https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

Open Access Earth System Science Data Discussions

Gap-filled subsurface mooring dataset off Western Australia during 2010-2023

Toan Bui[1], Ming Feng[1*], Christopher Chapman[2]

[1]CSIRO Environment, Indian Ocean Marine Research Centre, Crawley, WA, Australia

[2]CSIRO Environment, Hobart, Tasmania, Australia

Correspondence to: ming.feng@csiro.au

Abstract

Coastal moorings allow scientists to collect long-term datasets valuable in understanding shelf dynamics, detecting climate variability and changes, and evaluating their impacts on marine ecosystems. Continuous time series data from moorings is often disrupted due to mooring losses or instrument failures, which prevents us from obtaining complete and accurate information on the marine environment. Here, we present an updated version of the 14-year subsurface mooring dataset off the southwest coast of Western Australia during 2010-2023 (https://doi.org/10.25919/myac-yx60, Bui and Feng, 2024). This updated dataset offers continuous daily temperature and current data with a 5-meter vertical resolution, collected from six coastal Integrated Marine Observing System (IMOS) moorings at depths between 48 m and 500 m. Self-Organizing Map (SOM) machine learning technique is applied to fill in the data gaps in the previous version. The usage of the in-filled data product is demonstrated by detecting sub-surface marine heatwaves on the Rottnest shelf. The data products can be used to characterise subsurface features of extreme events such as marine heatwaves, and marine cold-spells, influenced by the Leeuwin Current and the wind-driven Capes Current, and to detect long-term change signals along the coast.

1 Introduction

Oceanography moorings use underwater instruments anchored on the sea floor to collect ocean currents, temperature, salinity, and other environmental parameters. Sustained long-term mooring observations serve as invaluable resources for environmental and climate research and play a vital role in calibrating and validating numerical models (Bailey et al., 2019). Typically, mooring services range from every 4-6 months in shelf waters to up to every 18 months in deep oceans (Sloyan et al., 2024).

The southwest Western Australian mooring array is part of the Integrated Marine Observing System (IMOS) program operated by CSIRO since 2009, designed to monitor the influences of the southward-flowing Leeuwin Current (LC) on the

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

Open Access Earth System Science Data Discussions

continental shelf (Chen and Feng, 2021). The anomalous meridional pressure gradient, associated with warm, low-salinity waters from the tropical Pacific Ocean entering the Indian Ocean through the Indonesian Archipelago, is the main driver of the LC (Thompson 1984; Feng and Wijffels, 2002; Godfrey and Ridgway, 1985). The strength of the LC varies seasonally due to variations in the alongshore winds (Smith et al., 1991). During austral summer, strong alongshore northward winds drive northward Capes Current in the mid-inner shelf (Fig. 1). The interannual variability of the LC is often associated with the El *Niño*–Southern Oscillation (ENSO), the current being stronger during La Niña and weaker during El Niño (Feng et al., 2003).

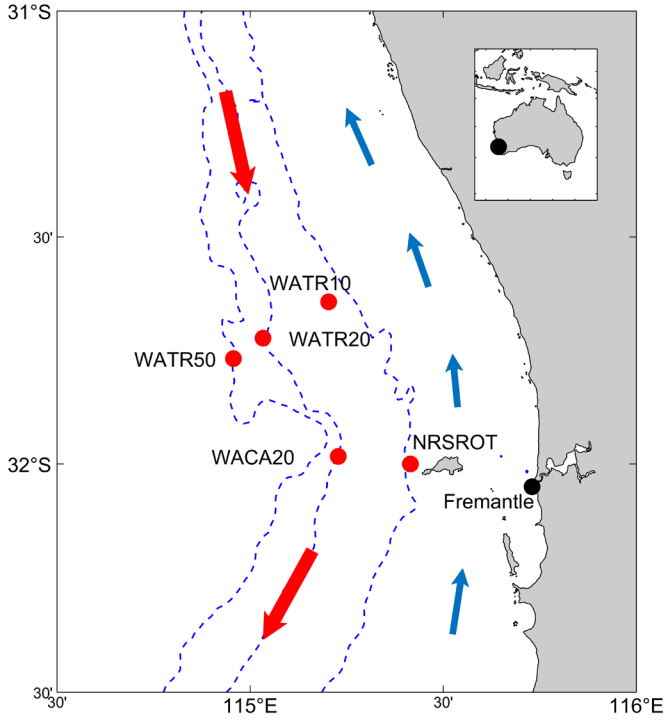

Figure 1. Bathymetry map and mooring locations (red circles) on the Rottnest Shelf. The red arrows denote the Leeuwin Current, while blue arrows show the direction of the wind-driven Capes Current. The three dashed lines present the 50m, 200m and 500m contours. The black circles denote the location of the Fremantle tide gauge station. NRSROT has two separate moorings.

The southwest Western Australian mooring array has helped scientists identify the key role of the LC in the development of marine heatwaves (MHW) off the coast (Benthuysen et al., 2014; Feng et al., 2013). The mooring data has also been employed by Feng et al. (2021) to detect abnormal cooling events off the coast over 2016-2019 (defined as the marine cold spell, MCS) when the thermocline depth was elevated due to the weakening of the LC during the El Niño events. The sustained IMOS mooring array encompasses six coastal moorings on the Rottnest shelf during 2010-2023, ranging from 50m to 500m (Fig. 1 and Table 1). The first version of the gridded data from these moorings was published by Chen and Feng (2021) and an extension was published by Bui et al. (2023). Mooring time series are susceptible to missing values due to mooring loss and

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

instrument failure. Strong currents can exert force on the mooring line, causing it to be pushed down in the water column, leaving data gaps near the surface (Sloyan et al., 2023). This paper introduces a new update of the mooring data, filling data gaps with a statistical method.

Table 1. Summary of coastal mooring stations.

Station	Latitude; Longitude	Station depth (m)	Temperature				ADCP			
			Instrument	Interval (min)	Mean sensor depths (m)	Data span	Instrument	Interval (min)	Bin numbers x bin size	Data span
NRSROT-Temperature	31.9900°S; 115.3850°E	61	SBE39 SBE37	5-15	27; 33; 43; 55	1/2010-5/2023				
NRSROT-ADCP	32.0000°S; 115.4170°E;	48					Signature 500	15	11x4m	8/2011-5/2023
WACA20	31.9830°S; 115.2280°E	200					Signature 250	15	41x5m	8/2011-5/2023
WATR10	31.6433°S; 115.2033°E	100	SBE39 SBE37	5-15	25; 30; 35; 40; 52; 70; 90	1/2010-5/2023	Nortek Aquadopp 400 kHz;	15	17x5m	8/2011-5/2023
WATR20	31.7233°S; 115.0333°E	200	SBE39 SBE37	15	25; 35; 50; 68; 100; 125; 150; 175	1/2010-5/2023	Signature 250	15	25x8m	8/2011-5/2023
WATR50	31.7683°S; 114.9567°E	500					Signature 55	15	26x20m	8/2011-5/2023

Various techniques have been employed to address gaps in mooring datasets. Sprintall et al. (2009) utilized a damped least square fitting method to fill substantial gaps in mooring current time series data, in estimating the Indonesian Throughflow transport. Wang et al. (2015) adopted a combination of data extrapolation, interpolation, and a least square regression model to fill in missing data recorded in the central equatorial Indian Ocean. More recently, Sloyan et al. (2023) experimented with a machine learning approach, the Self-Organizing Map (SOM), to fill data gaps in the East Australian Current mooring array.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

The choice of method depends on the characteristics of data loss, such as the duration of gaps or the depth range affected, as well as the intended analyses of the data.

SOM is a technique that projects high-dimensional input data onto a two-dimensional output space while preserving the topological structure of the input data (Kohonen, 1982). In SOM, units are organized so that similar units are positioned close to each other, while dissimilar ones are separated in the output data space. This method has found extensive applications in meteorology and oceanography (Liu and Weisberg, 2011), and can perform a range of tasks including clustering, data analysis and visualization, feature extraction, and data interpolation (Lobo, 2009).

Chapman and Charantonis (2017) utilized SOM to reconstruct deep current velocities in the Southern Ocean from surface data. They used densely observed surface velocities, sea surface height, sea surface temperature from satellites, and sparsely observed deep current velocities from Argo floats to train the map, then derived dense velocity fields at a depth of 1000m. Their method took advantage of local correlations in the data space to find the smallest Euclidian distance, weighted by the local correlations, between a vector with missing components in the data space and the SOM units, which increased the accuracy of the filled deep velocities.

This study employs the SOM method to fill in the data gaps in the southwest Western Australia mooring data, following the procedure in Chapman and Charantonis (2017), to generate a gap-free time series dataset. The use of the continuous dataset is demonstrated by examining several extreme temperature events that occurred in the region.

2 Data and methods

2.1 Moored instrument data

2.1.1 Temperature

The *in situ* temperature dataset is collected using Seabird Electronics instruments, including SBE37, SBE39, and SBE39 plus, with sampling intervals varying between 5 and 15 minutes (Table 1). To ensure data quality, the raw dataset underwent rigorous quality assurance (QA) and quality control (QC) procedures (Morello et al., 2014), utilizing the IMOS Mooring toolbox written in the MATLAB scientific programming language. Only data flagged as 1, indicating good quality, are retained for this analysis. The QC data are concatenated, and then linearly interpolated onto a grid of 5 m vertical resolution and averaged daily (Bui et al., 2023). The unfilled data are available in the CSIRO Data Access Portal (https://doi.org/10.25919/9gb1-ne81). For data completion, we use satellite sea surface temperature (SST) sourced from the Regional Australian Multi-Sensor SST Analysis (RAMSSA) version 1.0 (Beggs et al., 2011), to extend the temperature data at each mooring to the sea surface by linear interpolation. When minor gaps occur near the bottom, we use two available data points at the bottom of the vertical temperature profile to extrapolate linearly to the sea bottom. The resulting daily 5m-vertical

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

resolution temperatures at NRSROT, WATR10, and WATR20 moorings, spanning from January 2010 to May 2023, are presented in Fig. S1.

2.1.2 Velocity

The velocity observations on the IMOS mooring array are recorded by various instruments, including RDI Workhorse 300kHz/600kHz, RDI Long Ranger 75 kHz, Nortek Continental 190 kHz, Nortek Aquadopp 400 kHz, and Nortek Signature 55/250/500 kHz. These instruments typically sample at a 15-minute interval and are mounted in an upward-looking configuration above the bottom (Table 1).

The raw velocity data undergo quality control procedures, followed by concatenation and gridding into a daily grid with 5m-vertical resolution, as described by Bui et al. (2023). The velocity dataset comprises observations from five stations: NRSROT, WACA20, WATR10, WATR20, and WATR50. Initially, gaps in the time series are filled using linear interpolation if the temporal gap size is less than 3 days. Subsequently, for each velocity profile, gaps near the surface or bottom are filled using linear extrapolation, akin to the technique applied for temperature data. The meridional and zonal components of the velocity datasets, from August 2011 to May 2023, are presented in Figs. S2 and S3, respectively.

For the 2010-2023 period, the percentage of missing mooring data varies from 2% to 16% for temperature, and 12%-33% for velocity (Table 2). The largest percentage of missing data is at WATR20, situated near the core of the LC.

2.2 SOM method

To produce a gap-filled data product, we follow the method described in Chapman and Charantonis (2017), the Iterative Completion by Self Organising Maps (ITCOMPSOM). As discussed briefly in the introduction, this method “completes” a gappy dataset by first using available data to train a self-organising map (SOM), which effectively clusters the data into a set of N discrete states. These states can be represented as a 2-dimensional map, where neighbouring clusters are more similar to

Table 2. Percentage (%) of missing temperature and velocity for each mooring for the time period of 2010-2023.

	Temperature (%)	Velocity (%)
NRSROT	2	12
WACA20		19
WATR10	7	18
WATR20	16	33
WATR50		21

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

each other than distant clusters. Associated with each cluster is a *reference vector* that approximates the mean of all data assigned to that cluster and a weighted mean of data assigned to neighbouring clusters. After the map is trained, new data can be assigned to existing clusters by comparing the Euclidian distance in data space between that new data vector and the reference vector of each cluster. The cluster with the smallest Euclidian distance is known as the Best Matching Unit. Once a SOM is available, data vectors with missing components are presented sequentially, the Best Matching Unit is found, and the missing data is completed (in-filled) by replacing it with the relevant components of the reference vector of the BMU. For full details, see Chapman and Charantonis (2017).

A schematic of using the SOM method to fill gaps in the mooring dataset is shown in Fig. 2. We utilized the Vesanto et al. (2000) SOM toolbox for Matlab 5 in this study. The temperature or velocity data, along with ancillary data, are organized into data matrices. Ancillary data include day-of-year and daily Fremantle sea level (Fig. 1). Sea level data are obtained from the University of Hawaii Sea Level Center (https://uhslc.soest.hawaii.edu/). Fremantle sea level serves as a proxy for the annual and interannual variations of the Leeuwin Current (Feng et al., 2003). The temperature input matrix comprises 4869 rows (representing the number of time steps) and 77 columns (reflecting the number of different observations at each time step). Similarly, the velocity input matrix consists of 4292 rows and 361 columns. The temperature/velocity input matrix with missing values is indicated by square 1 in Fig. 2. Only fully available profiles in the input matrix are selected as the training data shown by square 2. Consequently, the number of rows in the training data is 3675 for temperature and 1146 for velocity.

The number of units in the SOM must be specified prior to the training process. Following several tests, we have chosen 1000 units for temperature, and 500 for temperature and velocity, respectively. This selection was based on the number of rows in the training data. We used a batch algorithm to train the SOM (Chapman and Charantonis, 2017). The training phase of SOM was done in two steps: the first rough phase, followed by a fine-tuning phase. In the first step, the neighbourhood radius and learning rate were set to high values to gain a general orientation of the map, while in the second step, they were set to smaller values to perform only fine adjustments on the SOM unit's position.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

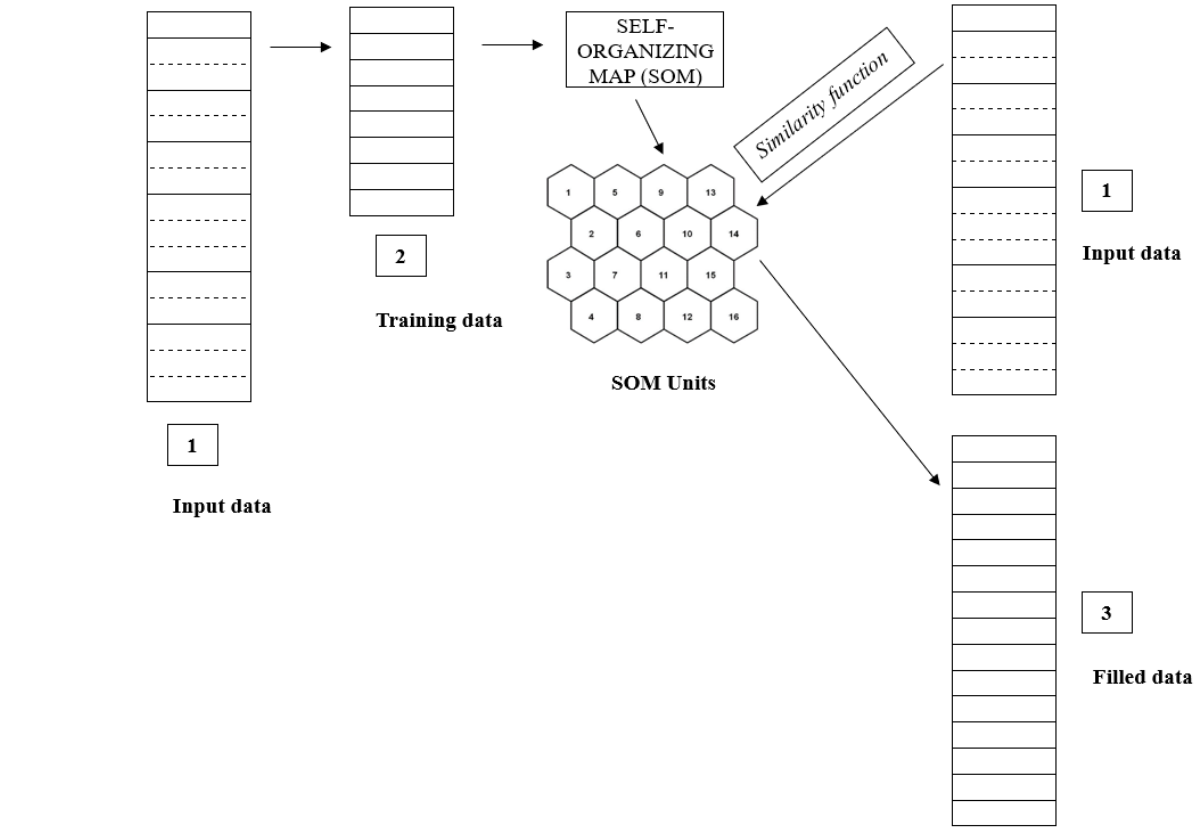

Figure 2. Schematic of the SOM method applied to fill gaps in the mooring temperature and velocity data. Square 1 denotes an input data matrix in which rows are daily time vectors, columns are observational variables. In square 1, solid lines present full available profiles, while dashed lines show the profiles including missing values. In square 2, only full data profiles are selected for training in SOM. In SOM, we pre-define the number of units, for instance 1000 units for temperature, and 500 units for velocity. Each SOM unit contains a reference vector. On the right hand side, each daily input vector in the input data matrix is assigned to each SOM unit using a similarity function defined by Chapman and Charantonis, 2016. Finally, we use the referent vector of each SOM to fill gaps in corresponding daily input vector, shown in square 3.

One important step was the assignment of each input vector to a specific SOM unit, *u,* shown on the right-hand side of Fig. 2. Firstly, we estimated the local correlations in the data space, represented by a *COR* matrix.

$$COR=1+\sqrt{\sum \mathrm{DAT\_cor}^2}, \quad (1)$$

Where *DAT_cor* is a correlation matrix among each normalized input vectors within a SOM unit.

Given with local correlations in the data space, we then calculated the minimum Euclidean distance between a normalized input vector X and the referent vector of the SOM unit, ref^u using a similarity function (Chapman and Charantonis, 2017). The similarity function is defined as:

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

Open Access Earth System Science Data Discussions

$$sim(X, ref^u) = \sum \left[\sqrt{(ref^u - X)^2}\right], \quad (2)$$

After determining the most appropriate SOM unit, the missing values in the input vector were extracted from the corresponding referent vector, providing the in-filled data shown in square 3 (Fig. 2).

2.3 Validation of SOM-based infilling technique

For mooring data, a failed mooring/instrument often results in a block of data being lost until the next deployment. To simulate this effect, we withhold temperature data at one site for 150 days from 1/1/2020 to 30/5/2020, which is roughly the length of one deployment cycle. We utilize temperature data at the other two sites to identify the best matching SOM units, to fill in the withholding data. At NRSROT, the R^2 and the root-mean-square-error (RMSE) between withheld and filled temperature data are 0.70 and 0.61°C, respectively; at WATR10, these values are 0.86 and 0.39°C, and at WATR20, they are 0.91 and 0.58°C, as shown in Fig. 3. For a different period-spanning from 10/1/2012 to 8/6/2012, with 150 days withheld, the comparisons yield RMSE values of 0.41°C at NRSROT, 0.36°C at WATR10, and 0.55°C at WATR20. If we repeat this process and validate the method against data included in training dataset, we obtain RMSE figures similar to those obtained from withheld data, indicating that the SOM method is not overfitting the dataset.

Using the same approach, we examine the accuracy of velocity data gap filling. Specifically, we consider the period from 5/2020 to 8/2020, during which velocity data at WATR50 within the depth range of 70-450m are withheld for 90 days. For the meridional velocity, R^2 and RMSE values between withheld and infill data are 0.63 and 0.12 m s^{-1}, respectively (Fig. 4a). For the zonal component, these values are 0.50 and 0.05 m s^{-1}, respectively (Fig. 4b). To determine if the SOM method overfits the data, we withheld velocity data from a different period spanning from 5/2012 to 8/2012. The resulting RMSE values for the meridional and zonal velocities are 0.13 and 0.06 m s^{-1}, respectively. These findings align with the RMSE from the validation data, indicating that the SOM method effectively avoids overfitting.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

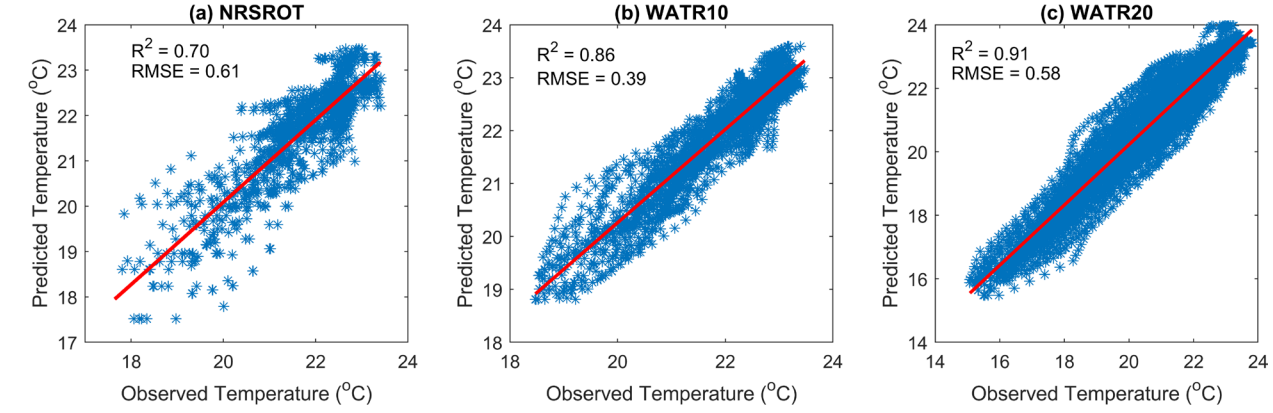

Figure 3. Observed and estimated temperatures at the three moorings between 1/1/2020 and 30/5/2020, a period of 150 days. The red lines are the line of best linear fit.

3 Data application

Having confirmed the effectiveness of the SOM method for filling missing values in a mooring dataset, we now employ all non-missing daily data to train the SOM, and then fill the data gaps. The filled temperature data exhibit consistent temporal and spatial variability (Fig. 5). The gap-filled data capture cold temperature events at WATR20 during early 2010 and mid-

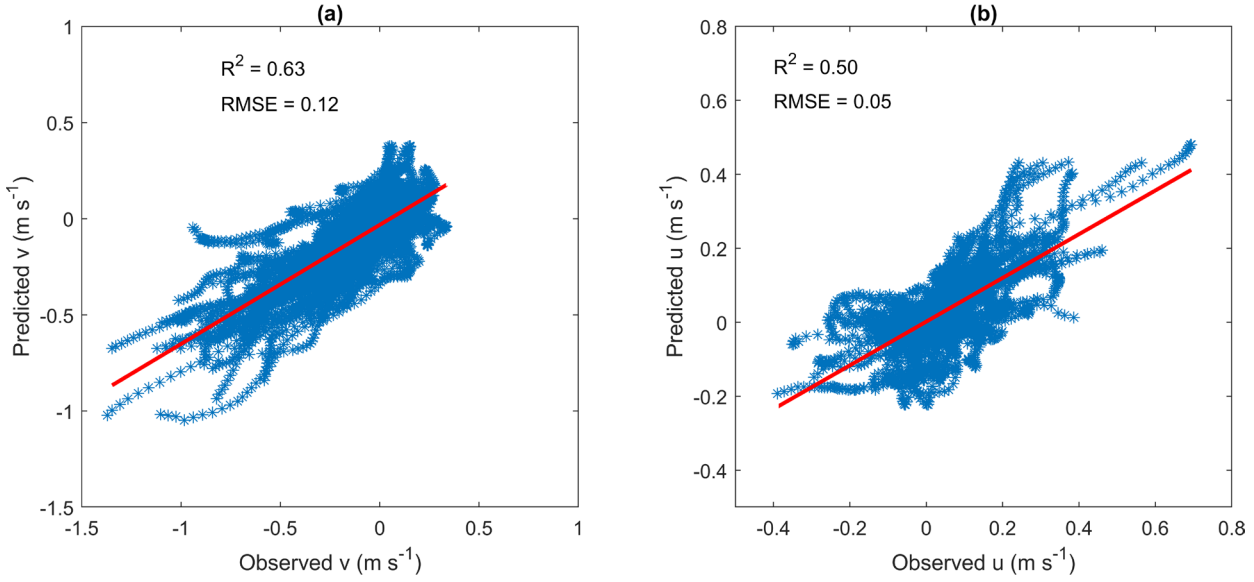

Figure 4. Observed and estimated (a) meridional and (b) zonal velocities at WATR50 between 5/2020 and 8/2020, a period of 90 days, and a depth range of 70-450 m. The red lines are the line of best linear fit.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

2016, coinciding with periods when the thermocline shoaled under the influence of El Niños, consistent with our understanding of the dynamics of the Leeuwin Current system (Feng et al., 2021).

The preprocessing of the input data via interpolation/extrapolation has dual advantages: (1) enhancing the accuracy of reference vectors in the SOM by increasing the number of good data profiles, and (2) reducing the potential for errors near the bottom depth in the input data. For example, without extrapolating the temperature data to the bottom, there are blocks of anomalous warm biases near the bottom depth in the filled data (Fig. S4).

Figure 6 compares the consistency between observed and filled temperature time series at three depths of the three moorings. The filled temperatures (shown in red lines) exhibit temperature variance similar to those of the observed time series. For example, at a depth of 95m at WATR10 towards the end of 2011, the filled temperature is anomalously warm, reflecting the enhanced Leeuwin Current system during a La Niña period (indicated by the red line rising above the black dashed line). Another example, at a depth of 190m at WATR20, during the beginning of 2010 and winter of 2016, the filled temperature was cooler than normal (indicated by the red line below the black dashed line) due to the shoaling of the thermocline towards the surface during El Niño episodes.

Continuous temperature time series are crucial for detecting subsurface marine heatwaves (MHWs) or marine cold spells (MCSs) that significantly affect marine ecosystems. Figure 7 shows the mean intensity of detected MHW or MCS events at WATR20 based on daily gap-filled temperatures. Following the intense MHWs during 2011-2013, MCSs occurred from 2016 to 2020, contributing to the recovery of impacted marine ecosystems (Fig. 7b). Many of the events are subsurface or bottom intensified, which are less detectable from ocean surface based on satellite data alone.

To highlight the role of data products in detecting subsurface marine heatwaves (MHWs), we examine several representative cases at three specific depths: NRSROT-40m, WATR10-80m, and WATR20-100m (Fig. 8). We also analyze the meridional component of velocity at these depths to explore the role of ocean currents in contributing to MHWs. A MHW at 40m depth at NRSROT lasted for 9 days in September 2020, with a maximum intensity of 1.5°C, and was classified as moderate strength (Category I) (Fig. 8a). During this period, the current was directed southward (Fig. 8b). A MHW at 100m depth at WATR10 lasted for a relatively longer duration of 20 days in September 2014, with a maximum intensity of 1.9°C, and was classified as strong (Category II). Although peak current occurred during the MHW event, it led to the peak temperature anomaly by 9 days (Fig. 8d). A MHW at 100m depth at WATR20 began on 13 August 2022, and lasted for 10 days with a maximum intensity of 1.4°C. Unlike the other events, the peak current led to the MHW timeframe, specifically on 10 August 2022. These observations suggest that strong southward currents often coincide with or precede MHWs by several days. This

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

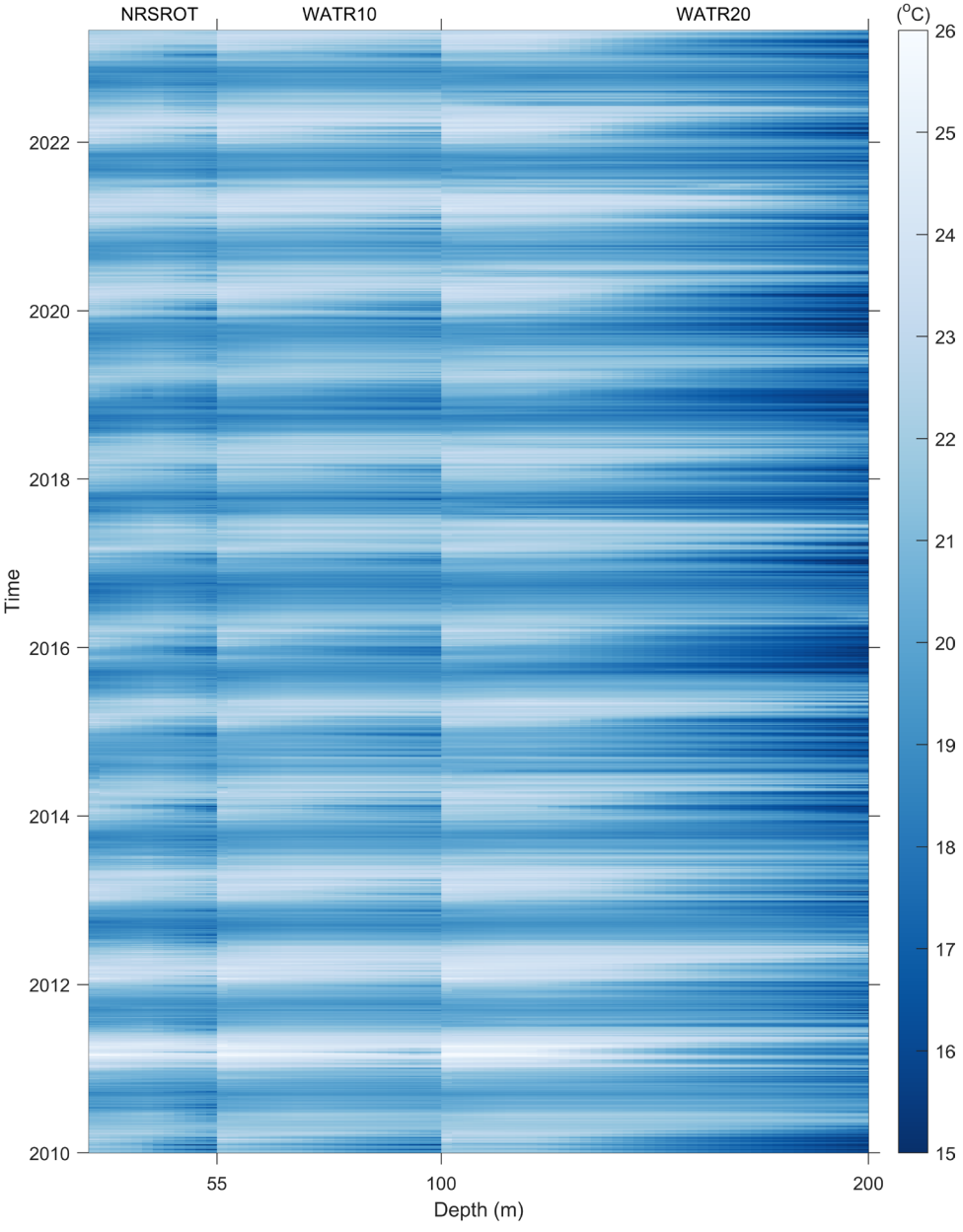

Figure 5. Data matrix of daily gridded, 5m resolution gap-filled temperatures for NRSROT, WATR10 and WATR20. The x axis shows the depth ranges of each moorings, while y axis presents time period from Jan 2010 to May 2023. Note that 0m follows directly after preceding mooring. The SST data are derived from the Regional Australian Multi-Sensor SST Analysis (RAMSSA) version 1.0. White spaces indicate missing observations.

183 indicates that the Leeuwin Current may be a significant factor driving subsurface MHWs on the Rottnest Shelf. Further research
184 is needed to clarify the underlying mechanisms.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

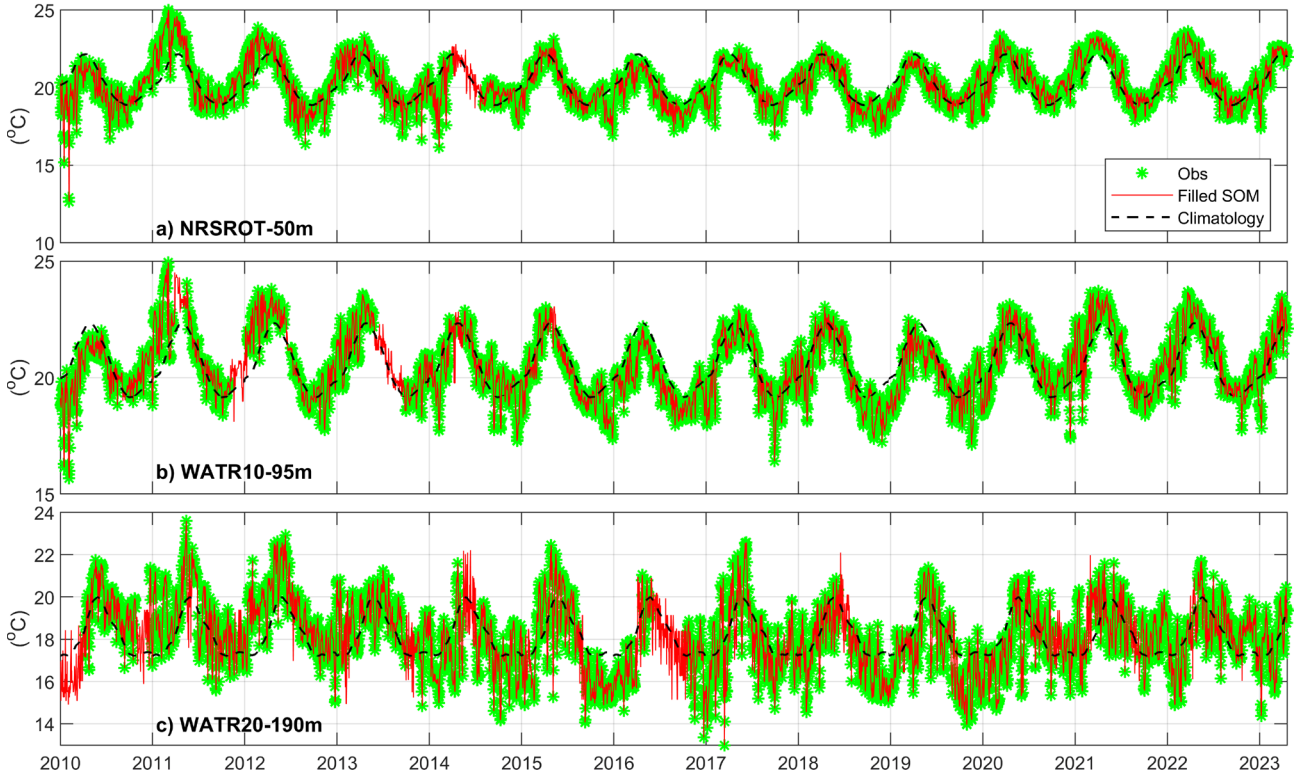

Figure 6. Comparison of observed and gap-filled temperature timeseries for a) NRSROT at 50m, b) WATR10 at 95m and c) WATR20 at 190m. The black dashed lines show daily climatological timeseries at corresponding depths. The climatological values are estimated from gap-filled data.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

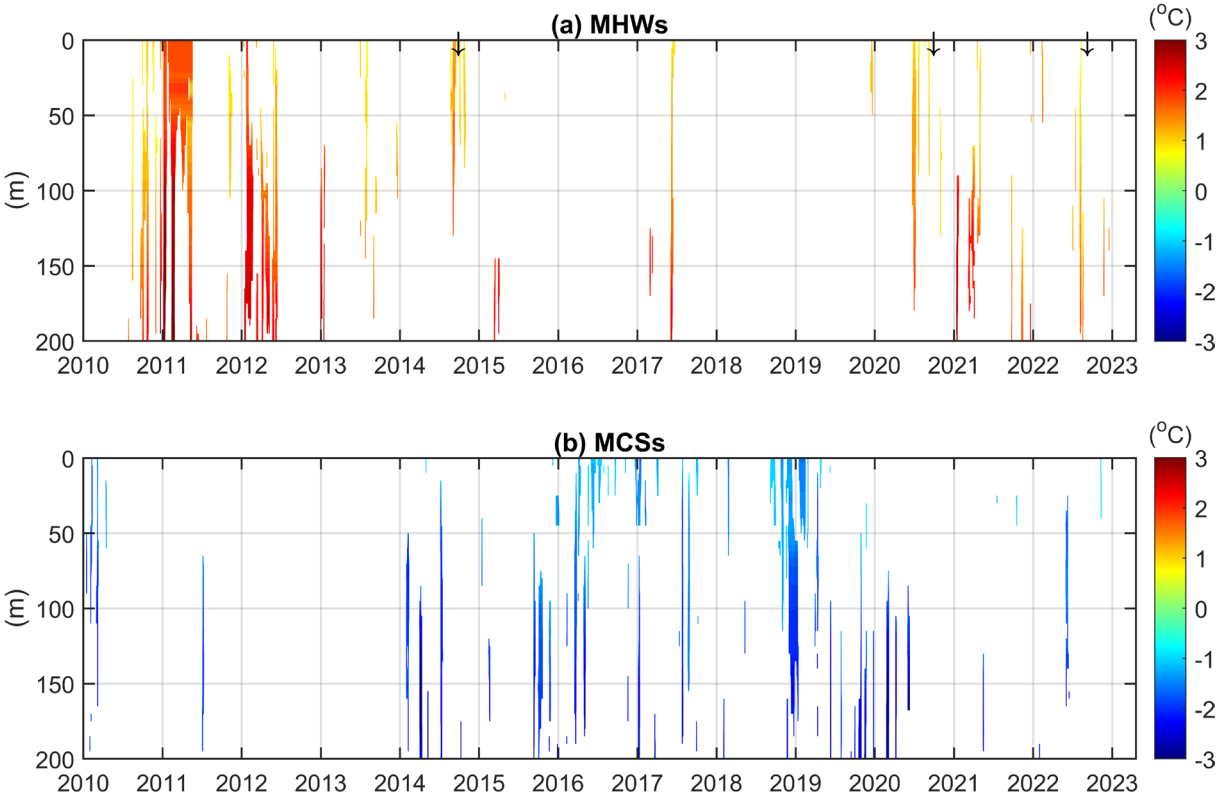

Figure 7. Mean intensity for individual (a) MHW and (b) MCS events at WATR20. Estimation is based on daily gap-filled temperature. The definition of each event follows Hobday et al. (2016). This plot is performed using Matlab code (Zhao and Marin, 2019). The threhold temperature indentifying a MHW or a MCS is set at the 90th and 10th percentile, respectively. Three arrows in (a) denote times of MHW events shown in Figure 8.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

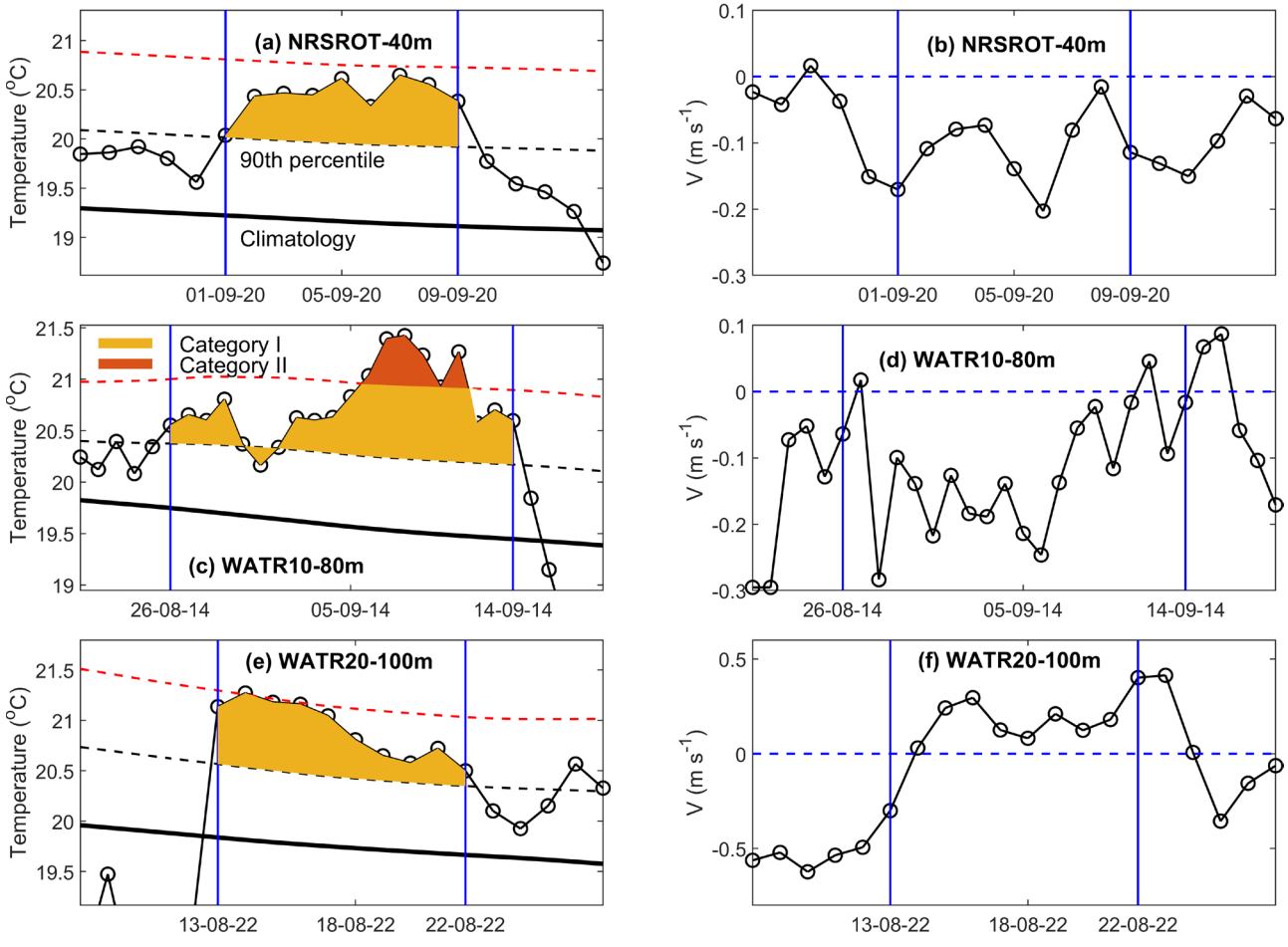

Figure 8. Left panels: Examples of marine heatwaves at NRSROT-40m (a), WATR10-80m (c), and WATR20-100m (e). Categories are moderate (yellow – category I) and strong (red – category II), as defined by Hobday et. al., 2018. In (a), the dashed red lines are estimated as twice the 90th percentile difference from the mean climatology value. Righ panels: Meridional component of the current velocity at the same time and depth as MHWs shown in the corresponding left panels. In all panels, vertical blue lines indicate the time frame of each MHW event.

Overall, the filled velocity data are consistent with temporal periods of data gaps both at the mooring location and at the adjacent mooring sites (Figs. S5 – S8). The observed mean vertical profiles agree well with those derived from the filled data (Fig. S7), indicating that the SOM method faithfully reconstructed the vertical structure of the LC.

The LC flows along the shelf break so that velocities measured at WATR20 and WACA20 are suitable for characterizing the LC. From the v-component data, the maximum mean southward currents recorded at WATR20 and WACA20 are 0.25 m s^{-1}, and 0.12 m s^{-1}, respectively (Fig. S7 d, b). Furthermore, the depths corresponding to these maximum values at the two stations are 80m and 100m, respectively. It can be inferred that the LC decelerates and deepens as it flows from WATR20 to WACA20. The irregular topography around the head of Perth Canyon may contribute to this disturbance (Fig. 1).

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

4 Data availability

The outcome of this research yields the in-filled data product, which is available at https://doi.org/10.25919/myac-yx60 (Bui and Feng, 2024). The product comprises continuous daily-5m resolution temperature and current variables (Table 2). All data products are available as NetCDF files. In addition to main parameters such as temperature and current, we provide quality control flags that indicate the original data sources. Specifically, we use seven flags for SOM-filled temperatures and four flags for SOM-filled currents, as detailed in Table 2.

We provide direct links to all datasets used in this study:

- The unfilled gridded data- https://doi.org/10.25919/9gb1-ne81 (Bui et al., 2023);

Table 3. Variables incuded in the in-filled data product.

Parameter	Variable name	Units	Description
Time	TIME		An array containing time information (days since 1950-01-01 00:00:00 UTC.
Depth	DEPTH	Meters (m)	An array containing depth levels.
Longitude	LONGITUDE	^{0}E	
Latitude	LATITUDE	^{0}N	
Temperature	TEMP	^{0}C	A matrix containing temperatures over the entire record for whole water column.
Temperature_quality control	TEMP_quality_control		A matrix containing flag values that indicate the original temperature data. 1: Observed temperature 2: SST 3: Interpolated temperature near surface 4: Extrapolated temperature near bottom 5: SOM filled temperature near surface 6: SOM filled temperature in sensor range 7: SOM filled temperature near bottom
U velocity	UCUR	m s^{-1} (true east)	A matrix containing current data over the entire record for whole water column.
V velocity	VCUR	m s^{-1} (true north)	
Current_quality_control	UCUR_ quality_control VCUR_ quality_control		A matrix containing flag values that indicate the original current data. 1: Observed current 2: Extrapolated current near surface 3: Extrapolated current near bottom 4: SOM filled current

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

-

- Satellite sea surface temperature from the Regional Australian Multi-Sensor SST Analysis (RAMSSA)- https://portal.aodn.org.au (Beggs et al., 2011);

- Fremantle sea level from the University of Hawaii Sea Level Center - https://uhslc.soest.hawaii.edu.

5 Code availability

We provide scripts in MATLAB to download and plot the data products. These scripts are available online (Bui and Feng, 2024), and are available under a Creative Commons Attribution 4.0 International license (CC BY 4.0).

6 Summary and discussion

In this research, we have employed a SOM-based method to fill significant temperature and velocity measurement gaps from a mooring array on the Rottnest shelf off southwest Western Australia that monitors the Leeuwin Current and associated shelf processes. We use daily temperature records from 3 moorings of approximately 14.5 years, and nearly 13 years of daily current velocity records from 5 moorings, in conjunction with daily SST and coastal sea level at Fremantle, to train SOM. Because this is a relatively small mooring array, we pre-process observational data using interpolation and extrapolation to have enough non-missing daily data profiles to train SOM. Evaluated with withholding data, the RMSE for temperature estimations at the 3 moorings are 0.61ºC at NRSROT, 0.39ºC at WATR10, and 0.58ºC at WATR20, respectively. The RMSE for the meridional (alongshore) and zonal (cross-shore) velocities are 0.1 m s^{-1} and 0.05 m s^{-1}. In addition, the data pre-processing brings better consistency between the observed and gap-filled data.

Our continuous daily data products reveal that many MHW and MCS events occur sub-surface, which are undetectable while using altimetry data. We also find that intense MHW events are frequently related to strong southward currents. Sustaining mooring observations into the future are needed to understand the long-term trends of MHWs and MCSs, as well as the factors driving extreme temperatures.

Addressing small gaps in the mooring data appears to be a crucial step before training SOM. We have tried two other options: assigning missing values as zeros or replacing them with climatological values derived from the original data. We have experimented with these two options with an iterative approach (e.g. Sloyan et al. 2023) but found that the filled temperature time series exhibits some inconsistency, such as a block of constant values or temperature inversions. Our option of pre-processing the observational data by filling small gaps increases the number of good profiles for training, e.g., 75% of temperature profiles are gap-free. The method can be easily applied to fill data gaps in shelf mooring arrays with small gaps in the vertical so that little errors are introduced from linear extrapolation. For complex mooring systems with enough

246 redundancy, the Iterative Completion Self-Organizing Maps (ITCOMPSOM) method outlined in Sloyan et al. (2023) could be
247 more useful.

248 **7 Supplement**

249 The supplement related to this article is available online at: https://doi.org/10.25919/myac-yx60.

250 **8 Author contributions**

251 MF conceptualized and designed the study. TB and MF performed the study, with SOM source codes provided by CC.
252 TB processed the data, produced the figures and first draft of the manuscript and associated data products; and MF and CC
253 reviewed and edited the manuscript.

254 **9 Competing interests**

255 The contact author has declared that none of the authors has competing interests.

256 **10 Disclaimer**

257 Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text,
258 published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications
259 makes every effort to include appropriate place names, the final responsibility lies with the authors.

260 **11 Acknowledgements**

261 CSIRO collected mooring data under the Integrated Marine Observing System (IMOS) program. IMOS is enabled by
262 the National Collaborative Research Infrastructure Strategy (NCRIS). Satellite SST sourced from the Regional Australian
263 Multi-Sensor SST Analysis (RAMSSA) version 1.0. Daily Fremantle sea level was downloaded from the University of Hawaii
264 Sea Level Center. We thank Ryan Crossing, Ian Darby, Mark Snell, Beau De Groot, and many others in the deployment and
265 recovery of the moorings, and Mark Snell and Miaoju Chen for quality control of the mooring data. We thank Bernadette
266 Sloyan for constructive discussion and for sharing the code employed by Sloyan et al. (2023).

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

12 References

Bailey, K., Steinberg, C., Davies, C., Galibert, G., Hidas, M., McManus, M. A., Murphy, T., Newton, J., Roughan, M., and Schaeffer, A.: Coastal mooring observing networks and their data products: recommendations for the next decade, Frontiers in Marine Science, 6, 180, https://doi.org/10.3389/fmars.2019.00180, 2019.

Beggs, H., Zhong, A., Warren, G., Alves, O., Brassington, G., and Pugh, T.: RAMSSA—An operational, high-resolution, regional Australian multi-sensor sea surface temperature analysis over the Australian region, Australian Meteorological and Oceanographic Journal, 61, 1, 2011.

Benthuysen, J., Feng, M., and Zhong, L.: Spatial patterns of warming off Western Australia during the 2011 Ningaloo Niño: Quantifying impacts of remote and local forcing, Continental Shelf Research, 91, 232-246, https://doi.org/10.1016/j.csr.2014.09.014, 2014.

Bui, T. and Feng, M.: Gap-filled, gridded subsurface physical oceanography time series dataset derived from selected mooring measurements off the Western Australia coast during 2009-2023, https://doi.org/10.25919/myac-yx60, 2024.

Bui, T., Feng, M., and Snell, M.: An updated dataset for "A long-term, gridded, subsurface physical oceanography dataset and average annual cycles derived from in situ measurements off the Western Australia coast during 2009-2020", https://doi.org/10.25919/9gb1-ne81, 2023.

Chapman, C. and Charantonis, A. A.: Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps, IEEE Geoscience and Remote Sensing Letters, 14, 617-620, https://doi.org/10.1109/Lgrs.2017.2665603, 2017.

Chen, M. and Feng, M.: A long-term, gridded, subsurface physical oceanography dataset and average annual cycles derived from in situ measurements off the Western Australia coast during 2009–2020, Data in Brief, 35, 106812, https://doi.org/10.1016/j.dib.2021.106812, 2021.

Feng, M. and Wijffels, S.: Intraseasonal variability in the South Equatorial Current of the east Indian Ocean, Journal of physical oceanography, 32, 265-277, https://doi.org/10.1175/1520-0485(2002)032<0265:Ivitse>2.0.Co;2, 2002.

Feng, M., McPhaden, M. J., Xie, S.-P., and Hafner, J.: La Niña forces unprecedented Leeuwin Current warming in 2011, Scientific Reports, 3, 1277, https://doi.org/10.1038/srep01277, 2013.

Feng, M., Meyers, G., Pearce, A., and Wijffels, S.: Annual and interannual variations of the Leeuwin Current at 32°S, J. Geophys. Res, 108, https://doi.org/10.1029/2002JC001763, 2003.

Feng, M., Caputi, N., Chandrapavan, A., Chen, M., Hart, A., and Kangas, M.: Multi-year marine cold-spells off the west coast of Australia and effects on fisheries, Journal of Marine Systems, 214, 103473, https://doi.org/10.1016/j.jmarsys.2020.103473, 2021.

Godfrey, J. and Ridgway, K.: The large-scale environment of the poleward-flowing Leeuwin Current, Western Australia: longshore steric height gradients, wind stresses and geostrophic flow, Journal of Physical Oceanography, 15, 481-495, https://doi.org/10.1175/1520-0485(1985)015<0481:Tlseot>2.0.Co;2, 1985.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., Benthuysen, J. A., Burrows, M. T., Donat, M. G., and Feng, M.: A hierarchical approach to defining marine heatwaves, Progress in oceanography, 141, 227-238, https://doi.org/10.1016/j.pocean.2015.12.014, 2016.

Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps, Bio. Cyber., 43, 59-69, https://doi.org/10.1007/Bf00337288, 1982.

Liu, Y. and Weisberg, R. H.: A review of self-organizing map applications in meteorology and oceanography, Self-organizing maps: applications and novel algorithm design, 1, 253-272, https://doi.org/10.5772/566, 2011.

Lobo, V. J.: Application of self-organizing maps to the maritime environment, Information Fusion and Geographic Information Systems: Proceedings of the Fourth International Workshop, 17-20 May 2009, 19-36,

Morello, E. B., Galibert, G., Smith, D., Ridgway, K. R., Howell, B., Slawinski, D., Timms, G. P., Evans, K., and Lynch, T. P.: Quality Control (QC) procedures for Australia's National Reference Station's sensor data—Comparing semi-autonomous systems to an expert oceanographer, Methods in Oceanography, 9, 17-33, https://doi.org/10.1016/j.mio.2014.09.001, 2014.

Sloyan, B. M., Cowley, R., and Chapman, C. C.: East Australian Current velocity, temperature and salinity data products, Scientific Data, 11, 10, https://doi.org/10.1038/s41597-023-02857-x, 2024.

https://doi.org/10.5194/essd-2024-449
Preprint. Discussion started: 10 October 2024

Sloyan, B. M., Chapman, C. C., Cowley, R., and Charantonis, A. A.: Application of Machine Learning Techniques to Ocean Mooring Time Series Data, Journal of Atmospheric and Oceanic Technology, 40, 241-260, https://doi.org/10.1175/Jtech-D-21-0183.1, 2023.

Smith, R. L., Huyer, A., Godfrey, J. S., and Church, J. A.: The Leeuwin current off western Australia, 1986–1987, Journal of Physical Oceanography, 21, 323-345, https://doi.org/10.1175/1520-0485(1991)021<0323:Tlcowa>2.0.Co;2, 1991.

Sprintall, J., Wijffels, S. E., Molcard, R., and Jaya, I.: Direct estimates of the Indonesian Throughflow entering the Indian Ocean: 2004–2006, Journal of Geophysical Research: Oceans, 114, https://doi.org/10.1029/2008jc005257, 2009.

Thompson, R. O.: Observations of the Leeuwin current off Western Australia, Journal of physical oceanography, 14, 623-628, https://doi.org/10.1175/1520-0485(1984)014<0623:Ootlco>2.0.Co;2, 1984.

Wang, Y., McPhaden, M. J., Freitag, H. P., and Fey, C.: Moored acoustic Doppler current profiler time series in the central equatorial Indian Ocean, http://doi.org/10.7289/V5HX19NP
2015.

Zhao, Z. and Marin, M.: A MATLAB toolbox to detect and analyze marine heatwaves, J. Open Source Softw., 4, 1124, https://joss.theoj.org/papers/10.21105/joss.01124, 2019.