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. However, we often cannot obtain continuous time series data from moorings due to mooring losses or instrument failures. Here, we present an updated version of the 14-year subsurface mooring dataset off the southwest coast of Western Australia (WA) 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 data captures the Leeuwin Current variability on the shelf from intraseasonal to interannual time scales. The data also capture the variability of the Capes Current, a wind-driven northward current on the middle shelf. The usage of the in-filled data product is demonstrated by detecting extreme temperature events 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 decadal change signals along the WA coast.

1 Introduction

Oceanography moorings are underwater instruments anchored on the sea floor that collect ocean currents, temperature, salinity, and other environmental parameters. Typically, mooring deployment periods range from 4-6 months in shelf waters to up to 18 months in deep oceans (Sloyan et al., 2024). 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).

28 The southwest Western Australian mooring array is part of the Integrated Marine Observing System (IMOS) program

29 operated by Commonwealth Scientific and Industrial Research Organisation (CSIRO) since 2009, designed to monitor the

30 influences of the southward-flowing Leeuwin Current (LC) on the continental shelf (Thompson, 1984; Chen and Feng, 2021).

31 The anomalous meridional pressure gradient, associated with warm, low-salinity waters from the tropical Pacific Ocean

32 entering the Indian Ocean through the Indonesian Archipelago, is the main driver of the LC (Feng and Wijffels, 2002; Godfrey

33 and Ridgway, 1985). The strength of the LC varies seasonally, most due to variations in the alongshore winds (Smith et al.,

34 1991). During austral summer, strong alongshore northward winds drive northward Capes Current in the mid-inner shelf (Fig.
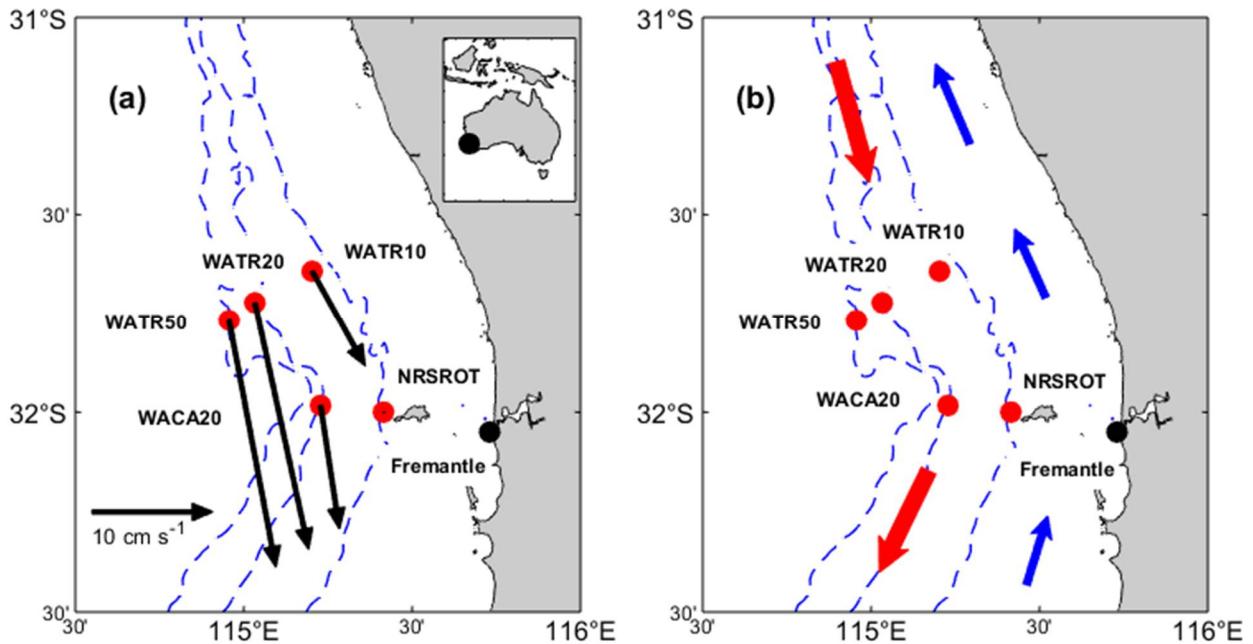
Figure 1. Bathymetry map and mooring locations (red circles) on the Rottnest Shelf. (a) Velocities estimated from measurements, with black arrows representing the mean state of vertically averaged velocities. The 0-200m average is used for the WATR50 mooring. The three dashed lines represent the 50m, 200m, and 500m contours. Black circles indicate the location of the Fremantle tide gauge station. Note that NRSROT consists of two separate moorings. (b) Schematic of shelf currents, with red arrows denoting the Leeuwin Current and blue arrows indicating the direction of the wind-driven Capes Current.

35 1). The interannual variability of the LC is often associated with remote signals from the Pacific, the El *Niño*–Southern

36 Oscillation (ENSO), the current being stronger during La Niña and weaker during El Niño (Feng et al., 2003).

37 The southwest Western Australian mooring array has helped scientists identify the key role of the LC in the development

38 of marine heatwaves (MHW) off the coast (Benthuysen et al., 2014; Feng et al., 2013). The mooring data has also been

39 employed by Feng et al. (2021) to detect abnormal cooling events off the coast over 2016-2019 (defined as the marine cold

40 spell, MCS) when the thermocline depth was elevated due to the weakening of the LC during the El Niño events. The sustained

41 IMOS mooring array encompasses six coastal moorings on the Rottnest shelf during 2010-2023, ranging from 50m to 500m

42 (Fig. 1 and Table 1). The first version of the gridded data from these moorings was published by Chen and Feng (2021) and an

43 extension was published by Bui et al. (2023). Mooring time series are susceptible to missing values due to mooring loss and

44 instrument failure. Strong currents can exert force on the mooring line, causing it to be pushed down in the water column,

45 leaving data gaps near the surface (Sloyan et al., 2023). This paper introduces a new update of the mooring data, filling data

46 gaps with a statistical method.

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69 **Table 1. Summary of coastal mooring stations. NRSROT: National Reference Station west of the Rottnest Island. WACA: Western**
70 **Australia Perth Canyon. WATR: Western Australia Two Rocks.**

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[a] SBE37[b]	5-15	27; 33; 43; 55	1/2010-5/2023				
NRSROT-ADCP	32.0000°S; 115.4170°E;	48					RDI Workhorse 600 kHz[c]; Nortek Signature 500 kHz[d]	15	11x4m	8/2011-5/2023
WACA20	31.9830°S; 115.2280°E	200					Nortek Signature 250 kHz[d]; Nortek Continental 190 kHz[d]	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[d];	15	17x5m	8/2011-5/2023
WATR20	31.7233°S; 115.0333°E	200	SBE39 SBE37	5-15	25; 35; 50; 68; 100; 125; 150; 175	1/2010-5/2023	Nortek Continental 190 kHz[d]; Nortek Signature 250 kHz[d]	15	25x8m	8/2011-5/2023

WATR50	31.7683°S; 114.9567°E	500					RDI Long Ranger 75 kHz[c]; Nortek Signature 55 kHz[d]	15	26x20m	8/2011-5/2023

a. SBE39 (and SBE39 plus) is a self-contained, autonomous temperature logger. (SBE: Sea-Bird Electronics).

b. SBE37 is a single-channel CTD (Conductivity, Temperature, Depth) sensor.

c. RDI ADCPs (Acoustic Doppler Current Profiles) are manufactured by Teledyne RD Instruments and comprise Long Ranger 75 kHz and Workhorse 600 kHz. (https://www.teledynemarine.com/rdi).

d. Nortek ADCPs are produced by Nortek group, including Signature 55 kHz, Continental 190 kHz, Signature 250 kHz, Aquadopp 400 kHz and Signature 500 kHz. (https://www.nortekgroup.com).

71  Various techniques have been employed to address gaps in mooring datasets. Sprintall et al. (2009) utilized a damped
72 least square fitting method to fill substantial gaps in mooring current time series data, in estimating the Indonesian Throughflow
73 transport. Wang et al. (2015) adopted a combination of data extrapolation, interpolation, and a least square regression model
74 to fill in missing data recorded in the central equatorial Indian Ocean. Cao et al. (2015) employed harmonic analysis and modal
75 decomposition to isolate the tidal currents for each mode and reconstruct the full-depth tidal currents in the northern South
76 China Sea. More recently, Sloyan et al. (2023) experimented with a machine learning approach, the Self-Organizing Map
77 (SOM), to fill data gaps in the East Australian Current mooring array. The choice of method depends on the characteristics of
78 data loss, such as the duration of gaps or the depth range affected, as well as the intended analyses of the data.

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

84  Chapman and Charantonis (2017) utilized SOM to reconstruct deep current velocities in the Southern Ocean from sea
85 surface observations. They used densely observed surface velocities, sea surface height, sea surface temperature from satellites,
86 and sparsely observed deep current velocities from Argo floats to train the SOM maps. Then, they derived dense velocity fields
87 at a depth of 1000m. Their method took advantage of local correlations in the data space to find the smallest Euclidean distance,
88 weighted by the local correlations, between a vector with missing components in the data space and the SOM units, which
89 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. The RAMSSA system combines SST data from infrared and microwave sensors on polar-orbiting satellites with *in situ* measurements to generate daily foundation SST. North of 40°S, RAMSSA is on average within ±0.07 °C of other multi-sensor SST analyses (Beggs et al., 2011). From conductivity-temperature-depth (CTD) profiles in the study region, ocean temperatures vary mostly linearly in the near-surface layer (top 30 m, below the foundation SST depth), so linear interpolation is an acceptable approximation.

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.

These procedures produce daily 5m-vertical resolution, gridded temperatures at NRSROT, WATR10, and WATR20 moorings, spanning from January 2010 to May 2023, as presented in Figure S1.

2.1.2 Velocity

The velocity observations on the IMOS mooring array are recorded by various instruments such as RDI and Nortek ADCP instruments, typically sampling at 15-minute intervals, and are mounted in an upward-looking configuration above the seabed (Table 1).

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

124   For the 2010-2023 period, the percentage of missing mooring data varies from 2% to 16% for temperature, and 12%-
125 33% for velocity at various moorings (Table 2). The largest percentage of missing data is at WATR20, situated in the core of
126 the LC system. The percentages of missing data tend to have high values near the surface and bottom layers of a mooring, due
127 to mooring movement and variations of deployment depth over time (Fig. S4 and S5).

128

2.2 SOM method

130   To produce a gap-filled data product, we follow the method described in Chapman and Charantonis (2017). As discussed
131 briefly in the introduction, this method "completes" a gappy dataset by first using available data to train SOM, which
132 effectively clusters the data into a set of discrete states. These states can be represented as a 2-dimensional map, where
133 neighbouring clusters are more similar to each other than distant clusters. Associated with each cluster is a *referentee* *vector*
134 that approximates the mean of all data assigned to that cluster and a weighted mean of data assigned to neighbouring clusters.
135 After the map is trained, new data can be assigned to existing clusters by comparing the Euclidean distance in data space
136 between that new data vector and the referentee vector of each cluster. The cluster with the smallest Euclidean distance is
137 known as the Best Matching Unit (BMU). Once a SOM is available, data vectors with missing components are presented
138 sequentially, the BMU is found, and the missing data is completed (in-filled) by replacing it with the relevant components of
139 the reference referent vector of the BMU. For full details, see Chapman and Charantonis (2017).

Table 2. Percentage (%) of missing temperature and velocity for each mooring for the time period of 2010-2023. Note that temperature profiles are not available at WACA20 and WATR50.

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

140   A schematic of using the SOM method to fill gaps in the mooring dataset is shown in Figure 2. We utilized the Vesanto
141   et al. (2000) SOM toolbox for MATLAB 5 in this study. The temperature or velocity data for each station, along with ancillary
142   data, are aggregated into data matrices. Ancillary data include day-of-year and daily Fremantle sea level (Fig. 1). Sea level
143   data are obtained from the University of Hawaii Sea Level Center (https://uhslc.soest.hawaii.edu/). Fremantle sea level serves
144   as a proxy for the annual and interannual variations of the Leeuwin Current (Feng et al., 2003). We have tested adding
145   alongshore winds to the data matrices, however, there is no improvement to the results, so wind data is not used in the SOM
146   calculation, as the wind information may have been integrated in the sea level data. The temperature input matrix comprises
147   4869 rows (representing the number of time steps) and 77 columns (reflecting the number of different observations at each
148   time step). Similarly, the velocity input matrix consists of 4292 rows and 361 columns. The temperature/velocity input matrix
149   with missing values is indicated by Dataset 1 in Figure 2. Only fully available profiles in the input matrix are selected as the
150   training data shown by Dataset 2. Consequently, the number of rows in the training data is 3675 for temperature and 1146 for
151   velocity.

152   The number of units in the SOM is specified prior to the training process. According to the literature, a small number
153   of SOM units is useful in capturing the general features of the system (Liu and Weisberg, 2011), while a larger number provides
154   more detailed information and is more suitable for data gap filling (Sloyan et al., 2023). In our case, where we aimed to capture
155   detailed information from the training data containing a large number of profiles, we opted for a larger number of units, 1000
156   units for the temperature data and 500 units for the velocity data. Using lower numbers of units only had minor effects on the
157   results. We used a batch algorithm to train the SOM (Chapman and Charantonis, 2017). The training phase of SOM was done
158   in two steps: the first rough phase, followed by a fine-tuning phase. In the first step, the neighbourhood radius and learning
159   rate were set to some high values to gain a general orientation of the map, while in the second step, they were set to smaller
160   values to perform only fine adjustments on the SOM unit's position.
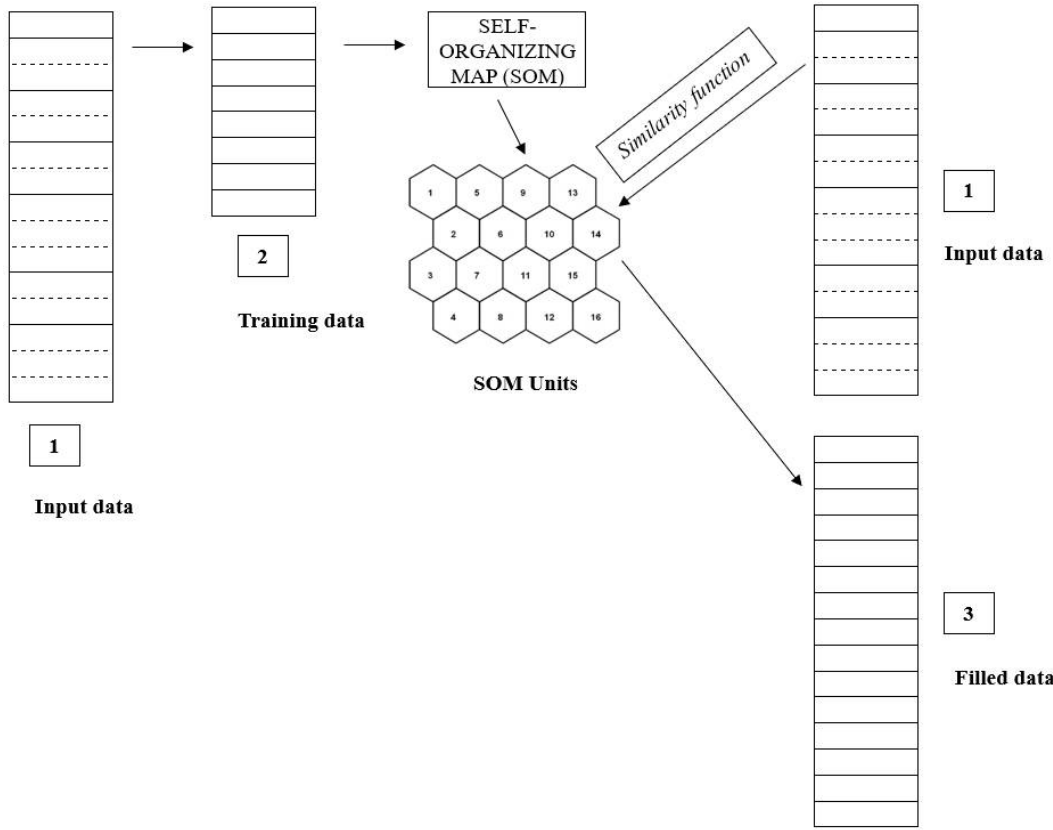
Figure 2. Schematic of the SOM method applied to fill gaps in the mooring temperature and velocity data. Dataset 1 denotes an input data matrix in which rows are daily time vectors, columns are observational variables. In Dataset 1, solid lines present full available profiles, while dashed lines show the profiles including missing values. In Dataset 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 referen~eet~ 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, 2017. Finally, we use the referent vector of each SOM to fill gaps in corresponding daily input vector, shown in Dataset 3.

161

162   One of the important steps was the assignment of each input vector to a specific SOM unit, *u,* shown on the right-hand

163   side of Figure 2. Firstly, we estimated the local correlations in the data space, represented by a $cor_{i,j}^{u}$~COR~ matrix.

$$cor_{i,j}^{u}\text{\sout{COR}} = 1 + \sqrt{\sum DAT\_cor^2}, \tag{1}$$

164   Where *DAT_cor* is a correlation matrix among each normalized input vectors within a SOM unit; $cor_{i,j}^{u}$ is the local correlation

165   matrix between the missing and the mean of all the observed training data within the SOM unit *u.*

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

$$sim(X, ref^u) = \sum_{i \in non-missing} \left(1 + \sum_{j \in missing} \left(cor_{i,j}^u\right)^2 \right) \times \sqrt{(X_i - ref_i^u)^2} ~~\sum \left[\sqrt{(ref^u - X)^2} \right]~~, \qquad (2)$$

Where X_i is the non-missing data in X, ref_i^u is the mean of all training data in the SOM unit u. 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, Dataset 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 Figure 3. Furthermore, we evaluate the ability of the SOM to reconstruct extreme temperature patterns. As shown in Figure S5, a comparison of the observed and SOM-derived temperatures at WATR20 during the validation period (January 1 to May 30, 2020) highlights this capability. Black crosses in both panels denote days identified as marine cold spells (MCS), defined as periods where temperatures fall below the 10th percentile for at least five consecutive days (Hobday et al., 2016). SOM-derived temperatures successfully captured three bottom-intensified MCS events as in observations, demonstrating the method's reliability in reconstructing extreme cold temperature patterns.

To assess potential overfitting, the SOM was tested on a separate period, spanning January 10 to June 8, 2012, with 150 days withheld from training. The resulting root mean square error (RMSE) values were 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.

To assess further the accuracy of the SOM method, we compare it with a simple climatology method over the same validation period, as shown in Figure S6. Overall, the mean vertical temperature profiles from the SOM method are closer to the observed data than those from the climatology method (Fig. S6a-c). As a result, the residuals from the SOM method, calculated by subtracting the filled SOM values from the observed temperatures, are smaller than the climatological residuals. Additionally, the standard deviation of the observed temperatures is closer to that of the SOM data, while it differs significantly

from the climatological values (Fig. S6d-f). These findings suggest that the SOM method is more reliable than the climatology method.

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.
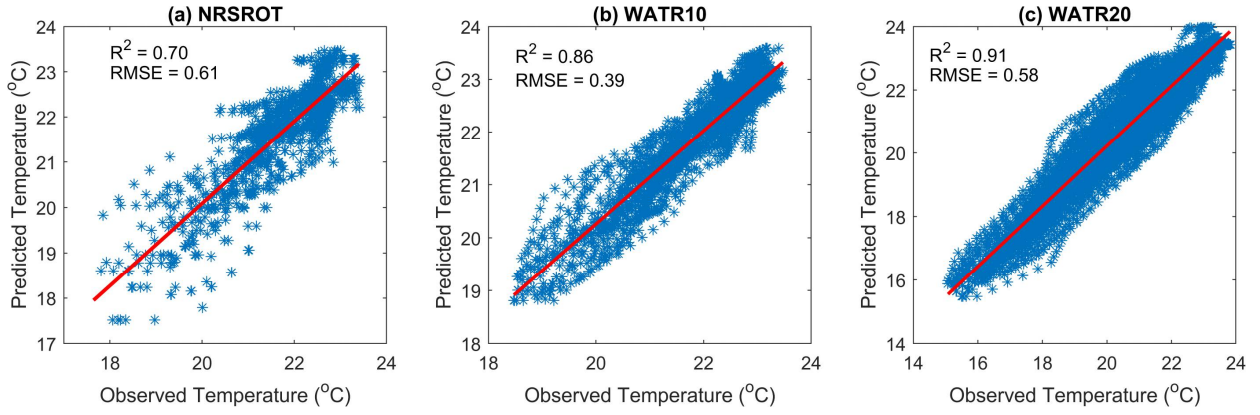
Figure 3. Scatter plots of observed and SOM-derived temperatures at the three moorings between 1/1/2020 and 30/5/2020, a period of 150 days. The red lines are the linear fits of the scatter plots.

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-
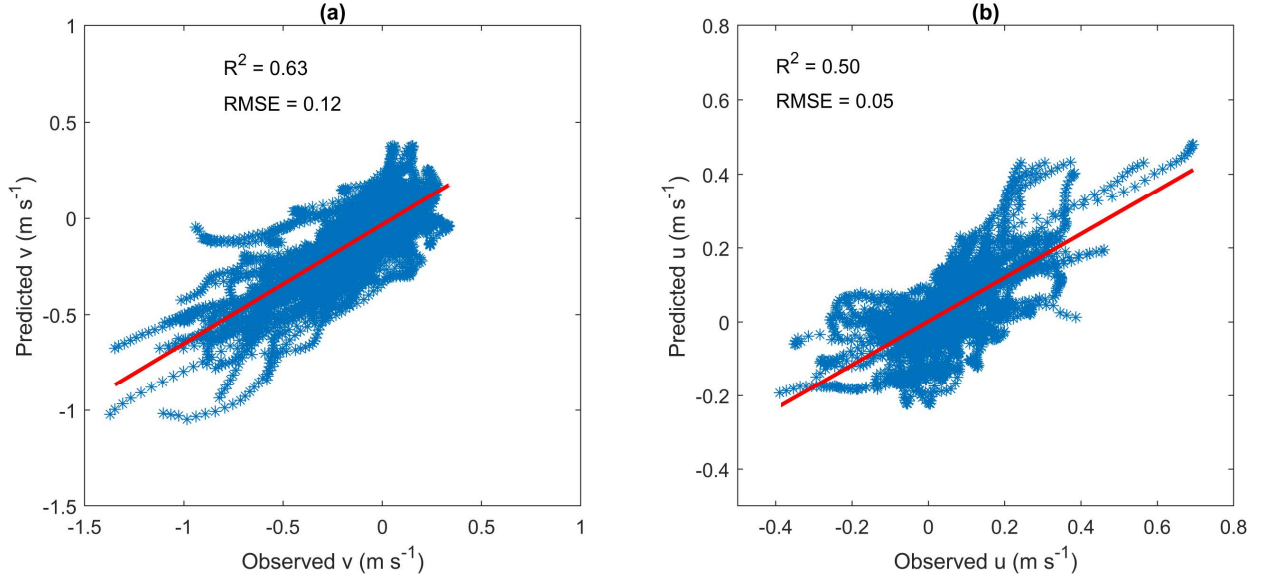
Figure 4. Scatter plots of observed and SOM-derived (a) meridional and (b) zonal velocities at WATR50 between 5/2020 and 8/2020, a period of 90 days, and a depth range of 70-450 m. The red lines are the linear fits.

208 2016, coinciding with periods when the thermocline shoaled under the influence of El Niños, consistent with our understanding
209 of the dynamics of the Leeuwin Current system (Feng et al., 2021).

210   The preprocessing of the input data via interpolation/extrapolation has dual advantages: (1) enhancing the accuracy of
211 refereneet vectors in the SOM by increasing the number of good data profiles, and (2) reducing the potential for errors near
212 the bottom depth. For example, without extrapolating the temperature data to the bottom, there are blocks of anomalous warm
213 biases near the bottom depth in the SOM-derived data (Fig. S7).

214   Figure 6 compares the consistency between observed and gap-filled temperature time series at three specific depths.
215 The filled temperatures (shown in red lines) exhibit temperature variance similar to those of the observed time series. For
216 example, at a depth of 95m at WATR10 towards the end of 2011, the filled temperature is anomalously warm, reflecting the
217 strengthened Leeuwin Current system during a La Niña period (Feng et al., 2013), as shown by the red line rising above the
218 black dashed line. Another example, at a depth of 190m at WATR20, during the beginning of 2010 and in the winter of 2016,
219 the filled temperature was cooler than normal (indicated by the red line below the black dashed line) due to the shoaling of the
220 thermocline towards the surface during El Niño episodes.

221   Continuous temperature time series are crucial for detecting subsurface marine heatwaves (MHWs) or marine cold
222 spells (MCSs) that significantly affect marine ecosystems (Smale et al., 2019). Figure 7 shows the mean intensity of detected
223 MHW or MCS events at WATR20 based on daily gap-filled temperatures. The definition of each MHW or MCS event is based
224 on Hobday et al. (2016). An MHW (MCS) event is classified as a thermal event when its temperature exceeds the 90th
225 percentile threshold (or falls below the 10th percentile threshold) for at least 5 days. Additionally, two consecutive events
226 occurring within a temporal gap of less than two days are considered a single combined event. This plot is performed using
227 MATLAB code for MHW/MCS detections (Zhao and Marin, 2019). Following the intense MHWs during 2011-2013 (Fig. 7a),
228 MCSs occurred from 2016 to 2020, contributing to the recovery of impacted marine ecosystems (Fig. 7b). Many of the events
229 are subsurface or bottom intensified, which are less detectable from ocean surface based on satellite data alone.

230   To highlight the role of data products in detecting subsurface marine heatwaves (MHWs), we examine several
231 representative cases at three specific depths of different moorings: NRSROT-40m, WATR10-80m, and WATR20-100m (Fig.
232 8). We also analyze the meridional component of velocity at these depths to explore the role of ocean currents in contributing
233 to MHWs. In this study, different categories of MHWs are defined based on multiples of the local difference between the
234 climatological mean and the 90th percentile (Hobday et al., 2018). The magnitude scale descriptors classify MHWs as moderate
235 (between 1–2 multiples×, Category I), strong (2–3×, Category II), severe (3-4×, Category III), and extreme (>4×, Category
236 IV). A MHW at- 40m depth at NRSROT lasted for 9 days in September 2020, with a maximum intensity of 1.5°C, and was
237 classified as moderate strength (Category I) MHW (Fig. 8a). During this period, the current was directed southward (Fig. 8b).
238 A MHW at 100m depth at WATR10 lasted for a relatively longer duration of 20 days in September 2014, with a maximum
239 intensity of 1.9°C, and was classified as strong (Category II) MHW. Although peak current occurred during the MHW event,

240   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

241   lasted for 10 days with a maximum intensity of 1.4°C. Unlike the other events, the peak current led to the MHW timeframe,

242   specifically on 10 August 2022. These observations suggest that strong southward currents often coincide with or precede

243   MHWs by several days. Further research is needed to clarify the impact of the Leeuwin Current in driving subsurface MHWs

244   on the Rottnest Shelf. In addition, we zoomed in on the SOM-filled temperatures from January to July 2011 when there was a

245   two-month gap at the WATR10 mooring (Fig. S8). The gap-filled temperatures at WATR10 (Fig. S8d) enabled us to detect the

246   MHW events across the water column.

247

248

249
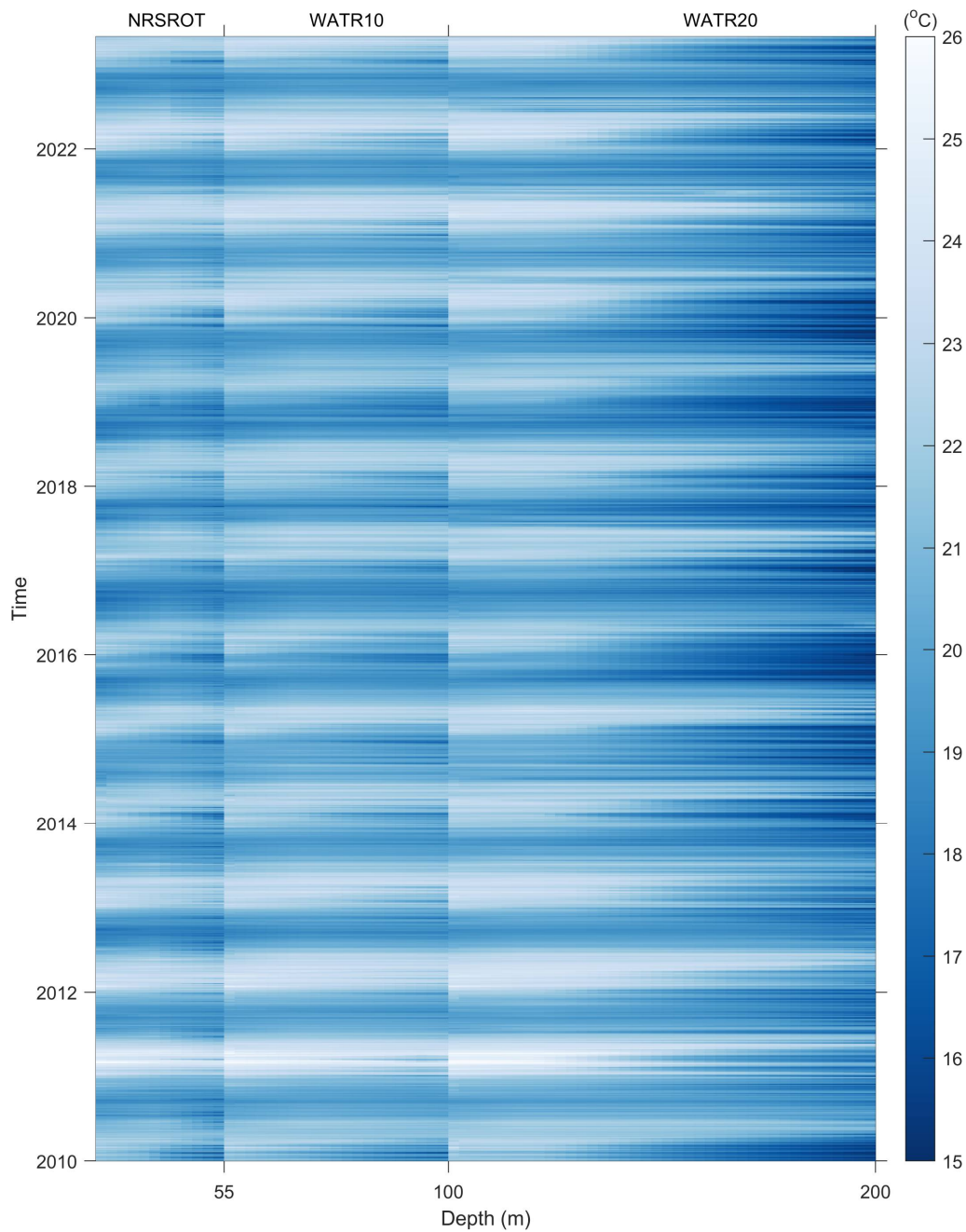
Figure 5. Data matrix of daily gridded, 5m resolution gap-filled temperatures for NRSROT, WATR10 and WATR20. The x axis shows the depth ranges of each moorings, while y axis presents time period from Jan 2010 to May 2023. Note that 0m follows directly after preceding mooring. The SST data are derived from the Regional Australian Multi-Sensor SST Analysis (RAMSSA) version 1.0.

250
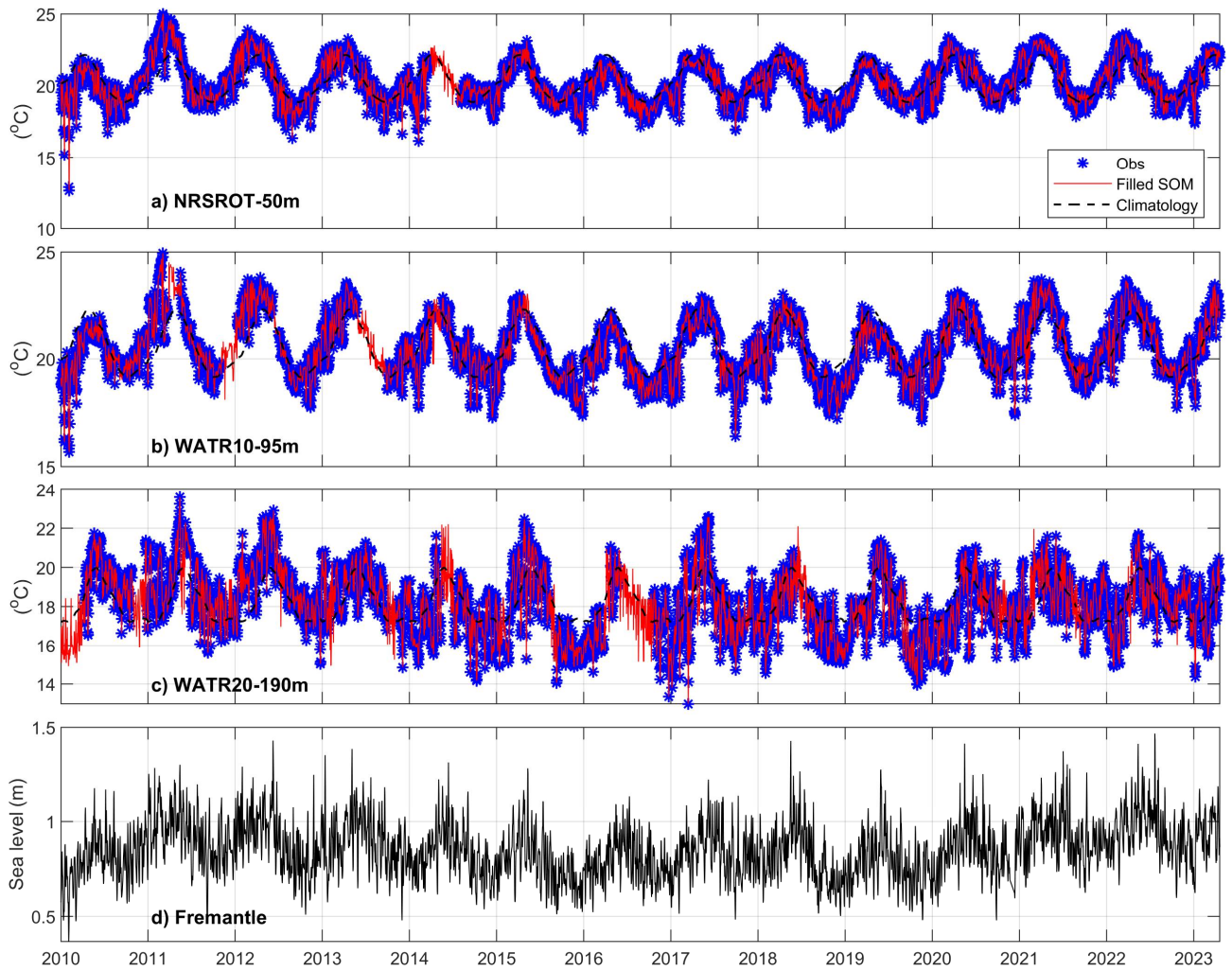
Figure 6. Comparison of observed and gap-filled temperature timeseries for a) NRSROT at 50m, b) WATR10 at 95m and c) WATR20 at 190m. The black dashed lines show daily climatological timeseries at corresponding depths. The climatological values are estimated from gap-filled data. The bottom panel shows Fremantle sea level timeseries.

251

252

253

254

255
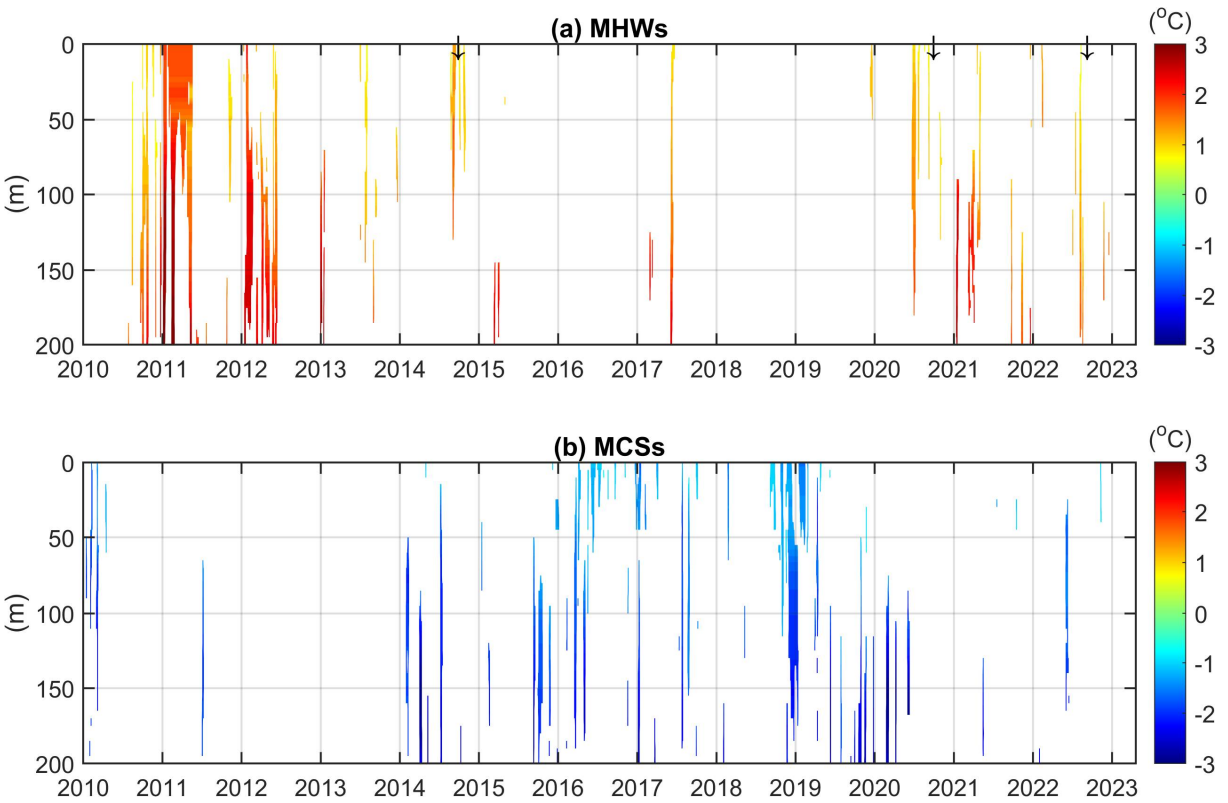
Figure 7. Mean intensity for individual (a) MHW and (b) MCS events at WATR20. Estimation is based on daily gap-filled temperature. The definition of each event follows Hobday et al. (2016). This plot is performed using MATLAB code (Zhao and Marin, 2019). The threshold temperature identifying a MHW or a MCS is set at the 90th and 10th percentile, respectively. Three arrows in (a) denote times of MHW events shown in Figure 8.

257

258

259

260

261

262

263
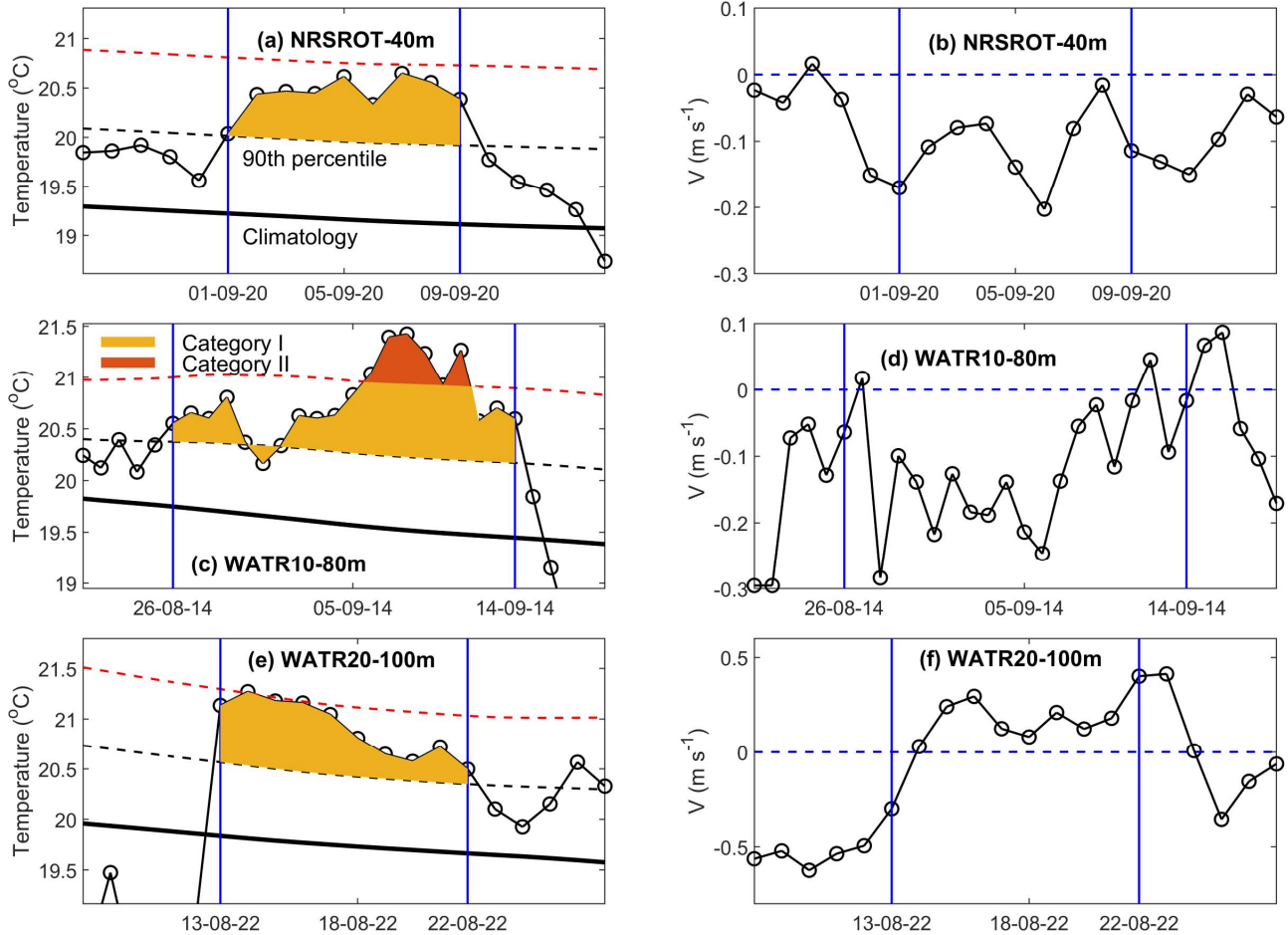
Figure 8. Left panels: Examples of marine heatwaves at NRSROT-40m (a), WATR10-80m (c), and WATR20-100m (e). Categories are moderate (yellow – category I) and strong (red – category II), as defined by Hobday et. al., 2018. In (a), (c), (e), the dashed red lines are estimated as twice the 90th percentile difference from the mean climatology value. Right 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.

264   Overall, the gap-filled velocity data are consistent with temporal periods of data gaps both at the mooring location and

265   the adjacent mooring sites (Figs. ~~S5~~S9 – ~~S8~~S11). The observed mean vertical profiles agree well with those derived from the

266   filled data (Fig. S~~11~~7), indicating that the SOM method accurately reconstructed the intricate vertical structure of the LC

267   system.

268   The LC flows along the shelf break, making velocities measured at WATR20 and WACA20 suitable for characterizing

269   its primary features. From the v-component data, the maximum mean currents recorded at WATR20 and WACA20 are -0.25

270   m s^{-1}, and -0.12 m s^{-1}, respectively (Fig. ~~S7~~S11d, b). Furthermore, the depths corresponding to these maximum values at the

271   two stations are 80m, and 100m, respectively. It can be inferred that the LC decelerates and deepens as it flows from WATR20

272   to WACA20. The irregular topography around the head of Perth Canyon may contribute to this disturbance (Fig. 1).

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 3). 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 3.

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	Days	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

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

283   -   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 13.5 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 0.61°C at NRSROT, 0.39°C at WATR10, and 0.58°C 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.

Since the strength of the LC is also influenced by local winds, we have evaluated the impact of including local winds during SOM training. Figure S12 presents the observed and reconstructed temperatures at the three moorings between January 1, 2020, and May 30, 2020, with local winds incorporated into the SOM training process. Compared to the case where local winds were excluded (Fig. 3), we found that including local winds resulted in a lower RMSE at NRSROT but a higher RMSE at WATR20. Overall, the differences were minimal. Local winds are important for the seasonal climatology of the Leeuwin Current, however, on interannual and intraseasonal time scales, the Leeuwin Current is more influenced by remotely forced coastal Kelvin waves, as reflected in coastal sea level variations (Feng et al. 2003; Marshall and Hendon, 2014). The effects of local winds may also have been integrated into the sea level variations.

SOM is an unsupervised learning method capable of capturing non-linear processes in the training data. Liu and Weisberg (2005) showed that the SOM method, unlike the linear empirical orthogonal functions (EOF), was able to reveal asymmetric features in the Florida Current system, such as variations in current strength and coastal jet location. However, as a statistical approach, it relies on enough realizations in the training dataset to properly capture the nonlinearity. In the Rottnest shelf region, several factors contribute to the non-linear variability in both temperature and velocity fields. Mesoscale eddies

can stem from the instability of the Leeuwin Current. Intense land-sea breezes during summer amplify near-inertial currents (Mihanović et al., 2016). Additionally, the strong shear zone between the Capes Current and the Leeuwin Current in summer, as well as interactions between the strengthening of the LC and the Perth Canyon in winter, can generate sub-mesoscale eddies (Cosoli et al., 2020). SOM may well capture the mesoscale processes in the LC. Due to their randomness, however, sub-mesoscale processes may not be fully captured in daily velocities. This is reflected in the lower R-squared values for velocities compared to temperatures (Figs 3, 4).

There are ~~systematic~~weak biases in the SOM-derived product, such as a warm bias at WATR20 during the validation period from January 1 to May 30, 2020 (Fig. S6c). ~~Notably,~~It is noted that this period ~~experienced~~coincided with multiple marine cold spells ~~the cold phase~~ (Fig. 7b). This systematic bias is likely due to the ~~smoothing~~nature of the SOM algorithm, which tends to underestimate the magnitude of extreme events while effectively capturing broader patterns. Future work could explore bias correction techniques to enhance accuracy.

Our continuous daily data products reveal that numerous MHW and MCS events occur sub-surface, which are undetectable while using altimetry data (Fig. 7). We also find that intense MHW events are frequently related to strong southward currents at the same depth (Fig. 8). However, the role of advection temperature due to shelf/slope LC current or warm-core eddies remains unclear. Future mooring observations are needed to better understand the characteristics 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 redundancy, the Iterative Completion Self-Organizing Maps (ITCOMPSOM) method outlined in Sloyan et al. (2023) could be more useful.

We have provided examples that highlight the advantages of using filled mooring data for end-users. The continuous daily temperature time series are essential for characterizing sub-surface marine heatwaves and cold spells on the Rottnest shelf, which can last from days to weeks. Furthermore, the gap-filled velocity time series from the mooring array allows researchers to capture episodical cross-shore and along-shore processes on the Rottnest shelf, offering valuable insights into the dynamics of the Leeuwin Current and Capes Current. These mooring data products, when combined with other observational platforms such as the IMOS glider program and surface radar observations, can be integrated into ocean-climate models to improve the accuracy of marine predictions for Western Australia.

7 Supplement

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

8 Author contributions

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

9 Competing interests

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

10 Disclaimer

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

11 Acknowledgements

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

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.

Cao, A.-Z., Li, B.-T., and Lv, X.-Q.: Extraction of internal tidal currents and reconstruction of full-depth tidal currents from mooring observations, Journal of Atmospheric and Oceanic Technology, 32, 1414-1424, 2015.

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.

Cosoli, S., Pattiaratchi, C., and Hetzel, Y.: High-frequency radar observations of surface circulation features along the south-western australian coast, Journal of Marine Science and Engineering, 8, 97, ARTN 97, 10.3390/jmse8020097, 2020.

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.

Hobday, A. J., Oliver, E. C., Gupta, A. S., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Holbrook, N. J., Moore, P. J., Thomsen, M. S., and Wernberg, T.: Categorizing and naming marine heatwaves, Oceanography, 31, 162-173, 2018.

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.: Patterns of ocean current variability on the West Florida Shelf using the self-organizing map, Journal of Geophysical Research: Oceans, 110, Artn C06003, 10.1029/2004jc002786, 2005.

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,

Marshall, A. G. and Hendon, H. H.: Impacts of the MJO in the Indian Ocean and on the Western Australian coast, Climate Dynamics, 42, 579-595, 10.1007/s00382-012-1643-2, 2014.

Mihanović, H., Pattiaratchi, C., and Verspecht, F.: Diurnal sea breezes force near-inertial waves along Rottnest continental shelf, southwestern Australia, Journal of physical oceanography, 46, 3487-3508, 2016.

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.

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.

Smale, D. A., Wernberg, T., Oliver, E. C., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L. V., Benthuysen, J. A., and Donat, M. G.: Marine heatwaves threaten global biodiversity and the provision of ecosystem services, Nature Climate Change, 9, 306-312, 2019.

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.