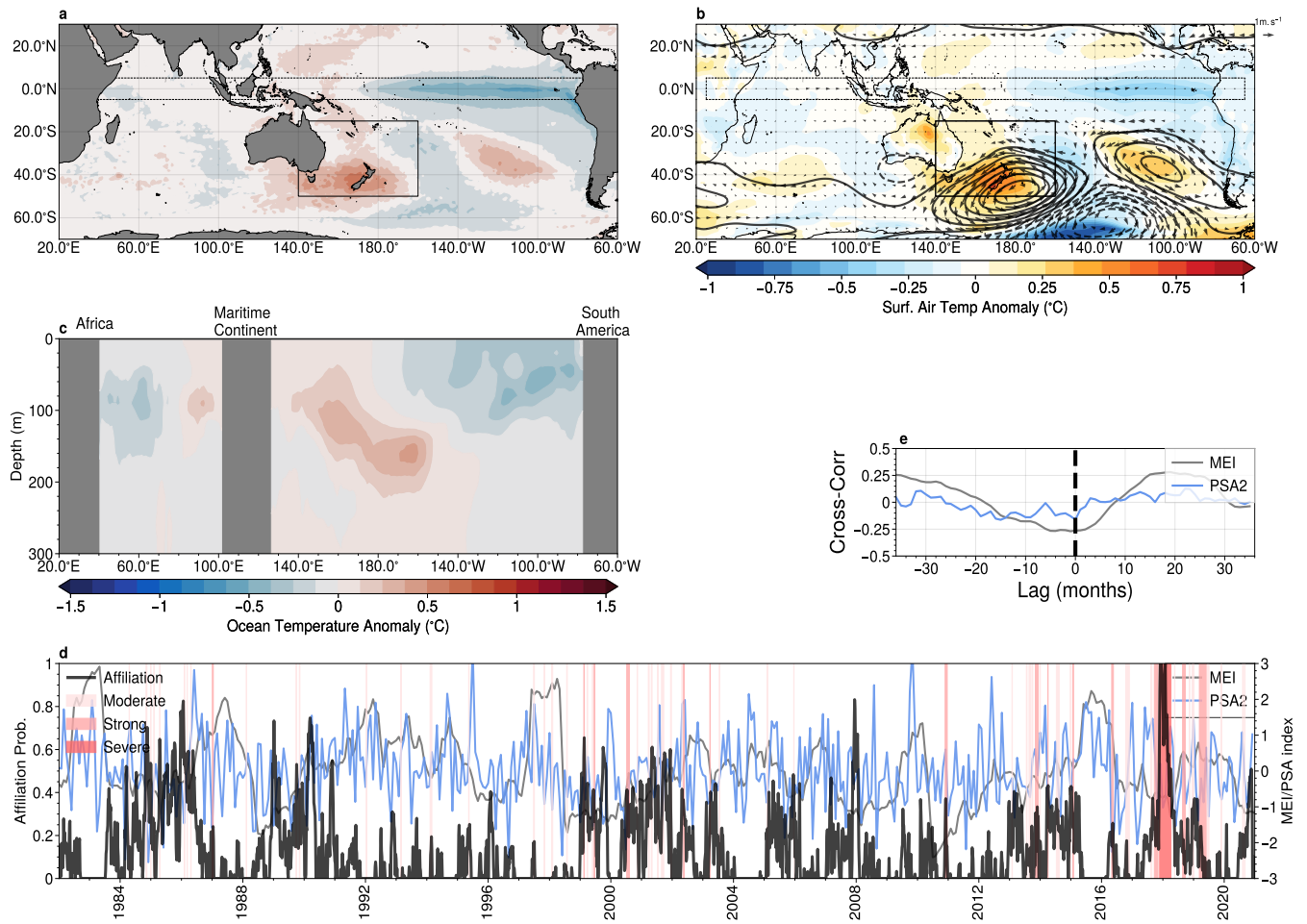


Figure 6: **Teleconnections Associated with Marine Heat Waves in the Tasman Sea:** As in Fig. 4, with the exception that the best-matching archetype is now archetype #2, and in panels **d** and **e** the blue line corresponds to the Pacific South America (PSA) index.

temperature anomalies in Figs 6a,c are weak, which suggests that forcing from La Niña conditions in the equatorial Pacific plays a only a secondary role to that of the intense atmospheric blocking.

Previous studies have also attributed marine heatwaves in the region to forcing associated with the Southern Annular Mode.^{41,42} However, the lagged cross-correlation shows no strong correlation between the affiliation time-series and the SAM index, with peak correlation coefficient of less than 0.1, which does not exceed the threshold for statistical significance. Additionally, the spatial patterns of anomalous atmospheric circulation are not strongly reminiscent of the SAM.³⁸ The spatial patterns are, however, similar to the anomalous atmospheric circulation associated with a mode of climate variability known as the Pacific South-America (PSA) pattern. The monthly PSA index,⁴³ is plotted together with the affiliation time-series in Fig. 6d. Despite the presence of high-frequency fluctuations, there is evidence of anti-correlation of the PSA index with the affiliation time-series. However, we note that the relationship is weak, with a peak correlation coefficient of ~ 0.15 .

The analysis conducted here suggests that although both classical La-Niña and PSA climate modes play a role on driving marine heatwaves around New Zealand, individually each of these climate drivers has only a weak influence, with the dominant role being played by localised atmospheric blocking high-pressure systems. However, blocking highs have long been associated with extreme weather events, including marine heatwaves,^{17,44} there is currently no generally accepted theory that completely explains their dynamics.⁴⁵ Further investigation using the techniques developed in this study could provide further insight into the dynamical origins of these phenomena and their interaction with broad-scale climate states.

Coral Sea and Great Barrier Reef Marine Heatwaves

For our final case studys, we turn our attention to marine heatwaves in the Coral Sea and Great Barrier Reef (GBR) to the north-east of Australia. Summertime marine heatwaves are known to induce mass coral bleaching events, as high ocean temperatures are a necessary (but not sufficient) condition for coral bleaching.^{46,47} The GBR suffered heat-induced mass bleaching events in 1998, 2002, 2006, followed by 3 events during the period 2016, 2017 and 2020.⁴⁸

In Fig. 7a, we show the composite average of anomalous SST for all summertime (December, January and February) marine heatwave events at a location representative of the “central” and “northern” regions of the GBR, as well as daily snapshots of anomalous SST for two marine heatwave events: March 2016 (Fig. 7b) and February 2020 (Fig. 7d), that were implicated in instances of mass coral bleaching. The composite average SST anomaly for all marine heatwave events shows the anomalies close to the Australian coastline, tending cooler

further offshore. A similar SST pattern can be seen in the daily snapshot of SST anomalies for the 2020 marine heatwave event (Fig. 7b). The 2016 event, in contrast, shows elevated SSTs extending further north and more broadly over the Coral Sea.

Unlike in the previous case studies, we find that at least 2 archetypes, archetypes #3 (Fig 7c) and #4 (Fig. 7e) are required to capture summertime marine heatwaves in the GBR region. These archetypal patterns show a similar spatial structure to the daily snap-shots corresponding to the peaks of the two events shown in Figs. 7b,d, and are more strongly expressed during the summer months (see Fig. 2c). The utility of using two archetypes becomes clear when we investigate the relationship between the affiliation time-series and the SST anomalies at the representative location (Figs. 7f,g), shows that marine heatwaves associated with major coral bleaching periods are well captured by one or the other archetype, with the only major exception being the 2017 coral bleaching event. Additionally, we note that marine heatwaves at this location *always* occur when the affiliation of one of our two best-matching archetypes exceeds 0.5.

We now investigate large-scale patterns associated with summertime marine heatwaves in the GBR region. The anomalous SSTs (Fig. 8a), and surface air-temperatures (Fig. 8b, colors) for archetype #4, which was strongly expressed during the severe 2016 coral bleaching event. The large-scale SST patterns shows strong positive anomalies (of 1–1.5°C) in the equatorial Pacific, characteristic of the mature phase of classical El-Niño conditions. Concurrent warm surface temperatures through the majority of the Coral sea, northern Australia and the maritime continent. Anomalous mid-troposphere circulation (Fig. 8b) shows further El-Niño like conditions, with weak westward trade-winds over the Coral Sea, reinforced by a blocking high pressure system over New Zealand. The presence of El-Niño conditions can also be seen clearly in the sub-surface ocean temperatures (Fig. 8c), with anomalously warm temperatures in the eastern basin, and corresponding cooler sub-surface temperatures in the western basin.

The affiliation time-series associated with archetype #4, shown in Fig. 10d, shows a clear positive correlation with the MEI, with the later tending to have a reach a maximum 2 month prior to a maximum in the affiliation probability. The lead-lag relationship with the MEI suggest that marine heatwaves in the GBR region may occur two to three months after the peak of El Niño events, which is consistent with the published literature.^{47,49}

We have performed a similar ‘zoomed out’ analysis of archetype #3, shown in Fig. 9. However, unlike with previous case-studies, the broad-scale spatial patterns associated with archetype #3 show no strong signatures in either oceanic or atmospheric fields near GBR region, with the exception of warm but relatively weak temperature anomalies in the central Pacific, which resemble the decaying phase of a central Pacific (Modoki) El-Niño. We note a weak pos-

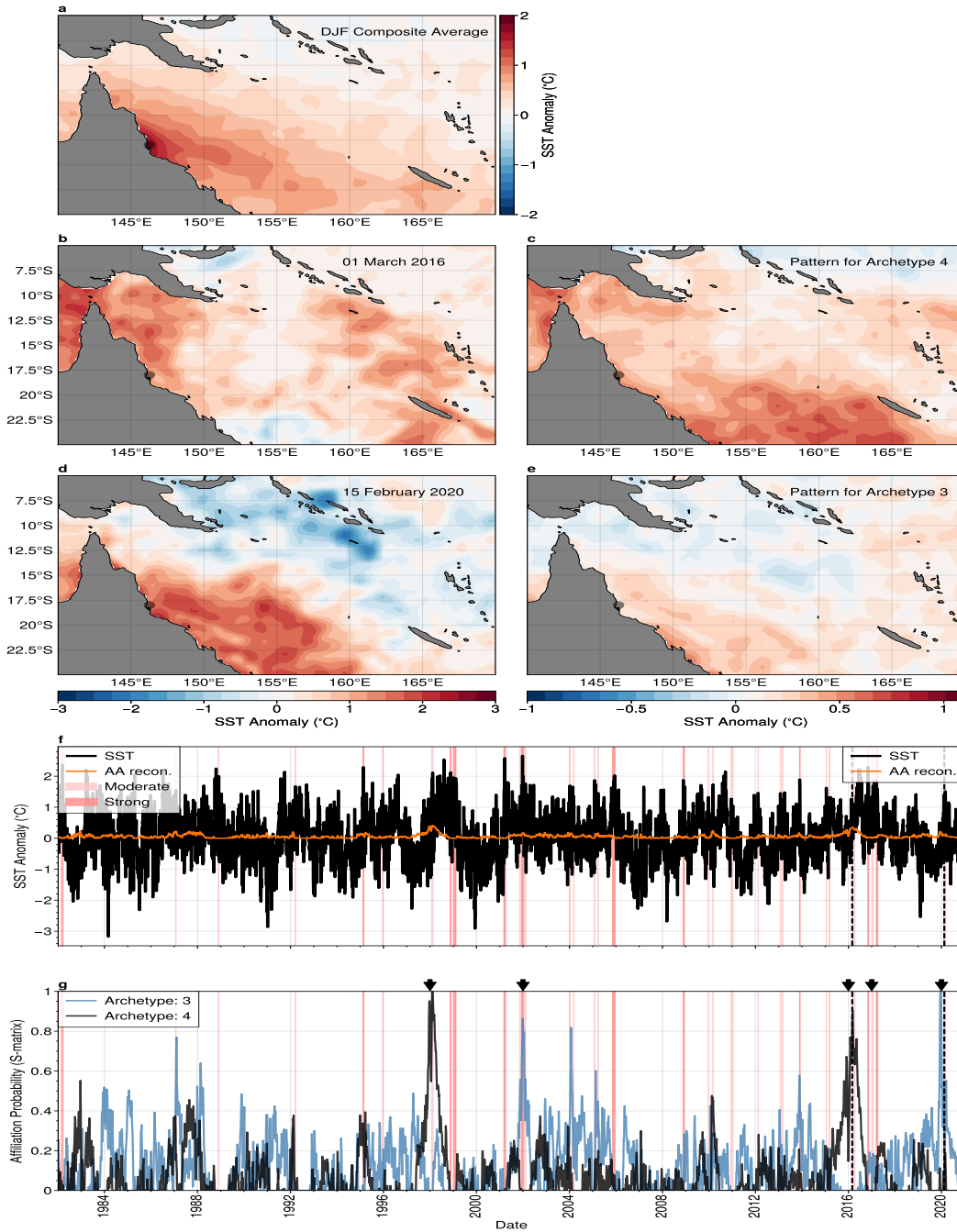


Figure 7: **The relationship between Marine Heat Waves and Archetypes #3 and #4 in the Great Barrier Reef region:** **a** Composite average of the SST anomaly for all summertime marine heat waves at a representative location (45.9°S, 171°E), indicated by the grey circle; **c, d** daily SST anomaly snapshot for the peak of 2016 marine heatwave; **c** the SST anomalies for best matching archetypal pattern for the 2016 event (archetype #4); **d, e** as in **c, d** for the 2020 marine heat wave; **f** time-series of SST anomalies (black) and the reconstruction from archetype #3 and #4 (orange) at the representative location shown in panels **a-c**; **e** time-series of archetype affiliation probability for archetypes #3 and #4. Colored bands in panels **d, e** indicate summertime (DJF) marine heatwaves occurrences, coded by the severity category. Black arrows in panel **g** indicate the occurrence of mass coral bleaching events.

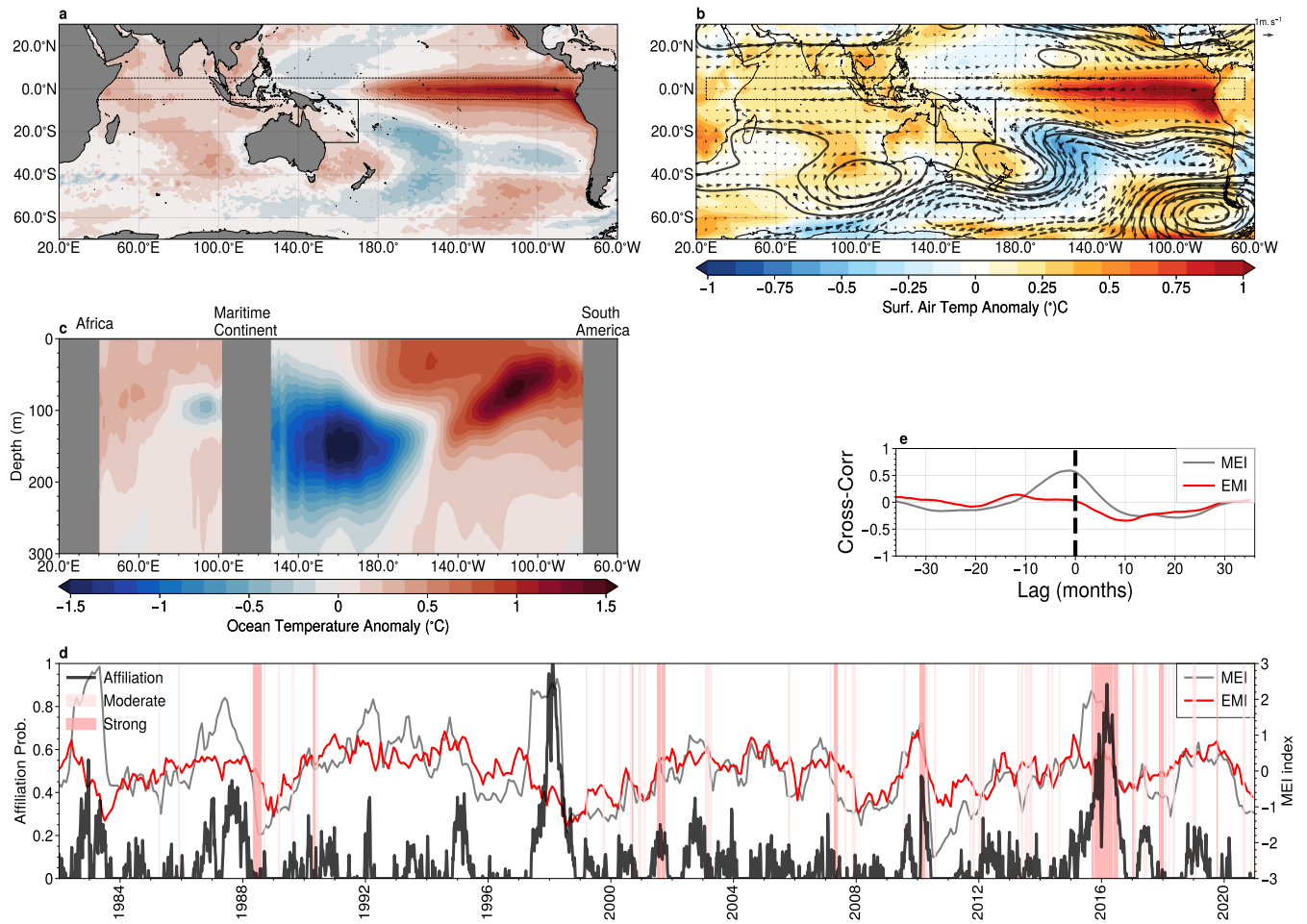


Figure 8: Teleconnections Associated with Marine Heat Waves in the Coral Sea and Great Barrier Reef: As in Figs. 4 and 6, with the exception that the best-matching archetype is now archetype #4, and in panels d and e the red line corresponds to the El-Niño Modoki Index (EMI) index.

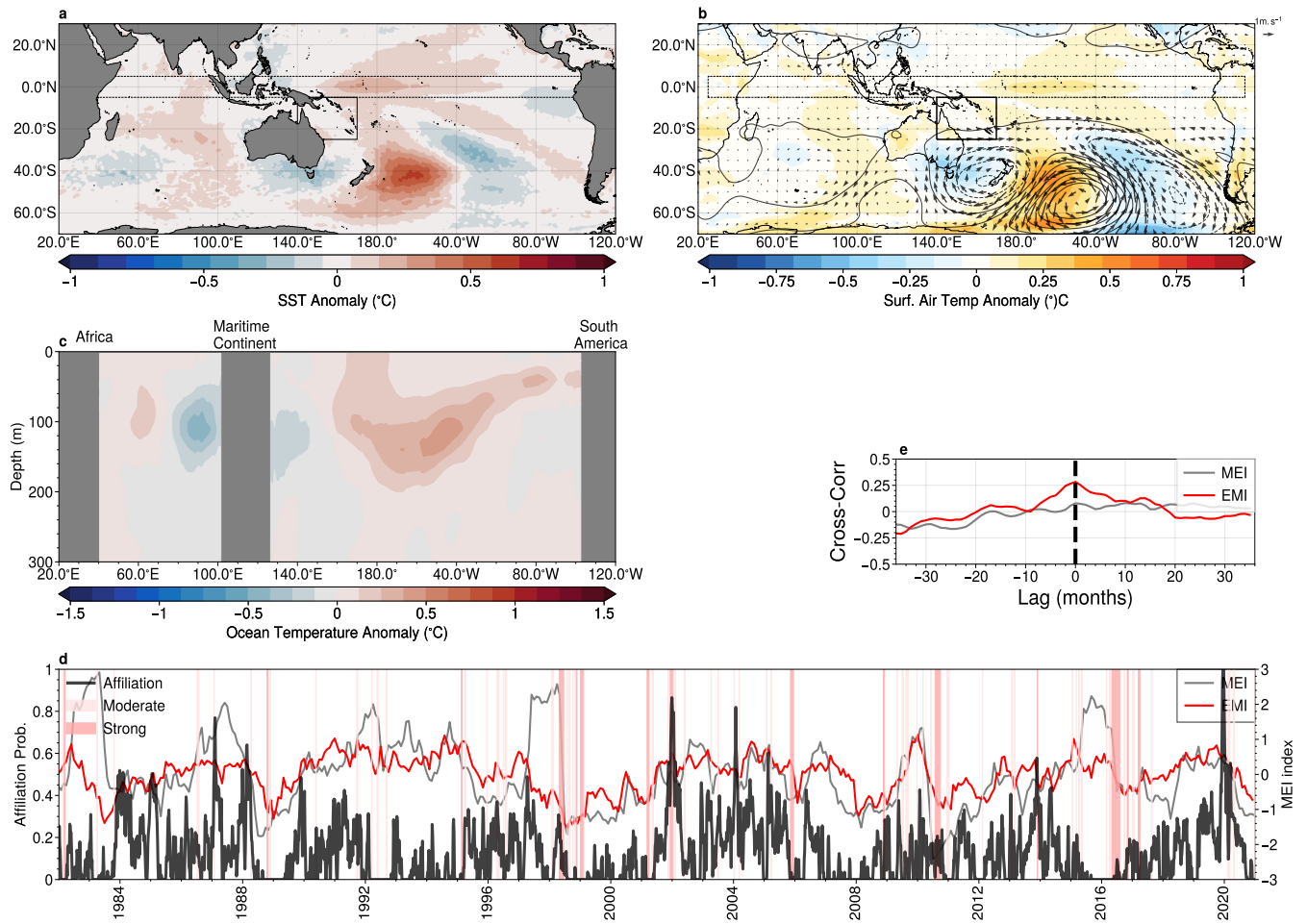


Figure 9: Teleconnections Associated with Marine Heat Waves in the Coral Sea and Great Barrier Reef: As in Figs. 8 with the exception that the best-matching archetype is now archetype #3, and in panels d and e the red line corresponds to the El-Niño Modoki Index (EMI) index.

itive correlation (Fig. 8b, maximum value of ~ 0.25 at zero lag) with the ENSO Modoki index and no significant correlation with the MEI.

While archetype #4 can be easily interpreted as an El-Niño like pattern, the interpretation of archetype #3 is more ambiguous, suggesting a role for local dynamics not identified by the large-scale patterns extracted by AA.⁵⁰

Additional Case Studies

In this study, we have focused on marine heatwaves that are efficiently described by AA. However, AA is capable of representing cold extremes as well. To illustrate this, two case studies of marine cold spells are included in the supplementary material. Additionally, it is important to note that the power of AA lies in its ability to recognise extreme states over large spatial scale and hence that extremes at a regional scale that are driven by local processes may not be well captured by AA. An additional case study is provided to demonstrate a case that is not well captured by AA.

Evaluation of Teleconnections associated with extremes in a Climate Model

Numerical ocean and climate models are a vital tool employed both to predict distinct extreme events on timescales of days to weeks,^{3,4,7} for projections of their statistics, such as frequency, intensity and duration, years and decades into the future as the climate warms^{20,51} or for climate change attribution of particular events.¹⁹ However, as is well known, climate models are imperfect representations of reality, and the representation of extreme events in numerical models is sensitive to the model resolution and biases.^{20,52} Climate models, particularly at the coarse resolution used for climate projections, typically do not capture the tail of the temperature probability distribution, and produce ‘extremes’ that are simply not as intense or frequent as in reality^{52,53} (see also supplementary material).

However, if climate models are able to capture the broad-scale fields and teleconnections associated with local extreme events, it may be less important that the model is incapable of representing the subtleties of those events at the local scale. For example, a model may well approximate the ENSO teleconnection patterns that are associated with increased probability of extreme events in certain regions, even if the model does not capture the localised extreme events themselves. Down-scaling may improve the representation of the local extremes, but only in the case that the teleconnections are captured by the low-resolution model.

Here, AA is employed to assess the capability of a coupled climate model to represent the extreme broad-scale patterns. We apply the technique to a long run of a climate model (the Australian Community Earth Systems Simulator

Decadal - ACCESS-D) with steady radiative forcing set at perpetual 1990 levels⁵⁴ (see methods). Eight archetypes are obtained from the final 39 years of detrended model SST anomalies (ie. the same length as satellite SST observations), over a domain identical to the observational case-studies.

We show the four climate model archetypal most similar to those utilised in the previous case studies in Fig. 10. For each archetypal pattern we show the large-scale SST anomaly (left column), the surface air-temperature anomaly and anomalous mid-tropospheric circulation (center) and the affiliation time-series along with the C-matrix weights used to construct the archetypes (right).

In the case of south-east Indian ocean (Fig. 10a–c) and Tasman Sea marine heatwaves, we note a strong similarity between the climate model archetypal patterns and those obtained from the observations (shown in Figs 4 and 6). In the case of the Southeast Indian region, the model archetypal patterns show similar anomalous SSTs along the west Australian coastline and cool equatorial Pacific SSTs, reflective of La-Niña like conditions. The model places the coolest SST anomalies further to the west than in the observations. The climate model also accurately reproduces the broad atmospheric circulation anomalies and surface air temperatures over the Australian continent (Fig. 10b). The archetypal patterns also reveal that the model captures conditions similar to those identified in the New Zealand case study (Fig. 10d–f). In particular, a large atmospheric blocking high pressure system that closely corresponds to the region of highest SST anomalies near New Zealand, although we note that the center of action is shifted significantly to the south and east when compared with observations.

However, when we consider teleconnections associated with extreme the model’s El-Niño like modes, shown in Figs. 10g–i and 10j–l, which are similar to those studied in the GBR case study, we find that the teleconnections are less reflective of observed patterns. In particular, we note that, in the equatorial Pacific, the high SST is significantly to the west of the observed position, which results in anomalously cool SSTs in the GBR and Coral Seas, instead. AA reveals clearly how biases in the representation of the equatorial Pacific impact the model’s ability to simulate important teleconnections.

To summarize, we apply AA to assess the ability of a climate model to accurately represent the extreme climate modes previously identified in our observational case-studies. Remarkably, the model produces modes that are similar to those obtained from the satellite observations. However, subtle model biases, such as the position of warm equatorial SST in the model’s simulation of El-Niño, can strongly influence the model’s teleconnections. The results of this analysis have implications for prediction of marine heatwaves at timescales longer than a few weeks, or projection of marine heatwaves in future climate states, as only those models capable of simulating the extreme climate states will be capable of predicting local extreme events related to the broad-scale

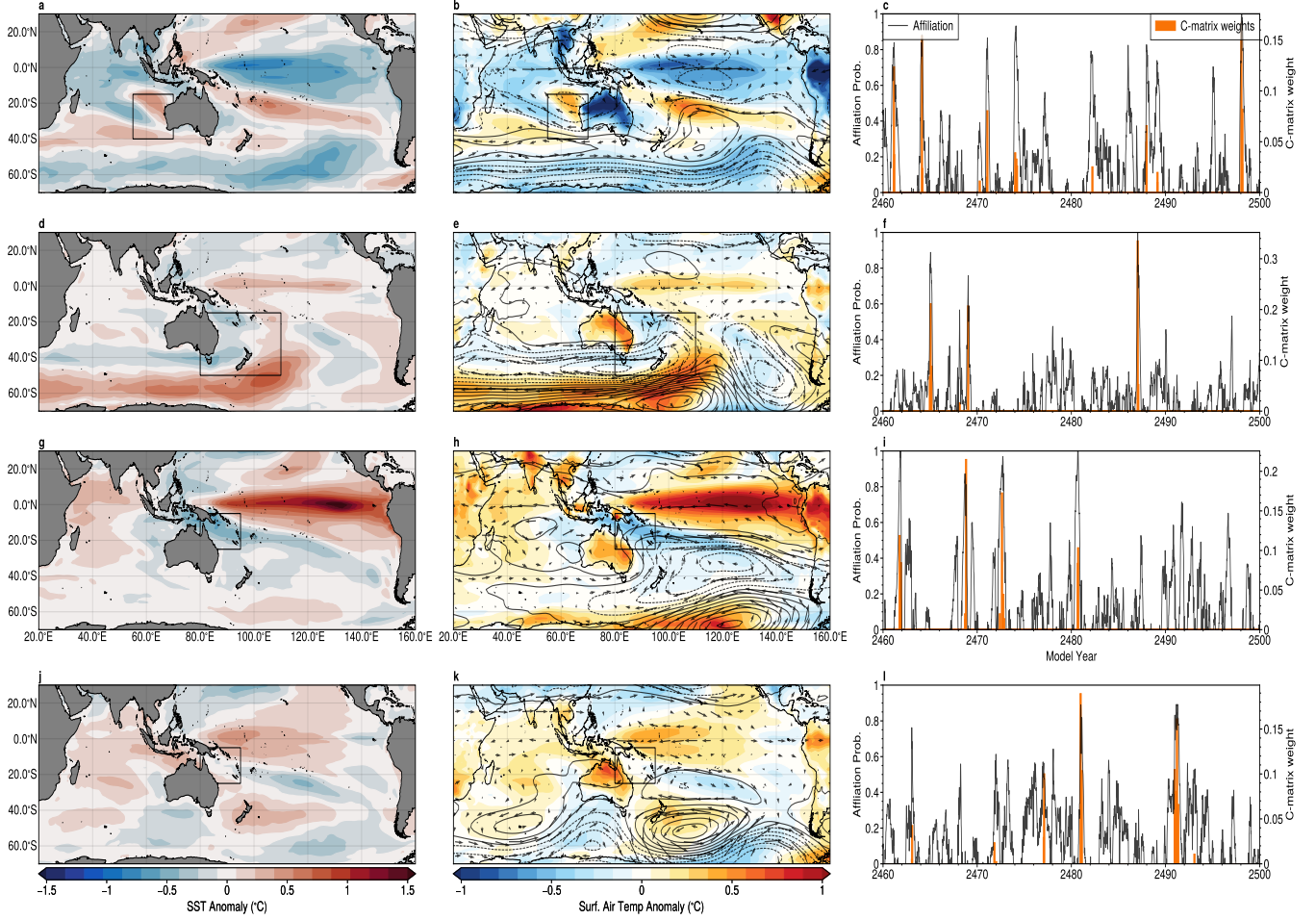


Figure 10: **Broad-scale patterns associated with marine extremes in a climate model:** anomalous SST (left), surface air temperature (colours) and mid-tropospheric (500hPa) atmospheric circulation patterns (center), and affiliation time-series (right) that correspond the archetypes that marine heatwaves in the **a–c** southeast Indian ocean, **d–f** southwestern Pacific near New Zealand; and **g–i** Great Barrier reef/Coral Sea region.

477 patterns, and there an 'outside-in' approach to assessing extremes in a climate
478 model is of obvious utility.

479 Perspectives and Conclusions

480 In this work, we demonstrate that our novel 'outside-in' approach to charac-
481 terizing extreme SST patterns using archetypal analysis (AA) is able to clearly
482 identify the relationships between large-scale oceanic and atmospheric condi-
483 tions and certain regional marine heatwaves. AA provides a minimal description
484 of these extreme regional events, reducing the complex multi-faceted system to
485 only one or two variables. The power of this approach is shown through several
486 case studies, that identify not just the climate mode most likely to be associated
487 with regional extreme SSTs (e.g. El Niño or La Niña), but also the atmospheric
488 and oceanic teleconnection patterns, and the temporal relationships between the
489 those climate modes and the expression of the archetype, as well as the impor-
490 tance of the flavour or phase of ENSO (i.e. classical or central Pacific/Modoki).

491
492 Of the events studied here, both El-Niño and La-Niña conditions, in their
493 various forms, are identified as the dominant influence on marine heatwave
494 occurrence in two of the three case studies (the southeast Indian Ocean and
495 GBR/Coral Sea), and perhaps a secondary influence in the 3rd (Southwest Pa-
496 cific/New Zealand events). However, our analysis has revealed that the flavour
497 and phase of ENSO plays an exceedingly important role as well. For example,
498 our investigation of the southeast Indian ocean region has shown that marine
499 heatwaves are frequent and intense in this region only during the central Pa-
500 cific/Modoki phase of La Niña. Application of our method to numerical model
501 output has presented an alternative methods for the assessment of the repre-
502 sentation of extremes in short term climate forecasts and climate projection.
503 Extremes

504
505 It is important to note that while AA efficiently describes the large-scale
506 patterns associated with extremes, it will not capture all individual events, par-
507 ticularly those driven by local processes. This should not be seen as a drawback,
508 as this fact provides a mechanism for distinguishing between events driven by
509 large-scale climate modes and those dominated by local processes. We have also
510 not attempted to use AA to diagnose the distinct physical drivers of the events
511 under study, and further work will prioritise blending process-based understand-
512 ing with data-driven approaches. Although applied to marine heatwaves in this
513 study, the approach can, in principle, be applied to a wide range of other phys-
514 ical phenomena, such as sea-level extremes or terrestrial heatwaves.

515
516 The novel approach presented in this study provides a viable and robust man-
517 ner for linking large-scale variability and regional extreme events that, in-turn,
518 provides an improved understanding of the links between large-scale drivers and
519 the local impacts.

520 Materials and Methods

521 Data

522 Sea Surface Temperature

523 In this study, we use a satellite derived sea-surface temperature product as our
524 primary dataset: version 2.1 of the *National Oceanic and Atmospheric Adminis-*
525 *tration's Optimum Interpolation SST Advanced Very High Resolution Radiome-*
526 *ter only* product (NOAA OISST-AVHRR only) which has data for the period
527 1st January 1982 to 31st December 2020 (hence 39 years)^{55,56} and anomalies are
528 computed relative to this period. Data are provided at daily output frequency
529 on a regular $0.25^\circ \times 0.25^\circ$ regular latitude/longitude grid.

531 Atmospheric Reanalysis

532 The atmospheric reanalysis used in this study is the *Japanese 55-Year Reanal-*
533 *ysis* (JRA55)^{57,58} provided on a $1.25^\circ \times 1.25^\circ$ latitude/longitude grid. In our
534 analysis, we use daily means of the original 6 hourly output, and restrict our at-
535 tention to the period 1st January 1982 to 31st December 2020 (i.e. an identical
536 periods to that of SST data).

537 Subsurface temperature

538 The subsurface temperature data employed is the optimally interpolated prod-
539 uct developed at the Scripps Institute of Oceanography,⁵⁹ which is based on
540 profiles obtained by the international Argo program.⁶⁰ This product provides
541 estimates of ocean temperatures from the surface to 2000db of depth, on a reg-
542 ular $1^\circ \times 1^\circ$ latitude/longitude grid. These data are only available from the year
543 2004 onward. The Argo temperature profiles were collected and made freely
544 available by the International Argo Program and the national programs that
545 contribute to it. (<http://www.argo.ucsd.edu>, <http://argo.jcommops.org>). The
546 Argo Program is part of the Global Ocean Observing System.⁶⁰

547 Climate mode indices

548 In our regional case studies, we correlated affiliation time-series for the best-
549 matching archetype with the various climate indices to illustrate the connection
550 between the extreme modes identified by the AA and more familiar climate
551 modes. The source of these indices are:

- 552 • **Multivariate ENSO (MEI) Index** (MEI): A measure of the variability
553 in the equatorial Pacific (30°S - 30°N and 100°E - 70°W) that uses principal
554 component analysis to combine 5 oceanic and atmosphere variables (sea
555 level pressure, sea surface temperature, zonal and meridional components
556 of the surface wind, and outgoing long-wave radiation) into a single in-
557 dex.^{39,61} Large, positive values correspond to El-Niño conditions, while

large negative values correspond to La-Niña conditions. Values are provided at monthly frequency from the December-January 1979 to present. These data were sourced from the National Oceanic and Atmospheric Administration’s Physical Science Laboratory (<https://psl.noaa.gov/enso/mei/>);

- **Marshall Southern Annular Mode (SAM) Index:** The strength of the SAM reflects the position and strength of the westerly winds that blow over the Southern Ocean between latitudes of 40°S and 65°. The Marshall SAM index is constructed from sea-level atmospheric pressure observations taken at 12 weather stations on the Antarctic continent and some sub-Antarctic islands. Values are available at monthly intervals. These data were sourced from the British Antarctic Survey (<https://legacy.bas.ac.uk/met/gjma/sam.html>);
- **Pacific South America Pattern Index:** A measure of the strength of a large, quasi-stationary wave train extending from Australia to Argentina. It is defined in this study as the 2nd (PSA1) and 3rd (PSA2) principal component time-series of the geopotential height anomalies at 500hPa from the JRA55 reanalysis in the southern hemisphere. We note that the interpretation of the PSA is complicated by its definition using statistical properties (ie. the principle components) and as opposed to a definition based on dynamics. In reality, the PSA may be composed of a superposition of travelling and stationary disturbances that can interact with each other,⁴³ as the PSA1 and PSA2 modes, while being in approximate phase quadrature, have a relatively low coherence in the relevant frequency bands.²⁶ In this study, we use the generally accepted definition by convention, but the reader should bear in mind that the exact nature of the PSA is still a matter of some debate.

Climate Model

we use a 2500 year long run variant of the GFDL Climate Model 2.1 (CM2.1),⁶² used in older versions of the Australian Community Climate and Earth System Simulator (ACCESS). The model uses the same atmospheric, land, and sea ice components as CM2.1 (AM2, LM2 and SIS respectively) but uses the Modular Ocean Model (MOM4p1). The ocean model grid is the tripolar ACCESS-o grid⁶³ with a nominal grid spacing of 1° but with a finer latitudinal grid spacing in the tropics and the southern hemisphere high latitudes. There are 50 vertical levels, with 10m grid spacing in the upper ocean, increasing to a maximum of 300 m. Subgrid processes for the ocean model are adopted from CM2.1, including neutral physics (Redi diffusivity and Gent-McWilliams skew diffusion), Brian-Lewis vertical mixing profile, Laplacian friction scheme and a K-profile parametrisation for the mixed layer calculation. The atmospheric model (AM2) has a grid spacing of 2° in latitude and 2.5° longitude, and 24 hybrid (sigma-pressure or terrain following pressure) vertical levels. Concentrations of atmospheric aerosols and radiative gases, and land cover are based on 1990 conditions. The model’s ocean temperature and salinity fields are restored to World Ocean Atlas 2013 (WOA13) climatology at depths below 2000m, with

601 a restoring time-scale of 1 year which improves the model’s representation of the
 602 upper ocean stratification, at the expense of suppressing variability with multi-
 603 decadal time-scales, which is not the focus of this work. The model achieves
 604 approximate statistical equilibrium after around 1500 years, with minimal drift
 605 in either temperature or salinity.

606
 607 Although CM2.1 is an older climate model, used within the Coupled Model
 608 Inter-comparison Project phase 3 (CMIP3), we have opted to use it in this
 609 project due to its low numerical cost and relatively good performance in repli-
 610 cating the broad-scale variability over the Australian region.

611 Methods

612 Archetype Analysis

613 Mathematical Formulation

614 The analysis undertaken in this study employs Archetypal Analysis – an ad-
 615 vanced data mining methodology that has been applied in fields ranging from
 616 marketing to astronomy. However, AA has only recently been applied to geo-
 617 physics problems. Here we give a brief description of the AA problem and its
 618 implementation.

619
 620 AA falls into a broad class of mathematical methods known as *matrix fac-*
 621 *torisation*. The goal of such methods is to represent a complex, high dimensional
 622 dataset as the product of several, simpler and lower dimensional datasets. In
 623 AA, for a given spatiotemporal dataset, $x(\mathbf{r}, t)$ represented as a *data matrix*
 624 $\mathbf{X} \in \mathbb{R}^{M \times T}$, where T is the number of time observations and M is the number
 625 of variables considered (i.e., number of grid-points in the SST dataset), we seek
 626 to find $P \ll M$ ‘archetypal’ states, z , that best represent the data:

$$627 \quad x_{r,t} \approx \tilde{x}_{r,t} = \sum_i^P z_{r,i} s_{i,t} \quad i \in [1, P] \quad (2)$$

628 where the subscript t refers to the time index, and the subscript m to the spatial
 629 index. $s_{i,t}$ is the affiliation probability of the i th archetype, which is subject the
 630 constraints:

$$631 \quad s_{i,t} \in [0, 1] \quad \text{and} \quad \sum_i^P s_{i,t} = 1. \quad (3)$$

632 The first of these constraints indicates that s can only take values between
 633 0 and 1, the second indicates that, at any given time, the sum of the affiliations
 634 across all archetypes is equal to one. Matrices with this property are known as a
 635 *left stochastic matrices*. Mathematically, we say that \tilde{x} is a *convex combination*
 636 of the archetypal patterns and corresponds only to an approximation of $x(\mathbf{r}, t)$.
 637 In AA, the archetypes themselves are written as required to reassemble the data.

638 To enforce this, the archetypal patterns are written as a mixture of the data
 639 themselves:

$$640 \quad z_{r,i} = \sum_t^T x_{r,t} c_{t,i} \quad i \in [1, P] \quad (4)$$

641 where $c_{t,j}$ are the mixture weights for archetype j , that have the constraints:

$$642 \quad c_{t,j} \in [0, 1] \quad \text{and} \quad \sum_t^T c_{t,j} = 1. \quad (5)$$

643 Matrices with this property are known as *left stochastic matrices*. Like with
 644 the affiliation probability, the c weights are constrained to take values between
 645 0 and 1. The weights associated with the i th archetype, $c_{t,i}$ sum to 1 over all
 646 time steps. In the case that the number of archetypes is equal to the number
 647 of time-steps in the dataset $x(\mathbf{r}, t) = \tilde{x}$, a trivial solution where each archetype
 648 corresponds to a the field at a single time-step.

649
 650 Combining Eqns. 2 and 4 gives:
 651

$$652 \quad x_{r,t} \approx \tilde{x}_{r,t} = \sum_i^P \sum_j^T x_{r,j} c_{j,i} s_{i,t} = \mathbf{XCS} \quad (6)$$

653 where we write the double summation as a matrix product between the original
 654 data matrix $\mathbf{X} \in \mathbb{R}^{M \times T}$, the *C-matrix* $\mathbf{C} \in \mathbb{R}_{\geq 0}^{T \times P}$, and the *affiliation matrix*
 655 $\mathbf{S} \in \mathbb{R}_{\geq 0}^{P \times T}$. The archetypal spatial patterns, such as those shown in Fig. 1, are
 656 given by:

$$\mathbf{Z} = \mathbf{XC} \in \mathbb{R}^{M \times P} \quad (7)$$

657 The problem is now: for a given data matrix \mathbf{X} , can we find the ‘best’ \mathbf{C}
 658 and \mathbf{S} matrices, which is accomplished through the minimization problem:

$$659 \quad \{\mathbf{S}, \mathbf{C}\} = \arg \min_{\mathbf{S}, \mathbf{C}} \|\mathbf{X} - \mathbf{XCS}\|_F \quad (8)$$

660 where $\|\cdot\|_F$ is the *Froebenius norm*, defined as the square root of the sum of
 661 the squared absolute value of all matrix elements.

662 While the manipulations written above may seem esoteric, the AA decom-
 663 position has a relatively straightforward interpretation. Eqn. 7 states that the
 664 spatial archetypal patterns are simply as an average of the original data weighted
 665 by the elements of the *C-matrix* (more specifically, a convex combination) of the
 666 original data, while Eqn. 6 shows that the original data can be approximated by
 667 an average of the archetypal patterns weighted by the elements of the *S-matrix*.
 668
 669

Geometrical Interpretation of Archetypes

In the main text, we have referred to the archetypal patterns as *extreme modes*. However, from the mathematical formulation of the problem above, it may not be readily apparent how we arrive at the statement.

Formally, the problem above can be shown to be equivalent to finding a discrete approximation to the *convex hull* of the dataset.^{22,24,64} The convex hull is defined as the smallest convex ‘envelope’ of a dataset, and can be considered to be the boundary of a (potentially high dimensional) dataset. Since the convex hull of a dataset and the convex hull of its extreme points are identical, approximating the convex hull is equivalent to finding the extreme points (or corners) of the data underlying distribution.

As first shown by Cutler & Breiman,²² the archetypal patterns are (approximately) located on the convex hull and are, therefore, approximations to the (high dimensional) extremes of the data. This astonishing result occurs due to the constraints imposed on the S and C matrices in Eqns. 3 and 5: that these matrices are *non-negative* and *stochastic*.⁶⁴

Numerical Implementation

The minimization problem posed in Eqn. 8 has no analytic solution for all but the simplest datasets and must be solved numerically in real-world applications. However, AA falls into a class of problems (non-negative matrix factorisation) that are known to be *NP-Hard*,⁶⁵ which implies that, in general, only approximations to the ‘true’ solution can be obtained.

An increasing number of open-source AA algorithms are freely available and have been implemented for most major computing language in use today. Throughout this work, we rely on the MatLab implementation, *PCHA*, by Mørup & Hansen,⁶⁶ whereby the optimization problem sketched in Eqn. 8 is efficiently computed through a simple, but robust, projected gradient method.

In order to deal with the high dimensionality of geophysical fields, we apply a modification of AA, coined Reduced Space AA (RSAA), introduced for the first time in⁶⁷ to reduce the spatial dimension of the problem and its computational burden. RSAA takes advantage of the invariance of the *Froebenius norm* in Eqn. 8 under unitary transformation:

$$\{\mathbf{S}, \mathbf{C}\} = \arg \min_{\mathbf{S}, \mathbf{C}} \|\mathbf{U}\mathbf{\Sigma}\mathbf{V} - \mathbf{U}\mathbf{\Sigma}\mathbf{V}\mathbf{C}\mathbf{S}\|_F = \arg \min_{\mathbf{S}, \mathbf{C}} \|\mathbf{\Sigma}\mathbf{V} - \mathbf{\Sigma}\mathbf{V}\mathbf{C}\mathbf{S}\|_F, \quad (9)$$

where the spatial patterns or Orthogonal Empirical Functions (EOFs) characterised by the unitary matrix $\mathbf{U} \in \mathbb{R}^{M \times R}$ can be factored out of the equation when a compact Singular Value Decomposition (SVD) is applied to the data matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}$. The optimization is only performed on the scaled Principal

Components (PCs), expressed as $\Sigma \mathbf{V} \in \mathbb{R}^{R \times T}$ in Eqn. 9, with $\Sigma \in \mathbb{R}_{>0}^{R \times R}$ the eigenvalue matrix where $R \leq \min\{M, T\}$ is the rank of the data matrix \mathbf{X} . To recover the archetypal patterns \mathbf{Z} , the solutions $\Sigma \mathbf{V} \mathbf{C}$ need to be left-multiplied by \mathbf{U} such that $\mathbf{Z} = \mathbf{U} \Sigma \mathbf{V} \mathbf{C}$. Typically, RSAA uses a low-rank approximation of \mathbf{X} , $\tilde{\mathbf{X}} = \mathbf{U} \Sigma' \mathbf{V}$, with $\Sigma' \in \mathbb{R}_{>0}^{R' \times R'}$ and $R' \ll R$.

When applied to detrended OISST daily pentad (5 day averages) anomalies in the Australasian region (60 – 0° S, 90 – 240° E), the dimension reduction step allows a $M=130349/R'=840 \approx 155$ -fold reduction in the number of variables, in our case grid points, the number of observations $T = 2849$ being left unchanged. The reduced rank $R' = 840$ is the number of retained PCs in the truncated SVD factorisation in Eqn. 9. R' corresponds to 95% of the total variance of \mathbf{X} . A similar approach and level of variance truncation are applied to the climate model data set.

Although the projection gradient algorithm used to solve Eqn. 9 can be shown to converge to a solution for a suitable initialisation, there are however no guarantees that this solution is optimal given the *NP-Hard* character of the problem. An iterative procedure is required and achieved by resorting to multiple initialisations. Here, we combine one clustering and random based initialisation strategies, whereby the data driven ‘**FurthestSum**’ procedure advocated by⁶⁶ is compared to 999 random initialisations prescribed by⁶⁸ based on ‘**coreset**’ construction for AA. The optimal solution across 1000 trials is kept as the final result. For each individual trial, the projection gradient algorithm PCHA is considered to have converged when the relative sum of square error stopping criterion reaches 10^{-8} .

Forming Composite Fields Using Affiliation (*S*-Matrix) and archetype weights (*C*-Matrix)

Once an affiliation time-series has been computed, it can be applied as a weight to form clusters or composites in order to identify the climatic states associated with any particular archetype. For example, in this study, we have extracted atmospheric and sub-surface ocean patterns associated with the extreme states identified in the SST by the AA in order to demonstrate the remote teleconnections that may influence the regional extremes. This utility arises from the interpretation of the affiliation time-series as the probability at time t that the data is associated with the i th archetypal pattern z_i :²⁴

$$s_{i,t} = Pr(z_i | x_t). \quad (10)$$

As such, the affiliation can be used to associate any dataset with the i th archetypal state.

To derive the spatial fields of a supplemental dataset (for example, atmospheric geopotential height at 500hPa) $y = y(\text{space}, \text{time}) = y_{m,t}$ associated

with the i th archetype, we simply compute the temporal average of y weighted by s_i :

$$\bar{y}_{m,i} = \frac{\sum_t^T y_{m,t} s_{i,t}}{\sum_t^T s_{i,t}} = \frac{\mathbf{Y} \mathbf{S}^T}{\sum_t^T s_{i,t}}. \quad (11)$$

In this study, AA is applied to an SST dataset with a temporal period from 1st January 1982 to 31 December 2020, at 5 day output frequency. As such, the affiliation time-series spans an identical time-period with identify output frequency. However, the atmospheric reanalysis spans a longer time-period (1958-present) with daily output frequency, while the Argo derived sub-surface temperature dataset spans a shorter time-period (2004-present) at monthly output frequency. In order to apply Eqn. 11 to these datasets, the JRA55 fields are first down-sampled to a 5 day output frequency by first low-pass filtering the data using a standard box-car filter with a cut-off period of 1/5 days, then sub-sampled to five day output frequency, truncated to the same temporal period as the SST data, which allows direct application of 11. In the case of the sub-surface temperature dataset, the affiliation time series is down-sampled and truncated to match that of the Argo product.

A similar procedure using the archetype weights (C -matrix) may also be enacted. However, since $\sum_t^T c_{i,t} = 1$, the weighted average is simply:

$$\bar{y}_{m,i} = \sum_t^T y_{m,t} c_{i,t} = \mathbf{Y} \mathbf{C} \quad (12)$$

Determination of the Statistical Significance of the Composite Fields

At present, the question of determining the significance level of archetypal patterns is unresolved. As such, in this paper, the statistical significance of the composite fields formed from the weighted averages is assessed by a simple, brute force Monte-Carlo technique. To begin, we generate synthetic S and C matrices by populating each element with a random number drawn from a uniform distribution between 0 and 1. The rows or columns are then appropriately normalised to apply that the constraints in Eqns. 3 or 5. We then form composite average fields using these synthetic matrices following Eqn. 11 (when testing the significance of the composites formed with the affiliation time-series) or Eqn. 12 (for testing the significance of composites formed using the C -matrix). The procedure is then repeated 1000 times and the 95% and 5% percentile computed. The spatial patterns obtained from the AA are then tested against these synthetic composites: a single pixel is considered ‘significant’ (with a 95% confidence level) if its value is less than the 5th percentile or greater than the 95th percentile.

Archetypal patterns and the associated composite averages are almost everywhere significant, as might be expected from a methodology that specifically extracts patterns associated with extreme states. As such, we have not included

regions of statistical significant on figures in the main text, to avoid visual clutter. These figures can be found in the supplementary material.

Identification of Marine Heat Wave/Marine Cold Spell Events

In this study, the definition of Marine Heat Waves (and Marine Cold Spells as discussed in the supplementary material) follows that of Hobday et al. 2018³² with a slight modification: a MHW is detected if the SST at a particular location exceeds the 90th percentile for a duration *of at least 10 days* (as opposed to the standard definition of 5 days). The temperature may briefly drop below the required thresholds for a period not exceeding 2 days and still be declared an extreme event. We have imposed a slightly more strict criteria on the persistence of events in order to eliminate multiple short duration, moderate intensity events that occur in near-coast regions that appear to be more a response to high-frequency “noise” in the SST that modulated low-frequency variability.

Data availability

We have made use of publicly available data only; no new data were generated as a result of this study. URL and DOIs for the relevant dataset can be found in the *Data* section above.

Computer code availability

Source code for the climate model used in the study can be obtained from the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory (<https://data1.gfdl.noaa.gov/CM2.X/>). Source code used for the generation of archetypes, written in the Matlab language, is freely available from the website of its author Morton Mørup (http://www.mortenmorup.dk/MMhomepageUpdated_files/).

Author Contributions

CC and DPM conceived the study. DPM was responsible for performing the archetype analysis. CC and DPM performed the analysis of the results, with input from all co-authors. CC was responsible for running the climate model experiments. CC lead the drafting of the manuscript, with inputs from all co-authors.

References

- ¹ Neil Holbrook, Alex Gupta, Eric Oliver, Alistair Hobday, Jessica Benthuyssen, Hillary Scannell, Dan Smale, and Thomas Wernberg. Keeping pace with marine heatwaves. *Nature Reviews Earth & Environment*, 1, 07 2020.

- ² Eric C.J. Oliver, Jessica A. Benthuisen, Sofia Darmaraki, Markus G. Donat, Alistair J. Hobday, Neil J. Holbrook, Robert W. Schlegel, and Alex Sen Gupta. Marine heatwaves. *Annual Review of Marine Science*, 13(1):313–342, 2021. PMID: 32976730.
- ³ Claire M. Spillman, Grant A. Smith, Alistair J. Hobday, and Jason R. Hartog. Onset and decline rates of marine heatwaves: Global trends, seasonal forecasts and marine management. *Frontiers in Climate*, 3, 2021.
- ⁴ Jessica A. Benthuisen, Grant A. Smith, Claire M. Spillman, and Craig R. Steinberg. Subseasonal prediction of the 2020 Great Barrier Reef and Coral Sea marine heatwave. *Environmental Research Letters*, 16(12):124050, December 2021.
- ⁵ Maxime Marin, Ming Feng, Helen E. Phillips, and Nathaniel L. Bindoff. A global, multiproduct analysis of coastal marine heatwaves: Distribution, characteristics, and long-term trends. *Journal of Geophysical Research: Oceans*, 126(2):e2020JC016708, 2021.
- ⁶ Ding Ning, Varvara Vetrova, Karin Bryan, and Sébastien Delaux. Deep learning for spatiotemporal anomaly forecasting: A case study of marine heatwaves. In *ICML 2021 Workshop on Tackling Climate Change with Machine Learning*, 2021.
- ⁷ Erik Behrens, Graham Rickard, Suzanne Rosier, Jonny Williams, Olaf Morgenstern, and Dáithí Stone. Projections of future marine heatwaves for the oceans around new zealand using new zealand’s earth system model. *Frontiers in Climate*, 4, 2022.
- ⁸ John M. Wallace and David S. Gutzler. Teleconnections in the geopotential height field during the northern hemisphere winter. *Monthly Weather Review*, 109(4):784 – 812, 1981.
- ⁹ Neil Holbrook, Hillary Scannell, Alexander Gupta, Jessica Benthuisen, Ming Feng, Eric Oliver, Lisa Alexander, Michael Burrows, Markus Donat, Alistair Hobday, Pippa Moore, Sarah Perkins-Kirkpatrick, Dan Smale, Sandra Straub, and Thomas Wernberg. A global assessment of marine heatwaves and their drivers. *Nature Communications*, 10, 06 2019.
- ¹⁰ R. Rodrigues, Andréa S. Taschetto, A. Sen Gupta, and G. Foltz. Common cause for severe droughts in south america and marine heatwaves in the south atlantic. *Nature Geoscience*, pages 1–7, 2019.
- ¹¹ Alex Sen Gupta, Mads Thomsen, Jessica A. Benthuisen, Alistair J. Hobday, Eric Oliver, Lisa V. Alexander, Michael T. Burrows, Markus G. Donat, Ming Feng, Neil J. Holbrook, Sarah Perkins-Kirkpatrick, Pippa J. Moore, Regina R. Rodrigues, Hillary A. Scannell, Andréa S. Taschetto, Caroline C. Ummenhofer, Thomas Wernberg, and Dan A. Smale. Drivers and impacts of the most extreme marine heatwaves events. *Scientific Reports*, 10(19359), 2020.

- ⁸⁴⁶ ¹² Edward N. Lorenz. The predictability of a flow which possesses many scales
⁸⁴⁷ of motion. *Tellus*, 21(3):289–307, 1969.
- ⁸⁴⁸ ¹³ J. J. Tribbia and D. P. Baumhefner. Scale interactions and atmospheric
⁸⁴⁹ predictability: An updated perspective. *Monthly Weather Review*, 132(3):703
⁸⁵⁰ – 713, 2004.
- ⁸⁵¹ ¹⁴ C. A. S. Coelho, C. A. T. Ferro, D. B. Stephenson, and D. J. Steinskog.
⁸⁵² Methods for exploring spatial and temporal variability of extreme events in
⁸⁵³ climate data. *Journal of Climate*, 21(10):2072 – 2092, 2008.
- ⁸⁵⁴ ¹⁵ Alan F. Pearce and Ming Feng. The rise and fall of the “marine heat wave”
⁸⁵⁵ off western australia during the summer of 2010/2011. *Journal of Marine*
⁸⁵⁶ *Systems*, 111-112:139–156, 2013.
- ⁸⁵⁷ ¹⁶ Jessica Benthuisen, Ming Feng, and Liejun Zhong. Spatial patterns of warm-
⁸⁵⁸ ing off western australia during the 2011 ningaloo niño: Quantifying impacts
⁸⁵⁹ of remote and local forcing. *Continental Shelf Research*, 91:232–246, 2014.
- ⁸⁶⁰ ¹⁷ Nicholas A. Bond, Meghan F. Cronin, Howard Freeland, and Nathan Mantua.
⁸⁶¹ Causes and impacts of the 2014 warm anomaly in the ne pacific. *Geophysical*
⁸⁶² *Research Letters*, 42(9):3414–3420, 2015.
- ⁸⁶³ ¹⁸ Emanuele Di Lorenzo and Nathan J. Mantua. Multi-year persistence of the
⁸⁶⁴ 2014/15 north pacific marine heatwave. *Nature Climate Change*, 6:1042–1047,
⁸⁶⁵ 2016.
- ⁸⁶⁶ ¹⁹ Eric Oliver, Jessica Benthuisen, Nathaniel Bindoff, Alistair Hobday, Neil
⁸⁶⁷ Holbrook, Craig Mundy, and Sarah Perkins-Kirkpatrick. The unprecedented
⁸⁶⁸ 2015/16 tasman sea marine heatwave. *Nature Communications*, 8:16101, 07
⁸⁶⁹ 2017.
- ⁸⁷⁰ ²⁰ Hakase Hayashida, Richard Matear, Peter Strutton, and Xuebin Zhang. In-
⁸⁷¹ sights into projected changes in marine heatwaves from a high-resolution
⁸⁷² ocean circulation model. *Nature Communications*, 11:4352, 08 2020.
- ⁸⁷³ ²¹ Charuni Pathmeswaran, Alex Sen Gupta, Sarah E. Perkins-Kirkpatrick, and
⁸⁷⁴ Melissa Anne Hart. Exploring potential links between co-occurring coastal
⁸⁷⁵ terrestrial and marine heatwaves in australia. *Frontiers in Climate*, 4, 2022.
- ⁸⁷⁶ ²² Adele Cutler and Leo Breiman. Archetypal analysis. *Technometrics*,
⁸⁷⁷ 36(4):338–347, 1994.
- ⁸⁷⁸ ²³ Christian Bauckhage and Christian Thureau. Making archetypal analysis prac-
⁸⁷⁹ tical. In Joachim Denzler, Gunther Notni, and Herbert Süße, editors, *Pattern*
⁸⁸⁰ *Recognition*, pages 272–281, 2009.
- ⁸⁸¹ ²⁴ A. Hannachi and N. Trendafilov. Archetypal analysis: Mining weather and
⁸⁸² climate extremes. *Journal of Climate*, 30(17):6927 – 6944, 2017.

- 883 ²⁵ Scott Steinschneider and Upmanu Lall. Daily precipitation and tropical mois-
884 ture exports across the eastern united states: An application of archetypal
885 analysis to identify spatiotemporal structure. *Journal of Climate*, 28(21):8585
886 – 8602, 2015.
- 887 ²⁶ James S. Risbey, Didier P. Monselesan, Amanda S. Black, Thomas S. Moore,
888 Doug Richardson, Dougal T. Squire, and Carly R. Tozer. The identification
889 of long-lived southern hemisphere flow events using archetypes and principal
890 components. *Monthly Weather Review*, 149(6):1987 – 2010, 2021.
- 891 ²⁷ Houk Paek, Jin-Yi Yu, and Chengcheng Qian. Why were the 2015/2016
892 and 1997/1998 extreme el niños different? *Geophysical Research Letters*,
893 44(4):1848–1856, 2017.
- 894 ²⁸ Carmen Boening, Josh K. Willis, Felix W. Landerer, R. Steven Nerem, and
895 John Fasullo. The 2011 la niña: So strong, the oceans fell. *Geophysical*
896 *Research Letters*, 39(19), 2012.
- 897 ²⁹ Alistair Hobday and Gretta Pecl. Identification of global marine hotspots:
898 Sentinels for change and vanguards for adaptation action. *Reviews in Fish*
899 *Biology and Fisheries*, 24, 06 2014.
- 900 ³⁰ Nick Caputi, Mervi Kangas, Ainslie Denham, Ming Feng, Alan Pearce, Yasha
901 Hetzel, and Arani Chandrapavan. Management adaptation of invertebrate
902 fisheries to an extreme marine heat wave event at a global warming hot spot.
903 *Ecology and Evolution*, 6(11):3583–3593, 2016.
- 904 ³¹ Eric C. J. Oliver, Michael T. Burrows, Markus G. Donat, Alex Sen Gupta,
905 Lisa V. Alexander, Sarah E. Perkins-Kirkpatrick, Jessica A. Benthuisen, Alistair
906 J. Hobday, Neil J. Holbrook, Pippa J. Moore, Mads S. Thomsen, Thomas
907 Wernberg, and Dan A. Smale. Projected marine heatwaves in the 21st cen-
908 tury and the potential for ecological impact. *Frontiers in Marine Science*,
909 6:734, 2019.
- 910 ³² Alistair J. Hobday, Eric C.J. Oliver, Alex Sen Gupta, Jessica A. Benthuisen,
911 Michael T. Burrows, Markus G. Donat, Neil J. Holbrook, Pippa J. Moore,
912 Mads S. Thomsen, Thomas Wernberg, and Dan A. Smale. Categorizing and
913 naming marine heatwaves. *Oceanography*, 31, June 2018.
- 914 ³³ Ming Feng, Nick Caputi, Arani Chandrapavan, Miaoju Chen, Anthony Hart,
915 and Mervi Kangas. Multi-year marine cold-spells off the west coast of australia
916 and effects on fisheries. *Journal of Marine Systems*, 214:103473, 2021.
- 917 ³⁴ Karumuri Ashok, Swadhin K. Behera, Suryachandra A. Rao, Hengyi Weng,
918 and Toshio Yamagata. El niño modoki and its possible teleconnection. *Journal*
919 *of Geophysical Research: Oceans*, 112(C11), 2007.
- 920 ³⁵ Shamal Marathe and Ashok Karumuri. 4 - the el niño modoki. In Swad-
921 hin Kumar Behera, editor, *Tropical and Extratropical Air-Sea Interactions*,
922 pages 93–114. Elsevier, 2021.

- 923 ³⁶ Ming Feng, Michael McPhaden, Shang-Ping Xie, and Jan Hafner. La niña
924 forces unprecedented leeuwin current warming in 2011. *Scientific reports*,
925 3:1277, 02 2013.
- 926 ³⁷ Takahito Kataoka, Tomoki Tozuka, Swadhin Behera, and Toshio Yamagata.
927 On the ningaloo niño/niña. *Climate Dynamics*, 43, 10 2013.
- 928 ³⁸ Clemens Spensberger, Michael J. Reeder, Thomas Spengler, and Matthew
929 Patterson. The connection between the southern annular mode and a feature-
930 based perspective on southern hemisphere midlatitude winter variability.
931 *Journal of Climate*, 33(1), 2020.
- 932 ³⁹ Klaus Wolter and Michael S. Timlin. Measuring the strength of enso events:
933 How does 1997/98 rank? *Weather*, 53(9):315–324, 1998.
- 934 ⁴⁰ Gareth J. Marshall. Trends in the southern annular mode from observations
935 and reanalyses. *Journal of Climate*, 16(24), 2003.
- 936 ⁴¹ M. Salinger, James Renwick, Erik Behrens, Brett Mullan, Howard Diamond,
937 Pascal Sirguey, Robert Smith, Mike Trought, Lisa Alexander, Nicolas Cullen,
938 Brian Fitzharris, Chris Hepburn, Amber Parker, and Philip Sutton. The un-
939 precedented coupled ocean-atmosphere summer heatwave in the new zealand
940 region 2017/18: drivers, mechanisms and impacts. *Environmental Research*
941 *Letters*, 14, 01 2019.
- 942 ⁴² M. James Salinger, Howard J. Diamond, Erik Behrens, Denise Fernandez,
943 B. Blair Fitzharris, Nicholas Herold, Paul Johnstone, Huub Kerckhoffs,
944 A. Brett Mullan, Amber K. Parker, James Renwick, Claire Scofield, Allan
945 Siano, Robert O. Smith, Paul M. South, Phil J. Sutton, Edmar Teixeira,
946 Mads S. Thomsen, and Michael C. T. Trought. Unparalleled coupled ocean-
947 atmosphere summer heatwaves in the new zealand region: drivers, mecha-
948 nisms and impacts. *Climatic Change*, 162(2):485–506, Sep 2020.
- 949 ⁴³ Terence J. O’Kane, Didier P. Monselesan, and James S. Risbey. A multi-
950 scale reexamination of the pacific–south american pattern. *Monthly Weather*
951 *Review*, 145(1):379 – 402, 2017.
- 952 ⁴⁴ Carmen Boening, Tong Lee, and Victor Zlotnicki. A record-high ocean bottom
953 pressure in the south pacific observed by grace. *Geophysical Research Letters*,
954 38(4), 2011.
- 955 ⁴⁵ Tim Woollings, David Barriopedro, John Methven, Seok-Woo Son, Olivia
956 Martius, Ben Harvey, Jana Sillmann, Anthony Lupo, and Sonia Seneviratne.
957 Blocking and its response to climate change. *Current Climate Change Reports*,
958 4:1–14, 09 2018.
- 959 ⁴⁶ Kenneth R.N. Anthony. Coral reefs under climate change and ocean acidi-
960 fication: Challenges and opportunities for management and policy. *Annual*
961 *Review of Environment and Resources*, 41(1):59–81, 2016.

- 962 ⁴⁷ Thomas Decarlo and Hugo Harrison. An enigmatic decoupling between heat
963 stress and coral bleaching on the great barrier reef. *PeerJ*, 7:e7473, 08 2019.
- 964 ⁴⁸ C. E. Langlais, M. Herzfeld, E. Klein, N. Cantin, J. Benthuyssen, and C. Stein-
965 berg. Oceanographic drivers of bleaching in the gbr: from observations to
966 prediction volume 2: 3d bleaching in the gbr: Development and analysis of
967 a 3d climatology and 3d heat accumulation bleaching products using ereefs.
968 Technical report, National Environmental Science Program., Cairns, 2021.
969 Report to the National Environmental Science Program.
- 970 ⁴⁹ Hamish McGowan and Alison Theobald. Enso weather and coral bleaching on
971 the great barrier reef, australia. *Geophysical Research Letters*, 44(20):10,601–
972 10,607, 2017.
- 973 ⁵⁰ Wenhui Zhao, Yi Huang, Steven Siems, and Michael Manton. The role of
974 clouds in coral bleaching events over the great barrier reef. *Geophysical Re-*
975 *search Letters*, 48(14):e2021GL093936, 2021.
- 976 ⁵¹ Eric Oliver. Mean warming not variability drives marine heatwave trends.
977 *Climate Dynamics*, 53, 08 2019.
- 978 ⁵² Gabriela S. Pilo, Neil J. Holbrook, Andrew E. Kiss, and Andrew McC. Hogg.
979 Sensitivity of marine heatwave metrics to ocean model resolution. *Geophysical*
980 *Research Letters*, 46(24):14604–14612, 2019.
- 981 ⁵³ John R. Lanzante, Mary Jo Nath, Carolyn E. Whitlock, Keith W. Dixon,
982 and Dennis Adams-Smith. Evaluation and improvement of tail behaviour
983 in the cumulative distribution function transform downscaling method. *Inter-*
984 *national Journal of Climatology*, 39(4):2449–2460, 2019.
- 985 ⁵⁴ Christopher C. Chapman, Bernadette M. Sloyan, Terence J. O’Kane, and
986 Matthew A. Chamberlain. Interannual subtropical indian ocean variability
987 due to long baroclinic planetary waves. *Journal of Climate*, 33(16):6765 –
988 6791, 2020.
- 989 ⁵⁵ Boyin Huang, Chunying Liu, Viva Banzon, Eric Freeman, Garrett Graham,
990 Bill Hankins, Tom Smith, and Huai-Min Zhang. Improvements of the daily
991 optimum interpolation sea surface temperature (doisst) version 2.1. *Journal*
992 *of Climate*, 34(8):2923 – 2939, 2021.
- 993 ⁵⁶ Richard W. Reynolds, Thomas M. Smith, Chunying Liu, Dudley B. Chelton,
994 Kenneth S. Casey, and Michael G. Schlax. Daily high-resolution-blended
995 analyses for sea surface temperature. *Journal of Climate*, 20(22):5473 – 5496,
996 2007.
- 997 ⁵⁷ Japan Meteorological Agency. Jra-55: Japanese 55-year reanalysis, daily 3-
998 hourly and 6-hourly data, 2013. Research Data Archive at the National Center
999 for Atmospheric Research, Computational and Information Systems Labora-
1000 tory, accessed 10 March 2021.

- 1001 ⁵⁸ Shinya KOBAYASHI, Yukinari OTA, Yayoi HARADA, Ayataka EBITA,
1002 Masami MORIYA, Hirokatsu ONODA, Kazutoshi ONOGI, Hirotaka KAMA-
1003 HORI, Chiaki KOBAYASHI, Hirokazu ENDO, Kengo MIYAOKA, and Kiy-
1004 otoshi TAKAHASHI. The jra-55 reanalysis: General specifications and ba-
1005 sic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*,
1006 93(1):5–48, 2015.
- 1007 ⁵⁹ Dean Roemmich and John Gilson. The 2004–2008 mean and annual cycle
1008 of temperature, salinity, and steric height in the global ocean from the argo
1009 program. *Progress in Oceanography*, 82(2):81–100, 2009.
- 1010 ⁶⁰ Argo. Argo float data and metadata from global data assembly centre (argo
1011 gdac), 2020. SEANOE. Accessed November 2021.
- 1012 ⁶¹ Klaus Wolter and Michael S. Timlin. El niño/southern oscillation behaviour
1013 since 1871 as diagnosed in an extended multivariate enso index (mei.ext).
1014 *International Journal of Climatology*, 31(7):1074–1087, 2011.
- 1015 ⁶² Thomas L. Delworth, Anthony J. Broccoli, Anthony Rosati, Ronald J. Stouf-
1016 fer, V. Balaji, John A. Beesley, William F. Cooke, Keith W. Dixon, John
1017 Dunne, K. A. Dunne, Jeffrey W. Durachta, Kirsten L. Findell, Paul Gi-
1018 noux, Anand Gnanadesikan, C. T. Gordon, Stephen M. Griffies, Rich Gudgel,
1019 Matthew J. Harrison, Isaac M. Held, Richard S. Hemler, Larry W. Horowitz,
1020 Stephen A. Klein, Thomas R. Knutson, Paul J. Kushner, Amy R. Langen-
1021 horst, Hyun-Chul Lee, Shian-Jiann Lin, Jian Lu, Sergey L. Malyshev, P. C. D.
1022 Milly, V. Ramaswamy, Joellen Russell, M. Daniel Schwarzkopf, Elena Shevli-
1023 akova, Joseph J. Sirutis, Michael J. Spelman, William F. Stern, Michael Wint-
1024 on, Andrew T. Wittenberg, Bruce Wyman, Fanrong Zeng, and Rong Zhang.
1025 GFDL’s CM2 Global Coupled Climate Models. Part I: Formulation and Sim-
1026 ulation Characteristics. *Journal of Climate*, 19(5):643–674, 2006.
- 1027 ⁶³ Daohua Bi, Martin Dix, Simon J. Marsland, Siobhan O’Farrell, H. A. Ab-
1028 dul Rashid, Petteri Uotila, Anthony C. Hirst, Eva Kowalczyk, Maciej
1029 Golebiewski, Arnold Sullivan, Hailin Yan, Nicholas Hannah, C. P. Franklin,
1030 Zhian Sun, Peter Vohralik, Ian Watterson, Xiaobing Zhou, Russell Fiedler,
1031 Mark Collier, Yimin Ma, J. M. Noonan, Lauren Stevens, Peter Uhe, Hongyan
1032 Zhu, Stephen M. Griffies, Richard Hill, Chris Harris, and Kamal D. Puri. The
1033 ACCESS coupled model : description , control climate and evaluation. *Aus-
1034 tralian Meteorological and Oceanographic Journal*, 63(1):41–64, 2013.
- 1035 ⁶⁴ Christian Bauckhage and Christian Thureau. Making archetypal analysis prac-
1036 tical. In Joachim Denzler, Gunther Notni, and Herbert Süße, editors, *Pattern
1037 Recognition*, pages 272–281, Berlin, Heidelberg, 2009. Springer Berlin Heidel-
1038 berg.
- 1039 ⁶⁵ Stephen A. Vavasis. On the complexity of nonnegative matrix factorization.
1040 *SIAM Journal on Optimization*, 20(3):1364–1377, 2010.

- 1041 ⁶⁶ Morten Mørup and Lars Kai Hansen. Archetypal analysis for machine learning
1042 and data mining. *Neurocomputing*, 80:54–63, March 2012.
- 1043 ⁶⁷ D. Richardson, A. Black, D. Monselesan, T. Moore, J. Risbey, D. Squire,
1044 and C. Tozer. Identifying periods of forecast model confidence for improved
1045 subseasonal prediction of precipitation. *JHM*, 22(2):371–385, 2021.
- 1046 ⁶⁸ Sebastian Mair and Ulf Brefeld. Coresets for Archetypal Analysis. In H. Wal-
1047 lach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett,
1048 editors, *Advances in Neural Information Processing Systems*, volume 32. Cur-
1049 ran Associates, Inc., 2019.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterialMarineHeatwavesArchetypescompressed.pdf](#)