Object-Based Evaluation of Seasonal-to-Multiyear Marine Heatwave Predictions

Jacob T. Cohen[1], LuAnne Thompson[1], Elizabeth Maroon[2], Anna-Lena Deppenmeier[3], Cassia Cai[1]

[1]School of Oceanography, University of Washington, Seattle, WA, USA, [2]Department of Atmospheric and Oceanic Sciences, University of Wisconsin–Madison, Madison, WI, USA, [3]University of Liverpool, Liverpool, UK

Corresponding author: Jacob T. Cohen (jtcohen@uw.edu)

Key Points:

- Assessment of predictive skill for marine heatwave (MHW) location, area, and intensity is performed using hindcast simulations.
- Errors in all attributes of predicted MHWs grow with lead time, exceeding the skill of a random forecast for at least 12 months.
- Errors in the intensity of predicted MHWs depend on target month, with higher skill in December–January and lower skill in August–October.

Abstract

Accurate and interpretable marine heatwave (MHW) forecasts allow decision makers and industries to plan for and respond to extreme ocean temperature events. Recent work demonstrates skillful pointwise prediction of MHWs. Here, we evaluate a method of detecting and predicting spatially connected MHW objects. We apply object-based forecast verification to the CESM2 Seasonal-to-Multiyear Large Ensemble (SMYLE) experiment, a set of initialized hindcasts with 20-member ensembles of 24-month simulations initialized quarterly from 1970–2019. We demonstrate that SMYLE predicts MHWs that occur near observed MHWs with high skill at long lead times, but with errors in location, area, and intensity that grow with lead time. SMYLE exhibits improved skill in predicting the intensity of MHWs in December and January, and worse skill from August to October. This work illustrates the capacity to forecast connected MHW objects and to quantify the uncertainty in those forecasts with potential applications for future community use.

Plain Language Summary

User-friendly forecasts of marine heatwaves, exceptionally warm areas in the ocean, allow local decision makers and industries in coastal areas to plan for extreme sea surface temperatures. To interpret these forecasts, we evaluate how well the forecast model predicts the location, size, and temperature of past marine heatwave events. Here, we examine marine heatwaves as spatially connected events, or objects, instead of looking at individual locations independently. We find that marine heatwaves can be predicted several months in advance, but with errors in location, size, and temperature. These prediction errors get worse as we predict marine heatwaves further in advance. Finally, predictions of a marine heatwave's temperature are more accurate when the marine heatwave occurs in December or January, and less accurate when the heatwave occurs

between August and October. Spatial marine heatwave predictions could be used to inform marine resource management efforts and to communicate uncertainty in operational forecasts.

1 Introduction

Emissions over the past century have led to irreversible climate change. The combination of long-term warming and extreme sea surface temperature (SST) events, known as marine heatwaves (MHWs), is pushing ocean temperatures to increasingly extreme highs (Deser et al., 2024; Frölicher et al., 2018; Oliver et al., 2018). These events cause devastating effects on biological communities and marine ecosystems already experiencing stress from pollutants and microplastics (Jacox et al., 2020; Mills et al., 2013; Smale et al., 2019; Smith et al., 2023). MHWs also have socioeconomic impacts including fish farm mortality and the closure of commercial fisheries, with economic costs that can exceed US$800 million (Smith et al., 2021). Accurate MHW forecasts could allow local decision makers and industries to plan for and respond to these events (Hartog et al., 2023; Hobday et al., 2024).

Understanding the mechanisms that drive MHWs is necessary to improve predictive capabilities. Global analyses reveal that MHWs are driven by local atmospheric and oceanic variability, including changes in shortwave radiation, surface fluxes, advection, and Ekman transport (Bian et al., 2023, 2024; Marin et al., 2022).These drivers are often associated with large-scale modes of climate variability like the Indian Ocean Dipole, the Pacific Decadal Oscillation, and El Niño–Southern Oscillation (ENSO) (Capotondi et al., 2024; Gregory et al., 2024; Holbrook et al., 2019; Sen Gupta et al., 2020). Focused studies identify drivers of individual MHWs, including those in the Northeastern Pacific (Amaya et al., 2020; Fewings & Brown, 2019; Scannell et al., 2020; Schmeisser et al., 2019), the Southwestern Atlantic (Manta et al., 2018), the Tasman Sea

(Kajtar et al., 2022; Oliver et al., 2017), the Indian Ocean (Qi et al., 2022), and off the coast of Western Australia (Benthuysen et al., 2014; Pearce & Feng, 2013).

MHWs are commonly defined as SST anomalies relative to a long-term climatology that exceed the local 90th percentile threshold of SST variability (Hobday et al., 2016). For monthly data, SST anomalies above a 90th percentile threshold for any duration are considered MHWs (Capotondi et al., 2024; Scannell et al., 2020). While useful for standardizing MHW identification, this definition fails to capture the spatial coherence of MHWs, which are spatially connected regions, or objects, that can change shape and move across ocean basins. The spatiotemporal evolution of MHWs has been the focus of numerous studies (e.g., Bonino et al., 2023; Scannell et al., 2024; Sun, Jing, et al., 2023). Treating MHWs as objects instead of as points introduces new ways to quantify prediction skill and assess sources of predictability.

Analyses of ensemble hindcast systems from the North American Multi-Model Ensemble (NMME) and the European Centre for Medium-Range Weather Forecast (ECMWF) demonstrate long-lasting skill for MHW predictions globally (De Boisséson & Balmaseda, 2024; Jacox et al., 2022). In these studies, a correct prediction only counts when the predicted MHW occurs at the precise location of the observed MHW. If the MHW is predicted in the incorrect location, the event is counted as an error twice: once as a missed prediction where the MHW was observed and again as a false positive where it was predicted. To address this “double penalty effect” (Rossa et al., 2008), the numerical weather prediction community developed forecast verification techniques that go beyond pointwise metrics (Dorninger et al., 2018; Gilleland et al., 2009, 2010). Here we apply an object-based verification technique to hindcast predictions of MHWs to evaluate how well the model predicts the location, overlap, area, and intensity of spatially connected extreme SSTs.

2 Data and Methods

2.1 Data

We analyze the Seasonal-to-Multiyear Large Ensemble (SMYLE), a set of hindcast simulations run using the Community Earth Systems Model (CESM2) with a nominal 1° horizontal resolution for models of each component (Danabasoglu et al., 2020; Yeager et al., 2022). The initial conditions for SMYLE come from a forced-ocean-sea-ice model (SMYLE FOSI). SMYLE FOSI is run from 1958–2020 using the Parallel Ocean Program version 2 (POP2) and the Community Ice CodE 5.1.2 (CICE5) forced by historical atmospheric conditions given by the Japanese 55-year Reanalysis (JRA-55-do; Tsujino et al., 2018). SMYLE consists of 24-month long forecasts initialized on the first of every February, May, August, and November from 1970-2019. Each forecast has 20 ensemble members. These data have been used to study predictability across the climate system, including MHWs and ocean acidification (Mogen et al., 2024), ecosystem stressors (Mogen et al., 2023), and the North Atlantic Oscillation (Dunstone et al., 2023).

The initialization month is defined as the month in which the forecast was initialized and the lead month or lead time as the time since initialization. Following Jacox et al. (2022), lead month 0.5 refers to the predicted monthly mean of the initialization month. The target month is the month being predicted, which varies for different initialization months and lead months.

We examine monthly means of the top layer (5 m depth) ocean temperature from SMYLE, which we refer to as SST. We compare the SMYLE forecast data to observations from NOAA's Optimum Interpolation Sea Surface Temperature v2.1 (OISST) product (Huang et al., 2021), which provides daily SST values at 0.25° resolution. Following Jacox et al. (2022), we resample the data to monthly frequency to match the SMYLE forecast resolution. We regrid the SST from both SMYLE and OISST to a common 1° x 1° grid and mask data poleward of 65° to avoid small grid cells at high latitudes. We also exclude the Black Sea, the Baltic Sea, the Red Sea, and the Hudson Bay.

2.2 Defining Marine Heatwaves in Time and Space

We define the point locations of MHWs for each month before defining spatially connected MHW objects (Fig. 1). In OISST, we remove the monthly climatology and linear trend calculated from the 30-year reference period of 1989–2018 to define SST anomalies. We then calculate a 90th percentile threshold for each month and location, above which any SST anomaly is considered an MHW. In SMYLE, SST anomalies (Fig. 1b) are calculated by removing a lead-dependent (24-month) climatology and linear trend for each initialization month based on the 1989–2018 reference period (Jacox et al., 2022). We then calculate a lead-dependent 90th percentile threshold to define the MHWs at each grid cell (Fig. 1c).

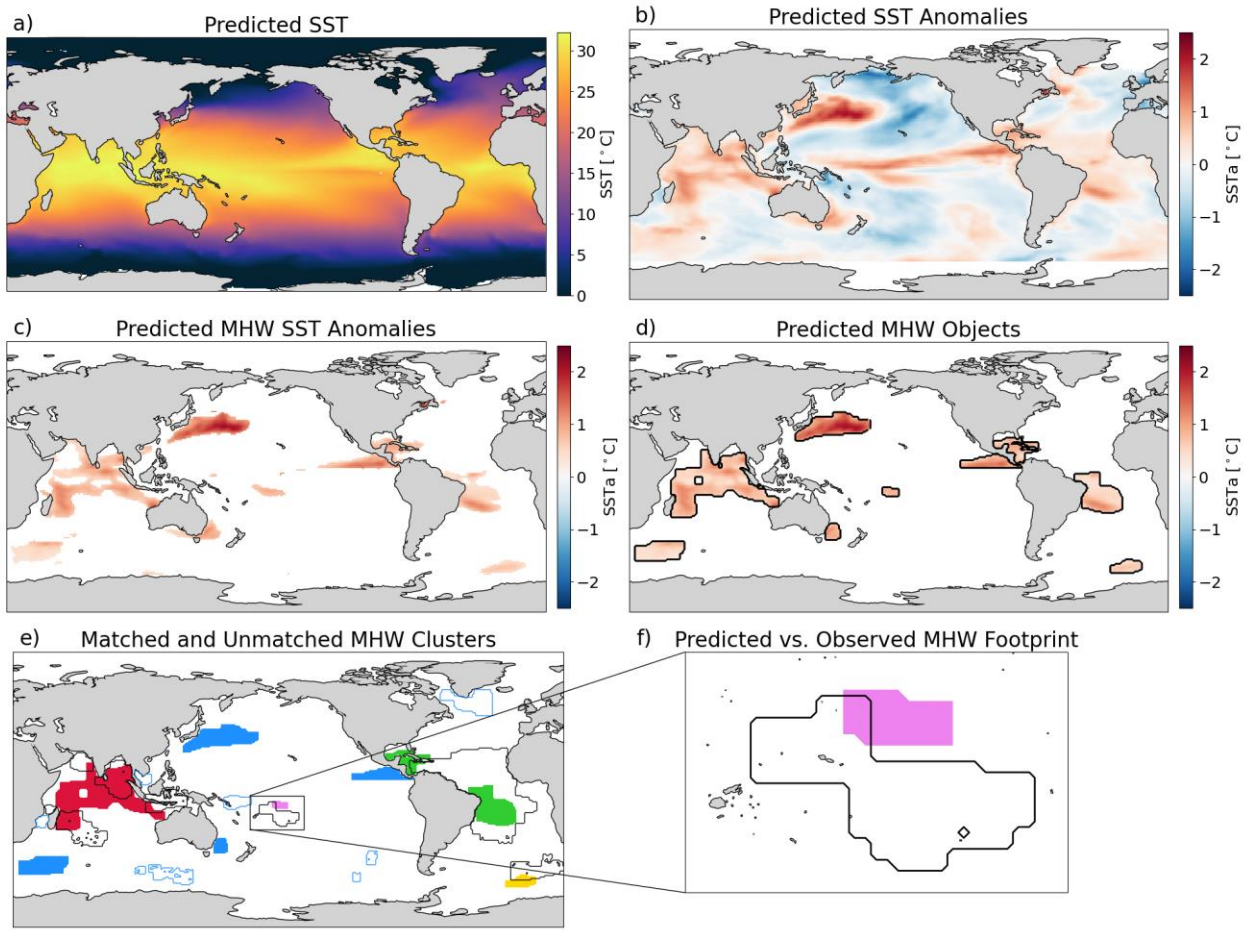


Figure 1. Identifying and matching MHW objects in an example forecast. In the first ensemble member prediction of May 2010 initialized on February 1, 2010 (3.5-month lead), we remove the climatology and trend from the (a) predicted SST to calculate (b) the anomalies. We apply the 90th percentile threshold to define (c) MHW point locations and use Ocetrac to define (d) the spatially connected predicted objects. (e) We match forecast clusters (red, pink, green, and yellow regions) with observed clusters (outlined contours) where an overlap exists. Where no overlap exists, forecast objects (blue regions) and observed objects (blue outlines) are unmatched. (f) For each matched forecast cluster, we consider the full forecast footprint (pink region) to be a correct prediction.

After defining MHWs at each ocean grid cell, we use Ocetrac, an MHW tracking tool that uses image processing techniques to smooth and label connected MHW grid cells (Scannell et al., 2024), to identify MHW objects from the MHW points in SMYLE and OISST (Fig. 1d; Text S1).

2.3 Verifying Marine Heatwave Forecasts

Object-based forecast verification requires a process to match objects between the forecast field (SMYLE) and the verification, or observation, field (OISST) to compare objects. We employ the Method for Object-based Diagnostic Evaluation (MODE; Davis et al., 2006a, 2006b) to perform object-based verification of MHW predictions. MODE has been used to analyze forecast skill of precipitation (Clark et al., 2014; Li et al., 2020), atmospheric rivers (DeHaan et al., 2021), drought (Abatan et al., 2018), cloud cover (Mittermaier & Bullock, 2013), and chlorophyll (Mittermaier et al., 2021).

MODE identifies, merges, and matches objects between the forecast and verification fields, and calculates attributes on simple objects, clusters, and cluster pairs (Text S1). MODE defines simple objects as individual objects in either field, clusters as groups of objects in one field that are paired to a cluster in the other field, and cluster pairs as sets of matched clusters in each field. We use the SST anomalies within the MHW objects identified by Ocetrac as the input field for MODE (Fig. 1d) and match a forecast object (SMYLE) and an observed object (OISST) at each timestep if they overlap in space (Fig. 1e).

2.4 Evaluating Forecast Skill

We evaluate forecast skill by assessing how often forecast MHWs are matched to observed MHWs and by examining the similarity of matched MHW clusters. For pointwise forecasts, contingency tables are used to measure how often a forecast is a hit (forecast yes, observed yes), a false alarm (forecast yes, observed no), a miss (forecast no, observed yes), or a correct negative (forecast no, observed no). In the object-based framework, we relax the constraint that a forecast MHW must occur at the same location as the observed MHW. Instead, we define the whole forecast MHW cluster to be a hit when the forecast cluster overlaps with an observed cluster (Fig. 1f). Thus, we define locations within matched forecast clusters as hits, locations within unmatched forecast objects as false alarms, locations within observed objects or clusters but not within matched forecast clusters as misses, and locations with no forecast or observed objects as correct negatives.

In the example in Figure 1f, a pointwise contingency table would determine that the pink region within the black contour is a hit, the pink region outside the black contour is a false alarm, the contoured white region is a miss, and everywhere else is a correct negative. The object-based contingency table, however, defines the entire pink region, whether a point is within the black contour or not, as a hit because it belongs to a matched forecast object.

We use the contingency table statistics to evaluate two metrics: the False Alarm Ratio (FAR) and the deterministic limit (T_{DL}). FAR quantifies the fraction of predicted events that do not occur and is calculated separately for each lead month:

$$\mathrm{FAR} = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}. \#(\mathbf{1})$$

Thus, higher values of FAR indicate weaker predictive skill. T_{DL} indicates the timescale at which MHWs can be reliably predicted. It is calculated as the latest lead time at which the number of hits (H) is greater than or equal to the number of misses and false alarms (X):

$$\boldsymbol{T_{DL}} = \mathbf{Lead}[\mathrm{H{=}X}]. \#(\mathbf{2})$$

We evaluate T_{DL} with both pointwise and object-based contingency table statistics.

These skill metrics convey how often SMYLE predicts MHWs but do not characterize how accurate the predicted MHWs are. We quantify this accuracy by calculating four pair attributes for each cluster pair: centroid distance, intersection over union, area ratio, and intensity ratio. The centroid distance of a cluster pair is the distance between the centroids of the forecast cluster and the observed cluster. The intersection over union is the overlapping area between the forecast cluster and the observed cluster divided by the area of the union of both clusters.

The area and intensity ratios are defined as the ratio of the areas or intensities of the forecast and observed MHW clusters belonging to a cluster pair, with the higher value being divided by the smaller value (Text S2, Fig. S1). For example, the area of the larger MHW cluster is divided by the area of the smaller MHW cluster regardless of whether the forecast MHW is larger or smaller than the observed MHW. The intensity ratio is similarly defined as the ratio of the median SST anomaly of the warmer MHW to the median SST anomaly of the cooler MHW. All statistics are calculated for the whole object cluster, not for individual objects within the cluster.

Perfect forecast skill is indicated by a centroid distance of zero (lower bound), an intersection over union of one (upper bound), an area ratio of one (lower bound), and an intensity ratio of one (lower bound). We average each attribute over every MHW cluster for each initialization month and lead time. Then, for each matched forecast MHW, we assign the value of its attributes to its spatial footprint and average over each initialization month and lead time to create spatial maps of each attribute. Mean values of ratio quantities are calculated as the geometric mean (Text S2), and all lead-dependent mean quantities are averaged over objects, not grid cells, so they are not area-weighted. All skill metrics are evaluated against the skill of a random forecast (Text S3). While Mogen et al. (2024) select the 97.5th percentile of skill scores from 1000 random forecasts, the computational expense of MODE limits our evaluation to a single random forecast. While this benchmark is less statistically robust, we expect that a larger sample size of random forecasts would yield qualitatively similar results.

3 Results

We evaluate the skill in predicting the presence of MHWs by evaluating the object-based FAR and by comparing the pointwise T_{DL} to the object-based T_{DL}. FAR is low at short lead times and increases globally with lead time, remaining lower (more skillful) than the random forecast FAR at most locations (Fig. 2a–d). At all lead times, FAR is lowest in the tropical Pacific Ocean, where the ENSO pattern of extreme SST anomalies is predicted better than the rest of the globe. Studies of pointwise MHW predictions also show higher skill in the tropical Pacific (Jacox et al., 2022; Mogen et al., 2024).

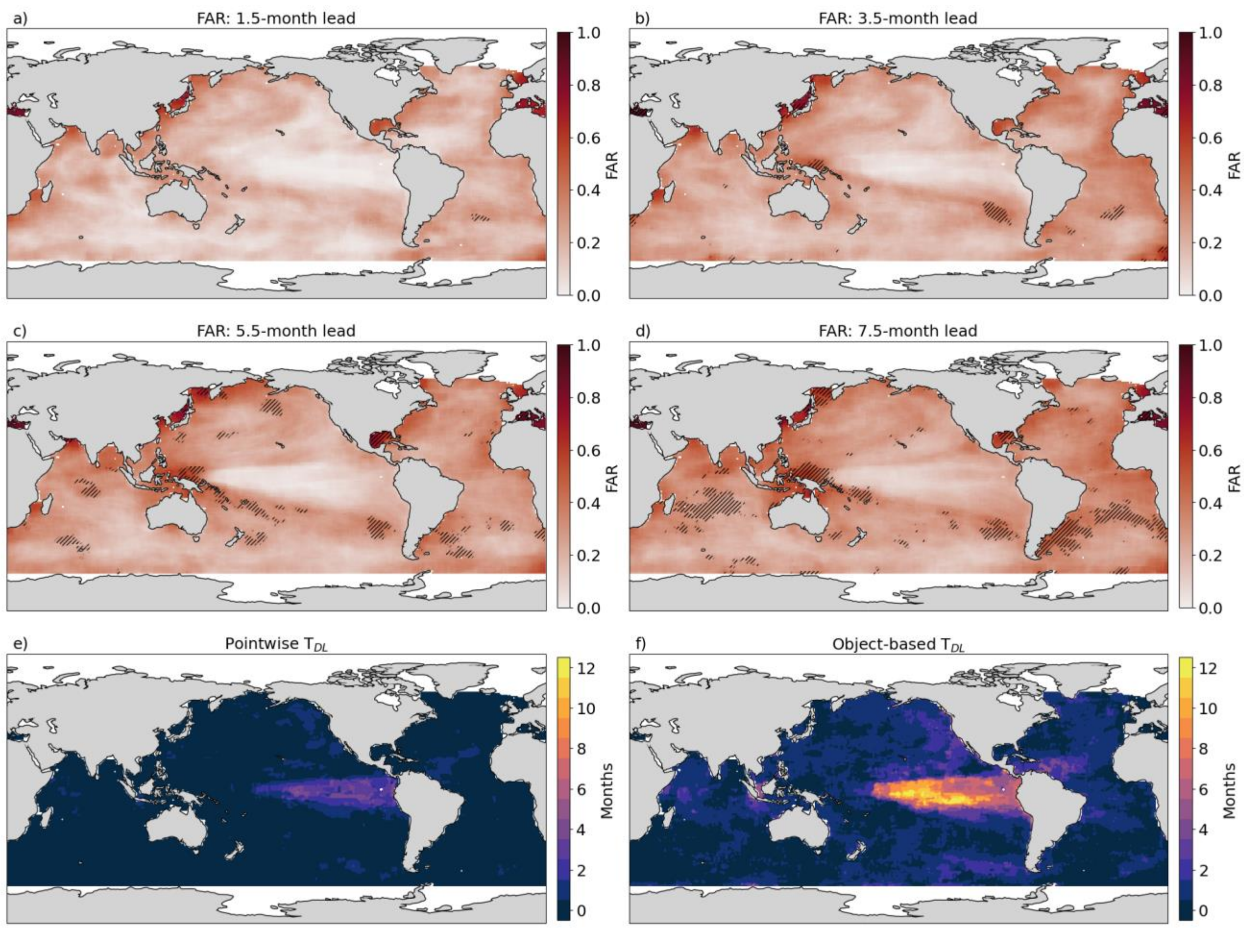
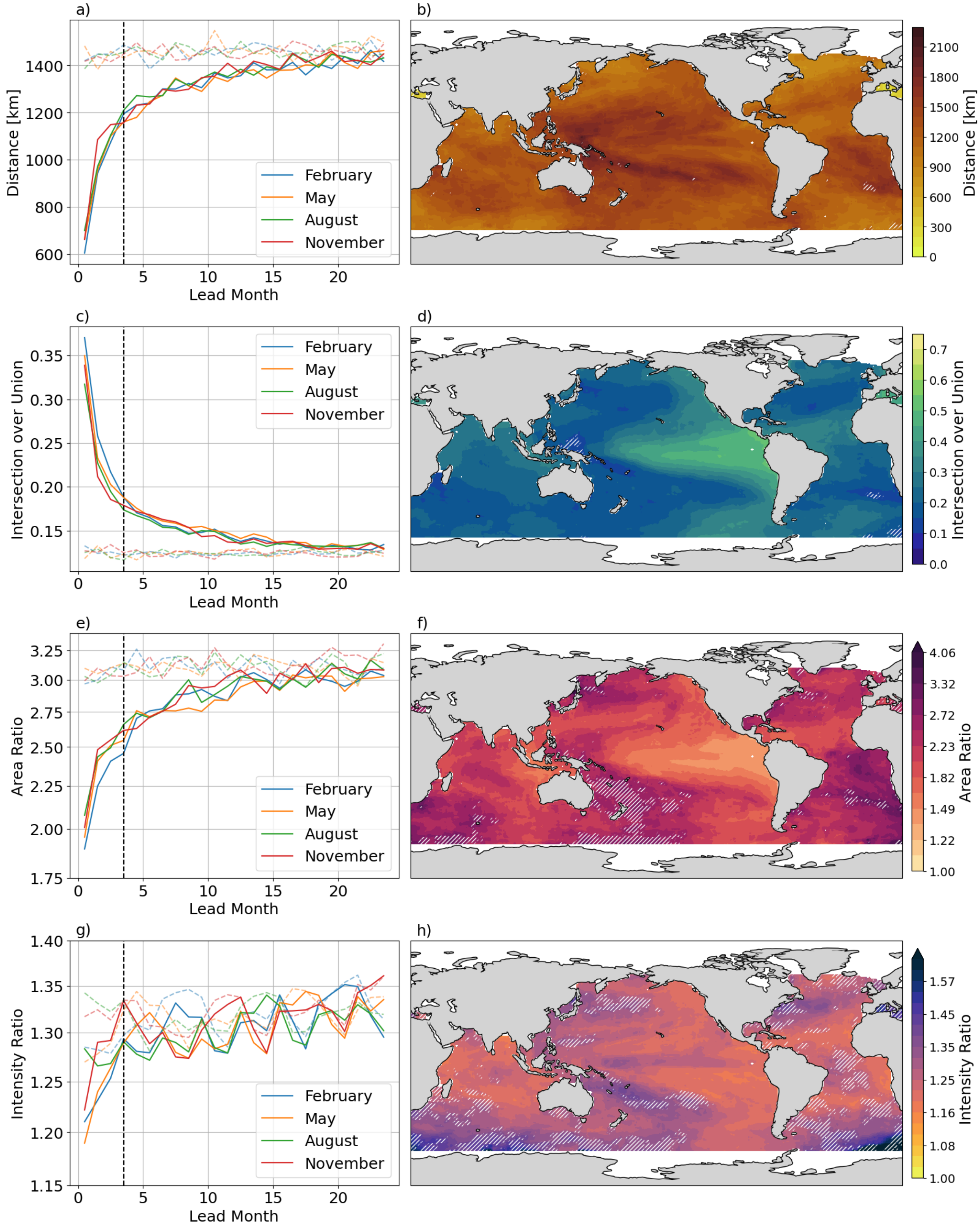


Figure 2. The false alarm ratio (FAR) and the deterministic limit (T_{DL}). (a–d) The 1.5-, 3.5-, 5.5-, and 7.5-month lead object-based FAR demonstrate that FAR increases globally with lead time. Hatched regions indicate FAR values greater than the random forecast's FAR values. While the pointwise T_{DL} (e) shows predictive skill past one month only in the tropical Pacific, the object-based T_{DL} (f) indicates multi-month skill globally.

The pointwise T_{DL} (Fig. 2e) indicates that SMYLE reliably predicts pointwise MHWs only in the first lead month, except in the tropical Pacific. The object-based T_{DL} (Fig. 2f), however, shows that MHW events are predicted to occur near observed events for one to two seasons globally, and up to a year in the tropical Pacific.

We next quantify the errors in forecast MHWs' location, overlap, area, and intensity to evaluate how well the predicted MHWs represent the observed MHWs to which they are matched. The mean forecast errors for each skill metric have similar overall spatial patterns at 3.5 lead months (Figs 3b, d, f, h). The lowest forecast errors occur in the tropical Pacific Ocean, suggesting that SMYLE not only predicts the presence of MHWs in this region (Fig. 2), but also predicts their locations, areas, and intensities well. Skill values are only averaged over matched forecast clusters, so interpretation of object-based skill metrics must incorporate the FAR and T_{DL} metrics. Because the verification data include a limited sample size of observed MHWs, other regions with especially high or low skill may be influenced by individual events that are especially well or poorly predicted. Evaluating drivers of individual MHWs could help determine why certain events are predicted better than others.

The centroid distance, intersection over union, and area ratio depend primarily on lead time (Fig. 3a, c, e). The intensity ratio depends on lead time for the first few months, after which it depends on the target month, indicating the importance of seasonal variability (Fig. 3g). The magnitude of each metric varies for MHWs of different sizes (Fig. S2) and depends slightly on the choice of Ocetrac radius (Fig. S3). The average matched forecast MHW has a centroid within 1400 km of the observed MHW's centroid, a 15–20 percent overlap with the observed MHW, an area less than 3 times smaller or larger than the observed MHW's area, and a median SST anomaly 1.25–1.35 times warmer or cooler than the observed MHW's median SST anomaly (Fig. 3a, c, e. g). All metrics exhibit high skill at short lead times and exceed the skill of random forecasts for the first 12 lead months.



237
238 **Figure 3.** Mean forecast errors in centroid distance (a, b), intersection over union (c, d), area
239 ratio (e, f), and intensity ratio (g, h). The left column (a, c, e, g) shows the mean error of each

metric averaged over all matched cluster pairs for each initialization month (colored lines) and lead time, while the right column (b, d, f, h) shows the mean errors for the 3.5-month lead predictions over all objects and initialization months, where lighter hues signify better skill. Dashed lines in the left column show the mean forecast errors of a random forecast, and hatched regions in the right column indicate skill worse than the random forecast.

We examine the intensity ratio by lead and target month to quantify its target-month dependence and seasonal variability (Fig. 4a). We then calculate the intensity ratio anomaly factor (Text S2), the factor by which the intensity ratio is greater than or less than the mean intensity ratio for a given lead month (Fig. 4b). Anomaly factors less than one (blue) mean the intensity ratio is more skillful and anomaly factors greater than one (red) mean the intensity ratio less skillful.

The intensity ratio for a given lead time is least skillful when the target month is between August and October, and most skillful when the target month is December or January. At lead times less than 4.5 months, initialization month plays an important role (Figs. 3g, 4b), after which the intensity ratio depends on the target month. The random forecast exhibits the same target-month improvements as the SMYLE predictions (crosses in Fig. 4b), indicating that the seasonal skill improvements are likely due to SMYLE's prediction skill of the overall SST distribution. Nevertheless, SMYLE's predictions of MHW intensity are more skillful than the random forecast's for the first 12 months (Fig. 4a). The target-month intensity ratio improvements are dominated by seasonal changes in the tropical and North Pacific, which display better skill in the boreal winter (when ENSO peaks) and worse skill in August–October (Fig. S4). This spatial skill pattern is consistent with previous results on ENSO-related prediction skill (e.g., Jacox et al., 2022; Shin and Newman, 2021).

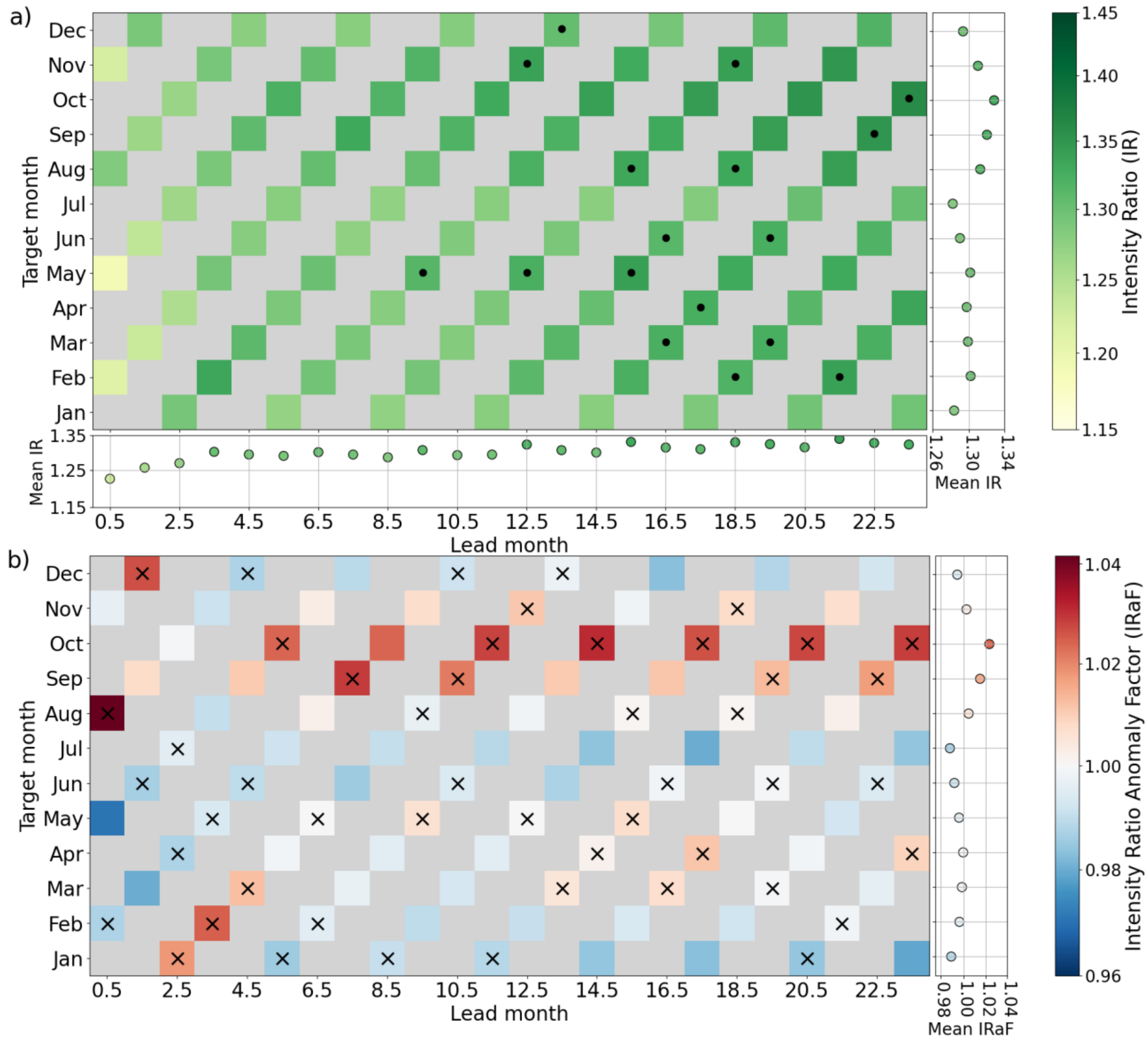


Figure 4. Dependence of the intensity ratio on lead month and target month. (a) The mean intensity ratio for each target month and lead month, with averages for each lead month shown along the bottom x-axis and averages for each target month shown along the right y-axis. Dots indicate intensity ratios greater than the random forecast's. (b) The same as in (a) but showing the intensity ratio anomaly factor relative to the lead-dependent mean. A negative anomaly factor (blue) represents better intensity ratios (closer to one) and a positive anomaly factor (red) represents worse intensity ratios (further from one). Crosses (X) indicate anomaly factors greater than the random forecast's. Panel (b) has no bottom scatter plot because the lead-dependent

mean intensity ratio anomaly factor is one at all lead months. Gray boxes in both panels are lead/target combinations with no data (for example, there is no 0.5-month prediction for June).

4 Conclusions

We demonstrate a novel evaluation method for verifying MHW predictions and quantify the skill in predicting the occurrence of MHWs and their associated attributes. By combining methods from both MHW detection and weather forecast verification we evaluate how well CESM2 SMYLE simulates and predicts MHWs. The results indicate that SMYLE forecasts effectively represent observed MHWs from OISST, extending previous evaluations of MHW predictions (De Boisséson & Balmaseda, 2024; Jacox et al., 2022) by predicting MHWs as objects instead of points. The method introduced here avoids incurring double penalties for the misplacement of predicted events and provides information about the types of errors that occur in MHW forecasts.

SMYLE demonstrates global seasonal to annual MHW prediction skill (Fig. 2). The error metrics examined here—centroid distance, intersection over union, area ratio, and intensity ratio—depend primarily on lead time and outperform random forecast skill for at least 12 lead months. Predictions of MHWs in December and January best predict MHW intensity, while predictions of MHW intensity between August and October perform worse, likely due to ENSO-related skill in the tropical and North Pacific.

A prediction with low error refers to a predicted MHW that has similar attributes to the observed MHW it is matched with. MHW predictions in locations and lead months with low error but high FAR values still exhibit weak predictive skill overall because the predicted MHW is unlikely to occur in the first place. Thus, forecast error metrics (Fig. 3) must be interpreted along with FAR (Fig. 2).

Both MODE and Ocetrac are designed to run on evenly spaced grids, unlike the native CESM grid or the interpolated 1° x 1° grid used here. This makes the presented approach problematic at high latitudes, so we constrained our analysis to latitudes less than 65°. MODE was also designed for regional grids instead of global grids, so it does not allow for periodic boundary conditions. This can cause some MHW events that span the southern coast of Africa or the Mediterranean Sea to be split into two events, which leads to more misses and less accurate predictions at the edges of the map (Figs. 2a, 3). We choose this longitude edge to minimize the impact of this aspect of MODE. Finally, we use permissive criteria for matching forecast and observed events: we require only that the objects overlap in space but do not require the objects to have similar areas or intensities. Moreover, while we allow for the separation of MHWs in space, we evaluate only the temporal co-occurrence of MHWs following other recent studies (e.g., Jacox et al., 2022; Mogen et al., 2024). A double penalty in time, when an MHW is predicted in the right place but the wrong time, could be addressed with temporal aggregation or object tracking (Clark et al., 2014).

Spatial forecasts have several potential applications for community use. Object-based definitions of MHWs and other ocean extremes could be used to improve our process-based understanding of the drivers and the predictability of extreme and compound events (Gruber et al., 2021; Mogen et al., 2024). Object-based seasonal forecasts may also be useful for open-ocean predictions relevant for dynamic ocean management (Maxwell et al., 2015) and living marine resource management, including monitoring and closures and annual catch limits (Tommasi et al., 2017). Spatial uncertainties and errors like those presented here must be incorporated for these operational applications. Spillman et al. (2025) identify the need for *useful* and *useable* MHW forecasts that present skillful information at relevant scales and with interpretable metrics.

Object-based predictions contribute to this goal by providing novel and intuitive ways to communicate MHW forecasts both to end users and to a general audience.

Acknowledgments

JTC, LT and CC were supported by NSF Award 2022874, EM was supported by NSF Award 2022740, and ALD was supported by NSF Award 2022842. JTC was additionally supported by the Graubard Fellowship in the Program on Climate Change at the University of Washington. We would like to acknowledge computing support from the Casper system (https://ncar.pub/casper) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. We thank Valentina Staneva and the UW eScience Institute for additional support.

Open Research

The data from the CESM2 SMYLE are available at: https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm2.smyle.html (Yeager, 2022). OISST v2.1 data (Huang et al., 2021) are available at NOAA/NCEI (https://www.ncdc.noaa.gov/oisst/optimum-interpolation-sea-surface-temperature-oisst-v21). The software for Ocetrac v0.1.5 is available at: https://github.com/ocetrac/ocetrac (Scannell et al., 2025). MODE is available through the METplus v6.0.0 package: https://github.com/dtcenter/METplus (Adriaansen et al., 2024). Model Evaluation Tools (MET) and METplus were developed at the National Center for Atmospheric Research (NCAR) through grants from the National Science Foundation (NSF), the National Oceanic and Atmospheric Administration (NOAA), and the United States Air Force (USAF). NCAR is sponsored by the United States National Science Foundation. The data and analysis scripts accompanying this manuscript are available on Zenodo (Cohen, 2025).

References

Abatan, A. A., Gutowski, W. J., Ammann, C. M., Kaatz, L., Brown, B. G., Buja, L., et al. (2018). Statistics of multi-year droughts from the method for object-based diagnostic evaluation. *International Journal of Climatology*, *38*(8), 3405–3420. https://doi.org/10.1002/joc.5512

Adriaansen, D., Biswas, M., Brown, B., Bullock, R., Burek, T., Fisher, H., et al. (2024). METplus Verification System Coordinated Release (Version v6.0.0) [Software]. Zenodo. https://doi.org/10.5281/ZENODO.10975507

Amaya, D. J., Miller, A. J., Xie, S.-P., & Kosaka, Y. (2020). Physical drivers of the summer 2019 North Pacific marine heatwave. *Nature Communications*, *11*(1), 1903. https://doi.org/10.1038/s41467-020-15820-w

Benthuysen, J., Feng, M., & Zhong, L. (2014). 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

Bian, C., Jing, Z., Wang, H., Wu, L., Chen, Z., Gan, B., & Yang, H. (2023). Oceanic mesoscale eddies as crucial drivers of global marine heatwaves. *Nature Communications*, *14*(1), 2970. https://doi.org/10.1038/s41467-023-38811-z

Bian, C., Jing, Z., Wang, H., & Wu, L. (2024). Scale-Dependent Drivers of Marine Heatwaves Globally. *Geophysical Research Letters*, *51*(3), e2023GL107306. https://doi.org/10.1029/2023GL107306

Bonino, G., Masina, S., Galimberti, G., & Moretti, M. (2023). Southern Europe and western Asian marine heatwaves (SEWA-MHWs): a dataset based on macroevents. *Earth System Science Data*, *15*(3), 1269–1285. https://doi.org/10.5194/essd-15-1269-2023

Capotondi, A., Rodrigues, R. R., Sen Gupta, A., Benthuysen, J. A., Deser, C., Frölicher, T. L., et al. (2024). A global overview of marine heatwaves in a changing climate. *Communications Earth & Environment*, *5*(1), 701. https://doi.org/10.1038/s43247-024-01806-9

Clark, A. J., Bullock, R. G., Jensen, T. L., Xue, M., & Kong, F. (2014). Application of Object-Based Time-Domain Diagnostics for Tracking Precipitation Systems in Convection-Allowing Models. *Weather and Forecasting*, *29*(3), 517–542. https://doi.org/10.1175/WAF-D-13-00098.1

Cohen, J. T. (2025). cohenjt/object-based-MHW-predictions: Object-based evaluation of seasonal-to-multiyear marine heatwave predictions (Version v1.0.0) [Software]. Zenodo. https://doi.org/10.5281/ZENODO.14758296

Danabasoglu, G., Lamarque, J. -F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916. https://doi.org/10.1029/2019MS001916

Davis, C., Brown, B., & Bullock, R. (2006a). Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Monthly Weather Review*, *134*(7), 1772–1784. https://doi.org/10.1175/MWR3145.1

Davis, C., Brown, B., & Bullock, R. (2006b). Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Monthly Weather Review*, *134*(7), 1785–1795. https://doi.org/10.1175/MWR3146.1

De Boisséson, E., & Balmaseda, M. A. (2024). Predictability of marine heatwaves: assessment based on the ECMWF seasonal forecast system. *Ocean Science*, *20*(1), 265–278. https://doi.org/10.5194/os-20-265-2024

DeHaan, L. L., Martin, A. C., Weihs, R. R., Delle Monache, L., & Ralph, F. M. (2021). Object-Based Verification of Atmospheric River Predictions in the Northeast Pacific. *Weather and Forecasting*, *36*(4), 1575–1587. https://doi.org/10.1175/WAF-D-20-0236.1

Deser, C., Phillips, A. S., Alexander, M. A., Amaya, D. J., Capotondi, A., Jacox, M. G., & Scott, J. D. (2024). Future changes in the Intensity and Duration of Marine Heat and Cold Waves: Insights from Coupled Model Initial-Condition Large Ensembles. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-23-0278.1

Dorninger, M., Gilleland, E., Casati, B., Mittermaier, M. P., Ebert, E. E., Brown, B. G., & Wilson, L. J. (2018). The Setup of the MesoVICT Project. *Bulletin of the American Meteorological Society*, *99*(9), 1887–1906. https://doi.org/10.1175/BAMS-D-17-0164.1

Dunstone, N., Smith, D. M., Hardiman, S. C., Hermanson, L., Ineson, S., Kay, G., et al. (2023). Skilful predictions of the Summer North Atlantic Oscillation. *Communications Earth & Environment*, *4*(1), 409. https://doi.org/10.1038/s43247-023-01063-2

Fewings, M. R., & Brown, K. S. (2019). Regional Structure in the Marine Heat Wave of Summer 2015 Off the Western United States. *Frontiers in Marine Science*, *6*, 564. https://doi.org/10.3389/fmars.2019.00564

Frölicher, T. L., Fischer, E. M., & Gruber, N. (2018). Marine heatwaves under global warming. *Nature*, *560*(7718), 360–364. https://doi.org/10.1038/s41586-018-0383-9

Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B., & Ebert, E. E. (2009). Intercomparison of Spatial Forecast Verification Methods. *Weather and Forecasting*, *24*(5), 1416–1430. https://doi.org/10.1175/2009WAF2222269.1

Gilleland, E., Ahijevych, D. A., Brown, B. G., & Ebert, E. E. (2010). Verifying Forecasts Spatially. *Bulletin of the American Meteorological Society*, *91*(10), 1365–1376. https://doi.org/10.1175/2010BAMS2819.1

Gregory, C. H., Artana, C., Lama, S., León-FonFay, D., Sala, J., Xiao, F., et al. (2024). Global Marine Heatwaves Under Different Flavors of ENSO. *Geophysical Research Letters*, *51*(20), e2024GL110399. https://doi.org/10.1029/2024GL110399

Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. *Nature*, *600*(7889), 395–407. https://doi.org/10.1038/s41586-021-03981-7

Hartog, J. R., Spillman, C. M., Smith, G., & Hobday, A. J. (2023). Forecasts of marine heatwaves for marine industries: Reducing risk, building resilience and enhancing management responses. *Deep Sea Research Part II: Topical Studies in Oceanography*, *209*, 105276. https://doi.org/10.1016/j.dsr2.2023.105276

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., et al. (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, *141*, 227–238. https://doi.org/10.1016/j.pocean.2015.12.014

Hobday, A. J., Spillman, C. M., Allnutt, J., Coleman, M., Bailleul, F., Blamey, L., et al. (2024). Forecasting a Summer of Extremes: Building Stakeholder Response Capacity to Marine Heatwaves. *Oceanography*. https://doi.org/10.5670/oceanog.2024.508

Holbrook, N. J., Scannell, H. A., Sen Gupta, A., Benthuysen, J. A., Feng, M., Oliver, E. C. J., et al. (2019). A global assessment of marine heatwaves and their drivers. *Nature Communications*, *10*(1), 2624. https://doi.org/10.1038/s41467-019-10206-z

Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., et al. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. *Journal of Climate*, *34*(8), 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1

Jacox, M. G., Alexander, M. A., Bograd, S. J., & Scott, J. D. (2020). Thermal displacement by marine heatwaves. *Nature*, *584*(7819), 82–86. https://doi.org/10.1038/s41586-020-2534-z

Jacox, M. G., Alexander, M. A., Amaya, D., Becker, E., Bograd, S. J., Brodie, S., et al. (2022). Global seasonal forecasts of marine heatwaves. *Nature*, *604*(7906), 486–490. https://doi.org/10.1038/s41586-022-04573-9

Kajtar, J. B., Bachman, S. D., Holbrook, N. J., & Pilo, G. S. (2022). Drivers, Dynamics, and Persistence of the 2017/2018 Tasman Sea Marine Heatwave. *Journal of Geophysical Research: Oceans*, *127*(8), e2022JC018931. https://doi.org/10.1029/2022JC018931

Li, L., Li, Y., & Li, Z. (2020). Object-based tracking of precipitation systems in western Canada: the importance of temporal resolution of source data. *Climate Dynamics*, *55*(9–10), 2421–2437. https://doi.org/10.1007/s00382-020-05388-y

Manta, G., De Mello, S., Trinchin, R., Badagian, J., & Barreiro, M. (2018). The 2017 Record Marine Heatwave in the Southwestern Atlantic Shelf. *Geophysical Research Letters*, *45*(22). https://doi.org/10.1029/2018GL081070

Marin, M., Feng, M., Bindoff, N. L., & Phillips, H. E. (2022). Local Drivers of Extreme Upper Ocean Marine Heatwaves Assessed Using a Global Ocean Circulation Model. *Frontiers in Climate*, *4*, 788390. https://doi.org/10.3389/fclim.2022.788390

Maxwell, S. M., Hazen, E. L., Lewison, R. L., Dunn, D. C., Bailey, H., Bograd, S. J., et al. (2015). Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Marine Policy*, *58*, 42–50. https://doi.org/10.1016/j.marpol.2015.03.014

Mills, K., Pershing, A., Brown, C., Chen, Y., Chiang, F.-S., Holland, D., et al. (2013). Fisheries Management in a Changing Climate: Lessons From the 2012 Ocean Heat Wave in the Northwest Atlantic. *Oceanography*, *26*(2). https://doi.org/10.5670/oceanog.2013.27

Mittermaier, M., & Bullock, R. (2013). Using MODE to explore the spatial and temporal characteristics of cloud cover forecasts from high-resolution NWP models. *Meteorological Applications*, *20*(2), 187–196. https://doi.org/10.1002/met.1393

Mittermaier, M., North, R., Maksymczuk, J., Pequignet, C., & Ford, D. (2021). Using feature-based verification methods to explore the spatial and temporal characteristics of the 2019 chlorophyll-<i>a</i>

bloom season in a model of the European Northwest Shelf. *Ocean Science*, *17*(6), 1527–1543. https://doi.org/10.5194/os-17-1527-2021

Mogen, S. C., Lovenduski, N. S., Yeager, S., Keppler, L., Sharp, J., Bograd, S. J., et al. (2023). Skillful Multi-Month Predictions of Ecosystem Stressors in the Surface and Subsurface Ocean. *Earth's Future*, *11*(11), e2023EF003605. https://doi.org/10.1029/2023EF003605

Mogen, S. C., Lovenduski, N. S., Yeager, S. G., Capotondi, A., Jacox, M. G., Bograd, S., et al. (2024). Multi-month forecasts of marine heatwaves and ocean acidification extremes. *Nature Geoscience*. https://doi.org/10.1038/s41561-024-01593-0

Oliver, E. C. J., Benthuysen, J. A., Bindoff, N. L., Hobday, A. J., Holbrook, N. J., Mundy, C. N., & Perkins-Kirkpatrick, S. E. (2017). The unprecedented 2015/16 Tasman Sea marine heatwave. *Nature Communications*, *8*(1), 16101. https://doi.org/10.1038/ncomms16101

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., et al. (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications*, *9*(1), 1324. https://doi.org/10.1038/s41467-018-03732-9

Pearce, A. F., & Feng, M. (2013). The rise and fall of the "marine heat wave" off Western Australia during the summer of 2010/2011. *Journal of Marine Systems*, *111–112*, 139–156. https://doi.org/10.1016/j.jmarsys.2012.10.009

Qi, R., Zhang, Y., Du, Y., & Feng, M. (2022). Characteristics and Drivers of Marine Heatwaves in the Western Equatorial Indian Ocean. *Journal of Geophysical Research: Oceans*, *127*(10), e2022JC018732. https://doi.org/10.1029/2022JC018732

Rossa, A., Nurmi, P., & Ebert, E. (2008). Overview of methods for the verification of quantitative precipitation forecasts. In S. Michaelides (Ed.), *Precipitation: Advances in Measurement, Estimation and Prediction* (pp. 419–452). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77655-0_16

Scannell, H. A., Johnson, G. C., Thompson, L., Lyman, J. M., & Riser, S. C. (2020). Subsurface Evolution and Persistence of Marine Heatwaves in the Northeast Pacific. *Geophysical Research Letters*, *47*(23), e2020GL090548. https://doi.org/10.1029/2020GL090548

Scannell, H. A., Cai, C., Thompson, L., Whitt, D. B., Gagne, D. J., & Abernathey, R. P. (2024). Spatiotemporal Evolution of Marine Heatwaves Globally. *Journal of Atmospheric and Oceanic Technology*. https://doi.org/10.1175/JTECH-D-23-0126.1

Scannell, H. A., Cai, C., Cohen, J. T., Staneva, V., Julius Busecke, Ryan Abernathey, et al. (2025, January 24). Ocetrac (Version v0.1.5) [Software]. Zenodo. https://doi.org/10.5281/ZENODO.14736769

Schmeisser, L., Bond, N. A., Siedlecki, S. A., & Ackerman, T. P. (2019). The Role of Clouds and Surface Heat Fluxes in the Maintenance of the 2013–2016 Northeast Pacific Marine Heatwave. *Journal of Geophysical Research: Atmospheres*, *124*(20), 10772–10783. https://doi.org/10.1029/2019JD030780

Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E., Alexander, L. V., et al. (2020). Drivers and impacts of the most extreme marine heatwave events. *Scientific Reports*, *10*(1), 19359. https://doi.org/10.1038/s41598-020-75445-3

Shin, S., & Newman, M. (2021). Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies. *Geophysical Research Letters*, *48*(10), e2020GL091886. https://doi.org/10.1029/2020GL091886

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., et al. (2019). Marine heatwaves threaten global biodiversity and the provision of ecosystem services. *Nature Climate Change*, *9*(4), 306–312. https://doi.org/10.1038/s41558-019-0412-1

Smith, K. E., Burrows, M. T., Hobday, A. J., Sen Gupta, A., Moore, P. J., Thomsen, M., et al. (2021). Socioeconomic impacts of marine heatwaves: Global issues and opportunities. *Science*, *374*(6566), eabj3593. https://doi.org/10.1126/science.abj3593

Smith, K. E., Burrows, M. T., Hobday, A. J., King, N. G., Moore, P. J., Sen Gupta, A., et al. (2023). Biological Impacts of Marine Heatwaves. *Annual Review of Marine Science*, *15*(1), 119–145. https://doi.org/10.1146/annurev-marine-032122-121437

Spillman, C. M., Hobday, A. J., Behrens, E., Feng, M., Capotondi, A., Cravatte, S., et al. (2025). What makes a marine heatwave forecast useable, useful and used? Progress in Oceanography, 234, 103464. https://doi.org/10.1016/j.pocean.2025.103464

Sun, D., Jing, Z., Li, F., & Wu, L. (2023). Characterizing global marine heatwaves under a spatio-temporal framework. *Progress in Oceanography*, *211*, 102947. https://doi.org/10.1016/j.pocean.2022.102947

Tommasi, D., Stock, C. A., Hobday, A. J., Methot, R., Kaplan, I. C., Eveson, J. P., et al. (2017). Managing living marine resources in a dynamic environment: The role of seasonal to decadal climate forecasts. *Progress in Oceanography*, *152*, 15–49. https://doi.org/10.1016/j.pocean.2016.12.011

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Yeager, S. G. (2022). Seasonal-to-Multiyear Large Ensemble (SMYLE) Experiment [Data set]. UCAR/NCAR - CISL - CDP. https://doi.org/10.26024/PWMA-RE41

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., et al. (2022). The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geoscientific Model Development*, *15*(16), 6451–6493. https://doi.org/10.5194/gmd-15-6451-2022